Marine Resources Assessment Group

for
Department for International Development
British Government

2003

Participatory Fisheries Stock Assessment Software

---

by Paul Medley

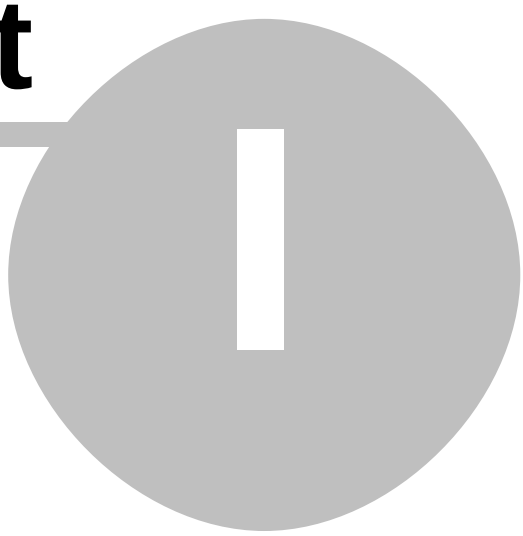
Version 1.0

Table of Contents

PFSA

Participatory Fisheries Stock Assessment
Software Manual

Part



1 Introduction

1.1 Welcome

The Participatory Fisheries Stock Assessment software supports a method for rapidly assessing fisheries, and making use of any available information. In particular the method encourages involvement of fishers in the assessment process, allowing them to influence decisions.

The software forms the analytical part of a methodology which includes interviews, fishing experiments and standard analyses involving time series data.

As with any analysis, the assessment is based on models, which in this case, are simple mathematical models describing the behaviour of the fishery. Models are simpler than the real world, although it is hoped that they capture the fishery's main behaviour and provide sensible advice.

Why use this method?

The approach has three distinct advantages over other stock assessment approaches.

1. You can involve the fishing community by using interview information. Even if these beliefs are unreliable, there is considerable political advantage in involving fishers in the assessment and they can see that their views are being taken into account. It is arguably necessary if co-management is being applied.
2. You can combine data from many sources, and in particular, you are able to use rapidly collected information and so may be used as a start point for an adaptive management system.
3. Combining sources also allows you to build up information for quite complex models. Breaking down complex models into simpler building blocks makes multispecies assessments easier.
4. The method applies decision analysis making use of utility (a measure of the desirability of an outcome) and risk to help in deciding management actions. This means the method can be used even when only a little information is available.

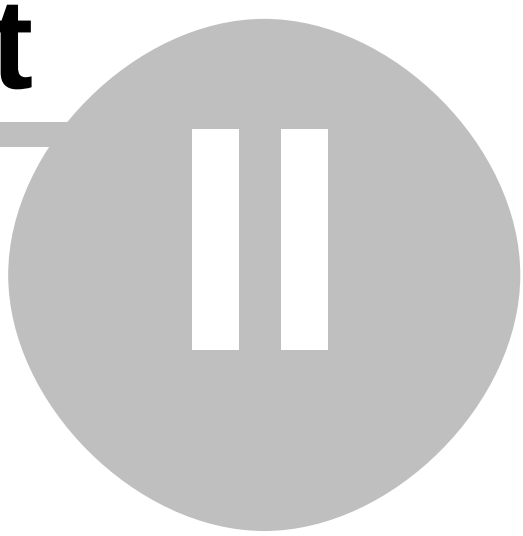
Getting Started

The methodology is not straightforward and it is unlikely you will be able to work the software without reviewing the manual. You can start by reading Getting Started and load a demonstration model. Otherwise make use of the help and manual, and read the technical sections as you work through the methods. Some familiarity with stock assessment and statistics is assumed.

PFSA

Participatory Fisheries Stock Assessment
Software Manual

Part



2 Software User Guide

There are four sections to the software. The main form connects the other three, and defines the simulation model which is used to generate the target and limit reference points. The probability form is the most complicated form as it implements all the probability modelling. The preference form deals with the fisher utility from a specific type of interview data. The management control forms defines how the fishery is controlled, the reference points and miscellaneous options for the simulations.

This is Version 1.0 of the software.

2.1 Getting Started

Before you are able to carry out any analysis, you must complete the following tasks.

1. You must define the simulation model. The analysis is based upon dynamic simulations of standard fisheries models. There are two models currently available which are robust enough to provide advice where little data is available.
2. You must define the number of gears and species in the fishery. The gears and gear names need to be listed. Remember you will need to provide at least some information on each gear. In general you should consider defining as few gear types as possible, combining similar gears where necessary. You will only be able to define species where your simulation model is the yield-per-recruit model. Again, you may not wish to include rare species or species of no commercial importance and limit each analysis to smaller species groups rather than trying to do all at once. More than one analysis may be carried out where species can be treated independently.
3. You must provide at least some information on each of the parameters in the simulation model. Information is provided in the form of parameter frequencies either provided directly or estimated from Monte Carlo simulations of models fitted to data. Developing the simulation parameter probability model is handled in the probability model editing form.
4. You must define how different outcomes should be scored. Ideally you collect information as preferences from fishers. Alternatively you can simply supply a price-cost ratio, which assumes a particular simple model. Preferences among different outcomes in the fishery are defined through the preferences editing form.
5. Finally you must define the type and values for fishery control variables. Stock assessment is only of limited value unless management is able to control the amount of fishing in some way. Control variables, such as fishing effort or quotas, define range of exploitation and what controls might be placed upon it. The types of controls available will depend on the simulation model. Only effort control can be usefully defined for yield-per-recruit. The types of controls, and the limits on these controls are defined in the controls editing form.
6. Once all the necessary information has been provided, you can carry out an analysis. The analysis runs simulations of the model to estimate two reference points to guide management. The reference points are a target, a control which best meets the economic and social needs of the fishers, and a limit which is more orientated to ecological needs and acceptable management risks. Several reference points may be obtained from different scenarios. Scenarios are designed to look at the stability of the results when using different data sources and controls.

2.2 Rapid Fishery Assessment

The following technique was developed to give rapid advice using fisher and other information suitable for small scale fishery assessments. The assessment should give an overall view on the state of stocks (what is and what is not known) and some advice on how management should proceed.

It is recommended that managers consider what policy they are trying to implement first. Important issues include the aims of the management and what management might use to reach those aims (mainly fishery controls). These assessments assume management wishes to protect the resource itself through a limit control and develop a target control which would most benefit the fishing community. The control can either be effort, catch quotas or closed areas. In all cases, it is assumed that managers will be able to implement and enforce these controls. This applies to all managers whether they are fishers or not.

Once a policy is decided, the primary aim should be to assess the fishery to see what needs to be done to achieve the policy objectives. It is assumed that a decision must be taken within the short term and not wait 5 or more years until enough scientific data have been obtained. This means gathering as much data as possible as quickly as possible and carrying out an assessment on these data.

Where there has been no previous data collection, it is recommended to focus on the logistic simulation model first. The logistic model is simple and captures the broad behaviour of a real biological system. It suggests that the biomass of the system should be kept above 50% of the mean unexploited biomass, which, without any detailed information, is a reasonable aim. Unless you have evidence that this model describes the dynamic behaviour of the fishery well, it is not likely that it will provide the optimum for the economics of the fishery. However, this is provides a reasonable start point and with more data better models might be found at a later date.

Data can be collected rapidly for the logistic model. The following is data is required for rapid assessment.

- Scaling data for the size of the fishery. This would define the fishery being assessed. Estimates of the total effort and catch within some relevant time period and the area being fished are the most useful information.
- Interview fishers to obtain a prior probability for the logistic model parameters and all gears. These data will need to be scaled up to the full fishery based on the estimate for the total effort.
- Interview fishers to obtain preference with respect to changes in the fishery catch and effort.
- Conduct a fishing experiment with the local fishers. This supplies information on current biomass and catchability and more deeply involves fishers in the assessment process.
- Close and then monitor the recovery of the fished area. This supplies information on productivity and the unexploited state.

These data will be adequate to provide assessment advice. In fact, even if only the interviews are carried out, some advice will be available. However, it should be emphasised any such advice is based on uncertain information, represented by the relevant probability distributions for the indicators. As such, this assessment should form part of an on-going adaptive management programme. Central to such a programme is the monitoring system (i.e. routine data collection) which should monitor the response of the fishery to management actions and measure the benefits and costs of the control.

In the same way as any other manager, fishers will need to make the final choice on what they plan to do in a co-management system. They will need to:

- Assimilate the assessment information, including the quantitative information and uncertainty. It is important fishers appreciate what they do not know and have a willingness to test their beliefs.
- Based on that assessment, they need to decide on the control they should apply and how long

they will re-apply it before a re-assessment. The time is important. A minimum time is required to test the result, and therefore the applied control should not be too onerous.

- They need to organise and co-operate with the monitoring. Monitoring need not just be limited to objective data (catch and effort), but could include personal attitudes why a control is or is not working well.
- Regular meetings should be held. In particular a re-assessment of the fishery should be conducted and it should be clear to fishers that they will have the opportunity to agree a different course of action.

2.3 Main Form

Access to all the method components can be found on the main form. The three main parts of the analysis can be accessed from main form either from the menu or through the buttons. However none of these are available until the analytical model has been defined.

When an action is not available it will be greyed out. If the analysis is unavailable, check that all necessary information has been provided as indicated by the checklist.

Model Type

You must first choose a simulation model type which will define both the parameters that require information, the available management controls and the relevant reference points. Two models are available: yield-per-recruit and the dynamic logistic model. You must then provide adequate information through the probability, preference and control forms to conduct a simulation.

Gear List

Next you must enter the gear names to be used in the analysis. You can add and delete the last gear in the list using the buttons (). To change the names, select the gear name and press F2. All names must be unique. As each gear type requires a parameter for each species, you should keep the numbers of gears as small as possible.

Species List

If you have chosen the yield-per-recruit model, you can enter more than one species and change the species names as for gear types (F2). All names must be unique. Each species has five population parameters and one parameter for each gear. The large number of parameters means that you can have no more than 10 species in any analysis for this version of the software.

Parameter List

The parameters for the current model with the specified numbers of gears and species are listed. Each parameter is given a standard name and description. You can edit the description. You can also edit the parameter name, but this is not recommended. The parameter names are standard for these models, and include subscript numbers for the gear and species where appropriate. It is worth reviewing all these parameters as you must provide at least some information on each one. The parameters are described in yield-per-recruit and dynamic logistic model.

Checklist

The current state of the model is recorded in the checklist. This indicates what information has been provided and is being used.

Probability Components refers to the probability models which must be defined for all parameters.

All Parameters Covered: All parameters in the model have a defined probability distribution.

Other probability distributions: Probabilities are listed, the check mark indicates whether they

are enabled (i.e. used).

Preference Components refers to the preference or utility score. There are two choices. A simple linear model describing the costs and benefits between relative changes in catch and effort. Alternatively, a more accurate quadratic model of fisher preferences using interview data.

Preference set: Whether the minimum preference information requirements have been provided.

Price-cost ratio: Whether the simple linear price-cost ratio is used.

Fisher preferences: The alternative model to the price cost ratio, which requires fisher interview data.

Default discount: A single overall default discount is used for all fishers.

Fisher discounts: Individual fisher discounts are used, obtained during interview.

Fisher importance: An importance score is used or not used to weight preferences among fishers. Fisher importance is usually defined as the importance of fishing to the fisher's income and the number of dependents they have.

Fisher preference data: Interview data is available for at least one fisher.

Control Components refer to the way the fishery will be controlled by management and also how the reference points will be defined. Unlike other assessment methods, the emphasis here is on management actions rather than status of stocks and the fishery. Hence, advice is given in terms of the chosen management action rather than more esoteric measures such as fishing mortality levels.

Controls set: Whether control variables have been set for the simulation. As well as a range of values over which to test the model, the levels of exploitation need to be set.

Effort, **quota** or **closed area**: The chosen management control is indicated.

Current control set: The current control has a value greater than zero.

Minimum/maximum Control set: the control value range has been set such that the maximum is greater than the minimum.

Current exploitation set: The current exploitation (i.e. effort) level needs to be set to allow comparisons with management changes. If effort is the control, this is automatically the current effort control level.

New Effort set: The new (maximum) effort level which the fishery will apply in future. This is only required if effort is not the control variable.

Scenarios

The results of the simulation are recorded in the scenario table. Each line in the table represents an analysis. Different analyses may be run with different data sources and assumptions. For example, different probability distributions may be enabled, fisher importance may be turned on and off, and so on.

A new scenario is created everytime you do an analysis in a new row at the end of the table and is automatically updated. You can supply your own descriptive name for each scenario in the first column and copy | paste the table into other software as text.

Menu

File

- **Open**: Opens a new stock assessment model.
- **Save**: Saves the current model and data to disk.

- **Save As**: Saves the current model and data to disk with a new name.
- **Close**: Closes the program without saving

Edit

- **Probability Model**: Displays the probability model edit form.
- **Preference Model**: Displays the preference model edit from.
- **Control Variables**: Displays the control variables edit form.
- **Save Scenario**: Adds a new scenario record to the table, and makes it the current scenario.
- **Delete Scenario**: Deletes the selected scenario record from the table.
- **Copy Scenario**: Copies the scenario table to the clipboard.

Projections

- **Analysis**: Carries out the simulations and analysis to obtain the limit and target reference points
- **Plot Projections**: Plots the simulated populations under the current level of effort. In most cases, the population moves to or remains at equilibrium.
- **Show Plots**: Displays the graphs form.

Help

- **Help**: Allows you to access this help file and information about the software.

2.4 Probability Models

The probability model edit form gives you access to the models and data used to model the probability distributions which encapsulate the risks in decision-making.

The model structure is displayed in the tree view. The model is displayed as the overall model and then a list of its component distributions. Each component distribution holds information on parameters as sets of parameter frequencies. These frequencies are treated as though they have been drawn from an underlying probability distribution. The probability distribution is then rebuilt based on the frequency using kernel smoothers. The individual probability distributions are combined to produce a posterior distribution for the model parameters, which summarises all information on what the parameters for this stock might be. The probability distributions are more properly called probability density functions or PDFs.

The joint PDFs can describe the probability for all the stock assessment model parameters simultaneously. This allows dependences (e.g. correlations and non-linear structures) between parameters to be modelled as well as the individual distributions. Each PDF need not have all parameters represented, however at least one PDF must have information on each parameter.

It is important that each PDF represented in the model is statistically independent. Independent data sources should be used in each model and not be used in more than one. More complex dependences among data may need to be modelled outside this software.

This version of the software implements two simulation models which can be defined in the main form.

The Schaefer or Logistic fish stock model. The model requires a minimum of 4 parameters: $Bnow$, r, $Binf$, and q. There will be a q parameter for each gear. The model describes changes in stock biomass over time, but does not differentiate the stock age or species. Instead, the fishable biomass is treated as a single unit.

The yield-per-recruit model describes the mortality and growth for single fish once recruited to the fishery. As well as q parameters for each gear, there are five population parameters for each

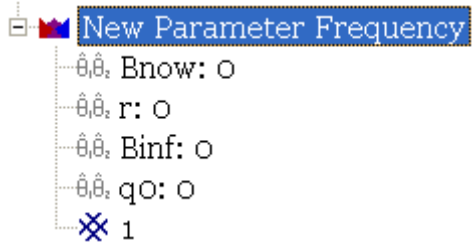
species: M, $a0$, K, $Winf$, $Wexp$. Unlike the biomass dynamics model, recruitment is not described but assumed to have a constant mean.

You can change the names of any of the models in the tree view for easier reference.

Probability Models

When you first open this form you will see a root for all models and the number of simulations (#). This represents the posterior PDF.

You must add frequencies to this to cover all parameters. Choose either New | Parameter Frequency or New | Parameter Frequency from Excel from the menu. In the former case you will need to select the parameters which will be included in this frequency. In the latter case, you should have selected the relevant data on the active worksheet in Microsoft Excel. The parameters and their mean values in the frequency are displayed. Until data are loaded, mean values are zero and there is only one data point (# 1). As long as there is no data on parameters they are disabled (greyed).



Once you have a frequency you can either load it with data directly or use a source model. For example, you might use Monte Carlo simulations (e.g. non-parametric bootstrap) fitting a model to catch and effort data in Excel using Solver then load the data from Excel. The set of bootstrap parameters can then be load as the new frequency using New | Parameter Frequency from Excel. Any parameter frequency can be loaded in this way, drawn from a known probability distribution, a likelihood (e.g. MCMC) or other source.

Source models allow you to generate parameter frequencies using common data and models. You can add a source model to any frequency by selecting it, then choose New | Source Model. The emphasis in this software is on robustness, so empirical rather than pure parametric methods are used to generate parameter frequencies. In particular, the non-parametric bootstrap is used as a proxy for model likelihoods. Alternative models and methods can be developed in other software, such as statistical or fisheries programs, and loaded into PFSA.

You will need to link the parameters between the parameter frequency and source model. In some cases this is done automatically, but you should ensure all such links are correct. You can drag and drop parameters between each other to build these links.

Once you have added a source model and provided it with the necessary data, you should Fit Model and Plot Observed Expected to ensure the source model is correct. Then Generate Parameter Frequency to provide the parameters to the parent parameter frequency model.

You can disable a parameter frequency, so that it is no longer used in deriving the posterior PDF. This allows you to see how different sources of information affect the results.

You can copy smoothing matrices between parameter frequencies by dragging and dropping the relevant parameter frequency icons ( ) with the target the one you want to copy to. The Recalculate Covariance Matrix and Fit Smoothing Parameter will be automatically turned off if you do this (see Edit Parameter Model).

There are a number of actions which load selected data from the active MS Excel spreadsheet. MS Excel must be running and the relevant data must be selected with the first row a name identifying the relevant parameters and the data in the column below. If the correct data is not selected, an error is raised and the operation is cancelled. You can also enter data directly into the software using data editing.

The following actions apply to the selected model in the tree view.

New

- **Frequency Data**: Creates a new parameter frequency. You must select the parameters in the frequency.
- **Source Model**: Creates a new source model to the selected parameter frequency.
- **Frequency Data Form Excel**: Creates a new parameter frequency. Data is automatically loaded from the selected area on the active MS Excel worksheet. The frequency data should be organised in columns, with the parameter name the first row in the column. Any incorrect spelling or incorrect selection will raise an error and cancel the operation.
- **Delete**: Deletes the selected model.

Model

- **Edit Data**: Brings up the edit data form for the selected model. Frequency data can be entered or changed in the edit view.
- **Edit Model**: Brings up the model edit form for the selected model. The model may be a source model or a parameter frequency model. In the latter case, you can view and edit the smoothing matrix for the frequency. It may be necessary manually to enter the smoothing matrix if there is too little data to fit it.
- **Fit Model**: Fits the selected model to the available data. The maximum likelihood fit results are displayed.
- **Generate Parameter Frequency**: This generates a new parameter frequency from the source model where appropriate. Parameter frequencies are generally produced through the non-parametric bootstrap method.
- **Update All Models**: Ensures all data is updated properly. Use this if the models do not appear to have reacted properly to changes you have made or errors have occurred.

Frequency

Dependent Probability

- **Add to Dependent Data List**: Indicates the selected parameter should be added to the list of parameters with observations, that is those with GLM data.
- **Make Parameter (In)dependent**: Switches the selected parameter from a independent (X) to dependent (Y) variable in the model and vice versa.
- **Add Parameter GLM Data from Excel**: Loads selected data from the active worksheet into the parameter observations. These data form the basis for linking known parameters to unknown parameters in the main parameter frequency (see Reading Dependent Probability Data from Excel).

- **Draw Posterior Sample**: With the top decision model selected (), program will draw a random sample of values from the combined enabled parameter frequencies. The size of the sample is set on the Management Controls form.

Excel

- **Add Frequency Data**: The parameter names must be exactly the same as for the assessment model being used (e.g. *Bnow*, *r*, *Binf*, and *q0*). The data is then loaded into an

additional probability model (see Reading Frequency Data from Excel).

- **Export Sample to Excel**: Exports the frequency to a new Microsoft Excel worksheet.
- **Add Smoothing Matrix**: Loads a smoothing matrix from Excel. The matrix must be square with the same number of rows and columns as the parameters in the frequency. The smoothing matrix is a covariance matrix, so the trace elements (diagonal) must be positive and the matrix must be symmetric. There should be no text in the selection.
- **Export Smoothing Matrix**: Exports the smoothing matrix to a new Microsoft Excel worksheet. Parameter names are also copied to the relevant columns.

- **Enable/Disable**: Enables or disables the selected parameter frequency. A disabled frequency is not used in drawing a posterior sample.

Population Model

- **Add Catch Effort Model**: Add standard catch-effort link models to the selected population model. A separate GLM will be added for each gear (catch effort data vector pair).
- **Add Population Index Model**: Adds a standard single parameter log-linear model to the population model(s). A separate GLM will be added for each species if a multispecies model is selected. You must supply the index as the Y variable.
- **Add Population Index GLM**: Add a population index link model to a selected single species population model. The model can be any support GLM.
- **Add Species**: Allows you to add a species to a multispecies population source model.

Excel

- **Add Catch and Effort Data**: Loads catch and effort data into the selected population model. The names of the columns do not matter, but the order of data is important. The data should be organise with effort data by gear followed by catch data by species then gear. The total catch for each species should be in first column of each species catch group (see Reading Catch and Effort Data from Excel).
- **Add Effort**: Loads effort data into the selected population model (see Reading Catch and Effort Data from Excel).
- **Add Catches**: Loads catch data into the selected population model (see Reading Catch and Effort Data from Excel).
- **Add Interview Data**: Loads interview data into the selected interview model (see Reading Interview Data from Excel).
- **Add Catch and Effort Model**: The action loads data from Excel as for Add Catch and Effort Data, then automatically sets up the catch and effort link model.
- **Add Index Model**: The action loads single column of survey data into a standard population index model (see Reading GLM Data from Excel).
- **Add Data to GLM**: Adds data from Excel to the selected GLM. The column names must match each corresponding model data vector or the operation will be cancelled.

Plot

- **Plot Kernel**: Plot the marginal data frequency and fitted kernel probability distribution. You can compare the data with the fitted PDF to check the fit is reasonable (see Getting Probability Distributions).
- **Plot Current Resource Status**: The plot shows the cumulative probability for the current resource status. In this application, status is the current estimated biomass as a proportion of the unexploited biomass. This is only available after successfully drawing a posterior sample.

- **Plot Observed Expected**: The plot shows the observed and expected values from a selected, fitted GLM. If the model is a catch-effort model, the observed and expected catches are displayed.

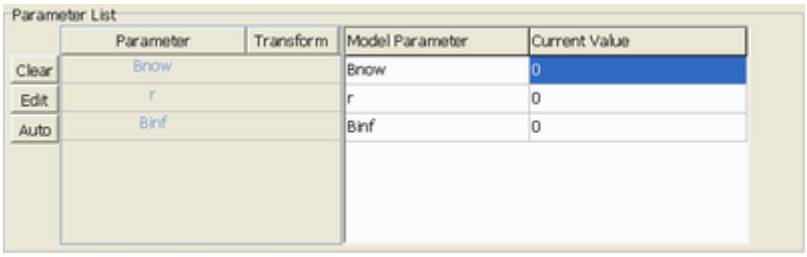
Help

- **Help**: Allows you to access this help file and information about the software.

2.4.1 Edit Source Model

After selecting New | Source Model, the source model edit form will be displayed. You must first of all choose the type of source model you wish to use. You will then be asked for defining information depending on the model chosen. In all cases you can change the name of the model and the accuracy. The accuracy governs the level of accuracy in the fit. If the fit is poor, you can try decreasing the value of the accuracy, although this will mean the model will take longer to fit.

You can also connect up parameters between the parent frequency and the source model. In some cases these links are made automatically, in which case you should make sure they are correct. The links are displayed as part of the parameter list. The linked parameters in the parent parameter frequency are displayed on the left. You can clear these, edit them, which brings up the Edit Parameter Links form, or auto link them which automatically links to parameters with the same name. You can also link parameters by dragging and dropping on the main model tree.



Once you have chosen a model and closed the model edit form, you will be able to open the edit form again, but not be able to change the model type. If you wish to do this, you must select the parent parameter frequency and then select New | Source Model again from the main menu.

Multispecies community model: The model allows you to fit several single species population models at the same time. This has the advantage of estimating the catchability parameters of several species simultaneously, so reducing their joint estimation error. The multispecies model is built from a statistical model describing species abundance. The only individual species population model currently supported is the linear depletion model, which is suitable for fishing experiments.

Single species model: The single species model describes how the population changes over time. Several models are available, unless the single species model is part of a multispecies model. The linear depletion and linear depletion with natural mortality describe population change with no recruitment. These models are most suitable for fishing experiments. The single species model owns its catch data, and if it is the source model rather than belonging to a multispecies model, owns its effort data too. There should be total catch data for each species, and effort data for each gear associated with separate catch data for each species-gear combination.

Generalised linear model: The generalised linear model (GLM) is the way all models are linked to data. They can be added to single species models or can occur independently if appropriate. You can add more than one link model to every population model. For example, you will probably need to fit catch and effort data, but you may also have stock survey data. This can also be included in a separate GLM.

Interview model: Estimates the parameters based on the stock assessment interview. This is most suitable for the logistic population model , although you might use it to supply priors for all catchability parameters. You cannot edit the interview model directly, but you can edit the interview data used to derive the parameters. For a technical description on how parameters are derived see Stock Assessment Interview.

Dependence probability model: This is another type of parameter frequency used specifically to build dependency between parameters. For example, you may have a set of growth and mortality parameters for a number of species closely related to the one you are interested in. You could just use this set as a prior probability. However you may also know the maximum length of your species and wish this knowledge to be impose upon the parameter set you have. This can be done by using the observed linear correlations, essentially by fitting linear models and estimating the growth parameter likelihoods dependent on the related species. Unlike the other source models, the parameters are estimated as model variables rather than parameters in models.

Dependence GLM source data: This marks the sub-set of parameters for which there are observations.

Dependence GLM: independent variable: Parameter data used as an independent variable. This parameter must be provided in the dependent probability parameter frequency.

Dependence GLM: dependent variable: Parameter data used as an dependent variable. This parameter is estimated in the dependent probability parameter frequency and need not be provided.

Parameters

All source models have parameters which can be linked to the kernel parameters.

Model Parameter: The parameter name is followed by the current fitted value.

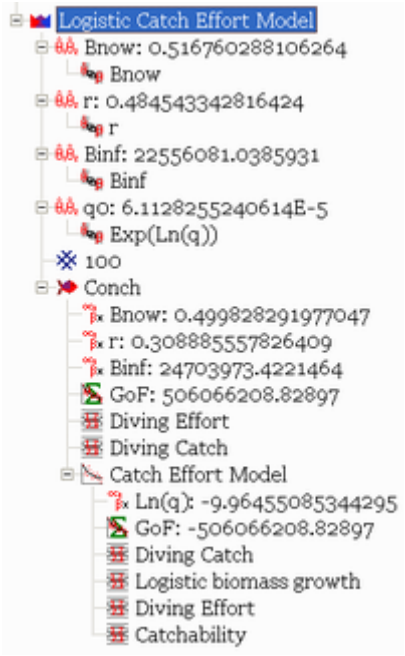
Parameter links: Indicates the name of the parameter which the PDF is linked to in the source model. Links may require a transformation in variable. In particular, catchability is estimated in the catch-effort models as a log-variable. The link can be transformed by selecting, then right-mouse-button click the parameter link and choose Exponent from the pop up menu.

Goodness of fit statistic: Where applicable, the goodness of fit statistic is displayed. It is based on the negative log-likelihood.

Data vector: These name the data vectors which the model uses or are available.

Parameters can be linked while editing or setting up the model or by dragging and dropping the parameters over each other. All parameters in the parameter frequency must be linked to the source model, except for the dependence probability model where linking is implicit.

Example

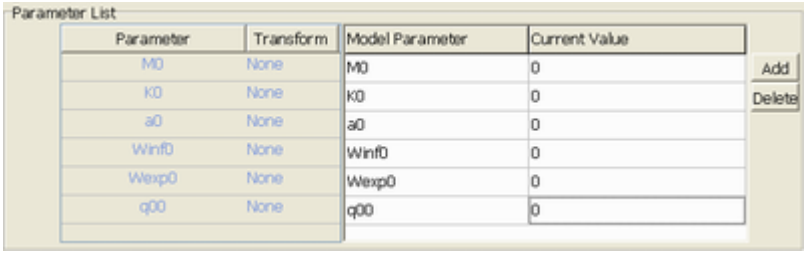


The "Logistic Catch Effort Model" parameter frequency has a single species model "Conch". The single species model is fitted to the catch and effort data through a catch effort model. This GLM fits Ln(q), so that the *q0* link must apply an exponent transform. Note that all the parameters are linked to the source model parameters. Additional population index models could be added to the population model if data were available. 100 parameter sets have been generated using "Generate Parameter Frequency".

2.4.1.1 Editing Dependent Probability Model

If you wish to build a dependent probability model, you must define the parameters in the model (see Dependent Probability Models). The basic parameters will be the same as for the parent parameter frequency. However, you can add (or delete) nuisance parameters that may be necessary to build your dependent relationships. For example, the maximum fish length may be a common variable of no direct use in the simulation model, but useful in building relationships between observed parameters amongst related species.

Nuisance parameters can be added and deleted from the parameter list. Core parameters from the parent parameter frequency cannot be changed. If you wish to add any nuisance parameters, they should be added at this stage. Nuisance parameters, such as length, are useful for building dependence between parameter estimates.

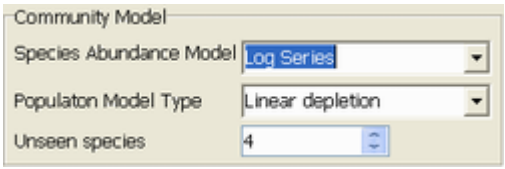


As for all other models, the edit form allows you to change the name of the model and its accuracy value.

2.4.1.2 Editing Multispecies Model

The multispecies community model fits several species simultaneously. Once you have chosen the community model, you must choose the species abundance model. In general, the geometric model is the best default model and should fit most species abundance data. If you choose the broken stick or log-series models, you can also specify the number of least abundant species (so-called veil-line). This will probably be necessary to fit the log-series model in particular, since the rarest species are unlikely to appear in catches. You can also fit the model as independent species, which would be the same as separate single species population models. See Multispecies Species Population Model for more technical information on these models.

The population model type is always the linear depletion model in this version of the software.



The parameters are the total population size of all species combined for the log-series and broken stick models, and the largest species population size and the proportional rank decrease for the geometric model. These parameters are not currently used in the simulation model. The advantage is that regardless of the number of species, only two parameters are required to estimate the initial population size.

2.4.1.3 Editing Single Species Model

You can choose between three population models. The linear depletion model has no recruitment or natural mortality. Catches are simply subtracted from an initial population size. It is most suitable for short term fishing experiments. An simple extension of this model allows for natural mortality, which may be more suitable over longer periods, although the natural mortality parameter may prove difficult to fit. Finally, the logistic difference model is provided, which is suitable for longer periods of catch and effort data, for example.

All single species models allow multiple gears, which default to the number of gears in the simulation model. You can add and delete gears for the source model independent of the simulation model. The number of gears has no affect on the number of parameters (which are incorporated in the link models) but they do affect the number of catch and effort data vectors which the model can reference. If you change the numbers of gears after adding catch-effort models, you will have to delete the relevant catch-effort models separately.

Each population model has its own set of parameters which are listed in the parameter box. These can be linked to the parameter fequency in the normal way.

2.4.1.4 Editing Generalized Linear Model

Generalized linear models are mainly used to link population models to observed data. In some cases, however, you may have data that allows you to estimate parameters directly. For example, fish count surveys might be used to estimate the current stock biomass (*Bnow*), which could be represented directly in a GLM format.

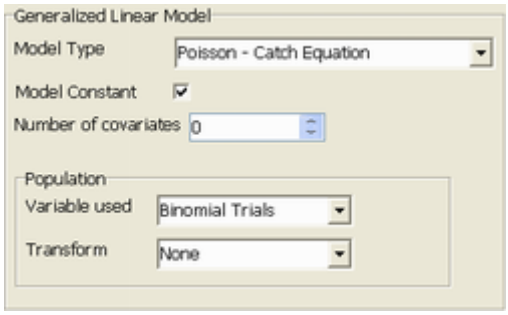
If you are unfamiliar with GLMs, it is recommended you only use the standard catch-effort and index models provided. They should be adequate in most cases.

In setting up a GLM, you must define the model type. This consists in defining the error model and the link function. Three types are available in this version. The normal linear model fits as a linear least regression. The Poisson log link is the standard log-linear model, where the effects are multiplicative. Finally, the Poisson catch-equation (also the link function is known as the complementary log-log link) fits a single gear catch equation. Multiple gear catch equation is not

available in this version of the software. As long as the fishing mortality estimates are not very high, separate single gear equations will give reasonably accurate estimates of catchability.

As well as the model type, you can choose to have a constant, which is estimated, and additional covariates. Each covariate and the constant will have a parameter associated with it, used to generate a linear predictor. The linear predictor is transformed by the link function to estimate the mean, which is fitted to the Y variate. You must supply a Y variate and the X covariate data for each model as required. Discrete factors are not supported in this version of the software.

If the model links a population model to observed data, you need to specify how the population size is used in the link model. It can appear either as an X covariate or as the "number of trials" parameter in the binomial model. The latter case is natural for catch effort models. You must also specify a transform. For example, if a log-linear model is being used, the variable may appear as a log-transform X covariate. See Generalized Linear Models for more technical information.



2.4.2 Edit Data

On choosing Model | Edit Data, the data for the selected model will be shown in a table. Data are organised into columns. Each column has headers describing the variate type (only relevant to GLMs), a variate name and a transform. Data is always shown in an untransformed state.

Absent data is shown as an empty cell.

Note that data may be shared between models, so only one copy need be edited. The data in the other models will be automatically updated. For example, catch and effort data are usually shared. If you edit the catch and effort data in a population model, all the generalized linear models which use that data will be automatically data.

Once you have finished editing the data, you must Save it. This will load the data into the relevant models. If you have not changed it or wish to discard changes press Cancel.

Frequency Data

Data from a parameter frequency model may take a long time to load if you are trying to look at frequency data with 1000 or more rows. You can edit the data. This is not generally recommended unless the data has been entered manually.

Variate Type	Frequency	Frequency	Frequency
Name	q00	q01	q02
Transform	Exponent	Exponent	Exponent
1	-5.82610978865899	-5.82334402859767	-5.82819087159123
2	-5.40621398909476	-5.38601383488616	-5.3281077810241
3	-5.85437321275175	-5.86153733456441	-5.76036092935865
4	-5.85400006105288	-5.86170032839212	-5.83736252493458
5	-5.82673029510062	-5.87351721403092	-5.83334695241251
6	-5.84081658764264	-5.842386795168	-5.81879566593935
7	-5.81760450277475	-5.82940434054563	-5.84594051417093
8	-5.85841472395223	-5.83341141993967	-5.83067162186813
9	-5.82072997902384	-5.83730008683555	-5.8391133546146
10	-5.87855320681762	-5.89049221718436	-5.86326177833538

If you wish to make many changes to frequency data, you may wish to consider Frequency | Export Sample to Excel, editing it in Excel, then reloading it to the parameter frequency using Excel | Add Frequency Data (see Reading Frequency Data from Excel).

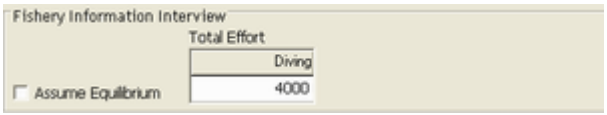
Catch and Effort Data

Catch and effort data is organised around population models. Multispecies models, if they are being used, hold the effort data, while the catch data can always be found by editing the individual single species models. Catches always start with the total catch, then the catch associated with each gear. In general, the total catch should be equal to or more than the sum of the individual gear catches. The total catch is used in the population model, where as the gear catches are only used in fitting catch-effort models. The following is an example single species model where there is no parent multispecies model.

Variate Type			
Name	Gear 0 Effort	Total Catch	Gear 0 Catch
Transform	None	None	None
1	50	2769.9938993909	2769.9938993909
2	50	2337.40894815913	2337.40894815913
3	50	2085.54488839427	2085.54488839427
4	34	1286.8893967261	1286.8893967261
5	96	2947.12854590523	2947.12854590523
6	50	1254.04640334748	1254.04640334748
7	27	517.355535034271	517.355535034271
8	55	1027.03872692135	1027.03872692135
9	50	835.410933919879	835.410933919879
10	22	296.160190158705	296.160190158705
11			

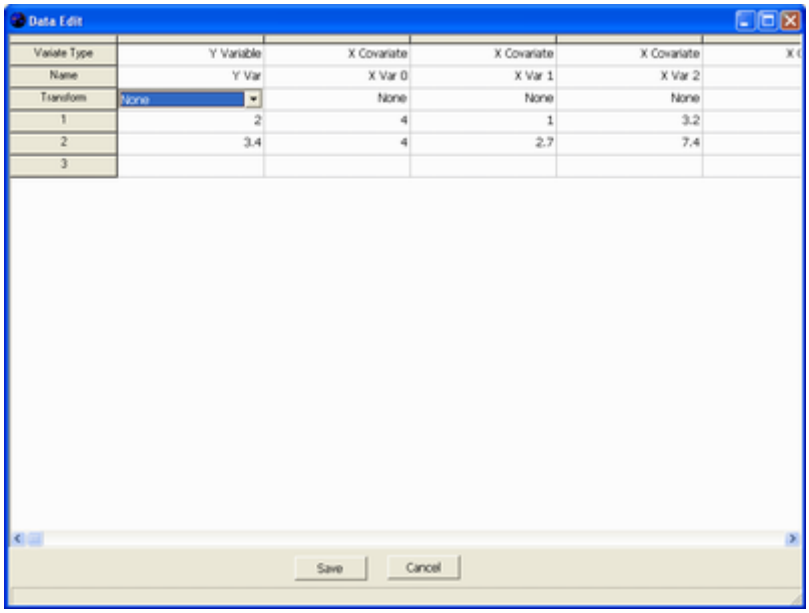
Stock Assessment Interview Data

As well as column data, the stock assessment interview model requires an estimate of the overall level of effort by gear when the interview was conducted. You should also record whether it is possible to assume the fishery is at equilibrium. In general, do not assume equilibrium in the calculations unless you are reasonably certain that catch rates have remained unchanged over a number of years.



Generalized Linear Models

You can edit generalized linear model data directly. In most cases this will be unnecessary. Unlike other data types, it is possible to change the names of the data vectors, transform the data and change the variable type. In this version of the software only one variable must be the dependent (Y) variable and all the others must be covariates (X variables). If you want to have an interaction term between covariates, you must enter their product (X1*X2) as a separate variable.



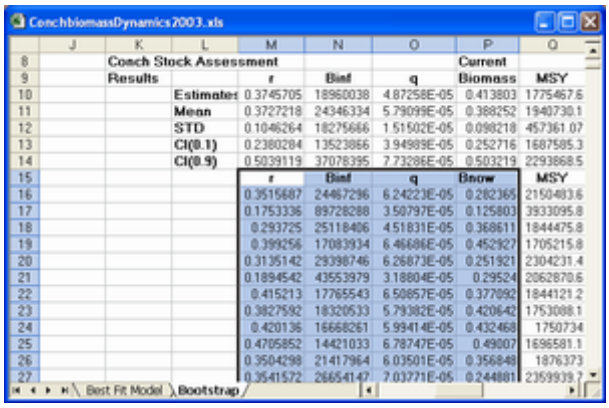
2.4.3 Loading Data From Excel

When loading data from Excel into any model, Excel must be open and the relevant data selected and on the active worksheet. Each selected column represents a data variable. The first row must contain the names of the data variables, with the actual data must lie below it.

The exception to this are parameter frequency smoothing matrices. These must a square area (columns=rows) with no text. The data read should be a covariance matrix, that is have a positive diagonal and be symmetric. You can Frequency | Excel | Export Smoothing Matrix and Frequency | Excel | Add Smoothing Matrix to move the estimated smoothing matrix to and from Excel.

2.4.3.1 Reading Frequency Data from Excel

The names heading each column must match exactly each parameter name in the selected parameter frequency model. Make the relevant worksheet the active sheet and select the names and the data. Choose Excel | Add Frequency Data from the menu, and the selected data will be loaded into the parameter frequency.



Important note: Be careful that the parameter values you are loading are compatible with the simulation model. The q parameter is from the catch equation. The current biomass is the biomass now, not the biomass from the start of the catch effort time series, for example. Also, you must be careful that the separate parameters are referencing the same fishery, that is the same area, set of boats and target species. It is easy to load parameters which are incompatible. If you do at best you will get an error message, at worst the analysis will be carried out but the results will be incorrect.

2.4.3.2 Reading Catch and Effort Data from Excel

You can order population models to load catch and effort data from an active Excel Worksheet. The data vectors must be in the correct order. The names heading each column need not match the data vector names in the selected population model. Essentially the first row is ignored. The column order should always start with effort, then the total catch and catches by species. There should be as many effort and catch variables as there are gears. There should be as many total catch / catch groups of variables as there are species.

For all models, the number of gears must be the same as that in the population model. If you have selected a multispecies model, the number of species must be the same as in the model. For example, there are two gears and three species catch and effort data in the following selection.

Gear 0 Effort	Gear 1 Effort	Gear 0 Total Catch	Species 0	Species 1	Species 2	Gear 1 Total Catch	Species 0	Species 1	Species 2
52	25	2769.99	1659.23	962.44	607.73	609.33	364.99	117.64	166.78
48	23	2337.41	1488.83	904.17	556.50	285.71	562.25	334.39	235.57
50	26	2085.54	1156.33	736.52	462.53	572.33	36.59	166.88	166.28
34	13	1286.89	716.56	431.19	233.85	485.99	101.27	78.04	18.93
96	47	2947.13	1669.24	1010.85	596.24	1089.95	109.02	153.89	102.94
50	25	1254.05	682.01	427.54	253.01	530.85	26.06	82.60	53.03
27	12	517.36	362.63	206.88	121.62	16.37	95.55	44.16	26.43
55	25	1027.04	639.33	409.66	225.13	232.71	136.16	157.42	54.36
50	25	835.41	500.87	285.21	202.14	300.33	115.89	56.67	90.40
22	9	296.16	211.82	142.25	62.39	41.86	71.29	69.36	11.51

Make the relevant worksheet the active sheet and select the names and the data. Choose Population Model | Excel | Add Catch and Effort Model from the menu, then the selected data will be loaded into the population model and standard catch-effort models will be automatically added for each gear and species. You can also choose Population Model | Excel | Add Catch and Effort Data from the menu which loads the data but does not then automatically add the models. Similarly ... Add Catches and ... Add Effort load catch and effort data separately.

2.4.3.3 Reading Interview Data from Excel

The data loaded into an interview model follows a standard format. See Stock Assessment Interview for more information on the interview questions. The names of the data variables (first row) are ignored, but the order of variables must follow the standard. The fisher name can be any identification label. The years fishing is not currently used, but might be used as a weighting index indicating experience in later versions. The main gear type for each fisher should be exactly the same name as a gear name used in the simulation model. The last year CPUE for the main gear, unexploited stock low and high CPUE and recovery period follow. This year CPUE for each gear in the fishery follows after that and should be in the same order as the gears in the simulation model. The following is example data for one gear.

Fisher Name	Years Fishing	Gear Type	Last Year CPUE	Virgin Low	High	Recovery Period	This Year CPUE
Christy Hall	40	Dive	1100	2500	3000	10	1400
Grovie Lockhart	10	Dive	500	500	600	1	500
Walter Moore	18	Dive	1000	1400	1500	1	1000
Alvin Parker	20	Dive	800	1500	1600	1	800
Henry Handfield	17	Dive	900	1500	2000	2	900
Marsha Pardee	11	Dive	2.5	4	7	10	2.5
Russell Jennings	20	Dive	200	450	550	1	200
Eroude Pierre	4	Dive	700	900	1000	1	600
Branford Hall	14	Dive	625	1000	1200	4	500
Harold Walkin	9	Dive	500	1500	1800	1	650
Daryl Hinson	15	Dive	900	1300	1700	1	600
Anthony Duncanson	10	Dive	600	1500	2000	1	600

After loading the data you will be requested to supply the current total effort for each gear in the fishery. This estimate should cover the entire fishery (including illegal activities). This is used as a scaling variable to raise the interview sample to total and mean values. You can also specify whether to assume equilibrium. In general, unless you are sure CPUE has not changed over the last several years, it is best not to assume the fishery is at equilibrium.

Although the total effort mightbe estimated from the interview data. It is better to get this information from other means (routine data collection), as interview data may be biased in this respect.

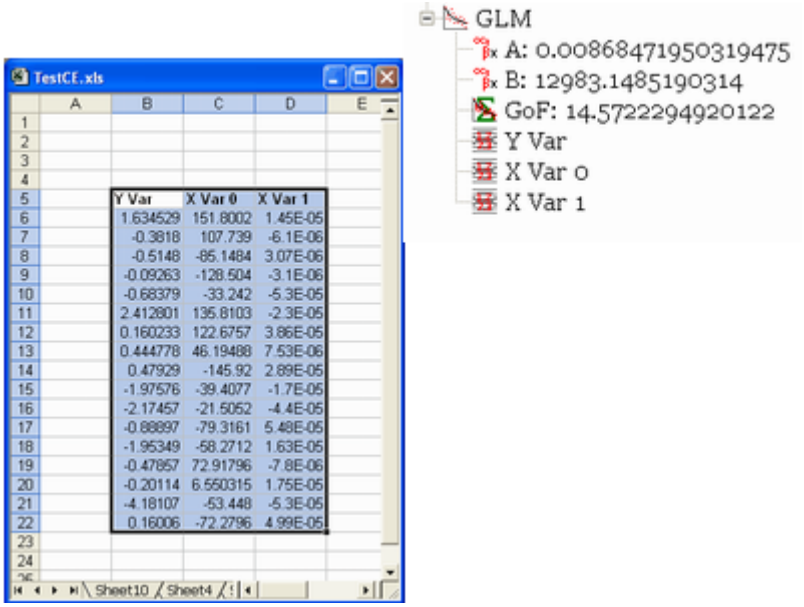
2.4.3.4 Reading Dependent Probability Data from Excel

Data for the dependent probability model can be read from Excel. These data are the independent and dependent variables used by the GLMs to estimate the dependent variables in the parameter frequency.

The names of the variables in the worksheet must match names of the relevant parameters in the dependent probability model, exactly the same as Reading Frequency Data from Excel. The GLM Data icon ( ) and the list of X variable icons ( ) will appear below the selected dependent probability model after these data are loaded.

2.4.3.5 Reading GLM Data from Excel

You can read in data to any GLM direct from Excel. The column names in Excel must match exactly the data vector names ( ). Rows with missing data will be removed from the fit. Catch and effort data should be added through the population model rather than the GLMs.



Y Var	X Var 0	X Var 1
1.634529	151.8002	1.45E-05
-0.3818	107.739	-6.1E-06
-0.5148	-85.1484	3.07E-06
-0.09263	-128.504	-3.1E-06
-0.68379	-33.242	-5.3E-05
2.412801	136.8103	-2.3E-05
0.160233	122.6757	3.86E-05
0.444778	46.19488	7.53E-06
0.47929	-145.92	2.89E-05
-1.97576	-39.4077	-1.7E-05
-2.17457	-21.5052	-4.4E-05
-0.88897	-79.3161	5.48E-05
-1.95349	-58.2712	1.63E-05
-0.47857	72.91796	-7.8E-06
-0.20114	6.550315	1.75E-05
-4.18107	-53.448	-5.3E-05
0.16006	-72.2796	4.99E-05

2.4.4 Edit Parameter Frequency

Model | Edit Model when a parameter frequency is selected will display the edit kernel form. You can change the amoothing matrix. As it is a covariance matrix, diagonal elements must be positive and the matrix must be symettical (cell[a,b]=cell[b,a]).

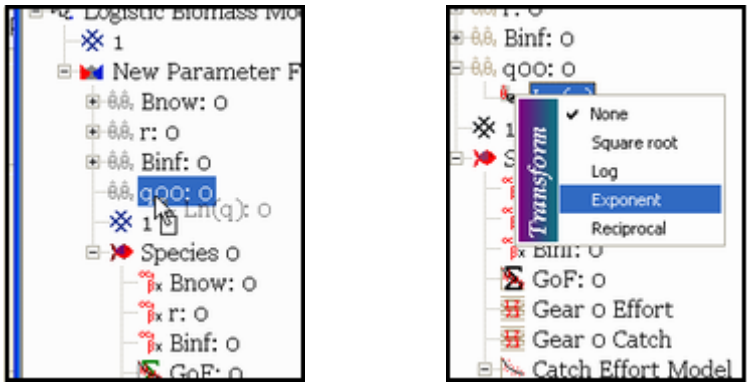
Supplying a covariance matrix is necessary if you have too little frequency data for the software to estimate the matrix. You can change and fix the matrix to particular values simply by editing the matrix. The correlations are set by calculating the covariance matrix from the data. If you wish to fix the correlations, make sure Recalculate covariance matrix is unchecked. The smoothing parameter scales the size of the matrix values. If you wish the software the absolute values supplied in the matrix, make sure Fit smoothing parameters is unchecked. Save to keep the changes, Cancel to discard them.

2.4.5 Edit Parameters

This dialog box is displayed when you choose New | Parameter Frequency. You can choose the parameters which will appear in the parameter frequency. Click on those parameters which you wish to select from all the available parameters in the list. Clicking on the tick ☑ selects all parameters, and on the cross ☒ deselects all parameters. Press OK to proceed, Cancel to stop the whole operation.

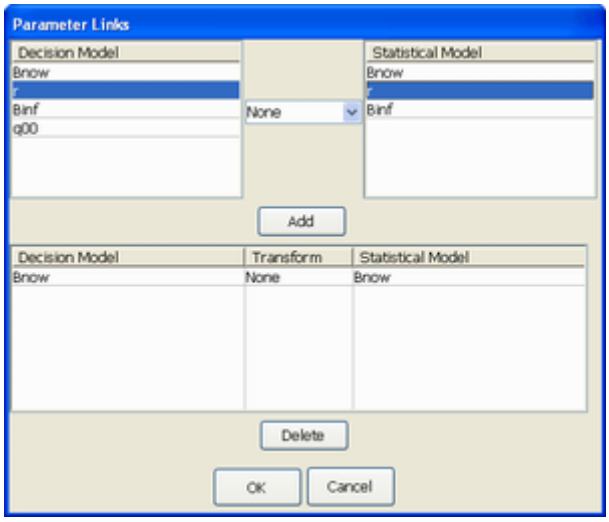
2.4.6 Edit Parameter Links

It is necessary to link source model parameters to the frequency model parameters so that parameters generated by the source model can be loaded into the parameter frequency. These links can be made in several ways. The easiest is to drag and drop parameters between the source model and its parent parameter frequency. You may then need to change the transform, as is required for the catch-effort model connecting q00 to Ln(q). Select the link then click the right mouse button to get the transform pop-up menu.



Source Model Edit

If you press Auto in the Parameters section of the Source Model edit form all parameters sharing the same name will be linked automatically. Clear will remove all links. If you press the Edit button, the link dialog will be displayed.



You should select the decision and statistical model parameters you wish to link and choose the appropriate transform. Press add and the linked pair will be added to the linked list below. Select one of the linked parameter pairs and press delete and the link will be removed. Press OK and the changes will be saved, Cancel and the changes will be discarded.

2.4.7 Building Priors

An important component of Bayesian statistics, which forms the foundation of the method described here, are prior probability distributions. Priors are the belief, including uncertainty, that you start with and update with scientific observations. There has been an on-going debate over priors in science as priors introduce subjectivity which science generally shuns. As a result, scientists have focused on finding uninformative priors which do not influence final results.

In this methodology, priors are seen as a benefit rather than a nuisance. They allow the stock assessment scientist to start the assessment process immediately and do not require a long wait before any advice can be given. However, this requires a reasoned approach to building informative priors, and care should be taken that they do not overwhelm the results when other data are available.

The most obvious source of priors are the fishers themselves. It is recommended that interview data is used for the logistic simulation model. With reasonable differences in opinion, interviews

still allow significant uncertainty, but involve fishers in the results. If you can show they have influenced the results, so that their opinions are demonstrably taken into account, they should be more likely to accept the final recommendations.

Many parameters in the yield-per-recruit models are not suitable for interview. It may be possible to develop methods in future to generate views on natural mortality and maximum size and so on, but these parameters are so far removed from fishers everyday experience that such priors may not informative. An obvious additional source of information would the experience of scientists in other countries with similar species.

Fishbase (www.fishbase.org) is a database of information on many fish species and in particular is a source of parameter estimates for growth and mortality parameters. These parameters can be easily downloaded and copied into a spreadsheet. Although the reliability of parameters may be questionable in many cases, they are probably a good way to build a prior probability which allow you to conduct a YPR analyses.

There is no standard way to do this, but the following techniques are suggested.

- If the species has many independent estimates for its parameters, you should load these directly into the frequency. The smoothed probability distribution would probably represent a reasonable prior as long as the estimates cover the range of environmental and ecological characteristics which apply to your fishery.
- You can build estimates of natural mortality using Pauly's (1980) empirical equation for each set of growth parameters and your fisheries mean annual water surface temperature (see Sparre and Venema 1992). As the regression is based on log-values, you should not find that resulting smoothing matrix is singular even though it is based on a linear regression. The equation has the form:

$$\ln M = -0.0152 - 0.279 * \ln L_\infty + 0.6543 * \ln K + 0.463 * \ln T$$

  where M and K are the mortality and growth rates (year^{-1}), L_∞ is asymptotic length (cm) and T is the mean annual water surface temperature (^{0}C).

- Winf can be found from L_∞ using weight-length conversion estimates. The weight-length conversion parameters are also available from Fishbase, but you can sample to get your own relatively easily. If you only use, for example, bootstrapped estimates of the weight exponent (Wexp), the model will use an implicit uninformative prior (uniform on 2.5-3.5). Alternatively, if you are using Fishbase estimates that are independent of the growth parameter estimates, you can sample them randomly to do the conversion.
- The age at recruitment can be found from information on the smallest fish in the catches, converted to age using the inverse von Bertalanffy growth equation for each set of growth parameters. Given a reasonable sample of the catch, it would be advisable to use a lower percentile (e.g. 5 or 10%) rather than the smallest individual as the mean size of recruitment.
- If the parameter set are too heavily correlated (e.g. if the smoothing matrix is close to singular an error will occur) even if you have a large number of observations, you could add small random numbers drawn from a normal distribution to the problem parameter which would reduce the correlation. For example, using Pauly's empirical equation ignores the observation errors in the equation's parameter estimates. It would be quite legitimate to add this error back in.
- The number of parameter estimates for many species will be too small to estimate the smoothing matrix. For these species you might consider using all similar species as a group. For example, species belonging to the same genus might be expected to have similar parameter estimates. There are three approaches to using this information.
  1. Simply use all similar species estimates as the prior. Many species would then share the same prior.
  2. Use all species estimates combined to estimate the smoothing matrix, then copy this matrix to those species with few frequency values using "drag and drop" on the probability form, or use Frequency | Excel | Export Smoothing Matrix and Frequency | Excel | Add Smoothing Matrix. The probability distribution would use the actual values as the mean, but spread the probability around these values using the overall smoothing matrix. You can also scale this matrix up to indicate greater uncertainty for the species if

the smoothing matrix does not cover the variation between the few parameter estimates.

3. Build a dependent probability model based on the parameter estimates. This uses correlations between parameter estimates to build a more realistic parameter set from the similar species parameters. It is useful where you have no parameter estimates for a species, for example, but wish to take account of its size (almost all species have an Lmax). Larger fish have a large *Winf* and may tend to grow more slowly, and so on. The method works by recalculating parameter estimates regressed towards the mean for that fishes size.

Catchability parameters might be estimated from interviews in the same way as for the logistic stock assessment or from fishing experiments. Unfortunately, catchability is dependent on more than just the gear type, some using other fishery's catchability is probably not advisable.

It is important to note that priors may favour parameter values far from the true value. They allow you to start the assessment process, but it is very dangerous simply to stop assessments at that point. Fishers' beliefs, like anyone else's (including scientists!), may well be incorrect and biased. In particular, fishers may well be optimistic over the productivity of their resource and in denial about the hard choices they must make to realise long term benefits. However, even in this case priors still provide a measure of how much scientific information may required to overcome this belief.

2.4.8 Constructing Source Models

Source models provide parameter frequencies. These parameter frequencies represent the relevant information of interest extracted from other models. Commonly such models are stock assessment models based on catch and effort data. However, in theory any models could be employed, although only a selection are supported.

Source models either are single entities, such as the stock assessment interview model or generalized linear model, or hierarchical models such as population model based assessments. Hierarchical models allow greater flexibility in designing stock assessments.

Population models can be non-linear and generally have few parameters. They define how the fish population changes over time, and are generally driven by the total catches. They are not fitted directly to data, but are connected to observations through generalized linear models. Generalized linear models can fit large numbers of parameters quickly. You can have more than one GLM for each population model.

Single species models have a two level hierarchy: the population model and then GLM link models. Multispecies models have three levels: the multispecies abundances, the single species population models and the GLM link models.

Population models hold the catch and effort data. Any other data, such as survey data, must be loaded into the generalized linear models themselves.

All parameters in the parent parameter frequency should be linked to source model parameters. Source models are generally straight forward. They are usually models fitted to data that share the same parameters as the target simulation model. Typically, for example, you would probably fit the logistic model to catch and effort data for the logistic simulation model. However, the dependent probability model is more complex as it can represent another set of parameter frequencies.

For more technical information on source models see Single Species Population Models, Multispecies Population Models, Generalized Linear Models and the Stock Assessment Interview.

2.4.9 Dependent Probability Model

Dependent probability models allow conditional probabilities to be modelled through a simple regression framework. The most likely use is in choosing growth parameters based on species size. This is most simply explained by an example.

Lethrinus borbonicus is similar to many other Lethrinidae, although it maximum size is smaller than many. No growth parameter estimates could be found for this species, but many were recorded for other species in the genus in Fishbase (www.fishbase.org). These other parameters were used as the basis for estimating *L.borbonicus* parameters. Although we could construct a prior probability from the raw parameter set, there will be a clear bias as most estimates occur for larger species. This would mean that the mean Winf parameter would be much higher than the true value and may bias the results (note however that as the simulation model deals with relative changes in yield, the absolute value of this parameter is not as important as first might appear). Other parameters may also be affected by size. M and K are often smaller for bigger fish, for example.

All parameter estimates can be loaded into a dependent probability model together with an indicator of size. The best size indicator to use is Lmax as this is available for all species. The common indicator can be used as the independent variable to regress the unknown parameter estimates for the species of interest. Lmax is added as a nuisance parameter when editing the dependent probability model. The original parameter estimates are loaded as GLM data most easily from Excel. You must then select the GLM Data parameters and switch the ones you want to dependent (Y) variables. Finally, you must supply the independent data for the species of interest on which the regression is based.

In the case for *L.borbonicus*, only the Lmax parameter was independent. Each dependent parameter has a linear model attached with a constant and slope parameter for the X variable. The independent data for *L.borbonicus* is a parameter frequency with one data value and a user defined smoothing matrix. These provide the mean and variance for a normally distributed value about the published Lmax for this species. The parameter frequency can be then generated by selecting the dependent probability model () and choosing Model | Generate Probability Frequency. This will firstly generate the Lmax parameters for *L.borbonicus* then bootstrap the parameter estimates for each of the GLM data regressions. The GLM data bootstraps share the same random residual selection, so that the correlations between bootstrapped estimates remain in tact. Each of the bootstraps is used to calculate the *L.borbonicus* dependent variable based on the random Lmax. If you view the generated frequency, you will see that the generated points are regressed towards values most appropriate for smaller species.

Even without a correlation between the dependent and independent variables, the bootstrap will reproduce the variability in parameter estimates. However, as regression towards the mean will reduce variance, it is best only used when parameter correlations are thought to be important.

2.4.10 Fitting Source Models

You can fit any model by selecting the model and choosing Model | Fit Model from the menu. On completing the fit, the maximum likelihood parameters should be displayed. Interviews and Dependent Probability models do not require fitting. Only GLM and population models with GLM link models can be fitted.

If an error message is displayed, you will need to take some action:

- **Population parameters are invalid**: If you have tried to fit a GLM link model, you will need to have valid parameters in the population model. Either edit the population model and manually provide valid parameters, or fit the population model rather than the link model.
- **Other errors**: Check you have data in all models in the hierarchy. Select each model and then select Model | Edit Data and Model | Edit Model from the menu. If a model has no data, add data or delete the model. If all appears correct, try selecting Update All Models from the menu.

Note that you should regularly save the model from the main form menu.

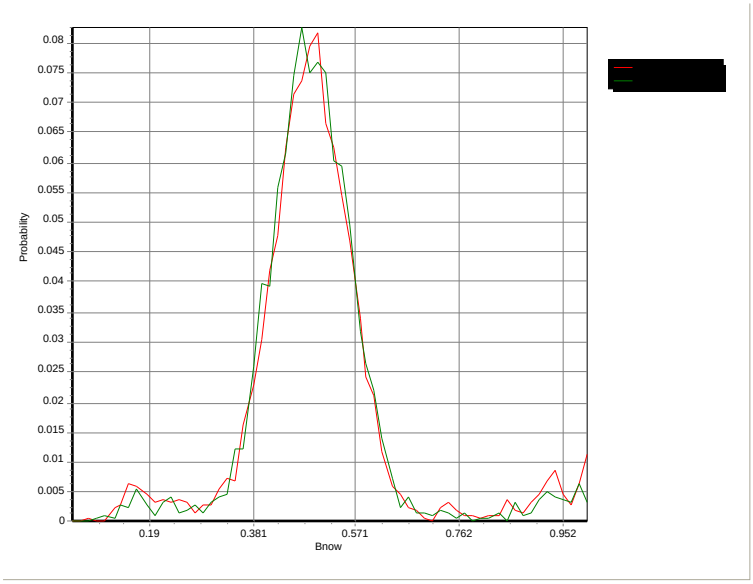
Once you have completed the fit, you should check it by looking at the observed-expected plots. Selecting each GLM model and select Plot | Plot Observed Expected from the menu. There should be a close correspondence between the observed (points) and expected (line) graphs. However, this software is not good for exploratory statistical analysis as there are no other diagnostic tools. You should already be reasonably confident that these types of models fit your data from plots in Excel, for example.

Once you are satisfied that the models fit, you can choose Model | Generate Parameter Frequency. This will refit the model the required number of times using a bootstrap process. The bootstrapped parameters will be loaded into the linked parameter frequencies. You can then view the fitted kernel probability distributions by selecting the parent parameter frequency and choosing Plot | Plot Kernel from the menu.

2.4.11 Viewing Probability Distributions

Once you have added frequency data to a parameter frequency, you should compare the data to the fitted PDF to make sure an error has not occurred in the fitting process. You can do this by choosing Plot | Plot Kernel from the menu. Both green (data) and red (PDF) lines should be displayed and should follow one another.

You can choose the variable plotted using the pick list control by the side of the graph. The principle component scores (PCA) used in the fitting analysis can also be viewed. The smoothing can be rescaled manually. This is not recommended for analysis unless the automatic procedure appears to have underestimated the smoothing parameter (i.e. the PDF is too spiky).



Correspondence between the data and the PDF will not be exact. Discrete data has been smoothed into continuous distribution, so that we would not expect them to follow exactly. Also, if there are several parameters, smoothing is taking place in several dimensions. If the data form a single homogeneous group, this should not create a problem. However it is possible that smoothing will allocate probability mass to space where data does not actually exists. This may be inappropriate, particularly for non-linear models. Inappropriate parameter combinations are removed during the simulation process, which reduces this problem. But you should nevertheless be aware of the potential problem where large numbers of dimensions (i.e. parameters) exist.

Viewing the principle components (PCA) should reveal independent PDFs as the PCA are not correlated. However, the correlations that have been "removed" are globally linear, so non-linear relationships between parameters may still cause problems.

In general, problems are more likely to occur with large smoothing parameters and few data frequency points. Large number of points with a PDF reflecting local parameter space will tend to be more robust to errors.

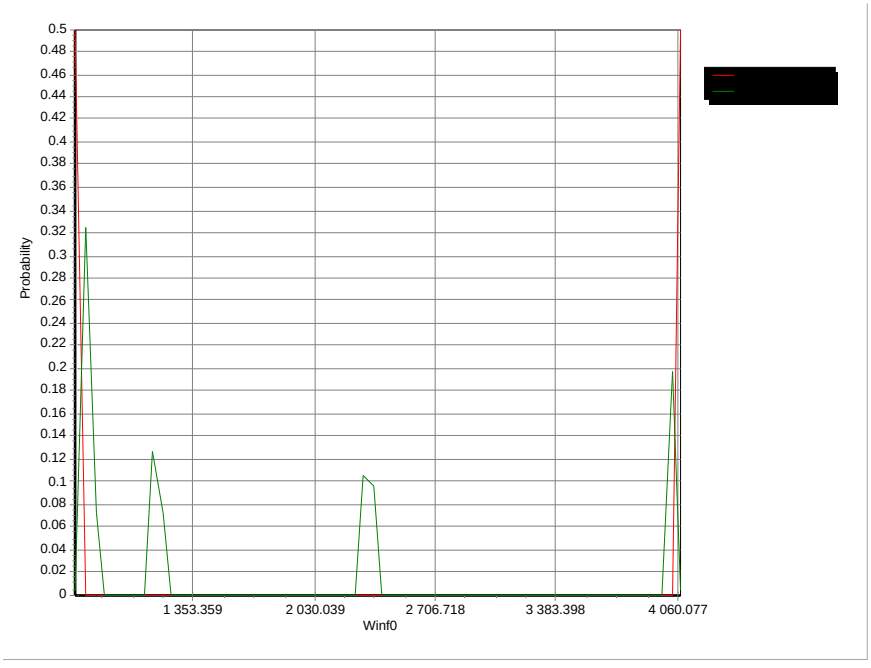
Smoothing Errors

Smoothing failure occurs when the fitting process degrades the smoothing parameter to a very low value. The software should detect this and prevents the parameter from going too low, but is effectively unable to estimate it. The PDF tends to become very spiky as the probability is gathered tightly around each observation (i.e. there is little or no smoothing). This can happen particularly where the number of parameter frequencies is small and the dimensions are high. In most cases smoothing failure does not matter much. For example, it is quite common under very high exploitation levels for the resource state smoothing to fail in the analysis. For the necessary statistics, this is not particularly important and generally occurs at an implausibly high exploitation

level. It would matter, however, if smoothing failed in one of the parameter frequencies used to generate the posterior. In this case, you are advised to check the data and to use the manual smoothing scaler available when you Plot | Plot Kernel. Unfortunately there is no objective way to choose the smoothing parameter under these circumstances, and so you should reapply different smoothing levels until it subjectively looks right. You may then wish to turn off the Recalculate covariance matrix and Fit smoothing parameters in the Edit | Edit Model for the selected parameter frequency to prevent the problem re-occurring.
In summary, if errors occur, it is suggested that you:

- check the data is correct and that the frequency data is reasonable.
- inspect each parameter frequency using Model | Edit Model. Check that the recalculate covariance matrix and fit smoothing parameters is turned on or off as required. These are sometimes turned off automatically.
- increase the number of parameter values being produced by the Generate Parameter Frequency from the source model.
- Reduce the numbers of dimensions by separating parameters or groups of parameters into separate parameter frequencies. This assumes such parameter frequencies are drawn from independent distributions.
- Apply a subjective smoothing multiplier to the smoothing parameters or supply a smoothing matrix from another source.

In some cases, the graph fails but no other error occurs. This should indicate that something is wrong in the analysis. For example, the smoothing matrix may be singular if parameter frequencies their linear correlation is close to 1.0. A similar problem can also occur if separate parameter frequencies do have have similar scales. The smoothing matrix must be inverted and numerical errors can occur. Separate parameter frequencies with estimates orders of magnitude different can produce this problem. In most cases there will be a problem with the underlying data and you should review this first as the data is probably incompatible. Otherwise you could try to edit the smoothing matrix to supply a variance for the problem variable. Typically, a singular smoothing matrix produces a graph with no PDF (red line) or the PDF shifted to the extremes of the distribution.



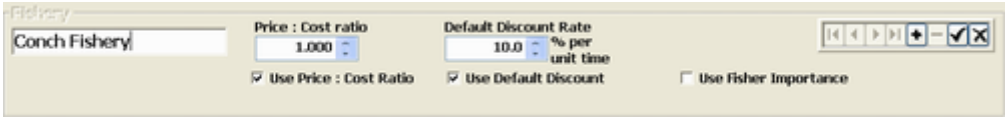
See kernel Model Fitting for more technical information on the fitting procedure.

2.5 Preference Models

The preference models form allows you to change the way the preference score between the simulation outcomes is calculated. The aim is to use a score which is not only higher for those outcomes which are prefered, but where the difference in the score between two scenarios is a measure of how much more one is preferred over the other.

The form has a header for global options and a fisher section to display data on the current selected fisher.
The global header allows you to change the options that apply to all fishers. There is a name for the fishery which can be entered but serves no specific purpose. You can indicate whether you wish to use the price cost ratio function by clicking on the check box () and enter a specific parameter (see below). There are two options for calculating preferences. One is to use the price cost ratio, a global linear function describing how preference catch and effort vary (see below) and is useful for exploratory analyses (see below). The much better option is to carry out interviews with fishers to obtain their preferences. If you do not wish to use the Price : Cost Ratio function you must have individual fisher data entered.

You can choose whether to use the default discount rate rather than the estimated individual fisher discounts. This may be useful for sensitivity analysis or preferable if the fisher discounts are not thought to be reliable. You must provide the global discount as a percentage per unit time. Finally you can choose whether you wish to use the fisher importance measure. This is used to weight each fisher's score. It is suggested that it should measure the contribution of this fishery to each interviewee's household income. It may be turned off as part of a sensitivity analysis or if the importance data is not considered relaible.



Navigation Buttons

If you have fisher data loaded, you can move between the records using the following buttons.

First: Moves to the first fisher in the list.
Previous: Moves to the previous fisher in the list. Disabled at the beginning of the list.
Next: Moves to the next fisher in the list. Disabled at the end of the list.
Last: Moves to the last fisher in the list.
Add: Adds a new blank record to the list of fishers ready for data entry.
Delete: Deletes the current fisher from the list.
Save: Stores the edited data. It is enabled only after changes have been made.
Cancel: Cancels any changes and reloads the original data from the fisher record. It is enabled only after changes have been made.

Save and Cancel also work for the header information.

Price Cost Ratio

The global price-cost ratio function requires a single Price : Cost Ratio parameter (PCR) which weights the proportion change in catch relative to the proportional change in effort from the current situation such that:

Score = PCR * Catch% - Effort%

If the score is proportional to profit, the weight might be calculated as the value of the catch divided by the catching cost: PCR = (Price*Catch)/(Effort*Cost). Clearly, the higher the PCR value, the more important changes in catch are to changes in effort. The default value is 1.0, so, for example a 10% increase in catch coupled with a 10% increase in effort will be viewed just as good as no change in either. The function is provided mainly as exploratory tool to allow some

analysis before interviews are completed.

Fishers

Each fisher interview contains information necessary to calculate the preference score. This consists of the the rank and score for each catch-effort scenario, the scores for each species (if appropriate) and any constraints of the catch, effort or CPUE that should be applied to the fisher.

Fisher

- **Name**: Identifier for the household (e.g. family name or interview number)
- **Importance**: Allows importance of fishers to vary from 1.0. For example, the number of dependents or the level of dependence on fishing as opposed to other economic activities may make some fisher's preference more important than others. The importance is used as a weight in the score calculation.
- **Discount**: The estimate of the discount rate as a percentage over the unit time for this fisher.
- **Constraints**: The range of catch and effort to which preferences may reasonably be applied. The minimum will be at or above zero for both variables. For example, a maximum effort in boats days may relate to the number of working days in the time period, such as 6 days in the week. Minimum CPUE defines the point when a fisher will stop fishing because income per day has fallen too low. Minimum catch would be the minimum income that a fisher would accept from this fishery in a unit time before he might give up fishing in this fishery. Maximum CPUE is the maximum catch rate that a fisher may be able to cope with his current gear and boat. Maximum effort is the maximum number of days a fisher can realistic fish in a unit time. The preference score is fixed at these boundaries, so for example the score will never fall below the zero-catch, zero-effort point.
- **Enabled**: Disable fishers to prevent them being used in the preference score calculation. This is useful to remove sets of fishers for sensitivity analyses.
- **Current Catch and Effort Reference Point**: The current levels of catch and effort are important for scaling purposes. You can see the reference points for each gear by clicking on the page tab. All changes are considered as changes from the current situation. You can enter the current days fishing and sets per day (effort = days*sets). The maximum sets per day is the maximum number of sets (trap hauls etc.) which can be applied in a day. "Sets" can be any effort unit, such as fishers per boat, traps or net hauls.
- **Catch and Effort Preference:** The catch-effort preference measures fishers attitude to different scenarios. Scenarios describe relative changes to the current catch and effort. These are ranked and scored through interviews (see Preference Data). You can only enter each scenario once in the list. Scores are integers from 0-9 and measure the difference from the scenario immediately below. Once complete, a graph will appear showing the score function for this fisher. You can rotate the graph using the arrow buttons ().
- **Catch Category Preference**: Each species needs a score giving a relative measure of importance. This score must reflect the importance of the species in the catches. This importance can be based on just catch or on catch value or other criteria.

2.5.1 Loading Preferences From Excel

Data can be loaded directly from MS Excel. The order of columns is fixed, with multiple columns being used for each gear and set of catch categories as appropriate. The order in the spreadsheet should be consistent with the model and parameters being used otherwise an error will be raised and the data will not be loaded. The names on the first row are ignored.

There is an implicit time unit (usually a month or year) over which effort and catch is measured. This should be consistent for all fishers.

The data format is consistent with a relational database table, with each record representing a

single fisher. The relevant range should be selected on the worksheet, with the first line being then field names (ignored), and the data in each record below. Each row will create a new record in the software.

The number of columns depends on the number of gears and number of species in the assessment. There should be Sets Per Day, Usual Effort, Usual CPUE and Max Sets for each gear. There should be species preference scores as the last columns for all species only if there is more than one species.

	Fisher Na	Importanc	Discount I	Sets per c	Usual Effo	Usual CPI	Max Sets	Min CPUE	Min Catch	Max CPUI	Max Effor	Scenario	Scenario	Scenario	Scenario
7	Walter Mo	2.4	14.7929	1	12	1000	1	0	0	1500	12	A	N	E	B
8	Alvin Parki	2.1	1.981179	1	12	800	1	0	0	2000	12	A	B	N	E
9	Henry Han	0.75	5.196305	1	18	900	1	500	0	2000	24				
10	Marsha Pa	0.01	68.13692	1	1	2.5	1	0	0	7	3	A	J	B	N
11	Russell Je	0.1	14.7929	1	3	200	1	0	0	600	8	A	J	B	N
12	Eroude Pa	2.8	61.73033	1	16	600	1	200	250	1400	16				
13	Branford H	1.5	1.981179	1	20	500	1	300	300	1000	20				
14	Harold Wa	4.8	3.923108	1	16	650	1	400	3200	1500	20	A	J	N	B
15	Daryl Hins	3	3.923108	1	5	600	1	200	200	1500	24	A	J	B	N
16	Anthony D	1.5	3.923108	1	8	600	1	200	0	1500	20	A	B	K	N
17	Oswald W	2.4	3.780454	1	16	550	1	200	0	1100	24				
18	Mack Smit	0.35	2.632811	1	12	750	1	400	450	1200	16	A	N	B	J
19	Tommy Pr	1	3.150996	1	12	900	1	300	8100	1000	12	A	N	B	E
20	Charles Lo	0.5	1.981179	1	12	600	1	500	0	2000	12	A	B	N	J
21	Anthony M	1.5	7.692308	1	16	600	1	0	0	700	16	A	B	N	J
22	David Forb	0.5	7.692308	1	14	200	1	0	0	1500	14				
23	Stewart Hu	2.1	0.995545	1	12	700	1	200	250	1300	30	A	B	J	K
24	Spice	0	0.995545	1	3	450	1	30	30	1000	20	A	H	N	B
25	Ammeth C	0.4	1.325186	1	16	650	1	300	0	800	20	A	N	B	O
26	Gordon Sn	0.2	27.3975	1	20	700	1	300	0	1200	24	A	J	B	N
27	Terence	1.2	2.632811	1	20	500	1	200	200	2000	24	A	B	N	J
28	Don Wilco	8	2.632811	1	20	700	1	150	200	1700	24	A	N	B	J
29	Rudolph S	7.2	2.632811	1	20	400	1	0	0	700	24	A	N	B	O
30	Alden Mon	0.8	5.826563	1	20	600	1	200	200	1000	24	A	N	B	J
31	Sherlock F	1.8	1.981179	1	25	700	1	300	0	1500	16	A	N	B	O
32	John Ewin	0	3.893437	1	2	50	1	25	0	100	4	A	J	N	B
33	Robbie Ro	1	2.632811	1	2	200	1	150	300	1500	20	A	B	H	N
34	Leroy Broc	0	3.923108	1	14	600	1	0	0	1000	25				

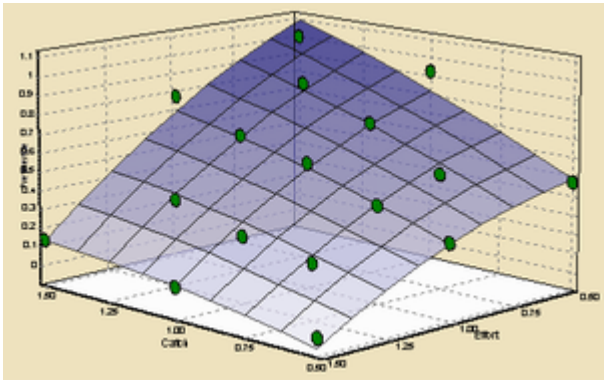
Field Order

Fisher Name: The name of the fisher
Importance: A score indicating the importance of this fisher to the overall score.
Discount Rate: Discount rate as a percentage rate per time unit
For each gear in the fishery four columns must be entered...
Sets per day: Effort units on each day (zero if gear is not used)
Usual Effort: Days fishing each time unit (zero if gear is not used)
Usual CPUE: The average gear catch rate (zero if gear is not used)
Max Sets Per Day: Maximum number of sets in each day (zero if the gear is not used)
Min CPUE: Minimum catch per day before fisher would leave the fishery
Min Catch: Minimum catch in a unit time
Max CPUE: Maximum catch per day a fisher could cope with using his current gear
Max Effort: Maximum effort
Scenario 1: First Scenario in rank order (always A)
Scenario 2: Second scenario in rank order
...
Scenario 16 Sixteenth scenario in rank order
Scenario 17 First Scenario in rank order (always C)
Score 1: Score of difference in preference between first and second rank
Score 2: Score of difference in preference between second and third rank
...
Score 15: Score of difference in preference between fifteenth and sixteenth rank
Score 16: Score of difference in preference between sixteenth and seventeenth rank
Species Pref 1: Preference score for the first species (not used if only one species)
Species Pref 2: Preference score for the second species (not used if only one species)
...

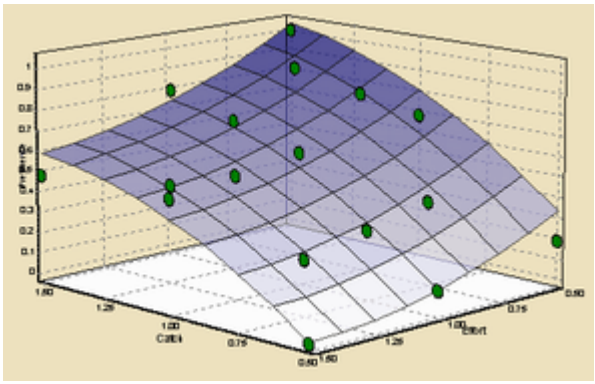
2.5.2 Preference Models

The absolute difference in scores is not important. The scaled difference between the best and worst scenarios is 1.0. A simple quadratic model is then fitted to the data. The model allows non-linear change and an interaction between catch and effort. The model provides a best fit to data points, which individually are probably not reliable.

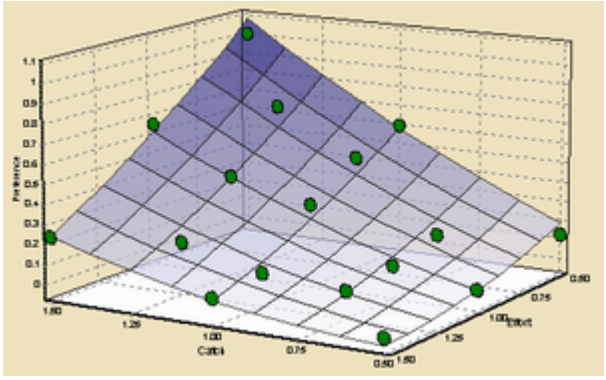
Three example preference curves fitted to interview data are presented below. They give a range of reaction to changes in catch and effort.



In the first case effort is risk adverse as the curve is convex. The fisher was relatively more worried about working harder than having to work less. In terms of catch, the relationship is linear, implying risk indifference. The more catch the better and each fish is worth the same whether the overall catch has been high or low.



In this case the catch is risk averse, so a loss of catch compared to the present is associated with a greater loss in utility. Conversely, the effort curve is risk seeking, with the relative gain in utility from lower effort exceeding the loss in utility with higher effort. However, note that this curve has a relatively high error, indicating possible inconsistency in the answers given.



In this case the fisher is risk seeking with respect to catch rate. The increase in utility with an increase in catch rate is much greater than the decrease in utility with an equivalent decrease in catch rate.

In terms of encouraging overfishing, risk seeking behaviour is only dangerous in relation to the catch. In a fishery where fishers express a much greater preference for high catches and relative indifference between current and lower catches, higher fishing is likely to occur. In terms of effort, there is more a concern over the general slope rather than risk behaviour. Indifference to more work implies less concern over having to increase effort to maintain income in the face of overfishing.

2.6 Management Controls

It is necessary to define the management controls which will be applied to achieve policy objectives for the fishery. These are set for simulated projections of the fishery, which are used to assess management outcomes. Whole sets of controls can be tested using this method to estimate expected outcomes even when there is a great deal of uncertainty as to the state and behaviour of the stock.

Controls

Type of control: Effort, quota or closed area controls are supported. Effort would represent limiting the numbers of fishers, days they fished, number of traps and so on. Quota is the total catch landed. This is usually only monitored for commercial fisheries with only a few landing sites. Closed area measures the proportion of the population protected from fishing, and would usually be implemented by a no-take zone. The logistic model can use all three controls. Yield-per-recruit can only apply effort control.

Current, Maximum, Minimum Controls: In the table, you must enter the current level of the control (i.e. current effort, quota or closed area) and the range (maximum and minimum) over which the control will be simulated. All comparisons will be made with current situation, so scores will measure relative improvements to the current situation. The current control need not be within the specified maximum-minimum range. If the control type is effort, values are required for each gear; if quota, values are required for each species, but if closed area, only a single control is required. All closed area controls must be between 0 and 1.0, being the proportion of the stock protected. Note that the effort minimum and maximum should, in general, not vary outside +/- 50% of the current value, as the preference information becomes extrapolated rather than interpolated at this point.

Effort: If the control type is not effort, you need to supply the current effort and new effort which will be applied in the simulations.

Current Reference Points: The current target and limit reference points are displayed. These are only valid if an analysis has been run. The limit reference point is calculated as the probability that the stock has fallen below a specified state. The limit probability and limit state must be provided as reference point parameters. In all cases here, the state of the stock is the biomass as a proportion of the unexploited biomass.

Simulation Parameters

Control Increments: This is the number of increments between the maximum and minimum control for which simulations are carried out. The more increments there are, the longer it will take to complete the estimate over the range of controls, but the more accurately the reference points will be estimated.

Number of simulations: The number of simulations affects the accuracy of the estimates. The more simulations, the more accurate the estimates, but the longer it will take to complete each analysis. Exploratory analyses could probably be undertaken with as few as 500 simulations, but final results should use considerably more (1000-5000 posterior draws). Ideally, the number of simulations should be increased until there is no discernible difference in the results between

separate posterior draws.

Projection Time: The projection time is the period the fishery is projected into the future on each simulation. High discount rates down-weight the later period of the projection. If the projection time is very long, each simulation takes longer to complete and later values may have little influence on the result. If the projection time is very short, decisions will be based only on short term outcomes, which may lead to unsustainable management. With reasonable discount rates, the projection time can be estimated so that further extension will have negligible influence on the results. This is done by pressing the calculator button(). The maximum projection time is 500.

See the Simulation Model for more technical information.

2.6.1 Effort Control

The effort control is the basic control which can be applied to either the yield-per-recruit or logistic population models. An effort control should control both fishing mortality (catches) and costs of fishing. Effort is used in the preference model, and in theory can be used to meet both management limits and management targets. In particular, for developing country fisheries, economic concerns are paramount and meeting such objectives are most likely to be achieved through effort control. Effort is more difficult to enforce, but under co-management a important benefit should be the ability to control fishing effort.

See Modelling Effort Control for more technical information on how this control is applied in the simulation.

2.6.2 Quota Control

The quota control applies to each species. Usually quotas are used in commercial fisheries, as all catches need to be measured and monitored for this control to be effective. Unless there is some critical control point, for example at export, quotas are difficult to enforce and generally not recommended for small scale fisheries. Furthermore, if a quota is set above the MSY, it will have no benefit at all, yet there will still be costs in managing it. Unless there is very good information on a fishery, quotas should be avoided as the sole control.

The quota is the total landed catch. Once the quota is reached effort is stopped. The new effort is the new level of effort applied to the fishery. You should set this relatively high so that the quota has some effect. If the quota is set too high, the new level of effort will become the control and there will be be no difference between quota levels. As for all fishing controls, the quota must be set low enough to have an impact on fishing.

See Modelling Quota Control for more technical information on how this control is applied in the simulation.

2.6.3 Refuge Control

You must specify the current, maximum and minimum proportion of stock which is to be protected. If you have no closed areas, the current proportion will be zero. You also need to set the current and maximum exploitation rate. The current level of exploitation is the current effort level and forms the baseline for comparisons. The new effort is the new level of exploitation which will apply

from the time the new control is implemented and is constant. This allows you to estimate what the level of effort might be after the control is implemented.

Management can provide a refuge from fishing by setting up closed areas or no take zones. In these areas, no fishing is allowed. Such zones may provide many benefits beyond that dealt with in this assessment model, and each of these benefits may be sufficient to justify a closed area. In particular, a no-take zone, if large enough, may maintain an unexploited habitat and ecosystem with which the fished area may be compared. This may not only preserve adult and juvenile fish. Where fishing causes habitat damage, or pollution and temperature effects may also be having impacts, such information is invaluable in helping management make decisions.

For current purposes, the refuge removes a proportion of the stock from fishing. This can only be done with the biomass dynamics model in the current software. (Refuges could only be interpreted as a change in fishing mortality in yield-per-recruit which is equivalent to effort change without any clear economic benefit.) The control is the proportion of the stock which has been protected from fishing. If the stock is evenly distributed, this could be the proportion of area of the fishing grounds which is closed. It is your responsibility to translate the proportion of stock protected to the actual practical implementation of a control.

The model makes two important assumptions about the way this control works. Firstly, the proportion of the stock denied to the fishers is assumed not to migrate to the fished area. The fished area only benefits from the biomass growth which is shared between the fished and closed areas. Secondly, because a proportion of the stock is removed from the fishing, the catchability will effectively fall by that proportion. Effort is assumed to remain unchanged (i.e. set as defined for the new effort in the control form). If effort remains unchanged, there must be a fall in catchability for the control to have any effect at all. If the primary effect of the control is to reduce effort, this should be modelled as an effort control rather than a closed area or refuge control.

In general in this model, a refuge control will not meet economic target controls unless overfishing is already taking place. Controlling effort directly is probably preferred for economic reasons. However, no-take zones can considerably reduce risks of a stock falling into an overfished state, and should always be considered as part of a set of management controls.

For more technical information on this implementation of refuge control see Refuge Control

2.7 Analysis and Simulations

To carry out an analysis, you need to:

- Regularly save the model in case of errors.
- Complete a posterior draw with as much information on the parameters as possible. You can check the population projections for the draws from the main form Projections | Plot Population Projections.
- Specify the preference scoring method. Use interview data if you have it.
- Specify the type of control you wish to apply and the range of the simulations.
- Choose Projections | Analysis to run the analysis. Depending on the options chosen, this may take a long time to run. The preference graph will be displayed while the projections are being calculated.
- Keep the number of simulations and control increments low to speed up the analysis for exploratory analyses only. Also, keep the range of control wide so that you can see how the simulation behaves.
- Once you are satisfied with the settings, increase the number simulations to 1000 or more and narrow the control range to the area of interest.
- On completing an analysis, the preference scores and resource state probabilities are displayed in the graphs. The target and limit reference points are recalculated and a new scenario is entered in the main form table.

Options which may be changed between scenarios include:

- Enable/disable particular data sources
- Enable/disable fishers
- Not use/Use importance
- Use default/Use fishers discount rates
- Use/Not use Price Cost ratio

2.8 Graphs

Several plots are available to explore the current analysis. In all cases you can zoom to any area on the graph by holding the left mouse button and mark the rectangle you want to zoom to. You can unzoom by marking a rectangle from the bottom right to the top left (i.e. reverse direction). On all 3D graphs, you can change the rotatation, zoom and elevation using the sliders on the right of the graph.

Population Projections

The population projections illustrate how the population might change if the current effort remains the same. Because the projection model is deterministic, the projections tend to move towards equilibrium smoothly. In the case of yield per recruit they will already be at equilibrium and no change will occur. They are only useful to check population parameters and the range of population sizes they predict. Fewer simulations and short projection time are recommended.

2.9 Interpreting Results

The results focus on the target and limit reference points, which indicate what management should do. These reference points incorporate probability, so that as long as the probability distribution is available, the reference points are defined. This allows action to be taken regardless of the state of knowledge.

When uncertainty is very great, the target reference point is likely to be higher (or lower in the case of closed areas) than the limit reference point. This is because the limit will have to fall much further to meet its limit probability criterion. In these circumstance the limit reference point will probably be too onerous for fishers, and an important aim should be to reduce the uncertainty in the assessment.

The assessment will have more uncertainty than that being modelled. This should allow some room for manoeuvre, so the estimated reference points should be guidance only. Whatever controls are chosen, they should have clear and transparent justification. Uncertainties in the assessment which need to be considered are:

- Whether the parameter probability distributions are good representations of the "true" probabilities. You should inspect the kernel plots to make sure both that the probability distributions are reasonable representations of the underlying data and that the actual values and spread ot the probability makes sense. Clearly, the valid estimates will depend on the source models and the various assumptions upon which they depend. You shold be familiar with fitting these types of model. It should be noted that as the posterior draws are only used for numerical integration reducing many dimensions to one or two, the final results will be reasonably robust to unbiased errors.
- Whether the preferences may be inaccurate. This largely depends on interview data. It may be worth while having a personal standard interview with reasonable scores and ranks for comparison with interview results. Otherwise, as long as the interviewees understood the choices, their views should stand. The sample of interviewees needs to be representative of all fishers. Individual interviews can be disabled to check their influence.
- Whether there are structural model errors. The most important inaccuracy arises from the

simulation model. The two models provided are standard fishery models, but they cannot be expected to provide accurate projections of what will happen, not least because the projections are deterministic. They instead are being used to provide a standard rational way to define reference points. Following the model advice may not optimise the fishery, but should move it towards the optimum, and as better information comes available, further improvements will be possible.

An important aspect of this approach to assessment is the use of probabilities rather than point estimates. This emphasises the uncertainty in fisheries and the importance of information. As a result, the assessment should be seen as part of an adaptive management process. The ability to generate prior probabilities and obtain reference points should not be seen as an end in itself, but simply as a start point so that assessments can be updated and improved as management is implemented.

2.9.1 Fisher Preferences

Preference score by fisher

This scores each level of control for each fisher. The preference score is the discounted sum of preference over projections for randomly selected model parameters. The control varies from the minimum to maximum control with the number of increments, all set in the control form. The maximum represents the action which results in the best expected outcome for that fisher. This average fisher score (black line) is simply the average of all the fisher preferences scores and is used to set the target reference point. The maximum average score is the action with the best expected outcome for all fishers and therefore chosen as the target. The target itself is estimated by interpolation between calculated points. The average score is weighted by the fisher importance if this is selected.

The scores are based on the changes in catch and effort from the current situation for each fisher. The delays between the current and equilibrium state are taken into account through the discounting. Although the equilibrium state may provide improved CPUE, there is often a period of lower catches before this can be attained.

If you click on one of the fisher (red) curves, the appropriate fisher who's preference it is will be displayed in the preference form. This is useful if you wish to check data or, for example, check whether particular fishers have undue influence by disabling them.

See Preference Data for more technical information on how the score is calculated.

2.9.2 Resource State

The smoothed frequency distribution of the resource state is displayed. The probability distribution represents the probability that the stock will be at the state indicated if the observation is chosen at random from all the simulation, time interval and species observations combined. That is, the PDF is built on the state frequency of all time periods, simulations and species combined. This means that the longer the time series, the more highly the equilibrium state will be represented. This state probability has proved to be a good indicator of the effect of controls and a convenient way of simplifying complex multi-dimensional data.

The probability of being below the specified limit state is flagged on the graph for each control level. This is automatically updated when the limit state is changed.

The resource state is defined here as the current biomass as a proportion of the unexploited biomass. Other state reference points are used, usually based on estimates of spawning stock. However, the emphasis here is on little information and requiring knowledge on the age of maturity and fecundity may make the assessments less reliable. Requiring the biomass to be above a particular level is a better aim unless the life history is very well understood.

You can change the limit state and the limit probability. The limit state change will be reflected on

the graph. For the logistic model, the appropriate limit state is 0.5. For yield-per-recruit, states below 0.5 may be considered particularly if mature fish are being exploited. The limit probability changes the reference point, but this is not reflected in the graph.

You can click on any line and it will be displayed by itself. Click on it again and you will return to displaying all control levels.

See Simulation Model for more technical information of the relevant population models.

2.9.3 Current Resource State

The graph shows the cumulative probability distribution for the current resource state. The resource state is the current biomass as a proportion of the unexploited biomass. The probability that the current state is below the limit reference point is displayed. The spread and steepness of the slope of the curve indicates the accuracy of the current estimate. A narrow curve increasing around the current estimate indicates high accuracy. This curve can be viewed whenever a valid draw is made on the posterior distribution.

See Simulation Model for more technical information of the relevant population models.

2.9.4 Reference Point Probability

The probability that the stock state falls below the limit state is displayed across the range of controls. The limit reference point is chosen when this probability equals the defined limit probability. This change is illustrated by a change in colour in the curve. You can change the limit state and the limit probability. The limit probability change is illustrated by the colour change. The lower the probability, the lower the limit control will be set. Changing the limit state will change the shape of this probability curve as different cross sections in the Resource State probabilities is taken. Again, this will affect the limit reference point.

There is no generally accepted way of choosing the limit probability. The value depends on how much risk management is prepared to accept. There is little point in choosing a value above 50% as this would be risk seeking. However, under significant uncertainty, being too risk averse will limit exploitation to very low levels. This would probably be unacceptable to the fishers. As a result you might be more interested in finding what the risk probability is at the target reference point (increase the limit probability until the limit equals the target reference point in the control form) and focus on how data collection might reduce this risk.

See Simulation Model for more technical information of the relevant population models.

2.10 Credits

The program was written by Paul Medley in Delphi (v 7.0) language (www.borland.com). The String grid component (TStringAlignGrid v. 2.1) was provided by Andreas Hörstemeir as freeware (www.hoerstemeier.com). The charting component (TChart) was provided by Steema (www.Steema.com). Many of the visible components were provided by LMD Innovative (www.lmd.de). All other parts of the software, including the technical routines, were written by Paul Medley (paul.medley@virgin.net).

The field work testing was conducted by Narriman Jiddawi, Paul Medley, Oliver Taylor, Saleh Yahya, Omar Amir and Kathy Lockhart. Field work was also supported by Hamad Khatib, Mohammed Sulieman and Rashid Juma.

Special thanks go to Dr Jiddawi and Dr Duby at the Institute of Marine Sciences, Zanzibar and Michelle Fulford-Gardiner and Judith Campbell of the Department of Environment and Coastal Resources, Turks and Caicos Islands for their co-operation.

The software links to Microsoft Excel (c) if it is installed. This is not an indorsement of Excel, but recognition that it is currently the most widely found spreadsheet software.

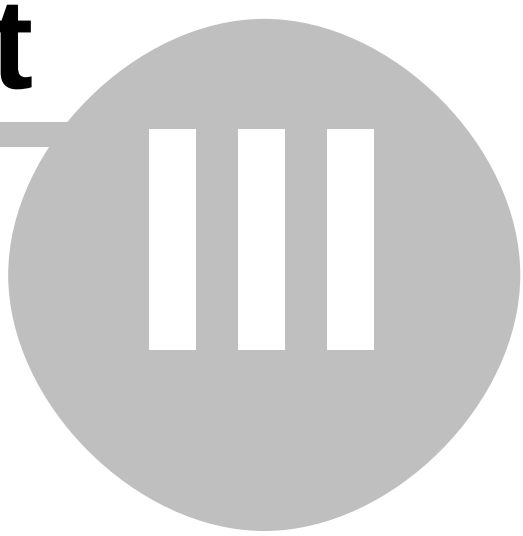
The help file was written by Paul Medley, Oliver Taylor and Gudrun Gaudian.

The project was funded by the UK Department for International Development.

PFSA

Participatory Fisheries Stock Assessment
Software Manual

Part



3 Technical Background

3.1 Overview

The decision model consists of two components:

- A probability model that primarily uses frequency data to probability density functions.
- A preference model that describes how individual households might feel about different outcomes from the fishery.

A optimal decision at any time is that which maximises the expected utility. The expected utility of a particular decision is found by multiplying the probability of each possible outcome by its utility and summing. While obtaining a exact value for expected utility is difficult under circumstances, it can be estimated relatively easily and accurately as the average utility from computer simulations.

The probability distributions of the relevant parameters are modelled from frequency data. This is a robust general technique similar to frequency histograms which can be similarly used to view probability distributions. Frequency data can be obtained from interview and stock assessment models. Strictly speaking, the frequencies are assumed to be drawn from prior and likelihood probability distributions.

The outcomes from each decision are modelled as simulations of results from applying the decision. For example, the decision might be to set an appropriate level of fishing effort. For each level of effort, the outcome is the level of effort itself, and the catch. Other indicators of outcomes may be chosen, but in general the main measures will be the work done (effort) and the production (catch).

The utility is approximated as the relative preference of the society between outcomes. Relative preference can be obtained directly from interviews, simply by asking how much people might prefer one outcome over another. However, problems arise as questions may be difficult to answer and inaccuracies may result. Furthermore, there are too many possible outcomes to ask about each separately. A model is required to predict these preferences both for the most difficult questions and the larger number of questions which cannot be asked.

3.2 Simulation Model

The decision model uses risk explicitly to define an optimum decision. This is the best decision under uncertainty, but may well turn out not to be the correct decision had better information been available. This underlines the fact that although the methodology will always produce advice even when very little is known, it is not a substitute for data collection and scientific stock assessment. It should, however, encourage information collection (as all information can be used) and better implementation of the results as fishers are involved in the whole process.

3.2.1 Logistic Simulation Model

Although a number of models exist for stock assessment, the biomass dynamics models possess an advantage in their simple demands for data (catch and effort) and in their basic assumptions. In multi-species fisheries, such as coral reefs, the model may be used to provide advice on the general productivity of the system and avoid trying to model hundreds of species. Models of the population dynamics of individual species could wait until better information comes available.

The simplest and most commonly used biomass dynamics model, the Schaefer model, provides advice on a limit reference point, the maximum sustainable yield (MSY). This limit reference point

can be used to restrict the risk of unsustainable fishing to an acceptable level.

In the difference equation form, the multi-gear logistic fisheries model is written as an equation describing how the population changes through discrete time (usually annual), as:

$$B_{t+1} = B_t + rB_t\left(1 - \frac{B_t}{B_\infty}\right) - C_t$$

$$C_{gt} = \frac{F_g}{\sum_g F_g}\left(1 - e^{-\sum_g F_g}\right)B_t$$

$$F_g = q_g f_{gt} \qquad (1)$$

where B_t is the stock biomass at time t, and C_t is all catches combined in the fishery, F_g = fishing mortality, q_g = catchability and f_g = effort for gear g. The model requires three population parameters: B_{now} = state at the start of the projection ($B_0 = B_{now} * B_\infty$), r = the rate of population growth, B_∞ = unexploited stock size, and as many catchability parameters as there are gear types. Apart from being slightly more accurate when fishing mortality is high, the catch equation avoids negative estimates for catches when fitting the model, so it is preferred to a linear catch model.

The state of the stock is defined as the biomass (B_t) divided by the unexploited biomass (B_∞). If the stock state falls below that required for the maximum sustainable yield (0.5), the stock is overfished.

3.2.2 Yield-per-recruit Model

Yield-per-recruit models focus on balancing the benefits from growth against losses from natural mortality. Growth is modelled as the weight form of the von Bertalanffy growth equation, which calculates mean weight as a function of age.

$$W_a = W_\infty\left(1 - e^{-K(a+a_0)}\right)^b \qquad (2)$$

where W_a = the weight at a years after recruitment, W_∞ = the asymptotic weight (W_{inf}), b = exponent converting length to weight (W_{exp}, usually close to 3.0), K = instantaneous growth rate, and a_0 = age at recruitment to the fishery such that the average weight at recruitment is derived from the model. This means that a_0 implicitly includes the growth model parameter t_0.
The yield-per-recruit combines the weight function with the negative exponential population model. Assuming knife-edge selection (i.e. all animals recruit to the fishery at the same age for all gears (g) and thereafter catchability is constant), the per-recruit stock biomass at equilibrium can be calculated as:

$$B = \sum_{a=0}^{A} e^{-\left(\sum F+M\right)a} W_\infty\left(1 - e^{-K(a+a_0)}\right)^b + \frac{W_\infty e^{-\left(\sum F+M\right)A}}{1 - e^{-\left(\sum F+M\right)}} \qquad (3)$$

The biomass is summed over discrete ages for simplicity to an age A where further growth is negligible and all fish can be combined into single "plus" group undifferentiated by size. Similarity with continuous recruitment can be improved by making the time units smaller. Fishing mortality is assumed linearly related to effort as for the logistic model (equation (1)). The unexploited biomass is found by setting $F_g = 0$ for all gears.
The yield is simply the catch equation summed over the age classes:

$$YPR_g = \frac{F_g}{F_g + M}\left(1 - e^{-\left(\sum_{g} F + M\right)}\right)B \quad (4)$$

Fishing mortality is assumed constant over age and size (knife edge selection). At equilibrium the total YPR remains constant and can be summed to infinity using the discount rate (as a series sum) over time.

This equation can also be adapted to a non-equilibrium system, where the fishing mortality regime has changed and the population is moving to a new equilibrium under the new regime. In this case, the initial population structure depends on the old equilibrium state (numbers in equation (3)):

$$N_a = e^{-\left(\sum_{g} F_{old} + M\right)a}$$

$$N_{a+t} = N_a\, e^{-\left(\sum_{g} F_{new} + M\right)t} \quad (5)$$

These equations can be substituted into the biomass and YPR population equations to get the incremental change in biomass and catch until the new equilibrium state is obtained. This approach includes an assessment of short term losses versus longer term gains often resulting from a decrease in fishing effort.

Unlike the logistic model, there is no pre-defined overfished state for yield-per-recruit biomass. In multispecies terms, YPR is carried out separately for each species. Clearly catchability, natural mortality and growth parameters are required for each species. Each species stock state is treated the same, so there is no discrimination between abundant and rare species. However, for the preference scoring, species can be weighted which takes account of their importance in the catches.

3.2.3 Effort Control

Within both models, controlling effort directly controls the fishing mortality. Fishing mortality measures the proportion of the stock which is being removed by fishing. Depending on productivity (measured as biomass growth), higher fishing mortality will depress stock biomass and reduce catch rates. Effort is also used as the basis of the preference score as it is well understood by fishers as related to costs in terms of labour, capital and operational costs. Although other controls are available, effort plays a central role in the simulation models and as such, is the preferred primary control, although others may play their part.

3.2.4 Quota Control

The catch quota control is applied as a future limit to catches. A new effort must also be supplied as the maximum effort. This is used to calculate catches. If catches exceed the quota, this maximum effort is scaled back to a level where the catches are met. This allows effort to change, but catches remain fixed if the effort is high enough to reach it and if the stock is not overfished. Setting the quota above the MSY means it will have no effect and the maximum effort control will apply.

3.2.5 Refuge Control

The refuge control (probably a closed area) indicates what proportion of the stock is protected from fishing. The control only applies to the logistic model. It is assumed that there is no adult migration between the protected and unprotected stock. Migration would reduce the effective refuge size. The two separate stocks are modelled independently. If there has been no previous refuge, both stocks will be at the same level. Once the control is applied the protected stock will rise to the unexploited level. The exploited stock will be subject to the new mortality based on a new effort level defined for this control.

The stock is initial split in proportion according to the control. The control splits the unexploited stock size and the recruitment between the refuge and exploited areas according to the control proportion.

$$R_{t+1} = R_t + \alpha r\left(R_t + B_t\right)\left(1 - \frac{R_t}{\alpha B_\infty}\right)$$

$$B_{t+1} = B_t + (1-\alpha) r\left(R_t + B_t\right)\left(1 - \frac{B_t}{(1-\alpha) B_\infty}\right) - C_t$$

$$C_{gt} = \frac{F_g}{\sum_g F_g}\left(1 - e^{-\sum_g F_g}\right) B_t$$

where R_t = refuge population, B_t = exploited population and a = proportion of the stock protected. Catch is only removed from the exploited population. This will result in an immediate decrease in catches after the control is introduced and effectively a decrease in catchability. There is a longer term gain in stock size as productivity is boosted by the refuge stock. As the model suggests, refuges are a good way to protect the stock and achieve the limit reference point. It is unlikely, however, that a target reference point above zero will be identified unless overfishing is already occurring. In general, an effort control will be better at achieving a target as it deals directly with economic issues.

Alternative models describing the effect of a closed area could be proposed. However, unless they limit the catch they will have no effect, so this loss cannot be avoided. Closed areas may also reduce effort and therefore costs, but this represents an indirect effort control. This is not necessarily optimised, and such a control may well not be universally popular or achieve fishers' objectives.

3.2.6 Projections

For each random set of parameters, the current effort is applied in the first year, to obtain a current catch. This generates a base line for comparison with the projection period. Thereafter the new control regime is applied and the population projected forward to obtain a time series of effort and catches. The effort and CPUE as a proportion of the current effort and CPUE is calculated. These data are used to calculated the preference for each fisher. The stock state is also recorded for each time period.

3.2.7 Reference Points

Target Reference Point

Indicators must be converted to measures of preference, so that risks can be properly assessed. For example, fishers may wish more to avoid low catches rather than make large catches, and hence be risk averse. This requires means converting indicators to some measure of utility (an economic measure of satisfaction).

The simulation model calculates the overall catch and effort for the fishery projection. These can

be converted to the relative change in CPUE and effort from the current CPUE and effort. These relative changes are assumed to apply equally to all fishers, so that if CPUE is 85% and effort 80% of the initial CPUE and effort, then the fishers CPUE is also 85% and 80% of his/her current CPUE and effort. The main assumption is that any effort or other control is applied proportionally to all fishers.

The optimum Bayesian decision is to choose the action that maximizes the expected preference. Using the preference data and model (see Preference), the discounted preference score can be summed for each simulation leading to a relative measure of how much preferred that outcome would be. The expected preference score is the average of the simulations where the simulation parameters are drawn at random from their posterior probability distribution.

The maximum is found by interpolating between the control increments using a fitted polynomial function. Finding the maximum by direct means would be very slow and produce an unnecessary degree of accuracy. If greater accuracy is required, the range of the control (minimum – maximum) can be reduced around the optimum point and/or the number of control increments can be increased.

Limit Reference Point

The limit reference point is designed to limit the chance of overfishing to some acceptable level. Overfishing is defined here as forcing the stock biomass below some limit state defined as the proportion of the unexploited biomass. The limit state may be set by the user, but is a generally excepted point for some models, most notably 50% for the logistic model. The probability is calculated as the chance that a scenario state taken at random from all scenario states combined over time, species and simulations, is below the limit state. This method, as well as working for the current simulations, will work with stochastic simulation models or under more complex management simulations. It could also be interpreted as the expected proportion of time that stocks will be spend in the overfished state under each management regime.

3.3 Probability Models

The ideas for the approach presented here originate with Press (1989), in which the author presented a method he used to estimate the probability of nuclear war. Nuclear war is similar to overfishing in that we do not want to have several observations before being able to estimate if and how it might occur. Press (1989) suggested using interviews with experts and kernel smoothers to generate a prior probability. This method was applied here to obtain a prior probability, but it was noticed that the approach can be easily extended to dealing with very many other sources of information.

Kernel smoothers provide the building block for probability density functions. Silverman (1985) provides a detailed description of the use of kernel smoothers in estimating densities in one dimension. This method has been adapted here to multiple dimensions.

3.3.1 Frequency Data

The probability distributions is modelled from frequency values. This is analogous to using frequency histograms. However, in contrast, the method allows small number of values to be used by smoothing the probability between points. This technique is known as kernel smoothing.

There are invariably several parameters for each stock assessment model. The parameter estimates are often correlated, so their probability distribution must be modelled simultaneously as a multidimensional function.

Once defined, several separate frequency distributions can combined to produce a final, complete probability distribution for the parameters, known in Bayesian statistics as a posterior distribution. Each frequency distribution may refer to one or more single parameters. For example, estimates

on current population size only may be available from population surveys.

Bayesian statistics demands a prior probability distribution representing belief. There have been considerable discussion on the appropriate prior distribution, and in the past a scientific constraint for objectivity has required any prior to have no influence on the final result. In management decision-making, it is recognised that subjective opinions should be accounted for, and good management often involves consultation. A problem has a arisen in marrying scientific assessment and subjective opinion and wishes. This methods addresses this problem.
The method can cope with missing data, but each series should have a minimum of 5 observations for each parameter.

3.3.2 Normal Kernel Smoothers

Given a set of frequency data, how can a probability density function be obtained? One option would be to fit a parametric distribution. This would require knowledge of the appropriate shape of the function. While in some cases we would be able to propose a function, such as the normal or log-normal, in many others it would not be possible. We would also run the risk of proposing an incorrect function and introducing structural error even if the distribution is parsimonious. Instead, a more general non-parametric technique using kernel smoothers is used.
Silverman (1986) provides details on kernel estimators for density functions. The basic aim is to estimate the probability density function from which the frequency has been drawn. There are two requirements. Firstly, a kernel function must be chosen. It has been shown that the particular choice of function is not particularly important in trying to estimate a density (Silverman 1986), so the function can be chosen more for convenience than mathematical requirements. The normal or Gaussian function was chosen for the current model for two reasons:

- The multivariate normal offers a simple way to calculate and maintain individual multidimensional kernel models through use of its covariance matrix. In particular, the posterior of a normal mixture can be calculated directly.
- Where very little data is available from interviews, for example, the normal distribution has a natural shape which it is assumed can represent an individual's subjective prior as well as building into a community density function once enough data are available.

Secondly, the method requires a smoothing parameter which controls the degree of spread of the density around each point in the frequency. This parameter is important. Not only does it change the look of the density, but it is a measure of the uncertainty associated with each point in the frequency and hence the frequency as a whole.
Each probability density function is represented by a smoothed probability distribution around a set of points. The points are either derived from interview, and represent the prior belief of interviewees (expert stakeholders / fishers) or derived from bootstraps from a fisheries model. Frequencies can be obtained by other means, but these are not supported by the PFSA software.
The discrete frequency data is smoothed by spreading the probability around each point using the normal kernel function (Figure 1).

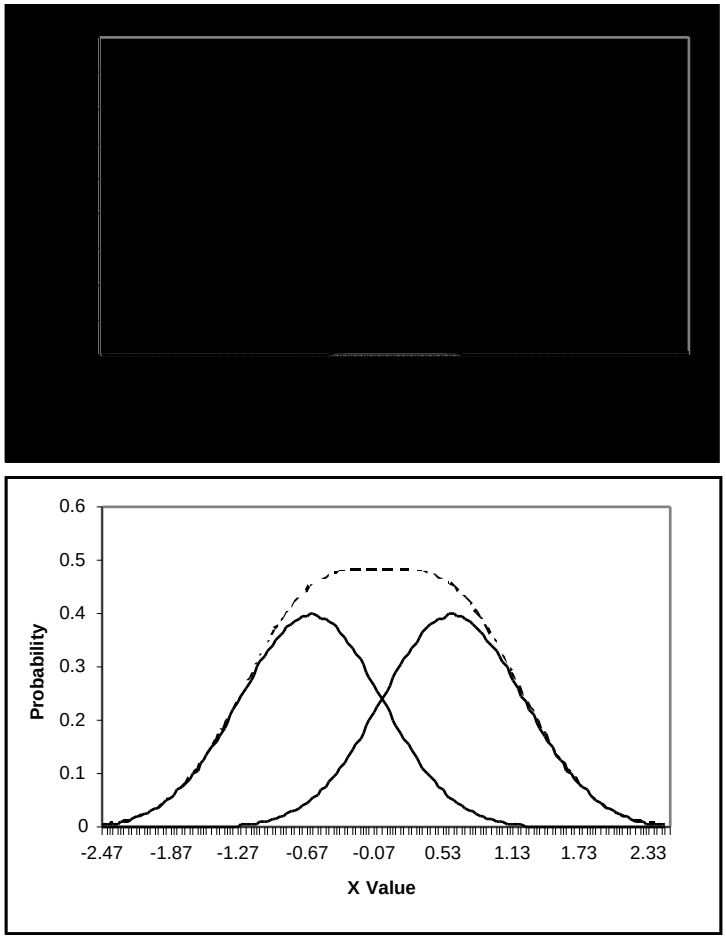


Figure 1 An example of two points forming a mixture distribution in one dimension. In the first, the smoothing parameter (Sigma parameter in the normal distribution) is relatively small and produces two modes. In the second, the smoothing is greater and a single flattened mode is produced.

3.3.3 Fitting the Smoothing Parameters

Estimating Parameter Frequency Covariance Matrix

It is clear from the above that the parameter frequency kernel covariance matrix is an important component of the posterior as it provides a weight for each parameter frequency. The covariance matrix is the fitted kernel smoothing matrix. The more heavily smoothed a frequency is, the lower weight it will have in the posterior.

The method applied is to reduce a multi-dimension frequency to a series of independent one dimension frequencies. Each of these can be smoothed separately, and then converted back to the original matrix. This can be achieved through linear transforms of the data. The transform chosen are the principle components which are a set of uncorrelated variables constructed from the original data.

The method is as follows:

1. Firstly the covariance matrix of the data is obtained by calculating the variances and covariances in the usual way.
2. Singular value decomposition (see Press *et al*. 1989) can then be used to reduce the covariance matrix into orthogonal matrices:

$$\Lambda = \mathbf{V}\ \mathbf{W}\ \mathbf{V}^{\mathbf{T}} \qquad (12)$$

3. **W** is the diagonal matrix containing the scaling terms for the independent PCA scores. The scores themselves can be calculated from the data and the linear terms in **V**. This is also particularly convenient because the inverse of the covariance matrix is simply the reciprocal of the scaling terms back into the equation.

$$\Lambda^{-1} = \mathbf{V}\ \mathbf{W}^{-1}\ \mathbf{V}^{\mathbf{T}} \quad (13)$$

4. The scaling values in the diagonal matrix **W** now become the smoothing parameters to be estimated. That is, the PCA score vector is calculated, the smoothing parameters for these scores are found and substituted for the relevant scale parameter in **W**. This is done for each PCA score vector (i.e. dimension). The smoothing matrix and its inverse can then be calculated from equations (12) and (13).

The square root of the covariance matrix is found using the square root method (Faddeeva, 1959) which works with positive symmetric matrices (i.e. covariance matrices).

The data are standardised using a mean and standard deviation calculated across all parameter frequencies to prevent numerical errors in the matrix routines. This has been found to work well. It eliminates scaling problems between parameters (e.g. q and B_{inf}) and differences among parameter frequencies should not be enough to create problems for a robust matrix decomposition routine unless there are significant incompatibility problems among data. As all data are scaled in the same way, the equations (7) to (11) defining the posterior distribution still apply and the linear scaling of the random variables can be easily reversed.

Estimating the Smoothing Parameters

The least-squares cross-validation method used for one dimension is described in detail in Silverman (1986). The idea is to find a smoothing parameter which minimizes the mean integrated square error between the estimated and true density. A score can be calculated using cross-validation, where each data point is removed in turn and the density from the reduced set becomes independent of the data point. Intuitively it can be seen that a good fit would be obtained by minimising the difference between the estimated densities and these independent values. In fact, the score is directly related to the error, so minimising the score minimises the squared error (see Silverman 1986). Where the number of frequency values is small, the least-squares score for the normal (Gaussian) kernel can be calculated directly as:

$$M_0(h) = \frac{1}{2\sqrt{\pi}\,n^2 h}\left(2\sum_i\sum_j Exp\left(-\frac{1}{4}\left(\frac{X_i - X_j}{h}\right)^2\right) + n\right) - \frac{4}{\sqrt{2\pi}n(n-1)h}\sum_i\sum_j Exp\left(-\frac{1}{2}\left(\frac{X_i - X_j}{h}\right)^2\right) \quad (14)$$

where X_i = the i^{th} data point, h = smoothing parameter, n = number of data points. For larger numbers of data (say, over 100), the score becomes time consuming (½ n (n-1) calculations). Instead of the direct score, an approximation is used which is close to M_0 for large samples. Again, the method is described in detail by Silverman (1986) and is based on the same score except n is substituted for (n-1) in equation (14) to create a score $M_1(h)$. In this form, fast Fourier transforms can be used to carry out the convolution between the data and the kernel. This reduces the score to a simpler exponential sum:

$$\left((b-a)^{-1} + M_1(h)\right) = (b-a)\sum_{l=1}^{M/2}\left(Exp\left(-\tfrac{1}{2}h^2 s_l^2\right) - 2\,Exp\left(-\tfrac{1}{2}h^2 s_l^2\right)\right)\left|Y_l\right|^2 + \frac{1}{\sqrt{2\pi}\,n\,h}$$

$$s_l = 2\pi l\,(b-a)^{-1} \quad (15)$$

where a and b are the interval range of the values, M is the number discrete transform components (an integer power of 2), h is the smoothing parameter and Y_l is the discrete Fourier transform of the discretized data. The data is discretized by allocating it to a fixed interval vector. The intervals are $\delta = (b-a)/M$, so that the k^{th} point has a value $t_k = a+k\delta$. For each data point X

lying between t_k and t_{k+1}, $n^{-1}\delta^{-2}(t_{k+1}-X)$ is added to k^{th} value and $n^{-1}\delta^{-2}(t_{k+1}-X)$ is added to the $k+1^{th}$ value. All points are added in this way. (Note: These weights are displayed in the software Plot | Plot Kernel graphs). The fast Fourier transform routine used is described in Press *et al.* (1989).

A parabolic interpolation method was used to find the minimum of the score (see Brent's procedure in Press *et al.* 1989). Once the minimum has been bracketed, the procedure finds it rapidly and does not require the differential of the function, making it more robust than many other methods. The start point for *h* is the normal distribution estimate: $1.06\, n^{-1/5}\sigma$ where σ^2 = PCA scale parameter. The start bracket is 0.25 and 1.5 the start value, which is extended to ensure it includes the minimum. To prevent degenerate behaviour, *h*, is given a lower limit of 10^{-3} of the standardised data.

Use of fast Fourier transforms makes fitting the kernels to even very large data sets very rapid. This technique is used in generating the stock state probability density functions, and only takes a small amount of time compared to the basic stock size and preference calculations.

3.3.4 Posterior Random Draws

The use of the multivariate normal kernel allows a relatively simple calculation of the posterior normal. Assuming equal weight to each point in each parameter frequency list, we choose a point at random in each list. These can then be combined to calculate a posterior kernel from which a random set of parameters can be drawn.

The multidimensional normal kernel function is given by:

$$K(\theta|\mu,\Lambda) = \frac{1}{(2\pi)^{d/2}|\Lambda|^{1/2}} \exp\left(-\frac{1}{2}(\theta-\mu)^T \Lambda^{-1}(\theta-\mu)\right) \tag{7}$$

The probability of a particular vector of values now depends upon the N points in the parameter frequency, as:

$$\Pr(\theta|X,\Lambda) = \frac{1}{(2\pi)^{d/2}|\Lambda|^{1/2} N} \sum_{i=1}^{N} \exp\left(-\frac{1}{2}(\theta-\chi_i)^T \Lambda^{-1}(\theta-\chi_i)\right) \tag{8}$$

where the covariance matrix (Λ) of dimensions *d* is chosen to smooth the density (see above). Combining the series to produce the posterior density is relatively simple. Given M densities, the posterior is given by:

$$\Pr(\theta|X_1,\Lambda_1,\cdots,X_M,\Lambda_M) \propto \prod_{j=1}^{M}\left[\sum_{i=1}^{N} \exp\left(-\frac{1}{2}(\theta-\chi_{ji})^T \Lambda_j^{-1}(\theta-\chi_{ji})\right)\right] \tag{9}$$

The probability is made up of a set of multinomial terms, each consisting of a unique combination of values taken from each of the density functions. A random term can be drawn from the posterior by choosing a random point from each density. These points combined from a posterior kernel with a mean and covariance based on the mixture property:

$$\Pr(\theta|\mu,\Lambda) \quad \propto \quad \exp\left(-\frac{1}{2}(\theta-\mu)^T \Lambda^{-1}(\theta-\mu)\right)$$

where

$$\mu = \left(\sum_{j=1}^{M} \Lambda_j^{-1}\right)^{-1} \quad \left(\sum_{j=1}^{M} \Lambda_j^{-1} X_{ji}\right)$$

$$\Lambda = \left(\sum_{j=1}^{M} \Lambda_j^{-1}\right)^{-1} \tag{10}$$

where X_{ij} is the i^{th} data point chosen at random from the parameter frequency *j*. The mean of the posterior is the weighted mean of set of individual vectors where the weights are the individual

inverse covariance matrix for each frequency. A random posterior parameter vector can be obtained by choosing a set of independent random normal variables of the same length as the vector, and applying the linear transform:

$$\rho = \mathrm{P}Z + \mu$$

where

$$\Lambda = \mathrm{P}^T\mathrm{P} \qquad (11)$$

So, P is equivalent to the square root of the posterior covariance matrix.

The posterior and the individual kernel covariance matrices and their inverses need be calculated only once before beginning a set of random draws. This makes the random draws themselves very fast.

Subsets of parameters can be dealt with separately. It is important to note that individual kernels do not have to have frequency data covering all parameters. The inverse covariance matrix for a subset of parameters has implicit rows and columns filled with zeros for those unrepresented parameters. These zeros indicate no information (infinite variance) and have no influence on the posterior.

Constraints

Parameters will often be constrained to particular ranges. Many population parameters will be constrained to positive values, and in many cases an upper bound is useful even if not strictly required. Parameter constraints are defined by the population models.

Constraints are dealt with by reflecting parameter estimates back into the valid region. Essentially, all the probability mass is conserved and gathered around the boundary which seems a reasonable representation. When choosing a random value, if it outside the boundary, it can be reflected back into the valid region, making the reflection algorithm very fast.

For each draw from the posterior distribution, the expected utility is calculated for each management action. This may involve complex calculations not only to get the output variables such as catch, but also to change this to the appropriate utility.

3.3.5 Summary

Given enough data in each frequency, the following standard method is applied:

1. All frequency data are scaled using the same global mean and standard deviation calculated from all frequencies.
2. For each frequency in the list:
    a. The data covariance is calculated from the data.
    b. The covariance matrix is decomposed using singular value decomposition that produces a set of scale parameters and uncorrelated PCA scores.
    c. For each PCA score
        i. The least-squares cross-validation score is either calculated directly if the number of frequency is small or on Fourier transformed data if the number of data is large. The data are discretized and a Fourier transform is applied carrying out the convolution between the data and the kernel.
        ii. The minimum least-squares cross-validation score is found by adjusting the PCA scale parameter using a standard minimisation routine.
3. The new rescaled covariance matrix and its inverse are recalculated from the decomposition with the new PCA scale values.
4. The overall inverse covariance matrix is found by summing the inverse matrix for each frequency. The covariance matrix is found inverting it. The square root of the matrix is found through a separate decomposition.
5. For each random parameter set required for the simulation:
    a. A sum vector is set to zero. For each frequency in the list, a point is taken from the frequency at random and multiplied by its inverse sigma matrix, then added to the sum vector. The resulting vector is the sum of the parameter values weighted by each inverse covariance matrix.

b. The sum vector is multiplied by the overall inverse matrix to get a random posterior mean point.
c. Independent random draws are made from the standard normal to fill a vector with the same length as the number parameters in the overall frequency. This is multiplied by the square root of the covariance matrix and added to the mean point to produce a random number draw from the posterior.

3.4 Source Models

Fitted models are structured as a linked hierarchy of sub-models. The structure allows greater flexibility, speeds up the fitting process and will allow easier development in future.
The basic structure is to have a multispecies model at the top level, if appropriate, the single species population models next and generalized linear models. There can be many species populations for each multispecies model and many generalized linear models for each single species model. The generalized linear models (GLM) link the population models to observations. The population models are more likely to be non-linear and more difficult to fit. By separating out the linear components, the overall model should become easier to fit. Furthermore, it should be easier to change a population model, for example, without changing the rest of the model structure, helping interaction with the software user, and easier to develop independently other models to fit in this hierarchy.
While the multispecies approach is new (see below), the separation of the single species model and GLM is a formal, more integrated approach of what is already commonly done (see Hilborn and Walters 1990; Hassen and Medley 2001). In many cases, a GLM is applied to observations to produce a population index. The population index is then used to fit the population model. While this pre-processing may be easier with some complex data sets, it introduces a redundant parameter and ignores possible parameter correlations with the population model.
The basic approach is to include the population size as a variable in the GLM. For any set of population parameters, the GLMs can be fitted to the population sizes. This is fast even if a GLM contains many parameters. A slower non-linear minimizer can then be used to minimize the fitted GLM log-likelihood with respect to the smaller number of population parameters.

3.4.1 Generalized Linear Models

McCullagh and Nelder (1989) provide a description of generalized linear models as implemented in the current software. GLMs consist of a linear predictor, link function and variance function. The link function describes the relationship between the mean and the linear predictor. The variance function depends on the error model chosen. With an identity link function and constant variance, the fit is standard least-squares. With other link functions and variance functions, the model must be re-weighted and fitted over a number of iterations. Under most circumstances, the number of iterations is small.
The least-squares estimates are found using singular value decomposition (SVD) to invert the information matrix (Press *et al*. 1989). SVD is slower than other methods, but robust. The iterative weights are calculated as described by McCullagh and Nelder (1989; page 40), who also give a justification.
Three links and errors are provided, although these can be easily expanded in future. The links represent the most commonly in fisheries.

Identity-normal: This is standard least squares regression.

Log - Poisson: This is the standard log-linear model. Linear terms are multiplicative. Independent variables can be linearized by taking their logarithms.

Complementary log-log - Poisson: This is the GLM form of the single gear catch equation:

$$\mu = (1 - \exp(\exp(\eta)))$$
$$\eta = \ln(\ln(1 - \mu)) \qquad (16)$$

where ì = expected value (probability a fish is caught) and η = linear predictor. The linear predictor is multiplicative, so log-effort is used as an offset (i.e. no parameter is fitted to it). The population size occurs as the binomial parameter, N, in this model, which most naturally would therefore use a binomial error. However, the variance function used is the Poisson, which is larger than the binomial variance. This is probably more suitable for over-dispersed data when N is being fitted rather than known, and when applying quasi-likelihood assumptions (see McCullagh and Nelder 1989 for a discussion of quasi-likelihood).

Currently linear predictors support a constant, covariates and an offset (variables with no parameter). Discrete factors are not supported yet. Although complex GLMs could be supported in future, the emphasis in the software is on simple single parameter GLMs, which are most likely for the fisheries data being considered.

3.4.2 Single Species Population Model

Logistic Model

The logistic model fitted to the data is the same as that used in the simulation model (see equation (1)). However the fitted catch-effort model is a GLM of the form:

$$\hat{C}_t = B_t\left(1-\exp\left(\exp\left(\ln(q)+\ln(f_t)\right)\right)\right) \quad (17)$$

This is fitted separately for each gear. The log-likelihood is the sum of the individual GLM log-likelihoods which are calculated for the Poisson and normal distributions as appropriate. Other GLMs can be added if population indices are available.

The three population parameters are fitted using the downhill simplex method of Nelder and Mead (1965; described in Press *et al.* 1989). While the method is slow, it was found to perform better on the logistic model than other methods (including Solver in MS Excel) even when the differential of the function was available. The logistic model can exhibit some difficult non-linear behaviour and a robust minimizer was preferred.

The parameters were given maximum and minimum limits to prevent unrealistic results. The initial population, B_0, is defined as the proportion of the unexploited size and therefore varies between 0 and 1.0. The intrinsic rate of increase produces erratic behaviour above 2.0. Estimates above 2.0 indicate a shorter time unit should be used. The unexploited biomass must be above the maximum observed total catch in any time period. Although the theoretical unexploited biomass could be infinite (or at least a small proportion of the mass of the earth), a limit was placed so that the maximum total catch would be no higher that 1% of the biomass. If catches do not discernibly decrease the resource size, the resource size estimate can become arbitrarily high. This is capped at a high level. If the estimate drifts to this level, the resource is hardly exploited at all, and this information is adequate for fisheries decision making. No boundaries are applied to the catchability parameter.

Table 1 Parameter limits for the fitted logistic model.

Parameter	Minimum	Maximum
B_0	0	1.0
r	0	2
B_∞	Max(Total Catch)	100* Max(Total Catch)

Linear Depletion Model

The simplest population model assumes a closed population with changes only coming about through catches (Leslie & Davis 1939):

$$N_t = N_0 - \sum_{j=0}^{t-1} C_j \quad (18)$$

where N_0 = the initial population size, N_t = population size in numbers on day t and C_j = the catch

on day j. Clearly, the initial population size must be greater than or equal to the sum of the catches.

Linear Depletion Model with Natural Mortality

The simple extension to the linear model allows for natural mortality as well as catches:

$$N_{t+1} = N_t e^{-M} - C_t e^{-M/2} \tag{19}$$

where M = the natural mortality rate. It is difficult to fit M in this model using pure maximum likelihood approach, so its practical usefulness is probably limited. No data was available to test it.

3.4.3 Multispecies Population Model

The obvious approach to modelling multispecies communities is to fit separate population models to each species with implicit fixed (natural mortality) or explicit variable species interactions. A significant problem with this approach is the large number of parameters which these models require when fitting to real data. Although this is not a theoretical a problem if sufficient data is available or the number of species being explicitly modelled is small, difficulties in data collection make such approaches impractical.

The most widely used approach to modelling communities has been to fit species abundance models. It has been demonstrated empirically that most, it not all, communities follow a consistent pattern (Magurran 1988). Species abundance models form the basis for the study and interpretation of species diversity and are often used to measure human impacts on species communities.

Previous methods to fit species abundance models have assumed the collection method of animals is not selective (e.g. Bulmer 1974). This is inadequate for many applications, including the analysis of species composition data in fisheries. In many cases, and particularly fisheries, it is the different species catchabilities that are most of interest.

Dynamic depletion models are an important class of models used in modelling fish populations and to estimate catchability. Depletion models require the number of individuals removed from population and an index proportional to the population size, both recorded over time. These models can be used to estimate current and past population sizes as well as catchability for single stocks (Hilborn & Walters 1992). A simple multi-species extension of depletion models allows multiple catchabilities to be estimated which would at least partially explain species composition. However a problem immediately arises in that, even if sufficient data is available, it is impossible to fit models where there is insufficient contrast (i.e. depletion) in the abundance index. This will be true for all species that are rarely caught, which may either be rare in the community or have a low catchability.

Estimates can be greatly improved if it is assumed a species population size is conditional on other species. Conditioning allows estimates for species having a good contrast to estimate catchability and initial population size to improve estimates of catchability in other species where depletion is not so clear. This is reasonable if there is some foundation for the observed abundance patterns in ecology or evolution. Most of these models are justified on the division of niche space (May 1975, Sugihara 1980), but agreement is not universal, particularly over the application of the log-normal (Ugland & Gray 1982).

For the current analysis the broken-stick and log-series abundance models were used, although other models such as the geometric or log-normal could equally be applied. These four models have been found to fit the widest variety of communities (Magurran 1988). The broken-stick model is appropriate where a single resource is being shared more or less evenly between species, and has most commonly been observed in narrowly defined communities of taxonomically related organisms (May 1975). The log-series (and in its deterministic version, the geometric series) is appropriate where fractions of the available resource have been pre-empted by species in sequence. The log-series has been most commonly observed where one factor dominates the ecology of the community, and can also be seen in small samples where only the commonest species of the log-normal are represented. These two models represent extreme cases in terms of evenness and the distribution of a resource among members of a community.

Multispecies Population Models

The simplest depletion model assumes a closed population with changes only coming about through catches (Leslie & Davis 1939). The multispecies form of this model is:

$$N_{it} = N_{i0} - \sum_{j=0}^{t-1} C_{ij} \quad (20)$$

where N_{i0} = the initial population size of species i, N_{it} = population size in numbers on day t and C_{ij} = the catch on day j. Using Equation (20) we can generate a set of population sizes (N_{it}) over T days with a set of i input parameters (N_{i0}) and the catch data (C_{ij}). We assume some community model for the initial population sizes.

For the broken-stick model, the number of individuals in the r^{th} most abundant of S species is defined as:

$$N_{r0} = \frac{N_T}{S} \sum_{n=r}^{S} \frac{1}{n} \quad (21)$$

where S = the number of species and N_T = total number of individuals of all species in the community.

For the geometric series, a species in rank r will have a population size N_{i0}, defined as:

$$N_{r0} = N_{00} \beta^{r} \quad (22)$$

where the β parameter is always less than or equal to 1.0.

Equations (21)-(22) can be used to provide the initial population size in Equation (20). Where the rank of a species is known, the joint likelihood can be calculated and all parameters fitted using normal methods. The problem is that ranks of species are not known and all species-rank permutations need to be considered.

Fitting the Model

Assuming the species abundances are independent, the parameter likelihood of a set of observed catches of S species can be estimated as the joint likelihood between the species abundance model and the population model. This assumes all S species have been drawn at random from the species abundance model, and does not prevent two species occupying the same rank. This is not a problem if all species are assumed to be equally catchable as the initial species abundances are forced to fit to the curve with the greatest likelihood across all ranks. This is the standard approach. If catchability is allowed to vary however, all species will tend to be mapped to the most probable rank. The way to address this is to allow only one species in each rank.

To model the dependence between species, the species abundance model is separated from the likelihood model. The species abundance model defines the unexploited stock size for each species rank, which can be used in the likelihood model. If the rank of each species is known *a priori*, the model is easy to fit through normal methods. However in practice each species rank would not be known and all possible ranks for each species need to be considered.

Calculating the likelihood of all species-rank permutations is not possible when the number of species is of any reasonable size, the number of permutations being the factorial of the number of species ($S!$). However the problem can be reduced to a combinatorial problem. For any set of model parameters, a likelihood matrix can be calculated with rows representing the species and columns the ranks, so that the likelihood of species i being in rank r can be found in the matrix cell x_{ir}. The problem then is to fill each rank with one species. Once a rank is filled by a species, that column and row is eliminated and the next species-rank must be chosen from the remaining reduced matrix. Each time this is done, the likelihood in the matrix cell can build up the product for this combination. This reduces the problem to one of combinations rather than permutations. The sum of all these likelihood combinations is known as the permanent of the matrix, which unfortunately has no simple method of calculation (van Lint and Wilson 2001).

The combinatorial likelihood was calculated using dynamic programming using a tree structure to process the matrix. The process state is defined as the filled ranks in the species abundance model. Equivalent state likelihoods are added together, so that at any stage a state is all filled rank combinations of the tested species. There are two advantages of calculating the permanent in this way.

Firstly, impossible species-rank combinations can be eliminated early in the process greatly reducing the number of combinations which have to be calculated. For example, the obvious condition that the total catch cannot exceed the initial population size means that the largest ranks must probably be filled by the most abundant species in the catches. The wide variation in species catches usual for a multi-species catch should considerably speed up the likelihood function evaluation.
Secondly, the method allows fast fitting of species specific parameters, such as catchability. The dynamic programming method allows the process to stop before the last empty place is filled. At this point there is one species left and as many states as there are species. Each state has one empty rank which will be filled by the remaining species and an associated likelihood representing all species-rank combinations for the filled ranks. The parameters associated only with this species can be fitted using the likelihoods as weights. Using a generalized linear model regression framework, maximum likelihood parameter fitting can become very fast as this is only an extension of weighted least-squares. The information matrix can be calculated from the sum-of-squares weighted by each rank likelihood. Furthermore, by backing up and working down through the process to work through each species, lower states need not be recalculated every time, again reducing computation. In practice, the estimates converge quickly.

Likelihood Matrix
The population size (Equation (20)) needs to be connected to the observed catches through a likelihood model. To deal with rare species where the population can be small, zero catches have to be accounted for. The Poisson likelihood is used as it parsimonious as well as allowing for discrete catch numbers and zero catches. The log-likelihood for a set of observed catches, C_{it}, of species i over T days is:

$$L_{ir} = \sum_{t=1}^{T} C_{it} \ln \mu_{irt} - \mu_{irt} - \ln \Gamma(C_{it}+1) \qquad (23)$$

The expected catches, μ_{it}, will depend upon the species abundance model for the initial population size:

$$\mu_{irt} = \left(N_{r0} - \sum_{j=0}^{t-1} C_{jt} \right) \left(1 - e^{-q_i f_t}\right) \qquad (24)$$

Because the error models are the same for both species abundance models, the log-likelihood can be used as a comparative goodness-of-fit statistic. However, the geometric requires two parameters, the broken stick only one. The advantage of the geometric is there is no need to know the numbers of species in the model, otherwise a veil-line accounting for unseen species may be required (Magurran 1988).
Restating this in generalized linear model terms, the population size can represent the binomial trials and the remaining catch term the complementary log-log function. In the latter case, the log-catchability is fitted as the constant with the log-effort as an offset in the linear predictor.
McCullagh and Nelder (1989) describe the weighted least-squares regression procedure for fitting these models. This is extended by using the additional likelihood weights when summing squares over ranks. Using generalized linear models also allows simple extensions to more parameters, other link functions, error models and quasi-likelihood assumptions.

3.4.4 Stock Assessment Interview

An example interview including the stock assessment component, is presented in Section 4.5 (The PFSA Interview Technique) and in Appendix 1. This includes notes on the meaning and interpretation of the different questions.
The time, catch and effort units need to be identified and used consistently for all interviews. This applies both to the stock assessment and preference components. If a fisher is more comfortable with different units, you will need to convert his answers. Units should identify those most easily understood by most of the interviewees. For example, a month may be better than a year in terms of assessing catch or effort.

- Identify the fisher's main gear, then last years CPUE (qB_{t-1}) and this year's CPUE (qB_t) for this gear.

- The current catch rates for all other gears used ($CPUE_i$).

- A catch rate range for the unexploited stock (U_l, U_h).
- The time for recovery (T).

The total effort in this fishery over the last year (f_{t-1}) and any other catches not accounted for in the fishery (L_{t-1}) have to be obtained from elsewhere.
The individual catch rates are regressed towards the mean of the sample. This is necessary as they are used as an estimate for the mean catch rate in the fishery although the question asks for the fisher's own catch rate. For the j^{th} fisher:

$$[\hat{q}B_t]_j = \left([qB_t]_j + \left(\sqrt{N}-1\right)\overline{qB_t}\right)\Big/\sqrt{N}$$

$$\text{where } \overline{qB_t} = \text{mean catch of the sample} \qquad (26)$$

These values can be used to calculate the parameters for each fisher based on the logistic population model. The intrinsic rate of increase (r) can be calculated by solving the non-linear projection equation for the unknown r:

$$X_1 = X_0\left(1+r\left(1-X_0\right)\right) \cdots X_T = X_{T-1}\left(1+r\left(1-X_{T-1}\right)\right)$$

$$X_0 = \frac{\hat{q}B_t}{\hat{q}B_\infty},\ X_T = \frac{U_l}{\hat{q}B_\infty},\quad \text{and } \hat{q}B_\infty = \frac{U_l+U_k}{2} \qquad (27)$$

X_0 is the current stock state, defined as B_{now} in the logistic equation.
With r defined, catchability can be estimated from the current catch rate and effort adjusted for stock change due to production and catch:

$$\hat{q} = \left(\frac{\left(\hat{q}B_{t-1}-\hat{q}B_t\right)}{S}+r\,\hat{q}B_{t-1}\left(1-\frac{\hat{q}B_{t-1}}{\hat{q}B_\infty}\right)\right)\Big/ f_{t-1}\hat{q}B_{t-1} \qquad (28)$$

This assumes a linear relationship between catch and effort, but should be a adequate approximation unless fishing mortality is high. The time S allows the time unit to be altered. For example, converting from a year to a month S is set to 12. This allows r to be rescaled between 0

and 2.0. Given the fisher's main gear catchability, , the unexploited stock size and

other gear catchabilities can be found.

$$B_\infty = \frac{\hat{q}B_\infty}{\hat{q}}$$

$$\hat{q}_i = \hat{q}\ \frac{CPUE_i}{\hat{q}B_t} \qquad (29)$$

If equilibrium is assumed, last year and this year's catch rates are the same. This leads to a simpler equation (28), but doesn't affect the draws from the bootstrap which will still allow non-equilibrium estimates.

3.4.5 Empirical Bootstraps

Empirical bootstraps were used to generate parameter frequencies. Press (1989) proposed bootstrapping as a robust technique to generate a non-parametric likelihood, although such approaches do not gain universal support from Bayesian statisticians (Gelman *et al*. 1995). For example, bootstrap estimates are not independent estimates, so they can only approximate the true likelihood and are invalid with small sample sizes. If the parametric likelihood is known, it will provide more accurate estimates than its non-parametric counterpart. Despite their problems, such non-parametric methods are still useful in an automated software system as they do not require the user to propose a parametric likelihood for their data and eliminate the potential error in making a poor choice. As the results are only dependent on the data, the results reflect the data quality rather than the choice of error model. In this sense the results are more robust.

The method is simplified because data fits are limited to the generalized linear models, hence the technique only has to be applied to them. The basic method is to use randomized residuals (Manly 1997). The standardized residuals are calculated as:

$$R_i = \frac{Y_i - \mu_i}{\sqrt{V_i}} \qquad (25)$$

using the i^{th} dependent variable observation (Y_i), the model best estimate (μ_i) and variance function (V_i) from the GLM. A new bootstrapped vector of dependent variable data (B_i) is created by adding a random residual (sampled with replacement) to the estimated value:

$$B_i = \mu_i + R_j \sqrt{V_i} \qquad (26)$$

Fitting the model to these bootstrapped data (B_i) produces bootstrap estimates. As long as the number of residuals is large, a large number of smoothed bootstrap estimates should be a reasonable approximation to the likelihood. Otherwise, it is still a measure of uncertainty, but more formal claims cannot be made.

The dependent probability model uses linear models to regress observed parameter frequencies dependent on a common set of independent variables. This uses GLMs described in the way already described. The only difference in the way the bootstraps are carried out. A set of GLMs will share the same independent covariates, but be fitted to separate parameters as dependent variables in the same record. The residuals between these models might be correlated as the dependent variables themselves may be correlated. Therefore the bootstrap residuals are random, but taken from the same records for all the models. That is, if a residual is selected from the 3rd data record for the first GLM, it selected from the 3rd record for all other GLMs. This means the bootstrapped estimated should, as far as possible, maintain correlations derived from the original data.

3.5 Preference

Although utility is a fundamental economic variable, it is difficult to obtain. Utility recognises that the value of any economic resource, including money, is not simply proportional to the amount you have. For example, a dollar is worth less to a rich person than a poor person. If a person's monthly rent is $50, then $49 on the day rent is due is worth much less than $50. The problem with utility is it changes from person to person and from time to time. However, any analysis will assume some form for the utility function, and usually in bioeconomics it is assumed the relationship is linear with profit.

For poorer fishers, a linear relationship with earnings may well result in poor decision-making. No account is taken of alternative employment (opportunity cost), fish caught for food, minimum income required by the household and so on. Any analysis trying to account for all these socioeconomic factors would become complex, expensive and impossible to implement.

To take account of preference, key comparisons between possible outcomes can be compared by interviews and their relative preference can be scored. Simultaneous comparisons between more than two possibilities are difficult. However, once a fisher understands the full implication of two different scenarios in the fishery, they should be able to qualitatively say which they prefer and whether that preference is very strong, very weak or somewhere between the two.

As well as obtain information from fishers on the requirements from the fishery, the interviews can be used to help fishers focus on the key issue of sustainability and what it means to them in terms of their own livelihoods.
Key variables which management should control and which are relevant to fishers are catch (e.g. annual earnings), effort (e.g. number of days they fish) and catch rate (e.g. the amount of fish they catch each day). Discussions with fishers should focus on these variables.

As well as preference, data is required to scale the effects of the total fishery on the individual fisher. This is done by scaling the changes in effort and catch rate as proportional changes to the current effort and catch rates which are available for both the whole fishery and each individual fisher. The current catch, effort and catch rates are used as reference points for the fisher to compare with alternative scenarios. There are also constraints to consider, such as the maximum and minimum catches which a fisher will take on a days fishing.

3.5.1 Scenarios

Scenarios represent possible changes in the catch and effort as they relate to the fisher. Changes are represented as +/-25% steps relative to the present and are constructed to maximise the information obtained for an information matrix. The scenarios, which were given a letter for easy identification, can be laid out in relation to the current catch and effort (I). Scenario C represents the worst case (lowest catch for the highest effort) and Scenario A the best case (highest catch for lowest effort).

Figure 2 The different scenarios used to assess fisher preference. The central scenario I represents the current situation with 4 fish and 4 boats representing the current catch and effort respectively. Effort and catch is decreased by 25% and 50% around this current value.

One scenario will dominate another where it is clearly better. If we assume higher catches are always better and higher effort always worse, any scenario where the catch is higher than or equal and effort is lower than or equal to another scenario will always be preferred. For example, **O** will always be preferred to **I**, as catch is higher and effort is the same. These dominance relationships can be used to rank all 17 scenarios more rapidly with the fewest number of comparisons. A represents the best, and C the worst scenarios, so it is only necessary to map all other scenarios between these two.

Scenarios can be ranked using a binary tree. The tree starts with seven scenarios already ranked according to the dominance relationships. Furthermore, scenarios may not need to be added at the apex, but further down applying dominance rules. Also as the tree nodes are completed, the rules can be applied to aid placement.

A near-linear (price-cost ratio type) score is obtained by the following order and scoring:

Rank	Scenario	Score
1	A	1
2	B	0
3	J	0
4	N	1
5	K	0
6	O	1
7	E	0
8	F	0
9	I	0
10	H	0
11	G	1
12	M	0
13	Q	1
14	L	0
15	D	0
16	P	1
17	C	

A completely linear score is not possible because scores can only be entered as integers. The price cost ratio can be used instead.

3.5.2 Preference Model

The additive nature of the scoring technique suggests a quadratic model of each variable with a single interaction term should be adequate in modelling the score. The model interpolates the score and smoothes through errors. Pure interpolation is too sensitive to errors. Even the ranking was not as reliable as was originally hoped (see below).
The simplest model to fit to the preference score is the quadratic equation:

$$1+U_1 = a_{10}+a_{11}x_1+a_{12}x_1^2 \tag{31}$$

The units are arbitrary, and the model can be scaled to any value. With two variables, and assuming utility independence, the model expands to:

$$(1+U_1)(1+U_2) = (a_{10}+a_{11}x_1+a_{12}x_1^2)(a_{20}+a_{21}x_2+a_{22}x_2^2) \tag{32}$$

This potentially has 8 parameters:

$$(1+U_1)(1+U_2) = \alpha_0+\alpha_1x_1+\alpha_2x_2+\alpha_3x_1^2+\alpha_4x_2^2 + \alpha_5x_1x_2+\alpha_6x_2x_1^2+\alpha_7x_1x_2^2+\alpha_8x_1^2x_2^2 \tag{33}$$

As the scores are arbitrarily scaled, so á$_0$ can be set to zero being the lowest scores. In practice, it would be difficult to fit all the remaining parameters to real data. They will be intrinsically correlated as they are fitted to same variables, albeit transformed. Therefore it seems sensible to focus on the lower order parameters, but allow at least one interaction term. Hence the last 3 terms (á$_6$- á$_8$) were not fitted (assumed to be zero). Further research may indicate more parsimonious or better representation of this utility curve.
The fishers current catch and current effort in the preference model is set to 1.0. So the scenario **I** is (1.0,1.0), scenario **G** is (0.5,0.5) and so on. The relative catch and effort for the fishery compared to the present can be calculated from the simulation model. This relative change is assumed the same for fishers. Given the overall catch and effort is set as c_t and f_t as proportions of the current catch effort at time t respectively, the fisher's score becomes:

$$U_t = \alpha_1c_t+\alpha_2f_t+\alpha_3c_t^2+\alpha_4f_t^2+\alpha_5c_tf_t \tag{34}$$

where the parameters are estimated from a least-squares fit to the scenario scores.
Where there are more than one species, the change in the overall catch (c_t) is calculated as the weighted average of the changes in individual species. The more important a species is to a

fisher the higher the weight. These weights could be the current proportion that each species makes up of the total catch or the catch value, or based on a preference score obtained in a similar way to the scenarios.

Price Cost Ratio

As an alternative to the interview preference, a simple linear price-cost function is provided. The global price-cost ratio function requires a single Price : Cost Ratio parameter (PCR) which weights the proportion change in catch relative to the proportional change in effort from the current situation such that:

$$U_t = \alpha c_t - f_t \quad (35)$$

If the score is proportional to profit, the weight might be calculated as the current value of the catch divided by the current catching cost: PCR = α = (Price*Catch)/(Effort*Cost). Clearly, the higher the PCR value, the more important changes in catch are to changes in effort. The default value is 1.0, so, for example a 10% increase in catch coupled with a 10% increase in effort will be viewed just as good as no change in either. The function is provided mainly as exploratory tool to allow some analysis before interviews are completed. Alternatively, even without other interviews being conducted, a user can conduct the full preference interview to obtain some reasonable preference curve.

Calculating the Discounted Preference Score

Given a time series of projected catch and effort changes (c_t, f_t), the time series of preferences can be obtained. The discounted mean preference score is calculated as:

$$U = \left(\sum_{i=1}^{n} \sum_{t=0}^{T-1} P_i U_{it} e^{-\delta t} + \frac{U_{iT} e^{-\delta T}}{1 - e^{-\delta}} \right) \Big/ n \quad (36)$$

where U_{it} is the preference score of fisher i at time t, P_i = the fishers importance (if used) and δ = discount rate. Importance weights a fisher's score, and could represent the importance of the fishery to his/her household income and the size of the household. The discount rate can be obtained for each fisher, or a global discount can be used. Note the sum only has to be continued until an equilibrium state is attained (i.e. c_t and f_t no longer change) at some time T, where after the infinite sum can be calculated. The mean score is the total divided by the number of fishers (n).

The target reference point is found by maximising this mean preference score.

3.5.3 Interview Design

Interviews are designed to obtain information on stock assessment and preference for each fisher. You should look at an example questionnaire and other notes before attempting interviews. The following information covers some elements of preparation and the questions you will need to ask. In particular, it will be necessary to rank and score the scenarios. There is a scenario binary tree to help with this task.

Identify the fishery and the fishers

The first step when undertaking the PFSA interviews is to identify the fishery of interest (reef fishery, pelagic fishery, long line, trap, gillnet etc) and deal only with the single fishery during the course of the interviews. Discussion involving other fisheries would only confuse matters.

Introducing the PFSA and encouraging participation

The PFSA interviews could be conducted by arriving at a landing site and beginning with available fishers. However, experience has shown that introducing the technique to fishers before conducting the data collection may ensure better participation and fit more within the framework of co-management which PFSA should help to promote. This can be achieved through village meetings or by involving PFSA in pre-arranged events such as workshops which may already take place. Organising a village meeting will vary between locations, though there is usually a

local protocol for establishing such events such as visiting the village She ha, fisheries officer, spokes person etc. This will allow a time and location for the introduction to be set and should ensure good participation.

Initial interviews

An important consideration when using the PFSA technique is to trial the method before collecting field data. The interview method can be learnt quickly, and conducting some preliminary trial interviews will aid the interviewer to establish their interview manner and identify potentially 'leading' or 'biased' presentation. Ideal subjects include persons with prior experience of the fishery in which you are interested (ex-fishers, fisheries officers, researchers etc).

During this period the researcher can also identify ways of increasing the rate of data collection by involving persons who have experience in the fishery. If the researcher does not have strong links with the fishing community, identifying key informants such as fisheries officers, beach recorders or a village spokes person/head-man to aid introductions to fishers. This can rapidly increase the number of fishers who will agree to participate, and will also reduce the time needed for locating individual fishers and thus the total time needed for the data collection phase of the project.

Other considerations at this point may include any logistics. In some locations fishers can be located on foot and are found in concentrations at landing sites or simply within the community. However, there may also be situations where fishers are more widely dispersed (particularly in artisanal fisheries) and less accessible. This may require that transportation be factored into the research programme as well as the additional costs that this may infer to the overall budget. Or there may be specific seasonal windows when data maybe most rapidly collected. This may include periods of rough weather or lunar phase when fishing activity is reduced and more fishers will be readily available for interviewing.

Interviews can also be aided by developing a list of all fishers in the fishery. This provides valuable information on how many fishers there actually are, will provide some indication of how many you intend to interview, whilst aiding fisheries officers or fishers to suggest who the next person to interview at a particular time should be.

The Interviews

Once a meeting or at least introduction to the work you plan to undertake has occurred, and the interviewer is happy with presenting the questions and recording the data then the interview phase proper can begin. Opening interviews may still be part of the learning process until the technique is completely familiar, though the researcher should be confident in the data collected by this point. If specific time periods for the interviews have been agreed with fishers then these should be adhered to.

Data can now be collected intensively by spending time in the field over a set time period (as maybe the case if the data collection is undertaken by a visiting researcher), or extended over a longer time frame if the researcher is a resident fisheries officer or similar. Ideally time in the field will be maximised so as to complete the data collection rapidly as part of the rapid stock assessment technique. Typically a researcher will visit an area daily and complete as many interviews as can be undertaken during a day. The number completed can be expected to vary depending of fisher availability. The location of the interviews may also vary. Ideally a fixed location including a desk will best serve data collection, but only where fishers are readily available. More often the researcher will conduct the interview after locating the fisher, with the interview taking place in a house or at some communal gathering point.

1.1.2 Interview Questions

The aim of the questionnaire is to extract from the fisher his/her view on the state of the stock, its productivity and preferences with respect to catch and effort and catch composition. The interview represents the core questions for developing prior probabilities and preference scoring for stock assessment. Additional questions could be added for other purposes, however the current questionnaire is already a considerable undertaking and additional questions would probably best form part of a separate interview. Most information is obtained indirectly. Direct questions, such as 'Do you think the stock is overfished', suffer not only from potential political bias, but also have an unclear meaning. However, indirect questions could lead to over-interpretation from the fishers' point-of-view. Care is needed in presenting the results of the analysis and in discussing their meaning.

Questions apply to one fishery only. Separate questionnaires should be conducted for each

fishery, although some data, such as preference information, may need to be only collected once. The following section introduces each question contained within the interview, the purpose of each question, and how the question may be presented by a researcher. Examples of question presentation and some additional alternatives are given where necessary, though some are considered straightforward. It may be useful for the researcher to have a copy of the actual interview to hand to aid this exercise. Key words are highlighted where necessary.

B.i Background

Q18) Including you, how many people are in your household?
Purpose: This should indicate all dependents on the fisher. This can be used in weighting the preference.
Presentation: Self explanatory
Q19) What proportion of your household income depends on your catch from this fishery?
Purpose: This should indicate the fisher's contribution from this fishery as a proportion of the household income. Income to the household from other people or from other fisheries must not be included in this proportion, only in the whole. This can be used in weighting the preference.
Presentation: Straightforward, though it can be beneficial to determine what other sources of income to the household exist through additional conversation.

B.ii Discounting

Q20) If you use an interest paying deposit account in the bank for your savings, what annual interest is paid?
Purpose: The discount rate is related to bank interest rates, loan rates and so on. However these are bound up with issues such as money supply, risk and other non-local effects. Although the bank rate can be used as an indicator of discount, it may be quite different to the true discount rate of the fishing community. Many fishers don't use bank accounts, or may not know the rate of interest.
Presentation: self explanatory

Q21) What is the time delay indifference point between current 1 month earnings now and 1 month earnings + 20%:
Purpose: This question aims to estimate the fisher's discount rate. The discount rate indicates the rate at which the future is devalued. Nobody realistically takes account in their day to day living of what will happen in thousands of years, and few of us take much account of what will happen beyond the next twenty years. Discounting is a simple way to adjust future values to represent more realistic estimates of true values. The discount rate is related to bank interest rates, loan rates and so on. However these are bound up with issues such as money supply, risk and other non-local effects. Although the bank rate can be used as an indicator of discount, it may be quite different to the true discount rate of the fishing community. It is therefore better, if a reliable method can be found, to obtain the discount rate from the fishers themselves.
Presentation: To obtain an estimate of a person's discount rate, it is necessary to separate it from other issues. In particular, in testing for indifference between to outcomes, only the time delay should vary, rather than the two scenarios being compared. This prevents the comparison being confounded with utility.

For example, a simple question would be: Which would you prefer more, $100 now or $120 in 1 year's time. If the interviewee prefers $120 in 1 year, the delay should be increased and the preference obtained until the approximate indifference point is identified. This can most easily and quickly be found by bracketing the point and repeated bisection (see box).
It was found in tests that the simple question posed above without further information did not work. Fishers found it difficult to think abstractly, so answers could be quite wild as they were interpreting the comparison in different ways. It is much better to find some activity which they actually do, such as saving schemes, and define two schemes which have a fixed quantified difference in payout which does not vary over time. By looking for the indifference point between schemes by varying the delay of the payout, the discount rate can be defined (see box).

B.iii Catch and Effort Preference

The catch and effort set consists of various scenarios representing the effort applied and catch obtained within the defined unit time.

The time unit is important as preference will vary with the time chosen. For example a fisher may prefer a high catch rate, but probably not if this was achieved by limiting his effort to one day a month. The time unit should be no less than a week, and no more than a year. In general, a month is probably the best measure as it allows more variability in effort and catch, but a unit should be chosen with which the fisher feels comfortable.

As in the discounting question, some level of abstraction is necessary to avoid fishers getting bogged down in the minutiae of fishing. Comparisons are always made with current practice and catch, including degree of variability. However, fishers will need to ignore the constraints, as these are taken into account elsewhere. For example, if a fisher cannot undertake more effort because of weather, we are still interested in his preference for doing so if this constraint was removed. This is because the preference for impractical scenarios still has an influence on the shape of the preference curve within the feasible region.

There are 17 scenarios with different levels of catch and effort measured as a difference from the current catch and effort levels for each fisher. The various catch scenarios are firstly ranked for preference. Then the relative scores between scenarios are recorded depending on how much one is liked over another. Scenario I represents the fishers current catch and effort.

Ranking the 17 scenarios is most quickly done using the binary tree. After comparing two scenarios, if the non-tree scenario is preferred it goes down the left branch and is compared with the next scenario in line, or if is less preferred it goes down the right branch. Comparisons continue until a free place in the tree is found.

The start points for each scenario in the tree are illustrated in the diagram. Only scenarios **E**, **G**, **F** and **H** could be compared to the current situation (scenario I). Scenario **B** starts with **N**; **J** and **K** with **O**; **M** and **L** with **Q**; and **D** with **P**.

In fact, scenarios **E**, **F**, **G** and **H** should all be worse options than the current situation unless there are constraints. For example, if the fisher prefers **G** to **I**, there is nothing stopping him reducing his effort and making scenario **G** his current option. He might not be able to do the same with scenarios **E** and **F** as his effort may be constrained by weather, availability and so on. So, although his preference should be for scenario **I** on all these initial comparisons, it is worth checking this first to ensure the fisher understands what is required of him.

It is important to note that some scenarios are dominated by others and comparisons need not be sought from fishers unless to check his/her understanding of what is required. For example, a fisher should clearly prefer any scenario where he catches more fish for the same amount of effort. The ranking can be speeded up by recognising dominance when it occurs.

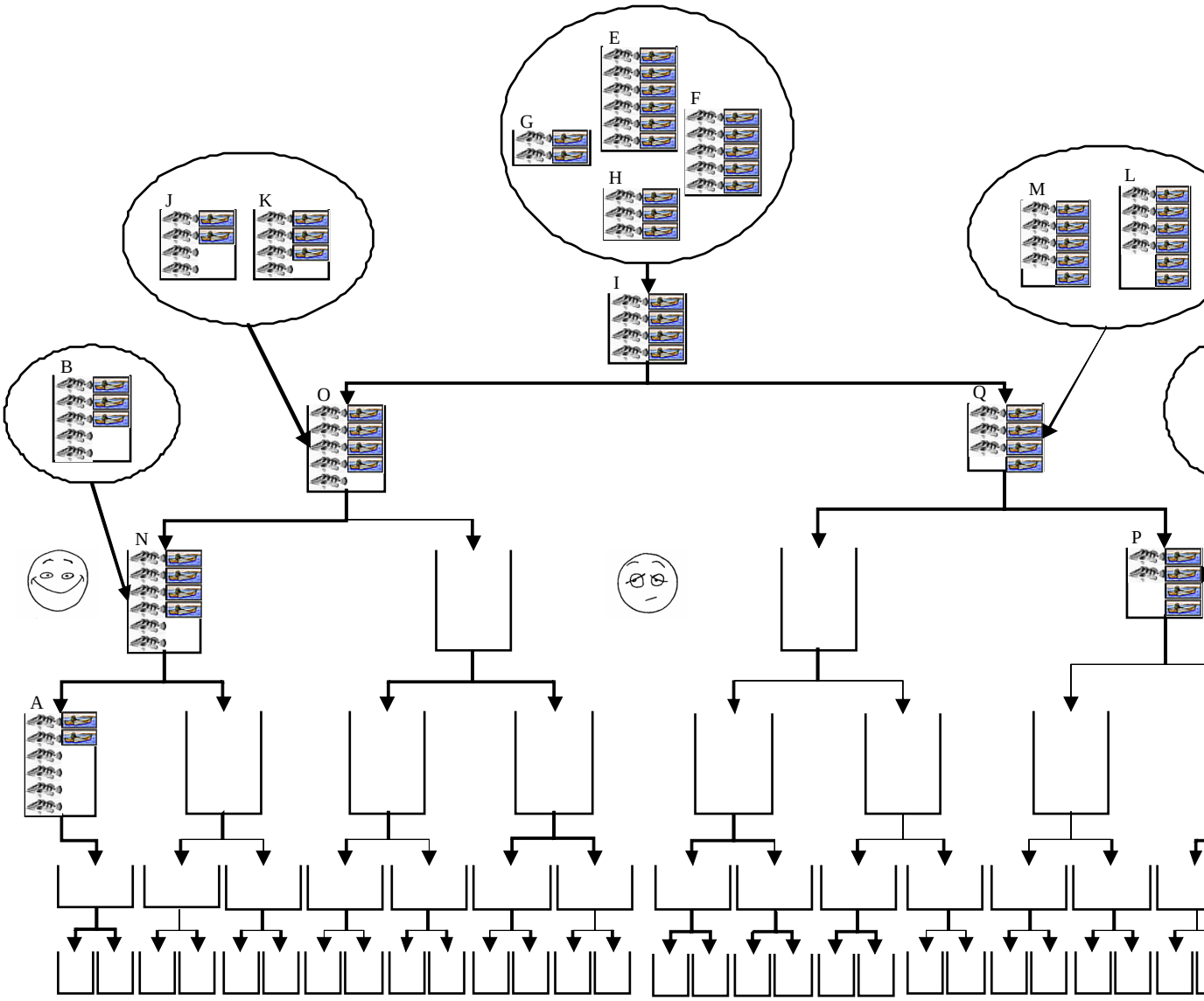
The binary tree only serves to aid ranking and has no other purpose.

Once all scenarios have been entered in the tree, the scenarios can be scored. During scoring it is worth confirming the rank order as with more thought a fisher may well change his mind. These are difficult questions that require consideration of many issues.

Scoring allows the fisher to indicate the degree of difference in preference between scenarios. It is quite possible that fishers are indifferent among some scenarios and have a strong preference among others within the ranking sequence. When ranking it should be made clear that they will have this opportunity. Therefore, they need not spend time ordering scenarios that they are essentially indifferent between.

3.5.3.1 Scenario Tree

Scenarios can be ranked most efficiently using a binary tree. Scenarios are ranked using pairwise comparisons and placed in the relevant tree position. If they are better they move to the right. If they are worse they move to the left. They are placed in the first free space.



3.6 References and Further Reading

Aiken, K.A., Kong, G.A., Smikle, S., Mahon, R. and Appledorn, R. (1999). The queen conch fishery on Pedro Bank, Jamaica: discovery, development, management, *Ocean & Coastal Management, Volume 42, Issue 12, December 1999, Pages 1069-1081*

Ali, A.B. and Lee, K.Y. (2003). Chenderoh Reservoir, Malaysia: a characterization of a small-scale, multigear and multispecies artisanal fishery in the tropics, *Fisheries Research, Volume 23, Issues 3-4, June 1995, Pages 267-281*

Bulmer MG (1974) On fitting the Poisson lognormal distribution to species abundance data. Biometrics 30: 101-110

Crosby, M.P., Brighouse, G. and Pichon, M. (2002). Priorities and strategies for addressing natural and anthropogenic threats to coral reefs in Pacific Island Nations. *Oceans & Coastal Management 45 (2002) 121-137*

Dovie, D.B.K. (2003). Whose involvement?—can hierarchical valuation scheme intercede for participatory methods for evaluating secondary forest resource use? *Forest Policy and Economics, Volume 5, Issue 3, September 2003, Pages 265-283*

Faddeeva VN (1959) Computational Methods of Linear Algebra. Dover Publications Inc. New

York..
Gaudian G, Medley PAH, Ormond RFG (1995) Estimation of the size of a coral reef fish population. Mar Ecol Prog Ser 56:13-27
Gelman A, Carlin JB, Stern, HS, Rubin, DB (1995) Bayesian Data Analysis. Chapman and Hall, London.
Hilborn R, Walters CJ (1992) Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman and Hall, New York
Keeney RL and Raiffa H (1993) Decisions with Multiple Objectives. Preferences and Value Tradeoffs. Cambridge University Press, Cambridge, UK.
Lassen, H and Medley, P (2001) Virtual Population Analysis. A practical manual for stock assessment. FAO Fisheries Technical Paper 400. FAO, Rome. 129 p.
Leslie PH, Davis DHS (1939) An attempt to determine the absolute number of rats on a given area. J Anim Ecol 8: 94-113
Magurran AE (1988) Ecological Diversity and Its Measurement. Chapman and Hall, London.
Mahon, R., Almergi, S., McConney, P., Parker, C. and Brewster, L. (2003). Participatory methodology used for sea urchin co-management in Barbados. *Ocean & Coastal Management 46 (2003) 1-25*
Manly, BFJ (1997) Randomization, Bootstrap and Monte Carlo Methods in Biology. Second Edition. Chapman and Hall, London.
May RM (1975) Patterns of species abundance and diversity. In: Ecology and Evolution of Communities. Eds: ML Cody and JM Diamond. Harvard University Press, Cambridge, MA. pp. 81-120
McCullagh P, Nelder JA (1989) Generalized linear models. Second Edition. Chapman and Hall, New York.
Moser, C.A. and Kalton, G. (1979). Surveys in Social Investigation. Galliard, Great Britiain.
Pauly, D. 1980 On the interrelationships between natural mortality, growth parameters and mean environmental temperature in 175 fish stocks. J. Cons. CIEM 39(2):175-192.
Pido, M.D. (1995). The application of Rapid Rural Appraisal techniques in coastal resources planning: experience in Malampaya Sound, Philippines, *Ocean & Coastal Management, Volume 26, Issue 1, 1995, Pages 57-72*
Press, SJ (1989) Bayesian Statistics: Principles, models and applications. Wiley Series in Probability and Mathematics, John Wiley and Sons, New York.
Press, WH, Flannery, BP, Teukolsky, SA and Vetterling WT (1989) Numerical Recipes in Pascal. The art of scientific computing. Cambridge University Press, New York.
Silverman, BW (1986) Density Estimation for Statistics and Data Analysis. Mongraphs on Statistics and Applied Probability 26. Chapman and Hall, London.
Sparre, P. Venema, S.C. (1992) Introduction to tropical fish stock assessment. Part 1. Manual. FAO Fisheries Technical Paper No. 306.1 rev 1. Rome, FAO. 376 p.
Sugihara G (1980) Minimal community structure: an explanation of species abundance patterns. Amer. Nat. 116: 770-787
Ugland KI, Gray JS (1982) Lognormal distributions and the concept of community equilibrium. Oikos 39: 171-8
van Lint JH and Wilson RM 2001 A course in combinatorics. 2nd Edition. Cambridge University Press.

Index

- A -

- B -

- C -

- D -

- E -

- F -

- G -

- I -

- K -

- L -

- M -

- N -

- O -

- P -

- Q -

- R -

- S -

- T -

- U -

- V -

- Y -

- Z -

PFSA