

Stock assessment of orange roughy in the Walter's Shoal Region

P.L. Cordue
Innovative Solutions Ltd
Consultant

Relates to agenda item: 4

Working paper Info paper

Abstract

A stock assessment is presented for the Walter's Shoal region (WSR) (defined to be the region enclosed by the rectangle 33 50' to 34 41' S, 44 00' to 46 00' E). Biological data were used in conjunction with a stock hypothesis, a catch history, and acoustic biomass estimates to perform a Bayesian stock assessment using NIWA's stock assessment package CASAL.

The absolute scale of the WSR stock is very uncertain because the true scale of the acoustic biomass estimates is poorly known. Virgin biomass (B_0) is very likely to be in the range of 25 000 – 90 000 t. However, it is certain (given the stock hypothesis) that 2017 stock status is above 50% B_0 (the upper bound of the biomass target range for New Zealand orange roughy stocks).

Recommendation

That the Stock Assessment Working Group

1. Accept the methods used to produce the stock assessment.
 2. Accept the base model and sensitivity models as presented.
-

Table of Contents

Executive Summary	2
Introduction	4
Methods	4
Stock hypothesis	4
Biological data	5
Catch history	6
Acoustic biomass estimates	8
Model structure	9
Estimation approach.....	10
Model runs	13
Projections	14
Results	14
Base model MPD fits	14
Deterministic MSY calculations and reference points.....	16
Base model MCMC estimates	17
Sensitivity analysis.....	25
Projections	27
Conclusions	30
Acknowledgements.....	31
References	31
Appendix A: Estimation of length-weight and growth parameters	33
Appendix B: Construction of the catch history	38
Appendix C: Acoustic biomass estimates and target strength.....	39
Appendix D: MCMC chain diagnostics for the base model.....	44
Appendix E: Marginal posterior distributions and MCMC fits and residuals for the base model.....	50

Executive Summary

Innovative Solutions Ltd (ISL) was contracted to provide a stock assessment for orange roughy in the Walter's Shoal region (WSR) (defined to be the region enclosed by the rectangle 33 50' to 34 41' S, 44 00' to 46 00' E). The area contains eleven underwater features/hills which are known to have spawning orange roughy. Sexed length-weight data have been collected from many of the features since 2004. In 2017, fish from one feature (Sleeping Beauty) were sampled for otoliths so that an age frequency could be constructed and growth parameters estimated. These data were used in conjunction with a stock hypothesis, a catch history, and acoustic biomass estimates to perform a Bayesian stock assessment using NIWA's stock assessment package CASAL.

The WSR stock was assumed to have multiple spawning sites within the WSR. All catch taken within the region was assumed to be taken from the stock and catches outside of the region were assumed to be from other stocks. A multi-area model was used so that acoustic biomass estimates from five features could be associated with particular model areas. These features were numbered rather than named to keep catch histories and acoustic biomass estimates confidential. The six un-numbered features that have spawning fish were represented by a single area in the model denoted as "Other".

The catch history is well defined from 2002 onwards but is very uncertain in the years 2000 and 2001 when a large number of vessels were fishing within the SIOFA area. There was little catch in the WSR before 2000. For the purposes of the stock assessment the best educated guess was used to increase the reported catches in 2000 and 2001 for the base model. Sensitivity runs were done at half and double the assumed increase in catches.

The true scale of the acoustic biomass estimates is uncertain as there are three factors that can change the absolute scale markedly. The base model uses a defensible combination of the three factors that delivers "middle level" estimates. There is also a Low model which combines the factors to produce acoustic biomass estimates that are 63% of the base model estimates and a High model that increases the base model acoustic estimates by 65%.

The main parameters in the models are virgin biomass (B_0), the year class strengths (cohort strengths), migration parameters (the proportion of the stock's mature fish that migrate to each of the spawning sites/numbered features) and the acoustic q (the proportionality constant that scales population mature biomass at a numbered feature to the acoustic biomass estimate at the feature). Each spawning site in the model was assumed to be fished during the spawning season with equal selection across ages. Natural mortality (M) was estimated but had a highly informed prior with a mean equal to 0.045 (the value used in New Zealand orange roughy stock assessments).

There was also an informed prior on the acoustic q which assumed that "most" of the spawning biomass had been surveyed at each feature in each year that there was an acoustic

biomass estimate. The prior was essentially the same one used in the New Zealand orange roughy stock assessments except it had a somewhat larger coefficient of variation (CV).

Stock status was considered relative to the biomass target range used in New Zealand for MSC certified orange roughy stocks of 30-50% B_0 . Bayesian estimation was used for all model runs with the median of the marginal posterior distribution used as the point estimate and 95% credibility intervals (CIs) calculated.

The absolute scale of the WSR stock is very uncertain because the true scale of the acoustic biomass estimates is poorly known. Virgin biomass (B_0) is very likely to be in the range of 25 000 – 90 000 t. However, it is certain (given the stock structure assumptions) that 2017 stock status is above 50% B_0 (the upper bound of the biomass target range for New Zealand orange roughy stocks).

Local depletion is only an issue for some un-numbered features that may have been heavily fished in 2000 and 2001. The numbered features are all estimated to be well above 30% in terms of local depletion (current spawning biomass divided by virgin spawning biomass for the given feature).

Five year projections, at the current level and distribution of catches across features, show no sustainability concerns. However, under the more pessimistic model assumptions, Feature 4 could become badly depleted if catches for that feature are increased. The challenge is to find a management mechanism that can allow an increase in stock wide catches while avoiding the severe depletion of any individual feature.

Introduction

Innovative Solutions Ltd (ISL) was contracted to provide a stock assessment for orange roughy in the Walter's Shoal region (WSR) which is part of SIOFA statistical area 2 (Figure 1). The area contains a number of underwater features/hills which are known to have spawning orange roughy. Sexed length-weight data have been collected from many of the features since 2004. In 2017, fish from one feature were sampled for otoliths so that an age frequency could be constructed and growth parameters estimated. These data were used in conjunction with a stock hypothesis, a catch history, and acoustic biomass estimates to perform a Bayesian stock assessment using NIWA's stock assessment package CASAL (Bull et al. 2012).

Methods

Stock hypothesis

The WSR stock was assumed to have multiple spawning sites within the WSR (defined to be the region enclosed by the rectangle 33 50' to 34 41' S, 44 00' to 46 00' E) (Figure 1). All catch taken within the region is assumed to be taken from the stock and catches outside of the region are assumed to be from other stocks. The region contained 11 named features from which spawning fish have been caught.

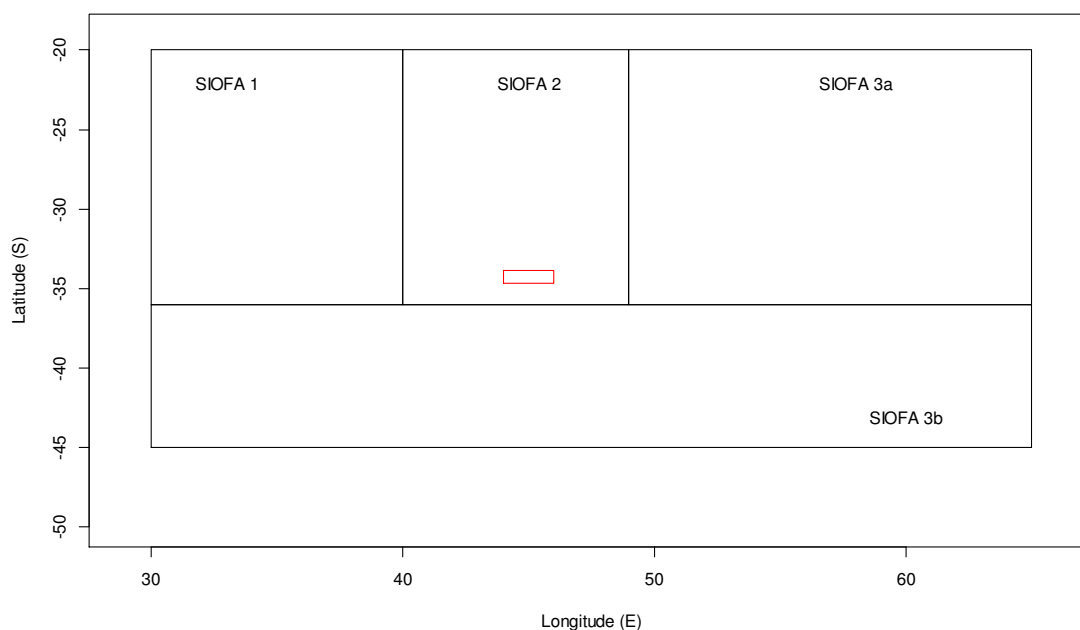


Figure 1: SIOFA statistical regions (1, 2, 3a, and 3b) and the Walter's Shoal region (red box) defined for the purposes of this stock assessment.

Biological data

A substantial quantity of sexed length-weight data were collected from 2004 onwards (see Appendix A). The data collected from features in the WSR all provide similar length-weight relationships (Appendix A) which provides some support for the fish forming an “homogeneous stock”.

The age-length data were collected from Sleeping Beauty in 2017 and were used to estimate von Bertalanffy growth parameters (Appendix A).

The age frequency produced by NIWA (Horn et al. 2017) was not stratified by sex which is better practice. Single trawl catches can often have very skewed sex ratios and it was accepted by New Zealand Working Groups that the sex ratio is so poorly estimated that it is best to assume a 50-50 ratio when constructing length frequencies. The same principle would seem to apply to age frequencies. The raw data were available so the scaled age frequency was recalculated using the same method as Horn et al (2017) except using the sex stratification (and equal weight for males and females) (Figures 2 and 3).

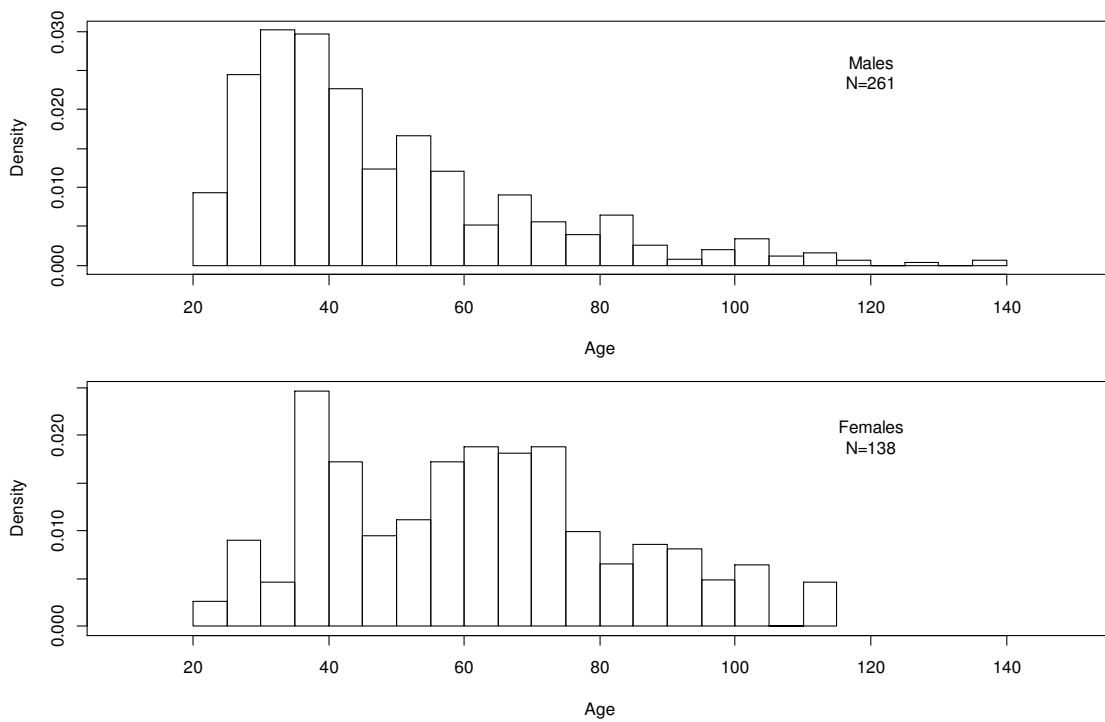


Figure 2: Scaled age frequencies by sex for Sleeping Beauty in 2017. N = number of fish.

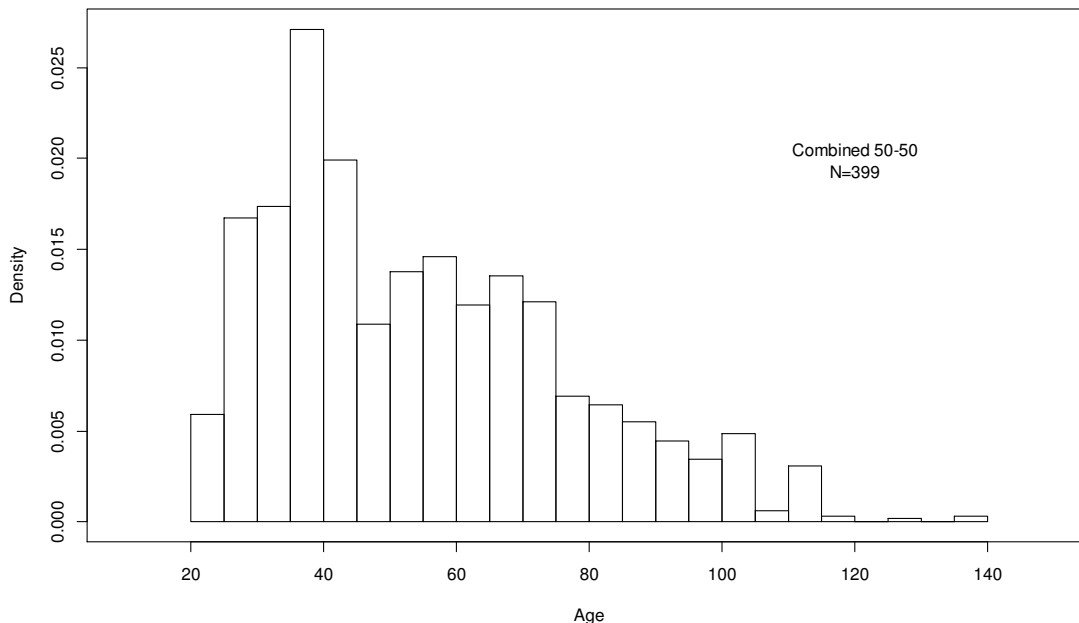


Figure 3: Scaled age frequency for both sexes combined for Sleeping Beauty in 2017. N = total number of fish. Males and females were given equal weight.

The age frequency shows fish up to 140 years old which is consistent with the value of natural mortality (M) used in New Zealand assessments of 0.045 (based on two separate estimates from age frequencies obtained for lightly fished stocks).

Catch history

The catch history is well defined from 2002 onwards but is very uncertain in the years 2000 and 2001 when a large number of vessels were fishing within the SIOFA area. There was little catch before 2000. See Appendix B for the details of the catch history construction.

For the purposes of the stock assessment the best educated guess was used to increase the reported catches in 2000 and 2001 for the base model and sensitivity runs were done at half (Low catch) and double (High catch) the assumed increase in catches (Figure 4). The catches are by calendar year (almost all of the recent catch is during a “spawning” period from June-August included).

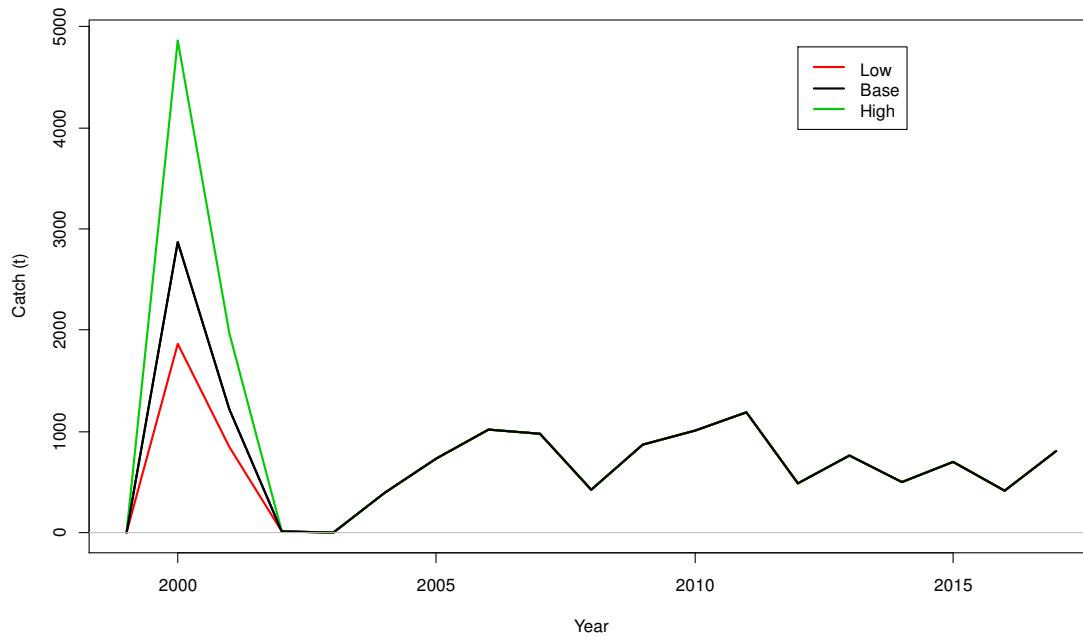


Figure 4: The total catch for the catch histories used in the base model and two sensitivities (Low has half the assumed catches in 2000 and 2001 compared to the base; High has double the assumed catches in 2000 and 2001 compared to the base).

For the stock assessment model the catch history was needed by feature for those features for which acceptable acoustic biomass estimates were available (five numbered features – names are not used to keep biomass estimates and catch histories for individual features confidential). All other features were combined into a single area denoted as “Other”.

In 2000 the largest catches from numbered features were from Features 3 and 5 with the largest catch from Other at about 1250 t (Figure 5). In recent years the largest catches have been from Feature 4 with over 500 t taken in 2017 (Figure 5).

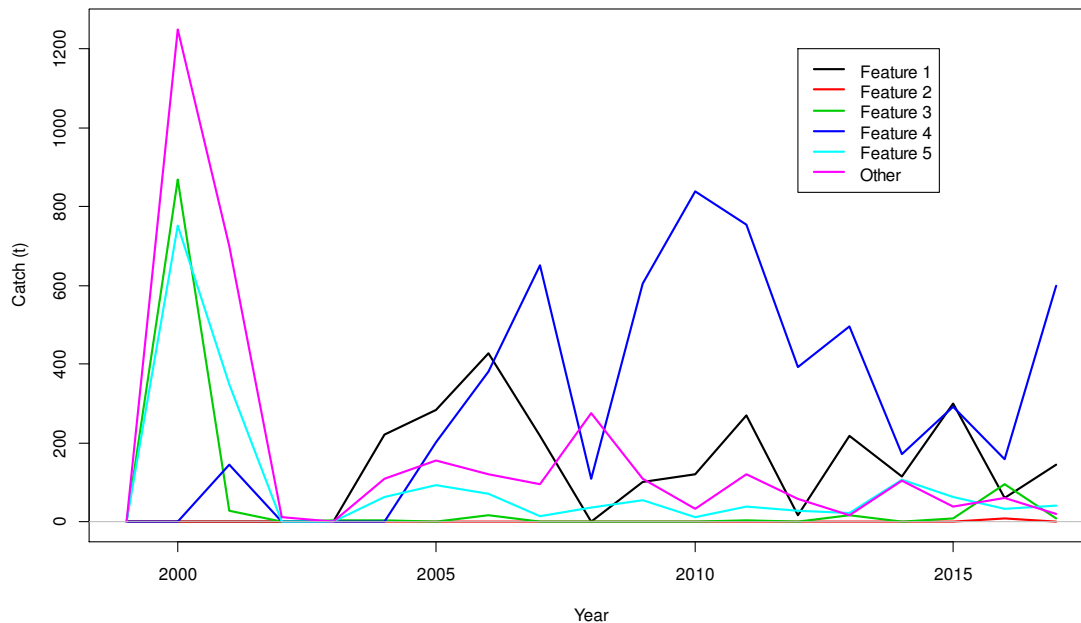


Figure 5: Base model: the catch histories by numbered feature and for Other (which represents all un-numbered features in the defined WSR).

Acoustic biomass estimates

Acoustic biomass estimates were available from some of the features in the WSR in some of the years since 2004 (Appendix C). Some of the biomass estimates were recently revised and these are the estimates used in the base model (Appendix C). In a sensitivity run, a much larger number of estimates was used.

Three different variations of the acoustic estimates were used in stock assessment models: low, middle, and high (Table 1). The different variations used combinations of alternatives: Doonan et al. (2003) or Francois and Garrison (1982) for the absorption coefficient; geostatistical or design based (“EDSU”) estimation; and the McClatchie-Kloser target strength (TS) relationship or the best fit 16.15 revised relationship (see Appendix C – some new TS data were collected). The existing biomass estimates used Doonan, geostatistics, and McClatchie-Kloser.

The “low” estimates were calculated by using Doonan, geostatistics, and the new TS measurement. For this variation the original estimates were reduced to 63% of the original value.

The “middle” estimates were calculated by using Doonan, EDSU, and best fit 16.15. For this variation the original estimates were used because the adjustments for EDSU and best fit 16.15 cancel out ($1.27 \times 0.79 = 1.00$) (see Appendix C).

The “high” estimates were calculated using Francois and Garrison, EDSU, and McClatchie-Kloser. The adjustment required is $1.3 \times 1.27 = 1.65$ (see Appendix C).

Table 1: Revised acoustic biomass estimates for features in WSR from Kloser et al. (draft). See the text for the low, middle, and high treatments. The features are identified by a number only for confidentiality.

Feature	Year	Low estimate (t)	Middle estimate (t)	High estimate (t)	CV (%)
1	2007	1829	2902	4790	11
	2015	2386	3788	6250	32
2	2015	1993	3164	5221	12
3	2015	2381	3779	6235	20
4	2007	4991	7923	13073	10
	2009	6689	10618	17520	30
5	2009	1138	1806	2980	21
	2011	1094	1737	2866	43

The WSR has 11 features from which spawning fish have been taken but only 5 of the features currently have acceptable revised biomass estimates. For two of the other features there are biomass estimates which are probably lower than they should be. There is an estimate for one feature which was pre-spawning rather than at the peak of spawning (895 t) and an estimate from another feature in 2005 when it is suspected that the transducer sensitivity was reduced (Scolding and Kloser 2018) (611 t).

If these two estimates are taken as representative of the other features (the six un-numbered features) then a rough estimate of the proportion of biomass covered by the six un-numbered features can be obtained. The average for the two surveys is 753 t. The total acoustic biomass estimate across the five numbered features (ignoring year and taking an average for each feature with multiple surveys) is 21 330 t. A rough estimate of the proportion covered by the six un-numbered features is $6 \times 753 / (6 \times 753 + 21330) = 17\%$.

For the base model it was assumed that the un-numbered features contained about 20% of the spawning biomass (when the stock was in a virgin state). This was implemented through an informed prior on the migration proportion to Other: Normal with a mean of 0.2 and a CV of 10%.

Model structure

A single-sex, multi-area model was implemented in CASAL (Bull et al. 2012).

A multi-area model was used to allow the acoustic biomass estimates for individual features to be associated with model areas. There are seven areas: Home, Other, and one for each numbered feature. All areas are spawning grounds except for Home which only has immature fish. Fish migrate from Home to one of the spawning areas when they become mature. Once at a particular spawning ground they stay there (this enables localized depletion to occur). The model keeps track of fish by maturity (mature or immature) and age (1-120 years with a plus group).

There is a fishery on each spawning ground which occurs at the end of the year after a full year of natural mortality. All spawning/mature fish are assumed to be equally vulnerable to the fishery independent of age.

The model population is initialised at equilibrium spawning virgin biomass (B_0) with stable age structure (and constant recruitment of R_0). The stock-recruitment relationship is assumed to be Beverton-Holt with a steepness (h) of 0.75 (a default value used in New Zealand orange roughy stock assessments). Recruitment to the population is modified from the stock-recruitment curve by the application of year class strengths (YCS) (an annual multiplier which represents cohort strength). See Bull et al. (2012) for a full set of model equations.

The model starts in 1885 to allow a large number of YCS to be estimated (using information from the age frequency). It runs through to the “current year” in the model of 2017. YCS are estimated from 1887 to 1992 inclusive. All other YCs are fixed at a value of 1 (which means in those years that recruitment is taken off the stock-recruitment curve). Growth and length-weight parameters are fixed in the model (see Appendix A).

The free parameters in the model (those that are estimated) are the YCS, B_0 , M , two maturation parameters (proportion mature at age is assume logistic in the virgin population), five migration parameters (there are six migrations of mature fish but the proportions must add to unity so there are only five free parameters), and the acoustic survey proportionality constant (q).

Estimation approach

The stock assessments were performed using the general Bayesian estimation package CASAL (Bull et al. 2012). The final assessments were based on the marginal posterior distributions of parameters and derived parameters of interest (e.g., virgin biomass (B_0), current biomass (B_{17}), and current stock status (B_{17}/B_0)). The marginal posterior distributions were produced using Markov chain Monte Carlo methods (hence termed “MCMC” runs). Preliminary analysis and some sensitivity runs were performed using just the Mode of the Posterior Distribution (MPD) which can be obtained much more quickly than the full posterior distribution (hence “MPD” runs). The MPD estimate is associated with the “best fit” that can be obtained – it is useful to check that the “best fit” is not too bad otherwise there would be concerns about the appropriateness of the model.

The philosophy behind Bayesian estimation is to update ones beliefs in a rational manner by the application of data. The initial or prior beliefs are represented by the prior distributions that are specified for each model parameter which is to be estimated. Bayes Theorem, of conditional probability, provides a “rational” mechanism for updating the prior beliefs based on the observed data. Application of the theorem produces posterior distributions which represent what one should believe given ones prior beliefs and the observed data. The updated beliefs, with regard to parameters or derived parameters of interest, are summarised

by statistics from the *marginal* posterior distributions. If a single point estimate is required, then a measure of central tendency for the marginal posterior distribution (either the mode, median, or mean) is calculated. Point estimates and credibility intervals (CIs) are produced from the MCMC samples (e.g., a two-sided 95% CI is constructed by excluding the lowest and highest 2.5% of the MCMC samples; the median is simply the median of all of the MCMC samples).

In New Zealand, the favoured point estimate is the median. I believe that this is preferred over the mean because of concerns that the mean may produce overly optimistic estimates of stock status if the marginal posterior distribution is skewed to the right. Of the three choices, the median will generally represent a middle ground between the mode and the mean (they will all be very similar for symmetric marginal posterior distributions and the median will generally be between the mode and the mean for right-tailed distributions).

The MPD point estimates of the parameters of interest may or may not be close to the medians of the marginal posteriors. It is only of academic interest as to how the MPD estimates and the MCMC medians relate to each other. Correct application of Bayesian estimation requires that posterior distributions are calculated. The only advantage of MPD estimation compared to proper Bayesian estimation is its speed. Outside of New Zealand, MPD estimates are often the full extent of a Bayesian estimation procedure. The general claim, in support of this approach, is that the MPD estimates and the MCMC estimates will be very similar. Of course, the veracity of this claim cannot be established unless the MCMC estimates are produced – in which case they should be preferred over the MPD estimates.

The primary source of biomass information in the model are the acoustic surveys of spawning biomass. For each survey, the spawning biomass estimate is included for the relevant feature as an estimate of *relative* spawning biomass rather than *absolute* spawning biomass. The estimates are not used as absolute biomass because there are two major sources of potential bias. The estimates may be biased low or high because the estimate of orange roughy target strength is incorrect. Also, the survey is unlikely to have covered all of the spawning stock biomass for that feature (i.e., the survey availability is unknown). The unknown proportionality constant, or q , for each survey is estimated in the model. To help with this estimation an informed prior is provided for q (which is assumed to be the same for each survey). The prior was based on the one used in New Zealand orange roughy stock assessments which is lognormal with a mean of 0.8 and a CV of 19% (Cordue 2014a).

Acoustic q prior

The two main sources of bias for acoustic survey estimates are fish availability and target strength (TS). (We will assume that any fish movement and double counting has been dealt with as has major contamination from other species.)

For orange roughy in the WSR the TS is very uncertain. Also, the appropriate absorption coefficient is debated by some. And, so is the appropriate analysis method (geostatistics or design based). These three factors are dealt with in the stock assessment by having three

different model runs. This is preferable to having a huge CV on the acoustic q prior for a single run.

The main source of bias for the acoustic q prior within a *given* model run is fish availability. However, even given the TS assumption for the model run there is still some potential TS bias. These two factors were considered by Cordue (2014a) and the result was a lognormal prior with a mean of 0.8 and a CV of 19%. The mean comes from the assumption that the TS is on average unbiased and that “most” of the spawning biomass was present when the survey(s) occurred. It cannot be assumed to be 100% as not all of the fish will be pluming all of the time (even if they are all on the ground at the same time). It will make little difference if a mean of 0.7 or 0.9 is used. Lower proportions are admitting that it was a poor survey (which contradicts the use of the survey estimate in the model). The CV used by Cordue (2014a) was just 19% which accounted for the potential TS bias and deviation from 80% availability. For this assessment a larger CV of 25% was used. A fairly arbitrary increase to allow for potentially larger bias (within the given run).

Natural mortality

Natural mortality (M) was estimated in the base model and most sensitivities with a Normal prior of 0.045 and a CV of 15%. The value of 0.045 has been used in New Zealand orange roughy assessments since 1994 when it was estimated from data collected on the north Chatham Rise in a 1984 trawl survey (Doonan 1994). The 95% CI was 0.03–0.06. There was a further estimate of M obtained from observer sampling in 1996 of the newly developed fishery in the Bay of Plenty (Doonan & Tracey 1997). In this case, M was estimated at 0.037 with a 95% CI of 0.02–0.06.

Year class strength estimation

The year class strengths (YCS) estimated in the model allowed the model to fit the full age range of the 2017 age frequency. Some very early YCS were estimated so that it had a chance of fitting the observation in the plus group.

The Haist parameterisation was used for all models (Bull et al. 2012) with a “nearly uniform” prior: LN(mode = 1, log-space s.d. = 4). This is the same approach taken in the 2014 orange roughy assessments in New Zealand and all subsequent orange roughy assessments (e.g., Cordue 2017). The near-uniform prior on the free year class strength parameters is un-informed compared to a lognormal prior which requires the specification of a variance (recruitment variability). For orange roughy we have very little information on recruitment variability (see Cordue 2017).

Informed maturation priors

From initial model runs it was apparent that the model was struggling to obtain sensible maturation parameter estimates (for both parameters the posteriors had long tails including very high values that were not credible). Therefore, the results of the five New Zealand orange roughy assessments were used to construct informed priors for the two logistic parameters. The median a_{50} estimates ranged from 32 to 41 with a mean of 37 years. The

median a_{1095} estimates ranged from 10 to 15 with a mean of 12 years. As the two parameters are correlated with large a_{1095} estimates only possible for large a_{50} estimates it was thought sufficient to primarily just restrict a_{50} . A Normal prior with mean 37 years and CV of 25% was used. For a_{1095} a weakly informed Normal prior with mean 12 years and CV of 90% was used (restricted to the range of 2.5 to 50 years).

Data weighting

The general approach taken to data weighting within the stock assessment was to down-weight composition data (just the single age frequency in this case) relative to biomass indices to allow any scale and trend information in the biomass indices to drive the assessment results. This is very much in the spirit of Francis (2011) who argued that composition data were generally given far too much weight in stock assessment models and were often allowed to dominate the signals from biomass indices. The age frequency was given an effective sample size of 40 which is the number of fish aged divided by 10. This puts it in line with the effective sample sizes used for age frequencies in New Zealand assessments (often the number of tows rather than the number of fish). The appropriateness of the weighting was checked by looking at the size of the Pearson residuals. Also, a sensitivity run was done where the weight was doubled to 80.

Model runs

The base model has been described above. It uses the middle catch history and the middle treatment of the acoustic data. The effective sample size for the age frequency was 40. There are informed priors on M , the acoustic q , the proportion migrating to Other (mean 20%), and the maturation parameters (means of 37 years and 12 years).

There were eight sensitivity runs performed:

- Low: This has the low treatment of the acoustic biomass estimates with only 10% of mature fish instead of 20% assumed to migrate to Other.
- High: This has the high treatment of the acoustic biomass estimates with 30% of mature fish assumed to migrate to Other.
- Uniform: This has a uniform prior on both maturation parameters.
- AF80: This has double the effective sample size on the age frequency (80 instead of 40).
- Low catch: The amount of catch added on to reported catch for 2000 and 2001 is half that assumed in the base model.
- High catch: The amount of catch added on to reported catch for 2000 and 2001 is double that assumed in the base model.
- Low, low M : The low treatment of the acoustic data and a fixed $M = 0.036$ (20% less than the mean of the prior in the base model).
- More aco: This uses a more extensive set of acoustic biomass estimates (that have not been revised/refined).

Projections

Projections were done for a period of 5 years (to the end of 2022) at the level of the current catches (for each numbered feature and Other). This was done for the Base model and the Low model. The projections were stochastic with new YCS sampled at random from all estimated YCS.

Results

Base model MPD fits

It is useful to check the MPD fits to see if they are sensible. If they are not then there must be some concerns about the model and changes to the model may be required.

The MPD fits to the acoustic biomass estimates look sensible (Figure 6). The fits are almost exact for the features where there is only a single estimate (this is to be expected) and also for Feature 5 which has two estimates that are very similar (Figure 6). For Features 1 and 4, which have two biomass estimates each, the fitted points decline a little while the estimates increase. However, the predicted values are within the 95% confidence intervals for all of the estimates (Figure 6).

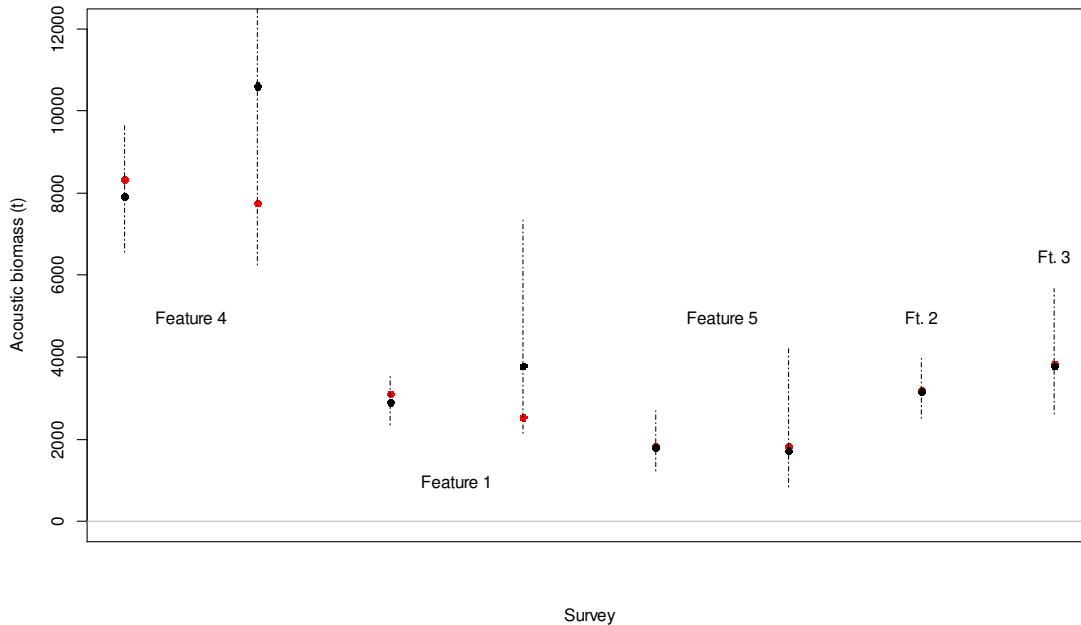


Figure 6: Base model MPD fits to the acoustic biomass estimates. The observations and the 95% confidence intervals are given in black and the predicted values are in red.

The MPD fit to the age frequency is adequate (Figure 7). The observations are very “spikey” and will never be fitted closely (particularly given the ageing error assumed in the model predictions). The fit of the model at the plus group of 120 years stands out but this is not a large residual – it is exaggerated in appearance because it is the accumulation of many age classes. Had the model been extended to include fish up to 150 years the fit would look far better but not actually be any better (and the model would run much more slowly).

The MPD estimated YCS show a smooth pattern with alternating periods of below average and above average cohort strengths (Figure 8). Not to be taken too seriously – we wait for the MCMC estimates.

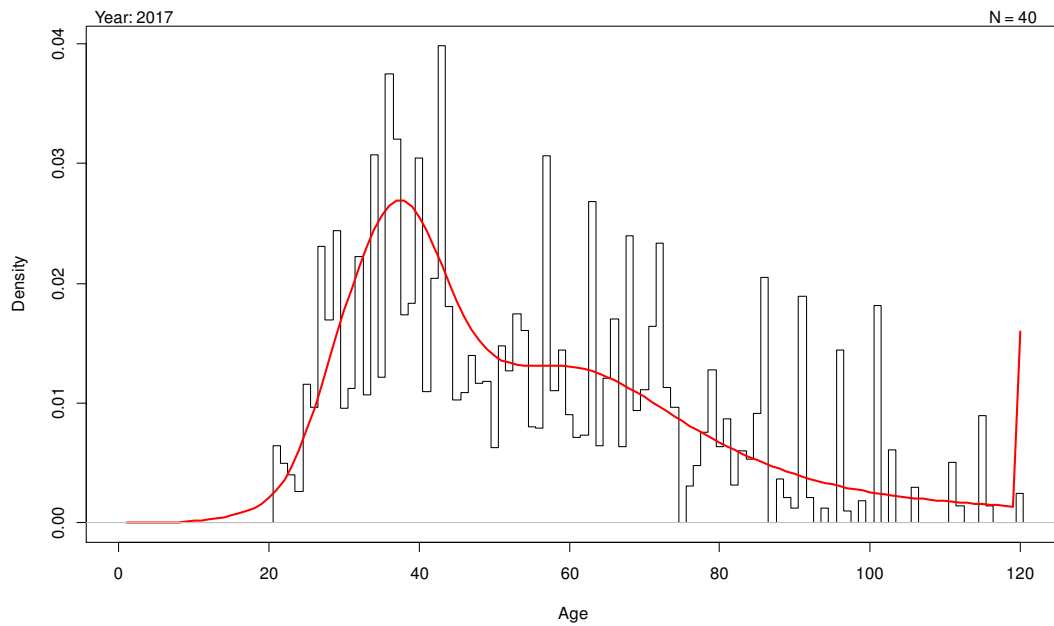


Figure 7: Base model MPD fit (red) to the 2017 age frequency from Sleeping Beauty (black histogram). The effective sample size is 40 which is the number of fish aged divided by 10.

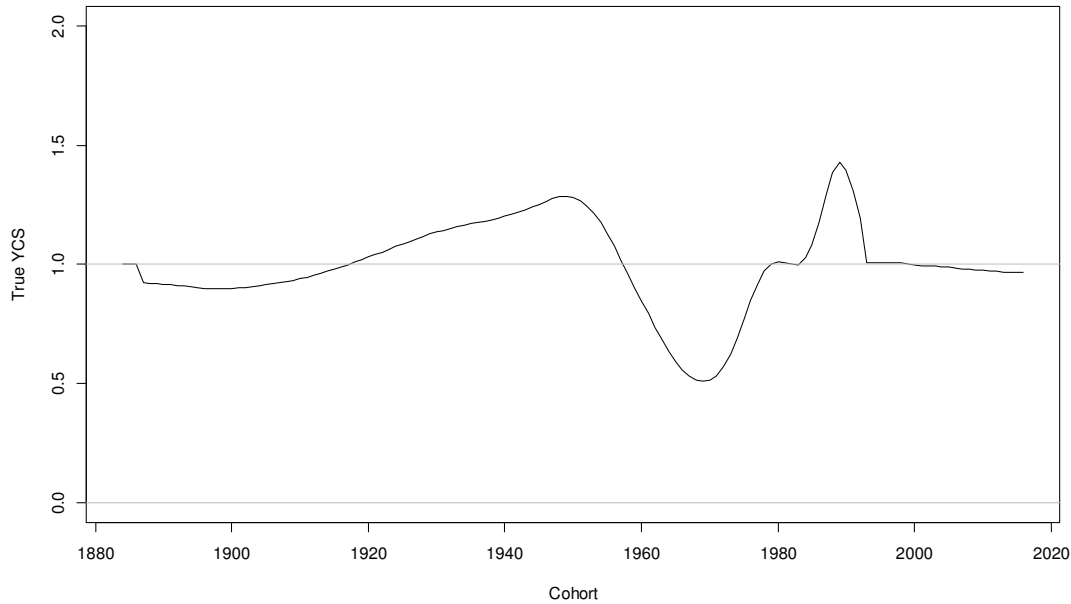


Figure 8: Base model MPD estimates of the true YCS (R_i/R_0).

Deterministic B_{MSY} calculations and reference points

Deterministic B_{MSY} has **not** been found to be a useful reference point for New Zealand orange roughy stocks. It is highly dependent on the stock recruitment relationship and is therefore very uncertain (see Cordue 2014b). Little is known about the stock recruitment relationship for any of the New Zealand stocks and even less is known for the South Indian Ocean stocks.

If a Beverton Holt stock recruitment relationship is assumed for the WSR stock then B_{MSY} is highly dependent on the steepness parameter (Table 2). It is not sensitive to the range of M values that can be expected for the stock (Table 2). It is also insensitive to the maturity parameters (Table 3).

Table 2: B_{MSY} (% B_0) calculated for combinations of M (natural mortality) and h (steepness in a Beverton Holt stock recruitment relationship). In the base model $M = 0.045$ and $h = 0.75$ (the default value used in New Zealand). For these calculations the maturity parameters were assumed to be the mean median values for the five assessed New Zealand stocks ($a_{50} = 37$ years, $a_{1095} = 12$ years). Sleeping Beauty growth and length-weight parameters were used.

M	h			
	0.65	0.75	0.90	0.95
0.036	28	23	16	11
0.045	28	24	15	11
0.054	28	23	15	11

Table 3: B_{MSY} , MSY , and U_{MSY} calculated for three alternative pairs of maturity parameters. Base model parameters were used for other parameters in each case.

Maturity (a_{50}, a_{1095})	B_{MSY} (% B_0)	MSY (% B_0)	U_{MSY}
30 years, 10 years	23.9	2.14	0.086
37 years, 12 years	23.6	2.25	0.091
45 years, 20 years	23.3	2.27	0.093

Given that B_{MSY} is very uncertain and could be close to or even less than $20\%B_0$ a point estimate of B_{MSY} is not suitable as a target or limit reference point (and U_{MSY} is not suitable as a target or limit exploitation rate). For the purposes of this assessment the reference points used in New Zealand for orange roughy are adopted: a limit reference point of $20\% B_0$ and a target biomass range of $30\text{--}50\% B_0$ (see Cordue 2014b). The associated target exploitation rate range is $U_{30\%B_0}\text{--}U_{50\%B_0}$ (where $U_{x\%B_0}$ is the exploitation rate that delivers a deterministic equilibrium spawning biomass of $x\% B_0$).

Base model MCMC estimates

Chain diagnostics

The MCMC chains ran relatively slowly because the model was run over many years, has many age classes, and has six migrations each year. Five chains of 2.5 million were run for the base model (and each sensitivity) with 1 sample in every 1000 retained and the first 500 retained samples discarded as a “burn-in”. Convergence of the chains was checked and appeared to be adequate when all five chains were combined for each run (see Appendix D for the base model chain diagnostics).

Marginal posterior distributions and MCMC diagnostics

A full set of marginal posterior distributions and MCMC fits to the observations and the MCMC residuals are given for the base model in Appendix E. In summary, the posterior distributions for parameters with informed priors appeared to be updated appropriately (none of the posterior distributions were in the tails of the informed priors).

For the acoustic q there was some updating with the posterior shifted to the left of the informed prior (Figure 9). There was a small amount of updating of for M with a slight shift to the left (Figure 10). There was no updating of the prior for the migration proportion to Other (which is as it should be) and the posteriors for the maturity parameters were not too different from the means of the priors but had much lower variances (see Appendix E).

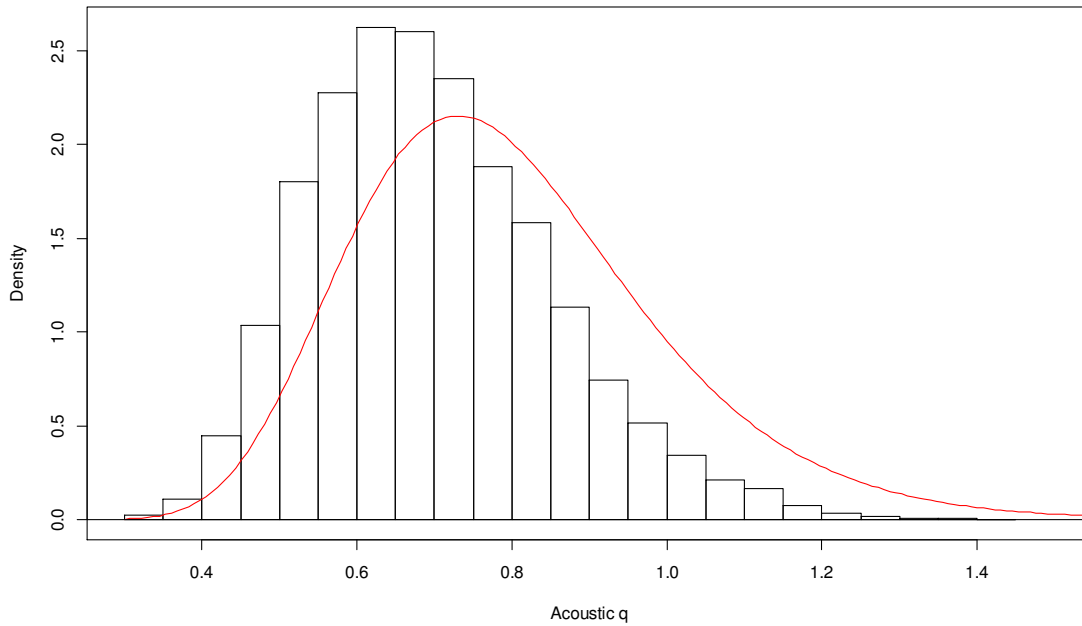


Figure 9: Base model: marginal posterior distribution of the acoustic q (histogram) with the informed prior (red line).

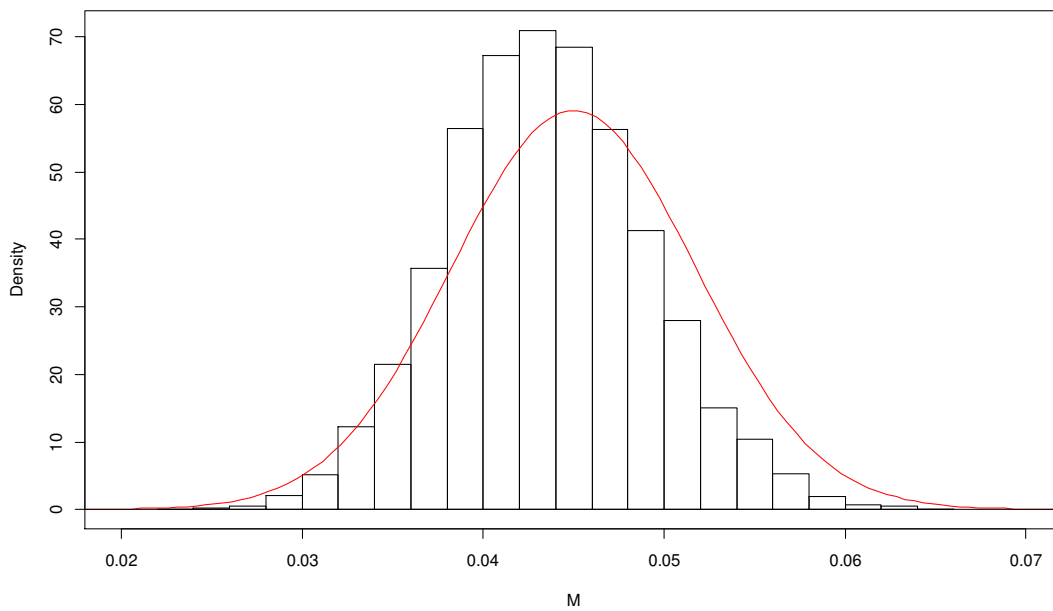


Figure 10: Base model: marginal posterior distribution of M (histogram) with the informed prior (red line).

The migration parameters had uniform priors and were estimated fairly precisely within the model (Figure 11).

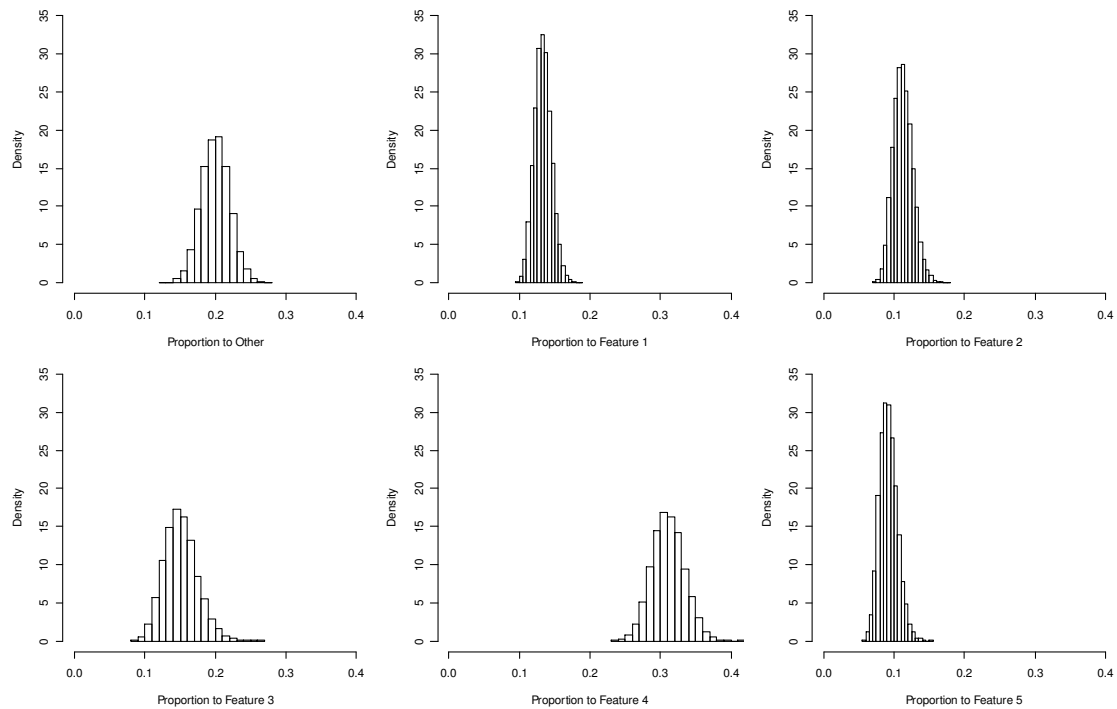


Figure 11: Base model: marginal posterior distributions of the proportion migrating to Other and each numbered feature. The same scale is maintained on both axes across all plots.

The estimated year class strengths (YCS) show almost no periods of strong or weak recruitment (Figure 12). Given the assumed effective sample size of 40, the age frequency is consistent with deterministic recruitment (as well as many other recruitment patterns that are not too dissimilar).

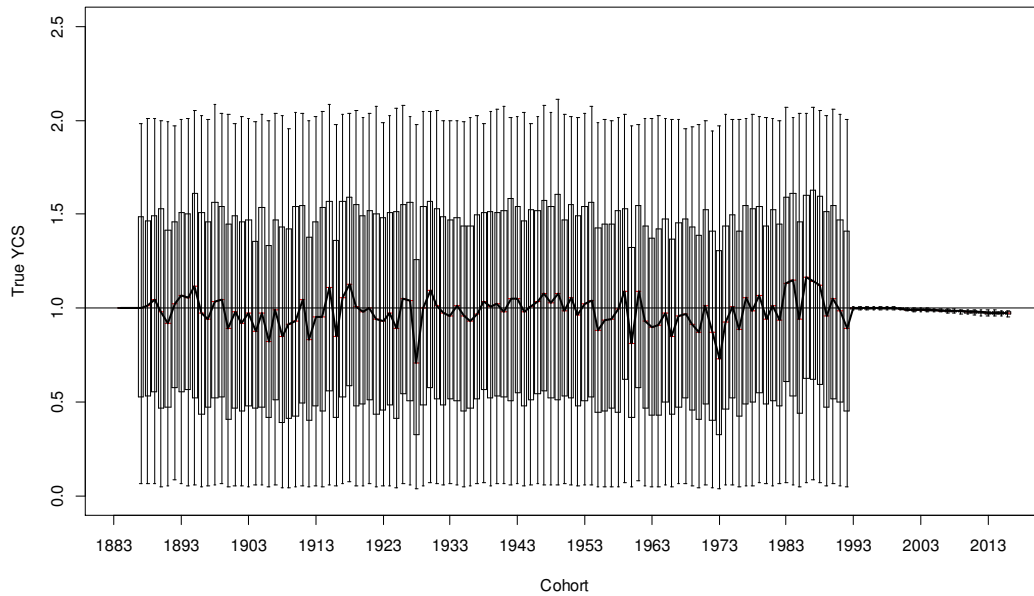


Figure 12: Base model: box and whiskers plot of the true YCS estimates (R_i/R_0). Each box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.

When the effective sampled size for the age frequency is doubled to 80 then the estimated YCS show more of a pattern (Figure 13). However, there is still little deviation in the median estimates from deterministic recruitment (all cohorts strengths equal to 1).

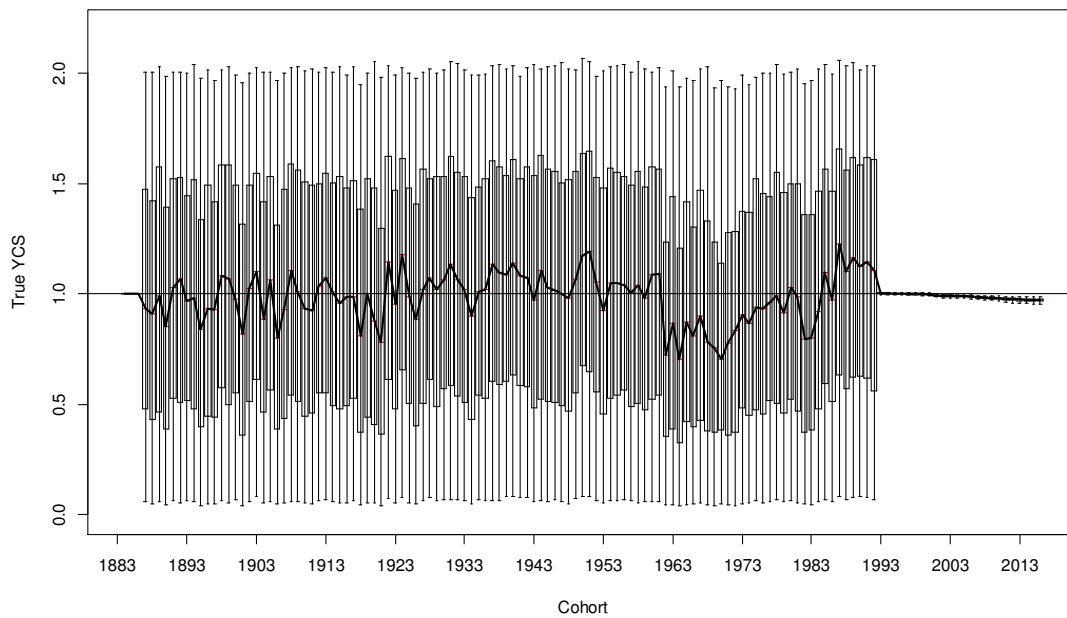


Figure 13: AF80 model (effective sample size for the age frequency doubled compared to the base model): box and whiskers plot of the true YCS estimates (R_i/R_0). Each box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.

The MCMC “fit” to the acoustic biomass estimates shows no problems. The key point is that there is overlap between the confidence intervals of the observations and the credibility intervals of the MCMC predicted values (see Appendix E). The normalised residuals for the acoustic data are primarily within 2 standard deviations of zero which is consistent with the CVs of the acoustic biomass estimates and the model assumptions (see Appendix E). The fit to the age frequency is also adequate as are the Pearson residuals (Appendix E).

Base model MCMC estimates

Virgin biomass was estimated at 43 000 t (95% CI: 29 000 t to 64 000 t)(Table 4). The acoustic q was estimated lower than the mean of the prior at 0.68 compared to 0.8 (Table 4). This is a similar result to that of the five New Zealand orange roughy stock assessments (see Cordue 2014a, Cordue 2017). M was barely unchanged from the mean of the prior as was a_{50} ; a_{1095} was slightly increased (from 12 to 14 years) (Table 4). The maturity ogive (logistic) suggests that fish may start to mature as young as 20 years and that 100% maturity may not occur until 60 years of age (Figure 14). This result is consistent with New Zealand orange roughy stocks (from which the prior was formed) and the observed age frequency (which was used in the model).

The point estimate of the migration proportion to Other was (of course) equal to the mean of the prior at 20% (Table 4). The other migration proportions, with the exception of Feature 4, were about 10% (Table 4). Feature 4, which has the largest acoustic biomass estimates, has a migration proportion estimated at about 30% (Table 4).

Table 4: Base model: MCMC estimates of the parameters (excluding YCS parameters). The median and 95% CI is given for each parameter. The second row are the migration proportions for mature fish from Home to the given ground expressed as percentages.

B_0 (000 t)	Acoustic q		M	a_{50} (years)		a_{1095} (years)		
43 29-64	0.68	0.44-1.05	0.043	0.033-0.055	37	29-47	14	5-25
Other	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5			
20 16-24	13 11-16	11 9-14	15 11-20	31 27-36	9 7-12			

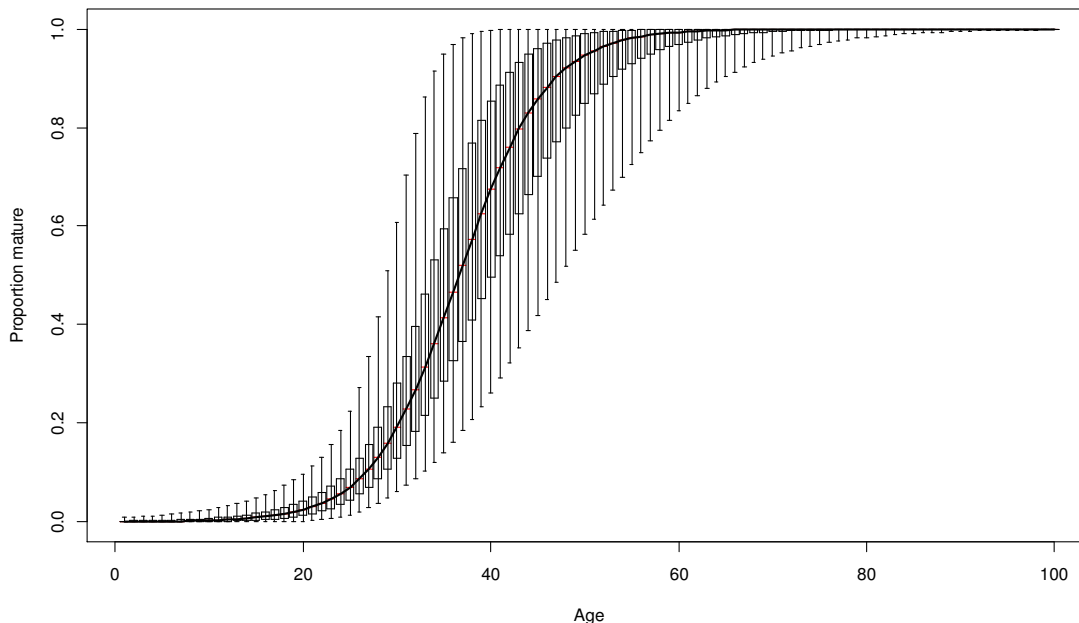


Figure 14: Base model MCMC: box and whiskers plot of the proportion mature at age in the virgin population. Each box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.

The current stock status is very healthy according to the model results with certainty (within the model) that current spawning biomass is greater than 50% B_0 (the upper bound of the target biomass range for the New Zealand HCR – see Cordue 2014b) (Table 5, Figure 15). The relative depletion levels at each numbered feature and collectively for the un-numbered features (Other) are also very healthy in the base model (Table 5, Figure 16).

Table 5: Base model: MCMC estimates of virgin spawning biomass (B_0), current spawning biomass (B_{17}), current stock status (B_{17}/B_0), and the probability of current biomass being above 30% B_0 or 50% B_0 . The median and 95% CI are given for each parameter. The second row gives the local depletion estimates for spawning biomass at each numbered feature and collectively the un-numbered features (Other). Local depletion is the current biomass at the feature divided by the virgin biomass at the feature (expressed as a percentage).

B_0 (000 t)	B_{17} (000 t)	SS17 (% B_0)	$P(B_{17} > 30\% B_0)$ (%)	$P(B_{17} > 50\% B_0)$ (%)	
43 29-64	32 19-53	76 63-87	100	100	
Other	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
75 60-87	66 51-79	99 90-107	89 80-98	66 49-80	71 57-83

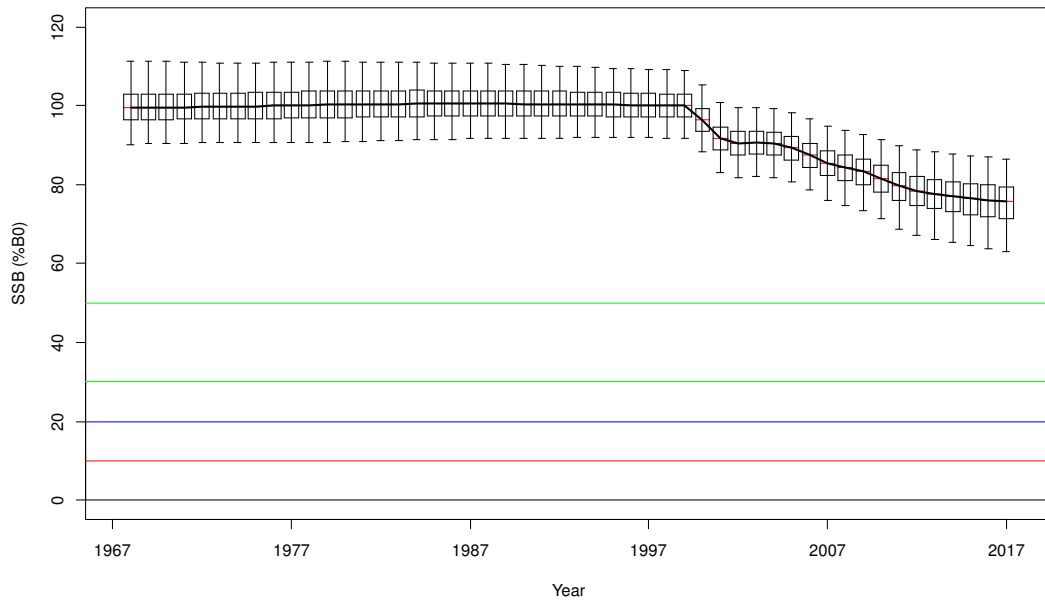


Figure 15: Base model MCMC: box and whiskers plot of the spawning biomass trajectory for the whole stock ($% B_0$). Each box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.

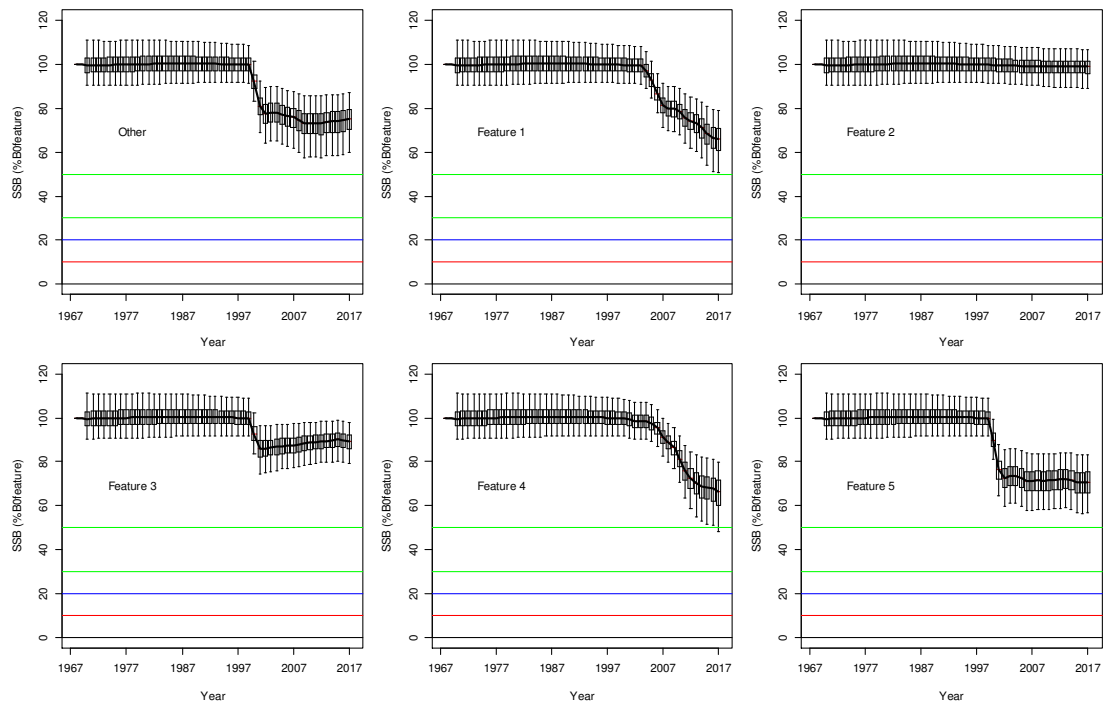


Figure 16: Base model MCMC: box and whiskers plot of the spawning biomass trajectory for Other and each numbered feature ($%$ of virgin biomass at the feature). Each box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.

Given the healthy stock status, not surprisingly the estimated exploitation rates for the stock are low except in 2000 when many vessels were fishing (Figure 17).

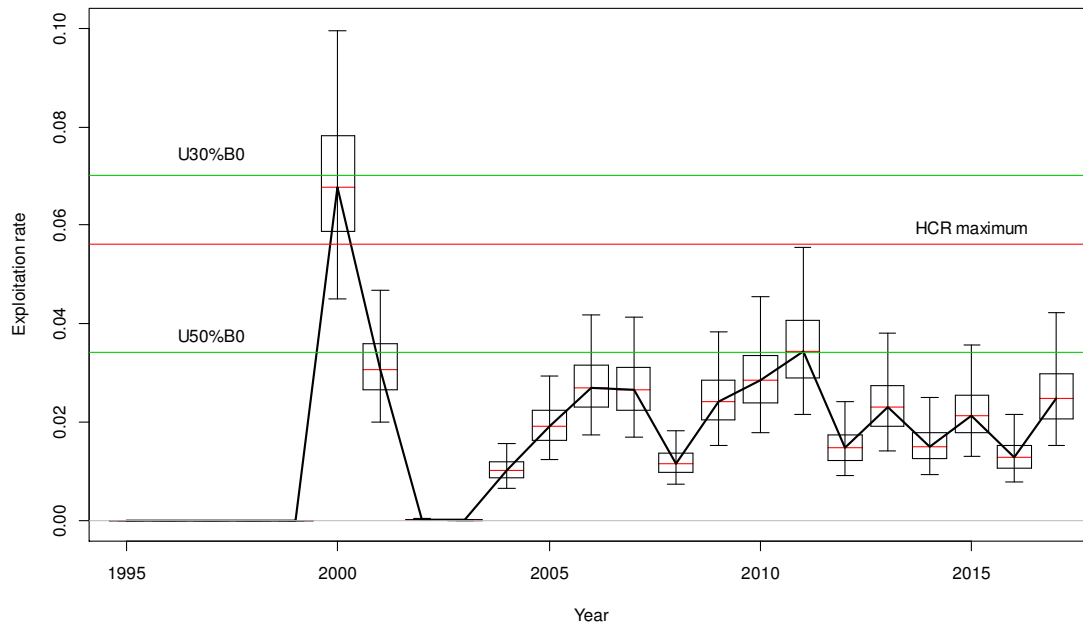


Figure 17: Base model MCMC: box and whiskers plot of the exploitation rate trajectory for the whole stock. Each box covers the middle 50% of the distribution and the whiskers extend to a 95% CI. The New Zealand HCR has a target biomass range of 30–50% B_0 . The associated target exploitation rate range of $U_{30\%B_0}$ – $U_{50\%B_0}$ is plotted (where $U_{x\%B_0}$ is the exploitation rate that delivers a deterministic equilibrium spawning biomass of $x\%$ B_0). However, the maximum exploitation rate allowed in the HCR is less than $U_{30\%B_0}$ and is marked by a red line.

The “snail trail” (a plot of the annual exploitation rate on the y axis and the annual stock status on the x axis) shows that the stock has never approached any region of the two-dimensional plane that would cause any concern with regard to sustainability of the stock (Figure 18). Exploitation rates are low and biomass levels are high.

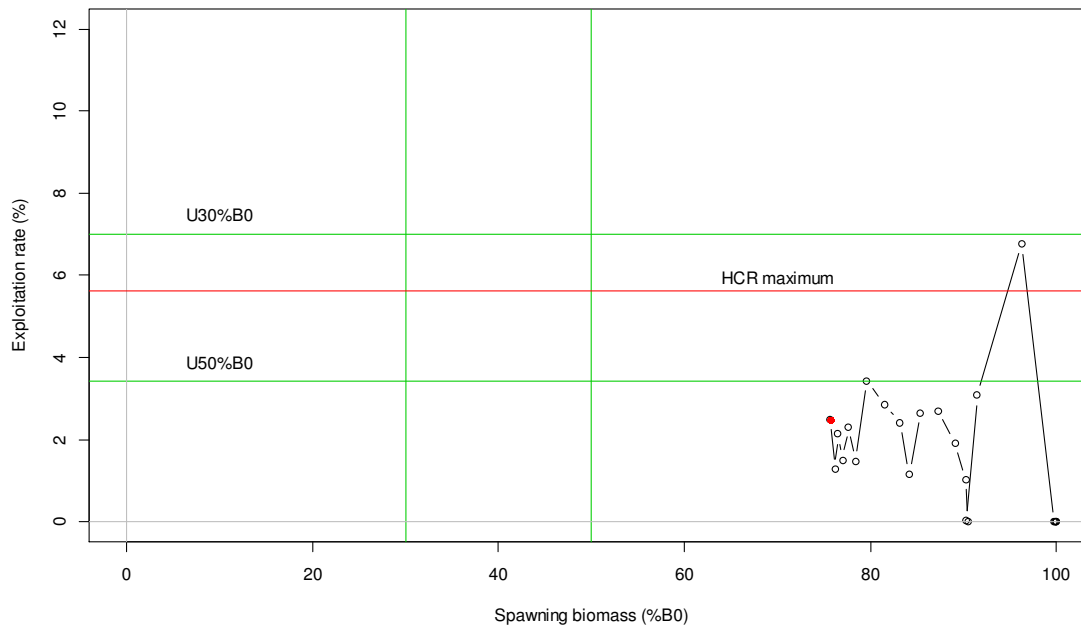


Figure 18: Base model MCMC: snail trail showing stock status and exploitation rate trajectories (medians) (current stock status and exploitation rate marked as the solid red point). The New Zealand HCR has a target biomass range of 30–50% B_0 . The associated target exploitation rate range of $U_{30\%B_0}$ – $U_{50\%B_0}$ is plotted (where $U_{x\%B_0}$ is the exploitation rate that delivers a deterministic equilibrium spawning biomass of $x\% B_0$). However, the maximum exploitation rate allowed in the HCR is less than $U_{30\%B_0}$ and is marked by a red line.

Sensitivity analysis

In addition to the base model eight MCMC sensitivity runs were also performed (see “Model runs” in the Methods section above).

A retrospective analysis was not performed because this is the first assessment that is really possible under the model structure that is used. If we step back one year to 2016 then the 2017 age frequency is not available and it would be necessary to assume a maturity pattern and deterministic YCS. This could be done, but then when we step back to 2014 we lose biomass estimates from two numbered features that were only surveyed in 2015. The structure would need to be changed so that there were only three numbered features. Whatever was assumed about the migration proportion to Other would determine whether the assessment was consistent with the current assessment or not.

The estimates of M are almost identical across all model runs (Table 6). This is primarily a reflection that there is little information in the data with which to estimate M . In the base model and other runs the posterior is slightly to the left of the prior but is little changed (see Figure 10 for the base model). The maturity parameters are not very well determined when uniform priors are used (see Uniform in Table 6). The righthand tails of the posterior “blow out” to unrealistically high values (this is the only effect – all other estimates are almost

identical to the base model). The use of the informed priors on all other runs keeps the posteriors in sensible space and there is little difference in the estimates of the maturation parameters across the runs (Table 6). The estimates of the acoustic q change in response to the treatment of the acoustic data as expected – being lower for the Low run and higher for the High run (Table 6). The acoustic q is highest when the more extensive set of original acoustic estimates is used but it also very uncertain (Table 6).

Table 6: MCMC estimates: medians and 95% CIs for the acoustic q , M , a_{50} , and a_{1095} for the base model and sensitivities (see the text for the description of each run).

	<u>Acoustic q</u>		<u>M</u>		<u>a_{50} (years)</u>		<u>A_{1095} (years)</u>	
	Base	0.68	0.44-1.05	0.043	0.033-0.055	37	29-47	14
Low	0.59	0.40-0.80	0.043	0.033-0.054	37	29-46	13	5-19
High	0.73	0.52-1.14	0.043	0.033-0.055	36	28-46	13	5-25
Uniform	0.69	0.45-1.06	0.044	0.033-0.056	38	28-56	16	5-36
AF80	0.68	0.43-1.03	0.043	0.034-0.053	35	29-45	12	5-21
Low catch	0.69	0.43-1.06	0.043	0.033-0.055	36	29-47	13	5-25
High catch	0.68	0.44-1.06	0.043	0.033-0.055	36	29-47	13	5-25
Low and low M	0.58	0.39-0.79		0.036	34	28-42	12	4-19
More acoustics	0.86	0.54-1.32	0.043	0.033-0.054	36	29-46	13	5-25

The estimated migration percentages are very stable across the runs with only the *assumed* migration percentages to Other showing any substantial variation (Table 7).

Table 7: MCMC estimates: medians and 95% CIs for the migration percentages to the numbered features and Other for the base model and sensitivities (see the text for the description of each run).

	<u>Other</u>		<u>Feature 1</u>		<u>Feature 2</u>		<u>Feature 3</u>		<u>Feature 4</u>		<u>Feature 5</u>	
	Base	20	16-24	13	11-16	11	9-14	15	11-20	31	27-36	9
Low	10	8-12	15	13-18	12	10-15	17	12-22	35	30-39	11	8-14
High	30	24-36	11	9-14	10	8-13	13	9-18	28	23-32	8	6-10
Uniform	20	16-24	13	11-16	11	9-14	15	11-20	31	27-36	9	7-12
AF80	20	16-24	13	11-16	11	9-14	15	11-20	31	27-36	9	7-12
Low catch	20	16-24	13	11-16	11	9-14	15	11-20	31	27-36	8	6-11
High catch	20	16-24	13	11-15	11	9-14	15	11-20	30	26-35	11	8-14
Low and low M	10	8-12	15	13-18	12	9-15	17	13-22	34	30-39	11	9-14
More acoustics	20	16-24	12	10-15	11	8-16	17	11-24	30	25-36	9	6-13

Estimates of B_0 do of course depend on the treatment of the acoustic data with the Low and High runs bracketing the base model (Table 8). Current spawning biomass and stock status also varies as expected with the treatment of the acoustic data. However, for all runs (even the very pessimistic “Low and low M ”) current stock status is almost certainly above 50% B_0 (Table 8).

Table 8: MCMC estimates: medians and 95% CIs for virgin spawning biomass (B_0), current spawning biomass (B_{17}), and current stock status ($ss_{17} = B_{17}/B_0$) for the base model and sensitivities (see the text for the description of each run). The estimated probability (%) of current spawning biomass being above 30% B_0 or 50% B_0 is also given.

	B_0 (000 t)	B_{17} (000 t)	ss_{17} (% B_0)	$P(B_{17} > 30\% B_0)$	$P(B_{17} > 50\% B_0)$
Base	43 29-64	32 19-53	76 63-87	100	100
Low	29 22-42	19 12-31	65 53-77	100	100
High	71 46-97	61 37-86	85 76-94	100	100
Uniform	42 29-64	32 19-53	75 63-86	100	100
AF80	43 30-67	32 19-55	74 62-85	100	100
Low catch	42 28-65	32 18-55	77 65-88	100	100
High catch	43 29-66	32 18-53	73 60-84	100	100
Low and low M	29 23-42	19 12-31	63 53-75	100	99
More acoustics	44 30-69	34 20-58	76 64-87	100	100

In terms of local depletion at the numbered features and Other (which represents all un-numbered features combined) it is only for Low and “Low and low M” that any features have any chance of being below 30% B_0 (Table 9). In both cases it is Other that may be experiencing a high level of local depletion (Table 9). That is just to say that some of the un-numbered features may have been fished to low levels in 2000 and 2001 depending on the model assumptions.

Table 9: MCMC estimates: medians and 95% CIs for the local depletion (%) of the numbered features and Other for the base model and sensitivities (see the text for the description of each run). Local depletion is the current spawning biomass (2017) divided by the virgin spawning biomass (for each feature or Other).

	<u>Other</u>	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
Base	75 60-87	66 51-79	99 90-107	89 80-98	66 49-80	71 57-83
Low	30 11-54	57 44-71	98 90-107	86 77-95	56 40-71	64 51-77
High	90 81-98	76 64-86	99 91-107	93 84-101	77 64-87	79 67-89
Uniform	74 59-85	65 50-78	97 88-105	88 78-96	65 48-79	70 56-82
AF80	74 59-85	65 50-78	97 88-105	88 78-96	65 48-79	70 56-82
Low catch	80 67-91	66 51-79	99 91-107	89 80-98	66 48-80	75 62-87
High catch	65 44-80	66 51-79	99 90-107	89 80-98	66 48-80	64 50-77
Low and low M	25 8-49	56 43-70	99 91-106	86 77-94	55 39-70	62 50-75
More acoustics	76 61-87	64 48-78	99 89-107	90 80-99	66 51-80	70 54-84

Projections

For the Base model a five year constant catch projection at current catch levels was performed. Overall stock status was maintained at acceptable levels being well above the 30-50% B_0 biomass range (Figure 19). Only Features 1 and 4 may possibly go below 50% in terms of local depletion (Figure 20).

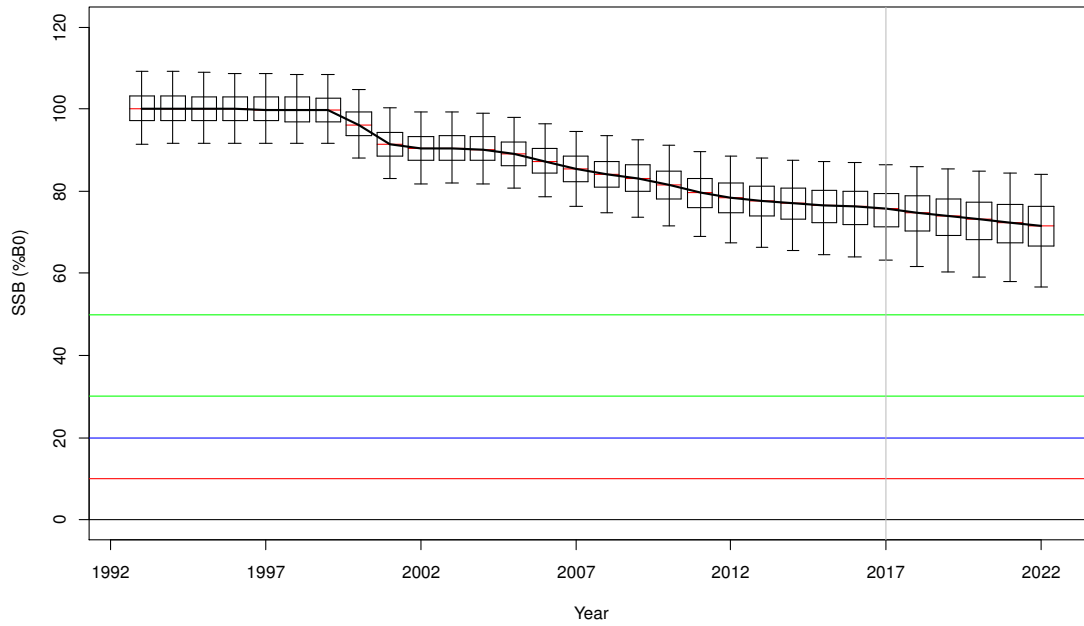


Figure 19: Base model MCMC: constant catch projection at current levels: box and whiskers plot of the spawning biomass trajectory for the whole stock. Each box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.

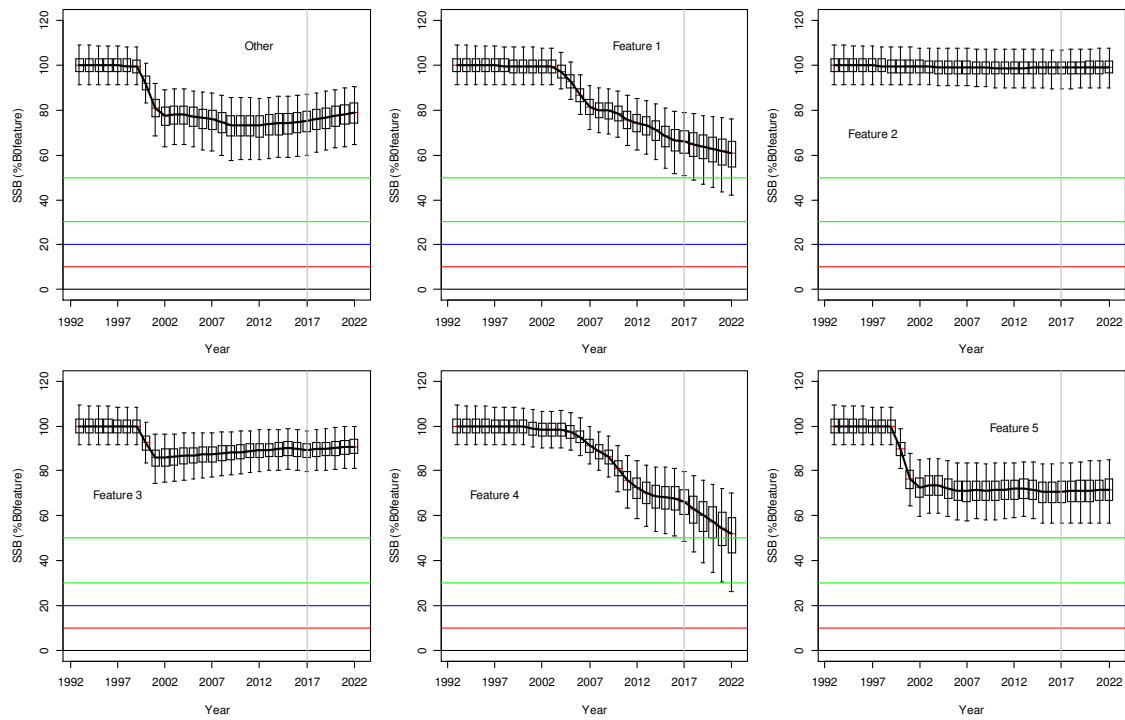


Figure 20: Base model MCMC: constant catch projection at current levels: box and whiskers plot of the spawning biomass trajectory for each numbered feature and collectively each un-numbered feature (Other). Each box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.

The constant catch projection for the Low model (which managers should keep an eye on – although it is unlikely it is still possible) shows that under the status quo catches that overall stock status is projected to stay well above the biomass range of 30-50% B_0 even under this pessimistic scenario (Figure 21). In terms of local depletion there is a possible problem with Feature 4 under this scenario if catches for that feature stay at the current level (almost 600 t) (Figure 22).

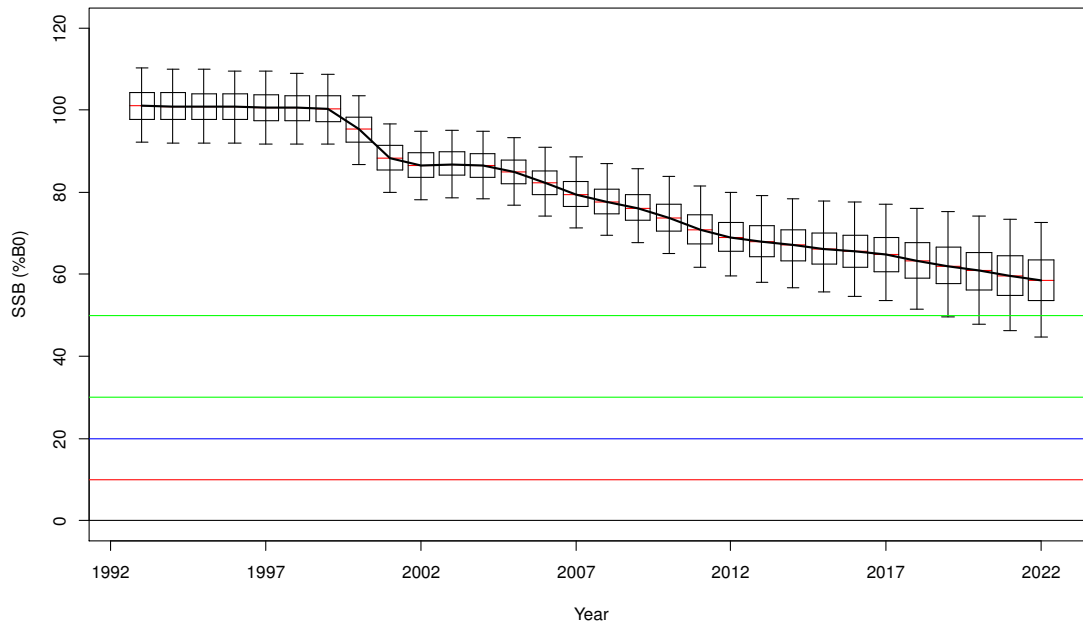


Figure 21: Low model MCMC: constant catch projection at current levels: box and whiskers plot of the spawning biomass trajectory for the whole stock. Each box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.

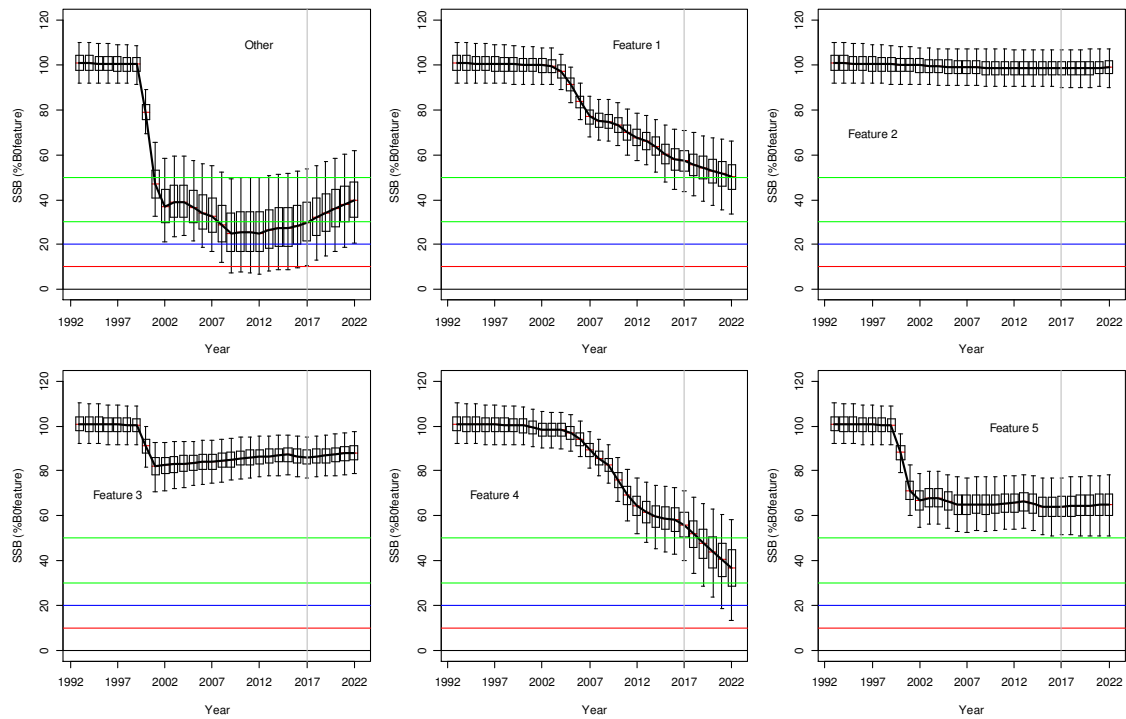


Figure 22: Low model MCMC: constant catch projection at current levels: box and whiskers plot of the spawning biomass trajectory for each numbered feature and collectively each un-numbered feature (Other). Each box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.

Conclusions

The absolute scale of the WSR stock is very uncertain because the proportionality constant for the acoustic biomass estimates is poorly known. Virgin biomass (B_0) is very likely to be in the range of 25 000 – 90 000 t. However, it is almost certain (given the stock structure assumptions) that current stock status is above 50% B_0 .

Local depletion is currently only an issue for some un-numbered features that may have been heavily fished in 2000 and 2001. The numbered features are all estimated to be well above 30% in terms of local depletion (current spawning biomass divided by virgin spawning biomass for the given feature).

Five year projections, at the current level and distribution of catches across features, show that there are no current sustainability concerns. However, under the more pessimistic model assumptions, Feature 4 could become overfished if catches for that feature are increased. The challenge is to find a management mechanism that will allow an increase in stock wide catches while maintaining the healthy stock levels at all individual features.

Acknowledgements

Many thanks to the Cook Islands delegation for the nomination for this project. Also, thanks to Graham Patchell for his many years of dedicated data collection that has made this stock assessment possible. Also, thanks to Graham for supplying much of the data that has gone into this assessment. Thanks also to Tom Nishida, Jon Lansley, and Pierre Peries for their efforts to facilitate the provision of data in a timely manner.

Finally, thanks to NIWA for the use of their excellent stock assessment package CASAL.

References

- Bull, B.; Francis, R.I.C.C.; Dunn, A.; Gilbert, D.J.; Bian, R.; Fu, D. 2012. CASAL (C++ algorithmic stock assessment laboratory): CASAL User Manual v2.30-2012/03/21. *NIWA Technical Report 135*. 280 p.
- Cordue, P.L. 2014a. The 2014 orange roughy stock assessments. *New Zealand Fisheries Assessment Report 2014/50*. 135 p.
- Cordue, P.L. 2014b. A management strategy evaluation for orange roughy. ISL Client Report for Deepwater Group Ltd. 42 p.
- Cordue, P.L. 2017. A 2017 stock assessment of ORH 3B Puysegur. Draft New Zealand Fisheries Assessment Report 2017/xx. 33 p.
- Doonan I.J. 1994. Life history parameters of orange roughy: estimates for 1994. New Zealand Fisheries Assessment Research Document 1994/19. 13 p.
- Doonan, I. J.; Coombs, R. F.; McClatchie, S. 2003. The absorption of sound in seawater in relation to the estimation of deep-water fish biomass. *ICES Journal of Marine Science* 60: 1047–1055.
- Doonan, I.J.; Tracey, D.M. 1997. Natural mortality estimates for orange roughy in ORH 1 (Bay of Plenty). New Zealand Fisheries Assessment Research Document 1997/26. 9 p.
- Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*. 68: 1124–1138.
- Francois, R. E.; Garrison, G. R. 1982. Sound absorption based on ocean measurements. Part II: Boric acid contribution and equation for total absorption. *The Journal of the Acoustical Society of America* 72: 1879–1890.
- Horn, P.L.; Doonan, I.J.; Maolagáin, C. Ó. 2017. Age distribution of orange roughy on the Sleeping Beauty seamount, Southern Indian Ocean. NIWA Client Report No: 2017410WN. 11 p.
- Jolly, G.M.; Hampton, I. 1990. A stratified random transect design for acoustic surveys of fish stocks. *Canadian Journal of Fisheries and Aquatic Sciences* 47: 1282–1291.
- Kloser, R.J.; Macaulay, G. J.; Ryan, T. E.; Lewis, M. 2013. Identification and target strength of orange roughy (*Hoplostethus atlanticus*) measured in situ *Journal of the Acoustical Society of America* 134: 97–108.
- Kloser R.J.; Scoulding, B.; Niklitschek, E.; Toledo, P.; Patchell, G.J. 2018. Orange roughy biomass estimation in SIOFA. Review of the use of acoustics from industry vessels. CSIRO. SAWG(2018)-01-INFO09.

- Macaulay, G. J.; Kloser, R. J.; Ryan, T. E. 2013. *In situ* target strength estimates of visually verified orange roughy. *ICES Journal of Marine Science* 70: 215–222.
- McClatchie, S.; Alsop, A.; Coombs, R.F. 1996. A re-evaluation of relationships between fish size, acoustic frequency, and target strength. *ICES Journal of Marine Science* 53: 780–791.
- McClatchie, S.; Macaulay, G. J.; Coombs, R.; Grimes, P.; Hart, A. 1999. Target strength of an oily deep-water fish, orange roughy. In situ target strength estimates of visually verified orange roughy (*Hoplostethus atlanticus*). Part I. Experiments. *Journal of the Acoustical Society of America* 106: 131–142.
- Niklitschek, E. J.; Patchell, G. J. 2015. Acoustic abundance indexes for orange roughy and alfonsino in the Indian Ocean (SIOFA) from industry acoustics 2004-2008. Southern Indian Ocean Fisheries Agreement, Scientific Committee Report. SC-01-INFO-15.
- Patchell, G.J. 2016. Biological data from orange roughy spawning stocks in the Southwest Indian Ocean. Southern Indian Ocean Fisheries Agreement, Scientific Committee Report. SC-01-INFO-16
- Roa-Ureta, R.; Niklitschek, E. 2007. Biomass estimation from surveys with likelihood-based geostatistics. *ICES J. Mar. Sci.* 64: 1723–1734
- Scouling, B.; Kloser, R. J. 2018. Review of SIOFA orange roughy (*Hoplostethus atlanticus*) acoustic data. CSIRO. SAWG(2018)-01-INFO10.

Appendix A: Estimation of length-weight and growth parameters

The length-weight relationship and von Bertalanffy growth parameters for the Walter’s Shoal Region (WSR) assessment were estimated from data collected at Sleeping Beauty. The equations were estimated by sex and then an average relationship was calculated for use in the single sex stock assessment model. Length-weight relationships were also calculated for other features in the WSR and compared to the Sleeping Beauty relationship.

Data

The data were supplied to ISL by Graham Patchell of Sealord Group (see Patchell 2016). For Sleeping Beauty there were 8321 length-weight records where fish were sexed. Also, there were 399 otolith readings for measured and sexed fish (Horn et al. 2017).

The WSR features for which sexed length-weight data were supplied, with the number of fish were:

Sleeping Beauty	8321
Boulder	5528
Sleepy Hollows	3768
Splitpin	1989
Porkys	964
Abby Road	1789
Coopaville	1572

Methods

The length-weight relationship was estimated separately for males and females using a log-log regression in R (using the function “lm”). An average relationship was then determined by averaging the male and female multipliers (the “a”s) and finding (by a least squares fit) the exponent (“b”) that gave the average fish weight at each length assuming males and females were 50/50. “Outliers” were removed before performing the regression (e.g., for Boulder these was a group of large females that were very underweight compared to the vast majority of the fish – most appear to have been sampled from just a single tow).

The von Bertalanffy parameter t_0 was assumed equal to -0.5 and k and L_{inf} were estimated by least squares (minimizing the sum of the squared differences between the predicted and observed values). A small negative t_0 gives a small positive length at age 0. It is necessary to fix t_0 otherwise it is confounded with k (given there are no very young fish in the observations). The average relationship was calculated by averaging the male and female estimates of L_{inf} and finding the k by least squares.

Results

The length-weight data for Sleeping Beauty, which were collected by random sampling, show that males and females are approximately equally represented in the catch (47% male, 53%

female). A single-sex model is normally used for orange roughy and it is standard practice in New Zealand to assume that the population is 50% male and 50% female (sex ratio is not well estimated in the population from catch data as individual trawl catches can often be very much skewed to one sex or the other). Therefore, age and length frequencies are usually stratified by sex and then each sex given equal weight in the final combined age or length frequency (used in the single-sex model). The requirement for length-weight and age-length relationships is that they will produce the correct average mean weight at age assuming males and females are 50/50 at each age.

Length-weight relationship

As expected, females were seen to grow to greater lengths than males and at given length were of greater weight (Figure A1). Because of the unbalanced nature of the data (only male fish at small lengths and only female fish at large lengths) the calculation of a length-weight relationship using all data and ignoring sex produces a distorted relationship (Figure A2). For given length, we want the average weight assuming a 50/50 split of sexes, so an average relationship is required (Figure A3).

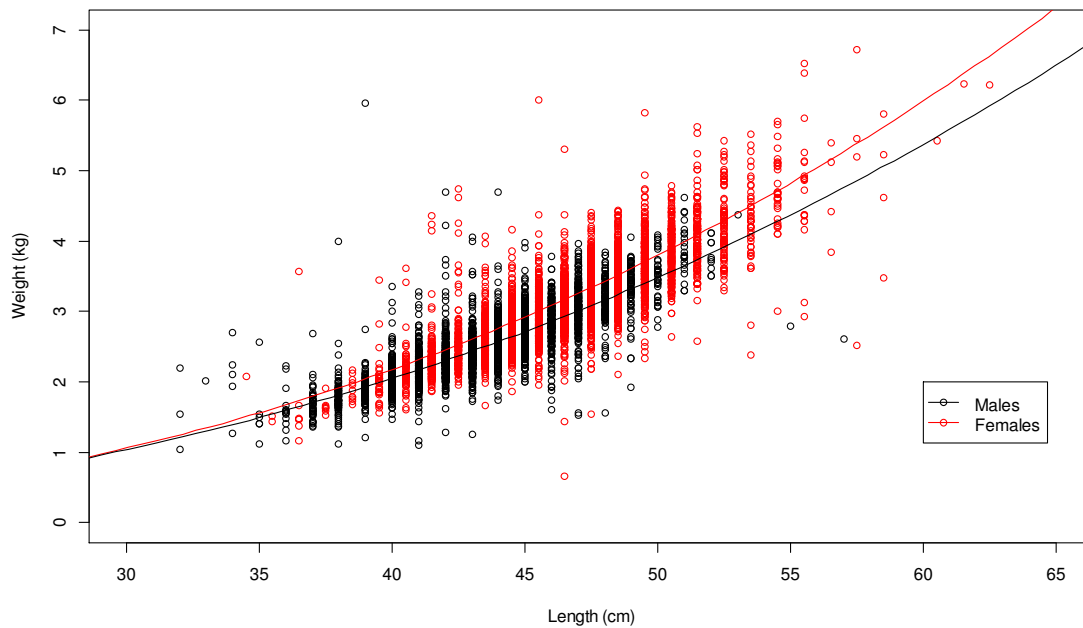


Figure A1: Sleeping Beauty: the length-weight data for males and females with fitted curves. The female lengths have been increased by 0.5 cm for presentation purposes only.

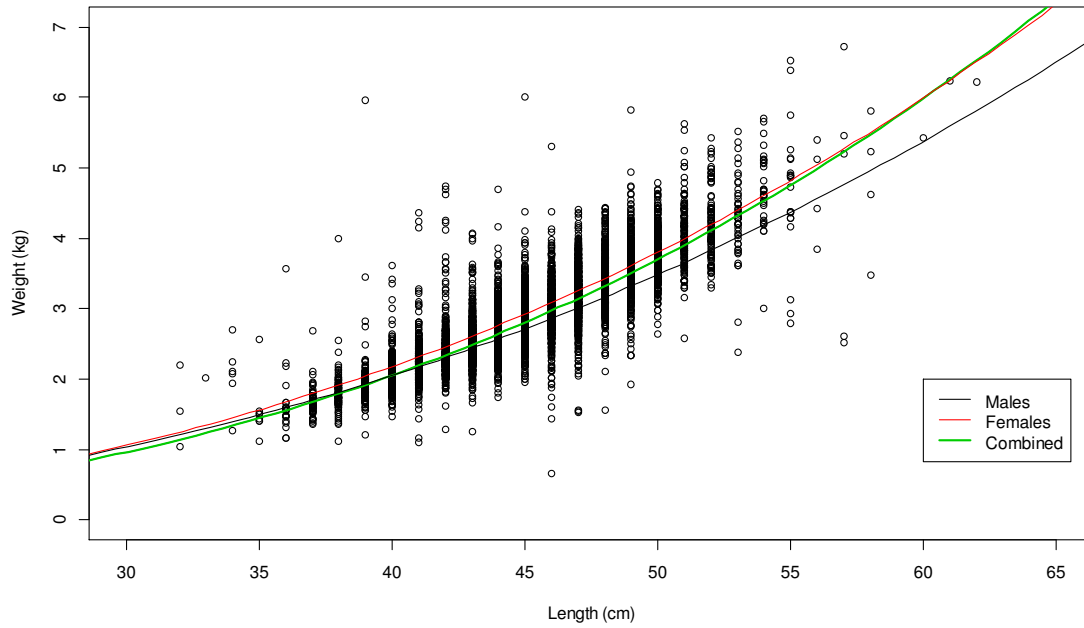


Figure A2: Sleeping Beauty: the length-weight data with the estimated relationships for males (black), females (red), and the full data set (green)(ignoring sex).

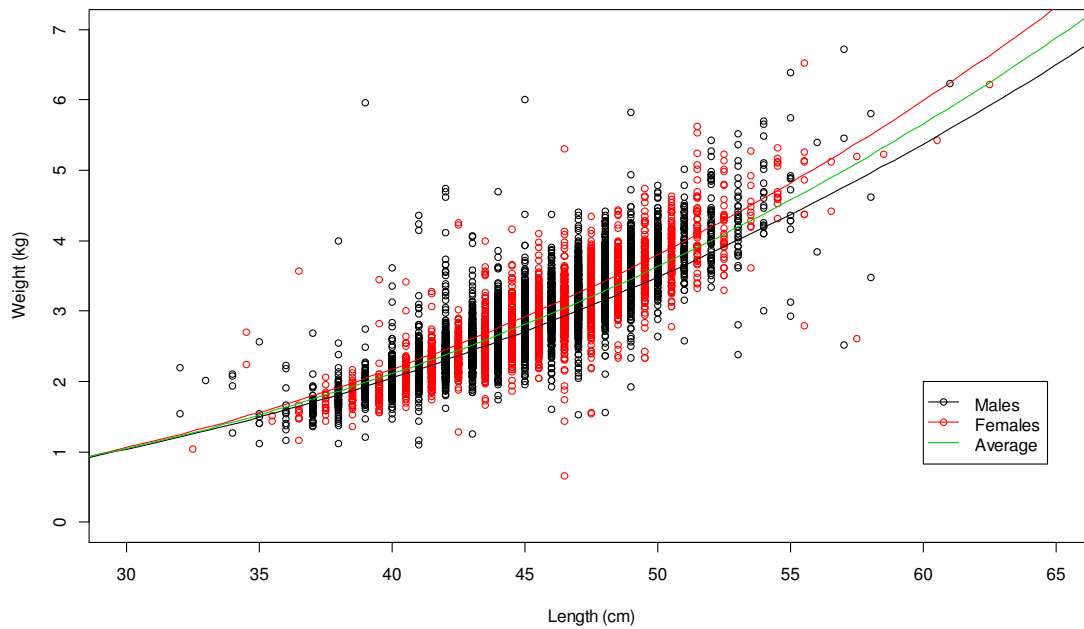


Figure A3: Sleeping Beauty: the length-weight data with the estimated relationships for males (black), females (red), and the average relationship (green). The female lengths have been increased by 0.5 cm for presentation purposes only.

For Sleeping Beauty, the average relationship is: $w = 0.000265 L^{2.436}$

This was derived from the male and female relationships:

Male: $w = 0.000316 L^{2.38}$

Female: $w = 0.000215 L^{2.50}$

Weight measured in kg and (standard) length in cm.

The full set of estimated relationships:

	Male		Female		Single sex	
	a (10⁻⁴)	b	a (10⁻⁴)	b	a (10⁻⁴)	b
Sleeping Beauty	3.16	2.38	2.15	2.50	2.65	2.436
Boulder	8.97	2.09	3.11	2.39	6.04	2.206
Sleepy Hollows	2.86	2.40	0.827	2.74	1.84	2.524
Splitpin	1.96	2.51	0.727	2.78	1.34	2.615
Porkys	5.53	2.23	2.02	2.51	3.77	2.339
Abby Road	2.60	2.43	1.16	2.65	1.88	2.520
Coopaville	1.17	2.65	0.804	2.75	0.989	2.695

The single-sex length-weight relationships for the different features (with the slight exception of Boulder) are all very similar across the range of lengths sampled (Figure A4).

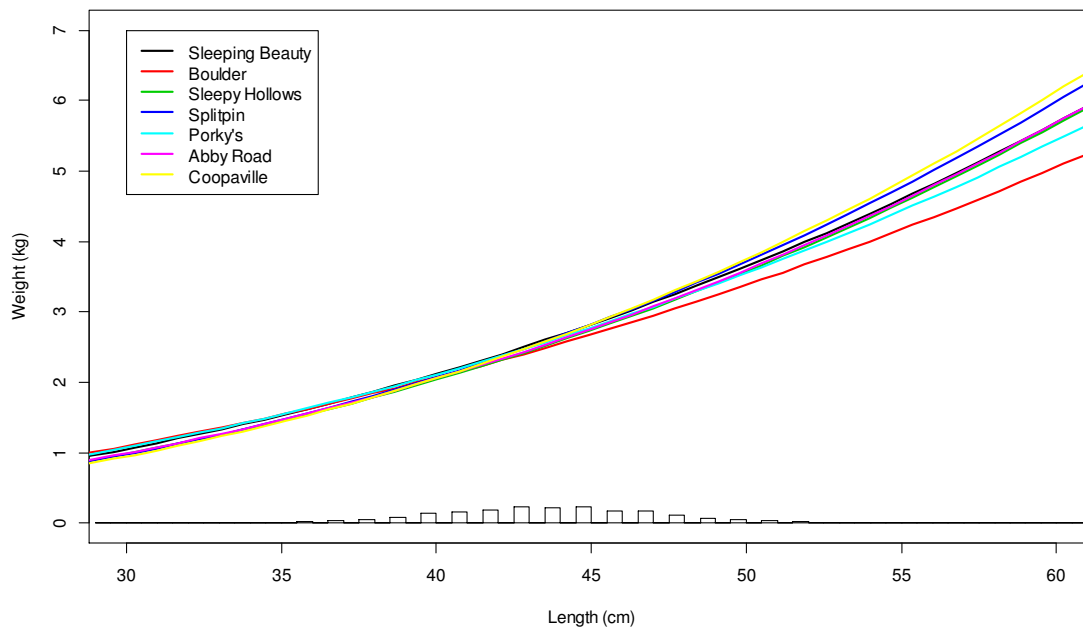


Figure A4: Single-sex length-weight relationships for the WSR features with a histogram of the length data,

Age-length data

The only age-length data come from the otolith readings that were done to construct a spawning age frequency for Sleeping Beauty in 2017 (Horn et al. 2017). Very young fish are of course absent from the spawning aggregations so the initial growth cannot be estimated. However, we only need growth parameters so that we can get mean length at age and hence mean weight at age for recruited fish. Therefore, it is adequate to assume a sensible value for t_0 and just estimate k and L_{inf} .

As expected, females are somewhat larger than males at given age (Figure A5).

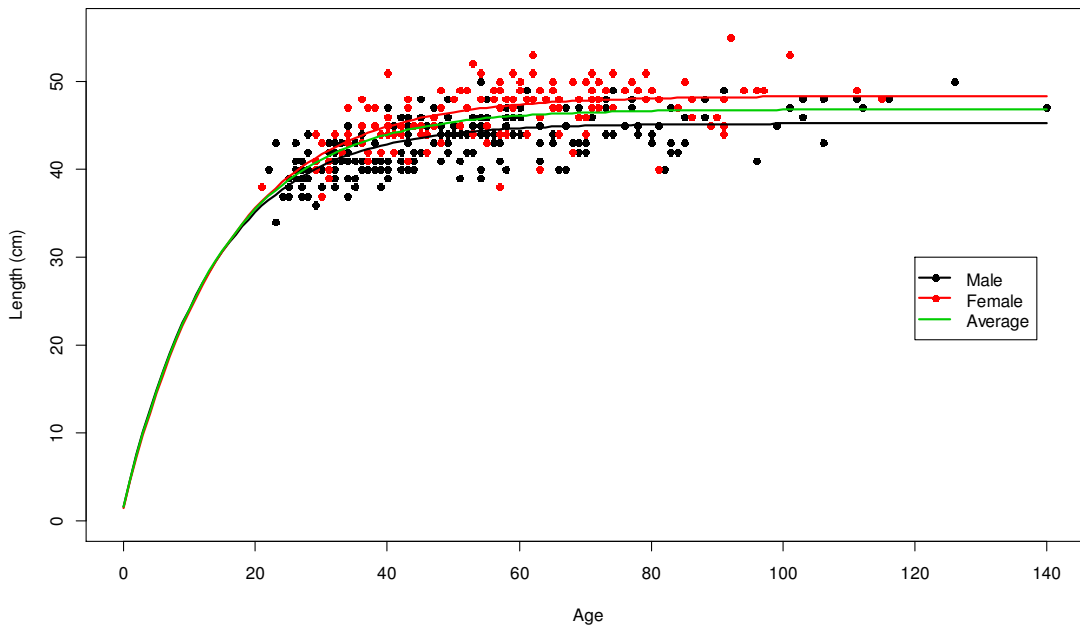


Figure A5: Sleeping Beauty: age-length data for males and females with fitted curves. The average relationship for a single-sex model is also plotted (green).

All relationships assumed $t_0 = -0.5$. The average relationship was derived by averaging the male and female L_{inf} and using least squares to estimate the k (which coincidentally just came out at the average of the male and female k s):

	k	L_{inf} (cm)
Male	0.073	45.2
Female	0.065	48.3
Average	0.069	46.75

Appendix B: Construction of the catch history

Catch and position data were supplied by SIOFA and by Graham Patchell from Sealord group. The SIOFA data were primarily from New Zealand, Australia, and Japan. There was an overlap in the New Zealand data and the Sealord data for one vessel in one year. The Sealord data were supplied as a catch history from 1999 to 2017 inclusive for each of 13 features in the WSR. The remaining data were tow by tow with catch, year, and position. Where accurate positions were given the catch was assigned to the nearest WSR feature within 5 n.m. of the starting tow position. Otherwise if the catch was in the WSR it was assigned to Others (where any catch not allocated to one of the five numbered features was accumulated).

There were a large number of vessels fishing for orange roughy in the SIOFA area in 2000 and 2001. It was clear that not all of the catch had been accounted for and 2000 t of catch was allocated to 3 features in 2000, and 750 t of catch was allocated to the same 3 features in 2001 (Graham Patchell, pers. comm.) (Table B1). The level of catch in these two years is unknown so sensitivity models were run at half and double of the assumed add-on catches.

Table B1: The catch (t) for each numbered feature from 1999 to 2017 inclusive for the WSR base model. The catch for all unnumbered features is accumulated in Others.

Year	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Others	Total
1999	0	0	0	0	0	0	0
2000	0	0	869	0	750	1250	2869
2001	0	1	28	144	348	700	1221
2002	0	0	0	0	0	11	11
2003	0	0	4	0	1	1	6
2004	222	0	2	0	63	110	397
2005	284	0	1	203	93	156	737
2006	428	0	16	382	71	121	1018
2007	217	0	0	650	15	96	978
2008	0	0	0	110	35	276	421
2009	100	0	0	604	56	108	868
2010	121	0	1	837	11	34	1004
2011	270	0	4	753	38	121	1186
2012	16	0	1	391	28	58	494
2013	219	0	18	496	21	16	770
2014	115	0	1	171	106	105	498
2015	301	0	8	292	62	38	701
2016	60	8	97	158	33	60	416
2017	145	0	8	599	40	19	811
Total	2498	9	1058	5790	1771	3280	14 406

Appendix C: Acoustic biomass estimates and target strength

Acoustic biomass estimates were supplied to ISL by Graham Patchell of Sealord Group (see Niklitschek and Patchell 2015). Acoustic surveys of features were conducted opportunistically during fishing trips using a calibrated acoustic system primarily during the orange roughy spawning season. Biomass estimates were calculated using geostatistical methods (Niklitschek and Patchell 2015, Roa-Ureta and Niklitschek 2007). Acoustic estimates were available for seven features within the WSR in years ranging from 2004 to 2015.

Recently, the acoustic biomass estimates were reviewed and, for selected surveys, revised estimates were calculated (Kloser et al. 2018, Scouling and Kloser 2018). Also, new target strength data were collected and analysed for orange roughy on Sleeping Beauty.

Main review results

The review found that one of the original survey estimates considered was much higher than it should have been due to fish movement and the calculation of the effective survey area. Two grid surveys had been combined and it was assumed that the fish were not moving. This resulted in a large survey area and some double counting of fish marks. In the reanalysis each grid survey was analysed individually and a much smaller estimate was obtained (approximately a factor of 3). Other revisions of survey estimates analysed using the same geostatistical methods resulted in much smaller differences (I noted four comparisons that could be made from the estimates supplied to ISL, revised estimates were 67%, 78%, 110%, and 111% of the original estimates.)

The review noted that for grid surveys that could be analysed by more traditional design-based methods (e.g., Jolly and Hampton 1990) that estimates were typically higher than those obtained from the geostatistical approach. For 14 survey estimates where traditional (“EDSU”) and geostatistical estimates were given, the ratio of the EDSU estimate to the geostatistical estimate ranged from 0.9–1.9 with a mean ratio of 1.27.

The formula used for the calculation of sound absorption was noted to be important. Doonan et al. (2003) was used but the alternative of Francois and Garrison (1982) would result in estimates being approximately 30% higher.

The target strength of orange roughy was also noted to be important as the existing relationships were based on much smaller fish. Importantly, new target strength data were collected on Sleeping Beauty in the WSR.

Length-target strength relationships for orange roughy

The length-target strength relationship for orange roughy used in New Zealand for many years is:

$$TS = 16.15 \log_{10}(L) - 76.81 \quad (\text{NZ relationship})$$

The slope of 16.15 comes from a draft of the paper McClatchie et al. (1999). When the paper was eventually published the slope of the relationship given in the paper was 16.374. This slope was used in conjunction with an orange roughy target strength estimate in Kloser et al. (2013) to provide the relationship used for the acoustic estimates provided to ISL:

$$TS = 16.37 \log_{10}(L) - 77.17 \quad (\text{McClatchie-Kloser relationship})$$

It happens that the two relationships are almost identical (see Figure B1).

With the new target strength data from Sleeping Beauty there are now three AOS estimates of orange roughy target strength at given mean length:

	Mean length (cm)	Target strength (dB)	Sample size
Macaulay et al. 2013	33.9	-52.0	24 fish, 83 echoes
Kloser et al. 2013	35.9	-50.7	110 fish, 401 echoes
Scolding and Kloser draft	45.0	-48.1	38 targets > 3 m

There was an estimate from tracking fish on Sleeping Beauty (-47.3 dB) but since it is only based on 6 fish the estimate is not used and it is preferred to use the estimate from 38 targets.

The best fit to the three points gives a relationship with a slope of 30.2 which is too high to be credible (e.g., see McClatchie et al. 1996). Two alternative fits were tried: the best fit given a slope of 20 (an often used default value) and fixing the slope at 16.15 (Figure B1):

$$TS = 20.00 \log_{10}(L) - 81.86 \quad (\text{slope fixed at 20})$$

$$TS = 16.15 \log_{10}(L) - 75.78 \quad (\text{slope fixed at 16.15}).$$

For fish of approximately 45 cm the first relationship implies that the biomass estimates should be reduced to 75% of the original value. For the second relationship, the estimates should be reduced to 79% of the original value.

Of course, it may be inappropriate to put the “small fish” and the “large fish” on the same length-target strength relationship. They are the same species but perhaps given the large difference in size a linear relationship (in log space) may not be appropriate. In that case, the best adjustment to the original estimates is to simply use the measurement of -48.1 dB which means the estimates should be reduced to 63% of the original value (a 2 dB difference for fish of 45 cm).

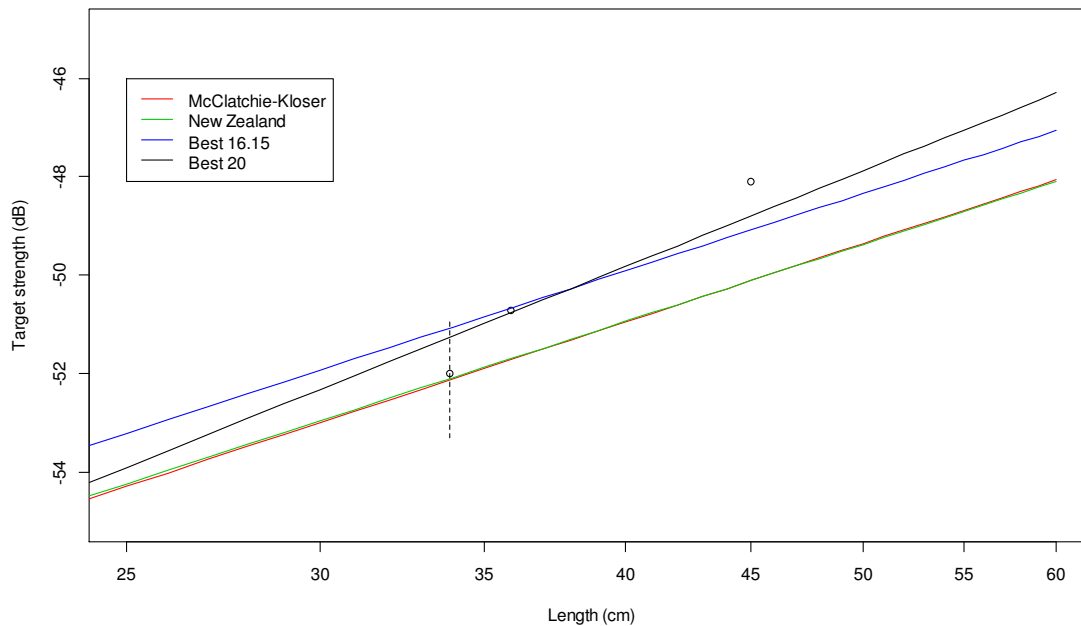


Figure B1: Least squares fits to the three AOS estimates of target strength for given mean length of fish. Also shown is a 95% CI for the Macaulay et al. (2013) estimate.

Acoustic biomass estimates

Three different variations of the acoustic estimates were used in stock assessment models: low, middle, and high. The different variations used combinations of alternatives: Doonan et al. (2003) or Francois and Garrison (1982); geostatistical estimates or EDSU; and the McClatchie-Kloser TS relationship or the best fit 16.15 revised relationship. The existing biomass estimates used Doonan, geostatistics, and McClatchie-Kloser.

The “low” estimates were calculated by using Doonan, geostatistics, and the new TS measurement. For this variation the original estimates were reduced to 63% of the original value.

The “middle” estimates were calculated by using Doonan, EDSU, and best fit 16.15. For this variation the original estimates were used because the adjustments for EDSU and best fit 16.15 cancel out ($1.27 \times 0.79 = 1.00$).

The “high” estimates were calculated using Francois and Garrison, EDSU, and McClatchie-Kloser. The adjustment required is $1.3 \times 1.27 = 1.65$.

In the main stock assessment models only the revised biomass estimates were used (Table B1). Any estimates from 2008 or 2005 were excluded. In 2008 there was a clear problem with the transducer and in 2005 there is a suspicion that transducer sensitivity was reduced

and the biomass estimates are lower than they should be (Kloser et al. 2018). Also, any estimates with very large CVs were excluded (a very large CV suggests that the fish were not well surveyed). Finally, only estimates that were noted as being at the peak of the spawning season were used.

Table B1: Revised acoustic biomass estimates for features in WSR from Kloser et al. (draft). See the text for the low, middle, and high treatments. The features are identified by a number only for confidentiality.

Feature	Year	Low estimate (t)	Middle estimate (t)	High estimate (t)	CV (%)
1	2007	1829	2902	4790	11
	2015	2386	3788	6250	32
2	2015	1993	3164	5221	12
3	2015	2381	3779	6235	20
4	2007	4991	7923	13073	10
	2009	6689	10618	17520	30
5	2009	1138	1806	2980	21
	2011	1094	1737	2866	43

In a sensitivity model, the original biomass estimates were used as there were far more of them and they included some potentially useful trend information within features (Table B2). They were only used with the “middle” treatment. When there are multiple surveys in the same year on the same feature then the average is used as the estimate.

Table B2: Original acoustic biomass estimates for features in WSR. These have received the “middle” treatment (which means they are unchanged). For some years for some features only revised estimates were made available. These are marked with an *. The features are identified by a number only for confidentiality.

Feature	Year	Estimate (t)	CV (%)	
1	2004	7459	51	
		6114	44	
		8923	37	
		9308	25	
		7951	19	
	2004 (average)			
	2007	2902	11	
	2009	3327	34	
2	2015	4542	32	
		3788	32	
3	2015	4043	21	
4	2015	5648	27	
		2009	5752	30
		9137	29	
		9562	40	
		17289	36	
	2009 (average)	9108	33	
	2010	10170	17	
	2010	13542	26	
	2010 (average)	8240	23	
	2015	10891	18	
5	2015	5269	30	
		2009	2501	34
		2011*	1737	43

Appendix D: MCMC chain diagnostics for the base model

The main models have several migration parameters and this makes the models relatively slow compared to non-migration models. For the New Zealand orange roughly assessments (non-migration models) typically 3 chains of 15 million were run with 1 in every 1000 samples being retained. This approach was not possible for the WSR models because they were too slow due to the migrations. Instead 5 chains of 2.5 million were run with 1 in every 1000 samples retained. The first 500 retained samples were discarded as a “burn-in” (the chains start at a random jump from the MPD estimates and therefore over-sample the region near the MPD at the start of the chain). Estimates and graphics were produced using the five concatenated chains (after burn-in) with a total sample size of 10 000.

Four diagnostics were used to judge whether the five chains provided an adequate sample of the posterior distribution (i.e., that there were no obvious signs that the chains had not converged). The first diagnostic is simply to plot the objective function value for each of the five chains (e.g., Figure D1). The move away from the MPD estimate is clearly seen at the start of the chains and a burn-in of 500 samples appears adequate. After the burn-in period the objective function appears to be in a steady stochastic state (Figure D1).

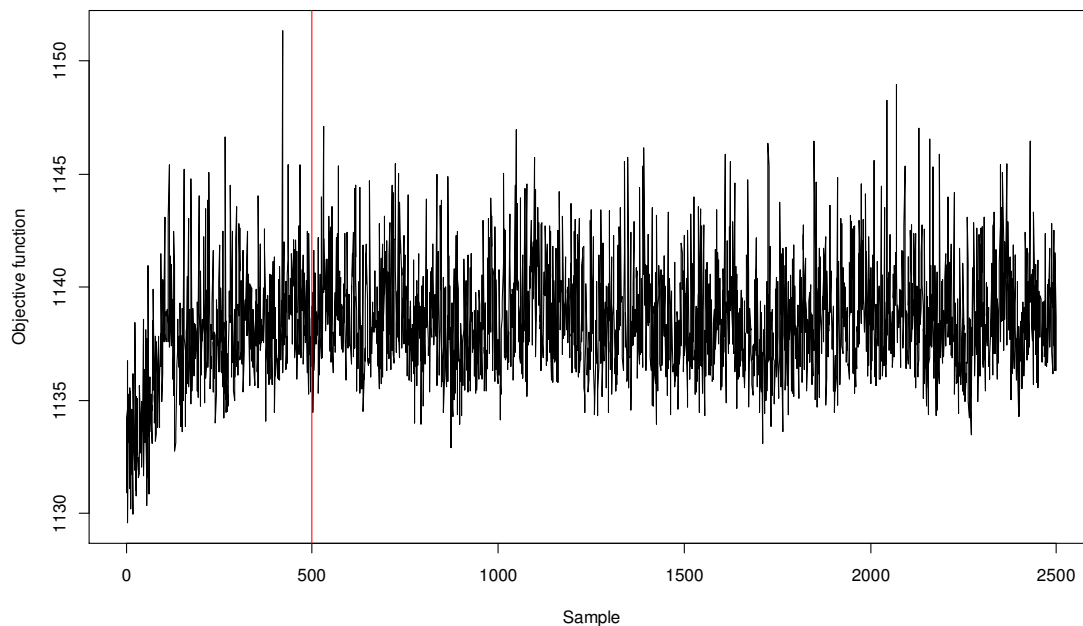


Figure D1: The objective function value at each retained sample for one of the five chains used for the base model. The vertical line marks the burn-in (samples to the left are discarded).

The second diagnostic is simply to view the chains of individual parameters (e.g., Figure D2-D6). The key point is that the chains have a relatively short period as they oscillate between low and high values. There are strong correlations but over the 5 chains combined there are

many oscillations between low and high values from the start to the end of the concatenated chain (so the chains appear to be long enough given that there are 5 of them).

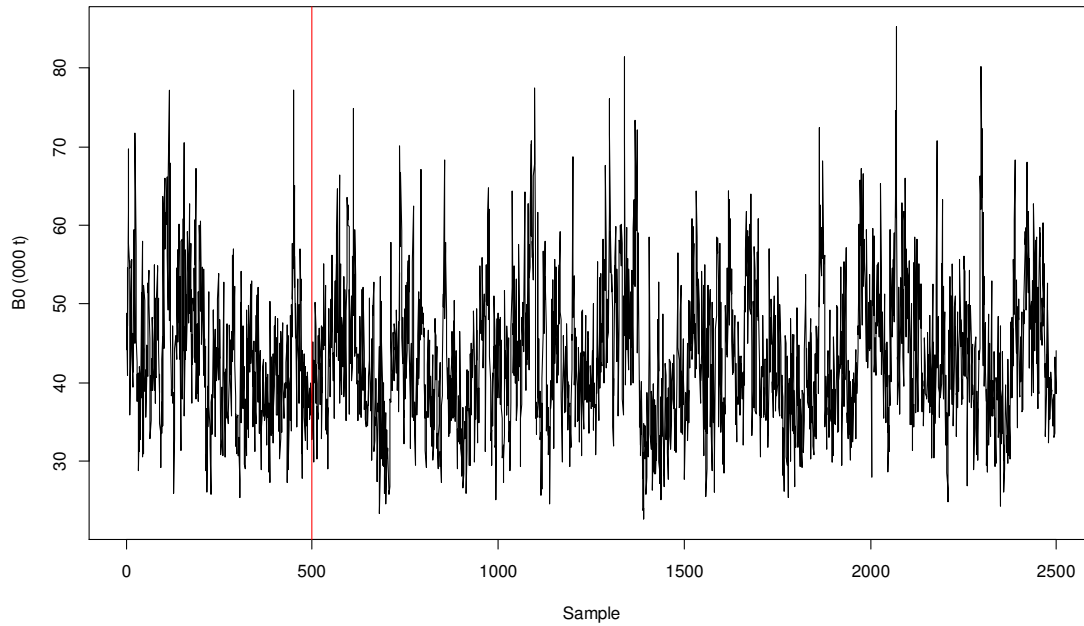


Figure D2: The samples of B_0 for one of the five chains used for the base model. The vertical line marks the burn-in (samples to the left are discarded).

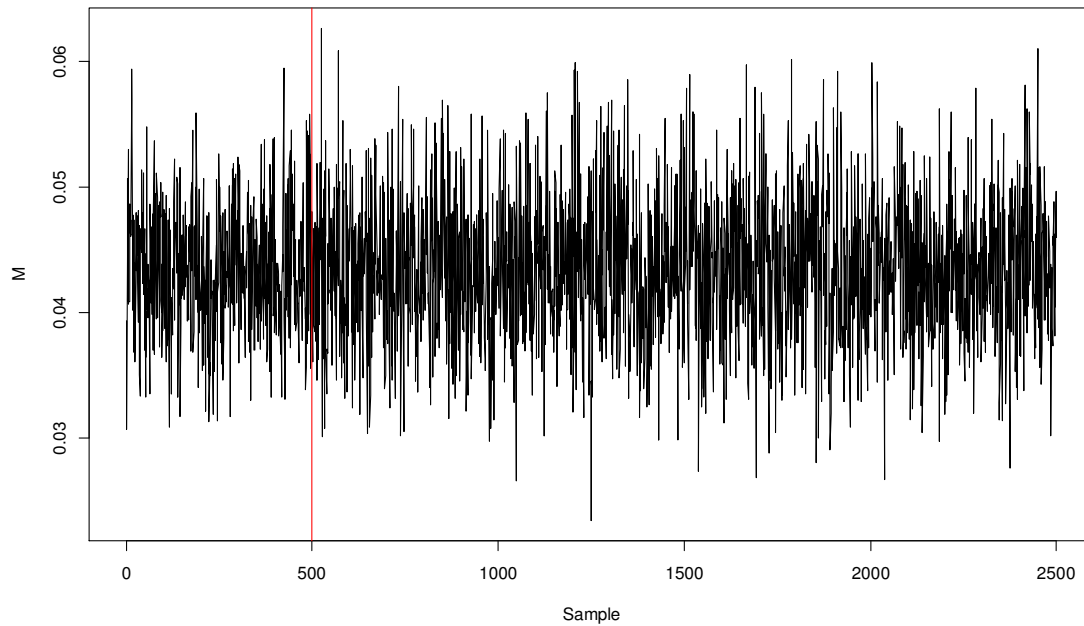


Figure D3: The samples of M for one of the five chains used for the base model. The vertical line marks the burn-in (samples to the left are discarded).

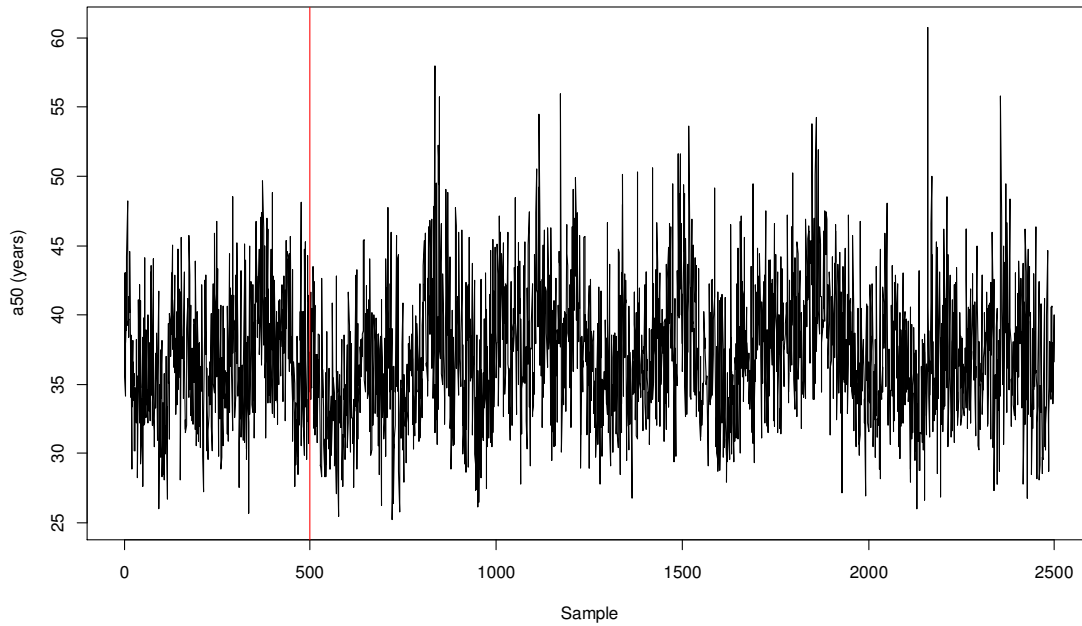


Figure D4: The samples of a_{50} for one of the five chains used for the base model (a_{50} is the age at 50% maturity). The vertical line marks the burn-in (samples to the left are discarded).

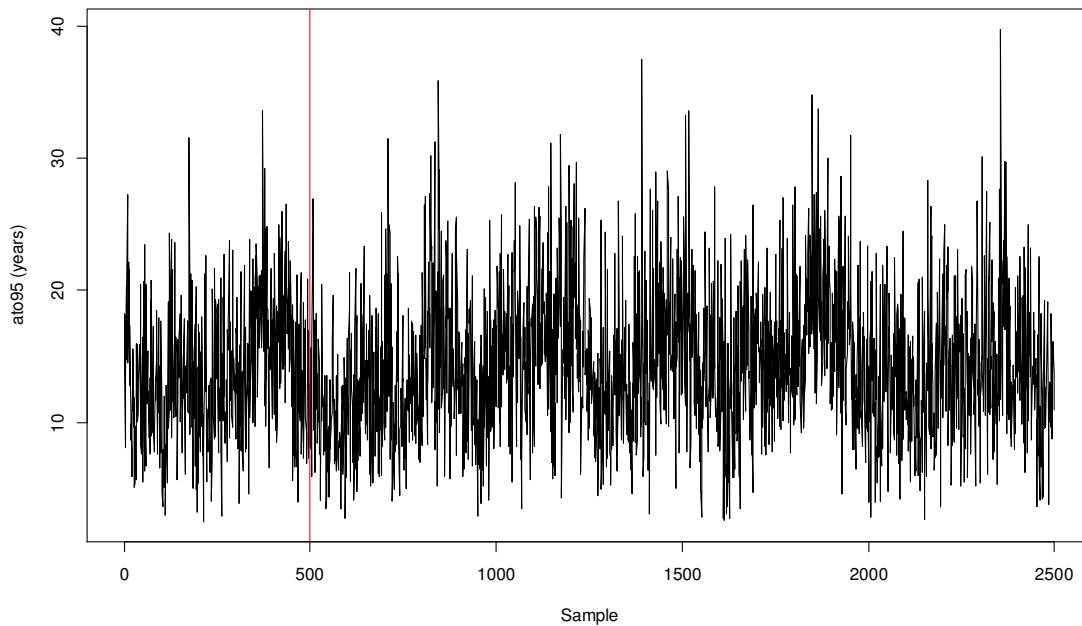


Figure D5: The samples of a_{1095} for one of the five chains used for the base model (a_{1095} is the difference between the age at 50% maturity and the age at 95% maturity). The vertical line marks the burn-in (samples to the left are discarded).

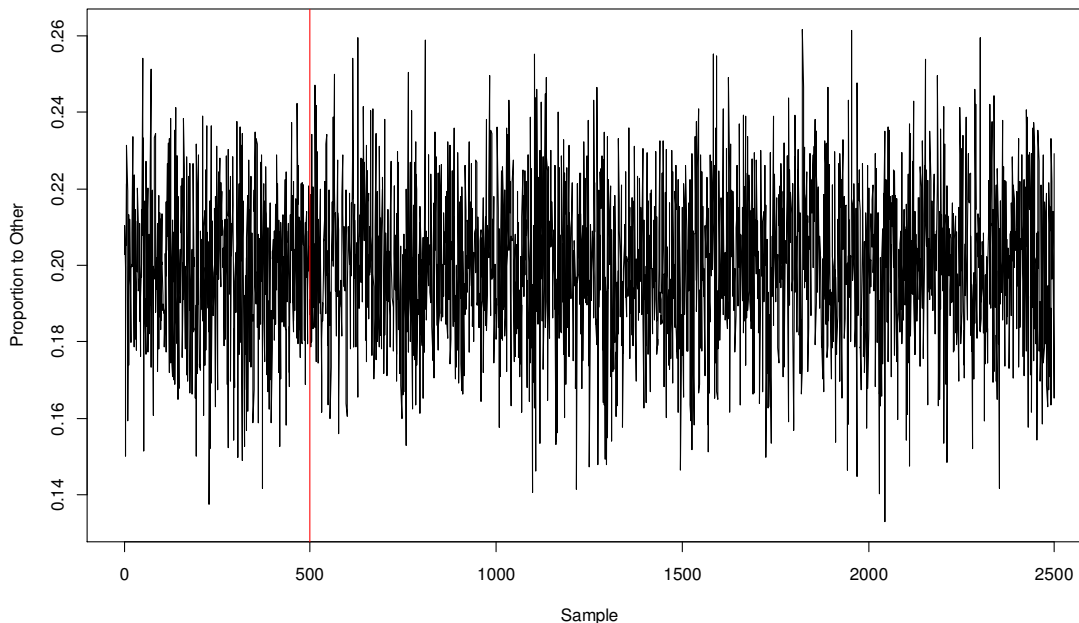


Figure D6: The samples of p_{other} for one of the five chains used for the base model. (p_{other} is the proportion of mature fish that migrate to “Other” in the model. “Other” is a catch-all for features not explicitly modelled.) The vertical line marks the burn-in (samples to the left are discarded).

The consistency of parameter estimates between the first half of the chains (after burn-in) and the second half of the chains was checked for the concatenated chain. For each parameter the mean was calculated for the concatenated burn-in segments (5 lots of 500), the concatenated first halves of the chains (5 lots of 1000) and the concatenated second halves of the chains (5 lots of 1000). These means were then standardised by dividing them by the mean for the full non burn-in segments of the chains (Figure D7). Deviations of more than 20% from the overall mean would be of some concern but the parameters show very little deviation except for some of the free year class strength parameters (Figure D7).

The final check is to compare histograms of the marginal posterior distributions of the main parameters and derived values over the 5 chains. The five chains are individually telling very much the “same story” for B_0 , M , and stock status in 2017 (B_{2017} as $\%B_0$) (Figures D8-D10 respectively). Small differences in median estimates across the chains are to be expected (e.g., Figure D10). The final estimate is from all 5 chains combined (which gives much higher precision for the point estimates and credibility intervals than any single chain).

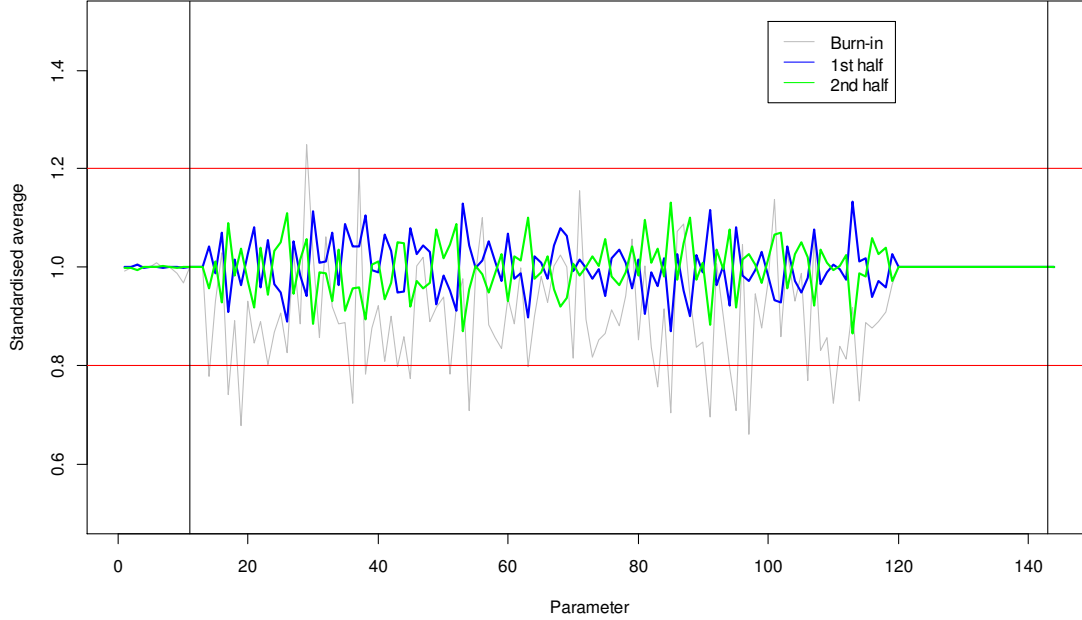


Figure D7: Standardised average values for each free parameter in the base model. The standardised averages are given for the burn-in segments (concatenated over the 5 chains), the first half of each chain (non burn-in, concatenated over the 5 chains), and the second half of each chain (non burn-in, concatenated over the 5 chains).

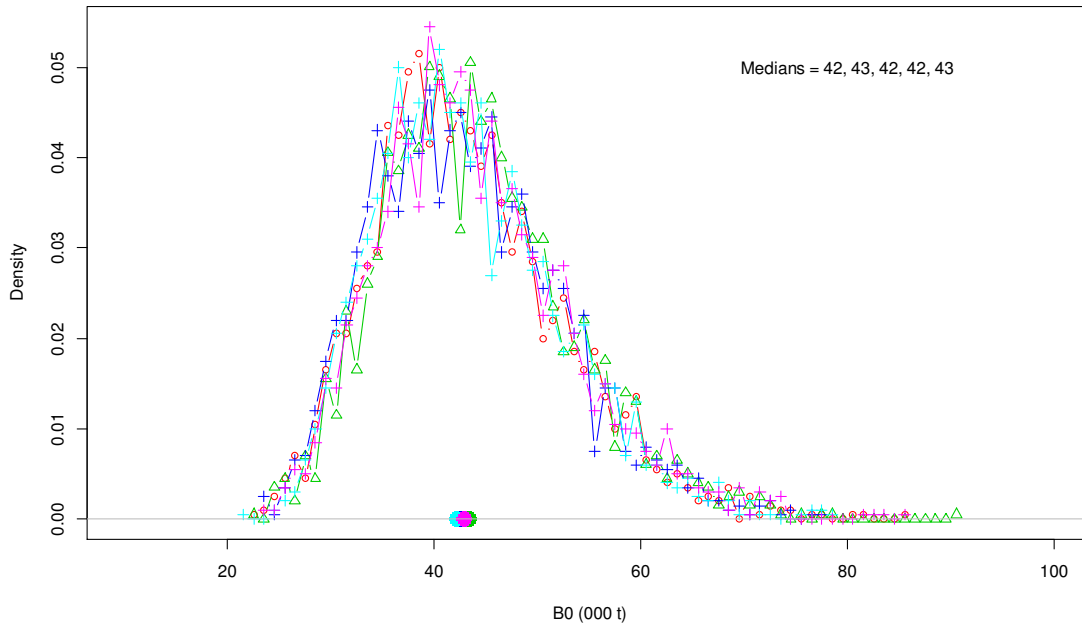


Figure D8: The samples from the marginal posterior distribution of B_0 for each of the 5 chains used in the base model. The solid dots mark the median for each chain and the values are given in the top right hand corner.

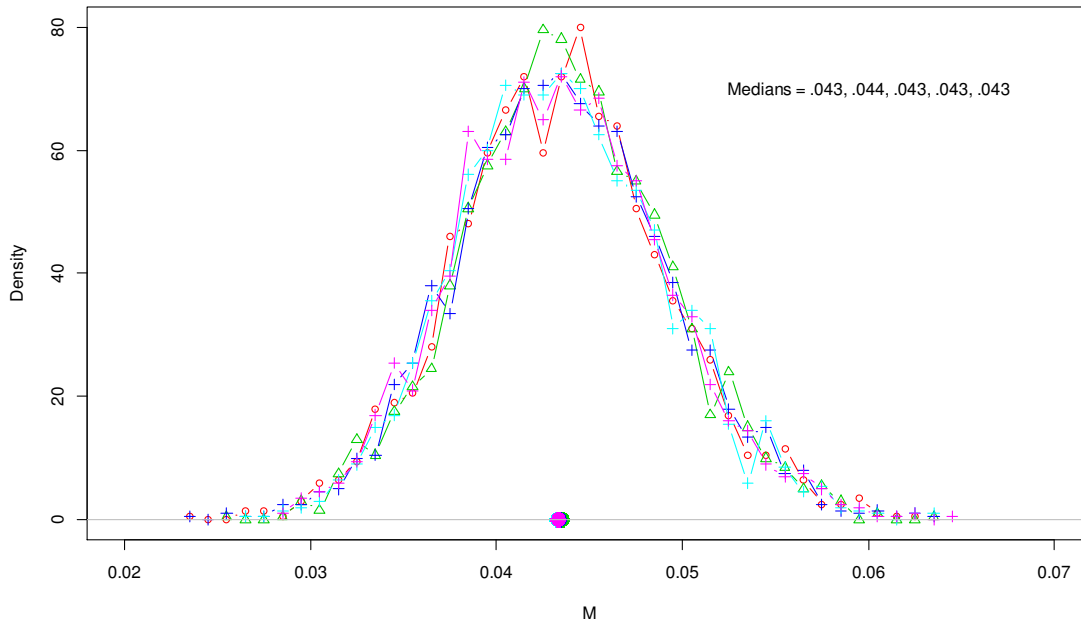


Figure D9: The samples from the marginal posterior distribution of M for each of the 5 chains used in the base model. The solid dots mark the median for each chain and the values are given in the top right hand corner.

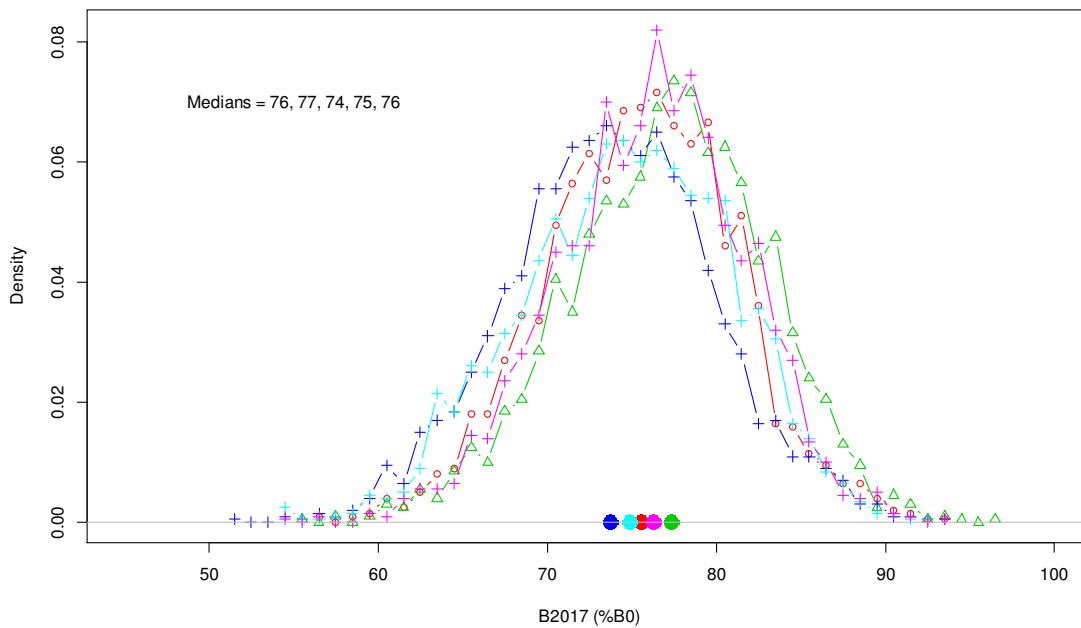


Figure D10: The samples from the marginal posterior distribution of stock status in 2017 (B_{2017} as $\% B_0$) for each of the 5 chains used in the base model. The solid dots mark the median for each chain and the values are given in the top left hand corner.

Appendix E: Marginal posterior distributions and MCMC fits and residuals for the base model

This appendix contains the full set of marginal posterior distributions for the free parameters in the base model together with the fits to the acoustic biomass estimates and the age frequency. The normalised residuals are shown for the acoustic estimates and the Pearson residuals are shown for the age frequency. Where informed priors were used they are plotted on the same graph as the marginal posterior distribution.

The posterior of the acoustic q is shifted to the left of the prior (Figure E1) showing that the model had some “information” with which to update the prior. Likewise, the model found some “information” in the age frequency with which to slightly modify the informed prior producing a slightly lower estimate than the mean of the prior (0.045 which is used in New Zealand) (Figure E2). “Information” is a tricky concept as the strength of the signals from data are dependent on model assumptions. The information is real in the context of the model but may not be fully valid in terms of the “real world”.

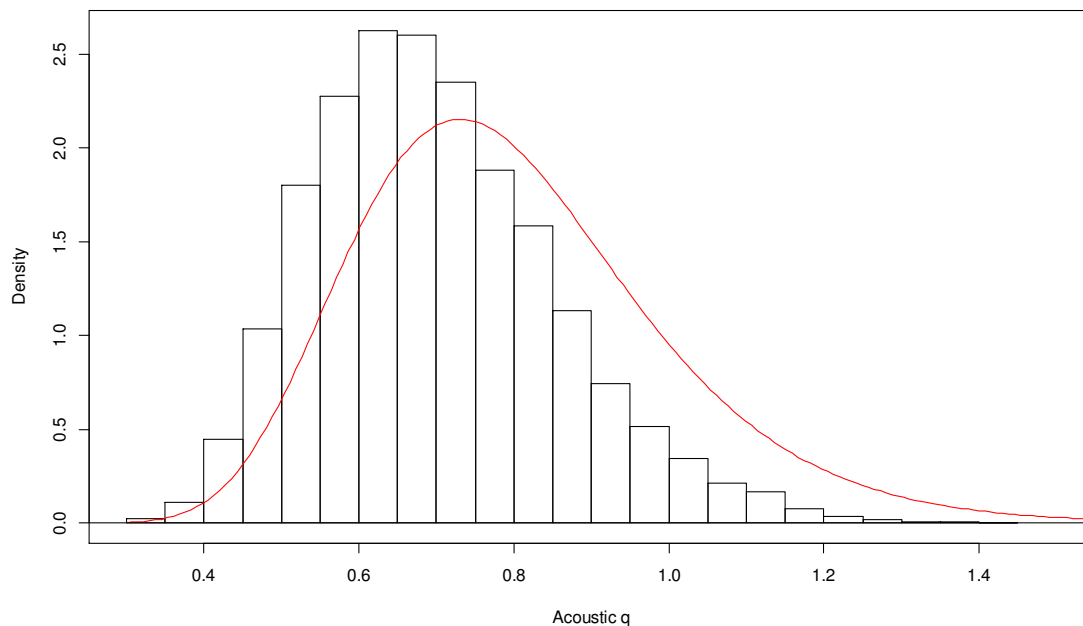


Figure E1: Base model: marginal posterior distribution of the acoustic q (histogram) with the informed prior (red line).

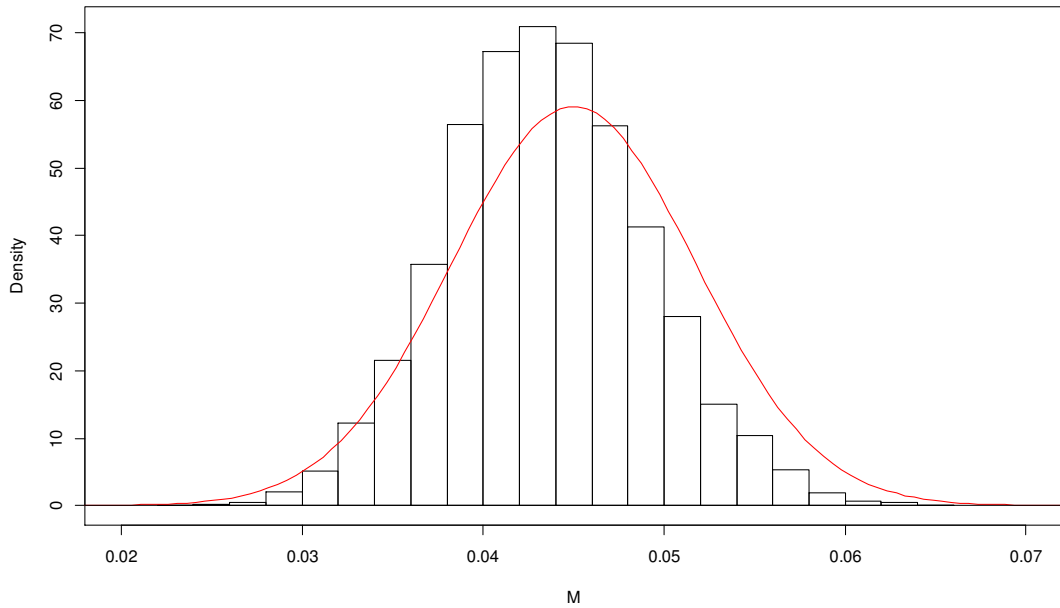


Figure E2: Base model: marginal posterior distribution of M (histogram) with the informed prior (red line).

The age frequency also provided information with which to update the informed prior on the age at 50% maturity (a_{50}) (Figure E3). The prior on the second maturity parameter (a_{1095}) was very weak and the posterior had somewhat less variance (Figure E4).

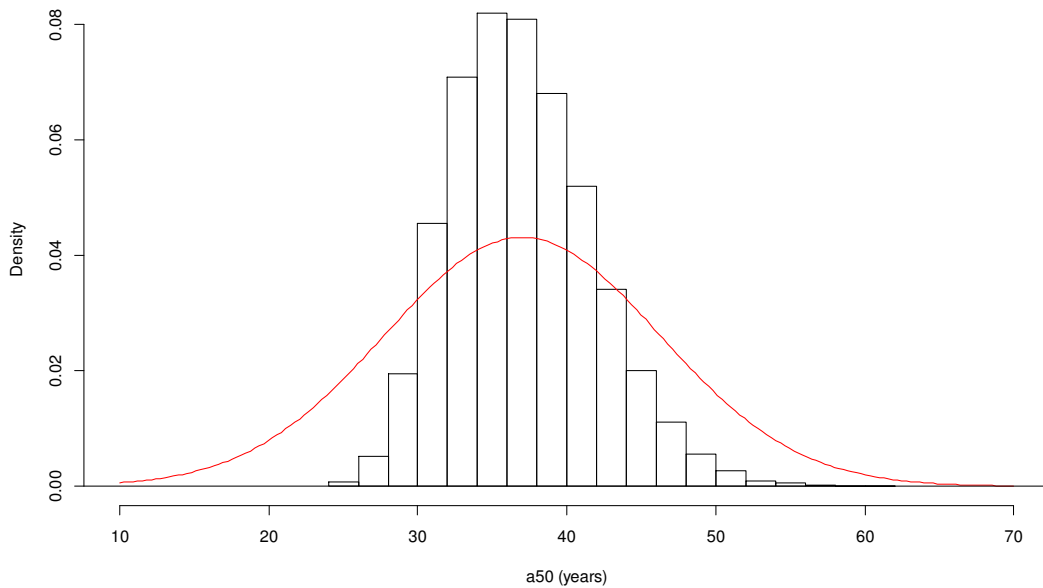


Figure E3: Base model: marginal posterior distribution of a_{50} (histogram) with the informed prior (red line).

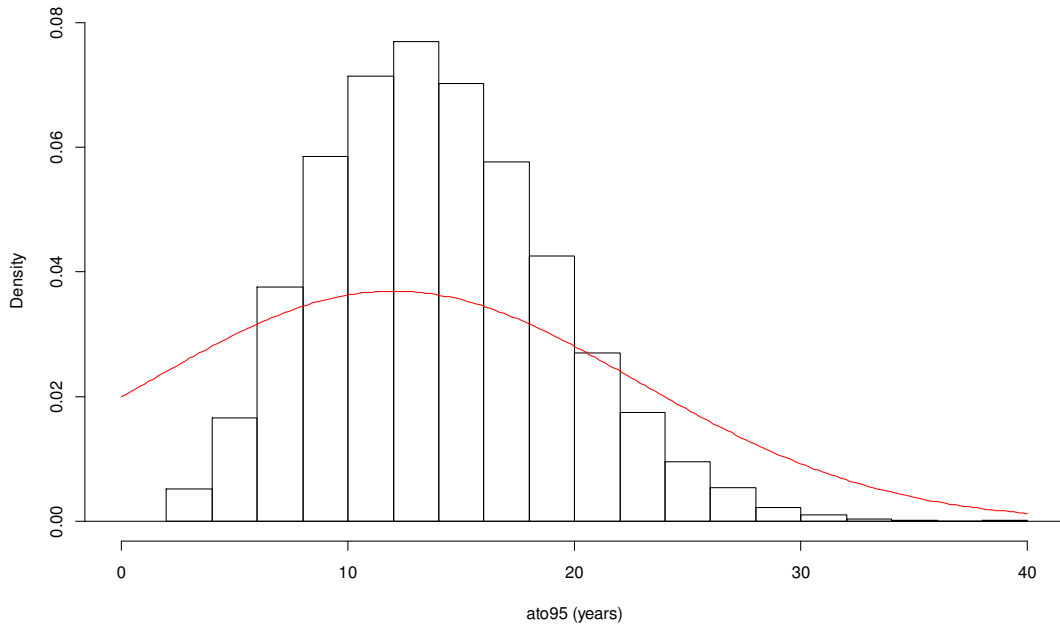


Figure E4: Base model: marginal posterior distribution of a_{t095} (histogram) with the informed prior (red line).

There was an informed prior on the proportion of the mature biomass that migrated to the catch-all region Other. For the base model the mean of the Normal prior was set at 20% (with a CV of 10%) and the model (quite rightly) did not find any information with which to update the prior (Figure E5).

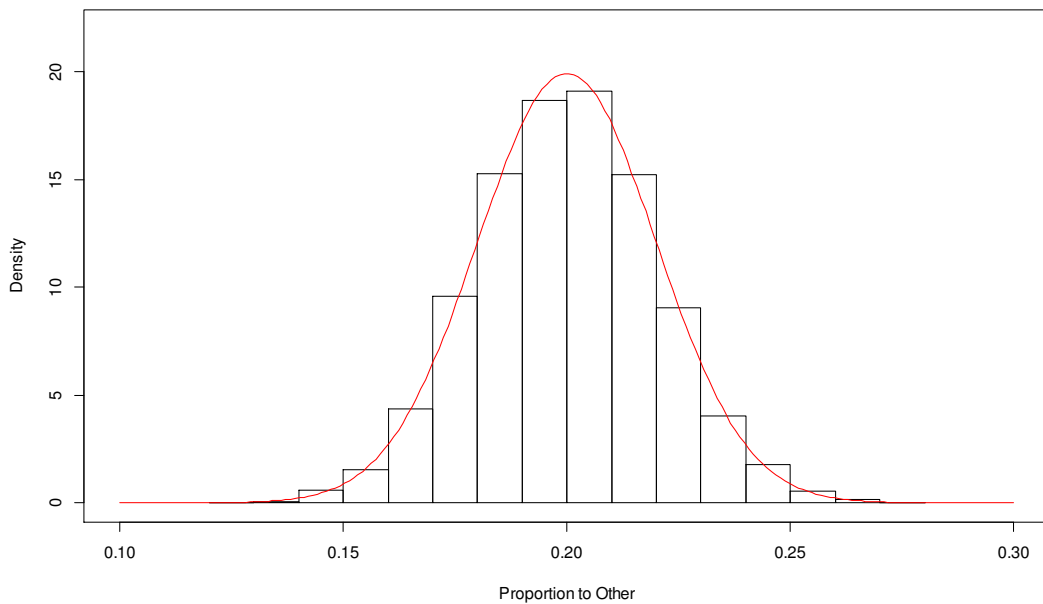


Figure E5: Base model: marginal posterior distribution of the proportion migrating to Other (histogram) with the informed prior (red line).

The other migration parameters had uniform priors and were estimated fairly precisely within the model (Figure E6).

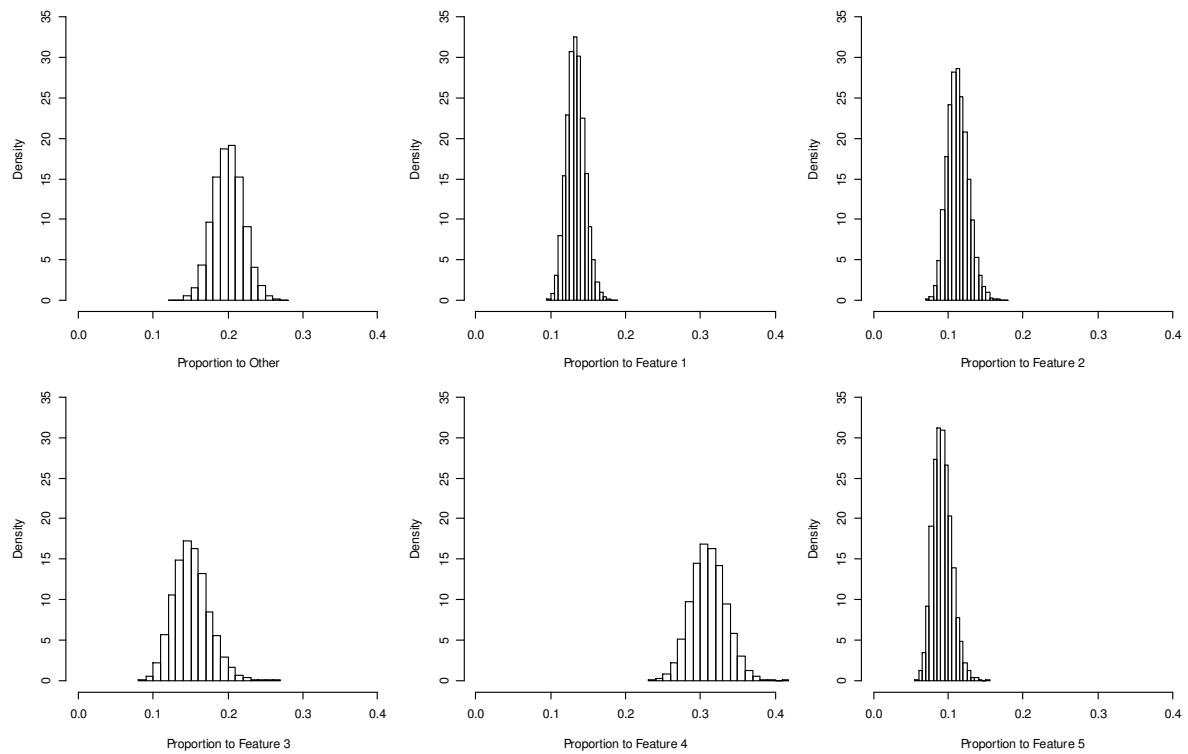


Figure E6: Base model: marginal posterior distributions of the proportion migrating to Other and each numbered feature. The same scale is maintained on both axes across all plots.

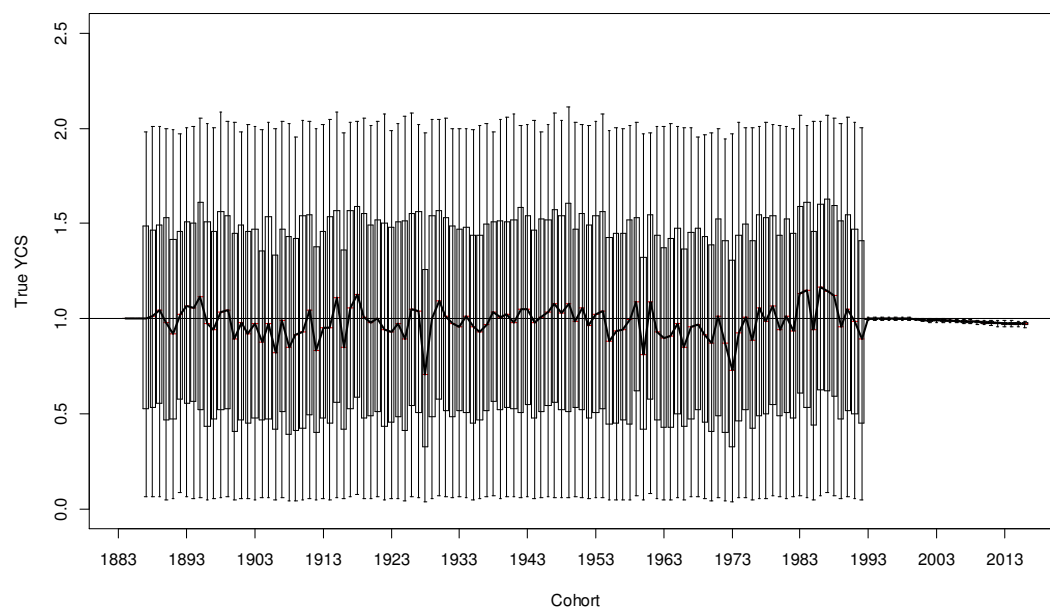


Figure E7: Base model: box and whiskers plot of the true YCS estimates (R_t/R_0). The box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.

The estimated year class strengths (YCS) show almost no periods of strong or weak recruitment (Figure E7). Given the assumed effective sample size of 40, the age frequency is consistent with deterministic recruitment (as well as many other recruitment patterns that are not too dissimilar).

The marginal posterior distribution of B_0 ranges from 20 000 t to just above 80 000 t with a mode at about 40 000 t (Figure E8).

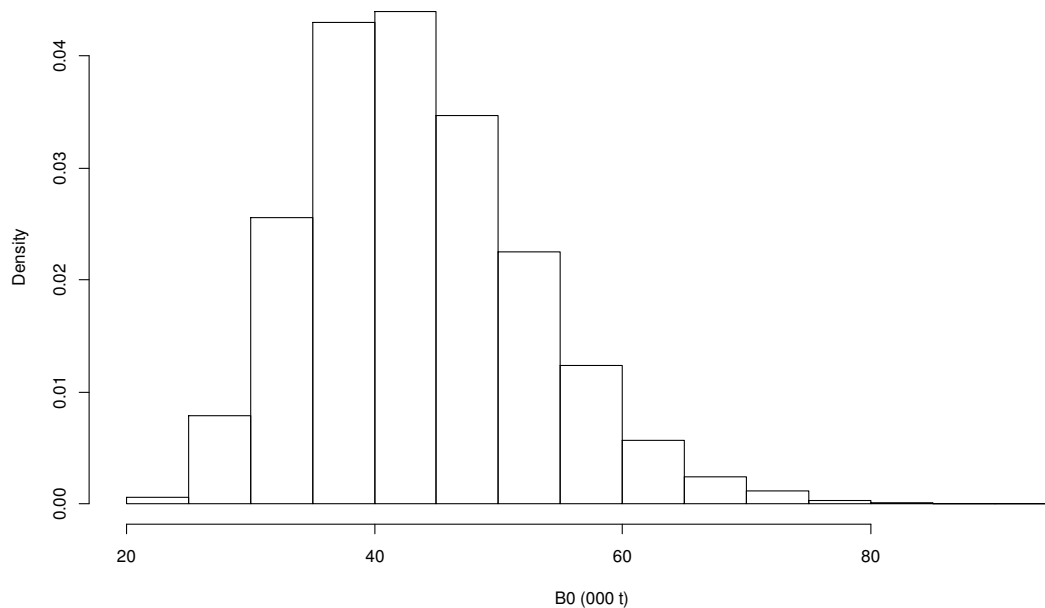


Figure E8: Base model: marginal posterior distribution of B_0 .

The MCMC “fit” to the acoustic biomass estimates shows no problems. The key point is that there is overlap between the confidence intervals of the observations and the credibility intervals of the MCMC predicted values (Figure E9). Both the confidence interval and the credibility interval are attempts at enclosing the true value. If they did not overlap there would be cause for concern. The fact that the observation is not enclosed by the credibility interval (in some cases) is not a cause for concern (the observation contains observation error due to the nature of the survey; the credibility interval is a product of the model and all of the data in the model – not just the single observation).

The normalised residuals for the acoustic data are primarily within 2 standard deviations of zero which is consistent with the CVs of the acoustic biomass estimates and the model assumptions (Figure E10).

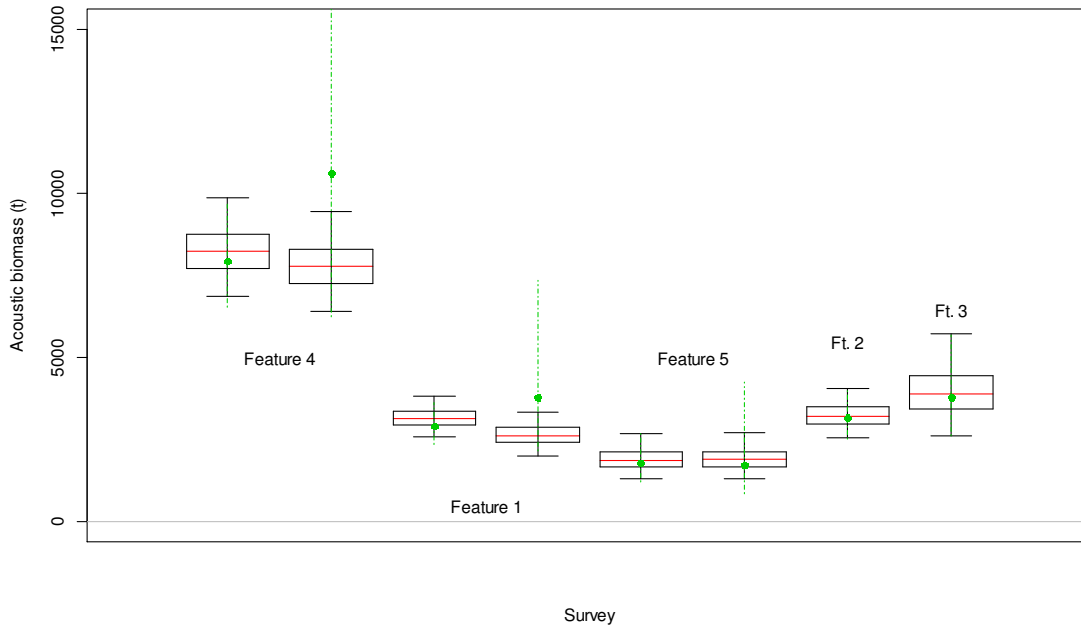


Figure E9: Base model: box and whiskers plots of the fits to the acoustic biomass estimates for Feature 1 (two estimates), Features 2 and 3 (single estimates) and Features 4 and 5 (two estimates each). The observation is the solid green point and the dashed line is the 95% confidence interval. The box covers the middle 50% of the distribution and the whiskers extend to a 95% CI (credibility interval).

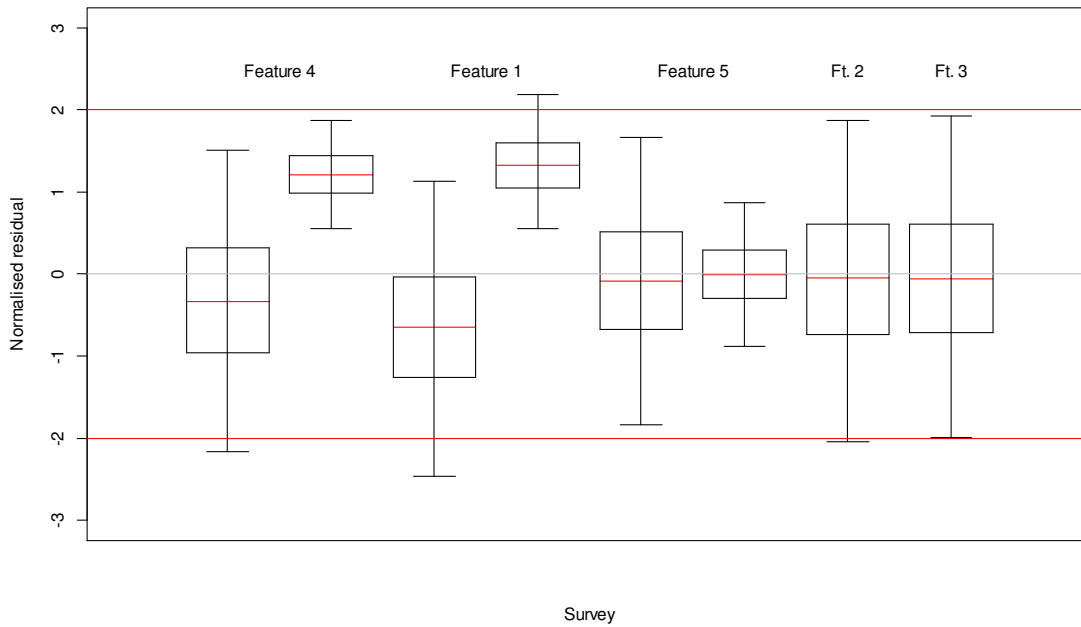


Figure E10: Base model: box and whiskers plots of the normalised residuals for the acoustic biomass estimates. The box covers the middle 50% of the distribution and the whiskers extend to a 95% CI. Horizontal lines are marked at -2 and +2 between which one would expect most of the residuals to lie if the CVs of the acoustic estimates are appropriate for the model.

The MCMC “fit” to the age frequency is adequate with a noticeable deviation from the observed value in the plus group (Figure E11). However, it is actually only a small residual and the other residuals are mainly contained within -2 to +2 suggesting that the effective sample size is appropriate (Figure E12).

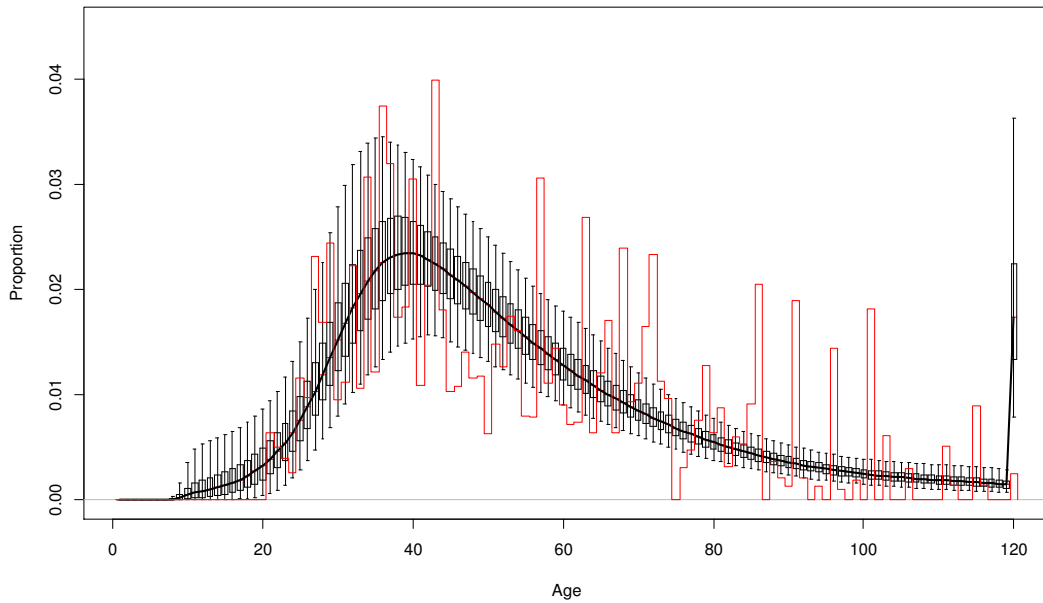


Figure E11: Base model: box and whiskers plots of the MCMC fits (black) for the age frequency (red). Each box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.

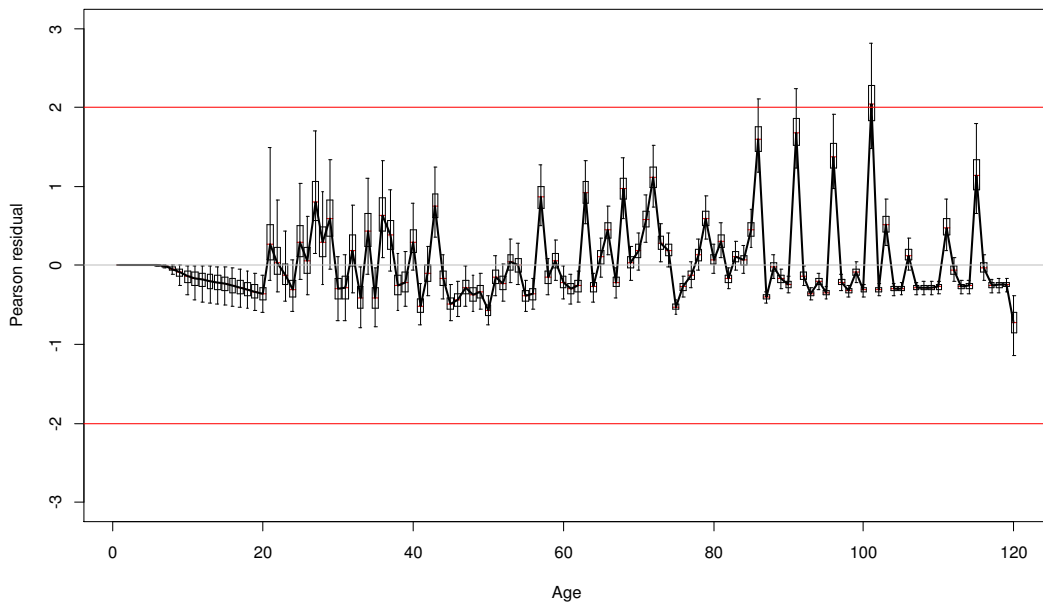


Figure E12: Base model: box and whiskers plots of the MCMC Pearson residuals for the age frequency. The box covers the middle 50% of the distribution and the whiskers extend to a 95% CI.