CCSBT-ESC/1107/13

Australian Government
Australian Bureau of Agricultural and Resource Economics and Sciences

Results of the performance of MP1

Barnes, B., Hillary, R., Tennant, P., Chambers, M., Preece, A., Davies, C., Begg, G.

July 2011

www.abares.gov.au
Science and economics for decision-makers

Contents

Summary

The inclusion of recent data into the operating model (OM) conditioning and management procedure (MP) projections has had an impact on MP1 performance, compared with previous results, under the scenarios identified for consideration. This paper explores the performance of MP1 with the updated OM. Although current spawning stock biomass (SSB) depletion has changed little, remaining below 10% of B_0 (CCSBT-ESC/1107/11), there are positive changes in recent recruitment and steepness that affect the projections and MP performance.

The Management Procedure Technical Working Group agreed intersessionally on four candidate MP scenarios for evaluation in terms of performance. However, given the more productive estimates of stock status and key parameters observed when conditioning the OM on the most recent data, MP1 could not tune to the IRP (Initial Reduction Period) scenario. Minimal difference in performance was observed for MP1 when tuning to 2035 (Run 1) or 2040 (Run 2) in terms of both short and longer-term SSB and catch, although with a 5000 t maximum TAC change and 2035 tuning year (Run 3) the initial TAC increase was slower and reduced variability in SSB later in the projection period.

The MP Technical Working Group also agreed on priority robustness trials; and MP1 (for the Run 1 scenario) was run on all these trials. Paying particular attention to the pessimistic trials (lowR, omega75, updownq, STwindows) MP1 appeared reasonably robust to both the lowR and omega75 trials, which were the most problematic in the previous MP testing work (CCSBT-ESC/1009/11). The updownq trial has an alternative view of the recent catch per unit effort (CPUE) increases (i.e. they are driven more by increased catchability than abundance) and this proved more problematic for MP1 than both the lowR and omega75 trials. The STwindows trial was by far the most problematic for MP1 (for the Run 1 scenario) given it has a significantly more pessimistic view of current SSB depletion and lower overall productivity, relative to the reference OM. As with the previous evaluation undertaken last year, MP1 was observed to be robust to trials relating to unreported catch levels (c0s1l1, c2s1l1, c3s1l1), alternative CPUE series (run3, run6, Laslett), and structural issues within the OM itself (mixtag, regime, aerflat).

1 Introduction

MP1 is a biomass random effects model based management procedure (MP) that uses scientific aerial survey and catch per unit effort (CPUE) data signals to adjust total allowable catch (TAC). Details of the specification are in Hillary and Preece (2011). The MP has been evaluated using projections from 2000 operating model (OM) scenarios sampled from the reference set grid of possible scenarios. For each individual historical scenario, the population is projected forward and the MP is used to adjust future TACs. MP1 is tuned so that 0.7 of the scenarios reach 20% of B_0 (unexploited spawning stock biomass) by the tuning year (2035 or 2040). This allows for a comparison in performance between tuned MPs. A set of runs with different tuning years and other constraints has been agreed intersessionally for initial examination at the 2011 Extended Scientific Committee (ESC) meeting. MPs tuned to these alternative constraints are then further examined using the robustness trials that allow for an examination of MP performance under more extreme, but plausible stock conditions (structural and statistical). The robustness trials have been grouped into 4 types, and the "pessimistic" robustness trials have been examined in more detail, as agreed by the Management Procedure Technical Working Group prior to the 2011 ESC meeting.

2 Candidate Management Procedures

The MP is tuned to reach an agreed target (0.2B_0) by the tuning year (2035/2040) and for the set of constraints described in Runs 1-5, for 70% of the 2000 projections. In the table below, ***Max TAC change*** refers to the maximum allowable change in TAC (3000 t/5000 t) from one TAC setting period to the next (3 years), ***lag*** refers to delay between the TAC setting and its implementation, and ***Prob*** refers to the proportion (70%) of the resultant trajectories that reach the agreed target.

The candidate set of MPs was decided intersessionally by the Management Procedure Technical Working Group to be:

Run		Tuning year	Prob.	lag	Max TAC change
Run 1		2035	0.7	1	3000
Run 2		2040	0.7	1	3000
Run 3		2035	0.7	1	5000
Run 4	IRP 3 yrs 3000t	2035	0.7	1	3000

In addition, a lag 0 candidate MP was examined:

Run	Tuning year	Prob.	lag	Max TAC change
Run 5	2035	0.7	0	3000

The first 3 candidate MPs have the 2011 TAC set to 9563 t and the lag year 2012 TAC set to 9449 t (consistent with ESC recommendations, paragraph 107, 2010 ESC Meeting Report), a fixed period of 3 years between TAC changes, a minimum TAC change of 100 t, and a maximum change of 3000 t or 5000 t. These three runs comprise the initial focus for MP testing. The fourth run was included to examine the effect of having an initial reduction period (IRP), with a fixed TAC of 3000 t from 2012-2014. And the fifth run was designed to evaluate the impact on MP performance of allowing for a change in TAC in 2012 (i.e. no lag).

At the ESC in 2010, both MP types (MP1 and MP2) had an additional constraint to not allow an increase in TAC in the first TAC setting period. This was agreed at the ESC meeting based on the estimated low spawning stock biomass (SSB), recent low recruitment and the potentially poor performance of both MPs under a number of “pessimistic” robustness tests. This additional constraint is not part of the candidate MP runs above. The impact of this constraint was explored and discussed below.

3 Specifications

There have been minor changes in the OM and MP1 structure since the last assessment in 2009, as well as the inclusion of more recent data:

- A new growth model has been incorporated into the OM (Eveson, 2011)
- A change in the grid for steepness and M0 has been incorporated, with sampling from higher steepness and M0 (by correlation) increased (Hillary et al., 2011, CCSBT-ESC 1107/11)
- MP1 has been modified to include the most recent data; thus it reacts sooner to positive and negative signals in the data (Hillary and Preece, 2011)
- Data input files for both the OM and MP1 have been updated to include recent data from up to 2011 for the scientific aerial survey, and 2010 for the other data sets including CPUE. (MP1 uses scientific aerial survey and CPUE data as input).

The recent OM reference set and stock status is described in Hillary et al. (2011) (CCSBT-ESC/1107/11), and all code used for the OM and projections is the most recent provided in June and July 2011 on the CCSBT website.

4 Constraints and Tuning

With the inclusion of new data (2009/2010-2010/2011), and for the previously agreed interim target SSB level of 0.2 B_0, the projected recovery is not consistent with the constraints on the MP in certain more restrictive scenarios. The combination of the tuning year, the tuning probability, the maximum change in TAC and the frequency of TAC setting constrain the potential range of behaviour of the MP, and if too restrictive can imply that the tuning target cannot be attained. Run 4 (the IRP candidate MP) does not converge, as the MP is unable to increase catches sufficiently over either of the tuning periods given the updated OM and resulting SSB projections. The other candidate MPs converge; however, if the additional constraint of no increase in catch in the first TAC decision year is imposed, then Run 1 cannot converge.

The updated OM grid (**basehupsqrt**) and constant catch projections (CCSBT-ESC 1107/11) indicate a more productive stock, relative to the previous OM, which results in an increased rate of future recovery. For example, for the zero and current TAC constant catch projections the interim rebuilding target will be attained by 2021 and 2026, respectively, at the 70% tuning level. Clearly from Figure 1, for both the 2035 and 2040 tuning years (Runs 1 and 2), MP1 increases median catches consistently at or close to the maximum amount at each TAC setting year to achieve the tuning target. When the MP is further constrained in the early years of the rebuilding period, as in the IRP scenario, it cannot increase the catches fast enough after the IRP to attain the relevant tuning target.

This issue of inconsistency between MP constraints, recent data and the updated OM is not MP specific. Exploration demonstrated that, with a first TAC increase in 2015, a three year TAC decision frequency and tuning year 2035, an increase in TAC of 3000 t is not adequate to meet the agreed target. This result is independent of the MP. Run 4 is more restrictive and thus cannot converge.

Within the boundaries already set by the Commission (in terms of tuning years and target probability) the results, using the most recent data suggest that the IRP scenario is not necessary for an MP to attain the rebuilding target. In the results for candidate MP1 Runs 1, 2 and 3 in the rest of this paper, there was no constraint imposed on changes to the TAC in the first decision year other than the maximum TAC change.

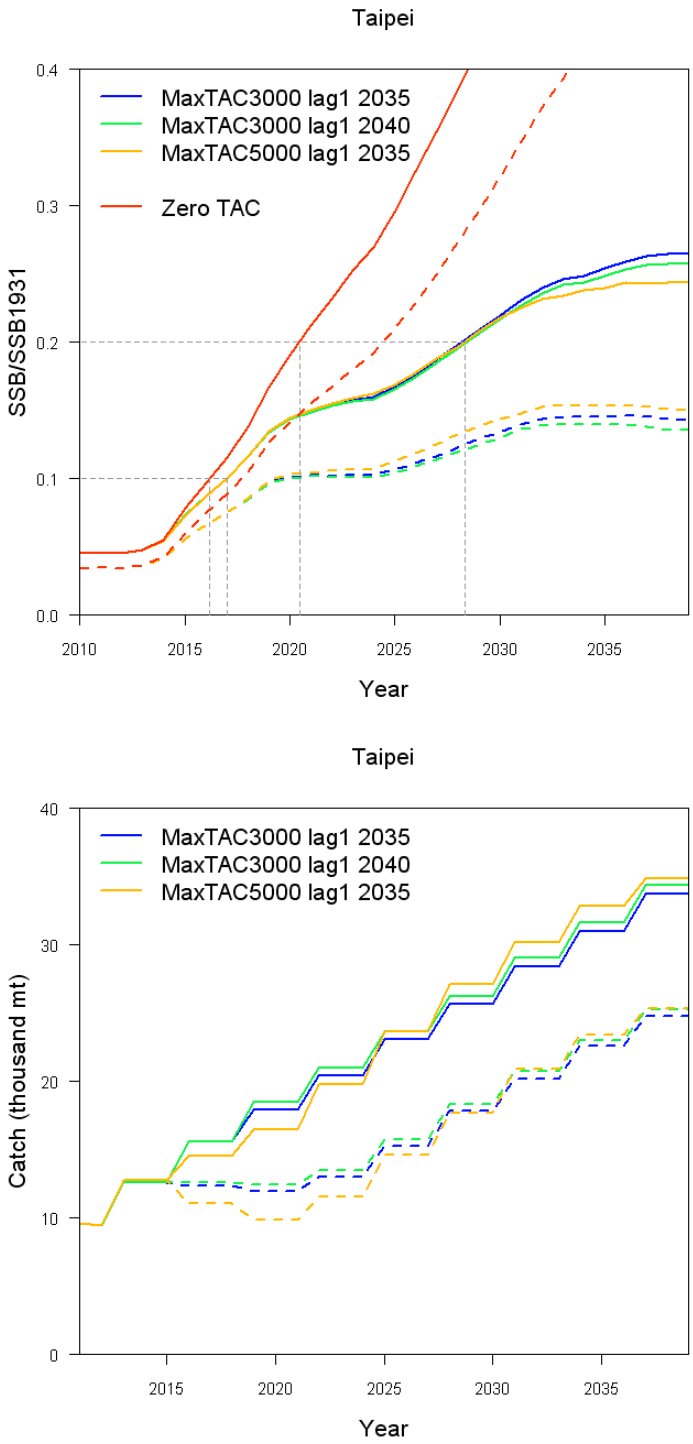


Figure 1: Spawning stock biomass (SSB) and TAC for candidate MP1 Runs 1-3. Solid lines represent medians and dashed lines the 10th percentile values.

5 Results for Reference Set

Candidate MP1 Runs 1, 2, 3 and 5

Figure 1 and Table 1 illustrate a comparison between candidate MP1 Runs 1, 2 and 3 in terms of SSB projections, TAC recommendations and target statistics. Further details (worm plots, statistics and trade-off plots) are given in Figures 2-4.

Runs 1-3 all tuned to the interim SSB rebuilding targets. Short-term SSB rebuilding is similar across Runs 1-3, with at least a 90% probability of increasing the SSB to 0.1 B_0 by the relevant year (2022 for Runs 1 and 3; 2025 for Run 2) (Figures 1-2). In contrast to the previous MP evaluation trials, none of the runs predict that TAC cuts in the first year are likely to be required. For Runs 1 and 2, median catches increase from the current level with time at a constant rate, at, or close to, the maximum amount. For Run 1, and the more pessimistic stock trajectories (Figure 1 and the worm plots in Figure 3), the rate of catch increase is slower (and later can decrease). In contrast, Run 3, the initial TAC is higher than in Runs 1 and 2, in the second TAC change the median increase is lower than for Runs 1 and 2, but eventually the median TAC exceeds the levels seen for Runs 1 and 2.

In the catch and SSB trade-off plot in Figure 4 it is evident that, although the uncertainty in average catch is higher for Run 3, the uncertainty in SSB is lower. This is the main difference between Runs 1 and 2 and Run 3. For Run 3, the median level of SSB obtained by the tuning years is lower than for Runs 1 and 2, but because the uncertainty is reduced and the lower percentiles are closer to the median (see Figure 1) the tuning target is still achieved.

The effect of lag was found to be negligible when Runs 1 and 5 were compared, and the results are illustrated in Figure 5.

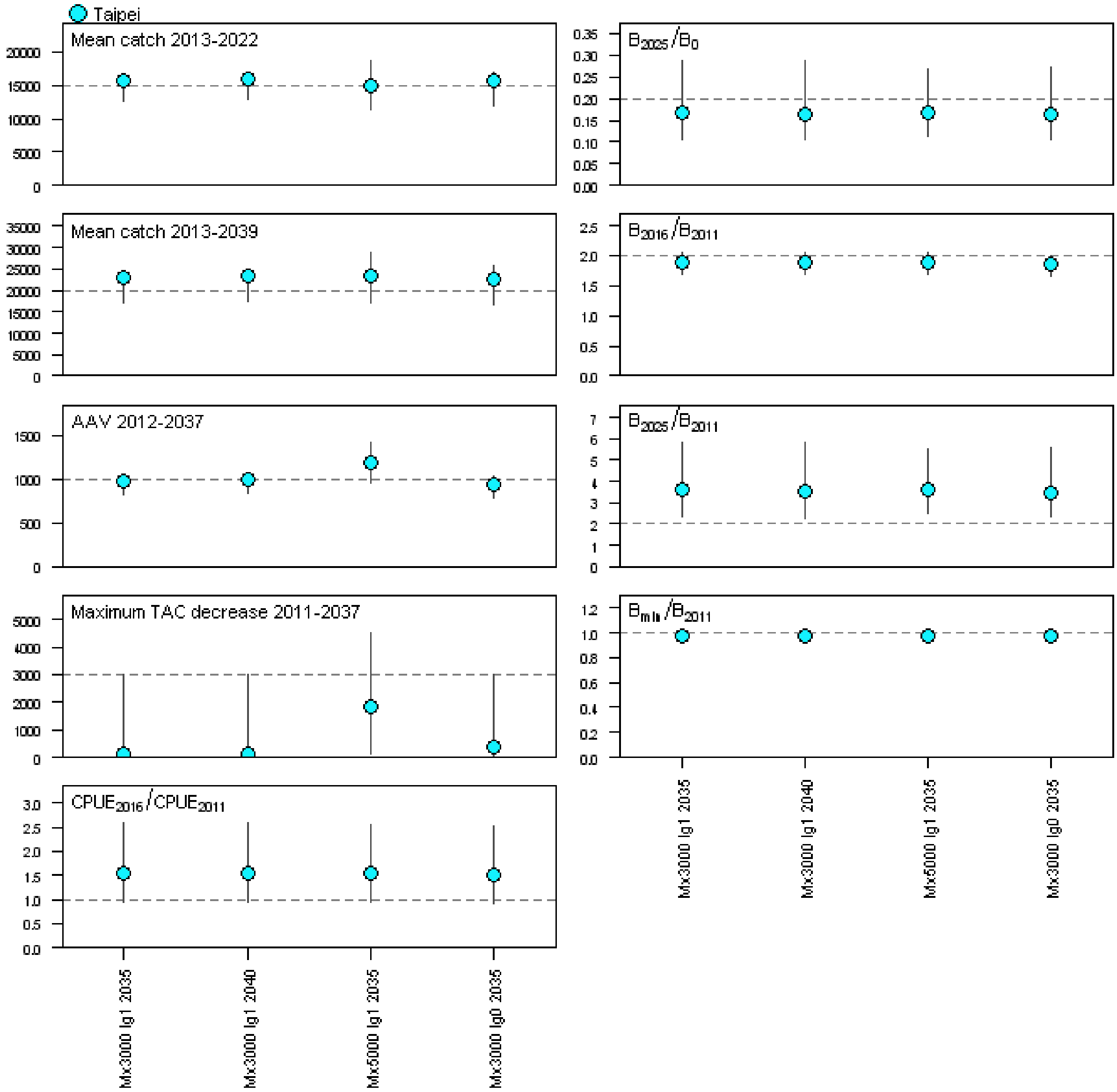


Figure 2: Comparison of performance statistics for candidate MP1 Runs 1-3 and Run 5. Circles are medians and whiskers are the 80% confidence intervals.

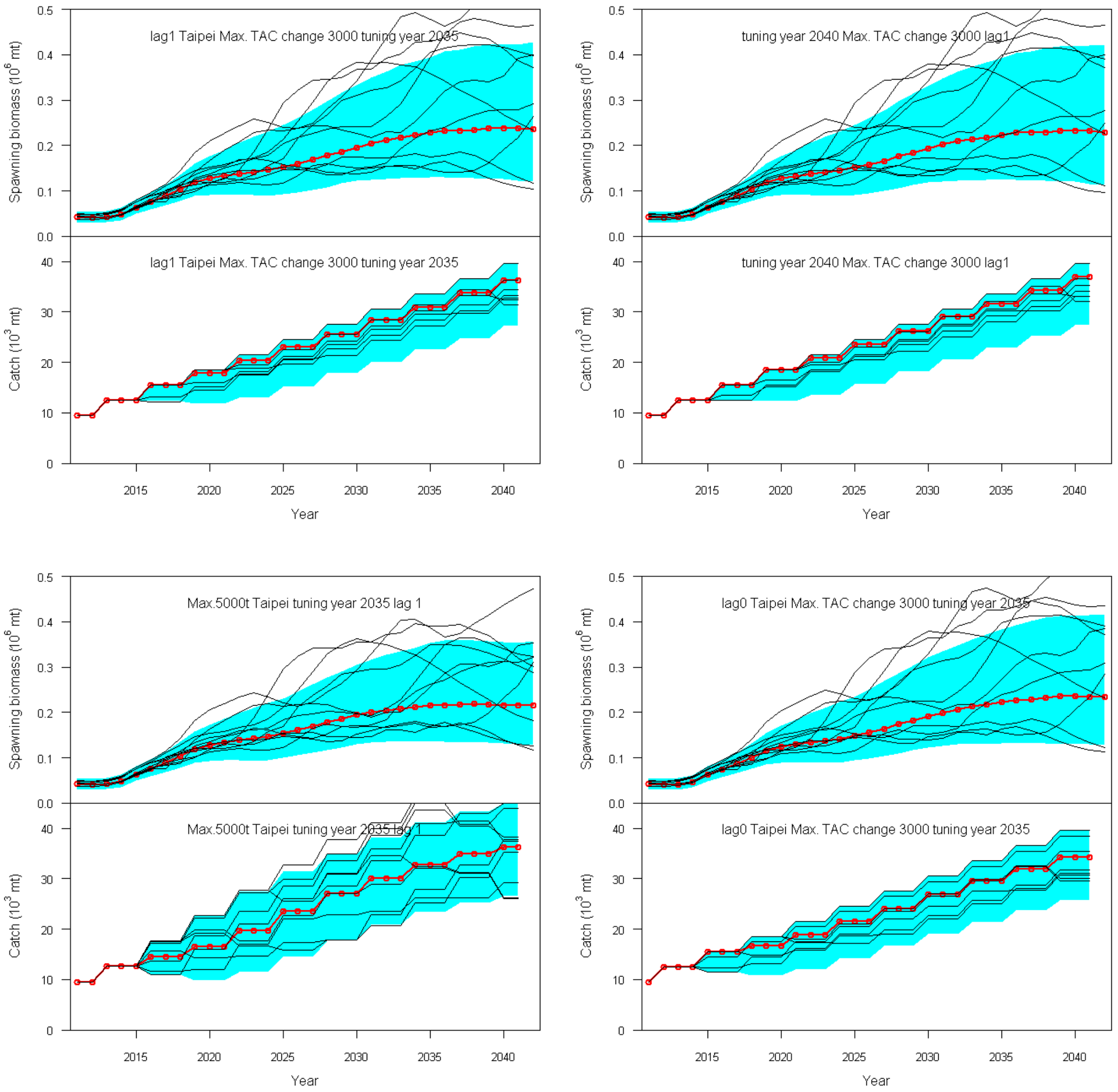


Figure 3: Worm plots of spawning stock biomass (SSB) and TAC for Run 1 (top left), Run 2 (top right), Run 3 (bottom left) and Run 5 (bottom right). Ten random simulated trajectories are given, with the coloured band the 80% confidence interval.

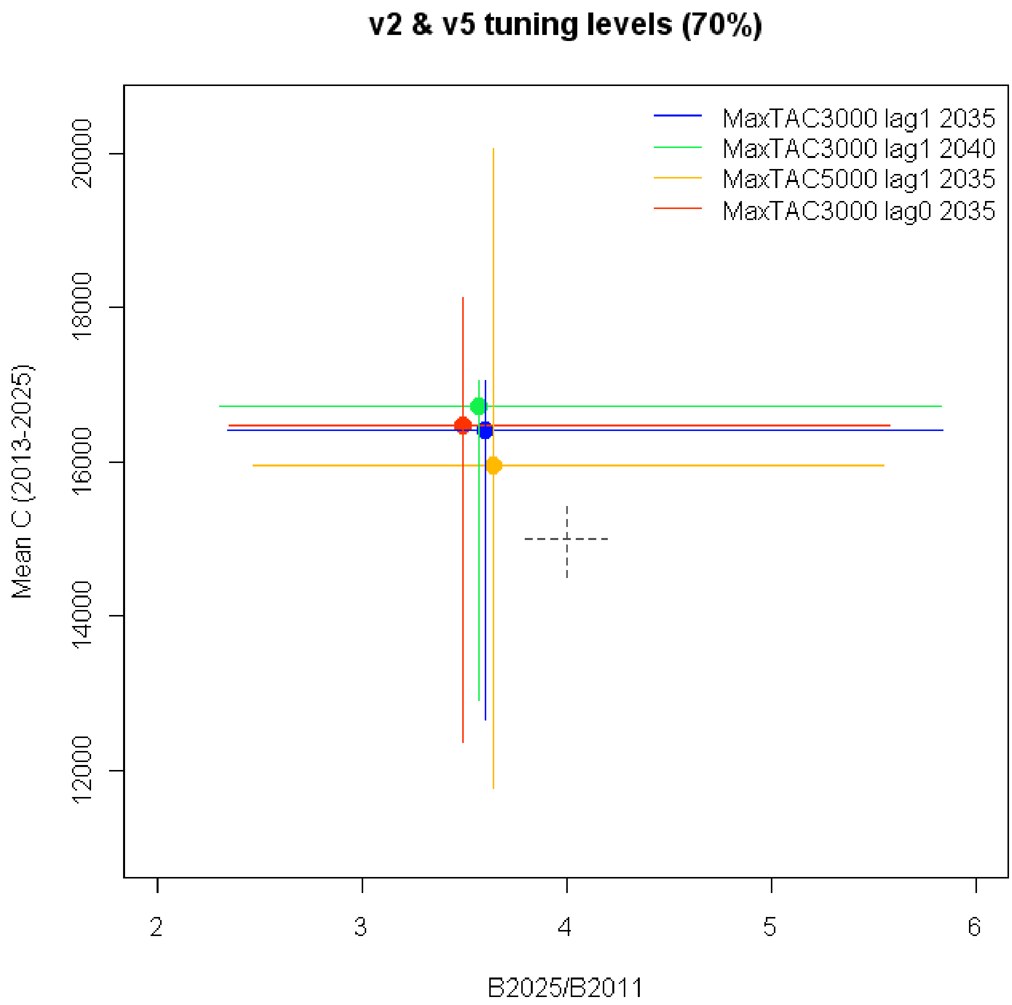
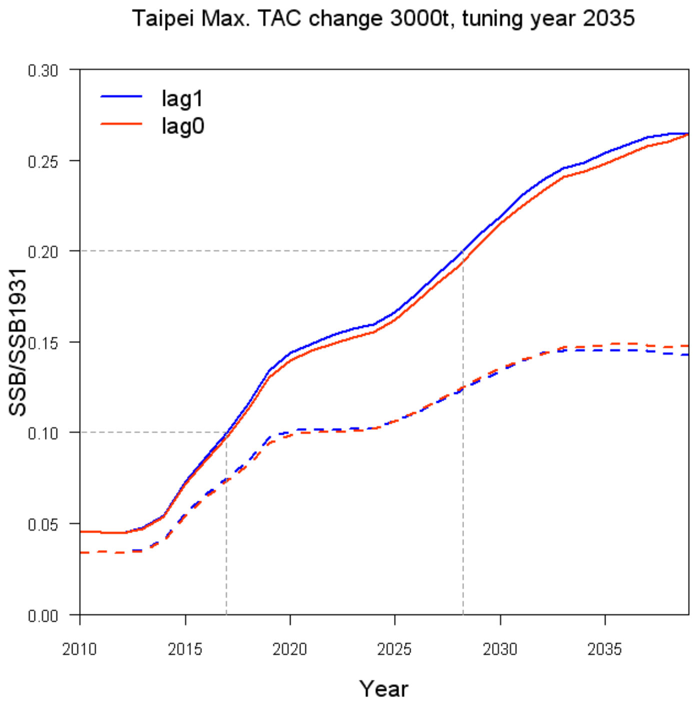


Figure 4: Interim-time-frame year-averaged catch versus average CPUE (relative to 2011) trade-off for the candidate MP1 Runs 1-3 and 5. Circles are medians and the whiskers are the 80% confidence intervals. (The dashed cross is included for reference for comparison with similar figures.)

Table 1: Performance diagnostics for candidate MP1 Runs 1-3 and Run 5. MPs are tuned to 70% probability of $B_{2035/2040} > 0.2B_0$.

	MP scenario	Tuning probability	Tuning year	Lag	Max TAC change	Check point year(t)	P[Bt>0.1B_0]	P[Bt>2B2010]	B2025/B2010	Median avg. catch (2013-2022)	Median avg. catch (2013-2039)	Year 1.2*B_{2010} reached (median)
1	Run1	0.7	2035	1	3000	2022	0.91	0.96	3.65	15731	22831	2015
2	Run2	0.7	2040	1	3000	2025	0.92	0.96	3.61	15957	23268	2015
3	Run3	0.7	2035	1	5000	2022	0.92	0.97	3.68	15062	23442	2015
4	IRP 3 years 3000 t	0.7	2035	1	3000	-	-	-	-	-	-	Unable to tune
5	Run5	0.7	2035	0	3000	2022	0.90	0.96	3.53	15690	22454	2015



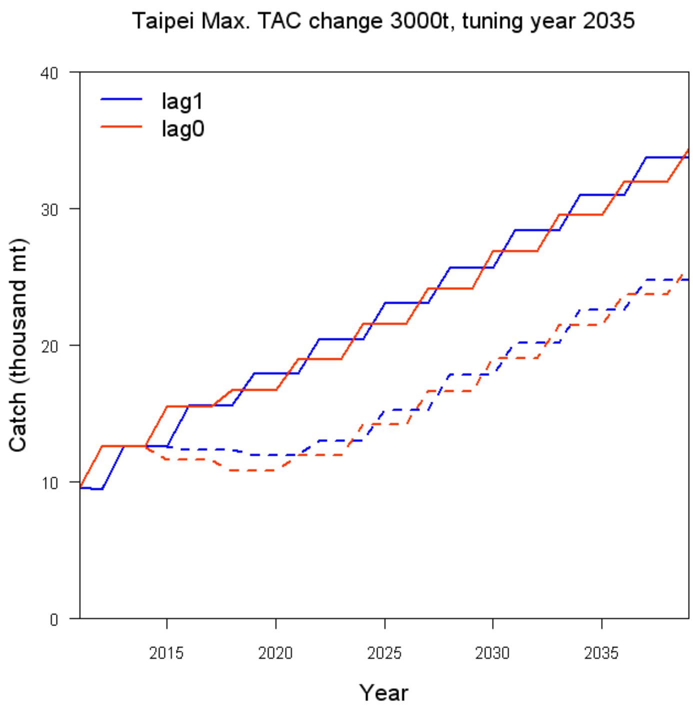


Figure 5: Spawning stock biomass (SSB) and TAC for candidate MP1 Runs 1 and 5, to establish the impact of lag. Solid lines represent medians and dashed lines the 10th percentile values.

6 Robustness Trials

A full set of robustness trials, with description, is given in Appendix 1. Each robustness trial corresponds to an alternative OM or changes to the underlying assumptions of the reference OM that is essentially 'hidden' from the MP (in all cases the MP run is that which was tuned to the reference OM). This allows for comparison of the performance of the tuned MPs under plausible but possibly more extreme conditions

Pessimistic robustness trials

Certain more ‘pessimistic’ trials were identified as high priority during intersessional communication (July 2011), and results for these trials are discussed below and illustrated in Figures 6-7. It should be noted that the same trials were carried out for Run 2, but they showed strong similarity with Run 1 outcomes and are thus not reported. They were also carried out for Run 3, and results are given in Appendix 2 for comparison purposes. The ‘pessimistic’ robustness trials considered here are:

- lowR – 4 years from 2011 where recruitment is 50% lower than predicted, uncorrelated with subsequent recruitments
- updownq – catchability goes up by 50% in 2009 and returns to normal in 5 years – originally designed to explore issues with change in behaviour of Japanese fishery but now included as an alternative explanation for the recent strong CPUE increases. Uncorrelated with subsequent CPUE observations.
- omega75 – Omega value of 0.75 rather than 1 i.e. CPUE non-linearity factor (although there is little support relative to linearity in the current reference grid)
- STwindows – alternative CPUE series that has been consistently lower over recent years than the others.

Note that, because of changes to the MP1 structure (it reacts more quickly to signals in the data), comparisons with previous results (in 2010) should be taken with caution.

lowR

From Figure 6, short-term SSB rebuilding decreases, relative to the reference case, as the low recruitments move through the juvenile population and into the adult biomass. From Figure 7, short-term (2013-2018) and longer-term (2019-2032) TACs are notably lower than in the reference case as the MP reacts to the negative signals appearing first in the scientific aerial survey and then in both the aerial survey and CPUE data. This reaction is also observable in the larger maximum TAC decrease statistics and the increased AAV statistics observed in the later period, which are linked to decreases in TAC not increases, as the negative signals have fully appeared in both the scientific aerial survey and CPUE data after this point. In terms of long-term SSB rebuilding we see that the lower levels of TAC chosen by the MP begin to offset the short-term effects of the low recruitments, and the tuning target is almost reached by 2035.

updownq

With the increase in catchability simulated in the robustness trial we can see a notable increase in the short-term future TAC levels chosen by the MP, relative to the reference case, for both the median and lower 5%ile (Figure 7). This is due to the MP acting on the upwardly biased CPUE by attributing it to an increase in exploitable abundance, and increasing catches accordingly. This has a notable negative effect on the short-term (2022 and for 0.1 B_0) and longer-term (2032 and tuning statistics) rebuilding of the SSB (Figure 6). Although average catches decrease as we remove the CPUE bias they are still somewhat higher than the reference case (Figure 7). This is driven by an upward bias entering into the MP estimates of mean recruitment biomass and biomass surplus growth (μR and μg) caused by the artificial CPUE increase. MP1 is adaptive and learns from the data over time, so the memory of this effect remains after the effect has gone from the actual data. From Figure 6 it is clear that, while relative to SSB in 2011 this robustness trial seems to have little impact, when assessing SSB rebuilding from either the 0.1 B_0 or 0.2 B_0 reference levels it performs noticeably worse than the reference case. This is because, for the updownq trial, recent (2008-2011) SSB and recruitment levels are all estimated to be lower than the reference case and so are the steepness levels and SSB depletion in 2011. Given there is no meaningful change in the estimates of B_0, the future levels of the SSB under this robustness trial are lower relative to B_0 than in the reference case, and in terms of the tuning level have a probability of less than 0.5 of increasing the SSB above the interim rebuilding target (and not 0.7 as in the reference case).

omega75

The omega75 robustness trial implicitly assumes a higher level of depletion in the historical exploitable biomass than the reference case, given the assumption of a sub-linear relationship between CPUE and abundance. The OM accounts for this with some lower individual recruitment estimates but mostly a lower estimate of steepness (mean of 0.71 versus 0.77). The resultant SSB depletion in 2011 is also lower, driven by the OM interpretation of a stronger decline in historical CPUE for this trial. One can see the effects of this in the MP projection in Figure 6, where the minimum future SSB is estimated to be further below 2011 than in the two previous robustness trials, and one can see the short-term rebuilding statistics suffer as well. The MP does not, however, begin to increase catches as quickly for this robustness trial (see average catch and lower short-term AAV statistics in Figure 7) as the increasing CPUE is not as strong a signal given the sub-linear CPUE-to-abundance assumption. The MP is then slower to react to the recovery, in terms of increasing TACs, and actually attains a better probability of attaining the tuning level in 2035 and SSB rebuilding level relative to 2011 than in the updownq trial.

STwindows

The STwindows CPUE series has been, for a number of years, much more pessimistic in terms of the relative level of historical depletion observed in the standardised CPUE. When conditioning the OM to this CPUE series we obtain marginally smaller estimates of steepness (mean of 0.74 versus 0.77) but noticeably lower estimates of current SSB depletion (mean of 0.03 versus 0.05), driven largely by lower estimates of recent recruitments. These low recruitments, slightly lower steepness and lower level of SSB depletion in 2011 all combine to produce the poorest SSB performance across all the pessimistic robustness trials (see Figure 6), with a probability of around 0.3 of attaining the interim rebuilding target by 2035. The MP does try to reduce catches to accommodate for the more pessimistic signals coming from the stock in the future (see Figure 7) but, given the differences between the reference and the STWindows-conditioned OM levels of initial stock status when the MP starts, it cannot do this enough to get close to the tuning targets.

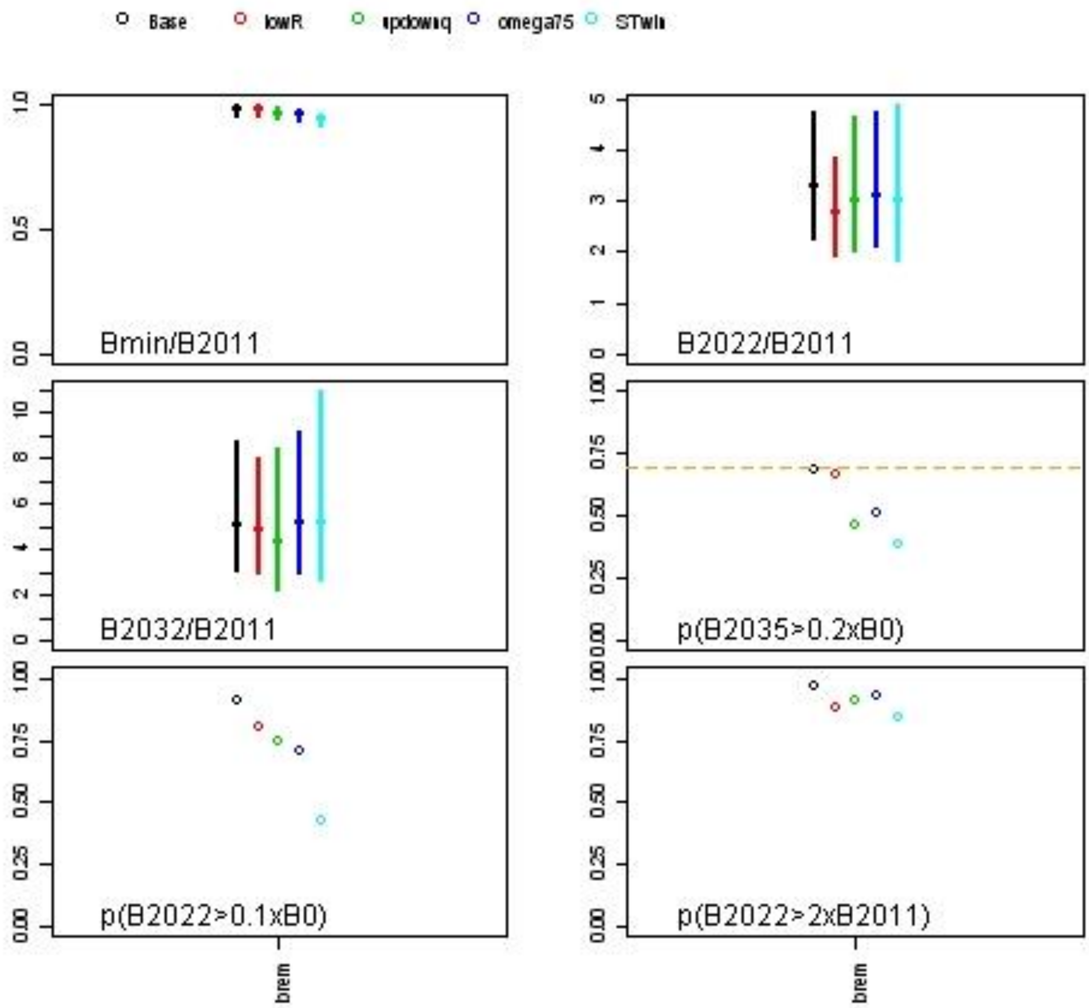


Figure 6: Spawning stock biomass (SSB) summary for MP1 on the Run 1 (base) specifications for the 'pessimistic' set of robustness trials. All statistics shown are the median and 80% CI and for the middle right tuning probability plot the dashed orange line denotes the actual tuning probability (0.7 in this case)

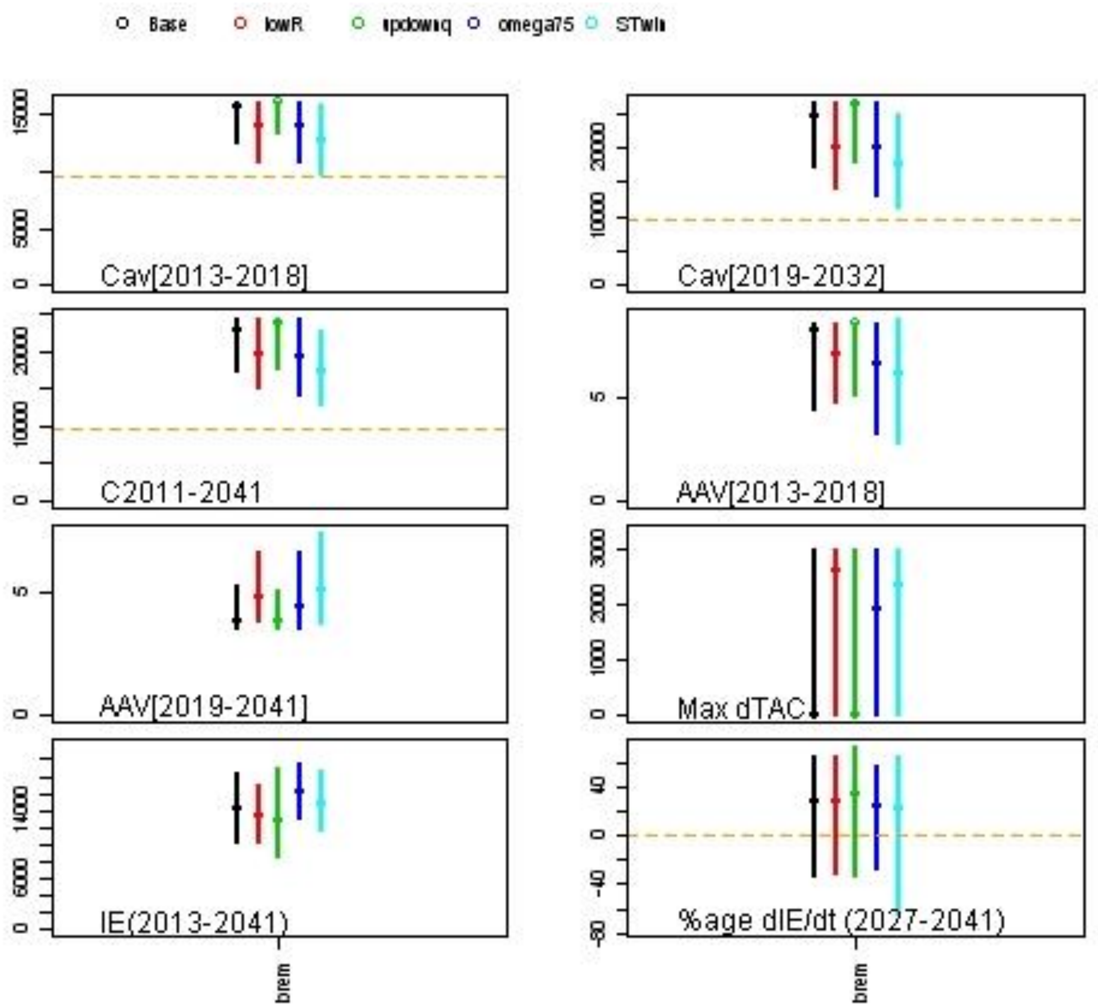


Figure 7: Catch summary for MP1 on the Run 1 (base) specifications for the 'pessimistic' set of robustness trials. All statistics shown are the median and 80% CI. For the three average catch plots (top left/right, middle left) the dashed orange lines denotes the current 9,449t catch.

Robustness of MP1 under constraints defined by Run 1

Pessimistic robustness trials

- In general, MP1 appears to be more robust to these trials than was observed in the previous MP trials (CCSBT-ESC/1009/11), and there is also a shift in the most influential trials.
- For the STwindows trial (alternative CPUE), the MP shows the worst short-term and long-term SSB rebuilding statistics. This is driven by the larger difference between the reference and STwindows-conditioned OMs, relative to last year. The STwindows trial does not have the strong CPUE increase observed in the reference OM, and this results in noticeably lower SSB depletion levels and slightly lower steepness levels relative to the reference case.
- The lowR scenario has a negative impact on the short-term level of SSB rebuilding, but the MP is able to act on the negative signals that move through the scientific aerial survey and CPUE data, reducing catches and almost achieving the tuning target by 2035.
- Unlike last year, the omega75 trial is not the most pessimistic for the MP. Lower estimates of steepness and a lower level of SSB depletion in 2011 contribute to poorer short-term performance of the MP. Conversely, the sub-linear CPUE-to-abundance relationship also does

not give the MP such strong increasing signals in the CPUE as the stock recovers, thereby slowing the rate of TAC increase later on. As a result, the SSB rebuilding of the stock is much improved on that observed last year, and while the MP cannot attain the tuning targets by 2035 it does have a probability of around 0.2 B_0.

- The updownq trial has been a focal point this year given the potential for an increase in catchability, not abundance, being a driver of the recent strong increases in CPUE (CCSBT-ESC/1107/11). The impact of this trial on the short-term rebuilding of the SSB was clearly negative, and comparable with the omega75 trial. What was perhaps more surprising was that the impact on longer term SSB rebuilding was worse than that of the omega75 trial, with the MP having a less than 0.5 probability of increasing the SSB to the interim rebuilding level by 2035. This was driven by bias introduced into the MP's estimates of key population parameters, which remains in the MP for some time after the cessation of the bias itself. This is driven by the adaptive nature of a model-based MP, such as MP1, and outlines the potentially more important issue of bias in the data (updownq) relative to more parametric/structural problems (omega75 and CPUE-to-abundance relationship) in relation to MP robustness.

Other robustness trials

- If we assume a lower proportion of estimated overcatch contributed to CPUE (c0s1l1), estimates of SSB depletion levels (relative to SSB_0 though not SSB_{2011}) tend to be higher than the reference OM, and the probability that SSB is above $0.2SSB_0$ in 2035 exceeds 0.7 by a small margin (Figure 8).
- If we assume a higher proportion of estimated overcatch contributed to CPUE (c2s1l1 and c3s1l1), estimates of SSB depletion levels (relative to SSB_0 though as with the c0s1l1 trial not B2011) tend to be below those in the reference OM, and the probability that stock is above $0.2SSB_0$ in 2035 is slightly less than 0.7 (Figure 8).
- For the LL1 fleet catch history, assuming Case 2 of the Market Report (CCSBT/0607/11; robustness trial c1s1l2) we observe greater SSB rebuilding, relative to both current levels and SSB_0, when compared with the reference case. The probability that SSB is above $0.2SSB_0$ in 2035 is slightly greater than 0.7 (Figure 8).
- Predicted future TACs set under MP1 were not substantially affected by assumptions about the proportion of overcatch that contributed to CPUE (Figure 9).
- For the Laslett CPUE series higher TACs would be expected than are expected under the base case (Figure 13). At the same time SSB rebuilding (relative to SSB_0) would be expected to be faster and greater in extent than predicted under the reference case (Figure 12).
- The overall performance of MP1 in terms of catch and stock rebuilding is predicted to be not substantially different from the base case for the alternative CPUE series run 3 or run 6 (Itoh, 2010, CCSBT/OMMP/1006/08) (Figures 12 and 13).
- We grouped the regime, mixtag, and aerflat trials into the "structural" group. For the regime shift scenario MP1 could not attain the interim SSB rebuilding target by 2035, caused by a smaller estimate of the regime-shifted SSB_0, though for the other trials the MP was able to meet the rebuilding targets. There was no observable effect on short-term rebuilding across all these trials. Catch performance across all these trials was very similar.

Robustness of MP1 under constraints defined by Run 3

The MP with constraints defined by Run 3, was also tested using the pessimistic robustness trials. The worst SSB performance observed was also for the updownq and STwin trials (see Figure 16).

The probability of reaching the target is approximately 0.2 for the updownq trial, and approximately 0.4. for the STwin trial. And the probability of reaching the interim target is approximately 0.6 for the updownq and lowest for the STwin trial at around 0.4.

This reduction in SSB rebuilding performance, for the updownq trial, is a direct result of the ability of the MP (for Run 3) to increase short-term TACs to much higher levels (see Figure 17) than for Runs 1 and 2 (with the 3000 t maximum change) given the upwardly biased CPUE data and the MPs interpretation of this as a strong increase in exploitable biomass.

7 Discussion

The performance of MP1 is determined by TAC change constraints, tuning period and probabilities, and the estimated greater productivity and higher recent recruitments from the OM (see Appendix 3). Thus, the changes in performance, relative to previous performance (CCSBT-ESC/1107/12), are a direct consequence of the reconditioned OM that includes the recent data. Compared with previous candidate MP results (SFMWG/1103/08), these results predict a high probability of attaining the short-term (and long-term) checkpoints, and without a substantial probability for initial TAC reductions, conditional on the current OM.

Runs 1-3 all tuned to their specific rebuilding criteria and, while minimal differences in performance were observed between them, Runs 1 and 2 tended to exhibit faster initial increases in TACs in the short-term (after the first TAC change in 2013), which was at, or close to, the maximum level of increase permitted. For Run 3 the initial TAC is marginally higher, but the median TAC changes are slower to increase, although eventually the median TAC exceeds the levels seen for Runs 1 and 2. In addition, Run 3 shows a noticeable probability for reductions in TAC in the second decision year following an initial increase. In terms of relative SSB rebuilding performance, Runs 1 and 2 reach a higher median level of SSB by 2035 or 2040, but for Run 3 the lower 10^{th} percentile is consistently higher in the later years.

For the ***lowR*** and ***omega75*** robustness trials, which were the most troublesome trials in the previous evaluation, MP1 (Run 1) performed fairly well almost attaining the tuning targets in the ***lowR*** case, and recovering the SSB to the interim rebuilding target with probability 0.5 in the ***omega75*** case. For the ***updownq*** trial, considered important given observed patterns in the most recent CPUE suggestive of catchability increases (CCSBT-ESC/1107/11), MP1 did not perform as well as for the ***lowR*** and ***omega75*** trials, and had a very low probability of rebuilding to the interim target for Run 3 trials. This suggested that bias in the data, rather than misspecifications in the CPUE-to-abundance relationship, might be a more influential factor on MP performance. The ***STwindows*** trial (an alternative CPUE series without as strong a positive signal in the recent years) suggests a considerably slower recovery (for Run 1), particularly in the short-term. This highlights the impact of changes in CPUE, and the effect of two years of data.

As in the previous evaluation work (CCSBT-ESC/1009/11), MP1 was observed to be robust to the various unreported catch, alternative CPUE, and OM structure robustness trials highlighted as a priority by the MP group.

Appendix 1 - Robustness trials for Run 1

Table 2: Robustness trials by group.

Grouping	Robustness Trial	Explanation
CPUE	Run3 Run6 Laslett	Alternative GLM model structure Alternative GLM model structure Alternative spatio-temporal effort distribution (most optimistic CPUE series)
Over-catch	c0s111 c2s111 c3s111	No over-catch to CPUE 50% over-catch to CPUE 75% over-catch to CPUE
Pessimistic	lowR updownq omega75 STwin	Recruitment ↓ 50% 2009-12 q 50% ↑ 2009 – 2013 Hyper-stable CPUE to biomass (CPUE proportional to 0.75 × abundance) Lowest of suite of CPUE series
Optimistic	troll Laslett	Inclusion of troll survey data Highest of suite of CPUE series
Structural	mixtag regime aerflat	Incomplete tag mixing – assumes surface fishing season Fs 50% higher than population. New B_0 (1978-2008) estimate Selectivity of 2-4 year olds in aerial survey is [1,1,1]

Overcatch group of robustness trials

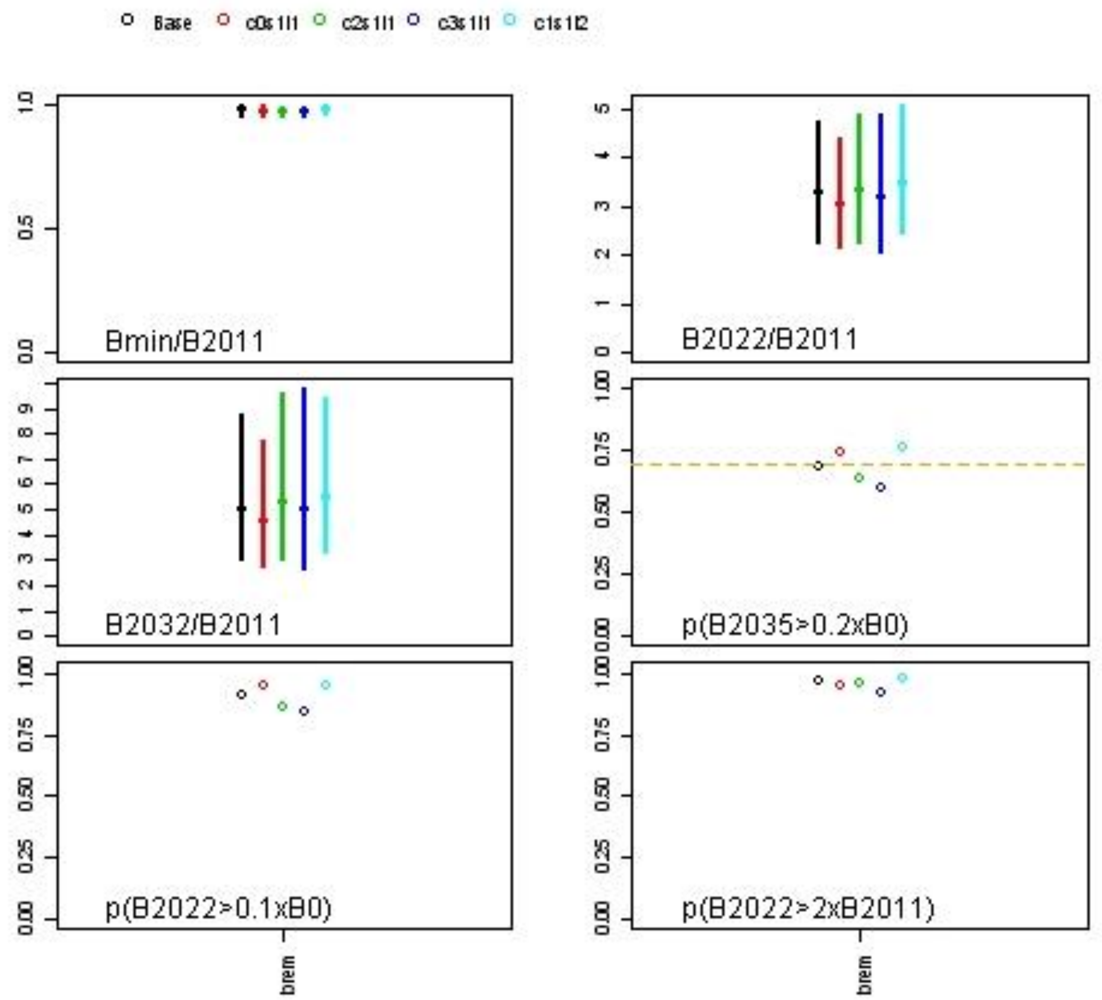


Figure 8: Summary of spawning stock biomass (SSB) performance statistics on the Run 1 (base specifications) for the overcatch robustness trials. The biomass ratio plots give the median and 80% confidence interval estimates. The dashed orange line plotted in the right centre panel denotes the actual tuning probability (0.7 in this case.)

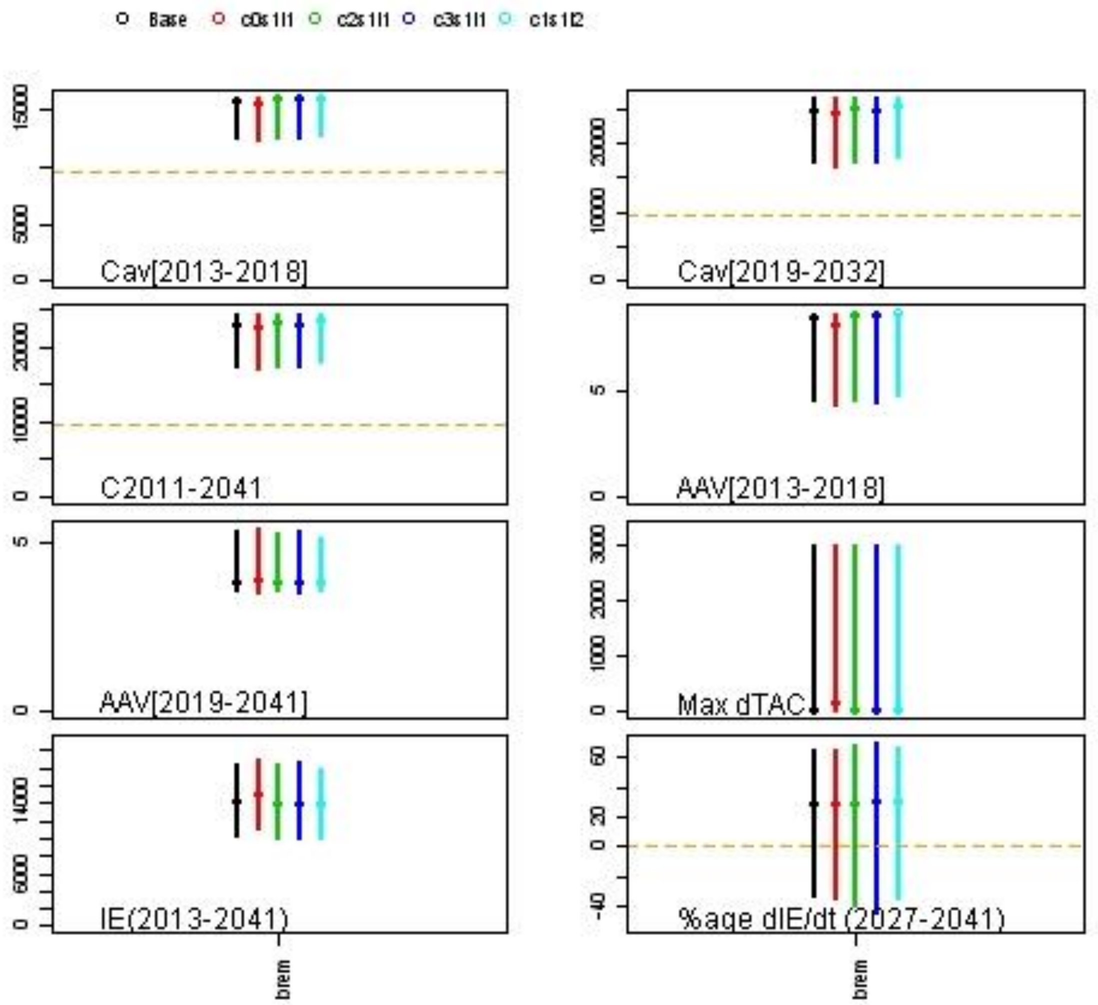


Figure 9: Catch summary for MP1 on the Run 1 (base) specifications for the 'overcatch' set of robustness trials. All statistics shown are the median and 80% CI and for the middle right tuning probability plot the dashed orange line denotes the actual tuning probability (0.7 in this case).

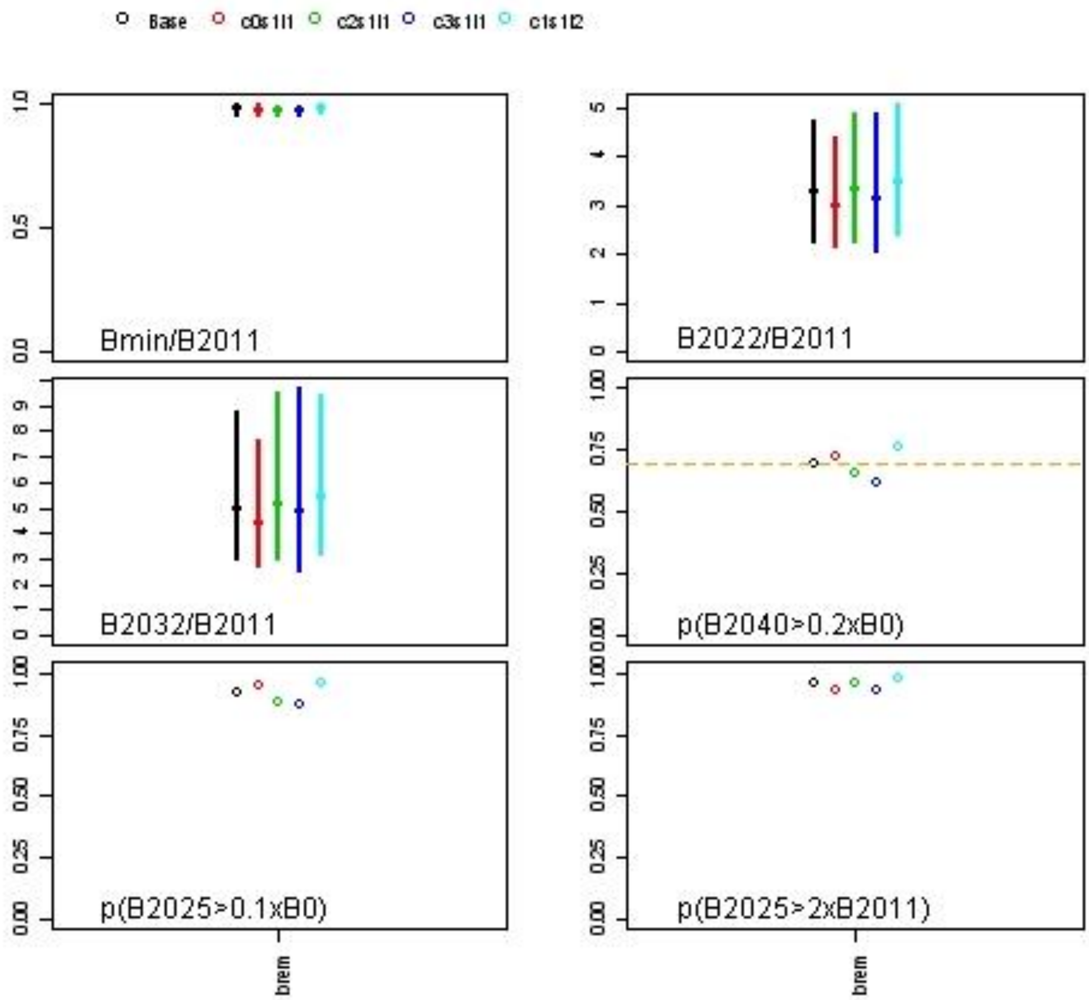


Figure 10: Spawning stock biomass (SSB) summary for MP1 on the Run 1 (base) specifications, except that tuning year is 2040, for the 'overcatch' set of robustness trials. All statistics shown are the median and 80% CI and for the middle right tuning probability plot the dashed orange line denotes the actual tuning probability (0.7 in this case).

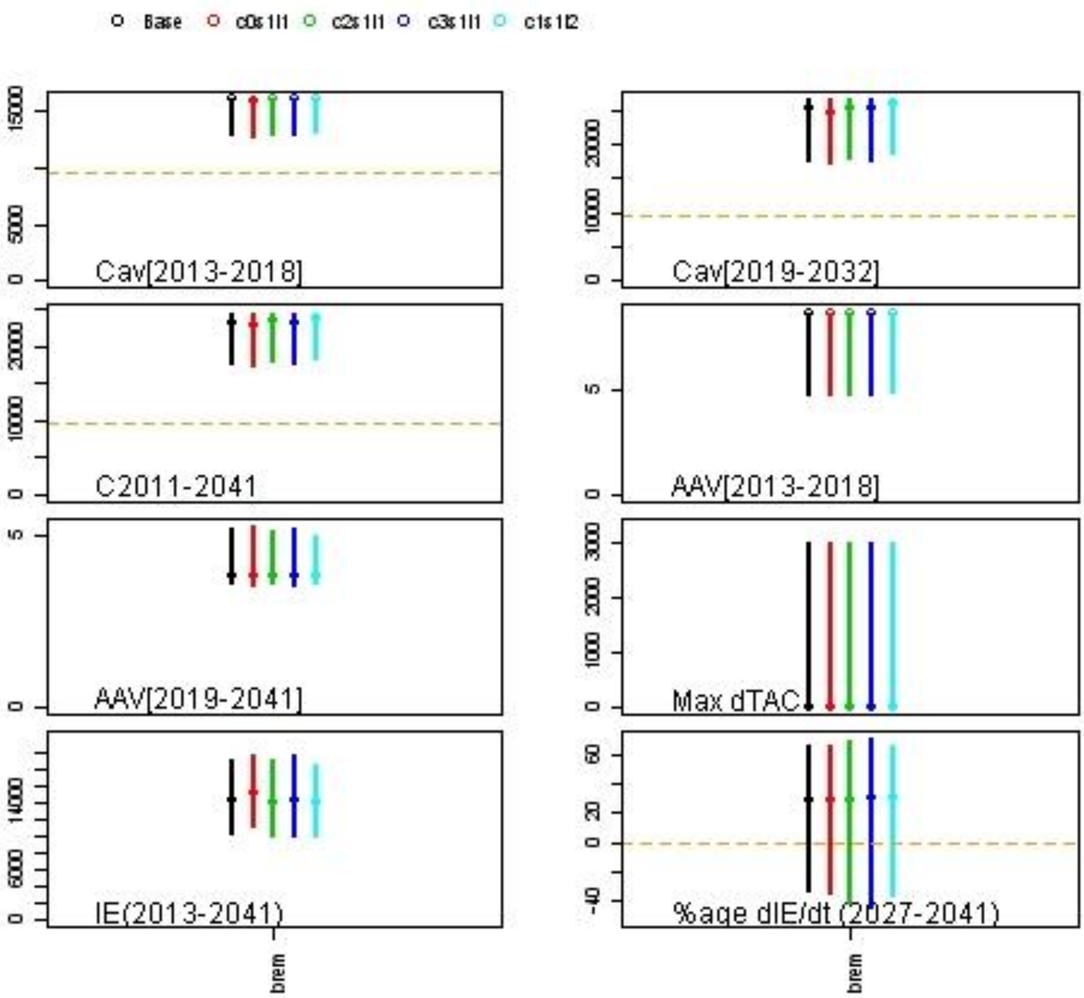


Figure 11: Catch summary for MP1 on the Run 1 (base) specifications, except that the tuning year is 2040, for the 'overcatch' set of robustness trials. All statistics shown are the median and 80% CI and for the middle right tuning probability plot the dashed orange line denotes the actual tuning probability (0.7 in this case).

CPUE group of robustness trials

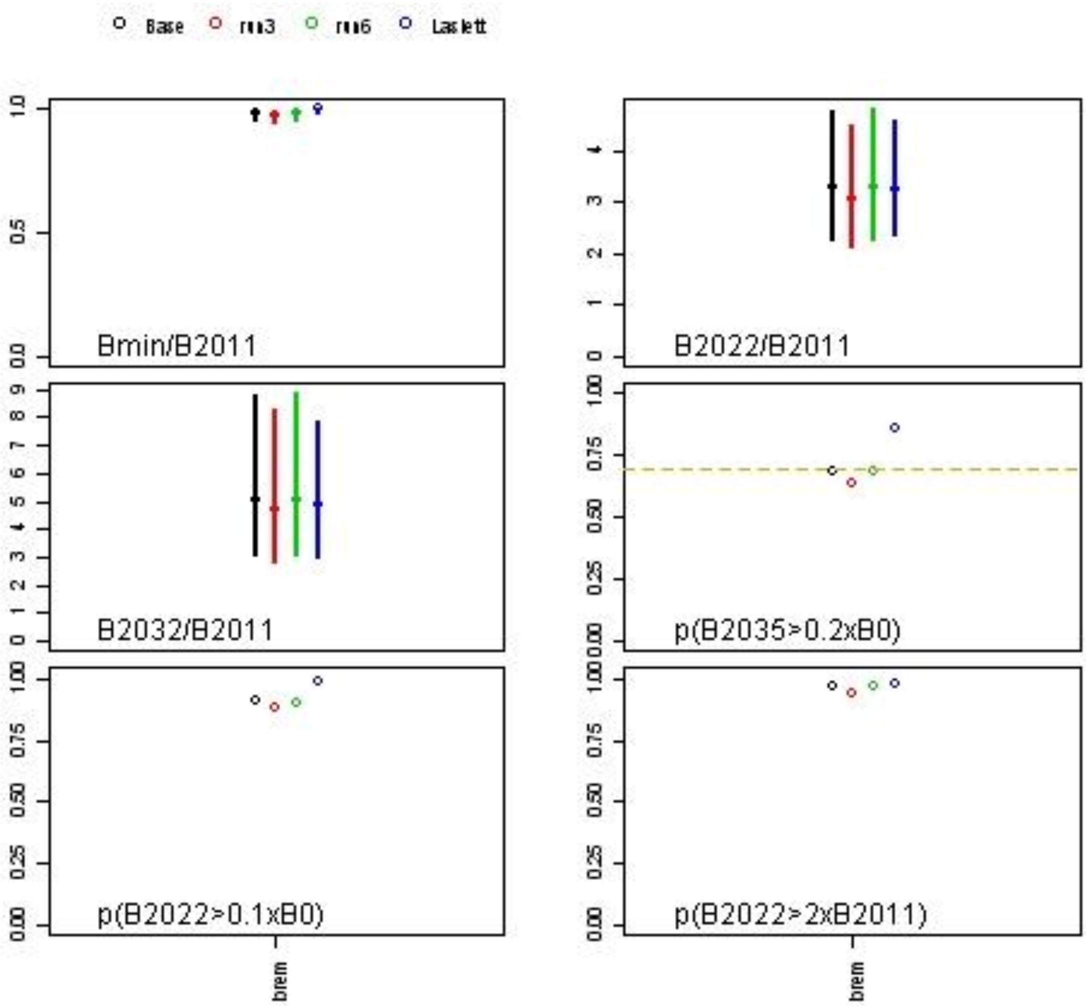


Figure 12: Spawning stock biomass (SSB) summary for MP1 on the Run 1 (base) specifications for the 'CPUE' set of robustness trials. All statistics shown are the median and 80% CI and for the middle right tuning probability plot the dotted orange line denotes the actual tuning probability (0.7 in this case).

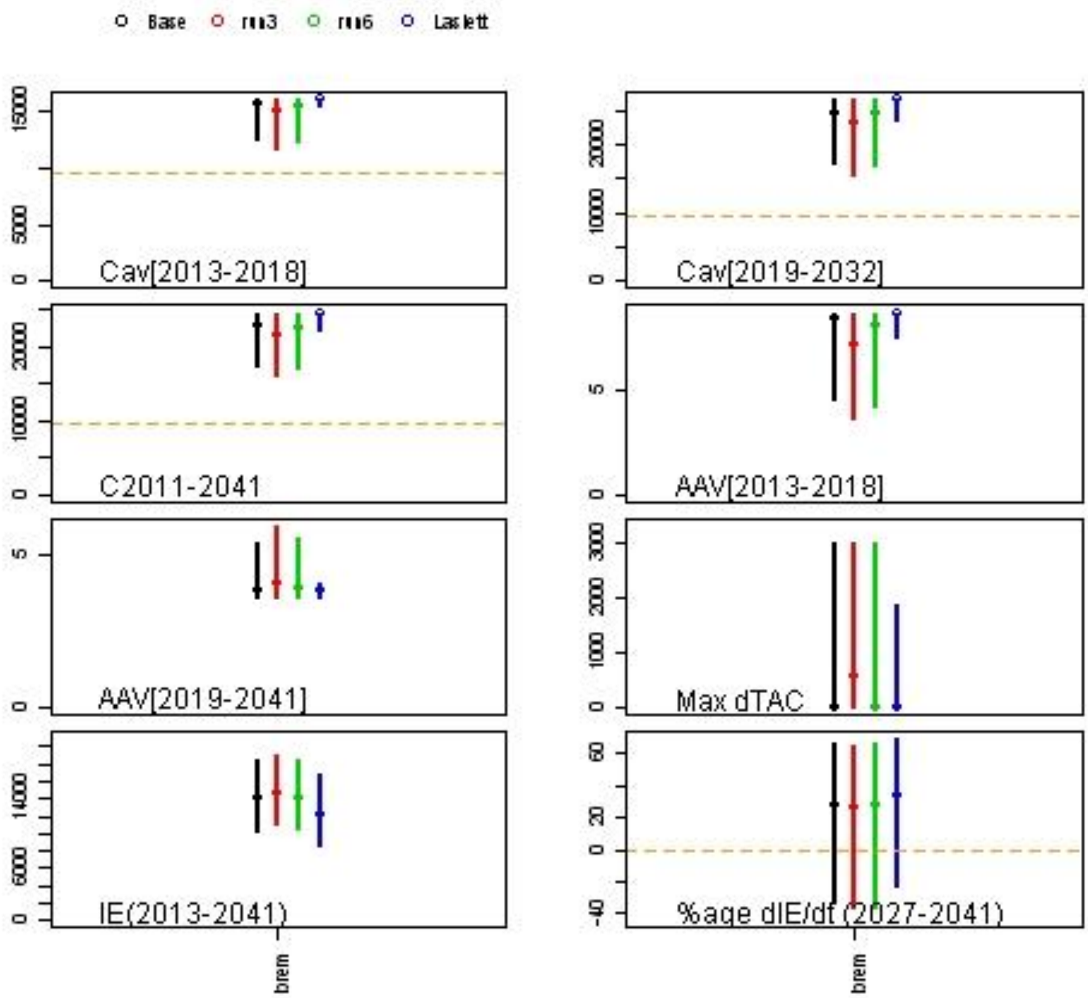


Figure 13: Catch summary for MP1 on the Run 1 (base) specifications for the 'CPUE' set of robustness trials. All statistics shown are the median and 80% CI and for the middle right tuning probability plot the dotted orange line denotes the actual tuning probability (0.7 in this case).

Structural group of robustness trials

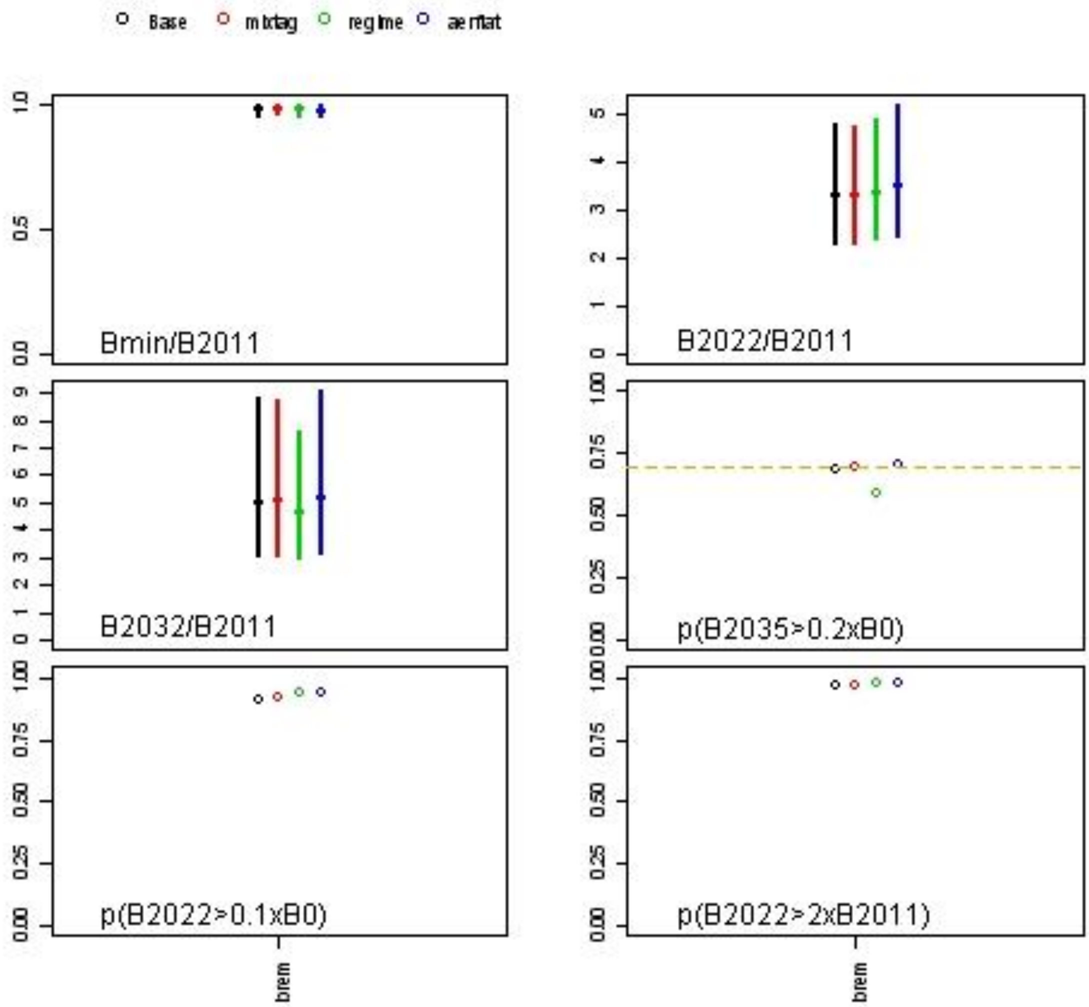


Figure 14: Spawning stock biomass (SSB) summary for MP1 on the Run 1 (base) specifications for the 'structural' set of robustness trials. All statistics shown are the median and 80% CI and for the middle right tuning probability plot the dotted orange line denotes the actual tuning probability (0.7 in this case).

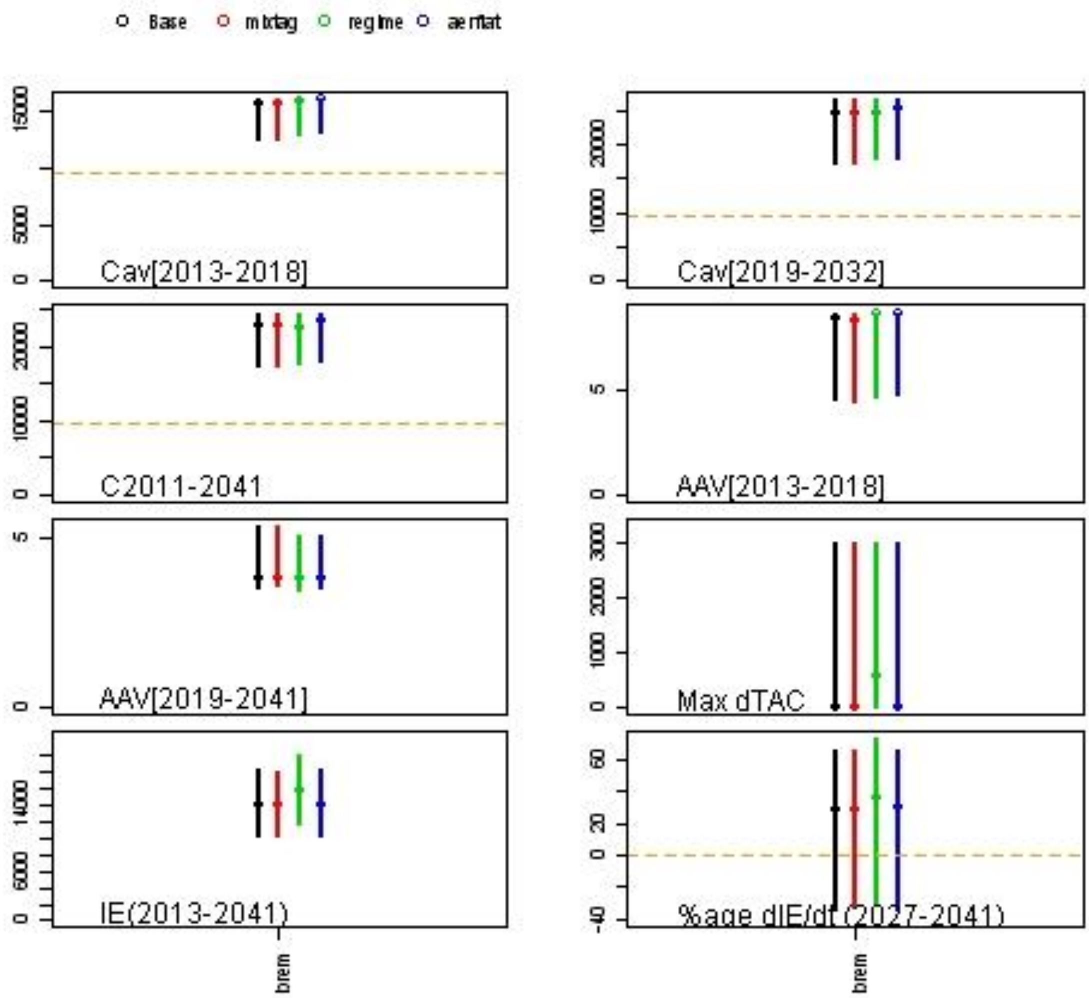


Figure 15: Catch summary for MP1 on the Run 1 (base) specifications for the 'structural' set of robustness trials. All statistics shown are the median and 80% CI and for the middle right tuning probability plot the dotted orange line denotes the actual tuning probability (0.7 in this case).

Appendix 2 - Pessimistic Robustness Trials for Run 3

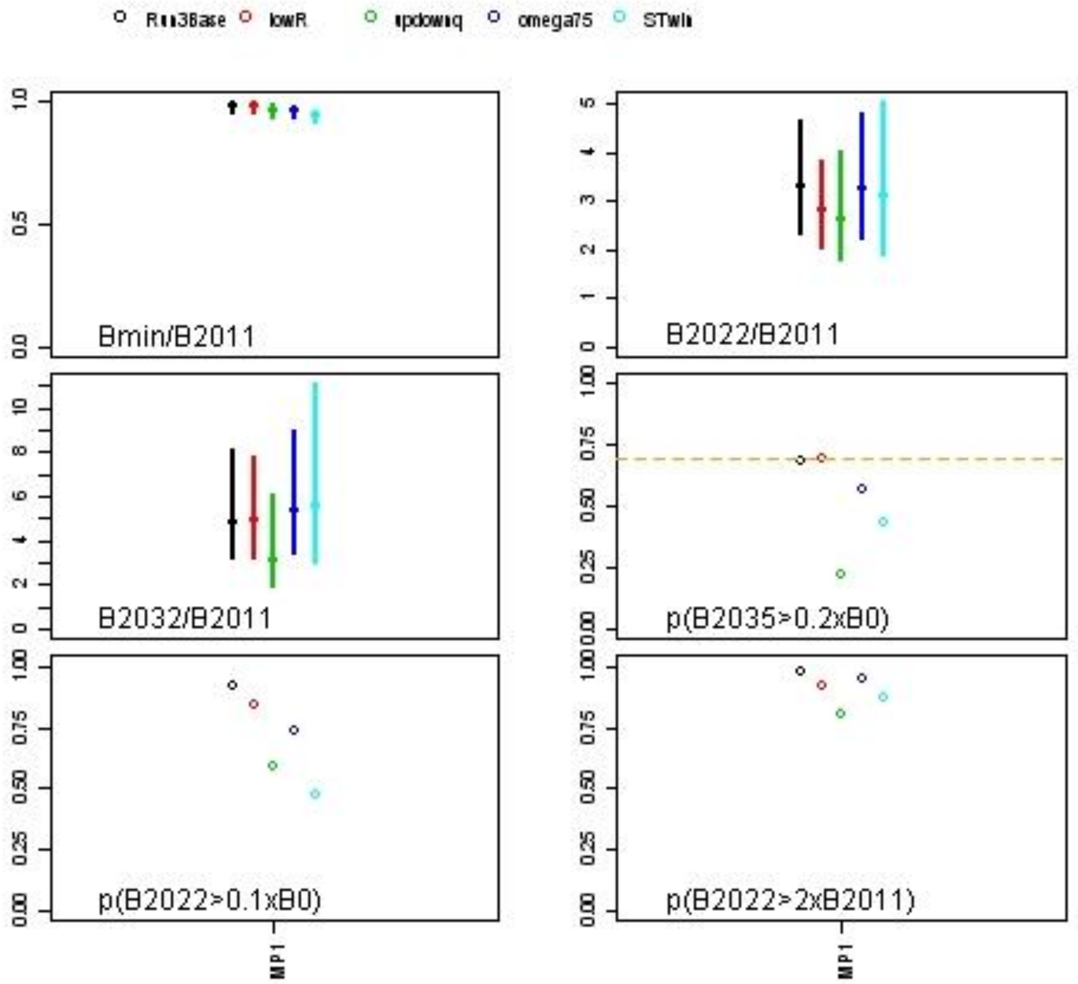


Figure 16: Spawning stock biomass (SSB) summary for MP1 on the Run 3 specifications for the 'pessimistic' set of robustness trials. All statistics shown are the median and 80% CI and for the middle right tuning probability plot the dashed orange line denotes the actual tuning probability (0.7 in this case).

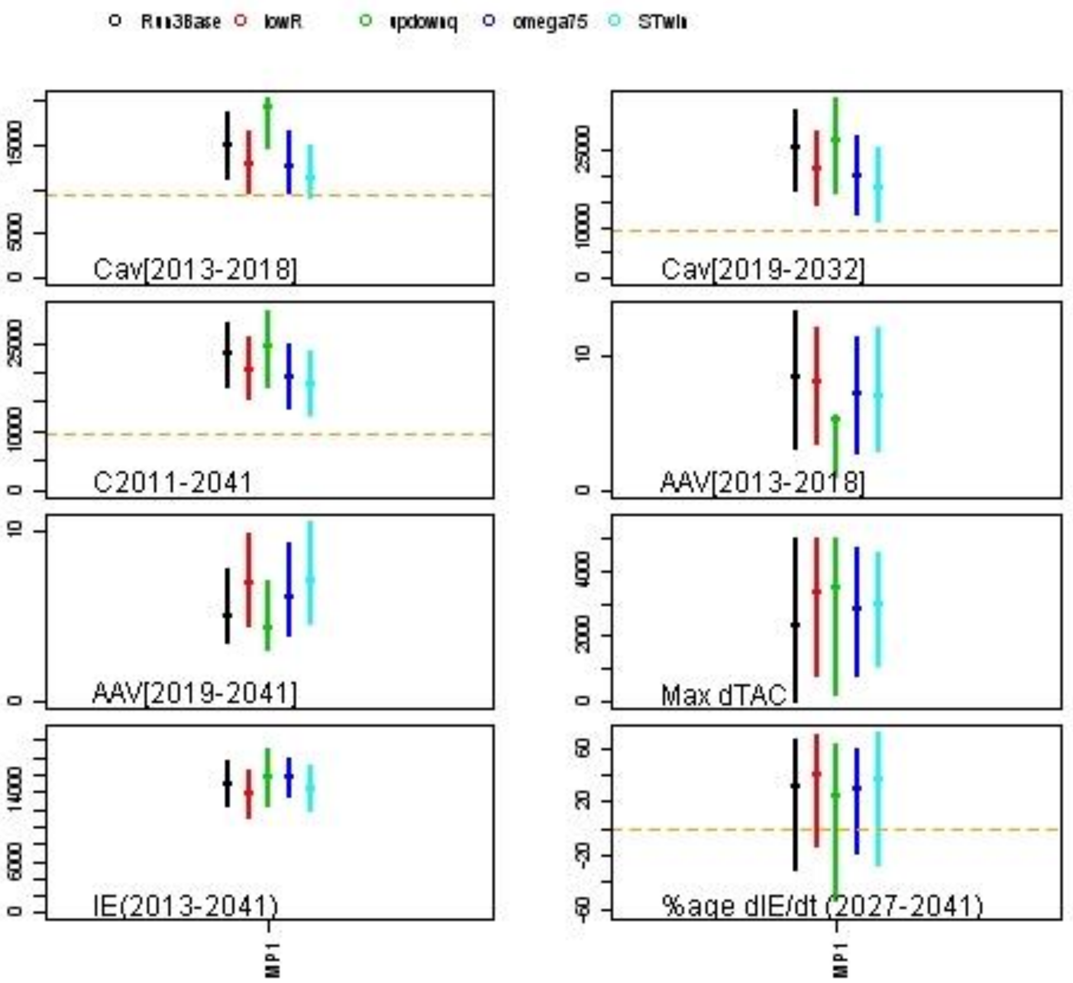


Figure 17: Catch summary for MP1 on the Run 3 specifications for the 'pessimistic' set of robustness trials. All statistics shown are the median and 80% CI. For the three average catch plots (top left/right, middle left) the dashed orange lines denotes the current 9,449t catch.

Appendix 3 - Performance statistics for full reduced set of MPs

Table 3: Performance statistics for the 70% tuning level for the full reduced set from the previous MP paper (CCSBT-SFMWG/1103/08). No limit refers to an MP with no constraint in the reduction (or increase) in TAC in the first TAC setting period.

	MP scenario	Tuning probability	Tuning year	Lag	Max TAC change	Check point year(t)	P[B_t>0.1B_0]	P[B_t>2B_{2010}]	B_{2025}/B_{2010}	Median avg. catch (2013-2022)	Median avg. catch (2013-2039)	Year 1.2*B_{2010} reached (median)
1	Run1	0.7	2035	1	3000	2022	0.91	0.96	3.65	15731	22831	2015
2	Run2	0.7	2040	1	3000	2025	0.92	0.96	3.61	15957	23268	2015
3	Run3	0.7	2035	1	5000	2022	0.92	0.97	3.68	15062	23442	2015
4	IRP 3 years 3000 t	0.7	2035	1	3000	-	-	-	-	-	-	Unable to tune
5	Run5	0.7	2035	0	3000	2022	0.90	0.96	3.53	15690	22454	2015
	No Limit	0.7	2035	1	3000	2022	0.89	0.95	3.52	16013	22380	2015
	No Limit	0.7	2040	1	3000	2025	0.90	0.95	3.45	16401	22731	2015
	IRP 2 years 3000 t	0.7	2035	1	3000	-	-	-	-	-	-	Unable to tune
	IRP 2 years 3000 t	0.7	2040	1	3000	-	-	-	-	-	-	Unable to tune
	IRP 2 years 5000 t	0.7	2035	1	3000	-	-	-	-	-	-	Unable to tune
	IRP 2 years 5000 t	0.7	2040	1	3000	-	-	-	-	-	-	Unable to tune
	IRP 4 years 3000 t	0.7	2035	1	3000	-	-	-	-	-	-	Unable to tune
	IRP 4 years 3000 t	0.7	2040	1	3000	-	-	-	-	-	-	Unable to tune
	IRP 4 years 5000 t	0.7	2035	1	3000	-	-	-	-	-	-	Unable to tune
	IRP 4 years 5000 t	0.7	2040	1	3000	-	-	-	-	-	-	Unable to tune

References

Anon. (2010) Report of the Extended Scientific Committee for the Fifteenth Meeting of the Scientific Committee, 4-9 September 2010, Taipei, Taiwan.

B. Barnes, M. Chambers, P. Tennant, R. Hillary, F. Giannini and G. Begg. (2011) Further Exploration of management procedure MP1 for southern bluefin tuna. CCSBT-SFMWG/1103/08.

P. Eveson. (2011) Updated growth estimates for the 1990s and 2000s, and new age-length cut-points for the operating model and management procedures. CCSBT-ESC/1107/9, Bali, Indonesia.

R. Hillary, F. Giannini, P. Eveson, M. Basson, C. Davies, B. Barnes and G. Begg. (2010) Results of the performance of the BREM suite of candidate management procedures. CCSBT-ESC/1009/11.

R.M. Hillary, A. Preece, C. Davies, B. Barnes and G. Begg. (2011) Reconditioning of the SBT model: exploratory data analysis, fitting performance and current stock status. CCSBT-ESC/1107/11, Bali Indonesia.

R.M. Hillary and A. Preece. (2011) Updated technical specifications and performance analyses for MP1. CCSBT-ESC/1107/12, Bali, Indonesia.

T. Itoh (2010) CPUE standardization up to 2009 data, CCSBT/OMMP/1006/08

X., Lou, T. Hidaka, A. Bergen and T. Kageyama. (2006) Independent review of Japanese southern bluefin tuna market anomalies. CCSBT/0607/11.