CCSBT-MPTM/0502/04

CSIRO
MARINE RESEARCH

Exploration of the SBT operating model with implications for the selection of the core set and robustness trials.

D. Kolody, M.Basson, A. Preece, J. Hartog, T. Polacheck

Prepared for the CCSBT Special Management Procedure Technical Meeting (MPTM)
15-19 February 2005, Seattle, USA

Exploration of the SBT operating model with implications for the selection of the core set and robustness trials.

D. Kolody, M.Basson, A. Preece, J. Hartog, T. Polacheck

Abstract

We explore a range of SBT MP operating model "core" set specifications to assist participants of the CCSBT Feb 2005 MP meeting to make a final decision for a reference set to evaluate candidate Management Procedures. We present results and suggest criteria which might assist the decision process. We focus most of our explorations on a few key assumptions that have been recognized as rather arbitrary in the past, but that potentially have substantial implications for estimated SBT dynamics. These alternative assumptions suggest that the current core proposal probably does not cover an adequate range of uncertainty, and tends to be relatively pessimistic. We propose some alternatives core sets that we consider to have better properties in terms of providing a reasonable coverage of the uncertainty. We do however note that whichever potential core set we consider, there are trade-offs, and all core sets have some undesirable features. The workshop needs to consider carefully on what basis to select the final reference set and these decisions should be explicitly documented for future reference. Additional details are provided in a separate annex.

Introduction

We have explored a range of alternative operating models for possible inclusion in the final reference set for evaluation of MPs. We were concerned that the current core set may not be a adequate reflection of the full uncertainty implied by the data and plausible model structures. There are scenarios, not in the current core set, that are both considerably more optimistic and more pessimistic in terms of recent trends or future stock status. Given this, we considered it important to further explore the original core and consider other options for the core set.

Some potential modifications to the CCSBT Operating Model Core set

There is a growing recognition that fisheries assessment models are often sensitive to arbitrary assumptions and that it is important in such cases to encompass a reasonable range of alternative assumptions/hypotheses to achieve robust results. The SBT conditioning model appears to be a good example of this problem. The SAG/SC/MP developers have long recognized that there are serious limitations to our ability to objectively quantify SBT historical and future dynamics using statistical methods, and common sense or "expert judgement" has a substantial influence on model proposals. Additionally, our ability to appropriately apply statistical approaches is limited by lack of appropriate sampling models for the underlying data inputs (e.g. likelihood functions for size, CPUE, tag return and reporting rates). With these things in mind, we propose a number of alternative (modified) reference case scenario for discussion at the Feb 2005 MP technical meeting.

Although a large number of runs with different assumptions and settings were done with the SBT operating model, it was not possible to explore all options and all interactions in the limited time. We note that there is no clear or obvious "best" core set; each set we examined had some good and some poor characteristics. Also, given the number of uncertainty dimensions and range of hypotheses within each of these that can have substantive effects, it is not feasible to include all of these within a single core set specification. As such, trade-offs will be required as to what elements should be included in the core to achieve a balanced representation of the underlying uncertainty. In doing this, there are, however, some general characteristics that are desirable to ensure robustness of an MP that is ultimately tested against the grid.

Our criteria for inclusion of CORE and GRID specifications in the Reference case roughly include:

- Tend to include plausible model hypotheses that represent a relatively high model likelihood
- Tend to include plausible model hypotheses with a relatively low likelihood if there is a general impression that the model should not be informative with respect to this particular dimension
- Tend to drop hypotheses if they add dimensionality to the GRID, without substantially altering the range of uncertainty represented in future (constant current catch projections) dynamics (e.g. recognizing that an increased density on the existing mode does affect the uncertainty distribution, this could in theory be represented by re-weighting the grid)
- Try to maintain a tractable number of GRID dimensions and levels, such that a sample size of 2000 projections will give a reasonable coverage of models
- Represent a reasonable probability that CPUE indices may increase or decrease somewhat over the next few years - i.e. explicitly admit that we may need to use some reverse engineering to ensure that the operating model is consistent with our strong prior that we can not predict with a high degree of certainty the directional change in CPUE in the near future (e.g. avoid a repetition of the problem observed in the 2004 MP meeting that could undermine credibility in the MP process).

Based on our explorations (detailed below and in Annex 1), the dimensions and levels that we would consider of primary importance include the 5 original dimensions and their previously defined levels:

- Three steepness levels (unchanged from 2004 SAG proposal)
- Three M(age = 0) (unchanged from 2004 SAG proposal)
- Three M(age = 11+) (unchanged from 2004 SAG proposal)
- Two omega (unchanged from 2004 SAG proposal)
- Five CPUE series

Plus the following three additional dimensions and levels within them:

- Three catchability age range assumptions (qa4-30, 6-18, 8-12) – expanding this dimension tends to support more pessimistic projections
- Two effective sample size assumptions (square root transformation, and 0.5*original) – expanding this dimension tends to support more optimistic projections
- Tags removed and tags included (with updated release/recovery data, reporting rate assumptions and objective function weighting)

This full cross of features leads to an unwieldy 3240 MPDs, so we explored a range of combination of these. Based on the exploration that we were able to complete we would suggest that the core/reference set be based on some combinations of the following simplifications to a full grid based on the above dimensions and levels within them, roughly in order of preference:

1) replace the 5 current CPUE series with the median
2) drop the square root transformation of effective sample size and keep the slightly more optimistic 0.5*original ESS weighting scheme which was based on an explicit analysis of sampling procedures
3) Drop the notag scenario in favour of the reporting rate 8 scaled (RR8s) tagging assumption, which is very similar in terms of projection behaviour (and actually uses the most substantial fisheries independent data that was collected to reduce assessment uncertainty!)
4) drop the qa6-18 assumption and possibly weight the qa4-30 assumption 2:1 relative to qa8-12 since qa6-18 and qa4-30 result in similar current catch predictions

One suggestion is to invoke the first 3 options above, which results in a manageable MPD grid of 162. In any case, we would suggest equal weighting of all levels for any of the new dimensions that were included in the final grid because we don't believe that the likelihoods can be informative for these cases. We would also suggest only minor modifications to the robustness set to reflect changes to the reference set, i.e. remove the qa8-12 and tag sensitivity tests, or possibly adding new sensitivity tests to encompass the more extreme CPUE series that are dropped from the core set.

Brief justification for the separate elements of the proposal is provided in Annex 1 of the document, but we present the characteristics of the proposal in this section. Australian participants have a complete set of the 3240 MPD cross above, plus additional scenarios described below, which can be analysed and graphically presented in a number of ways.

Characteristics of the potential new core sets

Figure 1,Figure 2 and Figure 3 illustrate the general problem that the current core set reflects only a subset of the uncertainty on the basis of the data and model structures that we consider as plausible and thus encompassing the range of uncertainty about the SBT stock and fishery. There are scenarios that are both considerably more optimistic and more pessimistic in terms of recent trends or future stock status. We caution that it simply was not possible to consider all possible combinations of factors and that the set of options we have presented represents a pragmatic approach for moving forward with items that we have identified as likely to be important. One problem that we encountered in trying to define what to include or remove from a potential core set was that substantial interactions were found among the elements within the full grid. We found that interactions among model components can be counter-intuitive. As such, a careful examination of the conditioning results for the final proposed reference set should be undertaken before signing off on a final specification at the February technical meeting (i.e. we would caution against inferring what the property of any proposed reference set may be based on examination of results from a sub-set or other combinations of previously examined core sets).

It was relatively straightforward to do conditioning runs and projections for alternative core sets base on the original grid (e.g. original grid with qa8-12 instead of qa4-30). It was less straightforward to combine results from more than one grid set without changing any of the existing code (e.g. combining qa8-12, qa6-18 and qa4-30) because it implies an expansion of the grid. It was also not straightforward to obtain weighted projection results for median CPUE only, since the default projection code resamples from CPUE 2-6. We chose to make these two types of combination outside of the conditioning and projection results and wrote R code to do so.

We show comparisons between some possible core sets (including the original core set with and without tag data). Table 1 lists the names and definitions of the various runs (also see Annex1).

Several types of plots were considered, but the main plots presented here are of two types:
(1) projection results (S2004/S2000, S2008/S2004 and S2020/S2004) summarised over a grid in terms of unweighted and weighted model selection. In the unweighted version (indicated by 'uniform' on graphs), each MPD result from the grid is included once, with constant current catch projections represented by the median of 25 stochastic replicates. In the weighted versions, the priors or priors*likelihood are used as in the proposed grid weighting scheme, with projections based on 2000 stochastic projections. The comparison shows the effect weighting has on the results. (e.g. Figure 1)
(2) expected number or proportion of samples for each level of the three factors m0, M10 and Omega (prior x neg. LLHD affecting sampling weight, marginalized over the over factors), together with the cumulative plot for sampling weights. This latter plot shows what % of models represent a given percentage of the samples. Steeper curves indicate heavy sampling from few models, whereas a gently rising curve indicates more evenly spread out sampling. (e.g. Figure 7)

In the format of the previous grid-axes specification table, the alternative proposal for a core grid is as follows – note that CPUE has essentially been dropped because we propose using median CPUE instead of the separate series, but is included in the table for clarity:

Factor	**No. of levels**	**Values**			**Prior**	**Weights**
Steepness	3	0.385	0.55	0.73	0.2, 0.6, 0.2	As in prior
M0	3	0.30	0.40	0.50	uniform	Prior × Likelihood-based
M10	3	0.07	0.10	0.14	uniform	Prior × Likelihood-based
Omega	2	0.75		1	0.4, 0.6	Prior × Likelihood-based
qa = CPUE age range	3	8-12	6-18	4-30	0.33, 0.34, 0.33	As in prior
CPUE	1	Median CPUE			1	1

Recall that the grid specification does NOT fully define a conditioning run. The other specifications relate to, e.g. weights used with age/size data, weights used with tagging data. These are listed below. We used this grid specification to explore two new cores One used the (Original sample weights /2), called "new core" and the other used the square root transformed weights, called "sqrt new core". In both these cases, tag data were included with a weight of 3.2 and new reporting rate 8s.

The table below summarises the assumptions for the new core options outlined above, together with assumptions for the original core, Cfullnotag:

Name	Weight to tagging	**Weight to Age/Len comp data**	Age plus for Indonesian selectivity	Indonesian selectivity	Time-Variability LL 1 selectivity
Core	very low (1000)	Reduced sample size, Sqrt(Original ESS weights) x 5	30	Variable	As in old reference set
New Core	3.2 (reporting rate 8s)	Reduced sample size (Original ESS weights x 0.5)	30 (unchanged)	Variable (unchanged)	As in old reference set (unchanged)
Sqrt new Core	3.2 (reporting rate 8s)	Reduced sample size, Sqrt(Original ESS weights) x 5	30 (unchanged)	Variable (unchanged)	As in old reference set (unchanged)

All other assumptions not mentioned above are also unchanged compared to the original core set (defined at SAG/SC 2004).

Figure 4, Figure 5 and Figure 6 show that not much is lost in terms of projection uncertainty when using median CPUE instead of the 5 series of CPUE. In comparison with the original core (Cfullnotag) the range of values for SSB2004/SSB2000 is in fact much more in line with the outcomes from the SAG/SC 2004.

There are, however, two important points to note with regard to "new core". First, there is a tendency (lower negative log likelihood) to prefer the higher M10 values, M10=0.14 in particular (Figure 8) compared to the original grid (Figure 7). A closer look at this revealed that it was not due to the inclusion or exclusion of the tag data, but rather due to the choice of weights (square root transformation or (original weights)/2) for the size and age frequency data. Second, the choice of weights also has an effect on the projection outcomes. The version with square-root transformed weights, "sqrt new core" (still combined over the three qa options) also tends to have higher expected N for higher M10, but almost equal for values of 0.1 and 0.14 (Figure 12). Figure 9, Figure 10 and Figure 11 show that the "sqrt new core" is more pessimistic than "new core" (original weights/2), but they are reasonably similar in terms of the width of the range of uncertainty. We note that the choice of weights is rather arbitrary and that one could argue that the choice of weights could be another axis in the grid. If this was done with median CPUE the grid would be 162x2=324 (versus the original grid of 270 models).

Although the plots of projected biomass (Figures 4-6, 9-11) do not suggest substantially different width of range of uncertainty for the various possible core sets, those which include the 3 qa options clearly incorporate more underlying uncertainty, but at the same time clearly suggest that the data are much more informative with respect to M10 and Omega (Figures 7,8,12). This makes it difficult to determine what combinations would form the most appropriate core set.

We also considered whether there were versions of the OM based on the original grid which might encompass the same amount of uncertainty and display similar characteristics as the new core sets. Figure 13, Figure 14 and Figure 15 show two such sets which could be considered as alternatives. They have different specifications for sample weights and qa-assumption relative to the original core. We note that they too suggest (Figure 16) that the data are informative with respect to natural mortality and Omega. Opting for one of these core sets would minimize the need for structural changes to existing software, but it might be worth examining frequency distributions rather than just boxplots of future dynamics for whatever case is selected to ensure that strange polymodal distributions are not arising.

Summary

To summarise, based on the results presented here (mainly in Annex1), we consider that the updated tag data should be included. Reasons for this are as follows. First, we hesitate to dismiss the the 1990s tagging data because the program provides the only source of (somewhat) fisheries-independent data in the assessment, it was collected and analysed at great expense, it was intended to reduce assessment uncertainty, and the methodology underpins current ongoing SBT tagging programs. Second, the analyses done here show that the tendencies to prefer high adult mortality rates relate primarily to the annual weights chosen for the size and age frequency data rather than the inclusion or exclusion of the tag data. We do, however, recognise that the tag data appear to be informative with respect to mortality. We suggest a weight of 3.2 and use of the updated reporting rate 8.

With regard to CPUE, we consider that there is more uncertainty in the other dimensions identified above (qa range, weights for age and size samples), and if those are included in the grid, it would be acceptable to use median CPUE. In that case, it may be worth considering robustness trials for some of the individual CPUE series.

In conclusion, the main decisions the workshop needs to consider to define a final core seem to be:
1. which sample weights to use and whether sample weight should be an axis in the grid
2. whether uncertainty in the catchability age range (qa-assumption) should be included as an axis in the grid
3. whether the CPUE axis in the original grid should be reduced to the median CPUE
4. whether tagging data should be included

The choices made with respect to the above will have implications for the range of uncertainty encompassed in the scenarios and the resulting relative weights given to different hypotheses about natural mortality and Omega (i.e. the number of resamples for the grid values) in the final reference set.

Comments on SAG 2004 OM Robustness Scenarios

The robustness scenarios obviously depend on the core set that is chosen as the final set. At this stage, we simply note that, given the results and proposals made above, there are three potential changes to robustness scenarios:
1. remove qa812 from robustness (becomes one of the dimensions of the proposed new grid)
2. remove ‘tag’ from robustness (becomes part of the default core)
3. if dropping 5 cpue series and using median on grid, consider a robustness trial based on conditioning on the cpue series found to have the most extreme behaviour (e.g. CPUE 4 and 5).

Names of core sets presented in the main document

(Also see final page of Annex 1 to this paper. Note in fig. labels: **CFull6*xxx*** is the same as **CFull*xxx*** (6 indicates that the median CPUE series was also run in the former case, but only included in results as indicated). A Core set is run on the grid of 3xh, 3xm0, 3xm10, 2xOmega values and Cpue series 2-6 or median (usually indicated in graphs)

CFullnoTag – default specification from SAG 2004 – all the following specifications are identical to this one, except for features described
CFullnoTag_qa812 – run with alternative q age specification

CFullnoTag_ESSOrig2 – Original ESS weighting – all downweighted by factor of 2
CFullTag8s = CFullTagRR8s – new scaled reporting rate 8, tag weight 3.2
CFullTag8s_ESSOrig2 = CFullTag8s + 0.5*original CA/CL effective sample sizes
CFullTag8s_qa812 = CFullTagRR8s + qa812 assumption
CFullTag8s_qa812_ESSOrig2 = CFullTag8s+ qa812 assumption + 0.5*original CA/CL effective sample sizes
Other combinations of qa and ESS assumptions following pattern above

Combinations of Core sets
new core = newcore = Cfull6tag8s_qa_EssOrig2 = combination of **CFullTag8s_ESSOrig2** (implicit qa4-30)**, CFullTag8s_qa618_ESSOrig2** and **CFullTag8s_qa812_ESSOrig2.** CPUE1 at the end or 'Cmed' at the start indicate results are for median cpue.

sqrt new core = Cfull6tag8s_qa_sqrt(ESS) = combination of **CFullTag8s** (implicit qa4-30, implicit square root transformed weights)**, CFullTag8s_qa618** and **CFullTag8s_qa812.** CPUE1 at the end or 'Cmed' at the start indicate results are for median cpue.

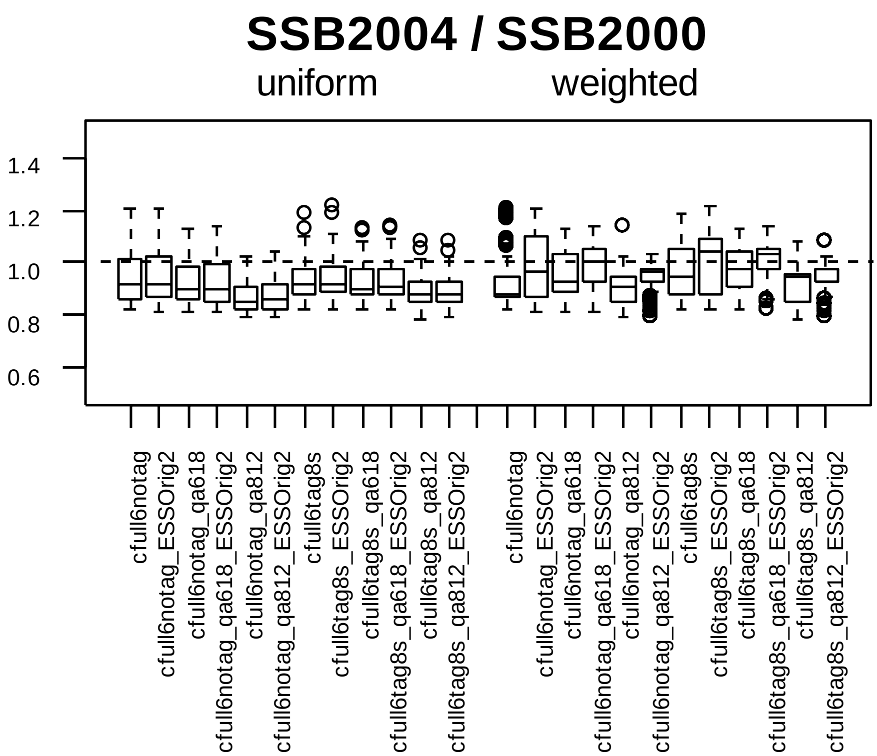

Figure 1. Boxplots of indices of recent trend under constant current catch projections comparing different SBT operating model "Core" specifications with the full "Grid" defined at the 2004 SAG. The left hand plots (uniform) are unweighted (i.e. all MPD model fittings are equally represented, except, for technical reasons, these plots are based on only the median CPUE series; projections represent the median of 25 runs). The right hand plots (weighted) are results from stochastic projections, numbers determined on the basis of priors and/or likelihoods according to the 2004 SAG proposal (including the 5 base case CPUE series). Boxes show the 25, 50 and 75%iles; whiskers show the range up to a maximum of 1.5*interquartile distance, with outliers beyond.

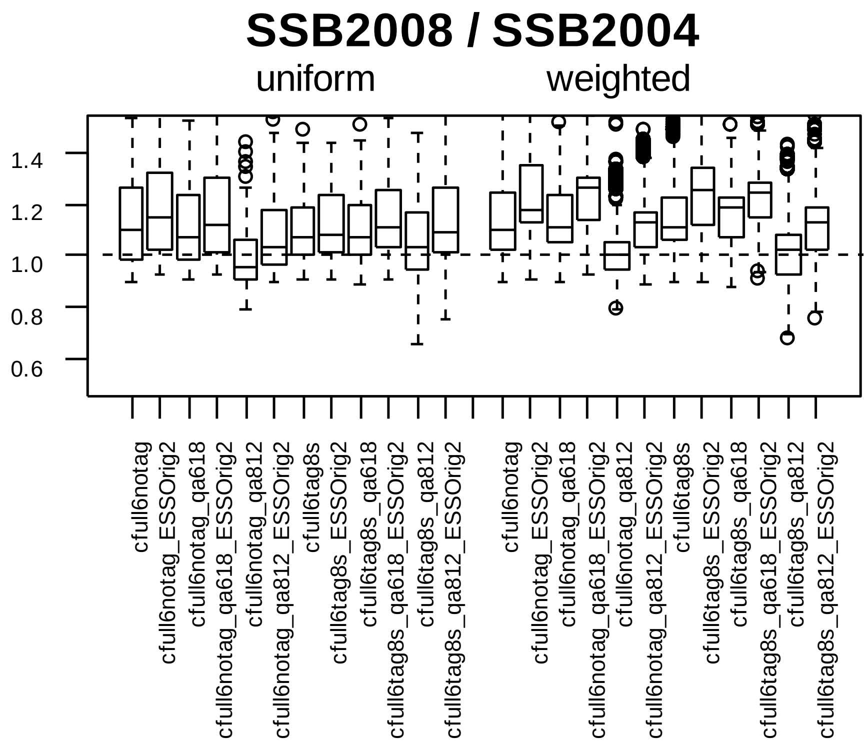

Figure 2 Boxplots of indices of S2008/S2004 under constant current catch projections comparing different SBT operating model "Core" specifications with the full "Grid" defined at the 2004 SAG. As in Figure 1.

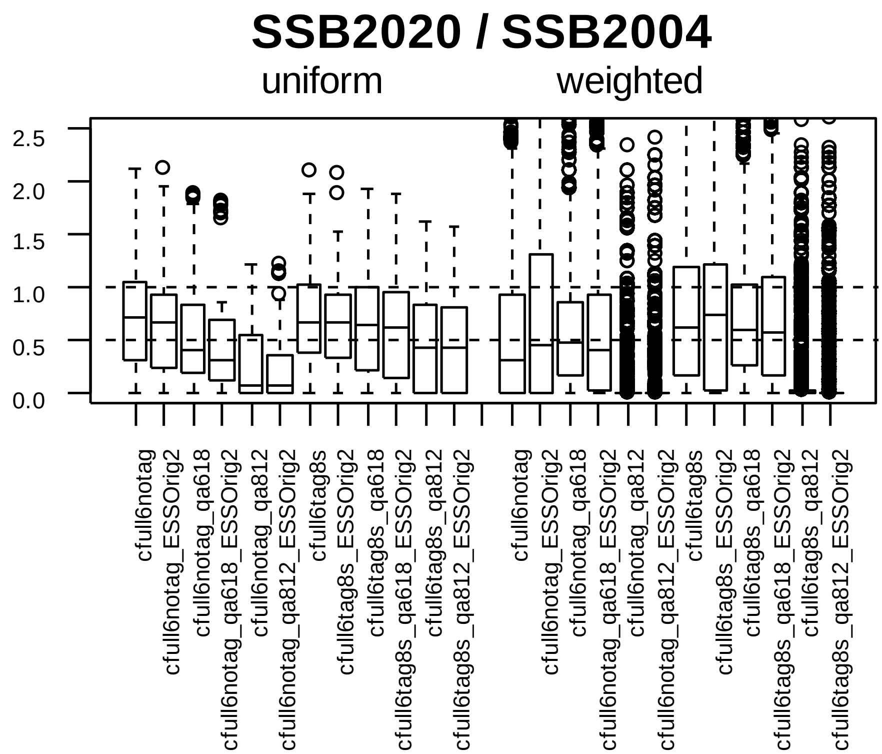

Figure 3 Boxplots of indices of S2020/S2004 under constant current catch projections comparing different SBT operating model "Core" specifications with the full "Grid" defined at the 2004 SAG. As in Figure 1.

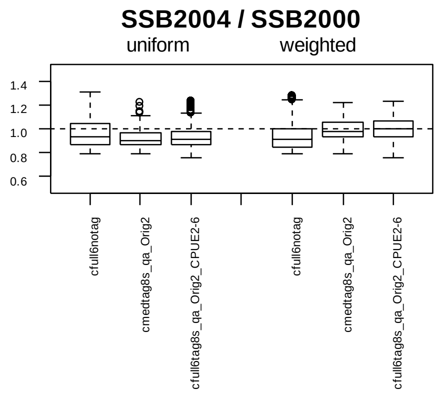

Figure 4. Boxplots of indices of recent relative spawning biomass under constant current catch projections comparing the original core set (cfullnotag) with the "new core" set based on median cpue and "new core" based on the 5 separate cpue series (2-6).

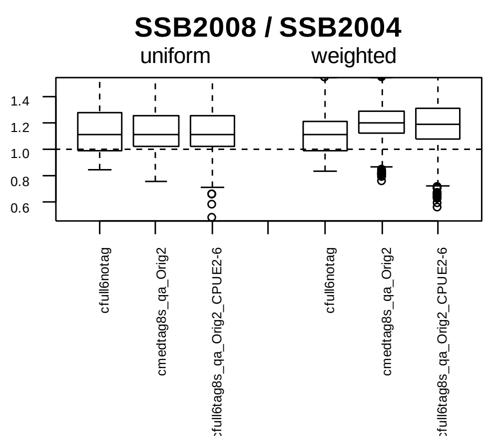

Figure 5 Boxplots of indices of S2008/S2004 under constant current catch projections comparing the original core set (cfullnotag) with the "new core" set based on median cpue and "new core" based on the 5 separate cpue series (2-6).

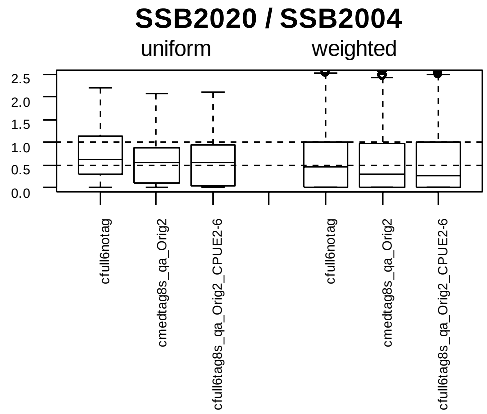

Figure 6 Boxplots of indices of S2020/S2004 under constant current catch projections comparing the original core set (cfullnotag) with the "new core" set based on median cpue and "new core" based on the 5 separate cpue series (2-6).

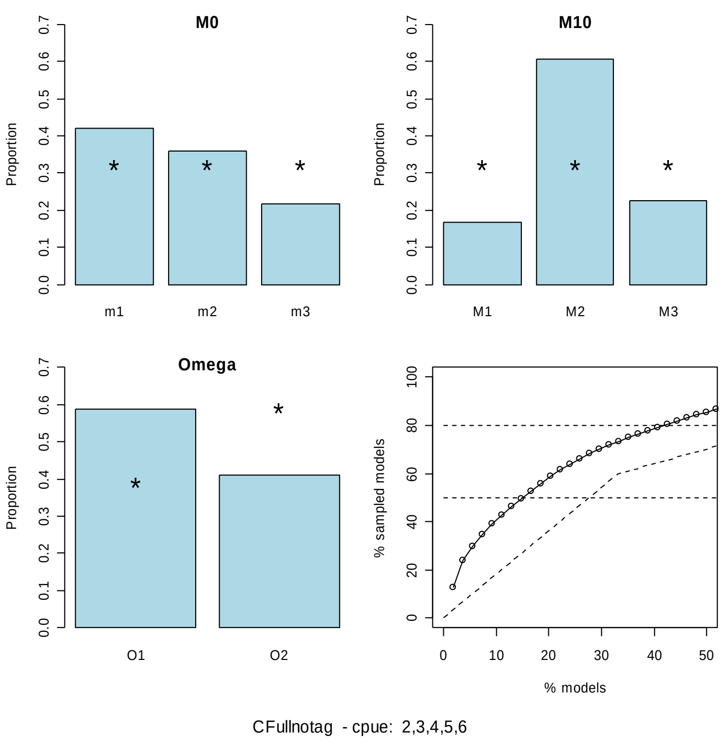

Figure 7 Expected proportion of resamples wrt m0, m10 and Omega (asterisks = priors), and the cumulative percentage resamples vs the % models the resamples are taken from, for the original core set. In bottom right plot, the sloping dashed reference line represents the curve that would be realized if all model likelihoods were equal (change in slope represents the effect of non-equal steepness priors)

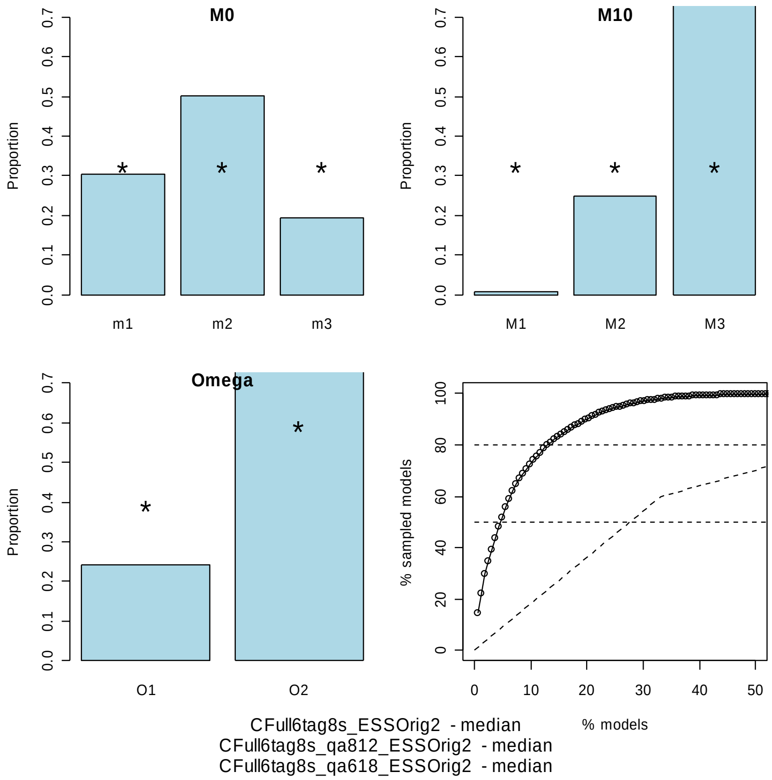

Figure 8 Expected proportion of resamples wrt m0, m10 and Omega (asterisks = priors), and the cumulative percentage resamples vs the % models the resamples are taken from, for "new core"

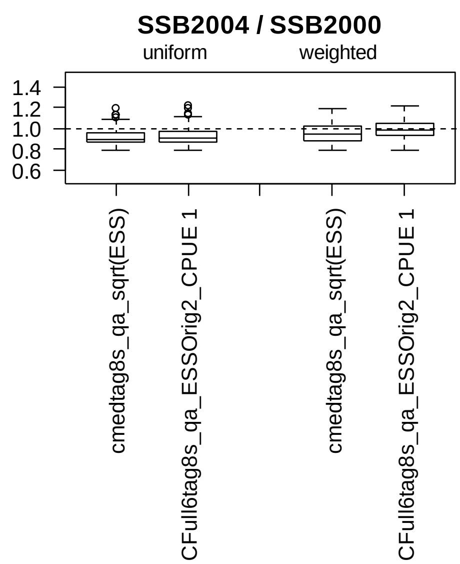

Figure 9 Boxplots of indices of recent relative spawning biomass under constant current catch projections comparing the "sqrt new core" set with the "new core" set.

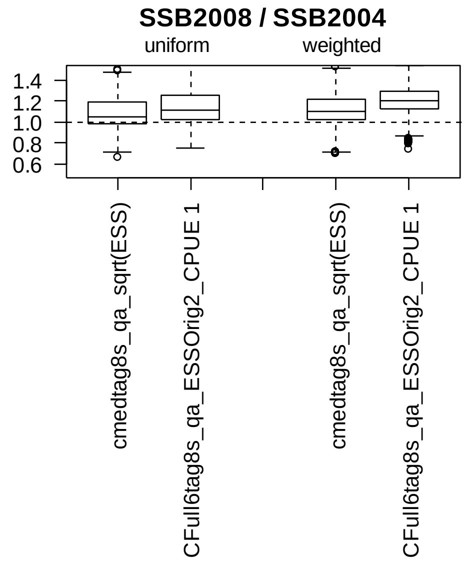

Figure 10 Boxplots of indices of S2008/S2004 under constant current catch projections comparing the "sqrt new core" with the "new core" set.

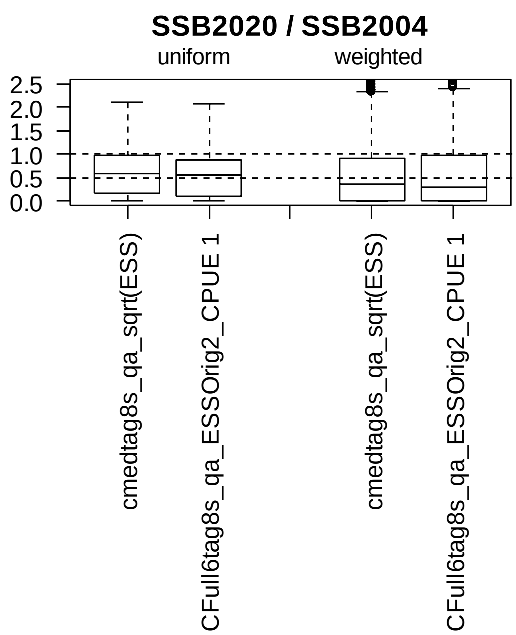

Figure 11 Boxplots of indices of S2020/S2004 under constant current catch projections comparing the "sqrt new core" with the "new core" set.

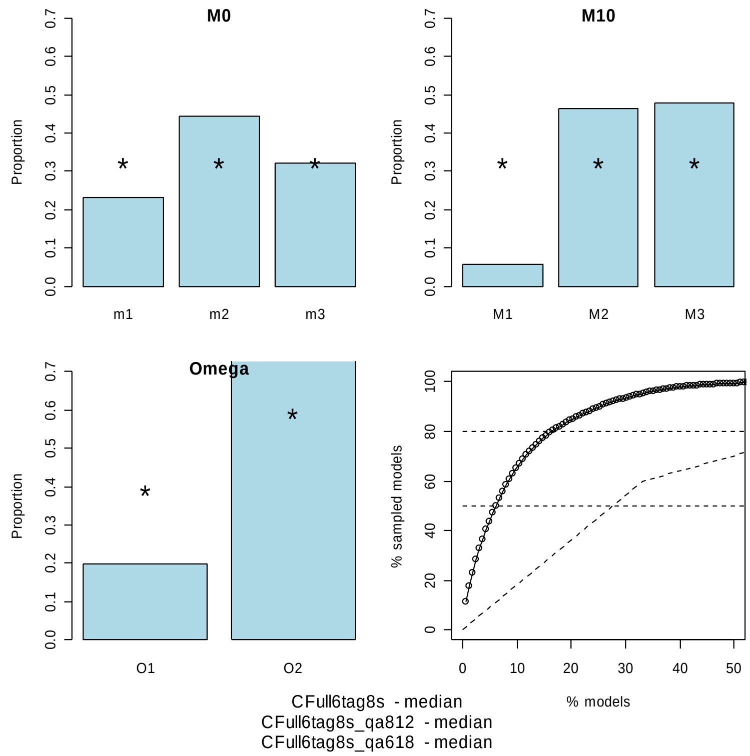

Figure 12 Expected proportion of resamples wrt m0, m10 and Omega (asterisks = priors), and the cumulative percentage resamples vs the % models the resamples are taken from for "sqrt new core"

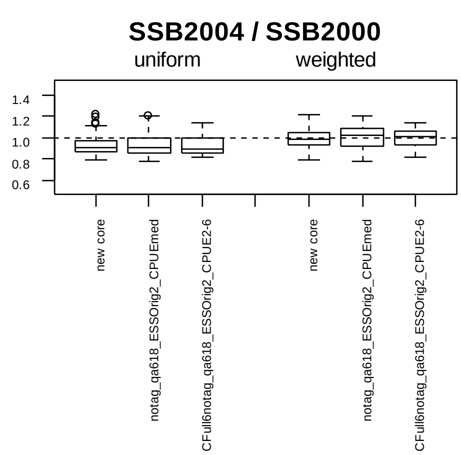

Figure 13 Boxplots of indices of recent relative spawning biomass under constant current catch projections comparing the "new core" set based on median cpue, and two similar alternatives which use qa6-18, original weights/2 (no tags) and either median CPUE or the 5 separate cpue series (2-6). See text.

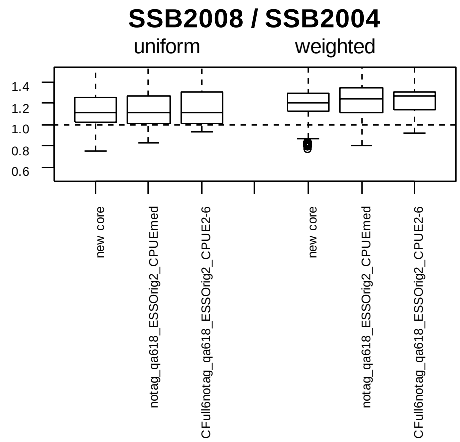

Figure 14 Boxplots of indices of S2008/S2004 under current catch projections comparing "new core" , and two similar alternatives with qa6-18, original weights/2 (no tags) and median CPUE or the 5 separate cpue series (2-6).

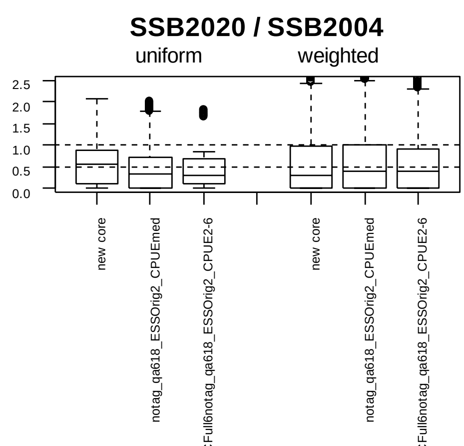

Figure 15 Boxplots of indices of S2020/S2004 under current catch projections comparing "new core" , and two similar alternatives with qa6-18, original weights/2 (no tags) and median CPUE or the 5 separate cpue series (2-6).

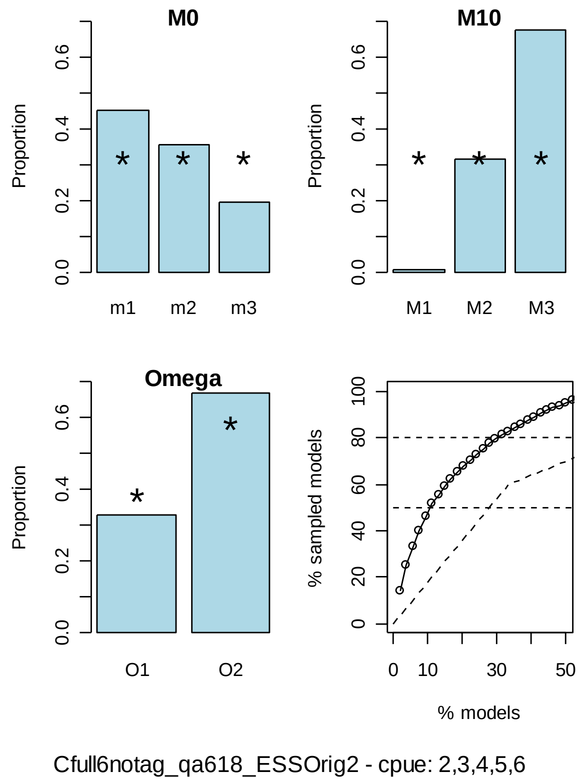

Figure 16. Expected proportion of resamples wrt natural mortality and Omega (asterisks indicate priors), and the cumulative percentage resamples plotted against the % models the resamples are taken from.

CSIRO
MARINE RESEARCH

ANNEX 1 to CCSBT-MPTM/0502/04: Details of OM explorations.

Exploration of the SBT operating model with implications for the selection of the core set and robustness trials.

D. Kolody, M.Basson, A. Preece, J. Hartog, T. Polacheck

**Prepared for the CCSBT Special Management Procedure Technical Meeting (MPTM)
15-19 February 2005, Seattle, USA**

ANNEX 1: Details of OM explorations.

Exploration of the SBT operating model with implications for the selection of the core set and robustness trials.

Natural Mortality Assumptions

Although we are not proposing a change to the way in which natural mortality is dealt with in the core set, we discuss some results from further investigation of the mortality assumption here. There seems to be some evidence for an increase in natural mortality among older age individuals in some species, such that it is conceivable that the assumption of M(age 11+) = M(a 10) is poor for SBT (perhaps due to spawning-related mortality, senescence or differences between sexes). We explored the implications of a couple of different M assumptions. This required small changes to sbtmod14, sbtProj112 and associated files.

The M25 assumption has M increasing linearly from age 25 to 30 by 0.05/y. This results in an improved overall likelihood over the flat M10+ scenario in most or all cases, seemingly driven primarily by the improved fit to the Indonesian fishery catch-at-age. In some cases, this assumption results in a relatively flat Indonesian selectivity, which has a certain appeal. But spawning ground selectivity is more often dome-shaped or monotonically increasing, with flatness a finely balanced transition point between the two (Figure 1).

The M18_Ind18 scenario was defined to satisfy the desire for a flat Indonesian selectivity, while admitting that there is an interaction with M. In this scenario, a linear trend in natural mortality from ages 18-30 was estimated as a free parameter, while Indonesian selectivity was constrained to be flat at 18+. The magnitude of the estimated slope of M18-30 was always positive, with magnitude inversely related to M10. Note that not all factors and projections were run for this scenario, and the model failed to converge in a couple scenarios with CPUE4 (ST window), comparative plots below are restricted to historical trends for CPUE(2,3,5,6).

Figure 2 shows the distributions of marginal likelihoods for (most of) CfullnoTag, M25 and M18_Ind18 (ignoring issues related to changes in numbers of estimated parameters). M25 appears to have the best fit in general. M18_Ind18 actually appears to have the worst fit to the spawning ground data (not shown), and does not represent much (if any) improvement from CfullnoTag (the increased freedom in M18+ estimation does not offset the constraint in selectivity for ages 18+). It might be worth noting that the relative likelihood (and hence weightings) for the different factors changes somewhat between CfullnoTag (favours low steepness and medium M10) and M25 (favours medium steepness and low M10). The unweighted CFullNoTag scenario is more pessimistic in terms of current biomass trend than M25 and M18_Ind18 (latter not shown), Figure 3.

One of the main reasons for exploring a U-shaped mortality curve, relates to the Indonesian fishery selectivity, which usually suggests substantial changes among older ages, despite their similar sizes and presumably behaviour. One way of quickly evaluating the shape for 324 models (6 cpue series) is to characterise it by ratios of selectivity at age 30 and age 18 = s30/s18, and s30/s28. Values greater than 1 can potentially mean an increasing selectivity curve. Out of 324 models, 81 (25%) had s30/s18>1 and 71 (22%) had s30/s28>1. On further inspection, we found that the expectedN of models with s30/s18>1.5 were between 1% and 14%, depending on the cpue series used.

Other M10+ scenarios were briefly explored, including estimating M11-M30 in two linear segments. The minimization was unsatisfactory in that the function minimizer could not locate objective function values as good as those known to exist from the explorations above; when given the best known values as a starting point, M estimates sometimes approached a lower bound of zero for some age classes.

Overall, we are hesitant to recommend expanding the MP operating model uncertainty to include alternative M(age 11+) assumptions because:

- The new alternatives explored are also arbitrary, and we do not think the estimation is reliable
- The medium to long term projected biomass indices suggest that the range of uncertainty is not extreme relative to other models explored
- The U-shaped mortality curves tested did not entirely solve the problem of dubiously variable Indonesian selectivity in older age classes

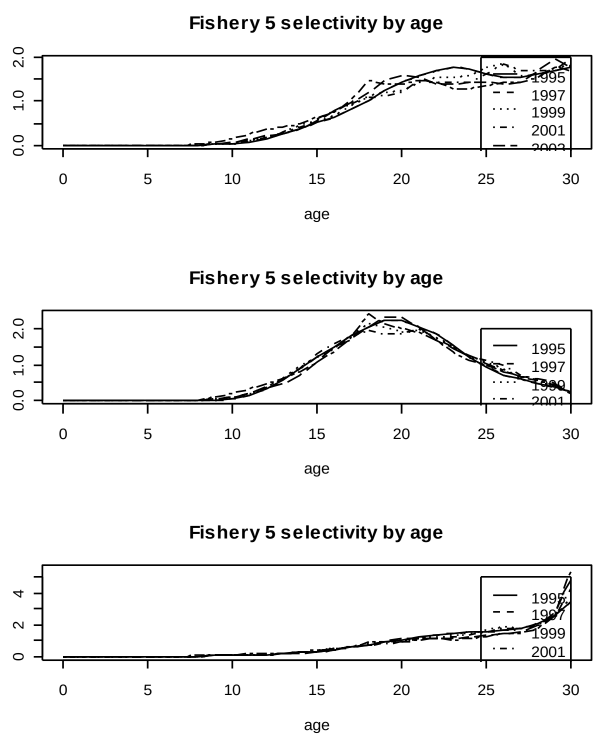

Figure 1 Typical Indonesian longline selectivity estimates from CFullnoTag_M25 (h3m1M3O2C6, h3m2M1O1C2, h3m3M3O2C3, respectively)

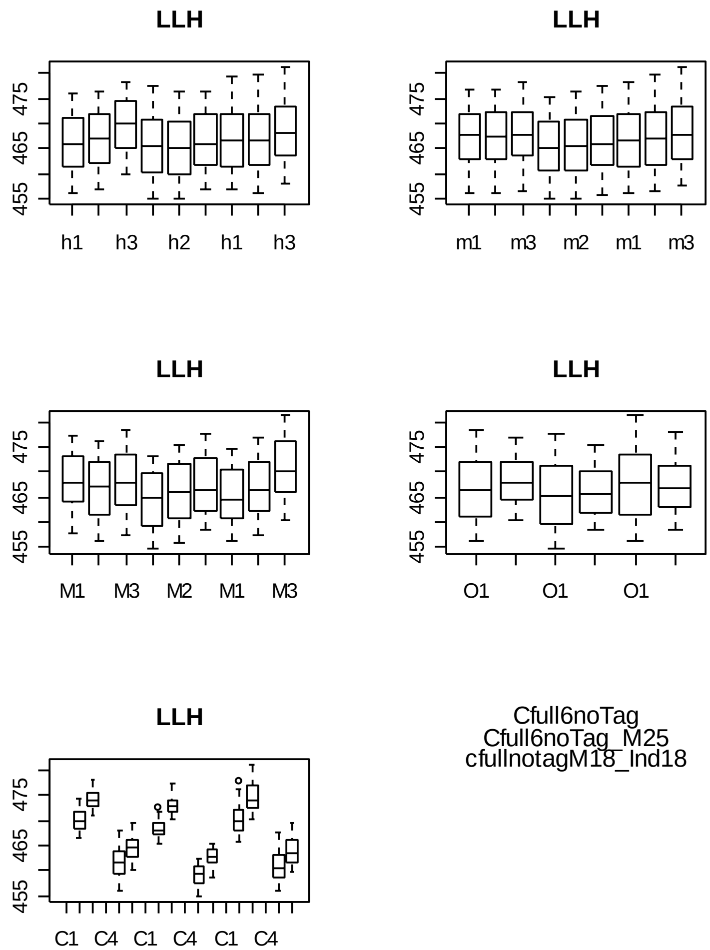

Figure 2. Negative log likelihood shown by the main factors in the grid (steepness, m0, M10, Omega and CPUE); only for CPUE series 2,3, 5 and 6 (due to convergence failures for series 4 in the CfullnotagM18_Ind18 runs). Values only directly comparable between factors within a model specification.

NB How to read these figures: 'Boxes' are shown for each level in the factor for the first model, then each level for the second model etc. For example, the panel for h (top left) is in sequence: first 3 are for Cfull6noTag with h1=0.385, h2=0.55, h3=0.73; the next three are for Cfull6noTag_M25 with h1, h2, h3, etc. Other panels are interpreted in the same way remembering that there are only 2 Omega values and 5 or 6 CPUE values (only 4 shown in figure 2 as explained in the legend).

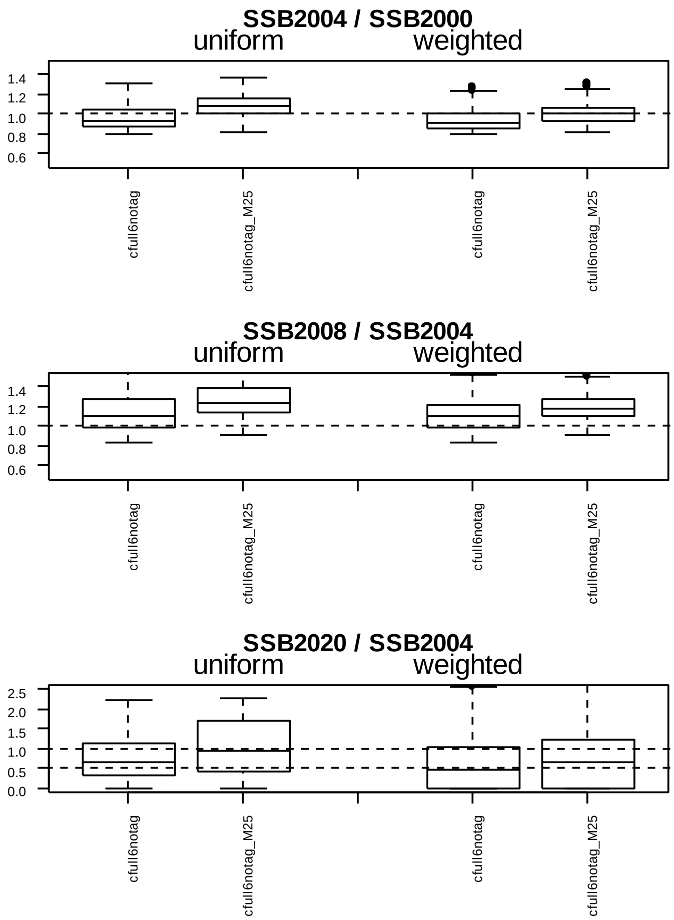

Figure 3. Comparison of current trend and projection results for Cfull6noTag and Cfull6noTag_M25 grid sets. Left side of the panel represents all MPDs weighted equally; right panel indicates the 2004 SAG weighting scheme with priors and/or likelihoods.

Does the CCSBT MP Conditioning Grid need to include 5 proposed CPUE Series?

The 5 different interpretations of CPUE as a relative abundance index (i.e. so-called nominal, Laslett core areas, Space-Time window, proxy B-ratio (w5) and proxy geo-statistical (w8)) appear by eye to be very similar, and given that there was no strong argument why any one should be preferable to the others, the median was initially accepted for the operating model conditioning. Explorations prior to and during the 2004 SAG indicated that the conditioning model was actually sensitive to the choice of CPUE series, as was resultant projections and MP performance. On the basis of these observations, it was decided that the OM should encompass the uncertainty represented by all 5 series. There are at least two reasons why we would prefer to return to the original decision of one CPUE series:

- The actual MP will be implemented with a single series (initially the median), but there is no plan to account for the inconsistency that arises between the conditioning to one series and implementation of another.

- In the current Grid specification, the multiple CPUE series accounts for the greatest proportion of the dimensionality of the current Grid, and we would argue that other sources of uncertainty would be more appropriate.

Figure 4 indicates the (negative-log) likelihood values of the CFull6_noTag MPDs, marginalized over the different factor levels. The likelihoods (here and in all other core sets examined) indicate that w5 and w8 are a better fit to the data than the other series, and the median is usually nearer these two than the others. The variability in likelihoods among CPUE series is much greater than the other factors. However, if short-medium term current catch projections are compared in the same manner, the variability among other factors (steepness and M10) appears to be more substantial than that due to CPUE (Figure 5 - 7). A similar comparison across a range of different plausible core specifications also suggests that the uncertainty due to factors (steepness and M10) is greater than the core specification (Figure 5 -7). However, we are left with the impression that CORE specifications are the dominant feature when it comes to actually comparing the weighted projections (Figure 8).

While we cannot deny that the choice of CPUE series is a source of uncertainty, we expect that it is of less importance than other factors under consideration, consumes a large portion of the practical grid uncertainty space, and creates a logical inconsistency between testing and implementation (i.e. where only one series can be used). However, we also recognize that the implications were not tested on an actual MP with dynamic TAC setting, and hence suggest that concerns might be addressed in sensitivity trials related to the most extreme CPUE series (e.g. CPUE4 and CPUE 5).

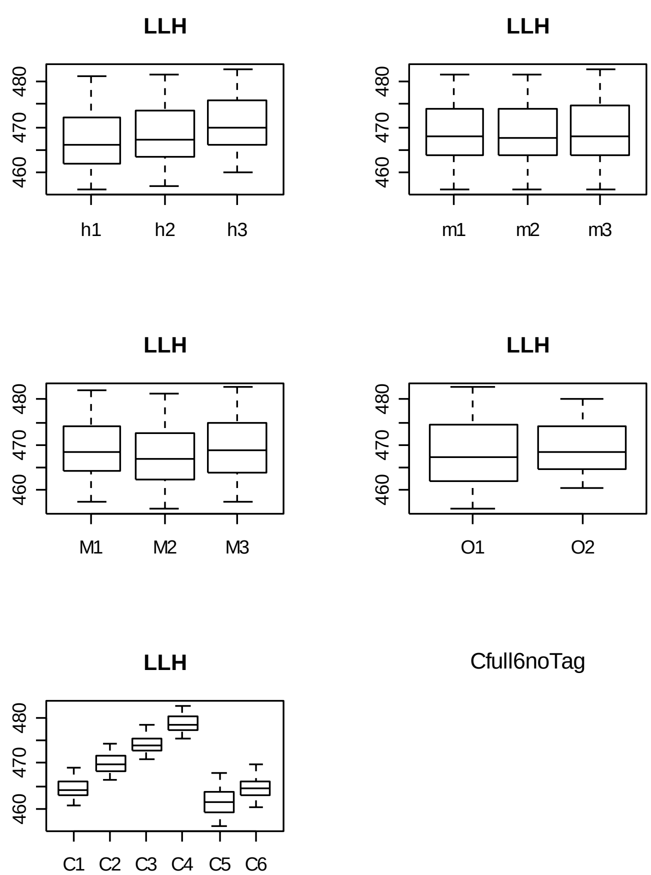

Figure 4 Negative log likelihood maginalized over the main factors in the grid (steepness, m0, M10, omega and cpue) for the Cfull6notag MPD fits (i.e. all cpue series, including the median as C1 in this case).

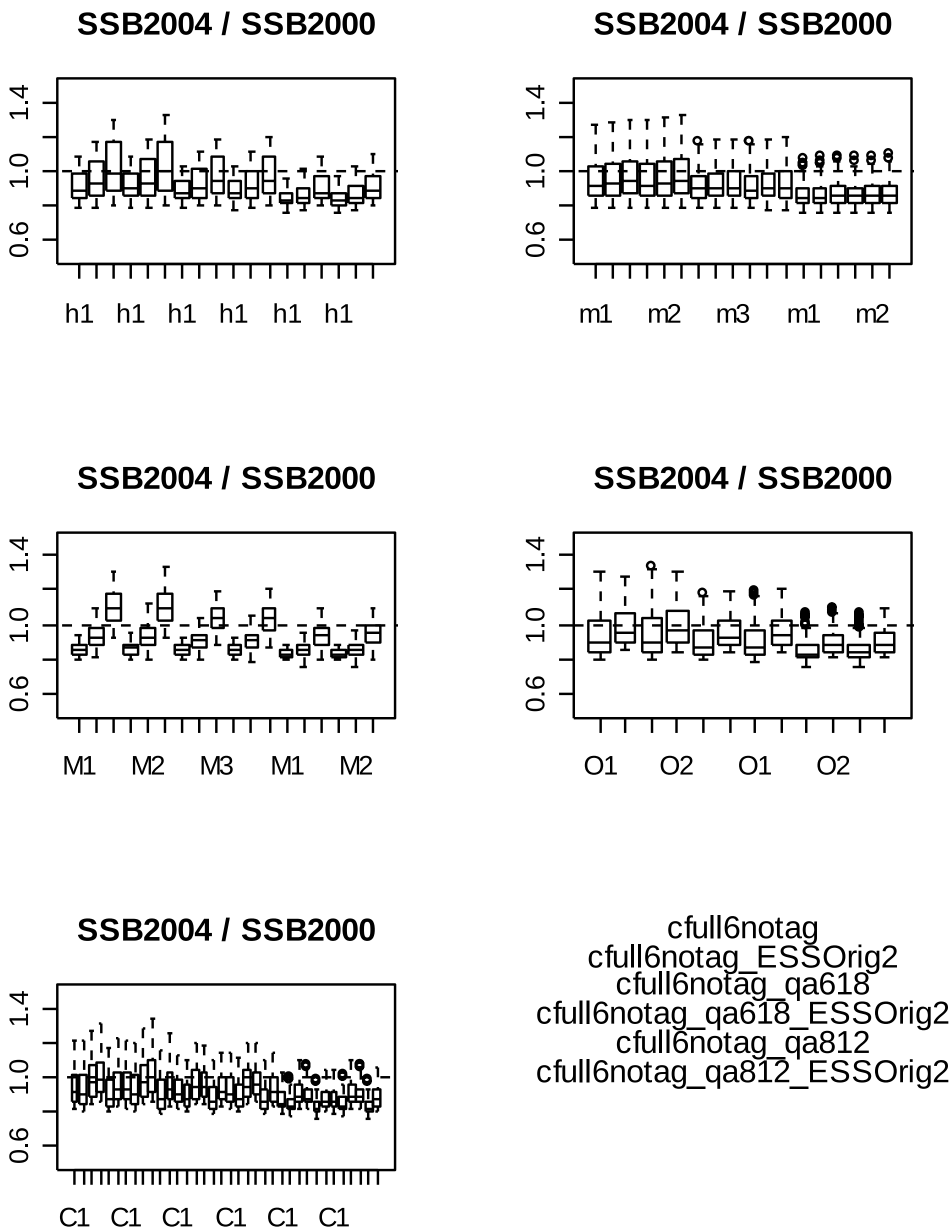

Figure 5 Unweighted projection results in terms of SSB2004/SSB2000 for a range of Core sets and all 6 CPUE series (C1=median). Each factor level is plotted for the first Core set (Cfull6notag), then the second core set (Cfull6notag_EssOrig2) etc, in the order listed in the blank panel.

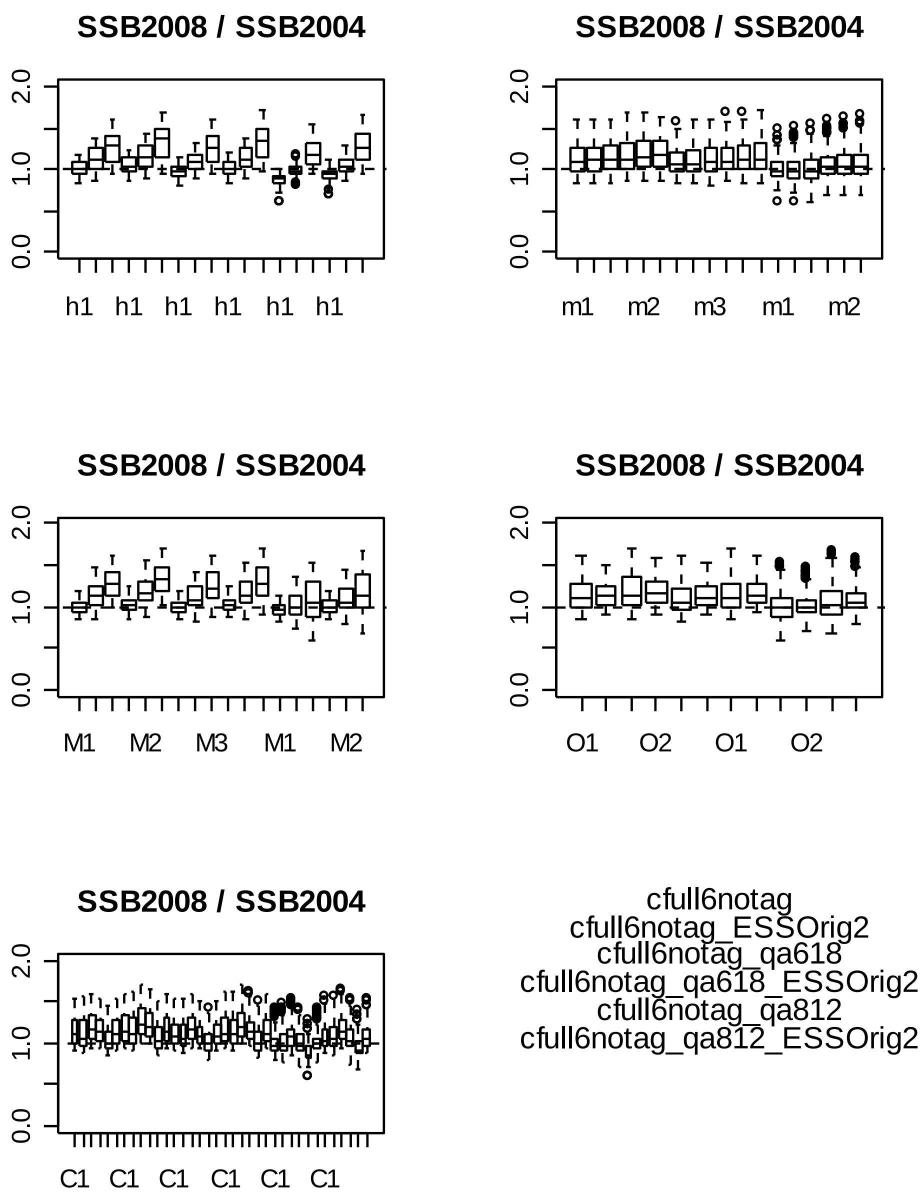

Figure 6 Unweighted projection results in terms of SSB2008/SSB2000 for a range of Core sets and all 6 CPUE series (C1=median). Each factor level is plotted for the first Core set (Cfull6notag), then the second core set (Cfull6notag_EssOrig2) etc.

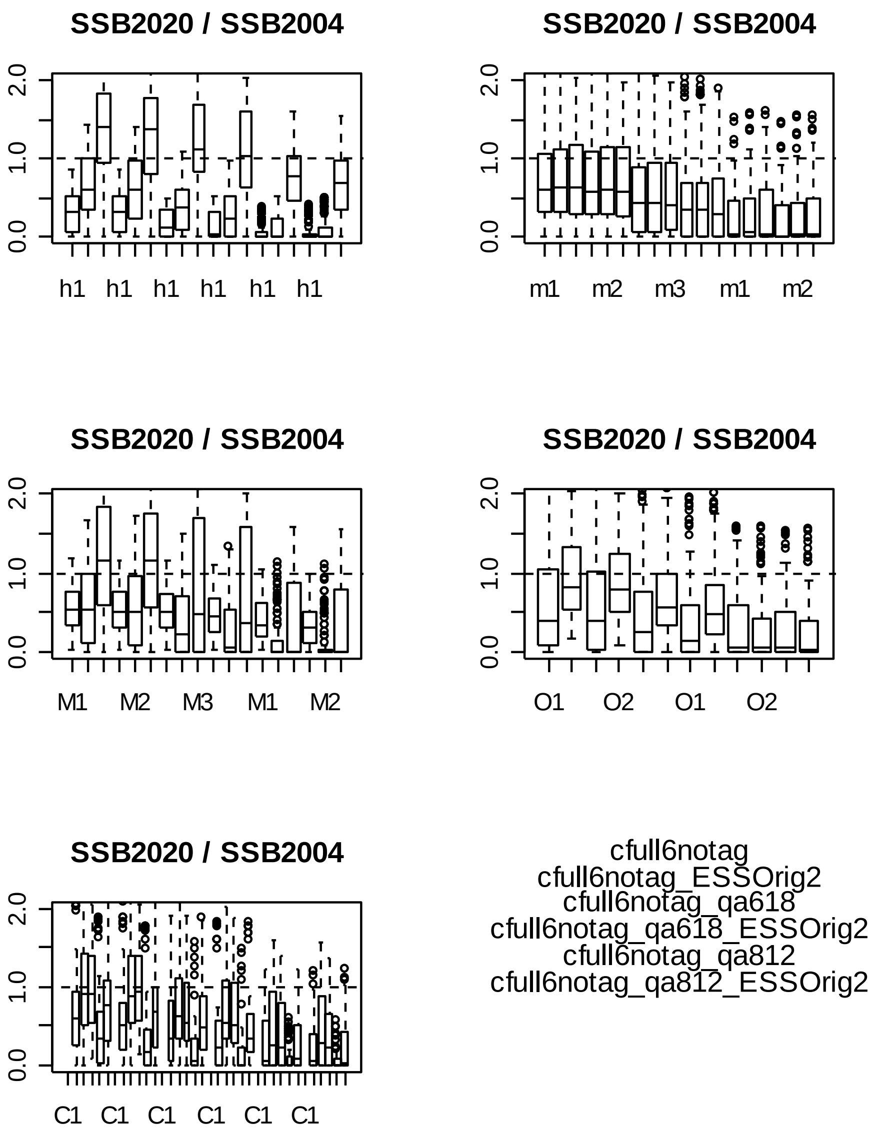

Figure 7 Unweighted projection results in terms of SSB2020/SSB2000 for a range of Core sets and all 6 CPUE series (C1=median). Each factor level is plotted for the first Core set (Cfull6notag), then the second core set (Cfull6notag_EssOrig2) etc.

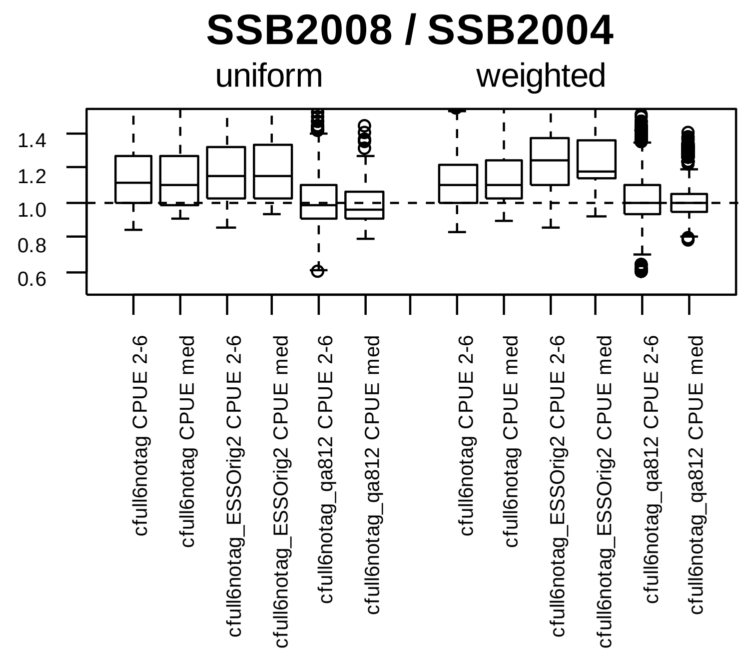

Figure 8 (a) Current trend and projection (b below) indices comparing a range of core sets and the differences between using CPUE series 2-6 and the median.

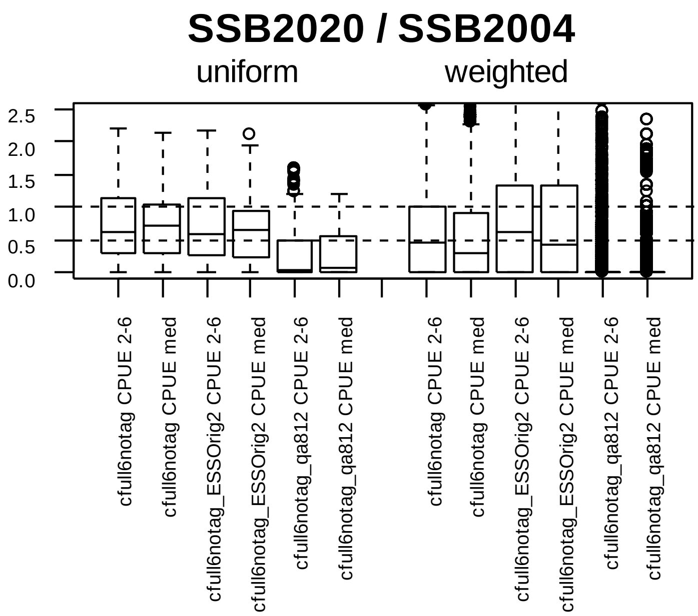

Figure 8b

Assumptions about the age range relating CPUE to abundance (i.e. the so-called qa8-12 vs 4-30 debate)

We cannot recall any explicit reason why the qa4-30 assumption (age range for catchability standardization) became the default, except in relation to the suggestion that the qa8-12 assumption was so pessimistic that it would automatically put the MP into the realm of exceptional circumstances if it was correct. If this latter assertion is actually true, it follows that there is already an implicit lower bound to what we are willing to accept as plausible, but this is an assertion that we are not comfortable with. The following plots illustrate performance of qa assumptions 4-30, 8-12 and intermediate specification 6-18. From these plots we observe:

- Qa8-12 provides the best fit to the data in terms of the likelihoods; (qa6-18 performance is intermediate,)
- Qa8-12 performance in the most pessimistic; (qa6-18 closely resembles qa4-30)

We find it difficult to justify ignore the uncertainty of this rather arbitrary assumption, in the manner proposed in the 2004 SAG and recommend that it should be promoted to the CORE.

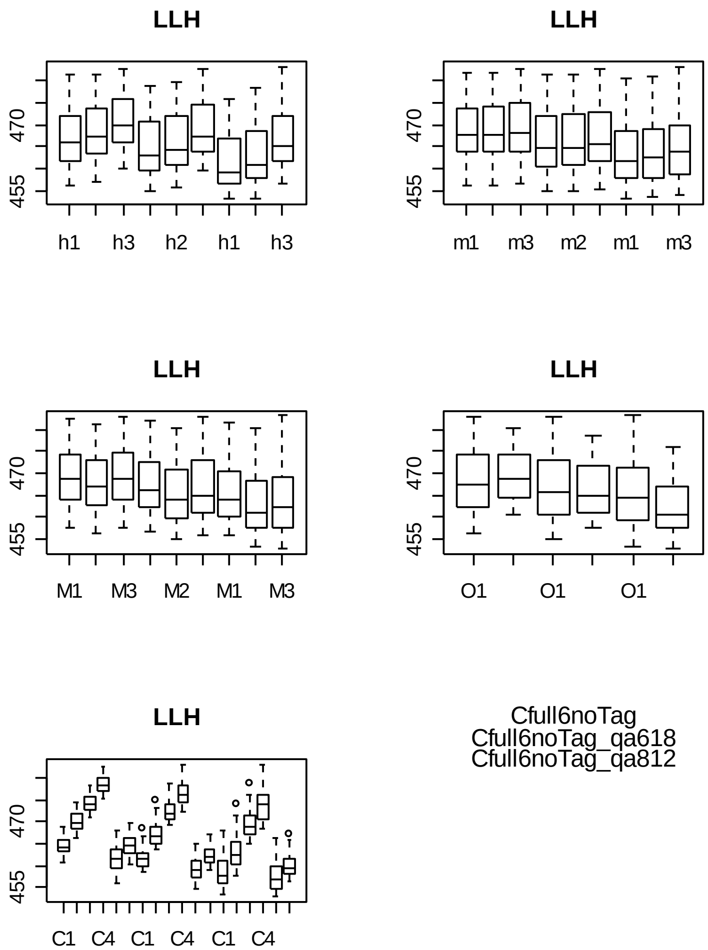

Figure 9 Negative log likelihood shown by the main factors in the grid (steepness, m0, M10, omega and cpue) for 3 core sets and all cpue series (C1=median CPUE).

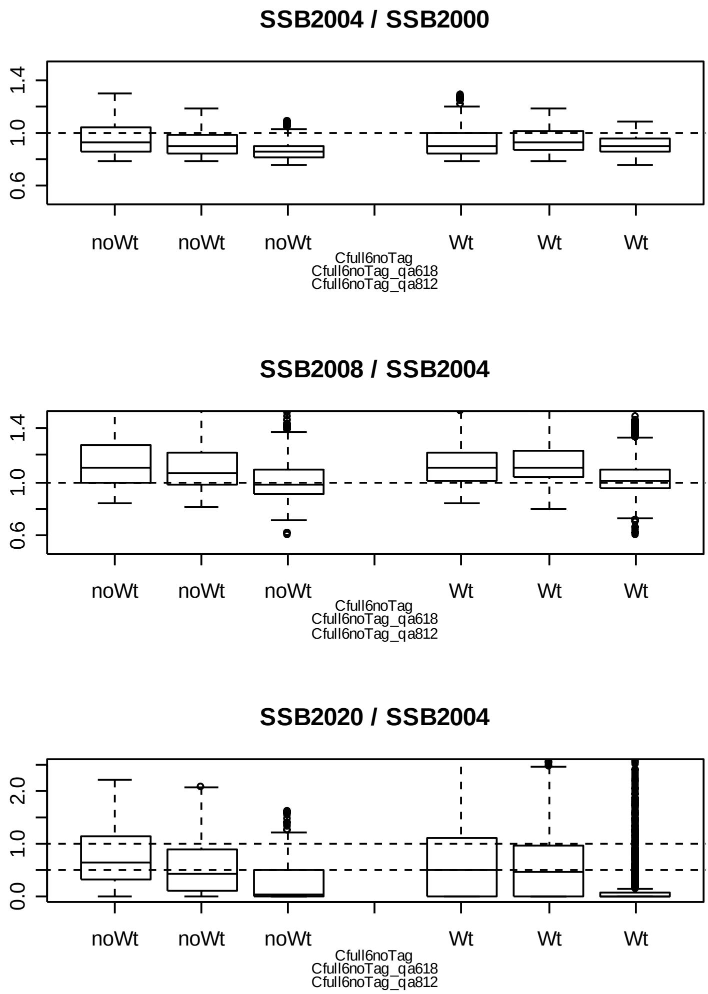

Figure 10 Unweighted (noWt) and weighted (Wt) projection indices for 3 core sets, cpue 2-6. Note that cfull6notag corresponds to the qa4-30 assumption.

Uncertainty in Projected CPUE

Figure 11 illustrates the distribution of projected CPUE trends over the short-medium term (under current catch) for a number of core sets. In all cases, the short term trend uncertainty looks roughly consistent with a flat trend and observation error CV around 0.2, and can adequately encompass a fairly large change in CPUE between 2003 and 2004. Of more concern perhaps is the low probability of CPUE increases being by 2008, particularly given that the majority of scenarios predict increasing biomass over the same period. Changes in selectivity might potentially be a concern.

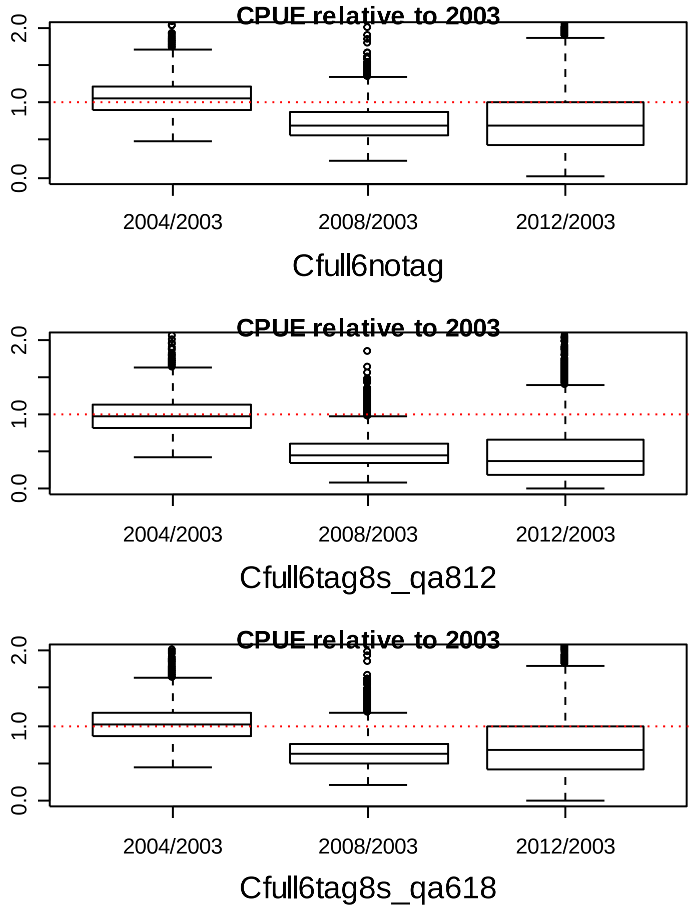

Figure 11 Projected CPUE relative to that in 2003 for 3 core sets. Results are unweighted and combined over CPUE series 2-6

Selectivity and CACL sample size Assumptions

There are quantitative analyses underpinning time series of effective sample sizes for some fisheries, but there remains the problem of how these data should be weighted relative to other components of the likelihood. Furthermore, there is a recognition that it may be desirable to adjust the relative effective sample sizes within the time series that are underpinned by analyses, to compensate for model structural limitations. In the case of SBT, there is a long running debate about how to handle the trade-off in variable selectivity over time and limited catch-at-length sampling. We considered the following 4 core set specifications:

1. the basic core set which uses the sqrt(5*original ESS) panel transformation (Cfullnotag; Cfull**6**notag is identical but means it was also run for the median CPUE)
2. the basic core set but with the ORIGINAL weights used prior to the panel's modification (Cfull6notag_ESSOrig)
3. the basic core set but with the ORIGINAL weights divided by 2 (Cfull6notag_ESSOrig2)
4. the basic core set (with panel weights), but with selectivity in fisheries 1 and 6 allowed to change every second year throughout the time series (Cfull6notag_Sel2)

The reasoning behind the ‘panel weights’ come from the Independent Panel’s document CCSBT-ESC/0409/42 (2004) which notes: 'We considered that the sample sizes for some of the size composition data in the final years (eg n=500 for LL1) were too large to be used in conjunction with constrained changes in selectivity as assumed in the model. They proposed using the square root of *n*, times 5 for all size and age frequency data, noting that this transformation reduces the overall sample sizes, retains some of the relative magnitude and reduces the contrast in sample sizes over time. We consider that the pattern of sample sizes over time is based largely on data, that the transformation unduly distorts this pattern. We therefore consider the alternative of the original weights divided by 2 to reduce the sample sizes (option 3 above) and to address the concern expressed above. Option 2 (original weights unchanged) are shown for comparison.

Figure 12 shows the projection implications of the 4 different plausible CA/CL effective sample size weightings and temporal variability in selectivity. In most cases, there are predictable impacts on model fit to data (i.e. higher ESS and more selectivity freedom results in a better fit to the CA/CL data). Less direct implications were not investigated. Figure 12 indicates that the default specification CFullnoTag is generally the most pessimistic of the 4 in terms of projections, particularly when likelihood weightings are invoked.

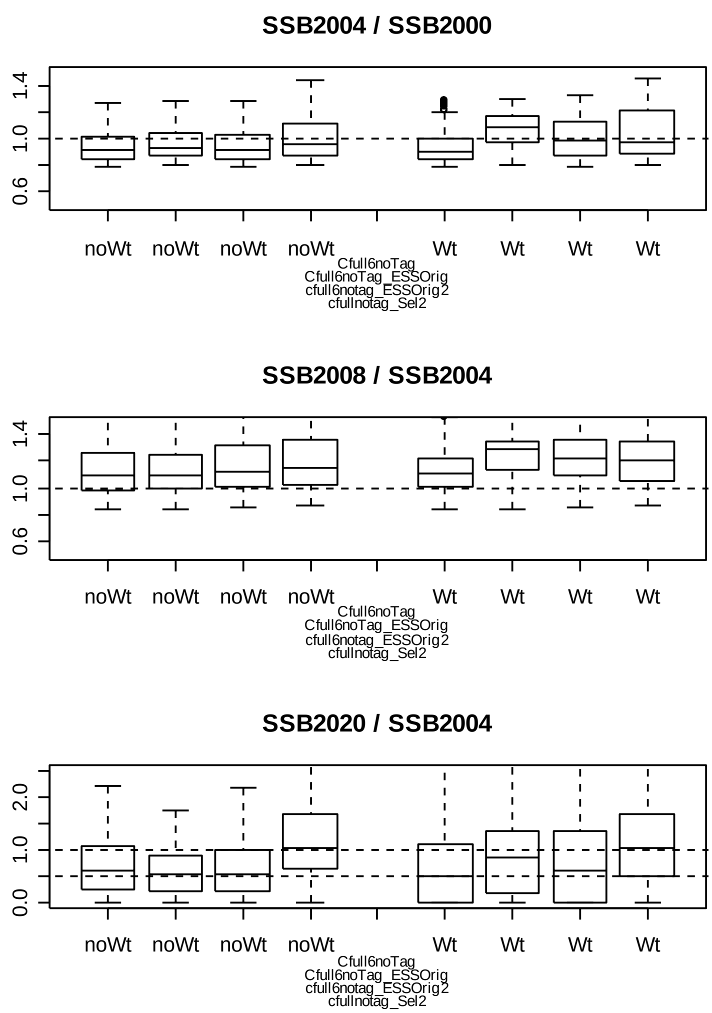

Figure 12 Unweighted (noWt) and weighted (Wt) projection indices for 4 core sets, cpue 2-6.

Inclusion of Tagging Data

We hesitate to dismiss the the 1990s tagging data because the program provides the only source of (somewhat) fisheries-independent data in the assessment, it was collected and analysed at great expense, it was intended to reduce assessment uncertainty, and the methodology underpins current ongoing SBT tagging programs. The reasons for discarding the tags (discussed at CCSBT SAG/SC in 2004) related to the suspiciously high precision with which M10 seems to be estimated when these data are included, as well as the notion that the estimates themselves were unrealistically high (in the judgement of some) at around 0.17. To further investigate this issue, we explored a number of avenues:

- Update tag release/recovery data to more accurately reflect the SBT operating model fishing seasons
- Update historical tag reporting rates from a new analysis to also more accurately reflect the SBT operating model fishing seasons
- Explore the implications of alternative tag weightings in the objective function
- Explore the sensitivity of results to different reporting weights
- Examine interactions among tag data and other model asssumptions

<u>Update tag release/recovery data</u>
The updated tag release/recovery data were distributed through the CCSBT Secretariat after the SAG/SC meeting in September 2004. These data are in the file sbtdata2.dat, and we propose that these are more appropriate to use than the old data (in sbtdata1.dat), because they are compatible with the SBT OM fishing seasons. Comparisons between the two datasets showed no substantial differences in results though the new data appear to fit somewhat better.

<u>Update historical tag reporting rates</u>
Updated reporting rates are presented in another document ("Updated estimates of tag reporting rates for the 1990s tagging experiments." J. Paige Eveson, Tom Polacheck. CCSBT-MPTM/0502/05). Reporting rates have were updated to match year definitions in the operating model (1st Nov-31stOct), and to use catch and size data used in the 2004 assessment (much of which was updated). Implications of the new reporting rates are considered below.

<u>Implications of alternative tag weightings in the objective function</u>
Note runs in this section were done with the new release/recover data (sbtdata2.dat), and the original reporting rates. Further investigation with tags included vs heavily downweighted revealed the following:

- tags included only has a preference for higher m0 and/or m10 with SOME components in the objective function (e.g. LL3 size frequency, Indonesian age frequency to some extent, CPUE and the penalty terms for selectivity change and sigmaR in the stock-recruit relationship)
- tags included has a preference for LOWER m values for other components (LL1, LL4 sizze frequencies and Surface fishery age frequency.)

Irrespective of whether tags are included or not, plots of objective function values for individual components at the different grid points (Omega, M10, m0, Steepness) show that some components are in apparent contradiction to one another. For example, LL1 size data and the recruitment penalty term favour low steepness, whereas Indonesian age data and LL3 size data favour high steepness. This is not a new observation, but again underlines how important the relative weighting of components are for the overall fit. In the overall fit with relative weightings as used in the most recent core sets, the runs with tags included do tend to prefer the higher m0's and mid-level m10s. These statements relate particularly to weights of 3.2 and the higher weight of 1.6 (which was used in exploratory work) for the tag component. However, runs with tag data included did not exclusively favour the highest mortality rates (see e.g. Figure 14).

In cases where the runs with tags had a higher objective function value (poorer fit) for the size/age frequency components, these components did not show any obvious misfit to the data. The tag and notag runs are almost indistinguishable by eye when looking at plots of fits to the size/age data or even residuals. The two components where the poorer fits become apparent from plots are in the CPUE (higher AC in residuals and a bit of an overall trend for with tag runs) and in the stock-recruit curve (again, higher AC).

The main problem with putting tags in at weights previously considered (ie 1.6 or 3.2) in the context of the core set for MP evaluation, is that tags are apparently informative with respect to mortality, m0 and m10. This makes sense, and is not surprising, but it means that the subsequent sampling over the grid can be less ‘spread out’ than for the notag runs (see below). In this particular context, one aim is to get reasonable coverage of the grid which is considered to span the plausible range of values. This can be remedied by further downweighting the tag data (e.g. 12.8) without weighting it ‘out’ entirely. Further consideration of this issue is given below where the new reporting rates have been explored.

<u>Sensitivity of results to different reporting weights</u>
A wide range of reporting rates (old and new) were run to compare effects and assess the sensitivity of results. Plots include results for old reporting rate 5, RR5 (i.e in CFullTag), old reporting rates 1 and 8 (RR1, RR8) and new reporting rates 1, 5 and 8 scaled (RR5s in CFullTagRR5s, RR1s and RR8s. Based on the discussion in Eveson and Polacheck (CCSBT-MPTM/0502/05), we consider that RR8s is the most “realistic” or plausible, while recognizing that there is substantial uncertainty in the estimates of reporting rates. We note that the actual values used for the reporting rates in the results presented here vary slightly from those given in Eveson and Polacheck, as the calculations done here were preformed with a preliminary set of estimates. Time did not allow us to repeat the calculations with the final estimates reported in Eveson and Polacheck (CCSBT-MPTM/0502/05).

Figure 13 shows the range of uncertainty in terms of projected SSB under current catch, for core sets based on different reporting rates. There are larger differences in projection indices for some of the other factors shown in Figure 1 than the differences implied by reporting rates. We therefore consider that reporting rate need not be an axis in the grid, and we also do not consider it necessary to incorporate this aspect into sensitivity or robustness trials.

It is interesting to note that the expectedN by factors on the grid can be rather different for the 5 different CPUE series. For example, expectedN for cpue 4 and 5 are very different with respect to juvenile mortality and Omega (Figure 15, CfullTagRR8s).

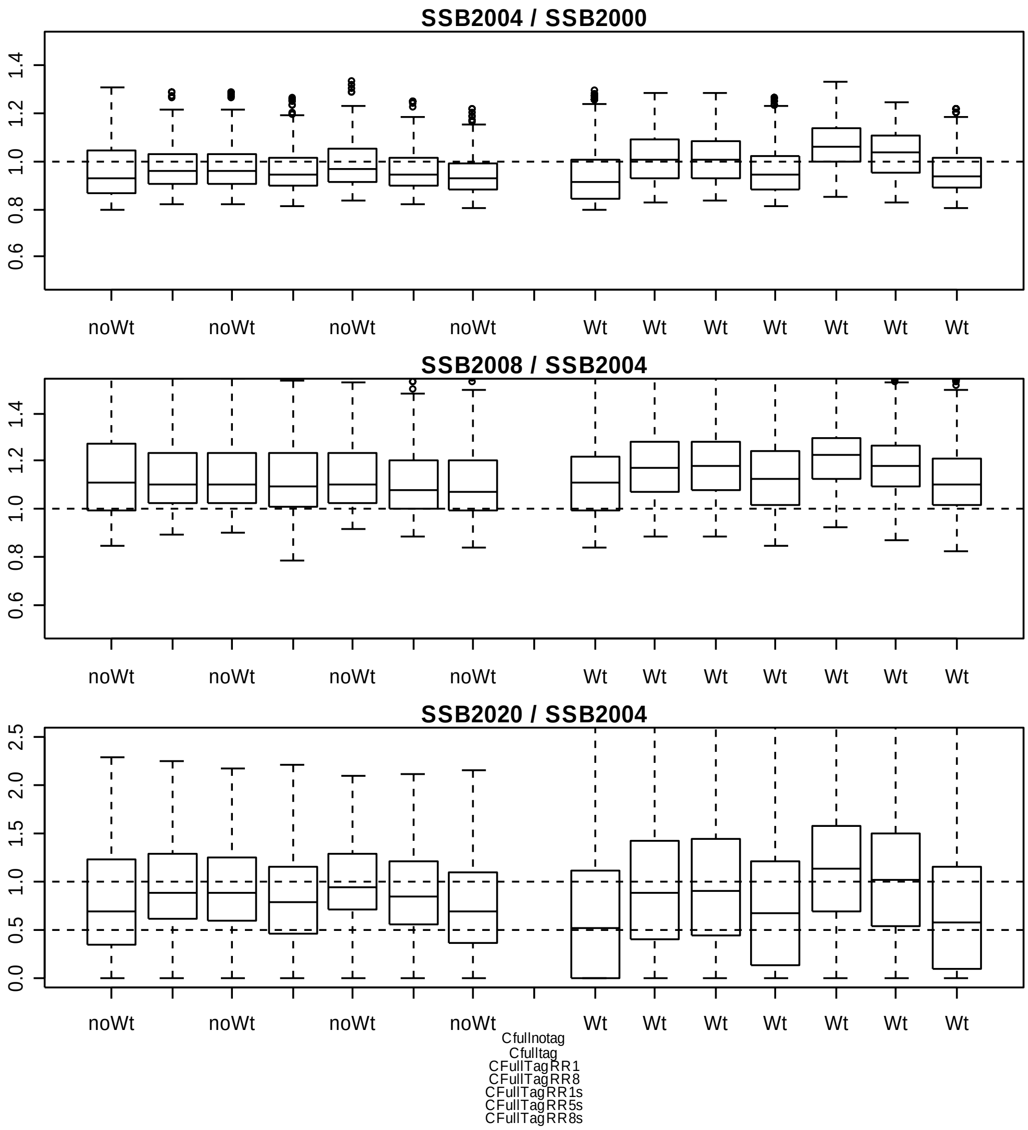

Figure 13 Unweighted (noWt) and weighted (Wt) projection indices for 9 core sets, using Cpue 2-6, and different reporting rates: Cfullnotag, Cfulltag (old reporting rate 5), CfulltagRR1 (old 1), -RR8 (old 8), -RR5s (new 5 scaled), -RR1s (new 1 scaled), and RR8s (new 8 scaled).

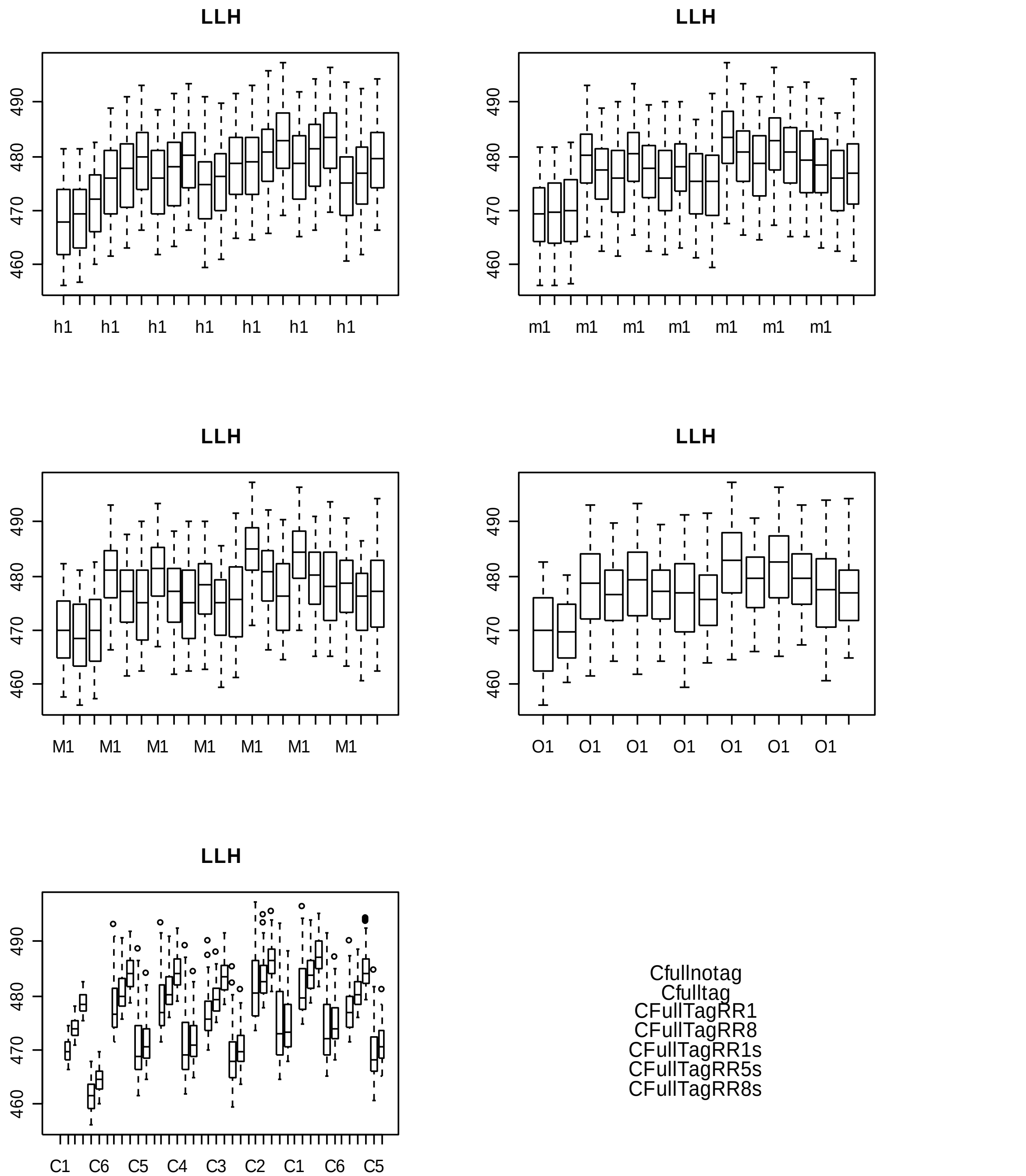

Figure 14 Negative log likelihood ranges by grid factors h, m0, m10, Omega and cpue 2-6 for the 7 core sets listed above. Note that likelihoods are not directly comparable between core sets, but are comparable within a core set.

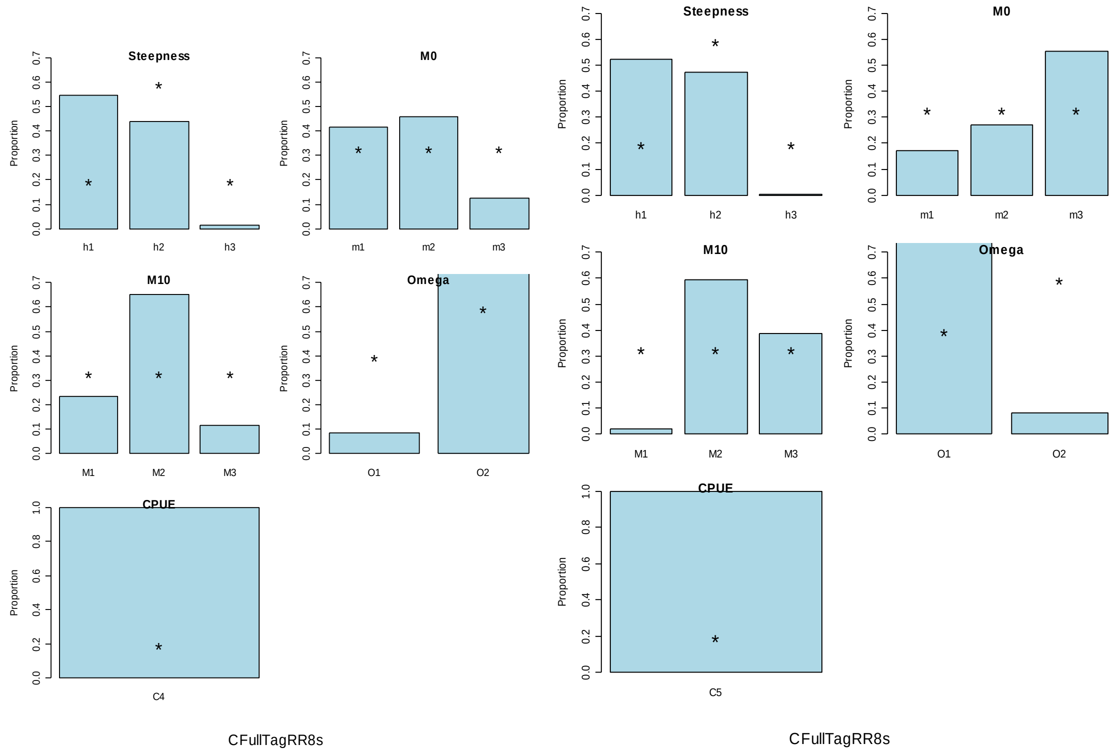

Figure 15 Expected proportion of samples by factor in the grid shown separately for cpue4 (left) and cpue5 (right). Bars are based on likelihood x prior whereas '*'s indicate priors, and bars indicate final weightings (except for steepness, in which the priors and final weightings are indicated by stars; bars indicate the frequencies that would have been observed for steepness, purely for the entertainment of those that are interested). CfullTagRR8s runs only.

List of Core/Grid Abbreviations mentioned in paper

Except where noted, all explorations were based on a "grid" of MPD model results estimated using the assumptions of the 2004 SAG (3 levels of steepness X 3 levels M0 X 3 levels of M10 X 2 levels of omega X 5 levels of CPUE). The following abbreviations represent orthogonal manipulations to the "core" set of remaining assumptions.

(Note in fig. labels: **CFull6*xxx*** is the same as **CFull*xxx*** (6 indicates that the median CPUE series was also run in the former case, but only included in results as indicated)

CFullnoTag – default specification from SAG 2004 – all the following specifications are identical to this one, except for features described
CFullnoTag_qa812 – run with alternative q age specification
CFullnoTag_ESSOrig – Original ESS weighting before the panel transformation
CFullnoTag_ESSOrig2 – Original ESS weighting – all downweighted by factor of 2
CFullnoTag_ESS4 – Panel ESS sqareroot transformation downweighted by factor of 4, all fisheries
CFullnoTag_Sel2 – number of selectivity temporal changes in fisheries 1 and 6 (?) doubled
CFullnoTag_M25 – M increases from ages 25 to 30 at 0.05 per year
CFullnoTag_M18_Ind18 – estimates the slope of natural mortality for ages 18-30 with a flat selectivity at ages 18+ imposed (*incomplete due to minimization failures in C4 scenario)
CFullTagRR1 – old reporting rate 1 (default in other runs is old RR5), tag weight 3.2
CFullTagRR8 – old reporting rate 8, tag weight 3.2
CFullTagRR1s – new scaled reporting rate 1, tag weight 3.2
CFullTagRR5s – new scaled reporting rate 5, tag weight 3.2

CFullTagRR8s – new scaled reporting rate 8, tag weight 3.2
CFullTag8s = CFullTagRR8s
CFullTag8s_ESSOrig2 = CFullTagRR8s + 0.5*original CA/CL effective sample sizes
CFullTag8s_qa812 = CFullTagRR8s + qa812 assumption
CFullTag8s_qa812_ESSOrig2 = CFullTagRR8s+ qa812 assumption + 0.5*original CA/CL effective sample sizes

Plus Other combinations of qa and ESS assumptions following pattern above