OCEANS AND ATMOSPHERE FLAGSHIP
www.csiro.au

Assessment of stock status of southern bluefin tuna in 2014 with reconditioned operating model

Rich Hillary, Ann Preece, Campbell Davies

CCSBT-ESC/1409/21
Prepared for the 19th CCSBT Extended Scientific Committee held in Auckland, New Zealand 1st-6th of Septmeber 2014.

Oceans and Atmosphere Flagship
Castray Esplanade, Battery Point, Hobart TAS 7000, Australia
Telephone : +61 3 6232 5222
Fax : +61 3 6232 5000

Copyright and disclaimer

Important disclaimer

Contents

Abstract

This paper details results and diagnostics relating to a full reconditioning of the SBT OM on the most recent data, and the first official reconditioning since the MP was adopted in 2011. The suite of standard population and fishery variables are summarised for the reference set of OMs, and a detailed analysis of the fitting performance of the model is undertaken for each of the data sets. Additionally, a suite of ranked robustness trials were also computed, and their stock status and MP performance statistics are compared with the reference set. Estimates of the biomass ratio (relative to unfished) range from 0.06-0.09 (biomass of animals 10+) and 0.08-0.12 (spawning abundance) with the suite of robustness trials resulting in similar and only slightly wider ranges. Estimates of the F_{msy} ratio were less than 1 but highly variable. The probability that the observed 2014 aerial survey is outside the simulated distribution used to test the MP is high for the reference set, but when considering the robustness testing in 2011, the observed point would be inside the range of values tested. Therefore, with respect to the data input series used in the MP, we conclude that exceptional circumstances have not been triggered for this, this year. Across all robustness trials MP rebuilding performance was generally quite good and arguably better than when adopted in 2011.

1 Background

While previous years have seen the development of the OM to incorporate new data sources - in particular the close-kin data [1, 2, 3] - 2014 is the first year for an official reconditioning of the SBT OM as per the MP implementation schedule. The last reconditioning was in 2011, for the finalisation of the MP testing, and the MP was implemented that year. In addition to the close-kin data, we now have aerial survey data up to and including 2014, Japanese long-line CPUE up to and including 2013 as well as the usual updated catch composition data for the various fisheries.

The paper is split into two main sections:

- **Reference set**: summary of all the main population and fishery variables as well as a detailed analysis of the fitting performance of the model for the reference set of OMs
- **Robustness trials**: summary of stock status and projection performance, under the adopted MP, for the set of (ranked) robustness trials defined in the OMMP meeting [4]

2 Reconditioned OM

2.1 Estimates for reference set

Figure 2.1 details the summary plot for the spawner abundance and recruitment dynamics for the reference set; Figure 2.2 details the associated level plot for all the key grid elements. Henceforth, the term biomass ratio refers to the ratio of biomass in the given year relative to the unfished state. Previous OMs have used the biomass of animals age 10+ as a proxy for SSB but with the inclusion of the close-kin data effective spawner abundance - a mixture of residence time, relative fecundity and resting times - is used as defined in [2]. For brevity, the usual likelihood profile plots for steepness, M_0, and M_{10} are placed in the Appendix and the results are summarised below:

- Steepness (h): preference in the actual observations is weak as before, but there is a general preference for medium to higher levels of steepness, coming from the catch composition and the aerial survey. The recruitment deviation penalty still strongly prefers lower steepness values so overall objective function is still weighted to towards lower values.
- M_0: Tagging data is the dominant information source here so, even with a weak penalty preference for lower values, the overall objective function is dominated by the tagging data's preference for medium to high values and is well defined (strongly unimodal)
- M_{10}: Tagging data are the dominant information source here as well, with a strong preference for

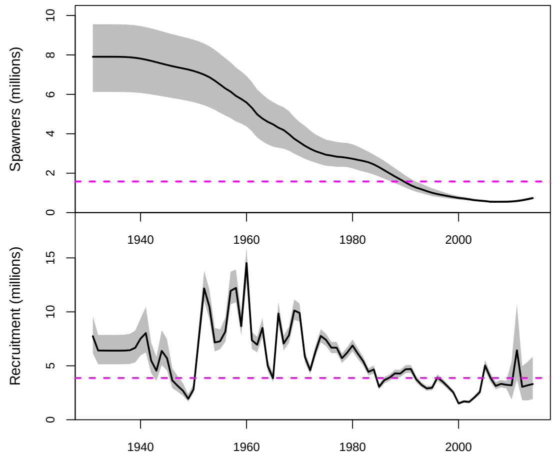

Figure 2.1: Spawner abundance (top) and recruitment (bottom) median (black line) and 90%CI (grey envelope) for the reference set of OMs. For the spawner abundance (not biomass of animals aged 10+) the dotted magenta line represents 20% of the median B_0; for the recruitment it represents 50% of the median R_0.

lower values. This is driven by strong decline in M from juvenile to sub-adult ages observed in the tagging data and given the assumed structure of the M-vector. The next strongest informant is the Indonesian age data with a preference for higher M_{10} values as seen previously. The close-kin data are more consistent with lower M_{10} values but more so at higher steepness levels, again as seen before.

Table 2.1 provides a general summary of the biomass ratio and MSY related variables for the reference set and, where feasible, with the reference set from 2011. The biomass ratio in the spawning population has a median value of 0.1; for the biomass of fish aged 10 and above this value is 0.07. Current estimates of the F_{msy} ratio (for ages 2 to 15, $F_{2014}/F_{\mathrm{msy}}$) have a median of 0.65. Current estimates of the B_{msy} ratio have a median of 0.38, with a median estimate of B_{msy} as a fraction of B_0 of 0.24. In general, the MSY estimates (apart from MSY itself) - especially the F and biomass ratios - are noticeably more variable than the biomass ratio estimates. When comparing with the same status statistics from 2011 the biomass ratios are higher, the F_{msy} ratios are lower and C_{msy} is approximately unchanged.

Variable	Estimate	2011
δ	0.1 (0.08-0.12)	-
δ_{10+}	0.07 (0.06-0.09)	0.05 (0.03-0.08)
F/F_{msy}	0.65 (0.39-0.98)	0.76 (0.52-1.07)
B_{msy}/B_0	0.24 (0.16-0.32)	0.22 (0.15-0.29)
B/B_{msy}	0.38 (0.26-0.7)	0.23 (0.15-0.32)
C_{msy}	33,242 (30,098-35,592)	34,500 (31,100-36,500)

Table 2.1: General summary of biomass ratio and MSY variables in terms of median and 90%CI in brackets for the reference set and for the 2011 reference set. In the above table δ refers to the biomass ratio in the spawning population, whereas δ_{10+} refers to the same but for the biomass of fish aged 10 and above.

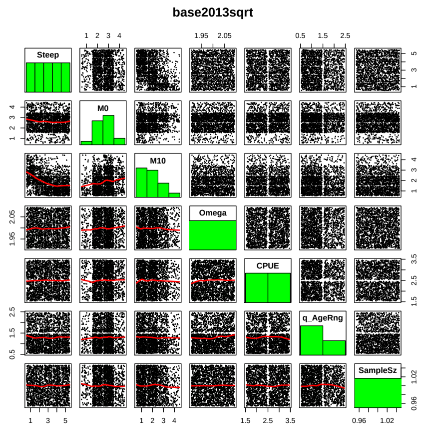

Figure 2.2: Level plot for the key grid elements in the reference set. Note M_0 and M_{10} **only** are resampled using objective function weighting.

2.2 Fits to data

The fitting performance of the OM for the reference set, summarised below, is detailed for each of the general classes of observations:

2.2.1 Catch composition

Figure 2.3 details the fitting summary for each of the long-line length composition data sets and the fits are shown only for those years in which the data are actually included in the likelihood (i.e. they have a non-zero effective sample size). In general, the data are well fitted but, as in previous years, with some notable years of misfit when effective sample sizes are very low - particularly for the LL2 and LL4 fleets. Nothing of serious concern, but likley a result of less than ideal sampling coverage and overall effort in these years.

Figure 2.4 details the fitting summary for the Indonesian and surface fishery age composition data. Both data sets are well fitted, though note that the 2013 selectivity for the Indonesian fishery was freed up so as to be able to fit the observed large increase in small fish in that age composition.

2.2.2 Abundance indices

Figure 2.5 summarises the predictive performance of the reference set in relation to the CPUE and aerial survey abundance indices. The process standard error for the aerial survey was increased from the previous value 0.18 to 0.22 as per the recommendations of the OMMP meeting [4], resulting in a more balanced p-value of 0.14. As noted before [3], for well established reasons we purposely underweight the CPUE and overweight the aerial survey and the predictive results merely reflect that in terms of post-fitting outputs. It should be noted that one should not really look at residual standard error (across years) for the survey given the complex covariance structure in the observation error model, and the same is true for counting numbers of data points inside/outside the approximate predictive 95%ile (i.e. proprtion inside/outside the predictive interval can be misleading with correlated data across years).

2.2.3 Mark-recapture data

The fits to the mark-recapture data, for the best fitting grid configuration and agggregated across tagger groups, can be found in Figure 2.6. The data are generally well fitted, with some mismatches but no obvious systematic problems for recapture histories on the same cohort.

The over-dispersion parameter, φ, was reduced from 2.35 to 1.82 for this reconditioning given a clear reduction in the residual variation in the overall mark-recapture residuals [4]. As a check, we can recal-

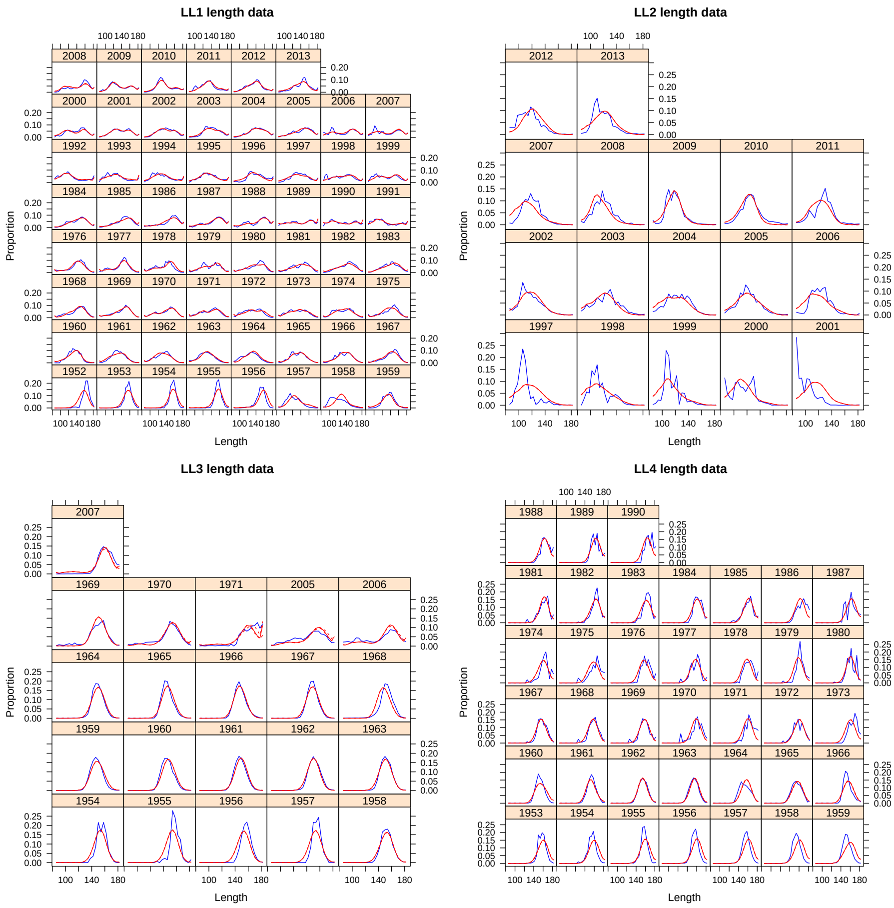

Figure 2.3: Observed (blue line) and predicted median and 90%CI (red full and dotted line) length composition for each of the long-line fisheries.

culate the new value of φ (across all grid elements) to ensure that the increase in weight to the tag data is “balance” - i.e. the output value is close to the input value. The median (and 90%CI) across all grid elements is $\widehat{\varphi} = 0.99\,(0.98 - 1.01)$. This suggests that the tag data, at the highest aggregation level, are seemingly well explained across the whole grid.

At the base aggregation level (tagger, cohort, age of release/recapture) a more detailed analysis of the over-dispersion in the mark-recapture data did not show any systematic fitting problems. Values for φ do vary around 1 at this level, as one would expect, but not greatly or with obvious trend. The final analysis was to fit a linear model to the full residual set with release age, cohort and tagger as factors and recapture age as a continuous variable (proxy for time-at-liberty). The only significant effect is for tagger group 6 that consistently results in recapture data that are less variable than for the other groupings, as seen in previous such analyses [3].

2.2.4 Close-kin data

Figure 2.7 details the fitting summary of the close-kin data (aggregated to the cohort and adult capture year level) for the best fitting grid configuration. In general the overall numbers of POPs and the trends are well fitted, with only one observation sitting outside the 95% CI. At the same aggregation level the posterior predictive properties of the reference set of OMs was explored and, more akin to what is done with the mark-recapture data, the variance in the standardised residuals of the close-kin data were examined. Both of these procedures are designed to detect the presence of over-dispersion in the close-kin data, and consequently whether we are over-weighting them in the likelihood.

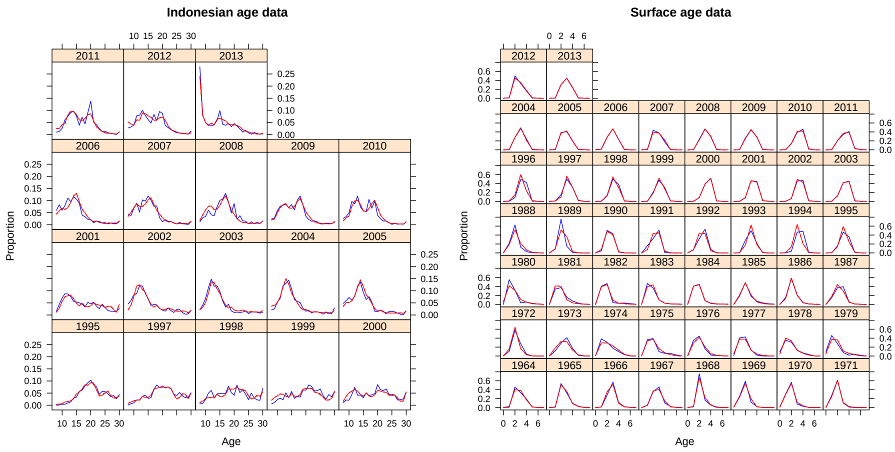

Figure 2.4: Observed (blue line) and predicted median and 90%CI (red full and dotted line) age composition for the Indonesian (left) and surface (right) fisheries.

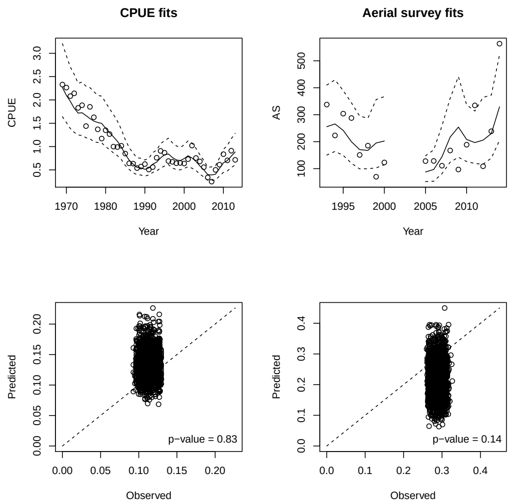

Figure 2.5: Predictive intervals (median and 95%iles) for the CPUE (top left) and aerial survey (top right) as well as the posterior predictive analyses (bottom).

Figure 2.8 summarises the predictive performance of the reference set on the close-kin data. For the posterior predictive analysis, the p-value of 0.35 suggests *slight* over-dispersion in the data; for the standardised residual bootstrap the estimates of the dispersion coeffecient are both ca. 0.83 suggesting the data are *slightly* under-weighted. Given two different methods show very slight but different directional shifts in the dispersion coefficient, we can probably be comfortable that the close-kin data are being weighted about right in the reference set of OMs.

3 Robustness trials

A suite of robustness trials were defined at the OMMP meeting and a low/medium/high ranking was assigned to each [4]. In this paper only those robustness trials not directly related to the current unaccounted for mortality work are explored. The comparisons with the reference set are primarily done on future performance of the MP in terms of biomass ratio levels and average catch. While some aspects of the historical estimates for each robustness trial are detailed (such as the current biomass ratio level) the

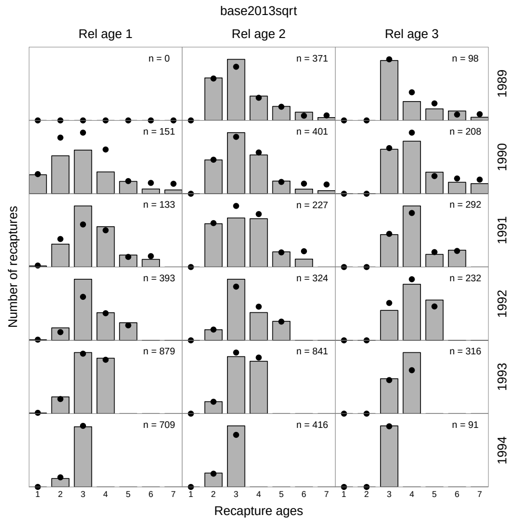

Figure 2.6: Summary of fits to mark-recapture data (aggregated across tagger groups) for the best fitting grid configuration.

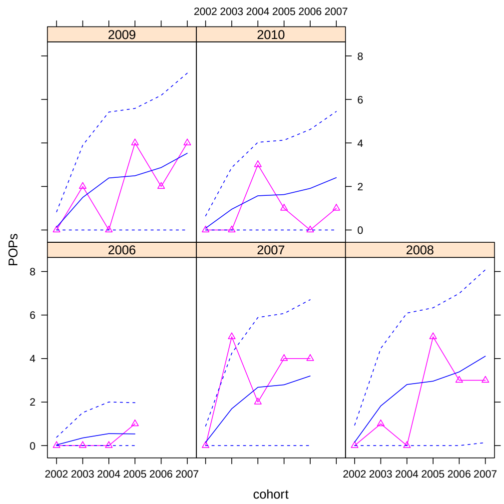

Figure 2.7: Summary of fits to close-kin data (aggregated to the juvenile cohort and adult capture year) for the best fitting grid configuration. The observed POPs (Parent Offspring Pairs) are denoted by the magenta triangles with the median and 95% CI of the predicted POPs given by the blue full and dotted lines, respectively.

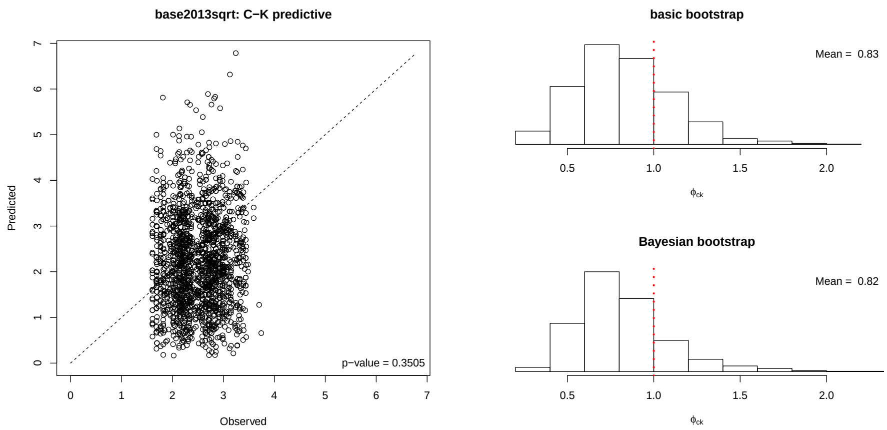

Figure 2.8: Posterior predictive analysis (left) and bootstrap of standardised residuals (right) for the close-kind data aggregated to the cohort and adult capture year level.

focus is placed on future performance under the adopted MP.

3.1 The 2014 aerial survey and exceptional circumstances

While the presence or absence of the 2014 aerial survey is explored in the robustness trials, it is included in all the projections, so we quickly address the rationale for this and how this datum relates to the potential for exceptional circumstances being invoked. Calculating the distribution of simulated 2014 survey points is simple (they can be found in the `.s7` files) **but** to compare it to the observed 2014 survey point we need to do some subtle rescaling. Recall the need for the `qratio` subroutine when calculating the global TAC from the MP. This deals with the year-to-year change of scale in the aerial survey (being a weighted sum not a mean-standardised index). To accurately compare the 2014 survey to the simulated 2014 survey points from the MP testing in 2011 we need to use this rescaling coming from the `qratio` procedure.

When doing this rescaling, the probability that the observed 2014 aerial survey was inside the simulated distribution from 2014 is 0.02 - so, in theory, this would trigger exceptional circumstances. However, in a similar line of reasoning used when exploring the same issue for the low 2012 aerial survey point, the following robustness trials explored in 2011 (**HighAerialCV, Laslett, troll**) all result in simulated aerial survey distributions that do contain the observed 2014 point. Therefore, one could argue that the 2011 testing work did encompass survey values such as that seen in 2014 and exceptional circumstances, in the context of the occurrence of observations in the input series outside the range used to test the MP, have not been triggered for the survey this year.

3.2 High priority

The robustness trials flagged as a high priority were: no aerial survey for 2014 (**noAS2014**), high-latitude CPUE series (**highlatCPUE**), reduced base CPUE series (**redbaseCPUE**), 0 (**C0S1L1**)and 50% (**C2S1L1**) contribution of LL1 over-catch to CPUE, and steepness resampled from the objective function (**esth**).

When comparing current to future biomass (of animals 10+) for the biomass ratio (Figure 3.1) the robustness trials are mostly more pessimistic (i.e. lower than the reference set) except for the high latitude CPUE and in particular the zero contribution of LL1 over-catch to CPUE trials. The range of current biomass ratio levels is 0.058-0.097 vs. 0.06-0.09 for the reference set only. There are no cases where the probability of the future biomass exceeding 20% of B_0 is less than 0.5 for these robustness trials (see Table 3.1).

In terms of the future catch/rebuilding trade-off (Figure 3.1) the only clear deviation away from the refer-

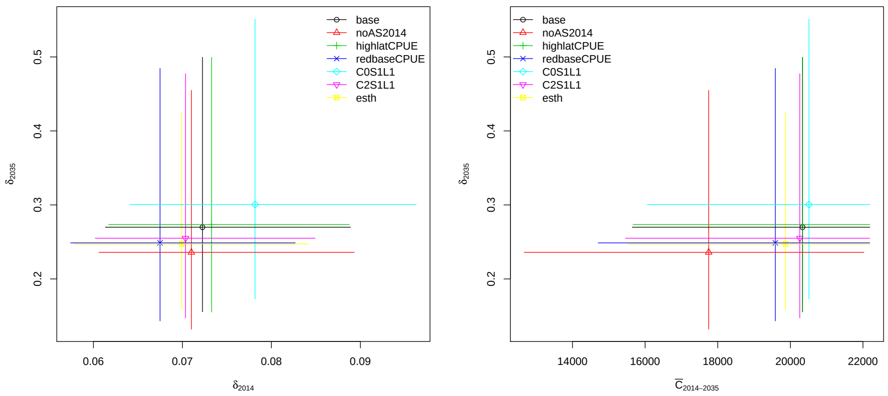

Figure 3.1: Current versus future biomass ratios (δ_{2014} vs. δ_{2035}, left) and future average catch versus future biomass ratios ($\bar{C}_{2014-2035}$ vs. δ_{2035}, right) for the high priority robustness trials. The points represent the median with the intervals the 90%CI.

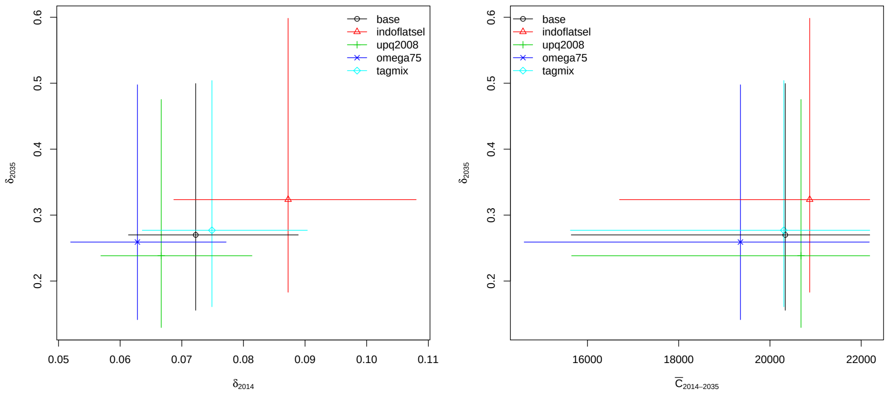

Figure 3.2: Current versus future biomass ratio (δ_{2014} vs. δ_{2035}, left) and future average catch versus future biomass ratio ($\bar{C}_{2014-2035}$ vs. δ_{2035}, right) for the medium priority robustness trials. The points represent the median with the intervals the 90%CI.

ence set is for the trial where the 2014 aerial survey is removed - this results in the lowest future average catch as well as rebuilding level. All the other robustness trials cluster fairly closely around the reference set.

3.3 Medium priority

The robustness trials flagged as a medium priority were: flat Indonesian selectivity from age 20 (**indoflatsel**), 25% increase in catchability of LL1 CPUE from 2008 (**upq2008**), sub-linear abundance to CPUE relationship for LL1 (**omega75**), inclusion of the trolling survey (**troll**), and incomplete mixing of the tags in the first year of release via higher Fs for tagged animals (**tagmix**). At the time of writing the paper, we experienced persistent numerical issues when trying to run the troll survey robustness test and, as such, it could not be included in these results.

When comparing current to future biomass (of animals 10+) for the biomass ratio (Figure 3.2) the robustness trials are evenly spread in terms of relative optimism/pessimism when compared with the reference set. The tag mixing trial is slightly more optimistic, with the Indonesian flat selectivity trial noticeably more optimistic. Both the LL1 catchability and CPUE-to-abundance trials are, as expected, more pessimistic than the reference set. However, in contrast to 2011 where these trials resulted in noticeably worse MP performance statistics [5] - less than 50% chance of rebuilding to the interim rebuilding target - both now have a greater than 50% chance of exceeding the target.

It is perhaps strange that the flatter Indonesian selectivity trial results in more optimistic performance, as the general effect in assessments has been that higher doming in selectivity results in more optimistic estimates of status and overall abundance. However, the SBT case is fundamentally different given the level of information on M and, crucially, absolute abundance information for the spawning stock in the close-kin data. Decreasing the doming of the selectivity in the Indonesian fishery accentuates the informative nature of that fishery's age composition data which suggests higher Ms for older animals, but this also switches the preference of the CPUE data from lower to higher M_{10}. In the reference set this preference is outweighed by the tag data's preference for lower values and the close-kin data's concommitance with those values, and little information in the CPUE or elsewhere. To fit the close-kin data, if we have higher M_{10} values then we must have higher "recent" recruitment levels to the spawning population to attain the same overall spawning abundance suggested by these data. Given the steepness is resampled from the prior there is some inevitable pivoting around this implied distribution, so we have a similar overall abundance, more optimistic biomass ratio levels recently, but the same steepness in the future. This results in both more optimistic biomass ratio levels in 2035 and higher average catches.

The range of current biomass ratio levels is 0.052-0.108 vs. 0.06-0.09 for the reference set only. There are no cases where the probability of the future biomass exceeding 20% of B_0 is less than 0.5 for these robustness trials (see Table 3.1).

OM scenario	$\delta_{2014,10+}$	$\mathbb{P}\left(\delta_{2035,10+} > 0.2\right)$
Base	0.07 (0.06-0.09)	0.75
	High priority	
noAS2014	0.07 (0.06-0.09)	0.63
highlatCPUE	0.07 (0.06-0.09)	0.75
redbaseCPUE	0.07 (0.06-0.08)	0.69
C0S1L1	0.08 (0.06-0.1)	0.82
C2S1L1	0.07 (0.06-0.08)	0.7
esth	0.07 (0.06-0.08)	0.73
	Medium priority	
indoflatsel	0.09 (0.07-0.11)	0.86
upq2008	0.07 (0.06-0.08)	0.63
omega75	0.06 (0.05-0.08)	0.69
tagmix	0.07 (0.06-0.09)	0.76
	Low priority	
highaerialCV	0.07 (0.06-0.09)	0.71
recdevAC	0.07 (0.06-0.09)	0.88

Table 3.1: Summary across all robustness trials for current (2014) biomass ratio in animals aged 10+ and the probability of meeting the Commission's rebuilding criteria by 2035.

3.4 Low priority

The robustness trials flagged as a low priority were: higher CV (40%) on the aerial survey process error (**highaerialCV**), upweight the close-kin data (**upwtck**), correlate future recruitment deviates to historical (**recdevAC**), and begin the OM from 1980 (**start1980**).

Only the aerial survey and recruitment deviate trials were run. For the upweighting of the close-kin data, this is complicated to do as it is not like the CV on an abundance index. By upweighting by a factor of 2 say, we are effectively saying we did twice as many comparisons as we did and we saw twice as

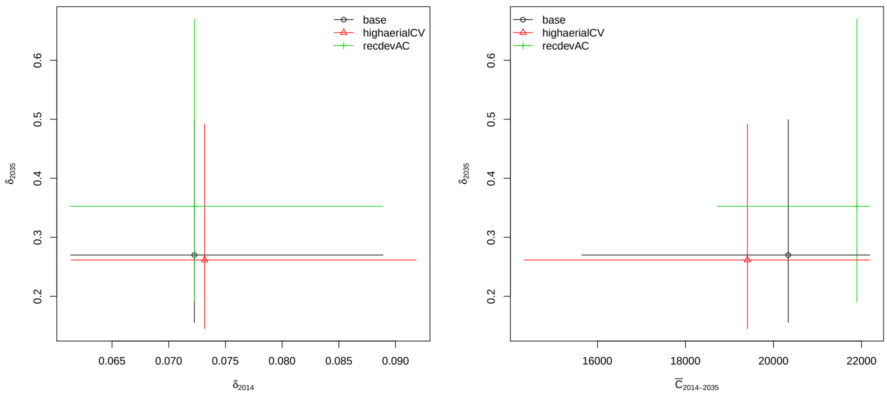

Figure 3.3: Current versus future biomass ratio (δ_{2014} vs. δ_{2035}, left) and future average catch versus future biomass ratio ($\bar{C}_{2014-2035}$ vs. δ_{2035}, right) for the low priority robustness trials. The points represent the median with the intervals the 90%CI.

many POPs. The overall abundance is effectively the same (driven by the ratio, which does not change) but the precision will be that much higher (approx. $\sqrt{2}$ times higher). The close-kin data are fitted by a combination of higher B_0, less biomass reduction relative to the unfished state, and lower M_{10} values (strongly backed by the tag data, in weak opposition to the Indonesian age data). The only data set that is significantly impacted by the inclusion of these data is the Indonesian age composition. The current reference set has no major problems predicting the overall level and variability in the close-kin data (Figures 2.7 and 2.8) so upweighting these data as a requirement does not seem plausible or necessary. In light of this rationale, we do not see a strong case for performing this robustness trial. For the 1980 start point for the OM, this would require significant changes to the OM code, and would require us to estimate an initial starting abundance (and age structure) *and* an estimate of B_0 for the stock-recruit relationship. Given issues already encountered with steepness (using all the historical data not just from 1980) this is likely to be a strongly over-parameterised OM. For this reasons we have not attempted the changes required to perform this robustness trial.

When comparing current to future biomass ratios (of animals 10+) (Figure 3.3) the high aerial CV robustness trial shows minimal divergence from the reference set. For the correlated future recruitment deviates the current biomass ratio is (by necessity) the same, but the future biomass ratio level is much higher as the high 2012 recruitment estimate driven by the survey results in higher subsequent unobserved cohorts and propagates into the spawning biomass.

The effect of this on the average catch levels is seen clearly for both cases: for the higher aerial CV there is less of a benefit from the lower estimated 2012 recruitment; for the correlated recruitment deviate trial the MP clearly reacts to the large increase in future mean recruitment with higher TACs while still obtaining higher rebuilding (suggesting no over-compensation in TAC increases).

As with the trials in the high and medium priority groups, there are no instances where the probability of rebuilding the stock to the interim target level was less than 0.5 (see Table 3.1).

4 Discussion

Overall, results are consistent with the results from recent developments of the OM to include new data sets and diagnostic methods [2, 3]. Stock status statistics indicate an improvement in status, relative to 2011, which is consistent with new information from the close-kin data and recent trends in major data sets (aerial survey, CPUE, Indonesian age structure). The estimated biomass ratio, for the reference set, ranges from 0.08-0.12 for spawning abundance and 0.06-0.09 for the biomass of animals aged 10+. The ratio of current median F to F_{msy} is around 0.65 and mostly less than 1 in terms of distribution, but highly variable (CV of 0.33). Current (i.e. dependent on recent selectivity profiles) estimates of C_{msy} are similar to 2011 i.e in the 30-35,000t range.

As noted previously [2], the combination of conventional tagging and close-kin observations for the same cohorts are most influential and significantly reduce overall uncertainty. This demonstrates the value of these types of absolute abundance and mortality data. The most recent aerial survey observation does not constitute exceptional circumstances for input data series for the MP, according to the same rationale used in 2012 (range in other robustness trials encompass the point); and there is time for two more observations of these cohorts (in the aerial survey and other data sources) before the next MP run for TAC setting. This will provide more information on the true strength of these year classes. There is still limited preference/information on steepness from the data, and also from the revised recruitment penalty, which is now adjusted for strong autocorrelation patterns. There is the prospect this will improve over coming years, given evidence of stronger year classes starting to reach the spawning stock and further strong recruitments moving into the exploitable population.

Across the robustness trials, current estimates of the biomass ratio appear both more optimistic and less sensitive than in 2011 [5]: in the reference set (for biomass of animals 10+) the range was 0.06-0.09; across all trials the range was 0.05-0.11. Projections for the reference set and the majority of robustness tests indicate the MP should perform as expected, and is still likely to achieve the Commission's rebuilding objective. Projected rebuilding performance is more optimistic for the reference set and all of robustness trials relative to 2011 - not one robustness trial had a less than 50% chance of rebuilding the stock to 20% B_0.

The relative information content of different data sets and the main remaining uncertainties (M-schedule, Indonesian selectivity, relationship between CPUE and abundance) suggests SRP efforts should be focussed on improving monitoring of the spawning ground (catch and catch composition, further close-kin and age structure samples) and abundance and mortality information on juveniles and sub-adults.

5 Acknowledgements

This work was funded by AFMA and CSIRO's Oceans and Atmosphere flagship.

References

[1] Bravington, M.V., Grewe, P., and Davies, C.R. (2012) Report of the close-kin project: estimating the absolute spawning stock size of SBT using genetics. *CCSBT-ESC/1208/19*.

[2] Hillary, R.M., Preece, A., Davies, C.R., Bravington, Eveson, J.P., and Basson, M. (2012) Initial exploration of options for inclusion of the close-kin data into the SBT operating model. *CCSBT-ESC/1208/21*.

[3] Hillary, R.M., Preece, A., and Davies, C.R. (2013) Data weighting, grid configurations and further exploration of the integration of the close-kin data into the SBT operating model. *CCSBT-OMP/1307/05*.

[4] Anonymous. (2014) Report of the Fifth Operating Model and Management Procedure Technical Meeting. *CCSBT*.

[5] Anonymous. (2011) Report of the Sixteenth Meeting of the Scientific Committee of the CCSBT in Bali, Indonesia. *CCSBT*.

Appendix

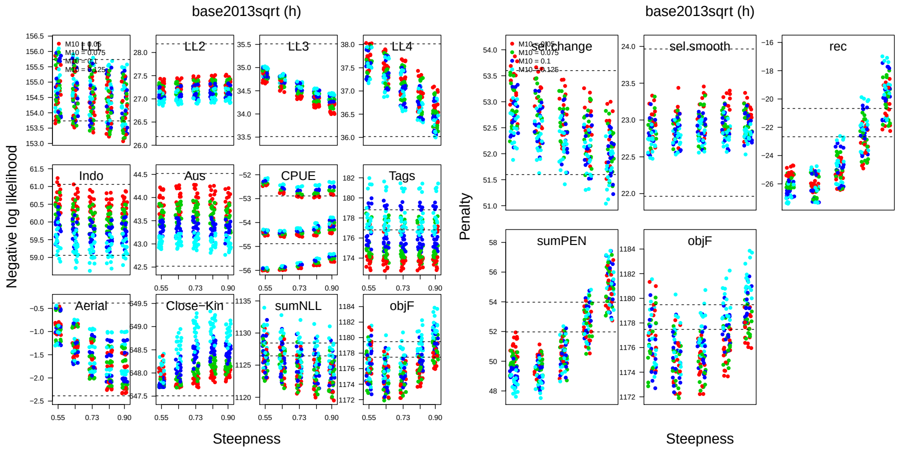

Figure 5.1: Likelihood (left) and penalty (right) profiles for steepness subsetted for M_{10} levels.

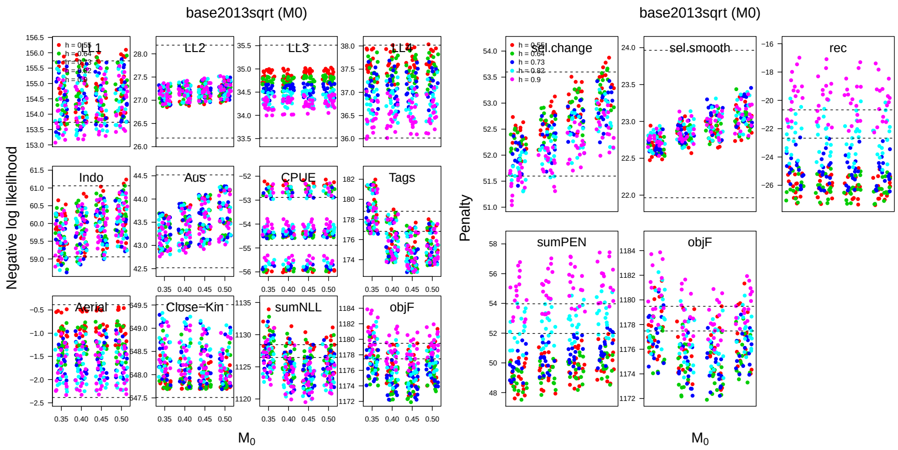

Figure 5.2: Likelihood (left) and penalty (right) profiles for M_0 subsetted for steepness levels.

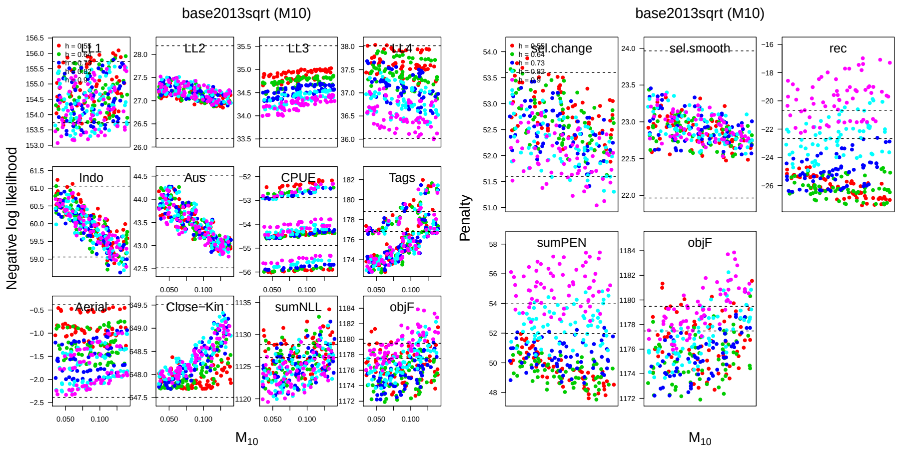

Figure 5.3: Likelihood (left) and penalty (right) profiles for M_{10} subsetted for steepness levels.

CONTACT US

t 1300 363 400
+61 3 9545 2176

e enquiries@csiro.au

w www.csiro.au

YOUR CSIRO

Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills for building prosperity, growth, health and sustainability. It serves governments, industries, business and communities across the nation.

FOR FURTHER INFORMATION

CSIRO Oceans and Atmosphere flagship
Rich Hillary
t +61 3 6232 5452
e Rich.Hillary@csiro.au
w