A proposal for guidelines for the presentation of fish stock assessment models to IOTC Working Parties.

Iago Mosqueira[1]

Cefas
Pakefield Rd.
Lowestoft NR 33 0HT
United Kingdom

1. INTRODUCTION

At the 9th session of WPTT a proposal was made for a set of guidelines on the presentation of stock assessment models to be prepared for presentation to the Scientific Committee. This document intends to serve as a starting point for further discussion on the issue of how to best present to IOTC Working Parties both the structure and dynamics of new assessment methods, and the results of their application to individual stocks. The guidelines are almost completely based on those provided by Gilbert & Dunn (2004), where examples of each individual recommendation can be found.

2. REQUIREMENTS

Although this document intends to provide only a suggested set of guidelines to be followed rather than a strict set of rules, certain requirements could be adopted as essential for stock assessment models to be admited and considered when establishing the status of a given stock.

The basic requirement is for the method to be sufficiently documented by following the guidelines provided here or any others with similar levels of detailed information on the structure and mechanisms employed.

Input and output files of all alternative runs or scenarios presented should be made available during the meeting for inspection by interested members and for later archiving by the Secretariat. These should ideally be stored together with a copy of the software used in the analysis. When this might not be possible due to licensing issues, a complete reference of the versions of both software and operating system employed should be made.

Software employed should ideally be open sourced using an appropriate license (1), or at least be made available to interested parties for inspection under a limited license. If closed source software is used, this should be clearly justified and sufficient tests as to its validity and reliability, under similar circunstances as those under which it will be used in IOTC-related work, should be carried out and its results made available. Comprenhensive testing, including regression testing and testing of the influence of various assumptions, is greatly encouraged in all cases. Peer review and testing of the method and its actual implementation should be considered when assesing software

---

[1] iago.mosqueira@cefas.co.uk

suitability.

The intention of these requirements is not to limit development of new models but to encourage proper and comprenhensive testing and documentation of assessment models and of their various software implementations.

3. GUIDELINES

These guidelines intend to cover a wide range of characteristics and options of most assessment models that are important when understading their behaviour and dynamics. The intention is to provide a checklist against which authors can check if a given model is sufficiently explained. Any set of guidelines finally adopted could be made available by the Secretariat both as an IOTC document and a set of word processor template for authors to use, for presentation of new assessment models and for application of those to a given stock.

3.1. Observations

Describe the available data and mention, if necessary, data sources or observations not included in the analysis. When refering to datasets provided by the Secretariat, indicate the date, coverage (years, fleets, areas, ...), and precise database (NC, CE, ...). Data sources not previously seen by a Working Party might need their own document presenting them. This includes standardized CPUE series or other data sources processed prior to use. Record the units of measurement of the various data sources, specially if they differ from those commonly used. Plots of time series should be provided if not available elsewhere.

3.2 Population dynamics

Describe the population dynamics that are modelled, in particular, describe the method of modelling including a description of the partition, annual cycle, and method of modelling particular population processes.

3.2.1 Partition of the population, if any

1. Age/length/sex based/plus groups
2. Maturity
3. Spatial structure.

3.2.2 Describe the annual cycle

1. Time steps
2. Age definition: nominal birthday when age is incremented
3. Annual growth assumptions.
4. Natural mortality function: stepwise or continuous; concurrent with fishing or not.
5. Fishing mortality function
6. Recruitment processes (stepwise or continuous)

7. Sequence (order) of processes (e.g., first spawning, then fishing, then migrations, etc.)

3.2.3 Dynamics of the fishery

Describe the dynamics of the fishery (selectivities, ogives, and other assumptions) and the method by which this was modelled

1. Catch equation, by fishery
2. Selectivity/partial recruitment: function (age/length based), parameters, bounds
3. Stock recruit relationships, environmental relationships; assumptions regarding mean for virgin population, initial population, modelled period and future.
4. Maturation curve
5. Migration curves
6. Details if any parameter is fixed (and details of how that value was chosen or otherwise estimated)

3.2.4 Statistical methods

Description of the formal statistical methods, including modelling methods, parameters and assumptions about parameters

1. Software name, version number, and source.
2. Describe the maximum likelihood/objective function, e.g., negative log-likelihood with penalty functions for year class strengths etc., if used.
3. Bootstrap assumptions, if used.
4. Bayesian (MCMC algorithm), if used.
5. Parameters estimated, and the number of such parameters

Describe the free parameters used by the model, including;

1. Parameter name
2. Description of the parameter
3. Details of the estimation bounds/functional relationships with other parameters
4. Details of the prior assumed (if any), and source of the prior
5. C.v., standard error, or variance parameters for likelihood terms
6. Weightings for likelihood terms
7. Adjustment of variance by scaling/adding process error
8. Penalties, etc.

Describe the derived parameters used by the model, including;

1. Parameter name
2. Description and definitions of derived parameters (be precise with those that have alternative definitions, e.g., B0, MSY, MCY, BMSY)
3. Details of any bounds/functional relationships with other parameters.
4. Details of any priors assumed (including source).

3.2.5 Scenarios and retrospective analyses

Alternative scenarios and retrospective analyses, including description of motivation for selection of initial/base cases, giving detail of how the alternative case assumptions differ from the initial or base cases. Description of any retrospective analyses, assumptions, and results.

3.2.6 Projections and assumptions

Describe the projections, definitions of parameters used in projections, and assumptions made;

1. Assumed future catch
2. Assumed (or predicted) future recruitment, including any variability and the basis for variability (bootstrap from estimated values, simulated from assumed distribution).

4. REFERENCES

Gilbert, D. J. & A. Dunn. 2004. *Guidelines for fish stock assessment: A checklist of information and reporting guidelines to working groups for the presentation of fish stock assessments.* Unpublished report. Available from IOTC Secretariat.

Guidelines for fish stock assessment

A checklist of information and reporting guidelines to working groups for the presentation of fish stock assessments

D.J. Gilbert & A. Dunn

INTRODUCTION

Methods for formally assessing fish stocks in New Zealand are becoming increasingly complicated and the use of specialised software packages to develop stock assessment models has become the norm. Fishery Assessment Working Groups are confronted with a wide range of model structures, using a variety of different and complex data that are presented in a variety of ways.

This document sets out a list of practical guidelines that are aimed at assisting presenters of information to Fishery Assessment Working Groups in deciding on the type and detail of information that may be expected.

The paragraph headings used here are intended for this document and would not necessarily be appropriate for a particular presentation or report. Not every item listed here would be required in all cases. These guidelines should be treated as a checklist and a presenter should choose the set of items that convey the key information best — being aware that some items will be inapplicable, irrelevant, or redundant. An oral presentation will necessarily be more selective, but in a written report the writer should have good reasons for omitting items.

The examples given are based on the 2002 stock assessment for hake on the Chatham Rise (albeit loosely, as some aspects particular to this stock have been glossed over in this example). This is a two-sex age-based model, with data from research trawl surveys, Ministry of Fisheries observers, and CPUE abundance indices. The model was fitted using Bayesian estimation in CASAL v 1.02 (Bull et al. 2002). Post model outputs were generated with S-Plus, with most output generated using the NIWA library of S-Plus functions.

The guidelines have been developed using standard principles for the presentation of statistical data and models. The essence of what to present should, ideally, be:

1. An overview of the problem
2. The available data
3. Models/methods employed in solving the problem
4. Assumptions required (and post-analysis examination of these assumptions)
5. Results that were generated
6. Interpretation of the results

Other useful guides on presenting and developing fish stock assessment models include National Research Council (U.S.), Committee on Fish Stock Assessment Methods (1998) and Mace et al. (2001).

1. INITIAL OVERVIEW

The idea here is to briefly present an overview of the objective of the stock model (e.g. to estimate current biomass, with respect to some reference biomass), and the species and stock characteristics that will form the detail behind the stock model. A summary of the species, biology, stock regions, methods of commercial and other catch, timing of the fishery, and any generally important species characteristics.

For example;

- Hake are a middle depths predator, found between 250 m (juveniles) and 1200 m. Trawlers typically catch them as a bycatch of hoki — although localised target fisheries exist, particularly to the northwest of the Chatham Islands.
- There are three QMAs for hake and the current stock hypothesis is that there are three stocks; HAK 7 (west coast South Island stock), HAK 4 and the western end of the Chatham Rise (Chatham Rise stock), and HAK 1 south of the Chatham Rise (sub-Antarctic stock). This stock assessment concentrates on the Chatham Rise stock, covering the QMAs HAK 4 and part of HAK 1.

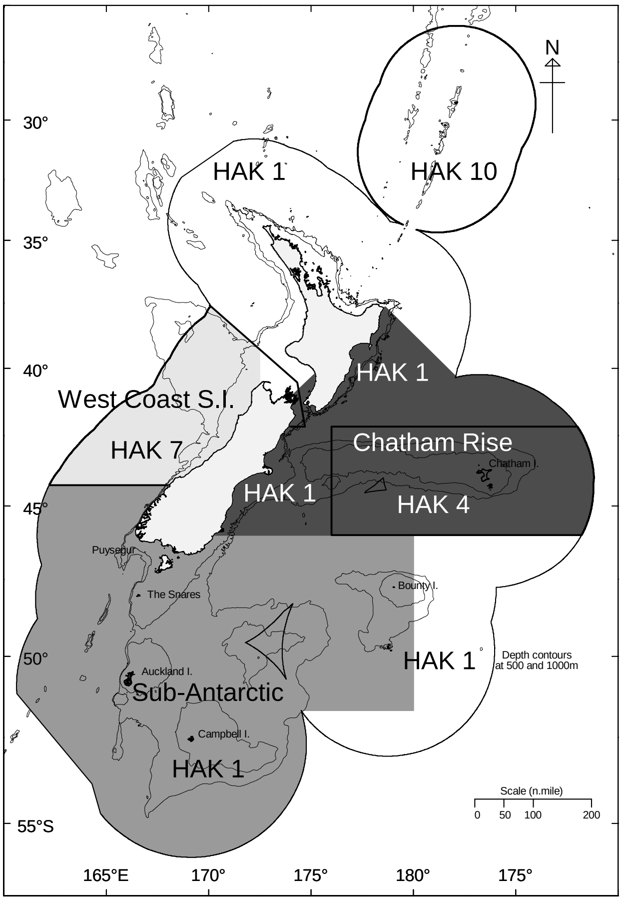

Figure 1: Quota Management Areas (QMAs) HAK 1, 4, 7, & 10; and the west coast South Island (light shaded region), Chatham Rise (dark shaded region), and sub-Antarctic (medium shaded region) hake stock boundaries assumed.

- The observations available include research trawl survey series (1992–2002) with biomass estimates and numbers-at-age, and recent commercial observer data. Most catch is taken during the summer months (>85%) by trawlers targeting hoki, ling, and hake. Catch histories are available since the early 1970s, when it is assumed that the commercial exploitation of hake began.
- Hake grow to a maximum of about 100 cm (males) or 120 cm, (females), and up to a maximum age of about 25 (although a few fish aged up to 30 have been found).
- Recent stock assessments have aimed at estimating (a) current biomass, with respect to virgin biomass B_0 (i.e., the biomass that would be expected under a constant, average, recruitment without any fishing mortality), and (b) future stock status under various future catch scenarios.
- Previous stock assessments have been uncertain, but indicate a B_0 in the range of about 30 000–50 000 t. Annual catches are in the region of about 3000 t, with a

TACC of 3500 t. Reported catches have often fallen short of the TACC by about 20% (See Dunn 2001).

2. DESCRIBING THE MODEL

2.1 Observations

Describe the available observations (and, where appropriate, observations that exist but are not used for the assessment), as a brief overview. Include a description of data sources, methods used to derive the observations, etc., Note that new observations (i.e., those that has not been seen by a working group previously) may need their own presentations.

Observations may include:

1. Catch history, by fishing method.
2. Time series of abundance estimates (absolute or relative), e.g. CPUE, research survey, tag-recapture, etc., Include a summary of the units of weight or number
3. Age, length, age by length distributions, including method of capture (commercial or research), and sampling method.
4. Source biological data, including data for length-weight relationships, maturity by age (or length), sex ratio by age (or length).
5. Ancillary information — same or similar species elsewhere that may be of assistance in interpreting data.
6. Unusual fishery or biological information that may be necessary to satisfactorily model the population dynamics, e.g., restrictions due to bycatch, high inter-annual variability in growth, etc.,

For example;

- Landings data were revised by Dunn (2003) to account for misreporting on the Chatham Rise.
- Research trawl survey biomass observations of hake are available from 11 *Tangaroa* surveys on the Chatham Rise from January 1992 to January 2001. The data used in the stock model include strata in the 200–800 m depth range, although two surveys have had additional strata in the 800–1000 m depth range on the northern Chatham Rise. Data from the deep-water strata are not used.
- Age observations are based on otolith readings using the method of Horn (1997), and by applying an age length key to the length data. Age observations from research surveys is available, but is based on a very limited number of data. Typically only 400–600 hake are caught in each research survey, although there is some evidence that year class estimates track between surveys in the estimates of numbers at age. (Tables and plots of the numbers, numbers by length, and numbers by age could be given here. The following plot has summarised the data from both sexes, research survey catch-at-age data, and commercial observer data onto a single graphic).
- Little empirical information exists on the shape and scale of selectivity curves for either commercial or research trawling.

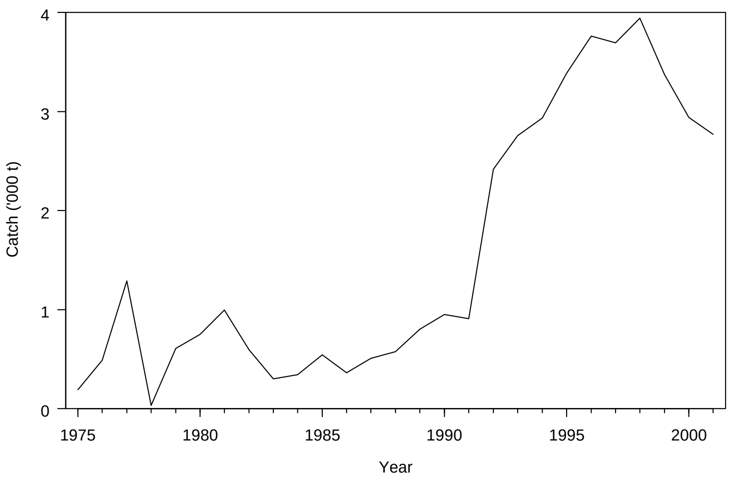

Figure 2: Reported landings (from Dunn 2003) for hake on the Chatham Rise 1974–75 to 2000–01.

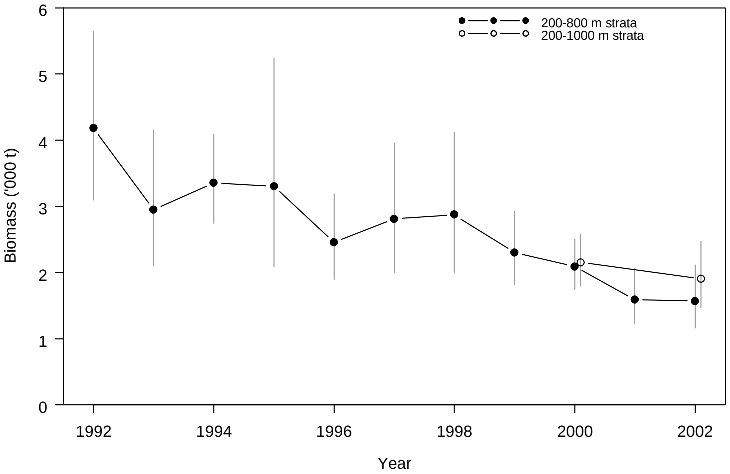

Figure 3: Hake biomass estimates from the R.V. *Tangaroa* surveys of the Chatham Rise, 1992–2002 for the January series, with approximate 95% confidence intervals.

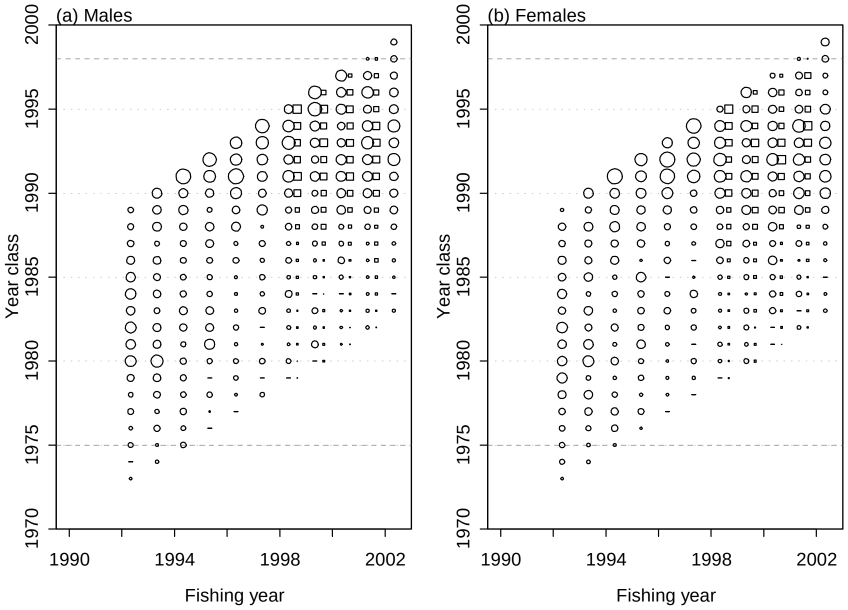

Figure 4: Age frequencies (ages 3 to 19) by year class and year (symbol area proportional to the proportions-at-age within sampling event) on the Chatham Rise for (a) research surveys (circles) and (b) commercial catch-at-age data (squares). Zero values are represented by a dash, and horizontal lines indicate the earliest (1975) and latest (1998) year class strengths to be estimated within the stock assessment model.

2.2 Population dynamics

Describe the population dynamics that are modelled, in particular, describe the method of modelling including a description of the partition, annual cycle, and method of modelling particular population processes.

2.2.1 Describe the partition

Describe the partition of population (if any):

1. Age/length/sex based/plus groups
2. Maturity
3. Spatial structure

For example;

- The stock assessment model partitions the Chatham Rise stock population into two sexes and age groups 3–30, with no plus group, i.e., fish older than age 30 are ignored by the model.
- The stock was considered to reside in a single area, with instantaneous spawning.
- Maturity was not modelled as a part of the partition, but was assumed to be a constant proportion of fish by sex and age. The maturity ogive represents the proportion of fish at each age that are assumed to be mature. Maturity values at age were assumed known, and were based on model fits to the 1998 HAK 7 assessment (Dunn 1998), using the commercial catch-at-age data under an assumption that the all year class strengths were 1. These values are broadly consistent with estimates from Colman (1998), who concluded that Chatham Rise hake reach 50% maturity at between 7–8 years.

2.2.2 Describe the annual cycle

Describe the annual cycle;

1. Time steps
2. Age definition: nominal birthday when age is incremented
3. Annual growth assumptions.
4. Natural mortality function: stepwise or continuous; concurrent with fishing or not.
5. Fishing mortality function
6. Recruitment processes (stepwise or continuous)
7. Sequence (order) of processes (e.g., first spawning, then fishing, then migrations, etc.)

For example;

- The model annual cycle was based on the fishing year and divided the year into three steps. The time steps and sequence of processes is given below.
- The proportion of males at recruitment was assumed to be 0.5 of all recruits, as there is no external data from which to estimate this value.
- Recruitment is assumed to occur at age 3, with year class strengths estimated using the parameterisation of Haist (as cited in Bull et al. 2002)
- Most hake catch (85%) on the Chatham Rise occurs during the summer months (October–February) as either targeted catch of hoki/ling targeted bycatch from large trawlers. Hence, while some targeting (and catch) of spawning fish occurs, we assume that the all fishing mortality occurs on the entire population during the first time step.

Table 1: Annual cycle of the stock model, showing the processes taking place at each time step, their sequence within each time step, and the available observations. Fishing and natural mortality that occur within a time step occur after all other processes, with half of the natural mortality for that time step occurring before and after the fishing mortality.

					Observations	
	Period	Processes (in the order applied)	*M*	Growth	Description	%*M*
1	Oct–Feb	Spawning, recruitment, & fishing	0.42	0.25	CPUE	20
					Jan. survey	100
2	Mar–Sep	None	0.58	0.50		
3	Sep	Increment age	0.00	0.00		

1. *M* is the proportion of natural mortality that was assumed to have occurred in that time step.
2. Growth is the fraction of age that is used to determine the additional growth that was assumed to occur at that time step.
3. %*M* is the percentage of the natural mortality in each time step that was assumed to have taken place at the time each observation was made.

2.3 Selectivities, ogives, and other assumptions

Describe the dynamics of the fishery, and the method by which this was modelled;

1. Catch equation, by fishery
2. Selectivity/partial recruitment: function (age/length based), parameters, bounds
3. Stock recruit relationships, environmental relationships; assumptions regarding mean for virgin population, initial population, modelled period and future.
4. Maturation curve
5. Migration curves
6. Details if the parameter is fixed (and details of how that value was chosen or otherwise estimated)

For example;

- The catch equation used was the instantaneous mortality (not Baranov) equation from Bull et al. (2002). This applies ½ of the natural mortality, then all fishing mortality, then the remaining ½ of natural mortality during each time step.
- The model used four selectivity ogives; male and female fishing selectivities on the Chatham Rise, and male and female *Tangaroa* survey selectivities.
- Research survey selectivities were initially assumed to be smoothed degree-3 polynomials, while fishing selectivities were assumed to be logistic. In all cases, male selectivity curves were estimated relative to female selectivity.

2.4 Modelling methods, parameters and assumptions about parameters

Description of the formal statistical methods, including;

1. Software name, version number, and source.
2. Describe the maximum likelihood/objective function, e.g., negative log-likelihood with penalty functions for year class strengths etc., if used.
3. Bootstrap assumptions, if used.
4. Bayesian (MCMC algorithm), if used.
5. Parameters estimated, and the number of such parameters

Describe the free parameters used by the model, including;

1. Parameter name
2. Description of the parameter
3. Details of the estimation bounds/functional relationships with other parameters
4. Details of the prior assumed (if any), and source of the prior
5. C.v., standard error, or variance parameters for likelihood terms
6. Weightings for likelihood terms
7. Adjustment of variance by scaling/adding process error
8. Penalties, etc.

For example;

- Biological parameters are assumed, with von Bertalanffy and length-weight parameters from Horn (1998) and estimates of natural mortality from Dunn et al. (2000).
- Variability in growth was assumed with the variability assumed to be lognormal with constant c.v.=0.10.
- No information is available on the stock-recruitment relationship, but the model assumed a Beverton & Holt stock–recruitment relationship with a steepness of 0.9.
- Lognormal errors, with known c.v.s, were assumed for all relative biomass and proportions-at-age observations. For the proportions-at-age observations, the lognormal assumption was slightly modified to make it robust against outliers. The negative log-likelihood was,

$$-\log(L)=\sum\nolimits_{i=1}^{n}\log(\sigma_i)-\log\left(\exp\left(-0.5\left(\frac{\log(O_i/qE_i)}{\sigma_i}+0.5\sigma_i\right)^2\right)+0.01\right)$$

where O and E are the observed and expected values, and σ is the standard deviation of the log of the error. The lognormal distribution is robustified by the addition of the term 0.01, and is a similar modification to that employed by Fournier et al. (1990) for the normal distribution, with the aim of minimising the influence of large absolute residuals (i.e., against type II deviations where observed values are either much higher or much lower than predicted values).

- The c.v.s available for the relative abundance and catch-at-age observations allow for sampling error only. However, additional variance assumed to arise from differences between model simplifications and real world variation, was added to the sampling variance. The additional variance, termed process error, was estimated in early runs of the model, using all available data, from MPD fits. Hence, the overall c.v. assumed in the initial model runs for each observation was calculated by adding process error and observation error. The process error added was a c.v. of 0.06 and 0.27 for the sub-Antarctic and Chatham Rise CPUE series respectively, and 0.52 and 0.22 for the sub-Antarctic and Chatham Rise observer proportions-at-age data (the same value for all ages). The process error estimated for the research survey relative abundance estimates and proportions-at-age data was zero. However, the resulting estimates of total c.v.s (process error plus sampling error) for the sub-Antarctic CPUE series was considered to be unreasonably low, and hence all subsequent model runs employed a process error of 0.2 for that series.
- The assumed prior distributions used in the assessment are given in the table below. Most priors were intended to be relatively uninformed, and were estimated with wide bounds. The exceptions were the choice of informative priors for the survey *q*s and natural mortality *M* (when estimated).
- The priors for survey *q*s were estimated by assuming that the relativity constant was the product of areal availability, vertical availability, and vulnerability. A simple simulation was conducted that estimated a distribution of possible values for the relativity constant by assuming that each of these factors was uniformly distributed. A prior was then determined by assuming that the resulting, sampled, distribution was lognormally distributed. Values assumed for the parameters were; areal availability (0.50–1.00), vertical availability (0.50–1.00), and vulnerability (0.01–0.50). The resulting (approximate lognormal) distribution had mean 0.16 and c.v. 0.79, with bounds assumed to be (0.01–0.40). The assumed distribution for all survey *q*s is shown as Figure 5. Selectivity priors, for all parameters, were uniform with bounds for the degree-3 polynomial set at (0.01, 5.00). Note that the values of survey relativity constants are dependant on the selectivity parameters, and the absolute catchability can be determined by the product of the selectivity by age and sex, and the relativity constant *q*.
- The prior on natural mortality, *M*, (when estimated) was determined by assuming that the current estimate of natural mortality was a reasonable approximation to the true value with the assumption that the true value could differ from the current point estimate by about 0.05, and not more than 0.1 (see Figure 5). Note that natural mortality was parameterised by the average of male and female, with the difference estimated with an associated normal prior with mean 0.0, standard deviation of 0.1, and bounds (-0.2,0.2).
- Penalty functions were used to constrain the model so that any combination of parameters that did not allow the historical catch to be taken were strongly penalised. A small penalty was applied to the estimates of year class strengths to encourage estimates that average to 1. In addition, in scenarios where selectivity ogives were not parameterised by logistic curves, selectivity estimates were penalised to encourage the shape of the age-based selectivity ogive to approximate a degree-3 polynomial.

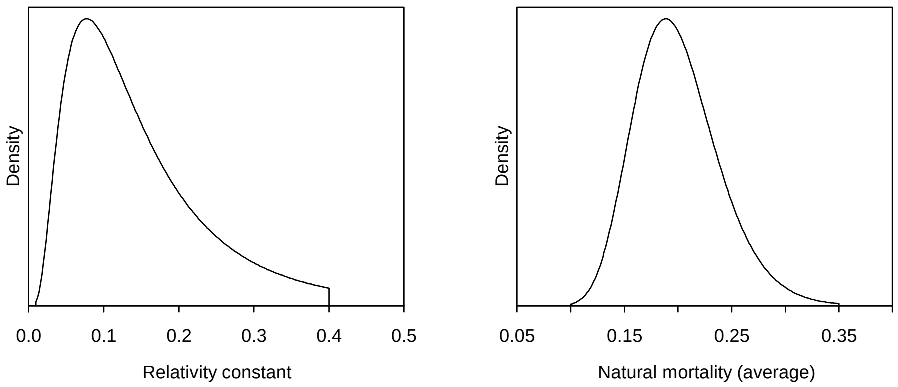

Figure 5: The prior distribution for (left figure) survey relativity constant (*q*) lognormal where μ=0.16, c.v.=0.79, and bounds (0.01,0.40), and (right figure) natural mortality (average), lognormal where μ=0.20, c.v.=0.20, and bounds (0.10,0.35).

Table 2: Fixed biological parameters assumed.

Parameter		Value		
Natural mortality			Males	0.20^{y-1}
			Females	$0.18\ y^{-1}$
Weight = a·(length)^b (Weight in g, length in cm)				
Males	a = 0.00249	b = 3.234		
Females	a = 0.00170	b = 3.328		
Von Bertalanffy growth parameters				
Males	*k* = 0.277	t_0 = –0.11		L_∞ = 90.3
Females	*k* = 0.202	t_0 = –0.20		L_∞ = 113.4
Steepness (Beverton & Holt SR relationship)		0.90		
Recruitment variability		0.6		
Proportion spawning		1.0		
Proportion of recruits that are male		0.5		
Spawning season length		0		
Maximum exploitation rate (U_{max})		0.5		
Maturity ogive (for ages 4–8 & 9+)	Male	0.02, 0.07, 0.31, 0.78, 1.00, 1.00		
	Female	0.02, 0.04, 0.07, 0.45, 0.86, 1.00		

Table 3: The assumed priors assumed for key distributions (when estimated). The parameters are mean (in natural space) and c.v. for lognormal; and mean and s.d. for normal.

Parameter	Distribution	Parameters		Bounds	
B_0	Uniform-log	–	–	2 500	250 000
CPUE *q*	Uniform-log	–	–	1e-10	1e-0
Survey *q*	Lognormal	0.16	0.79	0.01	0.40
YCS	Lognormal	1.0	1.1	0.01	100
M (mean)	Lognormal	0.20	0.20	0.10	0.35
M (difference)	Normal	0.0	0.05	-0.20	0.20

Describe the derived parameters used by the model, including;

1. Parameter name
2. Description and definitions of derived parameters (be precise with those that have alternative definitions, e.g., B_0, *MSY*, *MCY*, B_{MSY})
3. Details of any bounds/functional relationships with other parameters.
4. Details of any priors assumed (including source).

For example;

- Model parameters were estimated using Bayesian estimation implemented using the CASAL software (with the full details of the CASAL algorithms, software, and methods detailed in Bull et al. 2002). However, only the mode of the joint posterior distribution (MPD) was estimated in preliminary runs. For final runs, the full posterior distribution was sampled using Markov Chain Monte Carlo (MCMC) methods, based on the Metropolis-Hastings algorithm.
- Year class strengths were assumed known (and equal to one) for years prior to 1975 and after 1998, when inadequate or no catch-at-age data were available. Otherwise year class strengths were estimated under the assumption that the estimates from the model must average one.

2.5 Alternative scenarios and retrospective analyses

Description of motivation for selection of initial/base cases, giving detail of how the alternative case assumptions differ from the initial or base cases.

Description of any retrospective analyses, assumptions, and results.

For example;

- Selectivity runs were chosen to investigate aspects of the model where strong model assumptions may have influenced the model output, and where there was little evidence that the particular series of input data were well estimated.
- Sensitivity runs included:
    a. excluding the CPUE series (“no CPUE”) from model runs
    b. parameterising the research survey selectivities as logistic curves, rather than by using a penalty to encourage smoothed degree-3 polynomials (“logistic”)
    c. parameterising the fishery selectivities using a penalty to encourage smoothed degree-3 polynomials, rather than logistic parameterisations (“degree-3 poly”)
    d. estimating natural mortality *M* (“estimate *M*”).
- For each model run, MPD fits were obtained and qualitatively evaluated. Objective function values (negative log-likelihood) for each model run are shown below. MCMC estimates of the posterior distribution were obtained for all model runs. These are presented below. In addition, MCMC estimates of the median posterior and 95% percentile credible intervals are reported for the key output parameters. Summary plots of the base case MPD model fits are also given.

2.6 Projections and assumptions

Describe the projections, definitions of parameters used in projections, and assumptions made;

1. Assumed future catch
2. Assumed (or predicted) future recruitment, including any variability and the basis for variability (bootstrap from estimated values, simulated from assumed distribution).

For example;

- Five-year projected biomass estimates assumed that future catches are at the level of the current TACC (3500 t)

- Projections used the mean recruitment over the estimation period, adjusted according to the Beverton-Holt stock recruitment relationship, with the stochasticity in the projected year class strengths assumed to be lognormal with mean 1 and standard deviation 0.6.

3. DESCRIBING THE RESULTS

3.1 Parameter point estimates and distributions

Detail the estimates of parameters determined by the model, including;

1. Table or figures giving parameter estimates noting estimator type, i.e., MLE, MPD, or posterior mean. Note when a parameter is at a bound. It will often be desirable to also give distributional information such as: confidence intervals, marginal posterior credible intervals, or corresponding standard errors.
2. Tables of relevant non-fundamental parameter estimates noting estimator type. Again it will often be desirable to also give distributional information.
3. Model selection criteria: total likelihood, AIC, Bayes factor, etc,.
4. Table of pairwise correlations

Always include the following biomass estimates: B_0 (virgin stock biomass), B_{current} (current stock biomass), $B_{\mathrm{projected}}$ (projected stock biomass under specified catches). These should be presented as absolute values and as proportions of B_0. It will also often be desirable to include: B_{initial} (stock biomass at the start of the modelled period, if this differs from B_0), B_{lowest} (lowest historical stock biomass, if this differs substantially from B_{current}). It may sometimes be useful to also present spawning stock biomass or vulnerable stock biomass together with stock biomass.

Always include the following yield estimates: at least one version of maximum sustainable yield, e.g. *MSY, MAY, MCY*. The equilibrium stock biomass corresponding to the maximum sustainable yield presented, e.g. B_{MSY}, B_{MAY}, B_{MCY}. It will also often be desirable to include: the current surplus production (catch that would leave the stock at the current biomass), mean projected surplus production.

It will often be useful to present estimates of exploitation rates or fishing mortalities. These are often best presented as trajectories. Catch to mid-year stock biomass ratios are the easiest to interpret. Maximum fishing mortality or exploitation rates over age/sex or length/sex categories may also be presented.

Parameter-dependent plots include;

1. Selectivity curves
2. Length at age curves
3. Maturity ogives
4. Biomass trajectories

For example;

- MCMC chains were estimated using a burn-in length of 10^6 iterations, with every 10 000th sample taken from the next 10^7 iterations (i.e., a final sample of length 1000 was taken from the Bayesian posterior). Autocorrelations, and single chain convergence tests of Geweke (1992) and Heidelberger & Welch (1983) were applied to resulting chains to determine evidence of non-convergence. The tests used a significance level of 0.05 and the diagnostics were calculated using the Bayesian

Analysis Output software (Smith, B.J., 2001. Bayesian output analysis program. Version 1.00 user's manual. Unpublished manuscript. 45 p. University of Iowa College of Public Health. http://www.public-health.uiowa.edu/boa).

- The estimated median and 95% credible intervals for B_0, B_{2002}, and B_{2002} are shown in the table below.
- The estimated MCMC marginal posterior distributions for selected parameters for the Chatham Rise stock are shown in the figures below.
- Selectivities for males and females were divergent, with the selectivities for males higher than females in the research surveys at older ages (15+) in the January survey series. Female selectivities were significantly dome shaped, with female selectivity at age 12–14 about twice that at ages 15+. Both the male and female selectivities suggested that males and females were not fully selected by the trawl gear until aged at least 12+.
- Posterior density estimates of selectivities indicated considerable uncertainty in the estimates of selectivity by age and sex.
- Fishing selectivities were poorly estimated, with strong evidence that the selectivity of male fish was considerably higher at age than for females. There is no information outside the model that allows the shape of the estimated selectivity ogives to be verified.
- Year class strengths were poorly estimated for years where only older fish were available (i.e., before about 1982). More recent year class strengths appear well estimated, although the signal in the very recent age data is unlikely to be as strong as indicated by the model estimates.
- The year class strength estimates suggested that the stock is characterised by a group of strong year classes in the late 1970s, and again in the late 1980s and early 1990s, followed by a period of recent declining recruitment.
- Exploitation rates for the Chatham Rise appear to be increasing, with upper posterior near 0.4. These are defined as the maximum exploitation rate, over age and sex, in the fishery from fishing. This is indicative of a high, and possibly increasing, level of fishing pressure.

Table 4: Bayesian median and 95% credible intervals of B_0, B_{2002}, and B_{2002} as a percentage of B_0 for the base and sensitivity cases.

Model run	B_0	B_{2002}	B_{2002} (%B_0)
Base case	36 900 (31 600–43 200)	21 800 (16 100–28 900)	59.1 (49.5–68.3)
No CPUE	36 900 (31 800–44 100)	21 800 (16 000–30 400)	59.1 (49.8–70.1)
Logistic	32 100 (28 100–38 500)	17 200 (12 600–24 700)	53.4 (44.0–64.1)
Degree-3 poly	43 500 (34 900–68 100)	27 500 (19 000–49 800)	62.6 (52.8–74.5)
Joint fishing	37 800 (32 600–45 200)	24 200 (18 300–32 900)	64.1 (54.7–74.3)
Estimate *M*	36 000 (29 400–48 200)	21 800 (14 500–34 000)	60.4 (48.6–73.3)
No misreporting	43 600 (37 700–52 600)	26 100 (19 100–35 900)	59.6 (50.0–70.2)

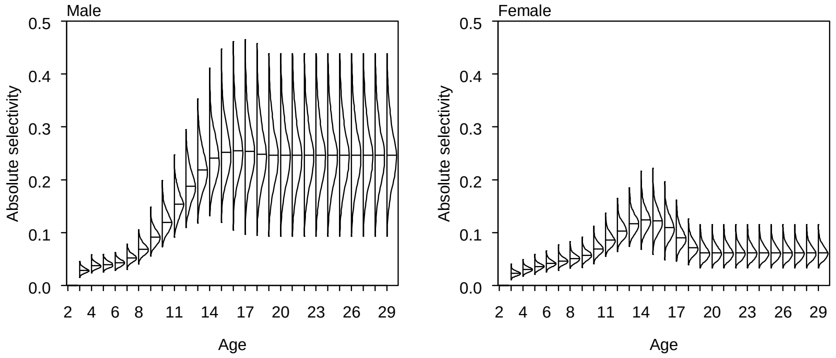

Figure 6: Estimated posterior distributions of absolute selectivity by age and sex, for the base case for the Chatham Rise January research survey proportions-at-age. Individual distributions show the marginal posterior distribution, with horizontal lines indicating the median.

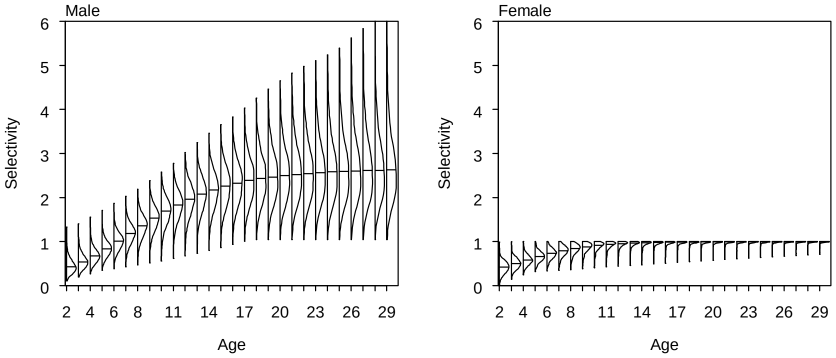

Figure 7: Estimated posterior distributions of selectivity by age and sex (relative to the maximum selectivity of females), for the base case for the Chatham Rise fishery proportions-at-age. Individual distributions show the marginal posterior distribution, with horizontal lines indicating the median.

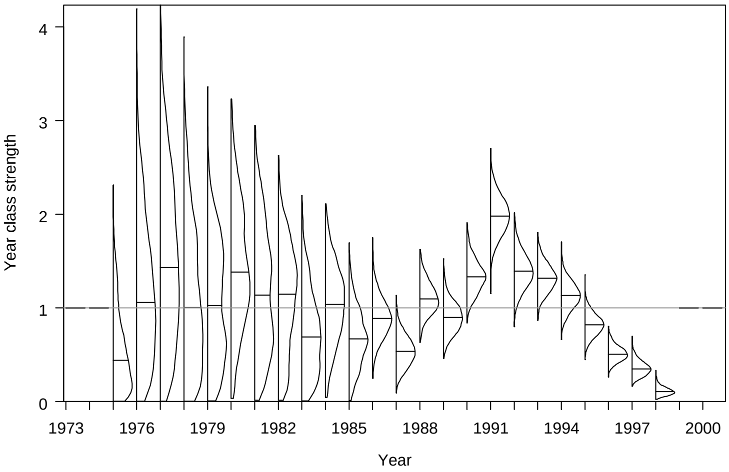

Figure 8: Estimated marginal posterior distributions of year class strengths for the base case for the Chatham Rise stock. Horizontal lines indicate the medians, and the grey horizontal line indicate a year class strength of one.

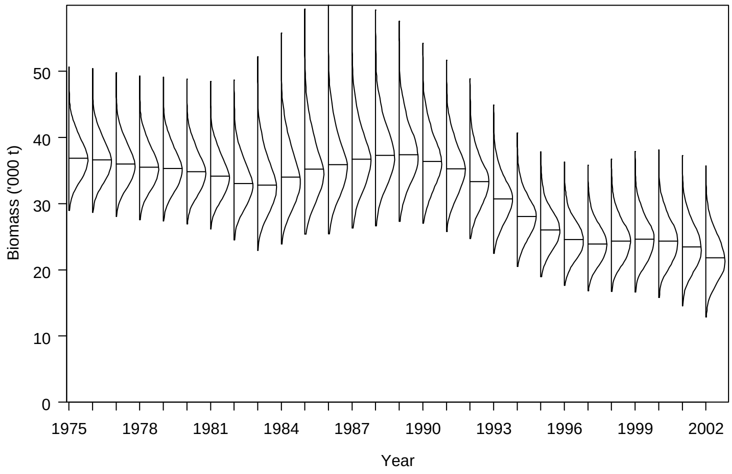

Figure 9: Estimated marginal posterior distributions of biomass trajectories for the base case for the Chatham Rise stock. Horizontal lines indicate the medians.

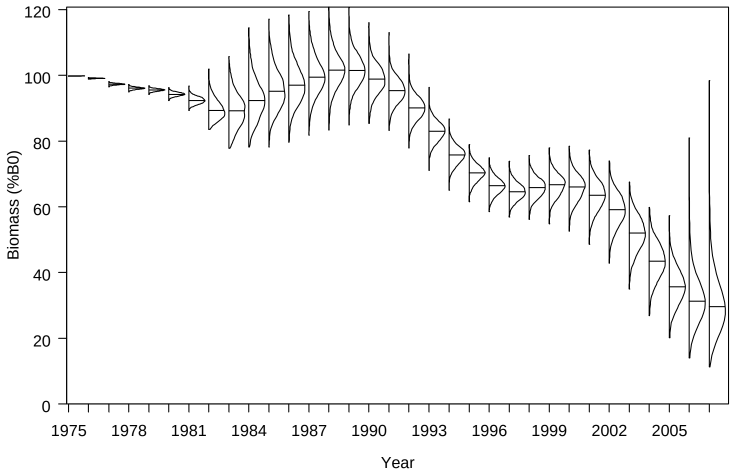

Figure 10: Estimated marginal posterior distributions of biomass trajectories (%B_0) for the base case for the Chatham Rise stock, projected to 2007 with future catches assumed equal to the TACC. Horizontal lines indicate the medians.

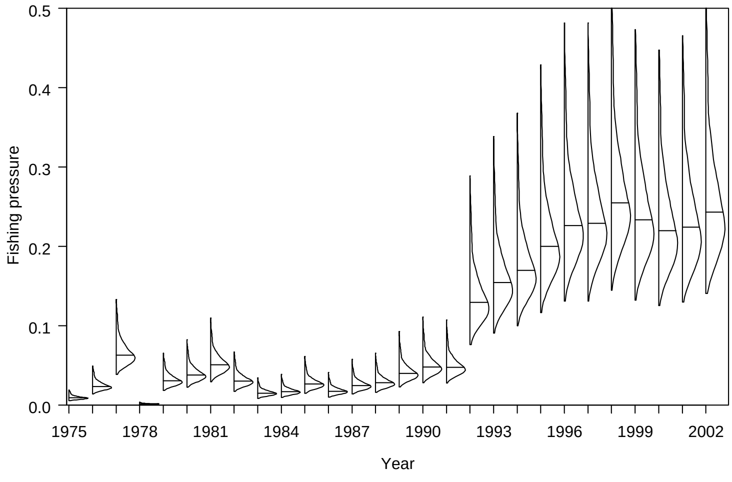

Figure 11: Estimated posterior distributions of exploitation rates for the base case for the Chatham Rise stock. Horizontal lines indicate the medians.

3.2 Model fit

Model fits can be evaluated in a range of different ways. Models fits are evaluated in order to determine if there is evidence that the assumptions of the model are adequate. The types of analyses required would usually include a selection of;

1. Diagnostic summaries and tables/plots, including;
    - Observed versus expected tables/plots
    - Residual versus appropriate covariate tables/plots (age, length, sex, cohort, sample year, vessel size, time of day, area, method)
    - Residual versus predicted tables/plots
    - Tables/plots of residual distributions and assumed distributions.
    - Quantile-quantile tables/plots
2. Tables/plots of means and standard deviations of residuals/standardised residuals for each data type.
3. Tables/plots of ancillary data that investigates structural assumptions, e.g., observed annual weight at age data versus predicted weight at age (often assumed constant).

Discussion should also include interpretation of the model fits and diagnostics. This should include an attempt to explore where model assumptions break down, and an explanation of why model fits may be poor or estimates of parameters may not be adequate. Specifically;

1. What can be inferred from the model fit diagnostics?
2. What diagnostics are consistent/inconsistent with the structural assumptions of the models and each other?
3. Are any subsets of the data inconsistent with the assumptions of the model?
4. Are the residuals consistent with the distributional assumptions?
5. Are there any patterns in the residuals?

For example;

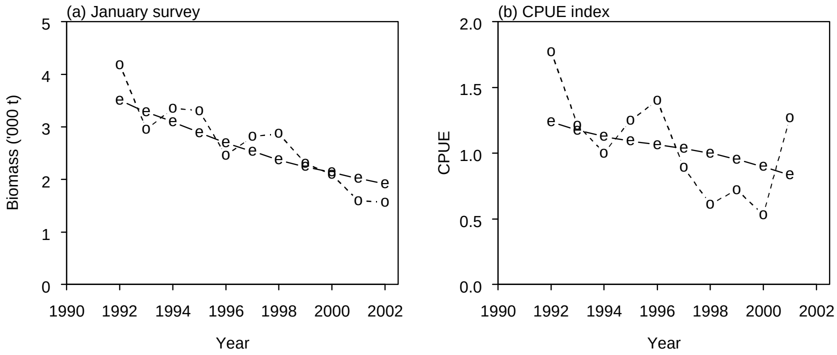

Figure 12: MPD (base case) fits to the (a) January research biomass index, and (b) CPUE relative abundance index, where 'o' indicated the observed value and 'e' indicated the fitted (expected) value.

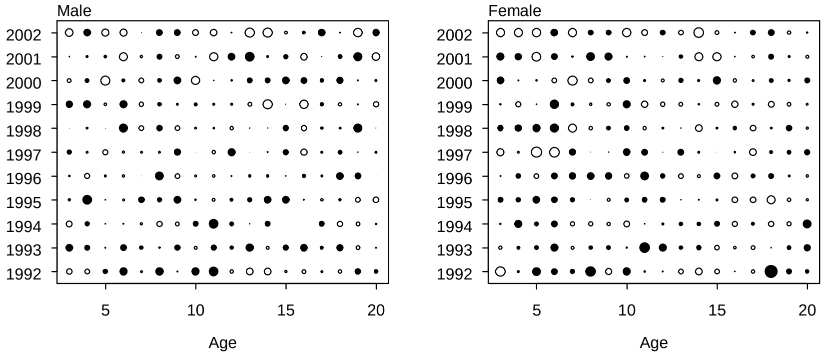

Figure 13: MPD (base case) residual values for the proportions-at-age data for the Chatham Rise January research survey series. Symbol area is proportional to the absolute value of the residual, with white circles indicating positive residuals and black circles indicating negative residuals.

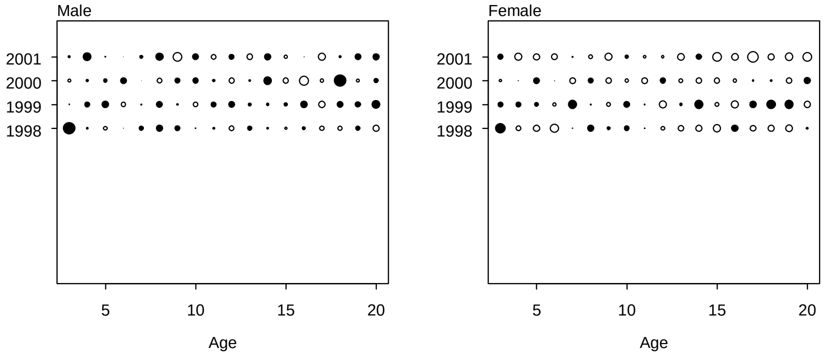

Figure 14: MPD (base case) residual values for the proportions-at-age data for the Chatham Rise observer sampling series. Symbol area is proportional to the absolute value of the residual, with white circles indicating positive residuals and black circles indicating negative residuals.

3.3 Parameter uncertainty

Discussion of the uncertainty assumed in the model, including the impact of priors and distributional/structural assumptions. This should include discussion of;

1. Effective priors (obtained from MCMC by fixing the total negative log-likelihood at 0). This is to check the effects of bounds, penalties, and the interaction of priors on both fundamental and non-fundamental parameters.
2. Either likelihood profile plots, marginal posterior plots (overlaid on priors), or bootstrap distribution plots of key parameters
3. Plot or tables of pairwise correlations

For example;

- Estimates of posterior distributions and assumed priors for the research survey relativity constant and estimates of natural mortality *M* are given in the figures below.

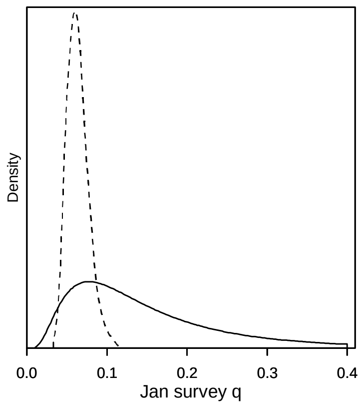

Figure 15: Estimated posterior distribution (dashed line) and prior (solid line) of survey relativity constant for the base case for the Chatham Rise January research survey series.

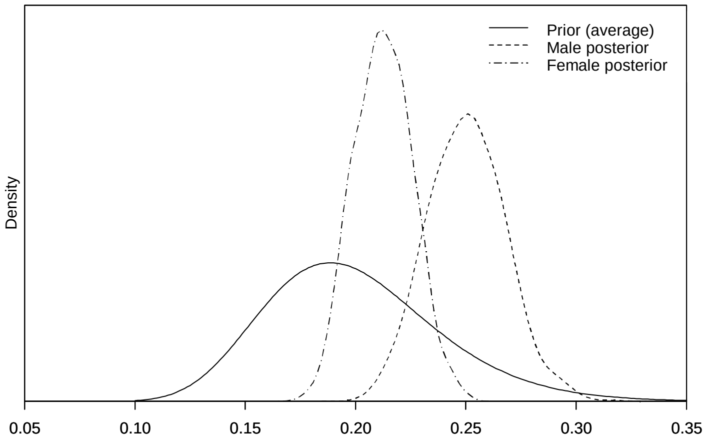

Figure 16: Estimated posterior distributions for the "estimate *M*" case for male and female natural mortality rates and the natural mortality prior (average of male and female *M*).

3.4 MCMC convergence statistics

Convergence of MCMC chains are difficult to ensure. Discussion of MCMC output should include some attempts to determine that MCMC output has converged, and that an adequate estimate of point estimates and credible intervals has been achieved. Include discussion of ;

1. Length of chain and sample frequency
2. Autocorrelation between samples
3. Table of convergence statistics
4. MCMC traces

For example;

- Convergence diagnostics for the Chatham Rise stock model are given in the table below. Diagnostics were run on chains of final length 10^7 iterations, after a burn-in of 10^6 iterations, after systematically sub-sampling ("thinning") to 1000 samples.
- Lagged autocorrelations can provide an indication of the level of mixing in the samples from the Markov chain. High values of autocorrelation at low lags indicate

slow mixing within the chain, and hence suggest that the chain may converge only slowly to the stationary (posterior) distribution.

- The Geweke (1992) convergence diagnostic is based on a test that compares the means of the first 10% and last 50% of a Markov chain. Under the assumption that the samples were drawn from the stationary distribution of the chain, the two means are equal and Geweke's statistic has an asymptotically standard normal distribution. The resulting test statistic is a standard Z-statistic, with the standard error estimated from the spectral density at zero. Values of the Z-statistic that have a p-value less than 0.05 indicate that, at the 5% significance level, there is evidence that the samples were not drawn from a stationary distribution.
- Heidelberger & Welch (1983) proposed two linked tests. The first is a stationarity test that uses the Cramer-von Mises statistic to test the null hypothesis that the sampled values come from a stationary distribution. The test is successively applied, first to the whole Markov chain, then after discarding the first 10, 20, etc., percent of the chain until, either the null hypothesis is accepted, or 50% of the chain has been discarded. If more than 50% of the chain is discarded, then the test returns a failure of the stationarity test. Otherwise, the number of iterations to keep are reported. The second test is the half-width test that calculates a 95% confidence interval for the chain mean, using the portion of the chain that passed the Heidelberger & Welch stationarity test. Half the width of this interval is compared with the estimate of the mean. If the ratio between the half-width and the mean is lower than 2% of the mean, the half width test is passed.
- No evidence of lack of convergence was found in the estimates of B_0, although some estimates of selectivity parameters and year class strengths showed strong evidence of lack of convergence. The constraint of degree-3 polynomials on selectivity functions and year class strengths having average 1 contributed to large correlations between some parameters, resulting in a lack of possible convergence. Trace diagnostics of B_0 is shown below.

Table 5: Autocorrelation at lags of 1, 5, and 10, for selected parameters from the base case MCMC chain for the Chatham Rise stock.

Stock	Parameter	Autocorrelation		
		Lag-1	Lag-5	Lag-10
Chatham Rise	B_0	0.075	0.021	0.049

Table 6: Percentage of parameters that passed the Geweke and Heidelberger & Welch convergence diagnostics for selected parameters from the base case MCMC chain for the Chatham Rise stock model (% refers to the percentage of parameters that passed each test).

Stock	Parameter	n	Geweke	Heidelberger & Welch	
				Stationarity	Half width test
Chatham Rise	B_0	1	100.0%	Passed	Passed
	Selectivity	38	78.9%	97.4%	100%
	YCS	24	79.2%	79.2%	87.5%

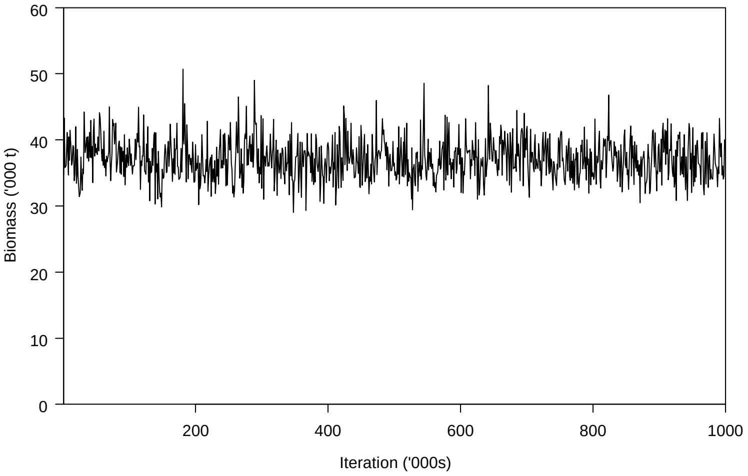

Figure 17: Trace diagnostic plot for base case MCMC chain for estimates of B_0 for the Chatham Rise stock model.

3.5 Interpretation of parameter uncertainty

What should we infer from the parameter uncertainty statistics? Describe any conclusions or hypotheses about parameters, given the parameter uncertainty.

4. DISCUSSION AND CONCLUSIONS

Discussion of the results of the analyses, exploring the implications of the results, and describing any external factors that may assist in drawing conclusions. Provide clear statements about the conclusions that can be drawn (and those that cannot) from the analysis. Where possible, it may be helpful to provide ideas or suggestions about how the analyses may be improved in future work, and how assumptions critical to the assessment could be validated.

For example;

- Information on the stock status of hake on the Chatham Rise appears reasonably strong. Biomass estimates from the Chatham Rise trawl series suggests strong evidence of a uniform decline in biomass, with biomass in 2002 at about half the level of 1992. While the general trend in the CPUE indices is consistent with this decline, the evidence that the CPUE indices are indexing abundance is weak.
- The use of a different model structure, revised catch histories, revised CPUE abundance indices, and addition of data from the 2001 trawl survey and two years of observer proportions-at-age have resulted in improved model performance since the previous assessment for the Chatham Rise stock. Biomass estimates from the previous assessment are consistent with the results from this assessment. That assessment suggested that B_0 (least squares estimate) ranged from 30 000–50 000 t, with the trawl survey selectivity ogive fixed to be fully selected for males at age 8. The (least squares) estimates of B_{2000} were about 50–70% B_0.
- If the model assumptions are correct, and the recent estimated year class strengths are as weak as have been estimated, then current catch levels are likely to continue to reduce the stock size for the Chatham Rise stock in the immediate future. But projections indicate that the rate of decline is slow. However, it would be advisable to review the assessment for hake on the Chatham Rise as new data become available,

as it is possible that some management action may be necessary at some stage in the medium term.

- The estimates of stock size and projected stock status rely strongly on the estimated shape of the selectivity ogives. The model suggested that maximum selectivity (for both trawl and fishing) appears to occur at ages about 12–14. These ages are higher than expected given the age at maturity, growth rates, and the maximum age of hake. Estimated differences in the relative selectivity between male and female hake are unusually high, although there is no information on the relative selectivity of hake to allow these estimates to be confirmed. If selectivity patterns were such that full selectivity occurred at a lower age than estimates, it is likely that estimates of current stock status would be lower.
- Estimates of the relative catchability constant (*q*) for the Chatham Rise trawl survey were close to the range of values assumed in the prior and appeared reasonable.
- Poor fits to the observer proportions-at-age suggest that better estimates of commercial catch-at-age should be of high priority. The model was not able to adequately explain the commercial proportions-at-age data without assuming relatively high additional variability. In addition, while evidence that strong and weak year classes track between trawl survey observations of proportions-at-age, there is little evidence that such patterns can be observed in the commercial proportions-at-age data. Annual changes in selectivity are possible in the hake fishery, especially as hake are taken predominantly as a bycatch species. It is possible that the lack of observed pattern in the observer proportions-at-age may reflect this. Annual changes in selectivity have not been modelled here, as the data are probably not adequate to allow sensible estimation of additional selectivity parameters.
- Estimates of recent recruitment on the Chatham Rise suggest strong evidence of collapsing in recruitment in recent years. While there is some suggestion in the trawl survey proportions-at-age data that the relative numbers of younger fish has declined, there is stronger evidence from the commercial catch proportions-at-age data. However, there has been considerable change in commercial fishing patterns on the Chatham Rise in the few last years. The introduction of the Hoki Fishery Management Company Ltd. guidelines to restrict the catch of juvenile hoki has limited the volume of fishing in and around the Mernoo Bank on the Chatham Rise. This *may* have led to a reduction in the catch of juvenile hake (although hake do tend to occupy deeper depths than hoki), resulting in an absence of juvenile hake in the sampled catch. If so, the apparent decline in recruitment may be a consequence of a substantive change in fishing selectivity on younger fish.
- In an analysis of the impact of changes in trawl survey design from 1996, Bull & Bagley (1999) concluded that there was a decline in proportions of older hake (defined by length class) on the Chatham Rise, and further, found evidence of a similar decline in the proportions of longer (older) hake in the proportions-at-length data from the commercial catch. Such observations are consistent with an exploited stock in decline.
- In both stock models, estimates of earlier levels of recruitment are poor. Inconclusive age data, particularly in older aged fish where few sampled are available resulted in a wide range of possible estimates of relative year class strength. Additional age data from surveys by the *Amaltal Explorer* and *Shinkai Maru* during the 1980s could provide useful information on the relative year class strengths in early years, although the relative selectivity patterns from these vessels is unknown and, moreover, is unlikely to be estimable. Age data from these surveys was not available for this analysis.
- The model structural assumptions are likely to lead to the Bayesian posteriors of current stock status under-estimating the true level of uncertainty. The projected stock status relies on adequate estimation of recent recruitment.

5. USEFUL REFERENCES

Bull, B.; Francis, R.I.C.C.; Dunn, A.; Gilbert, D.J. (2002). CASAL (C++ algorithmic stock assessment laboratory): CASAL user manual v1.02.2002/10/21. *NIWA Technical Report 117*. 199 p.

Mace, P.M.; Bartoo, N.W.; Hollowed, A.B.; Kleiber, P.; Methot, R.D.; Murawski, S.A.; Powers, J.E.; Scott, G.P. (2001). Marine fisheries stock assessment improvement plan: Report of the National Marine Fisheries Service National Task Force for improving fish stock assessments. *U.S. Department of Commerce, NOAA Technical Memorandum NMFS-F/SPO-56*. 69 p.

National Research Council (U.S.). Committee on Fish Stock Assessment Methods. (1998). Improving fish stock assessments. National Academy Press, Washington, DC. 177 p.