Ministry for Primary Industries
Manatū Ahu Matua

Rationale for year-class strength priors

New Zealand Fisheries Assessment Report 2015/62

R. I. C. C. Francis
D. Fu

ISSN 1179-5352 (online)
ISBN 978-1-77665-072-9 (online)

October 2015

Requests for further copies should be directed to:

Publications Logistics Officer
Ministry for Primary Industries
PO Box 2526
WELLINGTON 6140

Email: brand@mpi.govt.nz
Telephone: 0800 00 83 33
Facsimile: 04-894 0300

This publication is also available on the Ministry for Primary Industries websites at:
http://www.mpi.govt.nz/news-resources/publications.aspx
http://fs.fish.govt.nz go to Document library/Research reports

© Crown Copyright - Ministry for Primary Industries.

Contents

EXECUTIVE SUMMARY

Francis, R.I.C.C.; Fu, D. (2015). A rationale for year-class strength priors.

***New Zealand Fisheries Assessment Report 2015/62*. 23 p.**

In Bayesian stock assessment models, information beyond that contained in quantitative observations – of the fishery or the fish population – is presented in the form of prior distributions for the model parameters. Here we focus on the prior distribution for a specific type of parameter: the year-class strengths (YCSs). Until recently, all New Zealand assessments used a lognormal prior for YCSs, but a uniform (or near-uniform) prior was advocated for a 2013 orange roughy assessment and was subsequently used in other assessments. This raised the question, addressed in this report, as to what rationale should be used for the choice of a prior for YCSs.

We evaluated the information available for developing YCS priors. A literature review found theoretical support for a lognormal prior, and near unanimous use of this prior in both stock-recruitment studies and stock assessments. A collection of 157 stock-recruitment data sets was tested to see whether the associated YCS vectors were better described by a lognormal or uniform distribution. Less than 1% (3/390) of test results favoured the uniform distribution. This is less than the expected percentage of incorrect results (1.8%) if the true distributions were all lognormal.

We propose the following rationale for the choice of YCS prior. The default prior should be lognormal, but there are two situations in which an alternative distribution could properly be used. The first is where there is evidence that YCSs for the stock being assessed may have a distribution clearly different from the lognormal. The second is where its use is found necessary to obtain an acceptable assessment. To distinguish these two situations we say that the former results in a *genuine* prior and the latter in a *tactical* prior.

We reviewed five recent New Zealand assessment reports where a uniform prior was used and found that this prior was never justified according to our rationale. In particular, the differences between the base and alternative runs in the 2013 orange roughy assessment were shown to be caused more by different ways of parameterising the YCSs than the use of a uniform or lognormal prior.

As final asides, we offer some cautions about both the Francis and Haist YCS parameterisations in CASAL and the acceptability of recent orange roughy assessments.

1. INTRODUCTION

The Bayesian approach to statistics, which is often used in stock assessment models today (McAllister & Kirkwood 1998), allows information to be included in analyses in two distinct ways. As well as the usual quantitative observations – of the fishery or the fish population – it is possible to include other information (that may be more subjective or qualitative) about specific model parameters. This latter information is presented in what is known as a *prior* distribution for the parameter(s). The idea is that a prior can, and should, encapsulate all knowledge about a parameter beyond that which is contained in the quantitative observations. An outcome of the Bayesian approach is that, for each model parameter, the two sorts of information – from the observations and the prior – are combined in what is called a posterior distribution for the parameter.

This report focuses on how one should select a prior distribution for a specific type of parameter: the year-class strengths (YCSs). The YCS for each year indicates the extent to which the recruitment in that year is estimated to be above (YCS greater than 1) or below (YCS less than 1) average.

By far the most common prior distribution for YCS parameters is the lognormal (see below). This prior implies that recruitment is typically near average in most years, and very different from average in relatively few years. In some recent orange roughy assessments it has been suggested that it may sometimes be better to use a uniform (or near-uniform) prior for YCSs (Cordue 2014ab). This prior is very different from the lognormal, implying that all possible values of YCSs are equally (or nearly-equally) likely.

This report addresses the sole objective of Ministry for Primary Industries (MPI) project SCM201406: “Develop a rationale for the choice of a prior for year-class strength”. This choice of prior has been the subject of some discussion in New Zealand stock assessment Working Groups in the past two years. Though it has been shown that the choice can have a noticeable effect on the assessment results, no consensus has arisen as to the appropriate way to make the choice.

Our approach to the problem was tripartite. We first reviewed current knowledge about the statistical distribution of YCSs; then used this knowledge to develop a rationale for choice of YCS priors; and finally evaluated recent New Zealand stock assessments against this rationale.

2. REVIEW OF CURRENT KNOWLEDGE

2.1 Literature review

A multiplicative lognormal error has been very widely used in studies of stock-recruitment models (Allen 1973; Walters & Hilborn 1976; Peterman 1981; Walters & Ludwig 1981; Ludwig & Walters 1981; Archibald et al. 1983; Huang & Walters 1983; Hightower & Grossman 1985; Quinn & Deriso 1999; Haddon 2000; Myers 2001). In such studies Hilborn & Walters (1992) recommended that a lognormal distribution should be the starting assumption unless the evidence indicates otherwise for specific species. A common mathematical form in these studies is

$$R_y = \mathrm{SRR}(S_{y-a})\exp(\varepsilon_y)$$

which relates R_y, the recruitment in year y, to S_{y-a}, the spawning biomass a years earlier, via the mean stock-recruit relationship, SRR, and a normally distributed stochastic term ε_y. The assumption of a lognormal distribution has been supported by empirical evidence (Allen 1973), as well as biological realism (Quinn & Deriso 1999). The lognormal distribution was found to be better than the normal distribution in describing both the frequency distribution of observed year-class strengths for 18 fish stocks around the world (Hennemuth et al. 1980) and survival rates of three species of salmon (Peterman 1981). A theoretical justification for the use of this error distribution is that survival from spawning to recruitment can be considered as the combined effect of a series of independent environmental factors that affect mortality during early life stages (Walters & Hilborn 1976), so that ε_y, being the sum of multiple independent terms, should be approximately normally distributed according to the Central Limit Theorem (Stuart & Ord 1987). This interpretation of the lognormal error as arising from a combination of multiple environmental effects is biologically appealing as it implies that the recruitment is occasionally very large when all environmental conditions are favourable, and it also implies that the variance of recruitment will increase as stock size increases (Hightower & Grossman 1985).

In modern stock assessment programs, recruitment variation is very commonly assumed to be lognormal (Funk et al. 1998; Maunder & Starr 2001; Maunder & Deriso 2003; Maunder 2003, 2005; Breen et al. 2003; Haist et al. 2009; Maunder & Punt 2013; Punt et al. 2013; Deroba et al. 2015). However, it is not appropriate to ask what prior is used for YCSs in these programs because most of them structure their recruitment calculations in terms of log-space recruitment deviations (the above ε_y), rather than YCSs (which are equivalent to $\exp(\varepsilon_y)$). What these programs have in common is a component of the objective function (which we will call the *recruitment term*) that implies an assumed statistical distribution for recruitment. In CASAL the recruitment term is described as a prior on the YCSs, but different descriptors are used in other programs (e.g., it may be called a penalty, a constraint, process error, a recruitment regularity term, etc). In a survey of general-purpose stock assessment programs we found that all allowed a lognormal distribution for recruitment and, with the exception of CASAL, the only alternative was a log-uniform distribution (Table 1). CASAL allows various priors for the YCSs (including lognormal, normal, uniform, and log-uniform) simply because it aims to give the user a range of priors for all estimable parameters. We note in passing that some programs include additional recruitment structure such as auto-correlation (e.g., Chen et al. 2005; Craig 2012; Kleiber et al. 2013) and seasonal or environmental variation (Maunder & Watters 2003).

Table 1: Recruitment distributions allowed in some general-purpose stock assessment programs. LN, lognormal; LU, log-uniform (i.e., uniform in log space).

Program	Reference	Recruitment distribution(s)
AMAK	https://github.com/NMFS-toolbox/AMAK	LN
ASAP	Legault & Restrepo (1999); http://nft.nefsc.noaa.gov/ASAP.html	LN, LU[1]
A-SCALA	Maunder & Watters (2003)	LN
BAM	Craig (2012)	LN, LU[1]
CASAL	Bull et al. (2012)	Various
CASA	Sullivan et al. 1990	LN
Coleraine	Hilborn et al. (2003)	LN
iSCAM	Martell (2011)	LN
Multifan-CL	Fournier et al. (1998); Kleiber et al. (2013)	LN, LU[1]
SAM	Dickey-Collas et al. (2015)	LN
Stock Synthesis	Methot (2012); Methot & Wetzel (2013)	LN

[1]The LU distribution is invoked by setting a weighting term to zero and thus removing the recruitment term from the objective function (a step recommended against in the ASAP manual)

Two papers investigated the effect on estimation of omitting the recruitment term from the objective function. This omission is equivalent to assuming a uniform prior for whatever free parameters are used in the estimation of recruitment. Maunder & Deriso (2003) omitted the recruitment term (which they called a penalty – see their Eq. (2)) in a model in which the relevant free parameters were the R_y (rather than the ε_y or the YCS), so this implied that recruitment had a uniform distribution. They found that this model performed worse than other models. However, from the perspective of the present paper their results are inconclusive because they didn't consider models that were exactly the same as their preferred models (which they called Importance Sampling and Bayesian Integration) except for the lack of the recruitment term. For some of their simulations they generated recruitment deviations from a uniform distribution, but did so only as an "extreme sensitivity analysis". Dickey-Collas et al. (2015)

omitted the recruitment term in a model using the ε_y as free parameters, which implied that recruitment had a log-uniform distribution. They made the reasonable points that (i) conclusions from meta-analyses using recruitment estimates from diverse stock assessment programs are compromised by the fact that different programs make different assumptions about recruitment; and (ii) the "most appropriate way to parameterise recruitment depends on the final goal" (e.g., stock assessment or meta-analysis). Their suggestion that, for use in meta-analyses (or in studying specific cohorts), "recruitment time series free of distributional and parametric assumptions" would be preferable is debatable (it seems in conflict with the philosophy of Bayesian statistics to withhold information about the distribution of recruitments; and also, since dropping the recruitment term has a different effect depending on how recruitment is parameterised, the concept "distribution-free" is not well-defined in this context). However, from the perspective of the present paper the point to note is that there was no suggestion that the recruitment term be dropped in stock assessments.

2.2 Analysis of stock-recruitment data

We analysed data from the RAM legacy Stock Assessment database (Ricard et al. 2012) to see what support there was for uniformly distributed YCSs. This database is a compilation of stock assessment results for commercially exploited marine populations from around the world, covering over 300 fish and invertebrate stocks. It is intended to supersede the earlier Myers stock-recruitment database (Myers et al. 1995), which is no longer being updated.

The analysis used time series of spawning stock biomass (SSB) and recruitment derived from four data categories: integrated analysis, statistical catch at age model, survey index, and VPA. For those based on integrated analysis and catch at age methods, recruitments were most likely to have been generated from some form of stock-recruitment relationship (details are not available in the database). For these two categories, the SSB-recruitment series were restricted to the period in which recruitments were estimated from the model, where this information was provided (available for 51 time series); otherwise the time series were inspected visually, and if part or all of the series were clearly calculated from a stock-recruitment relationship, the series were not used (26 series were excluded). All the SSB-recruitment series from VPA-based assessments were considered. Further restrictions were imposed such that each series needed to have at least 15 years of data. In total 157 time series covering 7 taxonomic orders were analysed (not including the order Beryciformes, which includes orange roughy), with the longest time series having length 96 (Table 2). For our analyses, the categories "integrated analysis" and "statistical catch at age" were combined (and labelled catch at age); also the few "survey index" data sets were combined with the "VPA" category.

Table 2: Numbers of SSB-recruitment time series analysis by data category and taxonomic order. The number in parentheses is the maximum length of time series in each category.

Order	Integrated Analysis	Statistical catch at age	Survey index	VPA
Clupeiformes	1 (27)	12 (54)		13 (42)
Gadiformes	6 (32)	5 (49)	2 (22)	33 (77)
Ophidiiformes	5 (40)			
Perciformes	9 (95)	17 (79)		6 (75)
Pleuronectiformes	1 (96)	12 (53)		23 (49)
Scorpaeniformes	2 (39)	9 (47)		
Osmeriformes			1 (28)	
Total	24	55	3	75

In order to calculate YCSs we needed to fit some sort of stock-recruitment relationship (SRR) to each data set. We fitted two standard parametric SRRs – Beverton-Holt (Eq. 1) and Ricker (Eq. 2), fitted by maximum likelihood assuming lognormal residuals – and one non-parametric SRR, a smooth curve fitted using R function *lowess* (with default parameter settings). Although some of the fitted lowess SRRs may be biologically implausible (see Appendix A), they have the merit of requiring no assumptions about the distribution of the residuals. Note that it is not possible to fit SRRs assuming uniform residuals.

$$R = \frac{S}{\alpha + \beta S} \qquad (1)$$

$$R = \alpha S e^{-\beta S} \qquad (2)$$

When fitting the parametric SRRs, starting parameter values (α, β) were generated using three different values of steepness parameter (h): 0.2, 0.5, and 0.9, to maximise the chance to obtain a global maximum for the likelihood. For Beverton-Holt,

$$\alpha = \frac{B_0(1-h)}{R_0(4h)}; \qquad \beta = \frac{(5h-1)}{4hR_0} \qquad (3)$$

where B_0 and R_0 are the SSB and recruitment in an unfished population with deterministic recruitment (approximated by the maximum SSB and recruitment for each series). For Ricker,

$$\alpha = \frac{R_0}{B_0}(5h)^{5/4}; \qquad \beta = \frac{5\log(5h)}{4B_0} \qquad (4)$$

Estimates of Beverton-Holt parameters were made for 97 SSB-recruitment series for which convergence was attained and estimated α and β were both positive. Estimates for the rest of the series (29 from catch at age and 30 from VPA) appeared implausible and were therefore discarded. The reason that the Beverton-Holt relationship was not able to be fitted to a large number of those time series is that this SRR assumes that the expected recruitment increases monotonically with SSB, which was clearly inconsistent with some data series (e.g. series 59 and 98 in Figure A1). Estimates of Ricker parameters were made for 137 series (7 from Catch at age and 11 from VPA were discarded). Lowess lines were fitted to all time series.

For each fitted SRR, a vector of YCSs was calculated as the observed recruitment divided by that predicted by the SRR. A statistical test was then performed on each of these vectors to determine whether its distribution was more likely to be lognormal or uniform. The test statistic was ΔAIC, the difference between the AIC values (Akaike 1974) obtained from fitting the YCSs to lognormal and uniform distributions (R code for this test is given in Appendix B). In all but three of 390 tests the YCS vector was found more likely to be lognormal (Table 3). The test results did not seem to be much affected by data source (catch at age or VPA), SRR (Beverton-Holt, Ricker, or lowess), or taxonomic order (Figure 1).

We also calculated, by simulation, the probability that a test result would be 'uniform' if the YCS vector was lognormally distributed. This probability is strongly influenced by the length of YCS vector, and to a lesser extent by sigma (the standard deviation of log(YCS)) (Figure 2). Using the estimated probabilities for a range of sample sizes (assuming a sigma of 0.5, which is typical for these data – see Table 3), we calculated the expected number of 'uniform' test results and found that this was slightly higher than the actual number (Table 3).

We conclude that these data provide no support for uniformly distributed YCSs.

Table 3: Summary statistics for the YCS vectors (number of vectors, median number of years per vector, and median log-space s.d. ['sigma']; values in parentheses are ranges) and test results (actual number of vectors found more likely to be uniform, and the corresponding expected number if the vectors are actually lognormally distributed with sigma = 0.5).

	Catch at age			VPA		
	BH	Ricker	Lowess	BH	Ricker	Lowess
No. of vectors	49	71	78	48	66	78
No. of years	35 (24–96)	32 (16–96)	32 (16–96)	32 (15–77)	32 (15–75)	32 (15–75)
Sigma	0.54 (0.14–0.99)	0.49 (0.15–1.31)	0.47 (0.13–0.90)	0.54 (0.19–1.37)	0.52 (0.11–1.16)	0.50 (0.11–0.90)
Actual no. uniform	1	1	0	0	0	1
Expected no. uniform	0.7	1.2	1.3	0.9	1.4	1.6

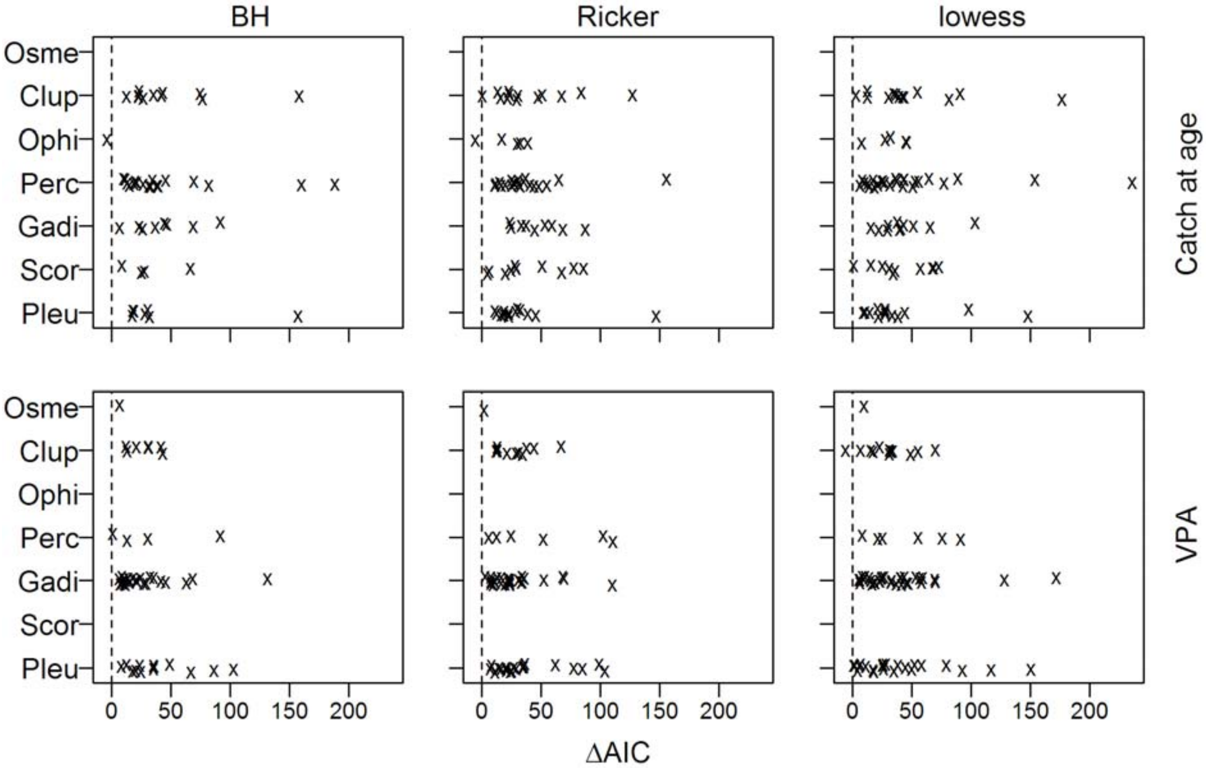

Figure 1: Values of the statistic, ΔAIC, of the test performed to determine whether each YCS vector is better described by a uniform (ΔAIC < 0) or lognormal (ΔAIC > 0) distribution. The results are organised by source data category (catch at age in upper panels, VPA in lower panels), fitted SRR (Beverton-Holt, Ricker, or lowess, in left, middle, and right panels, respectively) and taxonomic order (*y*-axis).

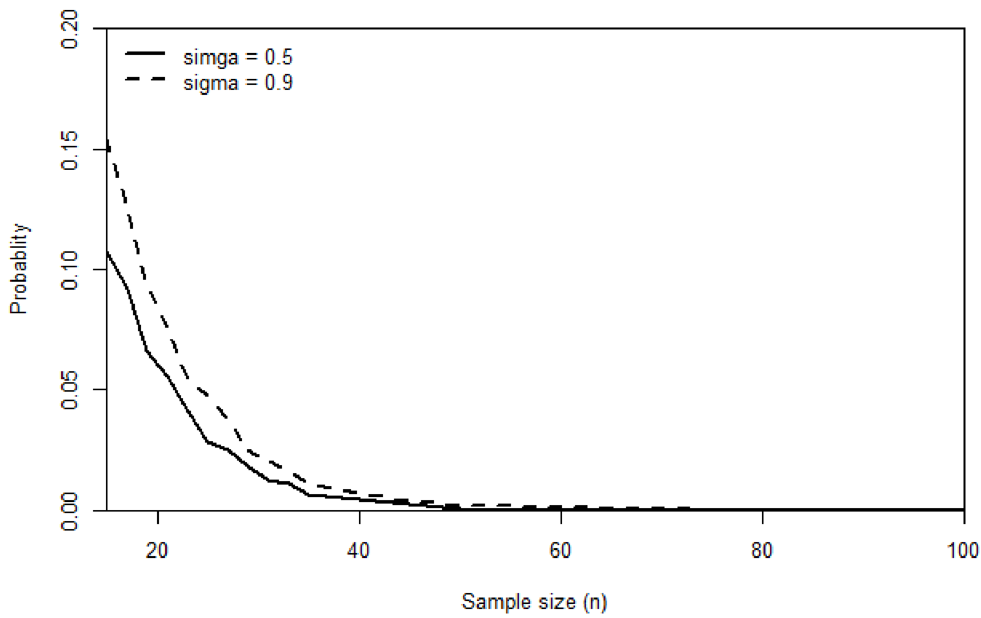

Figure 2: Estimated probability, as a function of sample size, of finding that a YCS vector is more likely to be uniformly distributed when it is actually lognormal. The probability is given for two values of sigma [s.d.(log(YCS))]: 0.5 and 0.9.

3. RATIONALE FOR CHOICE OF PRIOR

As noted above, priors provide a way to present information to a Bayesian model – beyond that contained in the data to which the model will be fitted – about the values of its parameters. For YCS parameters in stock assessment models we have shown that a lognormal prior has both theoretical support and very wide acceptance amongst fishery scientists. Thus, it is reasonable to suggest that the *default prior* for YCSs should be lognormal.

There are two distinct situations in which an alternative to this default could be justified. The first, and most obvious, is where there is evidence that YCSs for the stock being assessed may have a distribution clearly different from the lognormal. We will call this a *genuine* alternative prior to distinguish it from a *tactical* prior, which is one that we know to be false, but which we use in order to achieve an acceptable assessment. The reason that a tactical prior may sometimes be justified for stock assessments is that assessments are always badly underdetermined. That is, there is never enough information to unequivocally determine stock status. In order to be able to provide advice to fishery managers scientists deal with this problem by deliberately misspecifying the model, adding many assumptions that we know to be false (e.g., ignoring spatial structure; making natural mortality independent of time and age) but which allow us to estimate stock status and produce an acceptable assessment. Such misspecification is believed to be the reason why some models produce clearly implausible estimates of parameters such as natural mortality (Lee et al. 2011). It can also sometimes make it difficult for the model to fit biomass data, and in such cases it may be justifiable to take steps to force the model to fit, e.g., by assigning extra weight to these data (Francis 2011) or, where appropriate, using a tactical YCS prior.

Whether a tactical prior is justifiable in a specific stock assessment is, to a great extent, a matter of subjective judgement, in much the same way that many data-weighting decisions are inescapably subjective (Francis 2011). This is a consequence of the inevitability of model misspecification. However, we propose that there should be a *quid pro quo* for a tactical prior. That is to say, the negative effect of deliberately misleading the model with a false YCS prior should be balanced by some clear gain to the assessment. It is difficult to specify exactly what sort of gain is required but we suggest that a criterion could be that the assessment is acceptable with the tactical prior, but unacceptable without it. A good example (which is relevant to the first assessment discussed in the next section) would be if a good fit to a key biomass index was achievable only with the tactical prior. What would not be adequate justification of the tactical use of a uniform YCS prior would be an improvement in fit to age composition data, because this improvement is always expected when the default YCS prior, which is informative, is replaced by an uninformative prior.

4. EVALUATION OF RECENT ASSESSMENTS

In examining recent New Zealand assessments our aim was to evaluate both the justification given for the use of a (near) uniform prior for YCSs, and what differences this use made to these assessments. The first use was in 2013 for the MEC orange roughy stock (Cordue 2014a); in 2014 it was again used for four orange roughy (ORH) stocks (Cordue 2014b), as well as the OEO 4 smooth oreo stock (Fu & Doonan 2015) and hoki (HOK) (McKenzie 2015). As well as examining each published assessment we considered documents from the relevant meetings of the appropriate Fisheries Science Working Group (WG), because these sometimes provide detail not included in the published reports.

4.1 The 2013 MEC orange roughy assessment

Three different treatments of YCSs were considered in the 2013 orange roughy assessment (Cordue 2014a), and these differed in their ways of parameterising YCSs and/or in the prior used (Table 4). [A fourth potential treatment, Francis-uniform, was mentioned, but not tried, apparently on the erroneous grounds that it was "not possible" because of "a confounding between the Y_i and R_{mean}" Cordue (2014a, p. 14). In fact this apparent confounding is resolved by the prior on R_{mean}.]

Two decisions made about these alternative treatments are important for the present study. The first was to drop the Haist-lognormal treatment after preliminary runs. This decision was based on MPD results only, and was justified on the grounds that the main contrast amongst preliminary runs with the three treatments was between the other two treatments (Haist-uniform and Francis-lognormal). Further, "The Haist-lognormal runs were fairly similar to the Haist-uniform runs." (Cordue 2014a, p. 3). These are common and standard reasons for choosing which of many preliminary runs should be included amongst the final runs for an assessment. However, as we shall show shortly, the decision to drop the Haist-lognormal treatment had unfortunate consequences in limiting our ability to interpret results from this assessment.

Table 4: Five alternative treatments of YCSs in CASAL and the type of runs they were used for in recent stock assessments. '-', not used

			Type of run			
Label	Parameterisation[1]	Prior	ORH 2013	ORH 2014	OEO 2014	HOK 2014
Haist-uniform	Haist	Uniform	base	-	-	sensitivity
Francis-lognormal	Francis	Lognormal	alternative	-	-	sensitivity
Haist-lognormal	Haist	Lognormal	preliminary	-	alternative	base
Francis-uniform	Francis	Uniform	-	-	-	-
Haist-near-uniform	Haist	Nearly uniform[2]	-	base	base	-

[1]See section 5.4.2 in Bull et al. (2012)
[2]A very diffuse lognormal with mode 1 and log-space standard deviation 4

The second important decision was to use the Haist-uniform treatment in the base model and Francis-lognormal as an alternative. This was done on the basis that the former was "able to fit the data much better than the Francis [-lognormal] model. In particular, it had a better fit to the downward trend in the trawl survey biomass indices." (Cordue 2014a, p. 1). This better fit was very pronounced in the MPD runs (Figure 3), but was much less apparent in the more important MCMC runs (Figure **4**); indeed the notes from the 29 April 2013 WG meeting concluded that "Haist fits better at MPD ... At MCMC there is little to choose between them." (Tingley 2013).

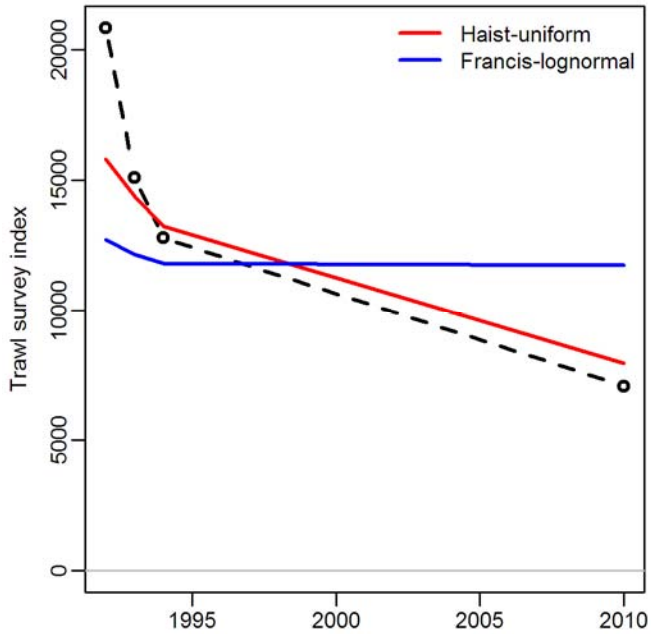

Figure 3: Observed (black points and lines) and predicted (coloured lines) trawl survey biomass from two MPD runs in the 2013 orange roughy MEC assessment. (This is a replotting of the right hand panel of figure 8 of Cordue 2014a).

Cordue (2014a) seemed to assume that the superiority in goodness-of-fit of the base model compared to the alternative was caused by the difference in priors: "The purpose of the uniform prior is to allow the model extra freedom in estimating YCS to fit the available data" (Cordue 2014a, p. 9; repeated at p.10). We could find no support for this assumption. There seemed no reason to believe that any difference between these runs should be due to the difference in prior (uniform vs lognormal) rather than that in parameterisation (Haist versus Francis). Indeed the opposite was suggested by the earlier statement, concerning the preliminary MPD runs, that "The Haist-lognormal runs were fairly similar to the Haist-uniform runs." (Cordue 2014a, p. 3). This statement is certainly consistent with our own evaluation of these preliminary runs (Error! Reference source not found.). It is also consistent with some results from the 2014 hoki assessment: in comparing posterior profiles on B_0 from the same three YCS treatments McKenzie (2015, p. 29) reported "Moving from the Francis to the Haist parameterisation substantially reduces the influence of the priors in the posterior profile, but going from lognormal priors to uniform priors makes little change (under the Haist parameterisation)." However, what is more important is to make the comparison on MCMC runs. To that end we did a Haist-lognormal MCMC run for this assessment (see Appendix C for details) and found that its fit to the trawl survey biomass indices was more similar to that for the Haist-uniform run than to that for the Francis-lognormal run (compare Figure 4 and
Figure 5). This supports the view that the important difference between the published Haist-uniform and Francis-lognormal runs was the YCS parameterisation, rather than the prior.

Table 5: Results of 21 visual comparisons of outputs from three preliminary runs in the 2013 MEC orange roughy assessment to see whether those from the Haist-lognormal run were more similar to those from the Haist-uniform or Francis-lognormal run. Comparisons were made from results presented in Cordue (2013).

More similar to Haist-uniform	More similar to Francis-lognormal	Not similar to either
11	2	8

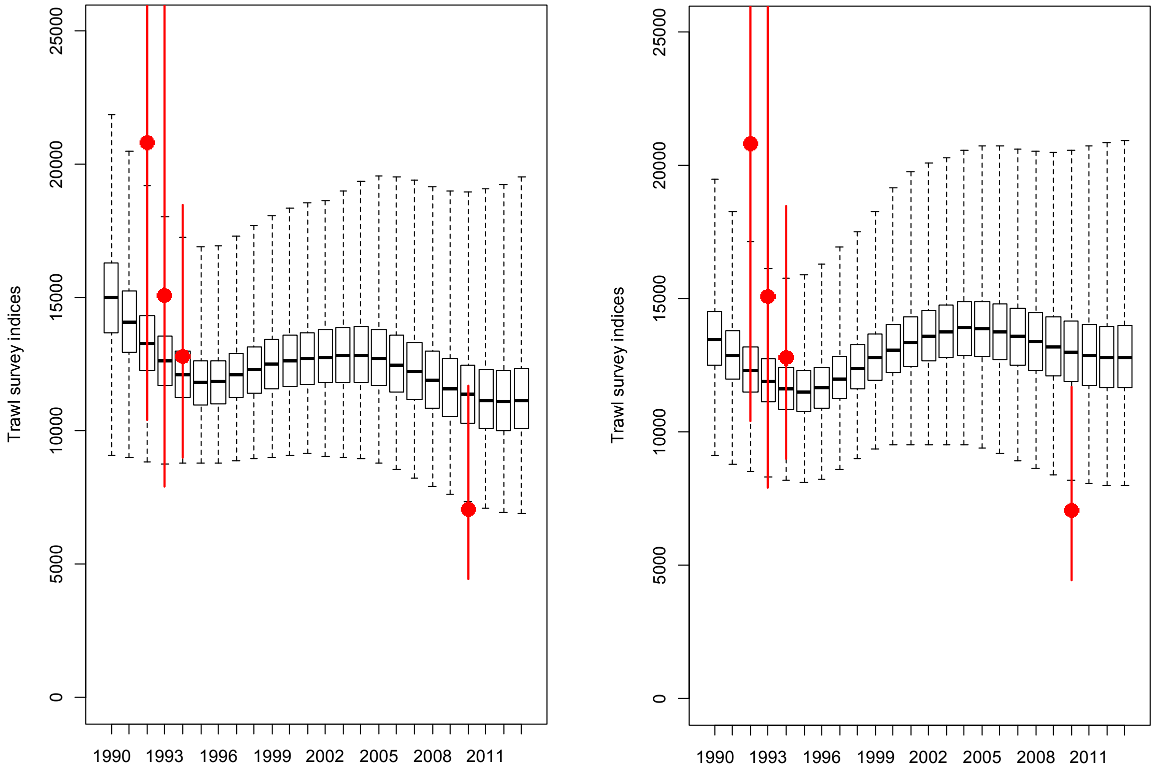

Figure 4: Observed (red, with 99% CIs) and predicted (black) trawl survey biomass from two MCMC model runs: Haist-uniform and Francis-lognormal. For the predictions, each box contains 50% of the distribution and the whiskers cover the full range. (Copied from right-hand panels of figures 25–26 of Cordue 2014a).

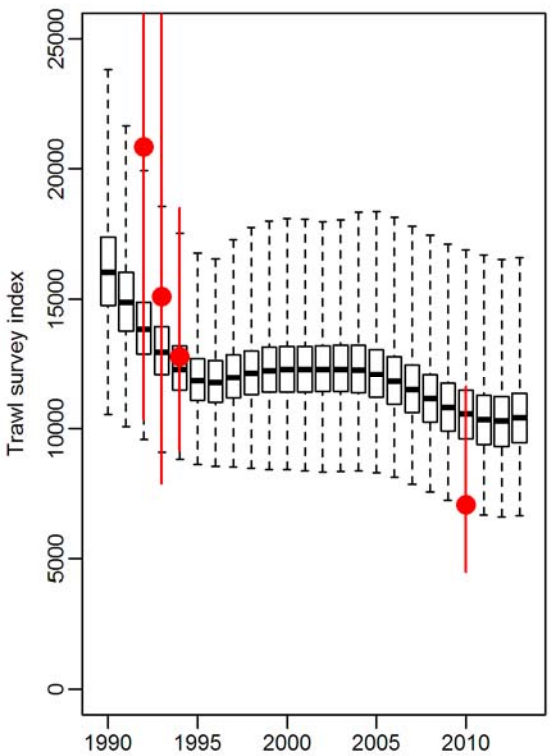

Figure 5: Observed (red, with 99% CIs) and predicted (black) trawl survey biomass from an MCMC Haist-lognormal run. Plotting conventions as in Figure 5Figure 4.

We conclude that, according to the rationale proposed in Section 3, the use of the uniform prior for YCS in a final model run in the 2013 orange roughy assessment was not justified. There was no claim (or basis for a claim) that this was a genuine prior, and no grounds to claim that it was necessary to obtain an acceptable assessment. The slightly better MCMC fit to the trawl survey biomass data from the base run, compared to the alternative run, was not a result of the uniform prior.

For the record, we note that the Haist-lognormal run estimated a more pessimistic status for the MEC stock than either of the published runs (Figure 6).

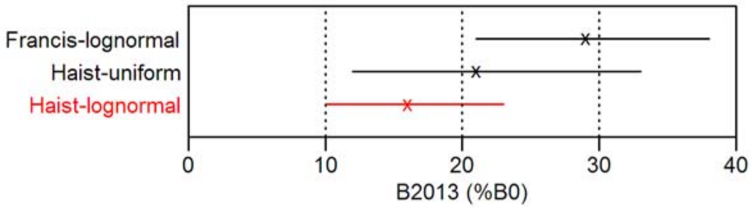

Figure 6: Comparison of Haist-lognormal model MCMC estimates of stock status [B_{2013} (%B_0)] (in red) with those from Cordue (2014a, table 7) (plotted in black). For each model 'x' is the median and the horizontal lines span the 95% confidence interval. Vertical broken lines indicate, from left to right, the management hard and soft limits and the biomass target.

4.2 The 2014 assessments

In concluding remarks in the 2013 orange roughy assessment report there was an acknowledgment of the need for more investigation of different YCS treatments: "It would be interesting to do more runs to explore the effect of the different parameterisations and priors. Certainly, Haist with lognormal priors might yield some insights." (Cordue 2014a, p. 14). However, the 2014 orange roughy assessments showed no evidence of such investigation. In these assessments, only one YCS treatment, the Haist-near-uniform, was used for all model runs. This was justified by the results of the 2013 assessment in which, it was said, "the alternative Francis parameterisation unduly restricted YCS estimates as evidenced by a poor fit to the trawl survey biomass indices ... In contrast the Haist parameterisation, with uniform priors, allowed an excellent fit at the MPD stage and an adequate fit at the MCMC stage." (Cordue 2014b, p. 7). No explanation or justification was given for the change of prior, from uniform to near-uniform. We conclude that the choice of prior in these assessments was unjustified. No claim was made that this prior was genuine, and because no other priors were considered it could not be claimed that the near-uniform prior was necessary to achieve acceptable assessments.

The 2014 smooth oreo assessment used two YCS treatments: Haist-near-uniform for the base model, and Haist-lognormal for an alternative run. The assessment report gave no justification for the choice of YCS prior in the base model beyond saying that the "preference of the Deepwater Working Group was to use a non-informative prior" (Fu & Doonan 2015, p. 35). In fact, the report cast some doubt about the wisdom of this choice, saying, in the same paragraph, that "some form of informative prior could help prevent the model producing implausible values of recruitment." No claim was made that the near-uniform prior was genuine, and it was clearly not necessary to obtain a satisfactory assessment because results from the base and alternative runs were quite similar. We conclude that the use of the near-uniform YCS prior in this assessment was unjustified.

In the 2014 hoki assessment a uniform YCS prior was considered (in the Haist-uniform treatment) first in some "pre-assessment" runs (based on the 2013 assessment model) and then in one of eight final sensitivity runs, in comparison with the base run, which used the Haist-lognormal treatment. Neither this sensitivity run (run 1.16) nor another using the Francis-lognormal treatment (run 1.17) were included in the final three runs for which projections were done. Evaluation of the effect of the uniform prior was restricted to consideration of a posterior profile (see quote in section 4.1 from McKenzie 2015) and many comparisons of model outputs (e.g., table 24 and figures 44, 51, 52, 56, 59, 60, McKenzie 2015). No goodness-of-fit comparisons were made. We conclude that the use of the uniform prior in this assessment was unjustified, though this was of little consequence because this prior was not used in model projections, which would have been the basis for management advice.

4.3 Two asides

There were two matters that arose during our review of recent assessments that lie outside the objective of this report, but which merit some mention.

4.3.1 The YCS parameterisations

Some investigations using the assessment model of Cordue (2014a) provided a deeper understanding of the Haist and Francis YCS parameterisations and led to some cautions about their use. We restricted attention to the two model variants that differed only in this parameterisation: the Haist-lognormal and Francis-lognormal.

A key difference between the two parameterisations is how they deal with what we might call *parameter ambiguity* - the fact that we can create multiple parameter vectors all of which produce exactly the same model outputs such as SSB trajectories and fits to the observations. For the Haist parameterisation, the ambiguity is restricted to the vector of free YCSs: we can multiply this vector by some arbitrary number without changing the model output. Two components of the objective function are used to resolve this ambiguity in the Haist-lognormal model: the YCS prior, and a vector-average penalty (which encourages the mean YCS to be close to 1). In the process of parameter estimation the model scales each potential YCS vector to minimise the contribution of these two objective-function components. For the Francis parameterisation, the ambiguity involves both the YCS vector and the Bmean (or Rmean) parameter: if we multiply the YCS by a constant and divide Bmean by the same constant the model output is unchanged. In the Francis-lognormal model this ambiguity is also resolved by two objective-function components, in this case the priors on the YCS and Bmean.

For each of the two models, we constructed a parameter vector which would produce almost exactly[1] the same output, and fit to the data, as the MPD parameter vector for the other model. We write the MPD vector for the Haist model as $\mathbf{P}_{\mathrm{H.mpdH}}$, and the corresponding vector for the Francis model as $\mathbf{P}_{\mathrm{F.mpdH}}$, where H and F denote Haist and Francis, respectively, and the first letter in the subscript indicates which model the parameter vector is used with. Note that $\mathbf{P}_{\mathrm{H.mpdH}}$ and $\mathbf{P}_{\mathrm{F.mpdH}}$ are very similar: they differ only in that the latter contains Bmean where the former has B_0; and the free YCSs in the latter are a constant multiple of those in the former (the multiplier happens to be 0.507). We used the objective function to resolve ambiguity in each of the cross-model vectors ($\mathbf{P}_{\mathrm{H.mpdF}}$ and $\mathbf{P}_{\mathrm{F.mpdH}}$).

The purpose of the YCS prior is to measure how likely the YCSs in each parameter vector are to have occurred given the prior distribution. Because exactly the same lognormal prior was used in both models, this measure ought to produce similar outcomes in the two models, but it doesn't (Table 6). The Haist-lognormal model says that the YCS vector at its own MPD is "better" (more likely to occur) than that at the Francis-lognormal MPD by 43 points [= (-12) - (-55)], whereas the Francis-lognormal model says that it is worse by 27 points [= (-65) - (-92)]. Thus, at least one of these parameterisations must be at fault in distorting the way that the YCS prior measures likelihood.

It appears that the problem is with the Francis parameterisation. With the Haist parameterisation, the recruitment in year y is calculated as R_y = $\mathrm{YCS}_{y\text{-}1}\, R_0$ SRR($\mathrm{SSB}_{y\text{-}1}$), where SRR is the stock-recruit relationship and YCS_y is a derived parameter, calculated by rescaling the free parameters $\{\mathrm{Y}_y\}$ to have mean 1 [i.e., $\mathrm{YCS}_y = \mathrm{Y}_y/\mathrm{mean}_{y'}(\mathrm{Y}_{y'})$]. Ideally, the prior would be applied to the $\{\mathrm{YCS}_y\}$, because they are scaled to have the desired meaning (e.g., $\mathrm{YCS}_y = 2$ means that the associated recruitment is double what would be expected on average). In fact, for technical reasons the prior is applied to the $\{\mathrm{Y}_y\}$, rather than the $\{\mathrm{YCS}_y\}$ (it was because of this drawback that the Francis parameterisation was

[1] The output is very slightly different because of the different way the two models treat the fixed YCSs

developed). However, this did not matter much in the MEC assessment because the effect of a strong vector-average penalty was to make Y_y very close to YCS_y. With the Francis parameterisation, $R_y = YCS_{y\text{-}1} R_{\text{mean}} \text{ SRR}(SSB_{y\text{-}1})$ and the $\{YCS_y\}$ are free parameters, which means that the interpretation of YCS_y is relative to R_{mean}, rather than R_0. Unfortunately, R_{mean} seems to have no biological meaning. Its value varies with each YCS vector, being adjusted to minimise the combined contributions to the objective function of the YCS and Bmean priors (so, for example, $R_{\text{mean}} = 4.3R_0$ at the Francis-lognormal MPD and $R_{\text{mean}} = 2.0R_0$ at the Haist-lognormal MPD).

We conclude that the Francis parameterisation undermines the YCS prior (so that it does not provide a reliable measure of how likely each potential YCS vector is) because the additional parameter (either Bmean or Rmean) rescales each potential YCS in a different way. The Haist parameterisation will avoid this problem as long as a strong penalty is used to encourage the $\{Y_y\}$ to have mean 1, as was done in the 2013 orange roughy assessment. A drawback is that strong penalties of this type can adversely affect MCMC convergence [it was perhaps for this reason that this penalty was omitted (without comment) in the 2014 orange roughy assessments]. It might be possible to avoid these problems with YCS priors by using something akin to the ADMB dev-vector, which is constrained to have mean 0 (see Fournier 2013), for YCSs in log space.

Table 6: Contributions of the YCS prior to the model objective function (i.e., the negative logarithm of the prior) for four parameter vectors. In each row of the table one parameter vector is the MPD vector for one model, and the other is the equivalent vector for the other model (see text for details).

	Haist-lognormal model		Francis-lognormal model	
	Vector	YCS prior	Vector	YCS prior
Haist-lognormal MPD	$\mathbf{P}_{\text{H.mpdH}}$	-55	$\mathbf{P}_{\text{F.mpdH}}$	-65
Francis-lognormal MPD	$\mathbf{P}_{\text{H.mpdF}}$	-12	$\mathbf{P}_{\text{F.mpdF}}$	-92

4.3.2 The 2013 orange roughy MEC assessment

We found two reasons to be concerned about this assessment. Because a primary requirement of an assessment is to reconstruct the biomass trajectory of the stock, it is particularly important that a biomass data set be well fitted (Francis 2011), especially if (like the orange roughy trawl survey data) it is fishery-independent and shows a strong trend. That this trend was not well fitted in this assessment is evident from Figure **4**, and was acknowledged in the assessment report: “the bulk of posterior density was not supportive of the trend in the trawl time series” (Cordue 2014a, p. 14). [We note in passing that there were two other fishery-independent biomass data sets in this assessment (from egg and acoustic surveys), but that these were deemed to be so weak that fits to them were unmentioned in the assessment report.]

Another reason to be concerned about this assessment it that it shows a pattern that has occurred, both before and since, in other orange roughy assessments: the assessment estimates a sharp decline in biomass followed by a period of rebuilding, with the latter not being supported by biomass data. This was first seen in the 2001 assessment of the South Chatham Rise stock, of which Francis (2001a, p. 3) said “There is some doubt as to whether the stock biomass is rebuilding (as the stock assessment model suggests) because none of the four CPUE series show any such rebuild.”. A similar conflict was seen in two independent assessments of the Northeast Chatham Rise stock in the same year (Figure 7). The pattern was repeated in 2006: “No model-based stock assessments were conducted for ORH 3B stocks from 2007 to 2013 inclusive. This was primarily because the 2006 stock assessment ... showed an increasing trend in biomass which was not supported by recent biomass indices” (Ministry for Primary Industries 2014). Cordue (2014b, p. 55) noted the same pattern in the 2014 orange roughy stock assessments: “For each of the four stocks, median stock status is trending upwards in recent years ... However, the biomass indices for the individual stocks do not generally show an upward trend over those years (or do not contain any trend information ...).”

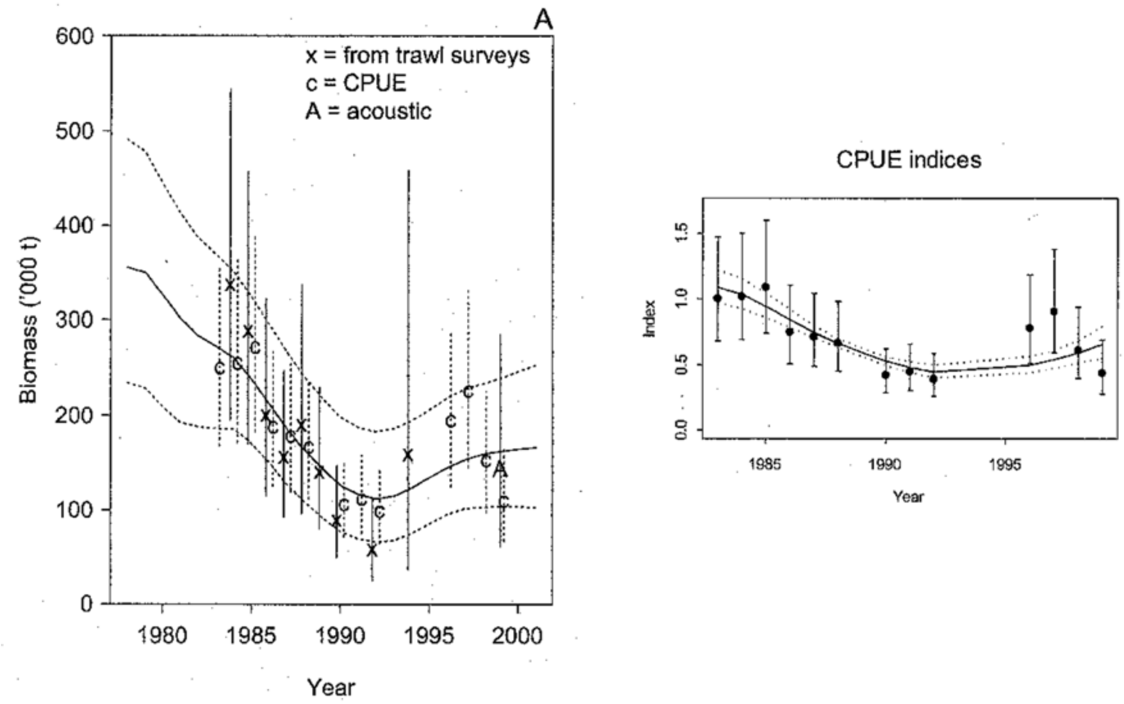

Figure 7: Reproductions of figure 4A of Francis (2001b) (left panel) and figure 2B of Smith et al (2002) (right panel) showing how, in both of these assessments of the Northeast Chatham Rise stock of orange roughy, the model's assessment that the stock was rebuilding was in conflict with the declining trend in post-1995 CPUE.

We are unable to offer any explanation for this repeated pattern. However, it suggests that our stock assessment models may be misrepresenting the productivity of orange roughy stocks in such a way that they incorrectly estimate that depleted stocks are rebuilding. Thus it would be sensible to be cautious about such rebuilding estimates, particularly when they conflict with biomass trend data.

5. CONCLUSIONS

We conclude that the default prior for YCSs in stock assessment models should be lognormal, but there are two situations in which an alternative distribution could properly be used: where there is evidence that YCSs for the stock being assessed may have a distribution clearly different from the lognormal; or where its use is necessary to obtain an acceptable assessment. Neither situation occurred in any of the recent New Zealand assessments in which a non-lognormal prior was used.

6. ACKNOWLEDGMENTS

This work was funded by the Ministry for Primary Industries under project SCM201406. We thank Ian Doonan for reviewing the manuscript.

7. REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. Automatic Control, IEEE *Transactions on Automatic Control 19(6)*: 716–723.

Allen, K.R. (1973). The influence of random fluctuations in the stock recruitment relationship on the economic return from salmon fisheries. *Rapports et procès-verbaux des réunions Conseil permanent international pour l'exploration de la mer.* 164: 350–359.

Archibald, C.P.; Fournier, D.; Leaman, B.M. (1983). Reconstruction of stock history and development of rehabilitation strategies for Pacific ocean perch in Queen Charlotte Sound, Canada. *North*

American Journal of Fisheries Management 3, 283–294.

Breen, P.A.; Kim, S.W.; Andrew, N.L. (2003). A length-based Bayesian stock assessment model for abalone. *Marine and Freshwater Research* 54, 619–634.

Bull, B.; Francis, R.I.C.C.; Dunn, A.; McKenzie, A.; Gilbert, D.J.; Smith, M.H.; 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.

Chen, Y.; Kanaiwa, M.; Wilson, C. (2005). Developing and evaluating a size-structured stock assessment model for the American lobster, *Homarus americanus*, fishery. *New Zealand Journal of Marine and Freshwater Research*, 39(3), 645–660.

Cordue, P.L. (2013). MEC ORH stock assessment: MPDs and preliminary MCMCs. DWWG2013-40. (Unpublished report held by Ministry for Primary Industries, Wellington.)

Cordue, P.L. (2014a). A 2013 stock assessment of Mid-East Coast orange roughy. *New Zealand Fisheries Assessment Report* 2014/32. 54 p.

Cordue, P.L. (2014b). The 2014 orange roughy stock assessments. *New Zealand Fisheries Assessment Report* 2014/50. 135 p.

Craig, K. (2012). The Beaufort Assessment Model (BAM) with application to cobia: mathematical description, implementation details, and computer code. SEDAR28-RW01. SEDAR, North Charleston, SC. 37 p.

Deroba, J.J.; Butterworth, D.S.; Methot, R.D.; De Oliveira, J.A.A.; Fernandez, C.; Nielsen, A.; Cadrin, S.X.; Dickey-Collas, M.; Legault, C.M.; Ianelli, J.; Valero, J.L.; Needle, C.L.; O'Malley, J.M.; Chang, Y.-J.; Thompson, G.G.; Canales, C.; Swain, D.P.; Miller, D.C.M.; Hintzen, N.T.; Bertignac, M.; Ibaibarriaga, L.; Silva, A.; Murta, A.; Kell, L.T.; de Moor, C.L.; Parma, A.M.; Dichmont, C.M.; Restrepo, V.R.; Ye, Y.; Jardim, E.; Spencer, P.D.; Hanselman, D.H.; Blaylock, J.; Mood, M.; Hulson, P.-J.F. (2015). Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. *ICES Journal of Marine Science*: 72(1): 19–30.

Dickey-Collas, M.; Hintzen, N.T.; Nash, R.D.M.; Schön, P.-J.; Payne, M.R. (2015). Quirky patterns in time-series of estimates of recruitment could be artefacts. *ICES Journal of Marine Science*: 72(1), 111–116.

Fournier, D.A. (2013). An Introduction to AD Model Builder for use in nonlinear modeling and statistics. Version 11.1 (2013-05-01), Revised manual (2013-06-24). Available from http://www.admb-project.org/.

Fournier, D.A.; Hampton, J.; Sibert, J.R. (1998). MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga. Canadian Journal of Fisheries and Aquatic Sciences 55(9)*, 2105–2116.

Francis, R.I.C.C. (2001a). Stock assessment of orange roughy on the South Chatham Rise. *New Zealand Fisheries Assessment Report 2001/27.* 25 p.

Francis, R.I.C.C. (2001b). Stock assessment for 2001 of orange roughy on the northeast Chatham Rise. *New Zealand Fisheries Assessment Report 2001/41*. 32 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.

Fu, D.; Doonan, I.J. (2015). Assessment of OEO 4 smooth oreo for 2012–13. *New Zealand Fisheries Assessment Report* 2015/7. 41 p.

Funk, F.; Quinn, T.J.; Heifetz, J.N.; Powers, J.E.; Schweigert, J.F.; Sullivan, P.J.; Zhang, C.-I. (1998). Fishery Stock Assessment Models. Edited by Alaska Sea Grant college Program Report No. AK–SG–98–01. University of Alaska, Fairbanks, Alaska.

Haddon, M. (2000). Modelling and quantitative methods in fisheries. Chapman & Hall/CRC, Boca Raton, Florida.

Hennemuth, R.C.; Palmer, J.E.; Brown, B.E. (1980). A Statistical Description of Recruitment in Eighteen Selected Fish Stocks. *Journal of Northwest Atlantic Fishery Science* 1, 101–111.

Haist, V.; Breen, P.A.; Starr, P.J. (2009). A multi-stock, length-based assessment model for New Zealand rock lobster (*Jasus Edwardsii*). *New Zealand Journal of Marine and Freshwater Research* 43(1), 355–371.

Hightower, J.E.; Grossman, G.D. (1985). Comparison of Constant Effort Harvest Policies for Fish Stocks with Variable Recruitment. *Canadian Journal of Fisheries and Aquatic Sci*ences 42(5),

982–988.
Hilborn, R.; Maunder, M.N.; Parma, A.M.; Ernst, B.; Payne, A.I.L.; Starr, P.J. (2003). Coleraine: a generalized age-structured stock assessment model. User's Manual Version 2.0. Accessed from http://fish.washington.edu/research/colerain/pdf/coleraine.pdf 1 May 2015
Hilborn, R.; Walters, C.J. (1992). Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Chapman and Hall, New York.
Huang, B.; Walters, C. J. (1983). Cohort analysis and population dynamics of large yellow croaker in the China Sea. *North American Journal of Fisheries Manage*ment 3, 295–305.
Kleiber, P.; Hampton, J.; Davies, N.; Hoyle, S.; Fournier, D.A. (2013). MULTIFAN-CL User's Guide, November 2013. Accessed from http://www.multifan-cl.org/. 1 May 2015
Lee, H.-H.; Maunder, M.N.; Piner, K.R.; Methot, R.D. (2011). Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. *Fisheries Research 109*: 89–94.
Legault, C.M.; Restrepo, V.R. (1999). A flexible forward age-structured assessment program. *ICCAT Collected Volume of Scientific Papers* 49 (2), 246–253.
Ludwig, D.; Walters, C.J. (1981). Measurement errors and uncertainty in parameter estimates for stock and recruitment. *Canadian Journal of Fisheries and Aquatic Sciences* 38(6): 711–720.
Martell, S. (2011). iSCAM Users Guide Version 1.0, Available from https://sites.google.com/site/iscamproject/. Accessed 21 October 2013.
Maunder, M.N. (2003). Paradigm shifts in fisheries stock assessment: from integrated analysis to Bayesian analysis and back again. *Natural Resource Modeling* 16(4), 465–475.
Maunder, M.N. (2005). Inter-American Tropical Tuna Commission Workshop On Stock Assessment Methods. La Jolla, California (USA) 7–11 November 2005.
Maunder, M.N.; Deriso, R.B. (2003). Estimation of recruitment in catch-at-age models. *Canadian Journal of Fisheries and Aquatic Sciences* 60(10), 1204–1216.
Maunder, M.N.; Punt, A.E. (2013). A review of integrated analysis in fisheries stock assessment. *Fisheries Research* 142(0): 61–74.
Maunder, M.N.; Starr, P.J. (2001). Bayesian assessment of the SNA1 snapper (*Pagrus auratus*) stock on the north-east coast of New Zealand. *New Zealand Journal of Marine and Freshwater Research* 35(1): 87–110.
Maunder, M.N.; Watters, G.M. (2003). A-SCALA: an age-structured statistical catch-at-length analysis for assessing tuna stocks in the eastern tropical Pacific Ocean. *Inter-American Tropical Tuna Commission Bulletin* 22(5): 435–582.
McAllister, M.K.; Kirkwood, G.P. (1998). Bayesian stock assessment: a review and example application using the logistic model. *ICES Journal of Marine Science 55*: 1031–1060.
McKenzie, A. (2015). Assessment of hoki (*Macruronus novaezelandiae*) in 2014. *New Zealand Fisheries Assessment Report* 2015/09. 68 p.
Methot, R.D. (2012). User Manual for Stock Synthesis Model Version 3.24f, Updated October 3, 2012. Accessed from http://ntf.nefsc.noaa.gov/Stock_Synthesis_3.htm.1 December 2014.
Methot, R.D.; Wetzel, C.R. (2013). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research 142*: 86–99.
Ministry for Primary Industries (2014). Fisheries Assessment Plenary, May 2014: stock assessments and stock status. Compiled by the Fisheries Science Group, Ministry for Primary Industries, Wellington, New Zealand. 1381 p.
Myers, R.A. (2001). Stock and recruitment: generalizations about maximum reproductive rate, density dependence, and variability using meta-analytic approaches. *ICES Journal of Marine Science*, 58, 937–951.
Myers, R.A.; Bridson, J.; Barrowman, N.J. (1995). Summary of worldwide stock and recruitment data. *Canadian Technical Report on Fisheries and Aquatic Sciences*, 2024: 327 p.
Peterman, R.M. (1981). Form of Random Variation in Salmon Smolt-to-Adult Relations and Its Influence on Production Estimates. *Canadian Journal of Fisheries and Aquatic Sciences* 38(9), 1113–1119.
Punt, A.E.; Huang, T.; Maunder, M.N. (2013). Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. *ICES Journal of Marine Science* 70(1), 16–18.

Quinn, T. J. II; Deriso, R.B. (1999). Quantitative fish dynamics. Oxford University Press, New York.

Ricard, D.; Minto, C.; Jensen, O.P.; Baum, J.K. (2012). Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish and Fisheries 13(4)*: 380–398.

Smith, A.D.M.; Punt, A.E.; Wayte, S.E.; Starr, P.J.; Francis, R.I.C.C.; Stokes, T.K.; Hilborn, R.; Langley, A. (2002). Stock assessment of the northeast Chatham Rise orange roughy for 2001. *New Zealand Fisheries Assessment Report 2002/25*. 30 p.

Stuart, A.; Ord, J.K. (1987). Kendall's advanced theory of statistics, Volume 1. 5th edition. Charles Griffin and Company, London.

Sullivan, P.J.; Lai, H.-L.; Gallucci, V.F. (1990). A catch-at-length analysis that incorporates a stochastic model of growth. *Canadian Journal of Fisheries and Aquatic Sciences*, 47 (1), 184–198.

Tingley, G. (2013). Deepwater fishery assessment working group, Notes of meeting of 29 April 2013. DWWG-2013/46**.** (Unpublished report held by Ministry for Primary Industries, Wellington.)

Walters, C.J.; Hilborn, R. (1976). Adaptive Control of Fishing Systems. *Journal of the Fisheries Research Board of Canada* 33(1), 145–159.

Walters, C.J.; Ludwig, D. (1981). Effects of Measurement Errors on the Assessment of Stock–Recruitment Relationships. *Canadian Journal of Fisheries and Aquatic Sciences* 38(6), 704–710.

APPENDIX A: Fits to SSB-recruitment data

Here were provide plots of the fits to the SSB-recruitment data sets, as described in Section 2.2.

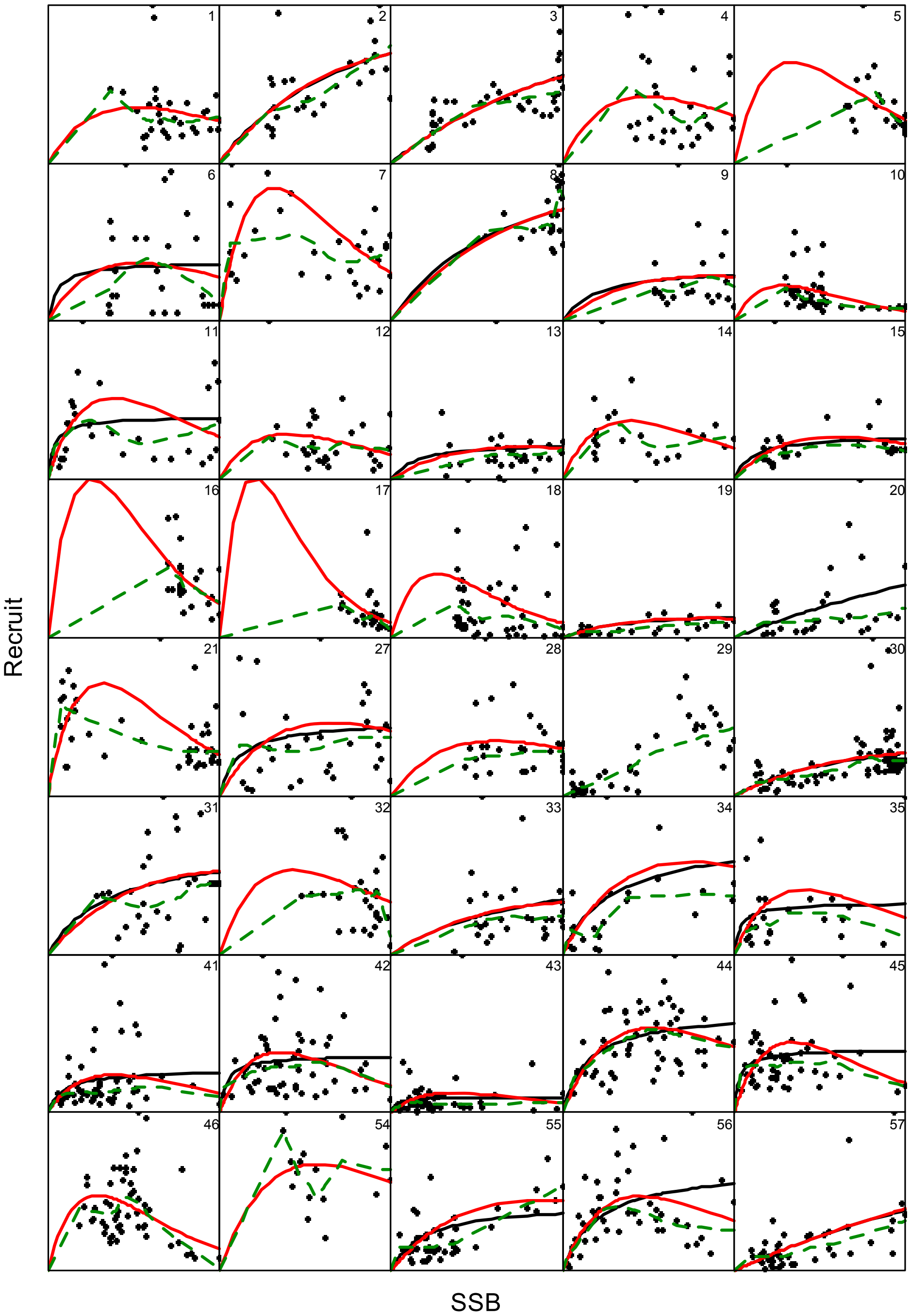

Figure A1: SSB-recruitment time series from the catch at age category (points) with estimated Beverton-Holt (black line), Ricker (red line) and lowess (green dashed line) stock-recruitment relationships (no line is shown where no satisfactory fit was found). In each panel the origin is at the lower left corner. The number in each panel identifies the time series.

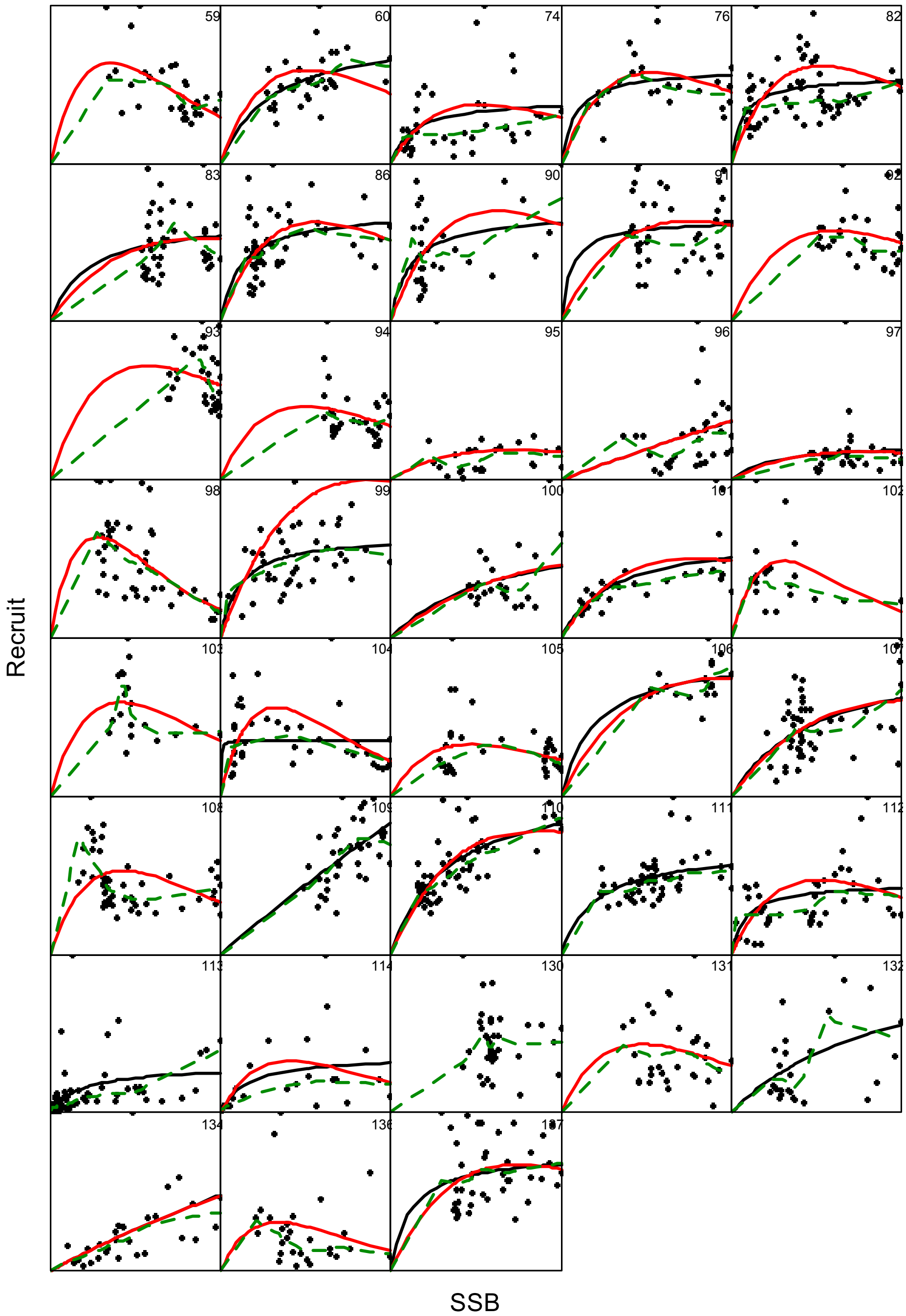

Figure A1 continued.

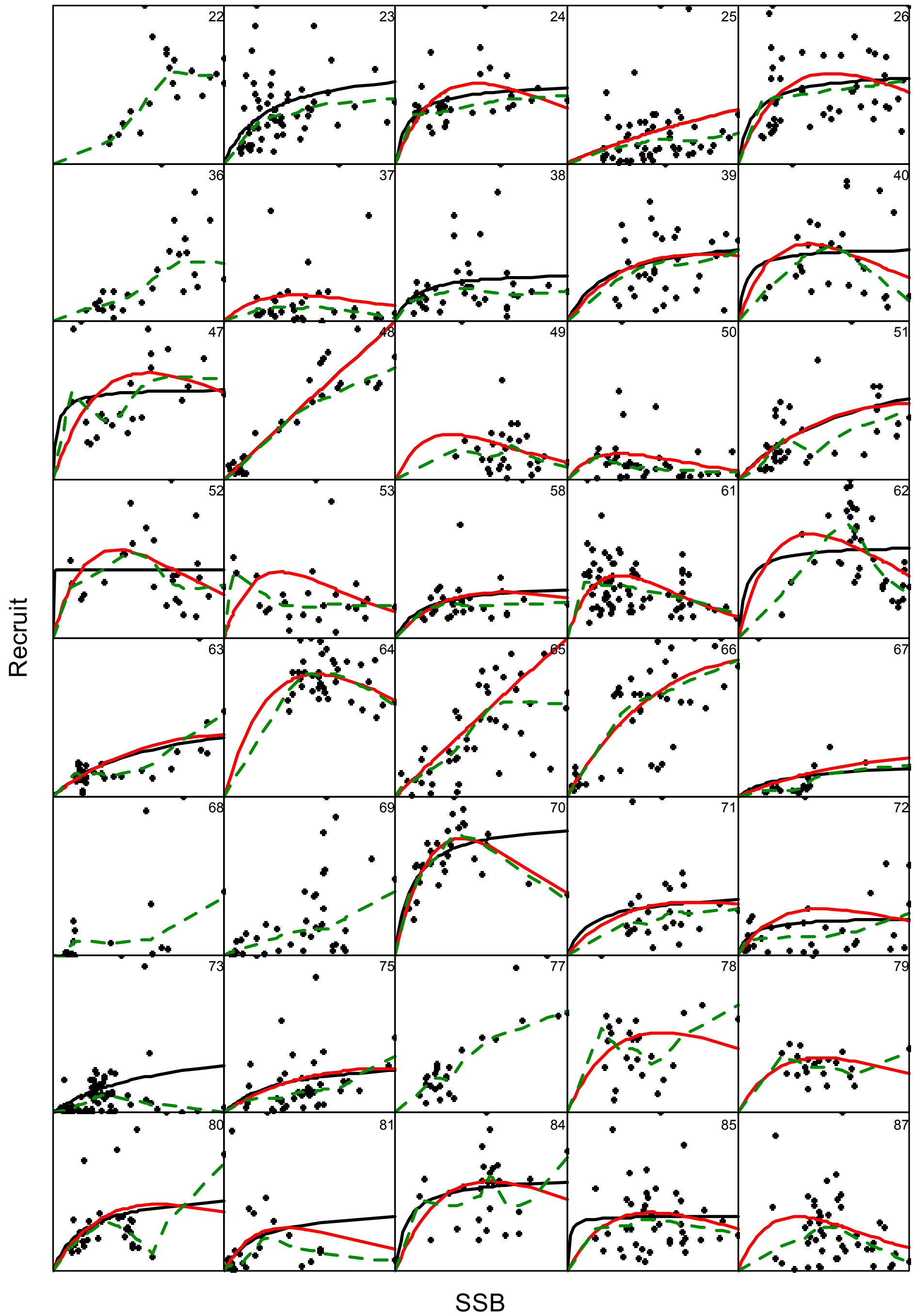

Figure A2: As in Figure A1, but for SSB-recruitment series from the VPA category.

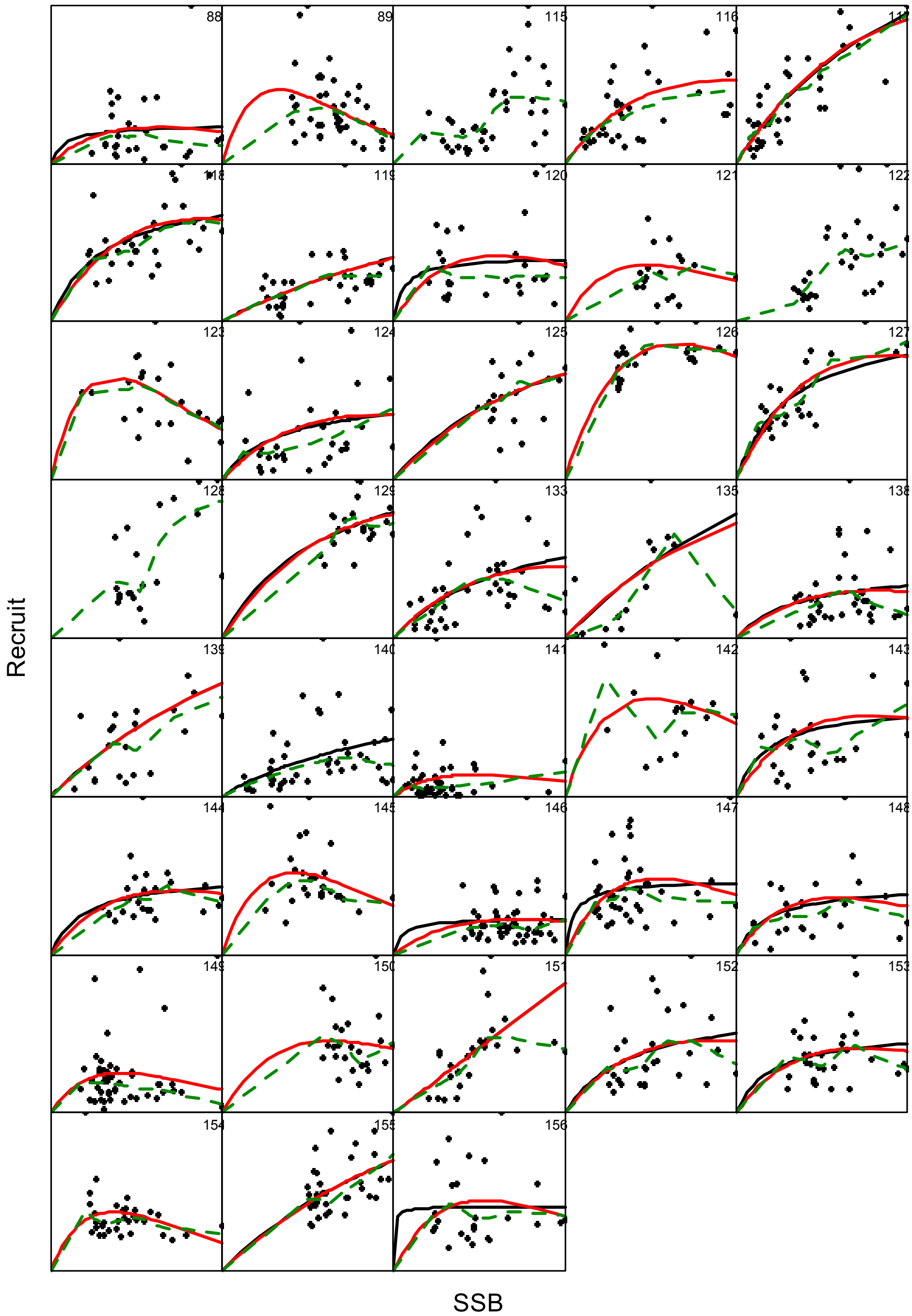

Figure A2 continued.

APPENDIX B: Statistical test on YCS vectors

Here is the R code for the statistical test described in Section 2.2.

```
TestYCS <- function(ycs)
  ## Uses AIC to decide whether a vector of YCS is more
  ## likely to be uniform or lognormal.
  ##
  ## Returns the AIC difference, which is positive (or negative) if
  ## the vector is more likely to be lognormal (or uniform)
  {
    Nycs <- length(ycs)
    ## Calculate NLL assuming a uniform distn U(0,B)
    ## where B = max(ycs)*(Nycs+1)/Nycs
    B <- max(ycs)*(Nycs+1)/Nycs
    NLL.unif <- Nycs*log(B)
    AIC.unif <- 2*(1 + NLL.unif)
    ## Estimate mu & sigma & calculate likl assuming LN with
    ## mean mu and sd sigma (in log space)
    NLL <- function(pars,dat){
      mu <- pars[1]
      sigma <- exp(pars[2])
      n <- length(dat)
      nll <- sum(log(dat))+0.5*n*log(2*pi)+n*log(sigma)+
        (0.5/sigma^2)*sum((log(dat)-mu)^2)
      nll
    }
    init.pars <- c(mean(log(ycs)),log(sqrt(var(log(ycs)))))
    pars <- nlm(NLL,init.pars,dat=ycs)$estimate
    NLL.LN <- NLL(pars,ycs)
    AIC.LN <- 2*(2 + NLL.LN)
    AIC.diff <- AIC.unif-AIC.LN
    AIC.diff
  }
```

APPENDIX C: A Haist-lognormal MCMC run

Here we provide some details about our Haist-lognormal MCMC run for the MEC orange roughy stock, mentioned in Section 4.1

Cordue (2014a) presented the CASAL input files for his base run, which used the Haist-uniform treatment of YCS, but did not describe exactly how these files differed from the corresponding files for any other model run. We made the following changes to the base run input files to create files for our Haist-lognormal run, which we assumed was similar to the published Francis-lognormal run in covering a shorter time span and estimating fewer YCSs than the base run.

In the population file
- changed @initial from 1882 to 1910
- in @recruitment block
    - changed first_free from 1881 to 1909
    - removed years 1881-1908 from YCS_years
    - removed the values from YCS corresponding to years 1881-1908

In the estimation file
- in the @estimate block for recruitment.YCS
    - changed 'uniform' to 'lognormal'
    - added subcommands "mu 1 1 ... 1" and "cv 1.53 1.53 ... 1.53"
    [NB, sigmaR = 1.1 and 1.53 = $(\exp(1.1^2) - 1)^{0.5}$]
    - removed the values from lower_bound and upper_bound corresponding to years 1881-1908
- in @vector_average_penalty change upper_bound to 87
- in @MCMC changed systematic to true

We followed the same approach to the MCMC analysis as Cordue (2014a, p. 9) except that our three chains were a little longer (15 million samples) and we subsampled the concatenated chains systematically, rather than randomly (as did Cordue 2014b). A diagnostic plot for these chains (Figure C3) shows that convergence was not ideal, but seemed better than was achieved for the published Haist-uniform and Francis-lognormal chains (cf figures 21–24, Cordue 2014a).

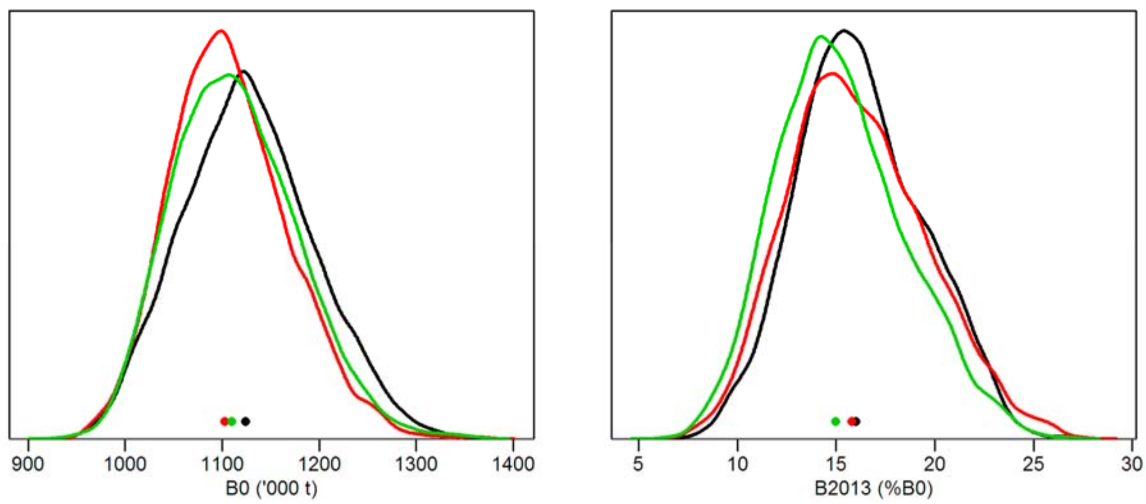

Figure C3: Diagnostic plot for the Haist-lognormal MCMC run showing marginal posterior distributions (coloured lines) and their medians (points) for each of three chains for B_0 (left panel) and B_{2013}(%B_0).