AN ASSESSMENT OF TOOTHFISH IN SUBAREA 48.3 USING CASAL

R.M. Hillary✉, G.P. Kirkwood and D.J. Agnew
Division of Biology
Faculty of Natural Sciences
Imperial College
London SW7 1NA
United Kingdom
Email – r.hillary@imperial.ac.uk

Abstract

This paper presents an assessment of the stock of toothfish around South Georgia (Subarea 48.3) using the CASAL stock assessment software (Bull et al., 2005). Detailed attention is given to the incorporation of as much of the available tuning data as possible, as well as a whole range of assessment sensitivities – to fixed parametric assumptions, model structures and alternative data scenarios. Given the integrated nature of the assessment, particular attention is given to rigorous statistical weighting of the various tuning datasets. Bayesian methods are used in the estimation procedure, and uncertainty in the dynamics is explored using Markov Chain Monte Carlo (MCMC) methods; methods for fast approximations to the more time-consuming MCMC tools and CASAL-specific convergence checking tools are also detailed. Finally, long-term yield calculations were undertaken, given the CCAMLR decision rules, for five main assessment candidates.

Résumé

Ce document présente une évaluation du stock de légine autour de la Géorgie du Sud (sous-zone 48.3), effectuée au moyen du logiciel CASAL d'évaluation des stocks (Bull et al., 2005). Les auteurs s'intéressent plus particulièrement à l'insertion du plus grand nombre de données d'ajustement disponibles et de tout un éventail de sensibilités des évaluations aux hypothèses paramétriques fixes, aux structures des modèles et à d'autres scénarios de données. Étant donné la nature intégrée de l'évaluation, une attention toute particulière est accordée à une pondération statistique rigoureuse des divers jeux de données d'ajustement. Les méthodes bayésiennes sont utilisées dans la procédure d'estimation et l'incertitude de la dynamique est explorée par les méthodes de Monte Carlo par chaîne de Markov (MCMC). Les auteurs décrivent des méthodes qui, par une approximation rapide, donnent des résultats proches de ceux obtenus par les outils MCMC, qui sont plus lents, et par les outils de vérification de la convergence spécifiques à CASAL. Finalement, des calculs de rendement à long terme sont réalisés, compte tenu des règles de décision de la CCAMLR, pour cinq propositions principales d'évaluation.

Резюме

В данной работе представлена оценка запасов клыкача в районе Южной Георгии (Подрайон 48.3) с применением программы оценки запаса CASAL (Bull et al., 2005). Подробное внимание уделяется использованию по возможности большего количества имеющихся настроечных данных, а также всего ряда чувствительности оценок – к фиксированным параметрическим допущениям, структуре моделей и альтернативным вариантам данных. С учетом комплексного характера оценки особое внимание уделяется строгому статистическому взвешиванию различных настроечных наборов данных. В процессе оценки используются байесовские методы, а неопределенность динамики исследуется с использованием методов цепей Маркова Монте-Карло (MCMC); подробно описываются также методы быстрой аппроксимации к требующим бóльших затрат времени программам MCMC и специфичные для CASAL методы проверки сходимости. В заключение с учетом правил принятия решений АНТКОМа проведены расчеты долгосрочного вылова для пяти основных возможных оценок.

Resumen

Este estudio presenta una evaluación del stock de austromerluza alrededor de las Islas Georgia del Sur (Subárea 48.3) mediante el software CASAL para la evaluación de stocks (Bull et al., 2005). Se hizo lo posible por incorporar la mayoría de los datos disponibles a la

evaluación, como también una gama de sensibilidades de la evaluación – a las suposiciones relativas a los parámetros fijos, a las estructuras del modelo y a otras condiciones referentes a los conjuntos de datos. Dada la naturaleza integrada de la evaluación, se presta particular atención a la ponderación estadística rigurosa de los diversos conjuntos de datos para las simulaciones. Se utilizan métodos Bayesianos en el procedimiento de estimación, y el efecto de la incertidumbre en la dinámica se explora mediante los métodos de Monte Carlo con cadenas Markov (MCMC); también se describen métodos para obtener aproximaciones más rápidamente que con las técnicas MCMC, y pruebas de comprobación de la convergencia específicas para CASAL. Finalmente, se calculó el rendimiento a largo plazo en las cinco simulaciones de evaluación principales, tomando en cuenta los criterios de decisión de la CCRVMA.

Keywords: stock assessment, MCMC, CCAMLR

Introduction

Until 2004, assessments and estimates of long-term yields for this stock were made using the generalised yield model (GYM) (Constable and de la Mare, 1996). The key input data were estimates of historical annual recruitment calculated from survey data using the CMIX (de la Mare, 1994) program. Agnew and Kirkwood (2004) showed that these recruitment estimates were unreliable, being almost certainly biased downwards, and therefore they were unsuitable for direct use in the GYM calculations. As a short-term ad hoc solution, Agnew and Kirkwood (2004) proposed rescaling the recruitment estimates so that the median 2004 vulnerable biomass calculated by the GYM coincided with either the median or a lower quantile of a mark–recapture estimate of vulnerable biomass. This would then allow calculation of long-term yields satisfying CCAMLR decision rules in the standard way using the GYM. Clearly, however, such an approach is not suitable in the longer term, if for no other reason than that there is now no guarantee that the recruitment estimates provide a consistent relative or absolute index of true annual recruitment.

CASAL (Bull et al., 2005) is an integrated assessment method, capable of fitting to a variety of different types of input data. For this stock, the available data include catch length-frequencies, standardised CPUE data, mark–recapture data and estimates of historical recruitment from surveys around South Georgia. This paper presents the results obtained using CASAL to fit all or a subset of these data. Particular attention has been paid to examination of a comprehensive set of diagnostics and sensitivity tests. For clarity, many of the results presented are point estimates from so-called 'MPD' (maximum posterior density) runs of CASAL, rather than those obtained using the full Markov Chain Monte Carlo (MCMC) procedure, which also takes a rather long time to run. However, results of full MCMC runs are presented for selected assessments that best meet all the goodness-of-fit criteria. A method is also outlined for approximating the MCMC sample, using the information from the MPD runs, which can substantially reduce the computing time for MCMC runs.

Input data

Official CCAMLR catch data were used to calculate catches in tonnes from 1985 to 2005[1] (Table 1). Catch proportions at length were calculated from length-frequency measurements of the catch weighted by catch size. In the early years (1985–1995) only fleet-based measurements were available, and these were often incomplete or concentrated within a single flag type in the fishery. Sufficient data to make reliable estimates of catch proportions were only available from 1992 onwards, and for many of these years the fleets had reported data at different resolutions (for example, fish measured to 3 cm below or to the nearest 5 cm) and standard length composition data had to be reconstructed, based on standard measurements. From 1997, observers were placed on all vessels, and the level of sampling of the catch and its consistency improved markedly. Observer data from 1992 to 2005 were used to produce catch-weighted proportions at length.

CPUE was standardised using a generalised linear mixed model (GLMM), with random vessel effects (Candy, 2004). No data are available for 1990 because the Russian fleet fishing in this year did not report haul-by-haul data to CCAMLR.

Mark–recapture data were obtained from the UK mark–recapture experiment in Subarea 48.3, previously described by Marlow et al. (2003). This provided a dataset of releases at length by year and

[1] All years refer to the fishing season that started in December prior to the year quoted. For example, year 2005 refers to the fishing season 1 December 2004 to 31 November 2005.

recaptures by length by year of recapture and year of release collated by 10 cm length class (Table 2). The scanned catch in each length class each year, required to estimate population size, was calculated by taking the total catch by licensed vessels divided by mean fish weight in the catch and separated to length classes using the catch proportions at length previously described.

The probability of a single tag loss was estimated using the subset of recaptured fish that were originally double-tagged. This confirmed previous estimates of tag loss of 0.06 yr^{-1} (n = 329, maximum time at liberty = 5.5 years). The probability of both tags being lost was therefore 0.0036 yr^{-1}. The immediate post-tag mortality was assessed by a multi-observer experiment (Agnew et al., 2006a) to be between 5 and 11%. A conservative value of 10% (i.e. immediate tagging survivorship of 90%) was used in CASAL and mark–recapture estimates of vulnerable biomass. Analysis of growth rates of tagged fish suggested that immediately following release there is a period of tag-related growth retardation, lasting approximately six months, which was also incorporated into the models. Observers are present on all vessels and a reward scheme is also is in operation. It is therefore assumed that tag reporting rates are equal to 1.

Estimates of recruitment by age and year class are dependent on bottom trawl survey data. Recent analyses of these data have shown that although individual cohorts of age 1–4 fish can be identified from peaks in the length-frequency data and estimates of their density and abundance made using the delta distribution mixture analysis of de la Mare (1994), the observation variance accompanying these estimates is so high as to mask the expected interannual progression of cohort densities (Agnew et al., 2004; Davies et al., 2004).

Natural mortality was assumed to be constant over all ages and years, at an assumed value of M = 0.165, which is consistent with previous assessments of this stock. However, this high natural mortality does not seem to be consistent with the low growth rates of toothfish, and therefore an alternative lower M was incorporated into sensitivity runs. The current growth curve for toothfish at South Georgia is a von Bertalanffy curve, with (k, t_0, L_∞) = (0.066, –0.21, 194.6). Again, there is some uncertainty about this, because the results of age-determination studies suggest a lower L_∞ (Belchier, 2004). An alternative growth model with lower L_∞ was incorporated into sensitivity runs.

The length–weight relationship previously used by CCAMLR was assumed to continue to hold in this analysis. The relationship is weight(kg) = $2.5e^{-5}$ length(cm)$^{2.8}$. Maturity data have in the past been problematic to parameterise because of the large difference between males and females. In 1997, an analysis suggested that for a combined male and female population, the logistic parameters[2] a and b were –6.38 and 0.0686 respectively, with L_{m50} = 93 cm (SC-CAMLR, 1997). These have been characterised in subsequent CCAMLR assessments as L_{m50} = 93 cm and length range for maturity of 30 cm (i.e. zero maturity at 78 cm and 100% maturity at 1 080 cm). These length–weight and maturity ogive data were used in the CASAL assessment.

CASAL model setup

The population model assumed for toothfish in Subarea 48.3, as specified in the CASAL population.csl input file, consisted of a single area, three-season age-structured model, assuming a Beverton-Holt stock-recruit relationship.

The first season was assumed to run from 1 December to 30 April, with recruitment occurring at the start of this season. The second season, where both fishing and spawning take place, lasted from 1 May to 31 August. The remainder of the year (1 September to 30 November) constituted the third and final season.

CASAL can handle a number of different fishing fleets fishing in different years with different selectivities. However, for each fleet, the selectivities-at-age should remain approximately the same in each year in which the fleet fished. Inspection of the catch-length frequencies (Figure 2a) indicated that there was a marked shift in length frequencies between 1997 and 1998, with length frequencies being quite similar both before and after that time. This in turn suggests that there was a change in selectivities-at-age between 1997 and 1998. Accordingly, two 'fleets' were identified: one consisting of all vessels fishing from 1985 to 1997, and the other consisting of all vessels fishing from 1998 onwards. It should be noted that one implication of having two fleets is that the standardised CPUE data have also to be considered as applying to the different fleets in the two sets of years (Table 3).

The specifics of the CASAL estimation routines are described in the manual (Bull et al., 2005). For the purposes of this paper, it is simply necessary

[2] The logistic curve is $p(l) = \frac{\exp(a+bl)}{1+\exp(a+bl)}$, where $p(l)$ is the proportion of mature fish at length l.

to specify the observation error model (probability distribution and error structure) assumed for each of the sets of input data to which the model is fitted.

The first input data are the annual catch proportions-at-length. These were assumed to be independently multinomially distributed, and for this an effective sample size for each year also needed to be specified. The effective sample sizes were estimated using the approach described in Dunn et al. (2005) specified below.

Given catch proportions-at-length, $p_{l,y}$, for each length l and year y, the CV-at-length across years (cv_l) of the proportions-at-length can be estimated by direct or bootstrapping methods. If the proportions-at-length are multinomially distributed, then the CVs should satisfy the following equation:

$$cv_l = \frac{\sqrt{\hat{N}_y p_{l,y}(1-p_{l,y})}}{\hat{N}_y p_{l,y}} \quad (1)$$

where $\hat{N}_y$ is the effective sample size in year y. Given estimates of the CVs in each length class, it is then simple to estimate the effective sample sizes in each year using a non-linear solver. The resulting effective sample sizes are shown in Table 4. Of particular note is the step jump in effective sample sizes that occurred in 1997. This year coincides with the rapid transition to full observer sampling (Table 1).

The CPUE data were assumed to be lognormally distributed, with mean proportional to the model-predicted vulnerable biomass, and variance consisting of two components, the first arising from observation error with a CV equal to that estimated from the GLMM (see Table 1), and an additional (estimated) process error. Given the two-fleet model used here, the CPUE series was split between the fleets in the same temporal fashion as the catch and length-frequency data.

The tagging data are described in more detail later on. As described in Agnew et al. (2004), tagged fish were assumed not to grow for half a year immediately after tagging. A tagging mortality rate of 0.1 yr^{-1} was assumed, and the detection probability of the tags was assumed to be one.

The relative abundance data from the CMIX outputs (Table 3) also come with a specified lognormal CV for each age class. Given this, it is assumed that each survey (by country) had a single catchability, and was lognormally distributed around the model-predicted numbers-at-age. With regard to the tagging data, in CASAL the probability of detecting or not detecting a tagged fish is modelled as binomial, and an over/under-dispersion parameter (applied to all recapture events, not individually) can be used to increase or decrease the weight given the recapture information in the likelihood.

Data weighting and process error

In the integrated assessment framework, data weighting and the distinction between observation and process error are very important concepts, because they are fitted to multiple datasets that may potentially have different implications for most likely parameter values. One standard approach to appropriately weighting the data is to use the principle of iterative re-weighting. In this, initial data weights are first set before starting the estimation (using, for example, the relationship in equation (1) to compute effective sample sizes) and then the same quantities are recomputed after an initial MPD run and the data re-weighted accordingly. In principle, this process would be repeated until convergence is achieved, but in practice one re-weighting proved sufficient. In the rest of the analyses, one re-weighting step was performed for each model, but after this initial re-weighting a check was carried out to ensure that the results only changed marginally with a subsequent re-weighting.

This takes care of the length frequencies and the tagging data, but the CPUE data weighting also needed to be considered. An observation error CV for each standardised CPUE value from the GLMM analysis was already available, but to make sure all potential sources of error were accounted for, a process error CV for the CPUE series was also estimated. This essentially accounts for any extra variance (on top of observation error) that may be required for the population model to interpret the CPUE data. This was done for all assessment runs.

Depending on the assessment model and data used, the parameters to be estimated were:

(i) the virgin spawning stock biomass, B_0;

(ii) for each fleet i, three parameters $a_{1,i}$, $s_{L,i}$ and $s_{R,i}$ defining the double-normal selectivity ogive

$$f(x) = 2^{-\left[(x-a_1)/s_L\right]^2} \text{ for } x \le a_1,$$

and

$$f(x) = 2^{-\left[(x-a_1)/s_R\right]^2} \text{ for } x > a_1;$$

(iii) for each fleet, the catchability coefficients (calculated as nuisance parameters) and process error CVs for the CPUE series;

(iv) the catchability coefficient (again calculated as a nuisance parameter) for the recruitment survey series, calculated for each country.

As this study follows the Bayesian paradigm, prior probability distributions for all the parameters being estimated also had to be assigned. As is customary, it is assumed that all the parameters are *a priori* independent, so that the combined prior distribution is simply the product of each of the individual prior distributions. For the catchability parameters, a log-uniform prior is assigned; this is considered to be the appropriate non-informative prior for scale parameters such as these (Jeffreys, 1961; Box and Tiao, 1973). The same type of prior is applied to the virgin spawning biomass parameter, B_0; it should be noted that a truly non-informative prior for such a parameter cannot be calculated analytically, but a log-uniform prior is more sensible than a uniform prior, and the influence of the prior on the results can always be monitored. Uniform priors were applied to the selectivity ogive and process error parameters and ranges for these priors were set suitably wide.

The CASAL input files for all the runs detailed are available on request from the authors.

Assessment results and sensitivity tests

In this section, the results obtained for a number of different CASAL runs are presented. While the MCMC elements of CASAL were used to determine the full posterior distributions for the baseline and selected alternative assessments in the section 'MCMC estimation of stock status', all results presented in the sections 'Baseline assessment point estimates' and 'Sensitivity analyses' were calculated by estimating the posterior mode (so-called MPD runs). These are very much faster to run, and are therefore ideal for exploratory analyses and sensitivity trials. As indicated earlier, a model in which there were two fleets with different selectivities and catchabilities was identified as the structural baseline assessment.

A final clarification, with respect to data weighting, is for the over/under-dispersion in the tagging data. For all the assessments detailed from here onwards, the estimated value of the ratio between observed and model-predicted dispersion was greater than one, albeit by not much in many cases, which suggested that the tagging data were in fact under-weighted if the distributional assumptions of the tagging data treatment in the likelihood were correct. Given the influence of the tagging data, which will become apparent later on, it was decided not to readjust the tagging weighting, and the dispersion correction factor was left at one in all cases.

Following this, for the baseline assessment only, a series of sensitivity trials was undertaken to examine the effect of:

- including/excluding the survey estimates of younger fish abundance;
- a single-fleet implementation of the assessment model;
- varying the rate of natural mortality, M;
- varying the steepness, h, of the stock-recruit relationship;
- inclusion of alternative growth models;

and to examine the relative influences of the data to which the model fits by:

- removing the CPUE data from the assessment;
- removing the tagging data.

Finally, because of its strong influence, the contribution of each individual year's tag–recapture results were examined. This also facilitates a better comparison with the vulnerable biomass estimates obtained directly from an independent analysis of the tagging data alone.

Baseline assessment point estimates

As described in the previous section, the baseline assessment originally selected attempted to fit, *inter alia*, a separate process error CV for each fleet CPUE series. In practice, MPD runs of CASAL consistently estimated the process error CV for the second (later) fleet CPUE to be at its lower boundary (set at 0.001). Therefore, the model was re-run under the assumption that there was no process error in the second CPUE series, while still retaining a process error term for the first CPUE series. This then became the baseline assessment. Point estimates of the parameters are shown in Table 6.

Figure 1 shows the estimated historical stock dynamics for the baseline assessment. The current (2005) estimate of the spawning stock biomass (SSB) is some 69% of B_0. The current vulnerable

biomass is around 76% of the initial vulnerable biomass, but it should be noted that this comparison is difficult to interpret, as different selectivities apply for the two fleets. What is noticeable is that there is quite a disparity in the magnitudes of the spawner and vulnerable biomasses. This is because of the strongly peaked estimated selectivities, especially for the later period. The two estimated selectivity curves are shown in Figure 2.

The year-class strength plot in Figure 1 shows the relative decrease in recruitment from R_0 over time resulting from the application of the stock-recruit relationship and its assumed steepness. There has been a small decrease. Recent harvest rates have been around 0.08–0.15.

The remaining figures illustrate the fits achieved to the various data sources. In Figures 3a and 3b, fits to the early and later CPUE series are shown. The fit to the early CPUE data is especially poor. This is also reflected in the estimated process error CV for this series of 0.40.

The fits to the length-frequency data for the two fleets are shown in Figures 4a and 4b. Assuming a different selectivity curve (Figure 2) for the two fleets has allowed very good fits to the length-frequency data.

Sensitivity analyses

Given the baseline assessment model used here, this paper now looks at the sensitivity to model structure, assumed fixed values of input parameters, and then to inclusion or exclusion of different datasets.

The first sensitivity test looked at relates to the estimation of toothfish recruitment, and the inclusion of toothfish abundance surveys in the assessment process. When allowing for the estimation of interannual variations from the stock-recruit curve, although the estimation produced a recruitment trend, it was strongly considered that this trend did not represent a believable recruitment pattern; the resultant stock dynamics did not seem sensible either. Figure 5 is a plot of the stock dynamics for this particular assessment model, and it can clearly be seen that there are estimated low recruitments in the earlier years, with subsequent higher recruitments estimated in the years before 1997 – this was the last permitted estimated recruitment, because the last survey began on age-3 fish in the year 2000. One major indication of a problem is the current trend in exploitable biomass. This is very much predicted to be on the increase, which is inconsistent with the CPUE data, and the fit to the current CPUE data is correspondingly poor. The reason for this strange pattern is that the stock must rebuild itself to a specified value of the vulnerable biomass by the years 2004 and 2005, strongly influenced by the tagging data; as a result, the higher recruitments estimated in the mid-1990s do precisely this, thus providing the observed pattern in the exploitable biomass.

One important question is why this recruitment pattern is being estimated. Figures 6a and 6b show bubble plots of the survey data and the length-frequency data. Two things are clear from these plots: the first is that there are no clearly identifiable cohorts moving through the survey data – nothing that would tally with the estimated recruitment trend; the second is that the length-frequency data display a very stable pattern, with no strong or weak cohorts moving through the data – especially not in the later years. However, the estimated recruitment trend improves the fit to both the later length-frequency data and the tagging data. It is hard to accept that the length-frequency data possess any true recruitment information, given the bubble plot in Figure 6b; it is even harder to accept the fact that the tag–recapture data hold any information on recruitment in these years (1985–1997), as recruits from 1997 would have already left the observed tag–recapture age range by 2004 and 2005. This is, in the opinion of the authors, a clear indication of an over-parameterised model, with the interannual stock-recruit deviations essentially being used to improve the fits to data that clearly possess no information on recruitment in these years. For this reason, no estimate of interannual recruitment variations was attempted, and the survey data were not included in any of the assessment runs.

When identifying the baseline assessment scenario, it was the marked change in length frequencies that occurred in 1998 that led to the definition of two fleets, early (pre-1998) and late (1998 and onwards). The substantial difference in the two estimated selectivities (Figure 2) and the excellent fits to the length frequencies (Figure 4) apparently confirm the wisdom of this choice. It is true, however, that this choice required the splitting of the CPUE series. By only including a single post-1996 CPUE point in the first-fleet CPUE series, it is possible that the decline suggested by a straightforward interpretation of the overall CPUE series may have been masked. To examine this possibility, an assessment was carried out in which it was assumed that there was only a single fleet with a single selectivity curve applying throughout the time series. As can be seen from Figures 7 and 8, the fits to length-frequency and CPUE data are poorer for the one-fleet model than those seen in the two-fleet model.

The fit obtained to the CPUE series is shown in Figure 8.

Changing the steepness from 0.8 to 0.7 or 0.9 has only a very small effect on the estimated SSB and vulnerable biomass, with virtually no change at all for the other estimated parameters. As expected, as steepness increases, the estimated decline in SSB decreases, but only slightly. This is not surprising, of course, since the baseline assessment has the SSB, at its lowest, at around 69% of B_0.

The growth parameters used for the baseline model do not fit recent data from the fishery particularly well. The simple least-squares, constant CV-at-age fit to Belchier's (2004) data results in parameters (k, t_0, L_∞) = (0.067, –1.49, 152.8). This is partially a result of the shape of the selectivity curve, which suggests that large fish will be underrepresented in samples of old fish taken from captures made by the longline fishery. But, equally, estimating the selectivity curve accurately requires knowledge of growth parameters. Candy (2005) suggested a method of estimating both selectivity and growth simultaneously, but SC-CAMLR (2005a) was unable to calculate an unbiased estimate of the growth parameters for South Georgia toothfish using his method. Until it is possible to reliably untangle the inter-dependence of these two functions, the ordinary least-squares model was adopted as a sensitivity test, it being the most parsimonious model based on the Belchier (2004) data.

The GYM assessment (SC-CAMLR, 2004) used a uniform distribution of M [0.13,0.2]. A single value of 0.165 (the mid-point of the GYM distribution) was used in the CASAL runs. The sensitivity of the baseline assessment to reducing the value of M to the lower of the previously assumed CCAMLR limits, 0.13, and to the upper end of that range (0.2) was examined. Varying the assumed fixed value of M to the lower and upper ends of the range used in previous assessments has a substantial effect on the results. For the high M, the estimated B_0 is substantially lower than the baseline estimate, but the changes in estimated vulnerable biomasses are much smaller, as are the changes to the estimated selectivity parameters. Again, there is a simple explanation for this. An increase (decrease) in M will decrease (increase) the SSB per recruit, ρ, which relates the initial equilibrium recruitment, R_0, to the virgin spawner biomass: $R_0 = \rho^{-1} B_0$. The process variable, R_0, and M set the initial population age structure and levels. Each of the datasets being fitted by CASAL provides information directly on the current and recent levels of vulnerable biomass – especially the tagging data. This in turn dictates appropriate values of R_0. Consequently, an increase/decrease in ρ, due to a change in M for example, will require a subsequent increase/decrease in the estimate of B_0.

As a final step in examining the sensitivity to M, it was attempted to estimate it along with the other parameters. Bounds of M of [0.05, 0.25] were used. The result was that either M hit the boundaries, or B_0 hit the boundaries for the reasons outlined above due to changes in M, and no reliable estimates were obtained.

The values of M used so far in the sensitivity trials are undoubtedly rather high for an animal with the longevity of toothfish and its relatively low growth rate. One means of comparing values of different biological parameters for a fish stock is to examine the so-called life-history invariants calculated using life-history optimisation techniques (Charnov, 1993; Jensen, 1996; Beddington and Kirkwood, 2005). In the context of the standard Beverton and Holt dynamics, there are three Beverton-Holt invariants:

$$M^*T_m = 1.65;\ M/K = 1.5;\ \text{and}\ L_m/L_\infty = 0.67$$

where T_m and L_m are the age and length at (knife-edged) maturity.

Table 7 examines the Beverton-Holt invariant values with different combinations of growth and maturity parameters. Reducing natural mortality has the largest single effect on the invariants, and the combination of parameters that is closest to the expected values is that with low L_∞ (152.8 cm) and low M (0.13). Accordingly, the combined low L_∞ and M = 0.13 were included in the sensitivity trials.

The second set of sensitivity analyses investigated the relative information content of the different datasets used in the estimation. This is done by omitting either the CPUE data or the tagging data from the full baseline datasets and then re-estimating the selectivities. Note that the length frequencies must always be included, since otherwise there will be no information available to estimate the selectivities.

Omission of the CPUE data effects a minimal change in the results, with slightly lower virgin and current SSB and vulnerable biomass. When omitting the tagging data, a much more noticeable reduction in virgin and current SSB and vulnerable biomass is achieved. It is thus reasonable to infer that the tagging data contain comparatively strong information on the current vulnerable biomass.

By comparing estimates obtained using all data up to 2005 with an assessment using only data to 2004, a short retrospective analysis was carried out. Given the decision earlier to use only tag recaptures in 2004 and 2005, this analysis can only go back one year.

Given the importance of the tagging data, it is of interest to examine their components in more detail. This study is now restricted to an estimation using only length-frequency data and tag recaptures either from 2004, 2005 or both 2004 and 2005.

Clearly, whilst the inclusion or exclusion of the two main datasets (2004 and 2005) has some effect, this is generally not large (Table 8). The very close correspondence with the main model (using all data) and one using only length frequency and the last two years' tagging data demonstrates the importance of the latter data to the model.

These results were compared to those obtained simply using the modified Petersen estimator described by Agnew and Kirkwood (2004). Seber (1982) gives the form of the Petersen estimator implemented using Bailey's binomial adjustment as

$$\hat{N}_Y = \frac{n_Y(c_Y+1)}{(m_Y+1)} \tag{2}$$

$$\mathrm{var}(\hat{N}_Y) = \frac{n_Y^2(c_Y-m_Y)}{(m_Y+1)^2(m_Y+2)} \tag{3}$$

where $\hat{N}_Y$ is the estimate of population size in year Y, n_Y is the number of marked animals in the population prior to taking the sample in year Y, c_Y is the number of animals in the sample in year Y (which equals the number caught in the fishery in year Y) and m_Y is the number of marked animals in the sample. Accounting for growth and selectivity, n is found as

$$n_Y = \sum_{y=1,a=1,z=1}^{Y-1,\max age,2} T_{y,a,z}(1-p)(1-l^z)^{Y-y}(e^{-M(Y-y)})s_{a+Y-y-r} \tag{4}$$

where $T_{y,a,z}$ is the number of fish tagged in month/year y of age a with z tags, $(1–p)$ is the proportion surviving the initial tagging (0.9), l is the tag-loss rate per year for a single tag (0.06), M is the natural mortality rate and $s_{a+Y-y-r}$ is the relative selectivity of fish that started off at age a in month/year y when they have grown older at year Y, compensated for tag-related growth retardation r (0.5 of a year). Note that fish that were tagged and recaptured in the same season do not appear in either the first or the second parts of equation (4); they are omitted from the calculations because there is insufficient time during a four-month fishing season for sufficient mixing to have occurred. The instantaneous date assumed for estimates was set to 31 June (mid-season), so as to approximate the same time settings as used in the CASAL model.

The modified Petersen method yielded estimates of vulnerable biomass of 50 600 tonnes (95% CI 36 400–64 700) for 2004 and 51500 tonnes (42 600–60 400) for 2005. Although these are a little higher than the estimates from CASAL (46 000–48 000 tonnes) (Table 9), the confidence intervals overlap with all the results shown in Table 10. The minor discrepancies are probably produced by a combination of slightly different estimation methods, slightly different handling of growth and mean weights, and the fact that CASAL is an integrated, rather than a single, assessment method. Note also that the Petersen results are not directly comparable with Table 9 in that those in Table 9 are for VB_{2005} using either 2004 or 2005 data, whereas the Petersen results are for VB_{2004} and VB_{2005} using 2004 and 2005 data respectively.

The tagging data provide highly consistent estimates of population size in 2004 and 2005, whether analysed by the Petersen or CASAL methods. Using either 2004 or 2005 data gives practically the same answer. This reasonably implies that tags are now effectively mixed in the toothfish population at South Georgia because tagging, releasing and recapture fishing effort has taken place over the whole distribution of the fishery and main distribution of adult toothfish (Agnew et al., 2006b). It could be argued that it could also imply that both estimates are biased in the same way, but the increased number of fish being tagged and returned over this two-year period would have presumably yielded inconsistencies, if there were any issues with the mixing of the tagged and untagged populations.

Summary

Given the many sensitivity trials undertaken, five scenarios were taken forward for MCMC runs, and calculating the long-term yield under the CCAMLR decision rules. These models were the following:

- baseline two-fleet model;
- single-fleet model, which is similar to the baseline in terms of results, but is structurally quite different;

- low L_∞ growth model, which could be considered to be a more pessimistic model than the baseline;
- low M, which could be considered to be a more optimistic model than the baseline;
- low L_∞ and low M, which is the combination of parameters that most closely satisfies the Beverton-Holt invariants.

In each scenario, the models were fitted to the catch-length frequencies, CPUE data and tagging data. It was decided to take these models forward for use in the yield calculations and MCMC runs because they represent the key subset of sensitivity trials which cover all uncertainties regarding both parametric assumptions and model structure.

MCMC estimation of stock status

CASAL supports the facility to extract a sample, using MCMC techniques, from the parameter and process variable posterior distribution. Until now, only the mode of this posterior distribution has been estimated, but to gain an insight into the inherent uncertainty involved in the assessment process, this MCMC feature of CASAL was also used for this assessment. For the baseline assessment, 1 000 000 samples from the posterior were drawn, after a burn-in period of 100 000 iterations, and thinned the resulting Markov chain by a factor of 1 000 to yield 1 000 samples from the posterior of interest. Given the computational intensity of this process, only one Markov chain was generated, but two convergence tests (see Appendix 1) were applied to the former and latter halves of this chain, along with a visual check on the chain time series and histograms, to check for convergence of this Markov chain on the posterior. Both of these convergence tests, as well as the more standard visual tests, were passed, and the authors were satisfied that the chain had indeed converged on the posterior.

Figure 9 shows the trace plot and histogram for B_0 coming from the MCMC estimation. There is clearly no prior forcing for this, or indeed any of the other parameters.

The computational burden of running the MCMC simulations is large (around 24 hours using a 4 GHz processor), but with the information coming from the CASAL MPD estimation runs, an MCMC sample that is an approximation of a true MCMC sample from the model posterior can quickly be generated. The theory is as follows: the posterior distribution of interest can be approximated by using a multivariate normal distribution (Bernardo, 2003), with the mean defined as being the posterior mode, and variance–covariance matrix given by the inverse Hessian of the posterior at the posterior mode – both of which are outputs from the CASAL MPD run. Generating draws from a multivariate normal distribution is comparatively easy, and it takes only around two seconds to generate 1 000 samples. This technique was used when looking at the projections for all models used in this paper, and it performed well, as long as there was no overly strong skew in the marginal posteriors of the parameters. As an example, the median and 95 percentiles from the full MCMC sample of B_0 from the baseline MCMC assessment were 177 340 (157 732–202 105), while from the approximate MCMC sample they were 177 568 (153 477–199 302). The difference in the quantiles of the two samples was never more than around 3%, and any such differences occurred largely in the tails of the distributions. It is worth noting that, even though the rigorous theory does not exist for this case, one could perhaps use a multivariate t-distribution instead, to reduce the discrepancy seen in the tails of the distributions of the two samples. What is clear is that, under certain caveats, this approximation method can drastically reduce the computational burden of producing the MCMC samples, and it also performs well in comparisons with the true MCMC samples. Table 9 summarises the MCMC results for the five main assessment models.

The results in Table 9 suggest that, in terms of SSB depletion at least, the alternative growth model case is the most pessimistic and the lower natural mortality case the most optimistic, with the combined alternative growth and natural mortality case being intermediate. The baseline and single-fleet assessments are roughly equivalent; clearly, although the choice of model is important, it is not as influential as the growth and mortality parameters. Figure 10 shows a plot of the median historic SSB for all scenarios for which MCMC runs were performed.

Long-term yield calculation

CASAL also allows for stochastic/MCMC projections for a given catch level, which makes it a potential tool for determining the long-term catch limit, based on the two CCAMLR decision rules:

1. The future SSB must not drop below 20% of the median B_0 more than 10% of the time.
2. The final SSB must have a probability of 0.5 or greater of being above 50% of the median B_0.

The maximum catch which satisfies both these conditions is the estimate of the long-term yield.

For the purposes of projections, only the MCMC samples were used for calculating the long-term yield. Stochasticity in historic and future recruitments was introduced in the projection simulations by using a lognormally distributed annual year-class strength multiplier for both historical and future recruitments. The value previously used for these stochastic recruitment variations was $\sigma = 0.8$ (based on a CV of 0.95 from the surveys), but it was agreed at the 2005 meeting of SC-CAMLR's Working Group on Fish Stock Assessment to use a lower value of $\sigma = 0.7$ to account for the fact that MCMC methods are used, and stock uncertainty is already included in the estimation process. It is felt that this leads to a pertinent discussion on this topic, with respect to using point-estimate and MCMC methods when performing projections.

As already detailed, it is felt that deviations from the stock-recruit curve could not be estimated/detected reliably, and so recruitment was, henceforth, determined by the stock-recruit relationship alone. Any uncertainty in the recruitment in the subsequent historic MCMC-predicted recruitment values should, therefore, be determined ultimately by the variance structure in the MCMC parameter samples – there are variations in recruitment, but the temporal trend in these variations cannot be quantified directly. This has implications, with respect to projections, as it is technically incorrect to then randomise the historic recruitments – if it is accepted that the data weighting has been done correctly, then the resultant posterior variance in indices such as recruitment has been fully realised. When projecting into the future, the only argument for randomising the recruitments is a precautionary one, but the correct value of the variance of the noise is far from clear.

The posterior CV in the historic recruitment values is around 10%; the largest estimated value of the CPUE process error CV (which can be, in part, attributed to variations in recruitment) is 0.45 – both of these are well below those assumed in the past for projections performed for this stock. The effect of applying lower levels of future uncertainty is not currently known, and there are no definitive answers or suggestions as to what the correct value might be. However, if one is using MCMC methods to explore uncertainty, then recruitment randomisation should not be performed historically, as this breaks the correlative structure in the MCMC samples, and is adding uncertainty twice, albeit in a different manner. The issue of how to introduce stochasticity into such models is also touched on, with respect to natural mortality, in the paper by Xiao (2006), and it is felt that it is a question that should be addressed when assessing and managing fish stocks in this manner. Table 10 shows the long-term yields, calculated using the CASAL projection option.

Discussion

This paper presents an assessment of the South Georgia stock of Patagonian toothfish (*Dissostichus eleginoides*) using the CASAL (Bull et al., 2005) stock assessment package. Previous assessments of toothfish at South Georgia have used a stochastic projection method, which had as its source data absolute estimates of recruitment strength obtained from trawl surveys. Such assessments had a number of drawbacks, including the difficulty of obtaining precise estimates of recruitment and the assumptions behind projecting recruitment at age 3 to the vulnerable biomass, mainly on ages 8–12. It is worth noting that, even though the fits to the survey data were very poor and these data were not subsequently used, the estimated values of q for each country's surveys ranged from 0.041 to 0.404 – all less than the value of 1 implicitly assumed in the GYM.

As this was the first fully integrated assessment of this stock, a wide range of sensitivity trials were undertaken, and attention was also directed to achieving the correct weighting for each of the datasets used in the fitting.

Some of the more standard sensitivity analyses, such as the removal of datasets and the presence of retrospective patterns, were performed. What was clear was that the tagging data exerted a strong influence over the dynamics – they contain strong current absolute abundance information. As a result, this imposed a consistent pattern on the historical dynamics. When the tagging data were omitted, the CPUE data and the length-frequency data yielded a similar, yet slightly smaller, estimate of current stock size; when the CPUE data alone were omitted, the estimated stock size barely changed, and the estimates were more precise when the tagging data were present. This should not be so surprising, as relative abundance data of this kind (the 'one-way trip' kind) and length-frequency data alone have, historically, sometimes proved to be unreliable sources of information on absolute stock size (Payne et al., 2005). Tagging data of the quality of reporting rate and empirical knowledge of other mark–recapture parameters enjoyed in the South Georgia tagging program would naturally represent more informative data with respect to absolute stock size than CPUE and length-frequency data alone.

Tagging data provided a very consistent view of current stock status. There was very little change when all returns data were included (from 2002 to 2005), or when only the 2004/05 returns were included. When the 2004 returns alone were used, the estimate of stock size was a little larger than when only the 2005 returns were included – but not by much. On the whole, it would have to be concluded that the tagging data are, thus far, giving a consistent picture of current stock size. This is reinforced by the conclusions of Agnew et al. (2006b) that the current tagging program is creating effective mixing between the tagged and untagged populations. Furthermore, there was good correspondence between the results of the modified Petersen mark–recapture estimate of current vulnerable biomass and the CASAL integrated model estimate, when the latter included tag data. A one-year retrospective analysis was performed (using only data up to and including 2004), and this gave a slightly higher estimate of virgin stock size, but very similar estimates of current stock levels.

Sensitivities to some of the parametric (value of natural mortality, value of the fixed steepness, growth parameters) and structural (single-fleet; two-fleet) assumptions of the assessment model were also performed. Natural mortality was fixed at its currently accepted lower and upper bounds (M = 0.13 and 0.2) and the steepness was also fixed at h = 0.7 and 0.9. For all these cases, the current estimates of exploitable biomass were very similar, as the tagging data are highly informative on these.

What was clear was that the assessment model is most sensitive to the growth curve and level of natural mortality assumed, yielding a range of current SSB depletion factors of 0.51–0.74. In particular, with respect to the two growth curves applied, for the historical growth curve, strongly dome-shaped selectivities were estimated. This raised the issue of a potential cryptic biomass never seen in the catches, and of selecting immature fish. However, when using the growth curve with a lower L_∞, the selection pattern was then estimated to be targeted largely at mature fish, and also suggested that it would then be very unlikely that any sort of cryptic population existed.

Consideration of the Beverton-Holt invariants led to the conclusion that the combination of low M and low L_∞ is more appropriate as a parameter set than the baseline or either low M or low L_∞ on their own. This is also the approach adopted by CCAMLR (SC-CAMLR, 2005b). The assessment using both low M and low L_∞ indicated that currently spawning biomass is at about 59% of its virgin level (B_0 = 109 047, B_{2005} = 63 690, VB_{2005} = 52 934). Accordingly, yield calculations project a trajectory of spawning biomass that has only a small slope, so that the spawning biomass is reduced to 50% of the virgin level over the 35-year projection window. Clearly, this assessment suggests that the South Georgia toothfish population is now almost fully exploited. This is not an unreasonable assumption, given the relatively long exploitation history of the stock (20 years).

Significant uncertainties continue to exist, particularly in the determination of natural mortality and growth rate. Some progress is possible on these issues. M can be estimated from mark–recapture data (Seber, 1982), although in this case the situation will be confounded by the previously noted interrelationship between selectivity and estimated growth parameters (Candy, 2005). Given the development of age–length keys and ageing of the recaptured fish, it is hoped that a method of estimating M from the mark–recapture data can be implemented. Additional age determination may also throw some light on an appropriate growth rate, but again this is confounded by selectivity. One way of avoiding this problem would be to re-cast the CASAL model in terms of age rather than length, and use data on removals-at-age (from random capture-at-age sampling or using age–length keys). A final issue that could be examined is the effect of the inclusion of sexual dimorphism in the model, given that the species is significantly dimorphic, demonstrated by different lengths at sexual maturity and growth parameters for males and females.

Acknowledgements

The authors would like to dedicate this paper to the memory of Dr Geoff Kirkwood, who sadly passed away before the paper was published, but did so much in inspiring and carrying out both the work detailed in this paper, and that which this paper has relied and built upon.

The authors would also like to thank the two reviewers of the paper for helpful and insightful suggestions.

References

Agnew, D.J. and G.P. Kirkwood. 2004. Assessment of the status of the toothfish stock in Subarea 48.3. Document *WG-FSA-04/82*. CCAMLR, Hobart, Australia.

Agnew, D.J., J. Moir Clark, R.C. Wakeford, M. Collins and M. Belchier. 2004. Survey

estimates of recruitment of toothfish in Subarea 48.3. Document *WG-FSA-SAM-04/16*. CCAMLR, Hobart, Australia.

Agnew, D.J., J. Moir Clark, P.A. McCarthy, M. Unwin, M. Ward, L. Jones, G. Breedt, S. Du Plessis, J. Van Heerdon and G. Moreno. 2006a. A study of Patagonian toothfish (*Dissostichus eleginoides*) post-tagging survivorship in Subarea 48.3. *CCAMLR Science*, 13: 279–289 (this volume).

Agnew, D.J., G.P. Kirkwood, J. Pearce and J. Clark. 2006b. Investigation of bias in the mark–recapture estimate of toothfish population size at South Georgia. *CCAMLR Science*, 13: 47–63 (this volume).

Beddington, J.R. and G.P. Kirkwood. 2005. The estimation of potential yield and stock status using life-history parameters. *Phil. Trans. R. Soc. Lond. B*, 360: 163–170.

Belchier, M. 2004. The age structure and growth rate of Patagonian toothfish (*Dissostichus eleginoides*) at South Georgia. Document *WG-FSA-04/86*. CCAMLR, Hobart, Australia.

Bernardo, J.M. 2003. Bayesian statistics. In: *Encyclopedia of Life Support Systems (EOLSS). Probability and Statistics*. Oxford University Press, London.

Box, G.E.P and G.C. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. Wiley, New York.

Brooks, S.P. and G.O. Roberts. 1998. Assessing convergence of Markov chain Monte Carlo algorithms. *Statistics and Computing*, 8: 319–335.

Bull, B., R.I.C.C. Francis, A. Dunn, A. McKenzie, D.J. Gilbert and M.H. Smith. 2005. CASAL User Manual v2.07-2005/07/06. *NIWA Technical Report*, 126.

Candy, S.G. 2004. Modelling catch and effort data using generalised linear models, the Tweedie distribution, random vessel effects and random stratum-by-year effects. *CCAMLR Science*, 11: 59–80.

Candy, S.G. 2005. Fitting a von Bertalanffy growth model to length-at-age data accounting for length-dependent fishing selectivity and length-stratified sub-sampling of length frequency samples. Document *WG-FSA-SAM-05/13*. CCAMLR, Hobart, Australia.

Charnov, E.L. 1993. *Life History Invariants: Some Explorations of Symmetry in Evolutionary Ecology*. Oxford University Press, New York: 167 pp.

Constable, A.J. and W.K. de la Mare. 1996. A generalised model for evaluating yield and the long-term status of fish stocks under conditions of uncertainty. *CCAMLR Science*, 3: 31–54.

Davies, C.R., E.M. van Wijk and A.J. Constable. 2004. Theoretical considerations for estimating the density of cohorts and mean recruitment of Patagonian toothfish (*Dissostichus eleginoides*) based on research trawl survey data. Document *WG-FSA-04/92*. CCAMLR, Hobart, Australia.

de la Mare, W. 1994. Estimating krill recruitment and its variability. *CCAMLR Science*, 1: 55–69.

Dunn, A., D.J. Gilbert and S.M. Hanchet. 2005. Further development and progress towards evaluation of an Antarctic toothfish (*Dissostichus mawsoni*) stock model for the Ross Sea. Document *WG-FSA-SAM-05/12*. CCAMLR, Hobart, Australia.

Jeffreys, H. 1961. *Theory of Probability*. Oxford University Press, London.

Jensen, A.L. 1996. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. *Can. J. Fish. Aquat. Sci.*, 53 (4): 820–822.

Marlow, T.R., D.J. Agnew, M.G. Purves and I. Everson. 2003. Movement and growth of tagged toothfish *Dissostichus eleginoides* around South Georgia and Shag Rocks (Subarea 48.3). *CCAMLR Science*, 10: 101–111.

Payne, A., G.P. Kirkwood, R. Hillary and D.J. Agnew. 2005. Age-structured production model for toothfish at South Georgia. Document *WG-FSA-SAM-05/18*. CCAMLR, Hobart, Australia.

SC-CAMLR. 1997. Report of the Working Group on Fish Stock Assessment. In: *Report of the Sixteenth Meeting of the Scientific Committee (SC-CAMLR-XVI)*, Annex 5. CCAMLR, Hobart, Australia: 239–425.

SC-CAMLR. 2004. Report of the Working Group on Fish Stock Assessment. In: *Report of the Twenty-third Meeting of the Scientific Committee (SC-CAMLR-XXIII)*, Annex 5. CCAMLR, Hobart, Australia: 339–658.

SC-CAMLR. 2005a. Report of the Working Group on Fish Stock Assessment. In: *Report of the*

Twenty-fourth Meeting of the Scientific Committee (SC-CAMLR-XXIV), Annex 5. CCAMLR, Hobart, Australia: 289–561.

SC-CAMLR. 2005b. *Report of the Twenty-fourth Meeting of the Scientific Committee (SC-CAMLR-XXIV)*. CCAMLR, Hobart, Australia: 647 pp.

Seber, G.A.F. 1982. *The Estimation of Animal Abundance and Related Parameters*. Second Edition. Charles Griffin and Co. Ltd., London: 654 pp.

Xiao, Y. 2006. Several traps in size-dependent population models published in fisheries journals. *Fish. Res.*, 77 (3): 265–370.

Table 1: Annual catches of toothfish, sample sizes for estimation of catch proportions at length (vessel data in italics, observer data in normal type), a generalised linear mixed model (GLMM) estimate of standardised CPUE and its CV.

Season	Catch (tonnes)	Number of fish measured	GLMM CPUE (kg/1 000 hooks)	CV (%)
1985	521	*2 103*	0.253	50.18
1986	733	*8 371*	0.369	49.22
1987	1 954	*0*	0.713	49.73
1988	876	*0*	0.885	36.11
1989	7 204	*0*	0.524	36.85
1990	7 222	*5 302*		
1991	3 531	*2 588*	0.565	36.47
1992	6 871	*20 138*	0.623	9.93
1993	7 039	*6 466*	1.067	9.74
1994	5 438	*11 698*	0.671	10.26
1995	4 998	*14 550*	0.554	9.26
1996	3 542	10 496	0.302	9.18
1997	3 812	82 887	0.259	9.13
1998	3 347	81 275	0.259	9.18
1999	4 303	55 074	0.280	9.14
2000	5 919	47 374	0.283	9.09
2001	4 234	74 056	0.244	9.09
2002	5 722	108 342	0.251	9.09
2003	7 513	86 549	0.261	9.06
2004	4 447	51 879	0.224	9.11
2005	3 000	40 909	0.212	9.18

Table 2: The release–recapture matrix for Subarea 48.3 used in the assessment.

Release year	Number released	Number recaptured by recapture year				
		2001	2002	2003	2004	2005
2000	135	1	1	3	1	2
2001	347		29	38	15	4
2002	401			42	8	16
2003	355				23	12
2004	2 914					93
2005	3 944					
Total		1	30	83	47	127

Table 3: CMIX-derived estimates of numbers-at-age (with their associated CV in brackets) from the South Georgia toothfish surveys. Small roman numerals indicate years for which two sets of survey estimates are available.

Age	Country	3		4		5		6		7	
1987	US	234 761	(0.04)	890 137	(0.34)	1 085 772	(0.16)	73 362	(0.93)	n/a	
1990	UK	83 320	(1.22)	1 106 314	(0.42)	648 050	(0.55)	356 427	(0.45)	143 496	(1.03)
1991	UK	3 605 231	(0.37)	225 789	(0.49)	236 894	(0.56)	1 617 542	(0.75)	2 254 195	(1.07)
1992	UK	525 799	(0.34)	5 957 678	(0.23)	306 371	(0.77)	579 621	(0.41)	n/a	
1994(i)	UK	1 465 903	(0.31)	1 312 447	(0.48)	1 570 898	(0.43)	92 880	(1.70)	76 727	(0.32)
1994(ii)	ARG	217 924	(1.42)	98 065	(1.59)	1 394 715	(0.20)	14 528	(7.25)	n/a	
1995	ARG	824 263	(1.66)	937 955	(0.57)	3 642 190	(0.26)	2 221 056	(0.24)	n/a	
1996	ARG	837 148	(0.32)	2 787 619	(0.37)	297 748	(0.80)	1 324 766	(0.41)	293 433	(0.75)
1997(i)	ARG	321 481	(0.71)	671 814	(0.31)	774 853	(0.38)	803 704	(0.50)	746 002	(0.43)
1997(ii)	UK	95 163	(0.52)	165 501	(1.88)	1 874 304	(0.37)	405 478	(1.65)	910 257	(0.41)
2000	UK	1 134 828	(0.34)	593 478	(0.36)	240 599	(0.72)	324 809	(0.78)	1 951 082	(0.17)

Table 4: Effective sample sizes for catch proportions at age.

1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
79	81	114	135	116	243	227	198	320	287	343	305	300

Table 5: Priors applied for the parameters estimated.

Parameter	Prior distribution
B_0	Uniform-log on (20000, 1000000)
q (for all catchabilities)	Uniform-log on (1e-11, 1e-1)
a_1 (Max. selection point)	Uniform on (1, 50)
s_L (Left selectivity decay)	Uniform on (0.05, 500)
s_R (Left selectivity decay)	Uniform on (0.05, 500)
Process error CV	Uniform-log on (0.01, 10)

Table 6: Point estimates of initial (B_0) and current (B_{2005}) SSB, initial (VB_0) and current (VB_{2005}) vulnerable biomass, and the selectivity parameters for the baseline assessment.

Assessment	B_0	B_{2005}	VB_0	VB_{2005}	Fleet 1 selectivities			Fleet 2 selectivities			Process error CV, fleet 1
					$a_{1,1}$	$s_{L,1}$	$s_{R,1}$	$a_{1,2}$	$s_{L,2}$	$s_{R,2}$	
Baseline	176 966	123 390	61 657	46 966	9.21	2.17	4.53	6.94	0.75	3.89	0.39

Table 7: Beverton-Holt invariants given the growth and mortality parameters tested in the sensitivity trials. L_m was 93 cm.

	Invariant			Reduction in normalised sums of squares from baseline (%)
	M^*T_m	M/K	L_m/L_∞	
Expected Value	*1.65*	*1.50*	*0.67*	
Baseline	1.58	2.50	0.48	
M = 0.13	1.25	1.97	0.48	55
Low L	2.06	2.46	0.61	8
Low L_∞ and M = 0.13	1.63	1.94	0.61	82

Table 8: Summary table of results for the sensitivity trials undertaken, with point estimates of initial (B_0) and current (B_{2005}) SSB, initial (VB_0) and current (VB_{2005}) vulnerable biomass, and the selectivity parameters.

Assessment	B_0	B_{2005}	VB_0	VB_{2005}	Fleet 1 selectivities			Fleet 2 selectivities			Process error CV, fleet 1
					$a_{1,1}$	$s_{L,1}$	$s_{R,1}$	$a_{1,2}$	$s_{L,2}$	$s_{R,2}$	
Baseline	176 966	123 390	61 657	46 966	9.21	2.17	4.53	6.94	0.75	3.89	0.39
One fleet	163 986	108 887	82 228	61 845	7.73	1.37	5.35	n/a	n/a	n/a	0.46
h = 0.7	178 792	124 648	62 295	46 972	9.2	2.17	4.53	6.95	0.74	3.88	0.39
h = 0.9	175 539	122 414	61 159	46 972	9.21	2.17	4.53	6.95	0.75	3.89	0.39
Low M = 0.13	267 934	197 672	64 861	49 815	9.02	2.15	4.26	6.92	0.76	3.8	0.39
High M = 0.2	121 212	79 160	64 287	48 193	9.41	2.2	4.87	6.94	0.76	4.56	0.39
No CPUE	175 930	122 344	61 296	46 596	9.21	2.17	4.59	6.94	0.74	3.89	n/a
No tags	140 688	87 448	50 280	34 947	9.05	1.99	4.94	7.01	0.78	3.96	0.38
Data to 2004	188 530	133 129	63 833	47 925	9.18	2.16	4.52	6.86	0.69	3.91	0.39
LF + tags 2004/05	174 598	121 460	60 830	46 636	9.22	2.18	4.51	6.94	0.74	3.88	n/a
2004 tags only	179 408	126 252	62 510	48 352	9.21	2.17	4.5	6.93	0.74	3.88	n/a
2005 tags only	171 639	118 513	59 797	45 582	9.23	2.18	4.51	6.94	0.74	3.88	n/a
Low L_∞	70 372	35 464	82 296	52 746	12.4	3.22	13.8	8.1	1.07	12.5	0.36
Low L_∞ and low M	109 047	63 690	78 502	52 934	10.9	2.96	8.90	7.5	0.92	8.1	0.37

Table 9: Median and 95% confidence intervals for the virgin SSB, the current SSB, the ratio of current to initial SSB, and the initial and current vulnerable biomass from the MCMC samples.

Model	B_0 (thousand tonnes)		B_{2005} (thousand tonnes)		B_{2005}/B_0		VB_0 (thousand tonnes)		VB_{2005} (thousand tonnes)	
Baseline	177	(157–202)	124	(105–149)	0.69	(0.66–0.74)	62	(55–70)	47	(40–56)
One fleet	164	(140–188)	109	(85–134)	0.67	(0.61–0.71)	82	(70–95)	61	(49–75)
Low L_∞	70	(62–78)	36	(27–44)	0.51	(0.44–0.56)	83	(72–97)	53	(44–65)
Low M	267	(235–300)	197	(163–230)	0.74	(0.71–0.76)	65	(57–73)	50	(42–58)
Low L_∞ and low M	109	(96–122)	64	(51–76)	0.59	(0.53–0.63)	79	(69–92)	53	(44–64)

Table 10: Long-term yields (in tonnes) meeting the CCAMLR decision rules, for each CASAL assessment model, using the MCMC CASAL projection method.

Model	MCMC projection
Baseline	5629
One fleet	5428
Low L_∞	3407
Low M	5876
Low L_∞ and low M	3743

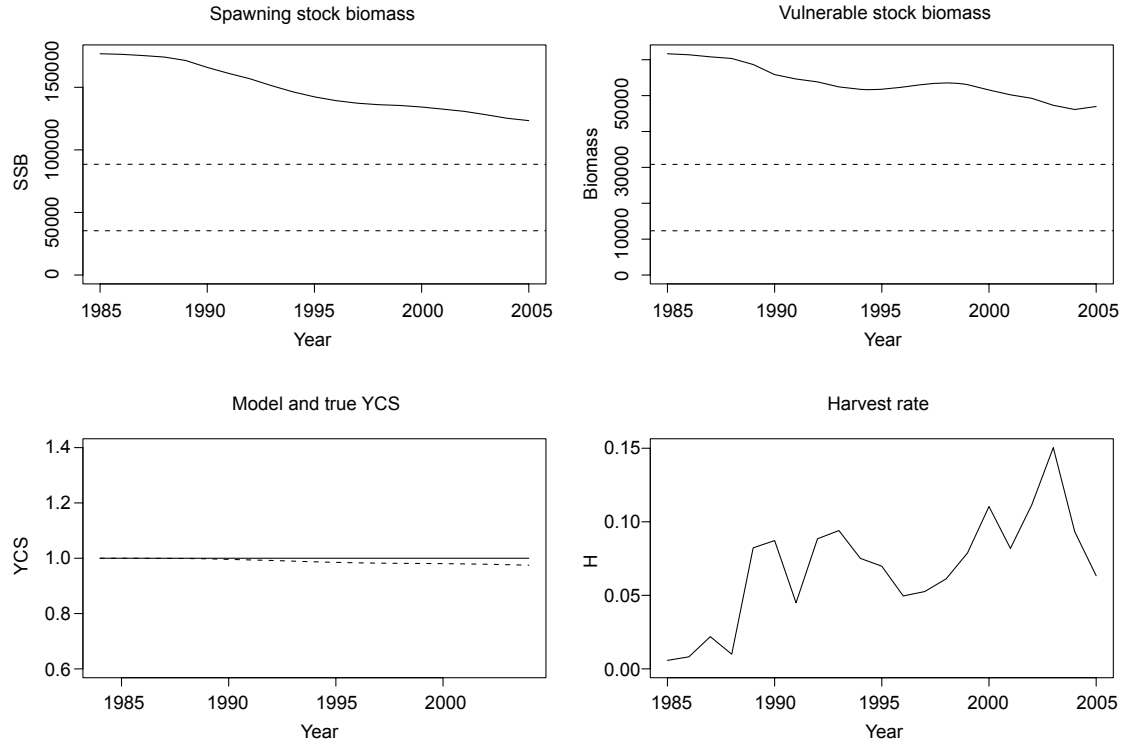


Figure 1: Historical stock dynamics for the baseline assessment. The upper and lower lines represent 50 and 20% of the virgin spawner biomass/vulnerable biomass respectively.

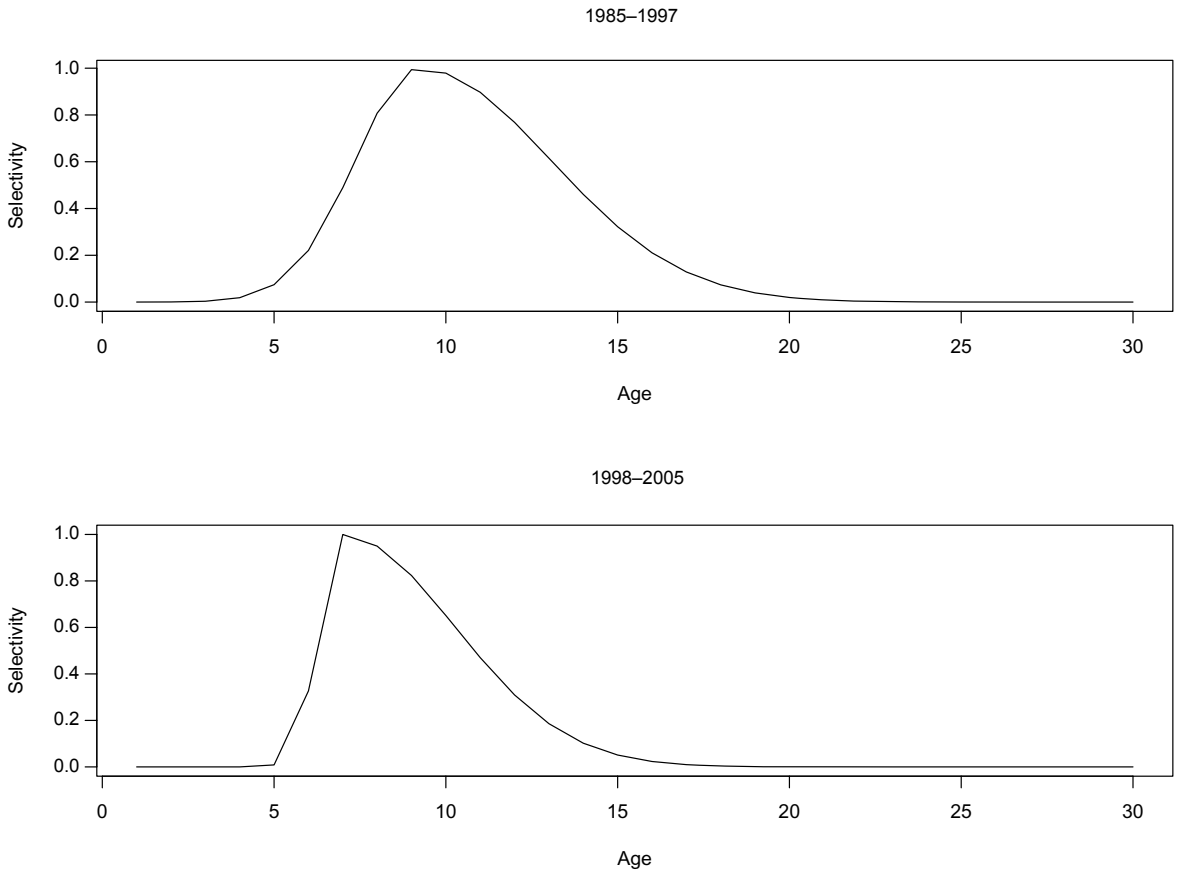
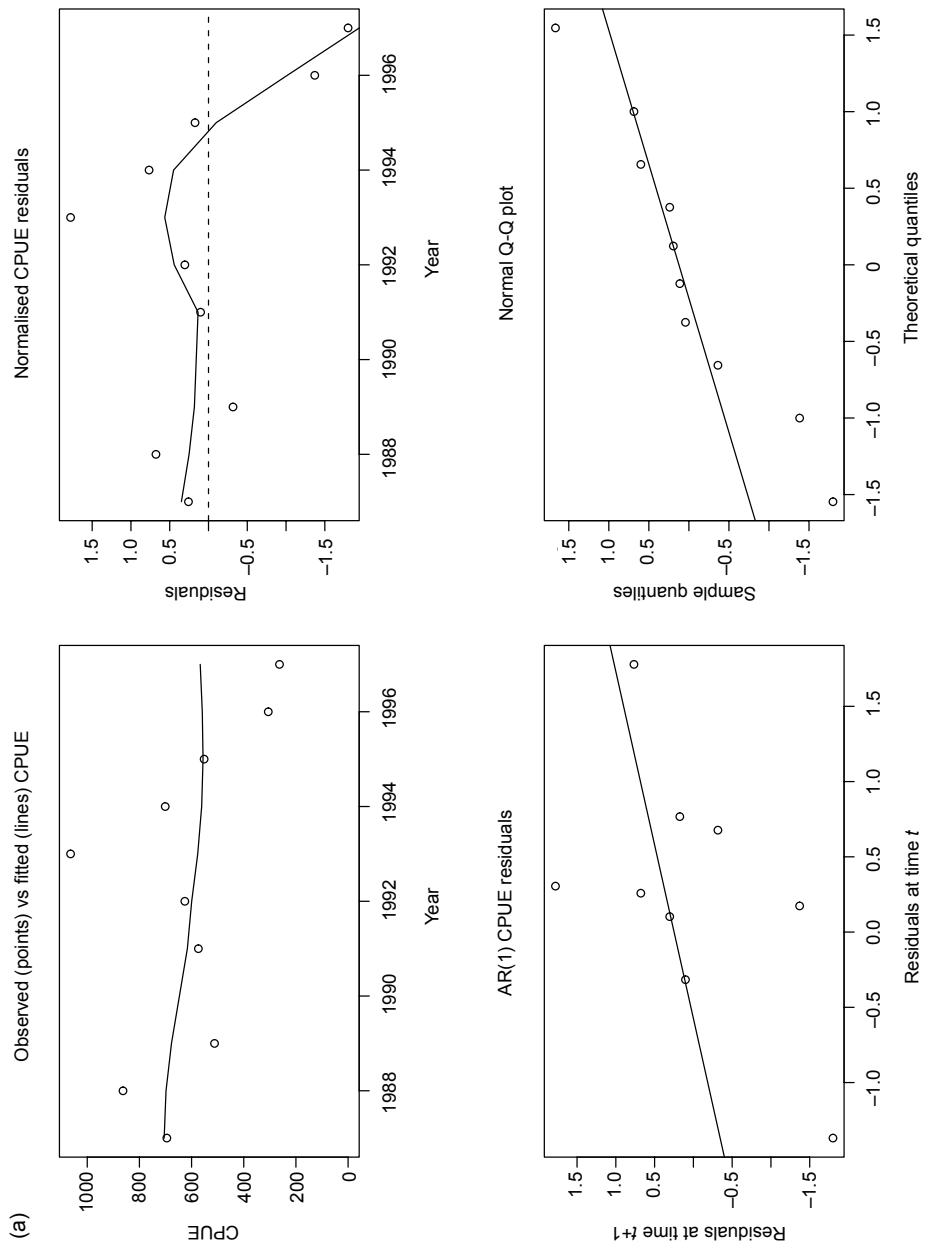


Figure 2: Estimated selectivity curves for the early (up to 1997) and later (1998 to present) fleets.


(a)
Observed (points) vs fitted (lines) CPUE
CPUE
Year
Normalised CPUE residuals
Residuals
Year
AR(1) CPUE residuals
Residuals at time t+1
Residuals at time t
Normal Q-Q plot
Sample quantiles
Theoretical quantiles

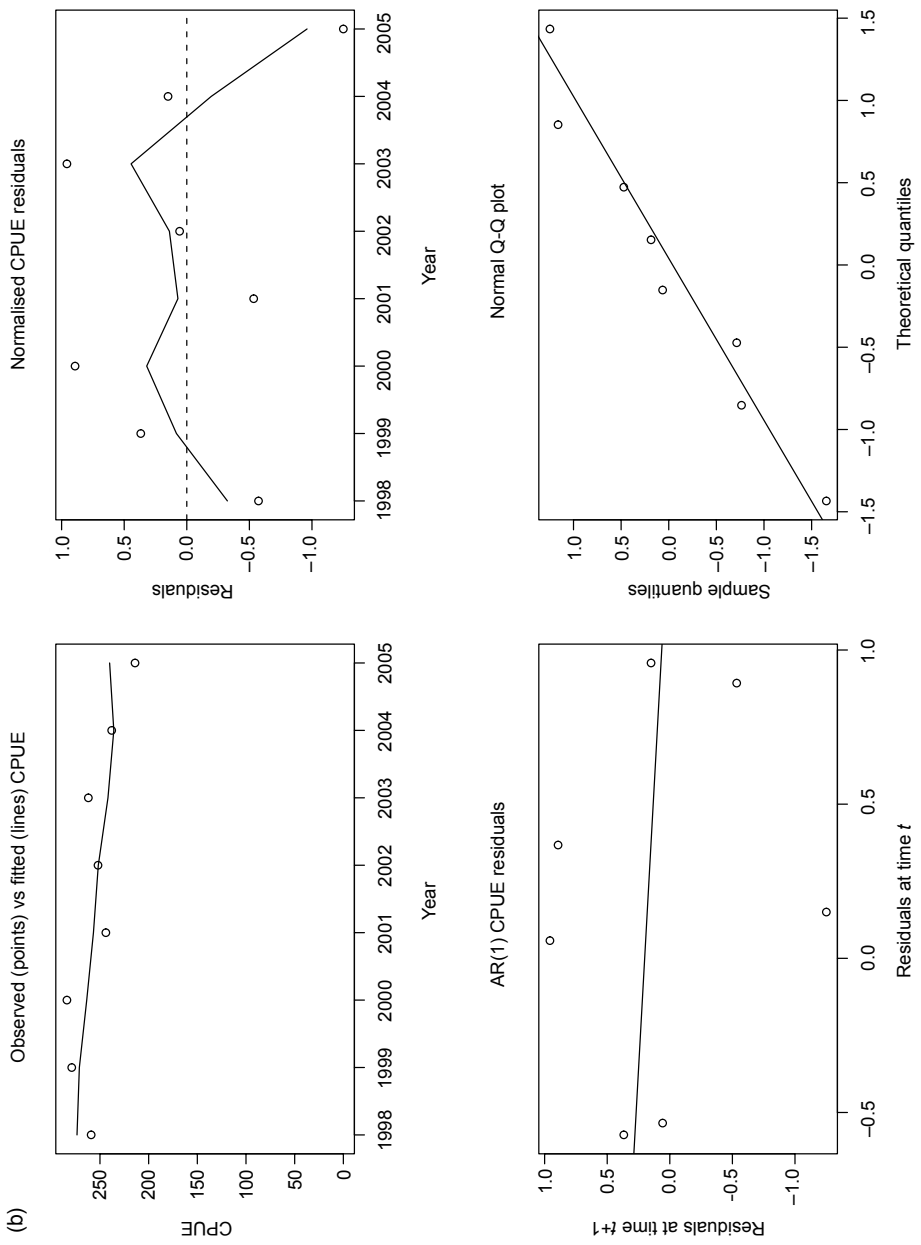
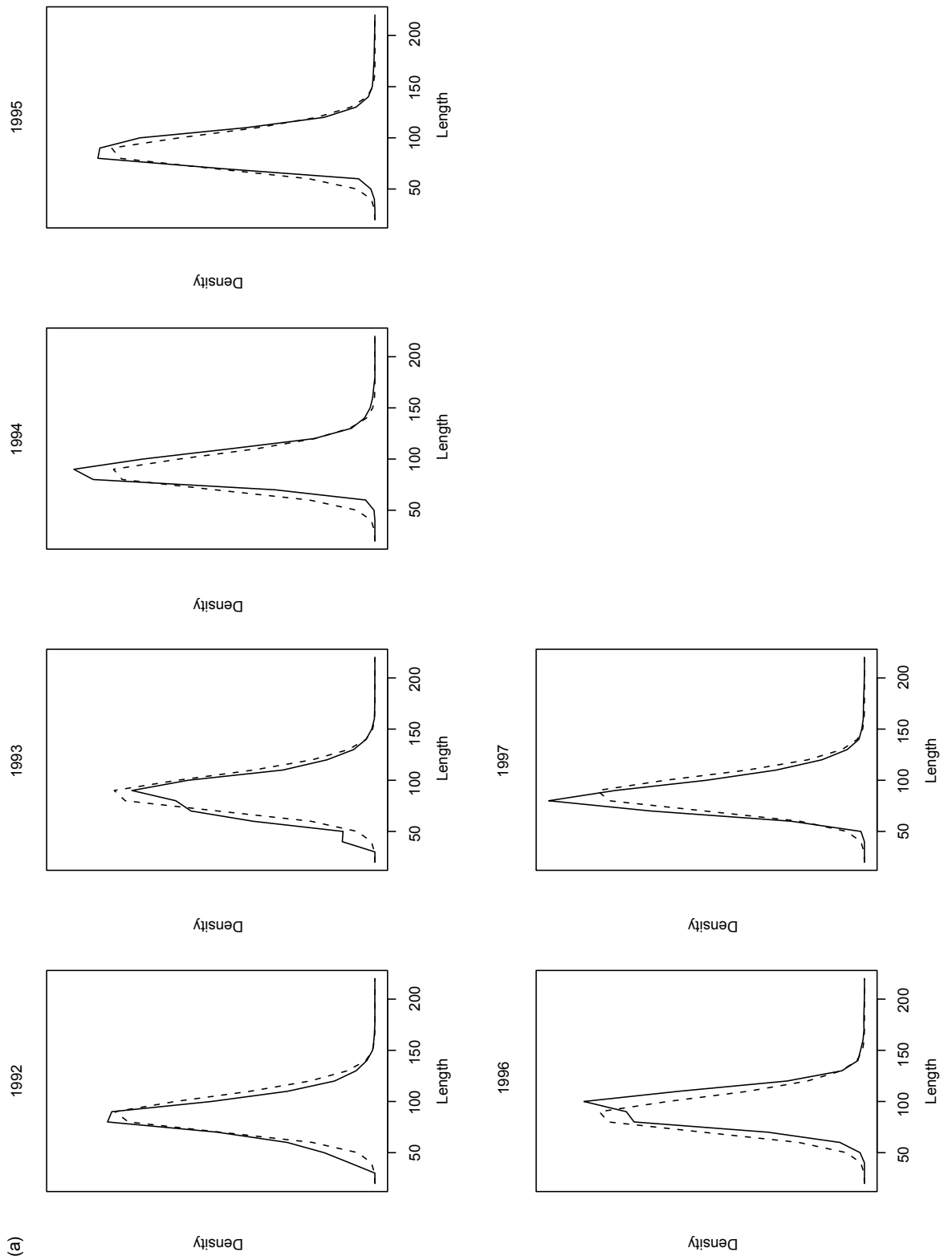


Figure 3: Fits to CPUE data for: (a) early fleet, and (b) later fleet.


(a)
1992
1993
1994
1995
1996
1997
Density
Length
50
100
150
200

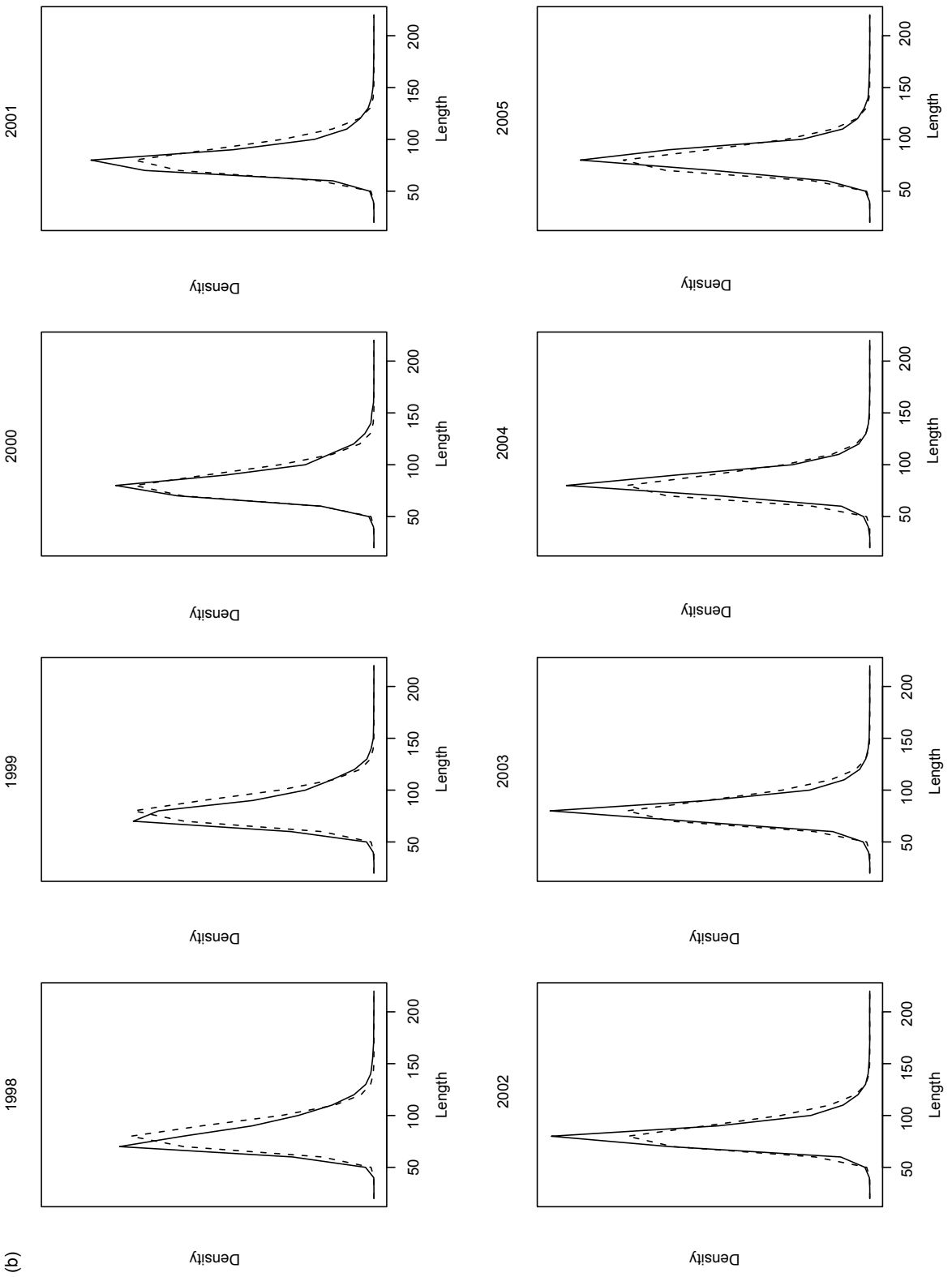


Figure 4: Fits to length-frequency data for: (a) early fleet, and (b) later fleet.

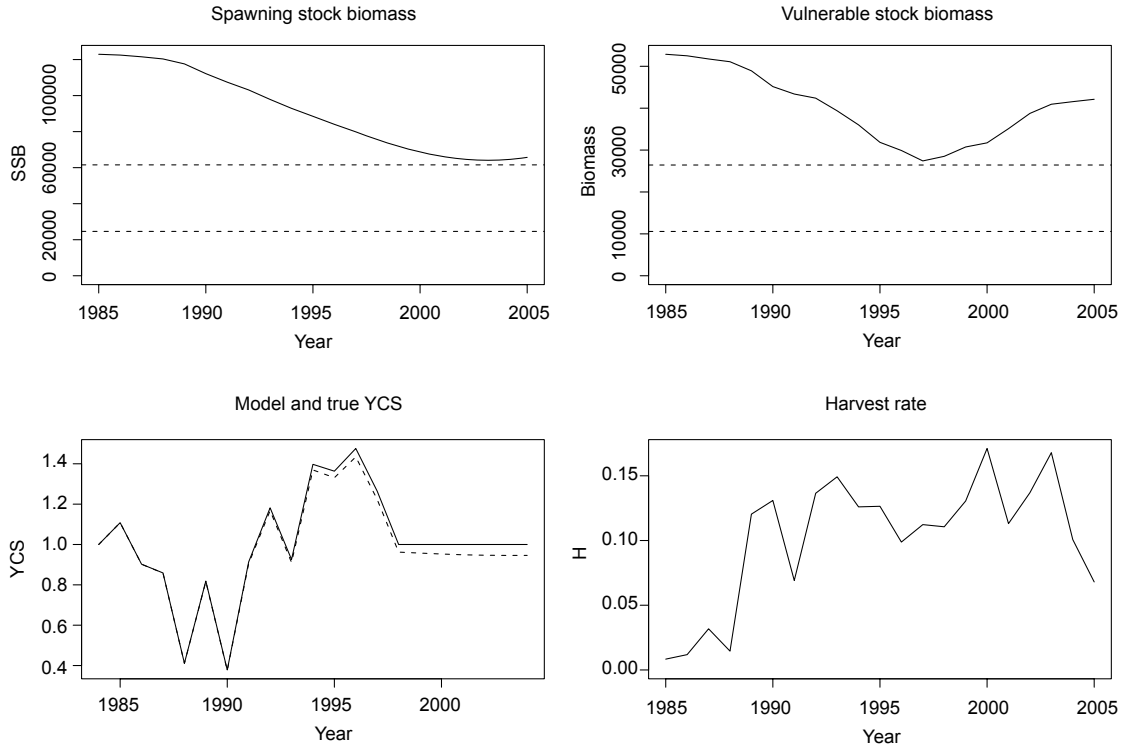


Figure 5: Stock summary plot with the survey data included.

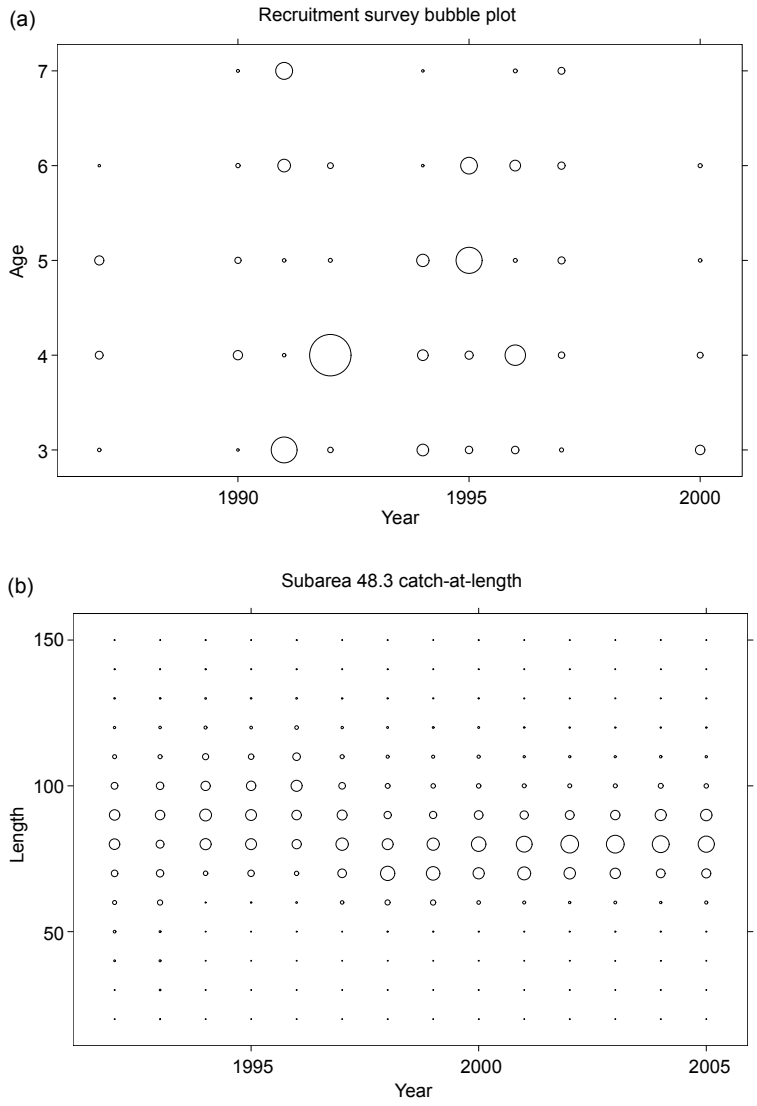


Figure 6: Bubble plots for (a) the survey data, and (b) length-frequency data used.

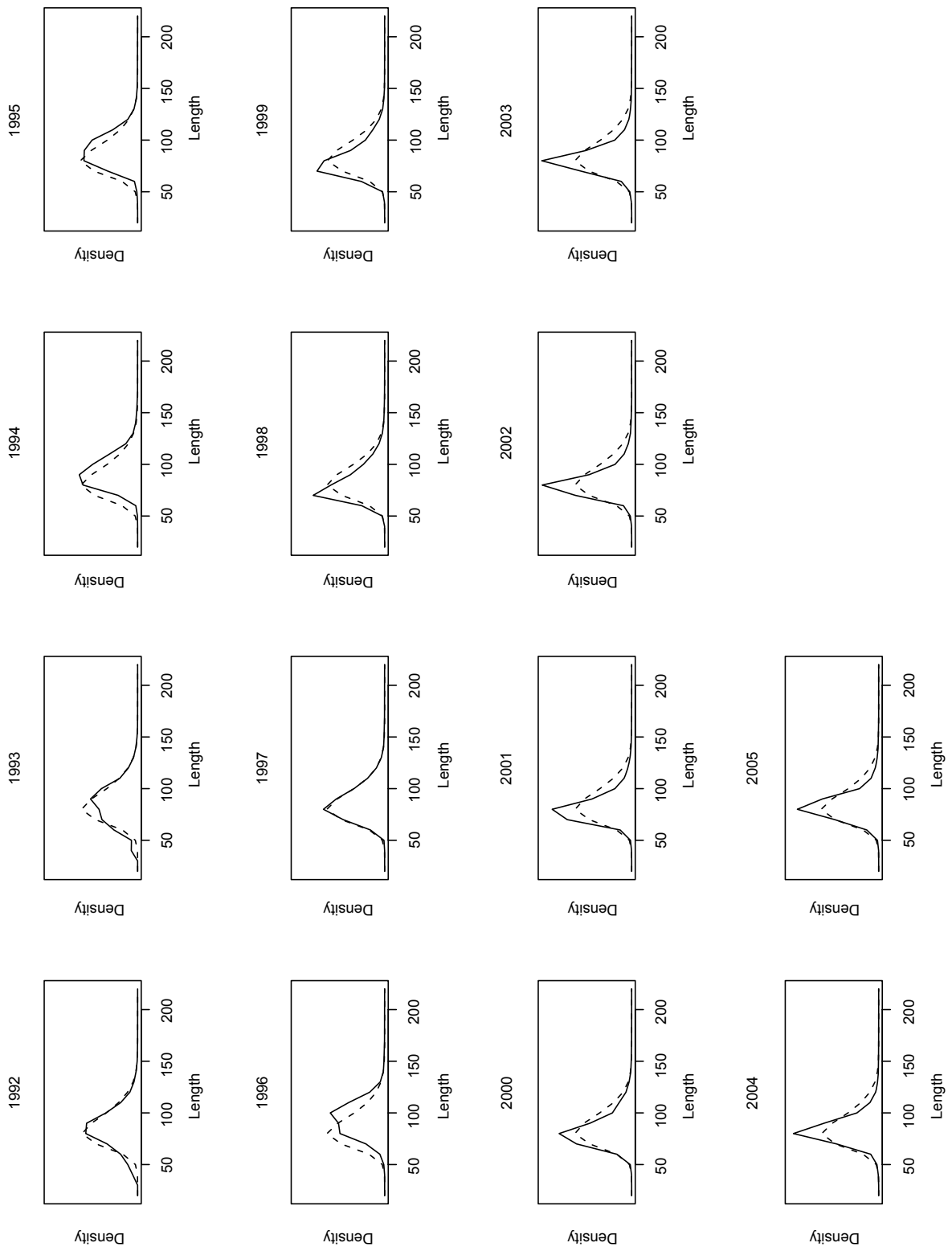


Figure 7: Fits to length-frequency data for the one-fleet assessment.

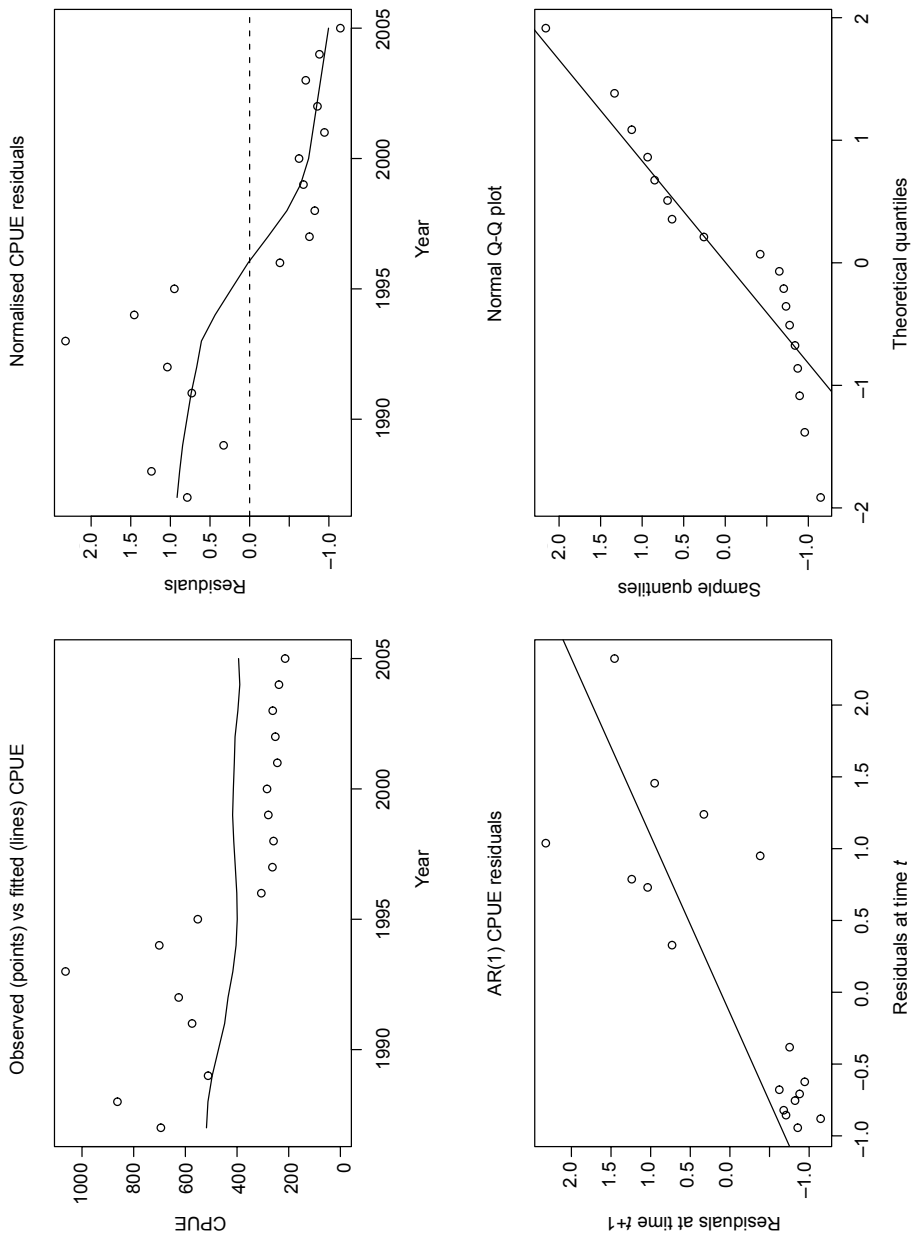


Figure 8: Fits to the CPUE data for the one-fleet assessment.

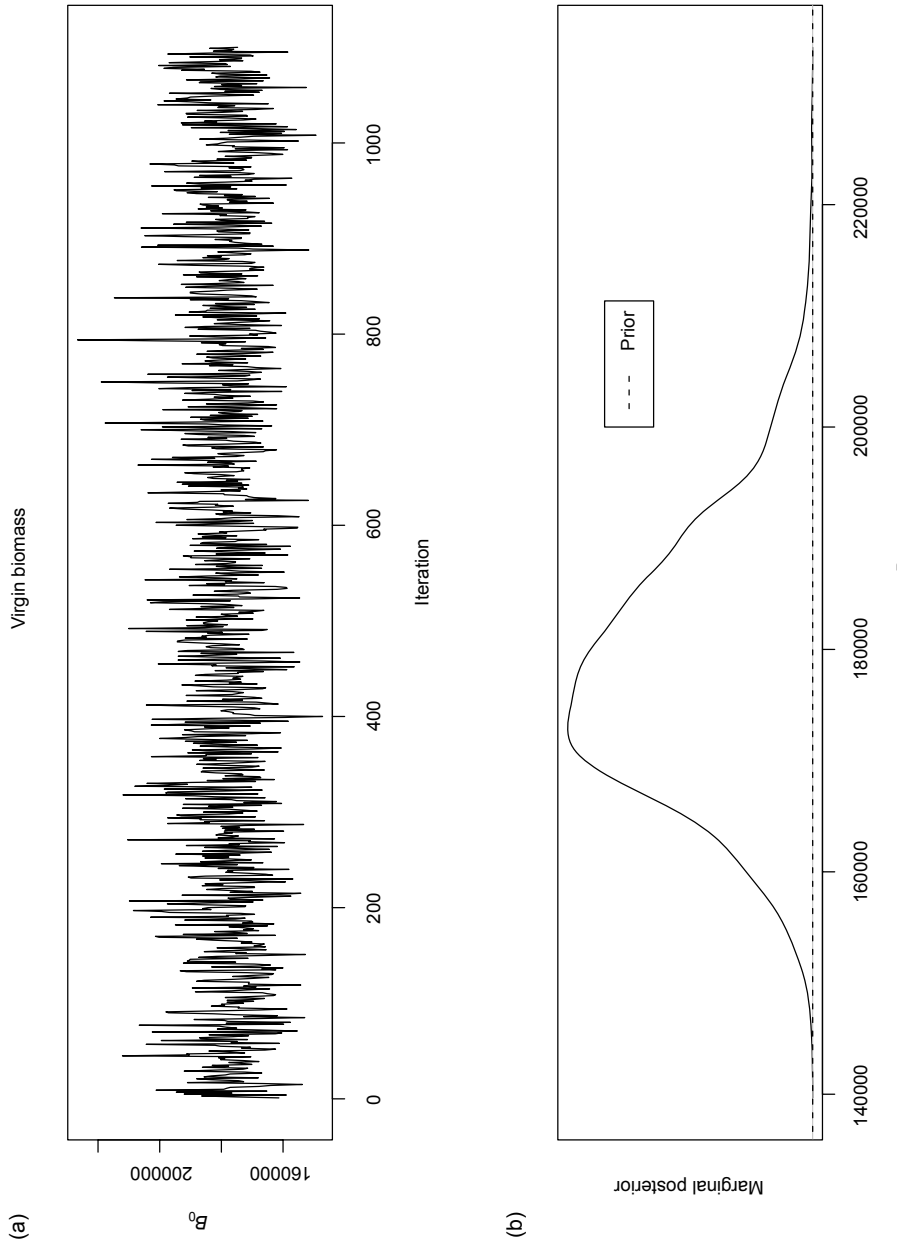


Figure 9: (a) Trace plot, and (b) marginal prior (dotted line) and posterior density (solid line) for the B_0 parameter.

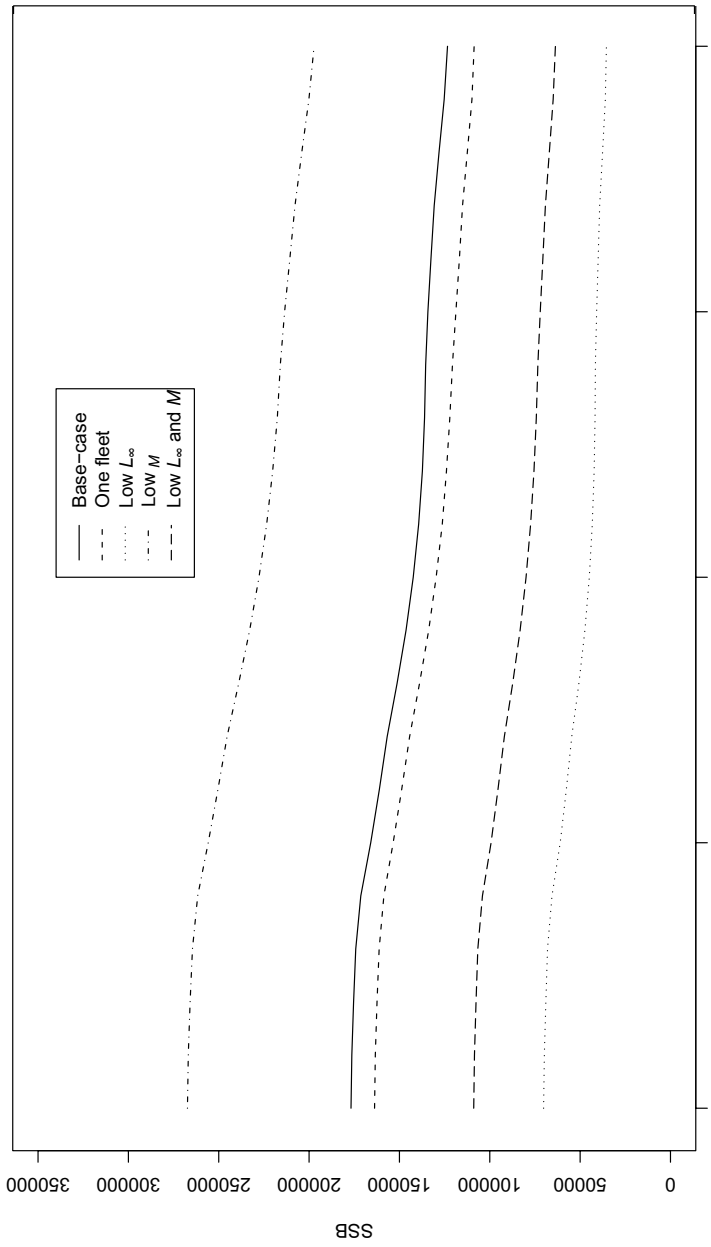


Figure 10: Median historical SSB trajectories for all five scenarios where MCMC runs were performed.

Liste des tableaux

Liste des figures

Список таблиц

Список рисунков

Lista de las tablas

Lista de las figuras

APPENDIX 1

MCMC CONVERGENCE CRITERIA

There are numerous ways of assessing the convergence of Markov chains on the distribution of interest; see Brooks and Roberts (1998) for a thorough review of many of these methods.

The first MCMC convergence test applied here is mentioned in Brooks and Roberts (1998), and uses the information stored in the quantiles of the distribution of interest. Basically, given two or more Markov chains, the difference between any two quantiles calculated for each of the Markov chains should be the same as the difference of these two quantiles for the chain made up of the concatenation of all the Markov chains being tested, if they all represent the same probability distribution. For this study, the 2.5 and 97.5 quantiles were chosen, as these wide quantiles are more likely to identify anomalous 'wandering' behaviour in the Markov chains, which is indicative of non-convergence. For the MCMC run performed for this assessment, all the parameters passed this particular test, with the ratio of the single and concatenated chains' quantile difference being very close to unity.

Most MCMC convergence diagnostics require only the Markov chains themselves to function. However, given that, in addition to the parameter MCMC samples, CASAL uses a complete parameter update Metropolis-Hastings MCMC algorithm, and outputs the value of the log-posterior for each state in the chain, an even simpler method could be used. As outlined in the paper by Bernardo (2003), the discrepancy $\delta(\pi_1, \pi_2)$ between two distributions $\pi_1(\theta)$ and $\pi_2(\theta)$ can be expressed in the following manner:

$$\delta(\pi_1, \pi_2) = \min\left\{\int_\theta \pi_1(\theta)\log\frac{\pi_1(\theta)}{\pi_2(\theta)}d\theta, \int_\theta \pi_2(\theta)\log\frac{\pi_2(\theta)}{\pi_1(\theta)}d\theta\right\} \quad (1.1)$$

This discrepancy measures how far apart the two distributions are – it is in fact the minimum of the posterior averaged log-ratio of the two densities. A standard result is that $\delta(\pi_1, \pi_2) = 0$ if $\pi_1 = \pi_2$, and this property will be used to derive an MCMC convergence indicator when the log-posterior information is available and a Metropolis-Hastings one-step update algorithm is used.

In the MCMC run performed here, there are, say, two chains, θ_i and υ_i, where $i = 1, \ldots, N$ is the number of MCMC iterations. It is naturally assumed that both these chains are drawn from the same posterior distribution, $\pi(.)$. If this is the case, then using both these Markov chains and the log-posterior information to express the discrepancy between the two manifestations of $\pi(.)$ should give an answer very close to zero. This is because we can express equation 1.1 in the following way:

$$\delta(\pi, \pi) = \min\left\{\pm\frac{1}{N}\sum_{i=1}^{N}\log\frac{\pi(\theta_i)}{\pi(\upsilon_i)}\right\} \quad (1.2)$$

and this should tend to zero as $N \to \infty$ and both θ_i and υ_i converge towards the distribution π. Given the log-posterior information coming from CASAL MCMC runs, this convergence checker can be implemented in a short line of R code, and it is very fast. Again, for our MCMC run this convergence criterion was satisfied.

The final MCMC convergence check should be a visual one: the trace plots of the Markov chains are required to be lacking any obvious trends or apparent 'wandering' behaviour, as well as a smooth resolution in either the histograms or density plots. The MCMC run in this study passes all three of these chosen performance indicators; as a result, the authors are satisfied that the Markov chain converged on the posterior distribution.