CCSBT-ESC/1009/11

Australian Government

Australian Bureau of Agricultural and Resource Economics – Bureau of Rural Sciences

Results of the performance of the BREM suite of candidate management procedures

Hillary, R.[1]
Giannini, F.[2]
Eveson, P.[1]
Basson, M.[1]
Davies, C.[1]
Barnes, B.[2]
Begg, G.[2]

Prepared for the CCSBT Extended Scientific Committee for the 15th Meeting of the Scientific Committee 4-10 September 2010
Taipei, Taiwan

[1] CSIRO Marine and Atmospheric Research, Hobart, Australia
[2] Department of Agriculture, Fisheries & Forestry, Australian Bureau of Agricultural and Resource Economics – Bureau of Rural Sciences, Canberra, Australia

Table of Contents

Abstract

A revised set of candidate management procedures (MPs), based around the Biomass Random Effects Model (BREM) estimation framework, were evaluated given the recommendations coming from the 3rd Operating Model and Management Procedure (OMMP) Technical Meeting. The original BREM harvest control rule was extended to allow for total allowable catch (TAC) stability effects (i.e. where current TAC is some proportion of the previous year TAC and some proportion of the TAC as predicted by the original MP) to smooth the catch trajectories and avoid overly strong increases in exploitation rates in later years. More reactive MPs cut catches harder earlier, mitigate the risk of further spawning stock biomass (SSB) decline better, and yield higher average catches in the medium to long term, relative to smoother/less reactive MPs which do not cut catches as early, do not rebuild as fast and yield lower medium- to long-term average catches. Less reactive MPs tend to stop the increasing of exploitation rates in later periods that is associated with more reactive MPs, but the memory effect must be very strong to do this, which negatively impacts short-term SSB rebuilding. Exploitation rates, for all alternatives, are unlikely to increase to unsustainable levels over the testing timeframe. The robustness trials with the strongest impact were the pessimistic and catch per unit effort (CPUE) related trials, whereas the trials relating to more optimistic levels of SSB plus the unreported catch, tag-mixing and regime-shift trials had little impact (across MP alternatives).

Introduction

This paper details the performance evaluation of the updated Biomass Random Effects Model (BREM) suite of candidate management procedures (MPs), given the recommendations at the 3rd Operating Model and Management Procedure (OMMP) Technical Meeting (Anon., 2010) and the revised set of robustness grids and MP settings. Both the updated performance assessment of the estimation model of the BREM suite of MPs and the technical specifications of the changes made to the core harvest control rule (HCR) can be found in document *CCSBT-ESC/1009/10.* The main recommended changes made by the OMMP group to the BREM MPs were as follows:

- To try and reduce the reactivity of the MPs by including a total allowable catch (TAC) smoothing/status quo term – essentially that the TAC in the year when a change occurs is some proportion of the previous TAC plus a change given by the underlying MP;
- To make the response of the MPs stronger than linear to biomass levels below the target level to hasten stock recovery and minimise risk;
- To change recruitment in the HCR to be of a target/limit-type form, relative to a historical average based on real scientific aerial survey data, and to be quadratic when below this level.

Changes to the BREM suite of MPs

As stated, the specifics of the changes made to the HCR to accommodate these recommendations can be found in *CCSBT-ESC1009/10* but the changes are a little more general than those requested by the OMMP Working Group and are worth stating again here:

- The TAC memory effect was included as follows: $TAC_{y+1} = \psi_y TAC_y + (1-\psi_y) TAC_{brem}$, where the memory weighting term ψ_y is between 0 and 1 and may change by year, and TAC_{brem} is the catch as predicted by the underlying BREM HCR (Eqns. 8-13, *CCSBT-ESC/1009/10*).
- The response of the HCR to levels of biomass above and below the target was permitted to be non-linear but related, in the sense that the power of the biomass ratio B_y/B^* (with B^* the "target" – the CPUE expected at the interim rebuilding target) is equal to $1+\varepsilon_b$ (ε_b>0, so stronger than linear) when the biomass is below the target, but equal to $1-\varepsilon_b$ (so weaker than linear) when the biomass is above the target (Eqn. 9, *CCSBT-ESC/1009/10*). The reason for this change was that a more reactive MP for SSB, although leading to faster rebuilding, increases the chance of much larger increases in catches (which could result in higher exploitation rates) later on, a problem outlined at the OMMP meeting. By making the described change, the reaction to levels above the target is softened which may help to ameliorate this type of dynamic.
- A similar power relationship was defined for future moving average recruitment levels above or below the historical average – i.e. the power of the recruitment ratio was set to $1+\varepsilon_r$ (ε_r>0) for future averages above the historical average and to $1-\varepsilon_r$ for future averages below the historical average (Eqns. 10-11, *CCSBT-ESC/1009/10*). This was flexible enough to still meet the quadratic response suggested by the OMMP group ($\varepsilon_r = 1$) whilst allowing the exploration of further options.
- A similar change to that made for the treatment of the recruitment parameters was made for using the biomass growth parameters, g_y, where a current moving (arithmetic) mean level was compared, as a ratio, with the mean observed level over the period for which there was real data. The TAC_{brem} predicted by the biomass and recruitment terms was then adjusted up or down based on the size of this ratio, weighted via an "influence" exponent, $\gamma \in [0,1]$, with $\gamma = 0$ denoting no influence at all on TAC through to $\gamma = 1$ denoting a linear response (Eqns. 12-13, *CCSBT-ESC/1009/10*).

Key alternatives explored with the new more general BREM framework

Although we have one generic MP, as defined in *CCSBT-ESC/1009/10* and above, we explored several different alternatives for the non-tuning parameters of the BREM HCR (i.e., all parameters except the reference catch level tuning parameter, δ). To be able to have different TAC memory weightings for different future years the following definition for ψ_y was used:

$$(1) \qquad \psi_y = \begin{cases} \psi^h & y < y_1 \ \& \ y > y_2 \\ \psi^l & y \geq y_1 \ \& \ y \leq y_2 \end{cases},$$

where ψ^h and ψ^l are the "high" and "low" memory weightings, respectively, and y_1 marks the year we change to the low weighting (the more active phase) and y_2 marks the year we change back to the high weighting (back to the more passive phase). Given this structure for the memory weighting-by-year the following 5 alternatives were defined:

1. **S1**: "weak" and fixed TAC memory effect of $\psi^h = \psi^l = 0.25$, medium biomass-trend and recruitment asymmetry ($\varepsilon_B = \varepsilon_r = 0.5$) with $\gamma = 1$.
2. **S2**: "medium" and fixed TAC memory effect of $\psi^h = \psi^l = 0.5$, medium biomass-trend and stronger recruitment asymmetry ($\varepsilon_B = 0.5$, $\varepsilon_r = 0.75$) with $\gamma = 1$.
3. **S3**: "strong" and fixed TAC memory effect of $\psi^h = \psi^l = 0.75$, medium biomass-trend and total recruitment asymmetry ($\varepsilon_B = 0.5$, $\varepsilon_r = 1$) with $\gamma = 1$.
4. **S4**: time-dependent TAC memory with passive phases between 2012-2015 and 2026-2039 (so $y_1 = 2015$ and $y_2 = 2026$) and a strong memory effect in the passive phase ($\psi^h = 0.85$) and a weak memory effect in the active phase ($\psi^l = 0.15$), with medium-strong and total biomass and recruitment asymmetry ($\varepsilon_B = 0.75$, $\varepsilon_r = 1$), respectively, and as before $\gamma = 1$.
5. **S5**: for reference the most BREM_1-like run with no memory effect at all ($\psi^h = \psi^l = 0$) and no asymmetry ($\varepsilon_B = \varepsilon_r = 0$) with $\gamma = 1$.

Alternative MP scenarios **S1**-**3** are intended to cover progressively stronger TAC memory whilst increasing the ability of the MP to react to negative biomass and recruitment trends – we know the trade-off between reactive and less-reactive MPs is in terms of short-term risk of further biomass decline. The 4th scenario (**S4**) is an extreme one – it was noted previously (see *CCSBT-OMMP/1006/05*) that the strongest TAC decreases are at the start of the time period so we strongly constrain the first few years of the MP to be very passive, give it a "reactive" period during which there are less constraints (beyond the maximum change) to the MP in terms of catch changes, then re-instate the passive phase after a set number of years to avoid rapid increases in the TAC if strong recovery occurs post-active phase. MP scenario **S5** was defined as a comparison, with this being the most similar the new BREM HCR can be to the old, reactive version presented at the OMMP meeting (BREM_1 in *CCSBT-OMMP/1006/5*).

Additional statistics generated given the OMMP discussions

Key concerns with the more reactive MPs presented at the OMMP meeting (2010) was that, while allowing for faster rebuilding, the later increases in catches might be faster than the stock recovery, with the potential to lead to eventual biomass decreases, albeit at very long timeframes (i.e. post 2040). While there is some difference in how an MP with an explicit "target" (i.e. a specific value/set of values of key MP variables that the HCR will try to attain over time such as the BREM MPs) versus MPs based on CPUE trends over time would act in such a scenario; this was the main motivator for requesting "smoother" MPs in terms of the catch trajectories. A form of exploitation rate would be a useful variable to monitor and detect such an effect but it is not trivial to generate a "mean" exploitation rate from the projection runs, given they are by fishery and are each linked to selectivity and spread across a wide range of age classes. One suggestion from the OMMP meeting was to use the *implied effort*: essentially catch divided by catch per unit effort (CPUE). This has a number of advantages as it's a simple and readily calculable proxy for exploitation rate (although it should not to be looked at too closely in the CPUE grid alternatives in some years or for the **omega75** option), and doesn't have the lag interpretation issues of say catch divided by spawning stock biomass (SSB). In this paper we use the implied effort as a proxy for exploitation and developed a simple statistic to explore the possible impacts of overly rapid increase in catches in the later years of the MP runs. To measure this effect we fitted a linear model to the implied effort series $E_y = TAC_y / CPUE_y^{LL}$ from 2025 to 2039 (the period during which the larger catch increases tended to occur). We then calculated the percentage change in the ratio of this fitted implied effort series, $\hat{E}_y$, from the start to the end of the period ($\left(1 - \hat{E}_{2025} / \hat{E}_{2039}\right) * 100$) to see whether the exploitation rate is increasing and, if so, by how much.

Results

The results section is split into two parts: (i) comparative performance of the BREM MP alternatives on the reference grid, and (ii) performance of the BREM MP alternatives across the robustness trials.

Performance on the reference grid

The performance of the 5 BREM MP alternatives (**S1** – **S5**) was tested on the reference grid. The OMMP Working Group recommended a default option for testing of MPs: tuning option 5 (70% chance that SSB will be above $0.2B_0$ by 2040 – noted as 5d in the results), a maximum TAC change of 3000 t, and a TAC-setting interval of every 3 years (denoted by 'd' in the results) with an implementation lag of 1 year. All MPs were to be tuned to this option on the reference grid, **base5h** (using the 5 levels of steepness). Comparisons were also made by the OMMP Working Group using tuning level 2 (70% chance that SSB will be above $0.2B_0$ by 2035 – noted as 2d in the results) with the other options as stated. Figures 1- 6 show the performance for each of the 5 MP alternatives for both of these tuning options. These two tuning options show similar relative performance between the MP alternatives. As was noted in *CCSBT-OMMP/1006/05*, the greatest difference in relative behaviour of the MP alternatives was seen when considering their performance for tuning option 3 (90% chance that the spawning biomass will be above $0.2B_0$ by 2035). Tuning level 3 is the most extreme tuning level provided by the Commission in terms of short-term catch reductions required to attain the target biomass level in the shorter timeframe and with the highest probability. As was noted in the OMMP (2010) report, tuning to this level provided little contrast among MPs and so its use to judge performance was considered uninformative. Hence our discussion will concentrate on analysing performance at the recommended default tuning option (option 5), with reference to tuning option 2 where significant differences in performance and/or interesting features occur. The graphics used to illustrate the more detailed performance of the MPs are fairly complex, and cover multiple timeframes. To summarise the key short-term performance of, and trade-offs associated with, the MP alternatives, Table 1 details for tuning level 5 (TAC option d, lag 1) the following six statistics:

1. *Short-term rebuilding statistic 1:* probability that the SSB is greater than 10% of B_0 by 2025.
2. *Short-term rebuilding statistic 2:* probability that the SSB is greater than twice the SSB of 2009 by 2025.
3. *Average TAC (short-term):* average TAC from 2013 to 2025.
4. *CV of the TAC (short-term):* coefficient of variation in average TAC from 2013 to 2025
5. *Average TAC (long-term):* average TAC from 2013 to 2039.
6. *CV of the TAC (long-term):* coefficient of variation in average TAC from 2013 to 2039
7. *Average CPUE:* average CPUE from 2013 to 2025
8. *CV of the CPUE:* coefficient of variation in average CPUE from 2013 to 2025.

Figure 1 shows the median and lower 10th percentile SSB relative to 2009 SSB, over the projected time period. The plot suggests two distinct groups of MPs in the 5 alternatives; **S1**, **S2** and **S5** (more reactive), and **S3** and **S4** (less reactive) in the other. The first group follows much the same trajectory for both the median SSB and the 10th percentile SSB. They indicate lower risk than the second group in terms of future SSB levels. The second group follows much the same trajectory for the median SSB but **S3** has the lowest 10th percentile SSB trajectory, suggesting that this scenario is the most risky of the 5 alternatives in terms of future recruitment and stock rebuilding, particularly in the short-to-medium term.

It is worth noting the "2d lag1" (tuning option 2 – 70% chance that SSB will be above $0.2B_0$ by 2035) plot in Figure 1 shows that, although **S3** and **S4** are considered more risky in the period leading to the tuning year (i.e. 2035), after the tuned year it is actually these alternatives that perform best in terms of SSB. This is likely due to both alternatives having a strong memory effect in this later stage, so less changeable TACs and the improving stock conditions have less of an effect on the TAC trajectory than for the other alternatives due to the high recruitment and biomass asymmetry in **S3** and **S4** – they are less reactive to trends above the target biomass and historical mean scientific aerial survey-predicted recruitment level. To some extent, this is behaviour emerging from the imposition of the tuning condition at either 2035 or 2040. Tuning ensures there is a long-term risk level (0.6, 0.7 or 0.9) at some specified time (2035 or 2040) at which *all* the MPs converge. These MPs have been developed with stock rebuilding as the focus, and so it is likely that a different type of MP would be required once the stock reached the interim target rebuilding levels.

Figure 2 shows the median and lower 10th percentile TACs set over the projected time period. For all alternatives, in 2013 and 2016, there are cuts in the TACs in the median and 10th percentile trajectories. MP scenario **S3** over this timeframe takes smaller cuts and consequently does not reach as high TAC levels pre-2025 as alternatives **S1**, **S2** and **S5** but in the longer term, does reach similar (if not higher) levels to **S4**. The difference between the two groups is clear from about 2025 onwards, with **S3** and **S4** having less of an increase in TACs in the later years than the other alternatives. Thus alternatives **S1**, **S2** and **S5** make bigger initial cuts in TAC, but provide lower medium to long-term risk to the stock than **S3** and **S4**.

Figure 3 shows the median and lower 10th percentile implied effort over the projected time period for all MP alternatives and for tuning options 2 and 5. For both tuning options all the MP alternatives result in a significant decrease (for both the median and lower 10th percentile) in the exploitation rate up to around 2020. For tuning option 5, as the stock recovery increases, and the more positive signals in the data are detected by the MPs, we observe a gradual increase in the exploitation rates which end in 2039 at higher levels for alternatives **S1**, **S2** and **S5** than for **S3** and **S4**. For tuning option 2 we only really see any future increase in exploitation rate for alternatives **S1**, **S2** and **S5** whereas for **S3** and **S4** they stay close to the 2020 levels until 2039. This is driven by the fact that (a) the tuning option 2 target is more conservative (with the reference catch part of the HCR being smaller than for tuning option 5), and (b) the stronger asymmetry and TAC memory for alternatives **S3** and **S4** acts to dampen the propensity of the integral BREM part of the HCR trying to increase TACs given strongly positive signals. For all MP alternatives and for both tuning options the median and lower 10th percentiles of the exploitation rate proxy are significantly lower in 2039 than in 2013.

Figures 4 and 5 summarise the trade-off between mean catch and relative SSB and CPUE (to 2009 levels) for the years up to the short-term check point. The plots also illustrate the 80% confidence intervals for these values and the distinct groups are again evident. As we have previously seen in *CCSBT-OMMP/1006/05*, Figure 4 suggests that greater SSB recovery (greatest for **S1**, **S2** and **S5**) is associated with decreased average catches over the short-term, but we should also highlight that this SSB recovery implies concurrent CPUE recovery. Thus the trade-off, from an economic viewpoint, may be less extreme than when just considering catch versus SSB recovery, given the expected higher efficiency in taking the TAC, although this argument only applies to fisheries where the production function is linear in catch, effort and biomass ($C = qEB$) – i.e. most likely the longline fisheries. However, increases in biomass will (on average) lead to increases in mean recruitment which will be felt first by the surface fishery.

Figures 6 and 7 provide more detailed performance statistics, Figure 6 relating to SSB and Figure 7 to catch. Figure 6 again indicates that **S3** and **S4** entail a higher risk in terms of SSB levels, with lower probabilities of recovery for both the short-term statistics ($p(B_{2025}>0.1\ B_0)$ and $p(B_{2025}>2\ B_{2009})$) and lower median and lower bound values of relative SSB when looking at the minimum SSB value, B_{2022} and B_{2032} (relative to 2009 levels). However, the 2032 performance is very close for tuning option 2, given that the tuning criterion forces the distributions very close together at this time because by 2035 they have to all possess the same probability of being above the interim rebuilding target. The TAC-to-SSB inconsistency statistic gives an indication of how often the direction of a change in TAC (an increase or decrease) does not match the underlying SSB signal. Though the medians for all alternatives are much the same, the longer lower tails for **S1** and **S5** indicate that these alternatives are more likely to get better consistency between TAC and SSB, although more so for tuning option 2 than for tuning option 5. This follows since these alternatives are the most reactive of the 5 MPs, with weak (**S1**) or no (**S5**) TAC memory effect, and so are able to respond to changes in SSB more quickly.

Figure 7 provides statistics on the performance of the MP alternatives relating to future catches. The short-term average catch statistics (2013-2018) indicate that the less reactive (stronger TAC memory) alternatives **S3** and **S4** have higher median average TACs, though all alternatives are below the current TAC. In the longer-term (2019-2032), and when looking over the entire time period (2009-2039), **S3** and **S4** have lower median average TACs, though this difference is a lot less pronounced when viewed over the full time period. The medians and lower bounds of the average catch for the two longer time-periods (2019-2032 and 2009-2039) suggest that **S3** and **S4** are more likely to have average TACs that are lower than the other alternatives, although more so for tuning option 2 than 5 (see Figure 7).

The decreasing inter-annual variation (AAV) statistics clearly reflect the decreasing level of reactivity between **S1**, **S2** and **S3** (**S3** has the lowest level of variation in TACs and is the least reactive of these 3 alternatives). This statistic is a good indicator of the stability of the TAC trajectories and of key concern to industry, where too strong a change in TAC (up or down) can result in capacity-related and other economic problems over a number of years. This pattern is also evident in the maximum TAC decrease statistic. Further, variation in the AAV and maximum TAC decrease statistics declines with increasing memory effect. With less reactivity, this is to be expected. Scenario **S5,** with no TAC memory effect, is comparable to **S1** (which has a low memory effect) in AAV levels and maximum TAC decrease. However, the statistics for **S4** are worth noting. Although in 2012-2015 **S4** has a passive phase (i.e. strong memory effect) this does not seem to have as much impact as the reactive phase in the short-term (2016-2018) as indicated by **S4** having similar median levels of AAV for the short-term (2013-2018) statistic to the more reactive **S5** and **S1**. Moreover, **S4** has the highest median TAC decrease of all alternatives, with the median coinciding with the 90th percentile. For the short-term AAV statistic this is also the case. It is possible that high TAC cuts are made by **S4** at the beginning of the reactive phase (see Figure 2) which could explain the short-term AAV result. The long-term AAV statistic shows the overall effect of **S4**, which includes a combination of active and passive phases, and has the median AAV levels being more comparable to **S2** and so less than **S1** and **S5**.

The median values for the implied effort percentage change indicate that **S3** and **S4** limit the increases in exploitation rate in the later years more so than the other alternatives as expected, although more so for tuning option 2 than 5 (see Figure 7). These alternatives are also more likely to have larger negative values. Given the increasing state of the stock abundance and catches in this period, even at the lower 10th percentile (see Figures 1 and 2), the strong TAC memory effect and pre-programmed HCR asymmetry more frequently results in trajectories where exploitation rate is decreasing over this period. It should be noted that the "saw-tooth" pattern in the implied effort exploitation rate proxy (see Figure 3) is caused by two things: (i) the 3 yearly changes in TAC, and (ii) around 2014-2017 the exploitable biomass begins to increase so that when, on average, the TAC begins to increase we see a sudden increase in the exploitation rate followed by a reduction for the following two years. This is because the same TAC is being taken from an increasing exploitable biomass for two further years, which equates to a decreasing exploitation rate.

Performance on the robustness trials

To compare and contrast the performance of the 5 MPs, we present detailed performance summaries only for tuning option 5 (70% chance that SSB will be above $0.2B_0$ by 2040), given this was recommended as the default tuning level in the OMMP report. The key robustness trials explored are detailed in Table 2 and were grouped into five general types:

1. CPUE: includes those trials which focus on the interpretation of CPUE in the period following the change in fishing behaviour of the longline fleet in 2006: **upq**, **updownq**, **downq**, **downupq**.

2. Pessimistic: includes the **omega75**, **STwin**, and **lowR** trials.
3. Optimistic: includes the **troll** and **Laslett** trials.
4. Unreported catch: includes the **c0s1l1**, **c1s1l2**, **c2s1l1**, and **c3s1l1** trials.
5. Structural: includes the **mixtag** and **regime** trials.

The first four groupings are self-explanatory but it is perhaps worth explaining the motivation of the fifth topic; structural. The operating model (OM) in its base form has no spatial structure and assumes the key stock-recruit parameters, virgin biomass and steepness, to be constant across time. The **mixtag** trial may refer to the mixing of the tagged SBT into the general population but in reality it is a trial that explores the possibility that not all of the SBT juvenile population passes through the Great Australian Bight (GAB). The regime trial tackles another point of frequent discussion; how reliable and applicable is an estimate of B_0 based on data from over 50 years ago to the present given the changes observed in growth and so on. While the first of these issues, tag mixing, has been explored in terms of the implications for stock status, the second has not, and looking at the effects in terms of MP performance can be more informative than simply addressing the issue from an OM conditioning perspective only. The same could be argued for the unreported catch trials as well.

As for the performance of the alternative MPs on the reference grid, the graphical depictions of performance are detailed and cover multiple parameters and time-frames. Using the six summary statistics used to construct the reference grid summary performance table (see Table 1 and previous section), Table 3 details the performance of the 5 MP alternatives averaged across all the relevant robustness trials.

Figure 8 details the key SSB and catch performance statistics of each of the 5 MPs for the CPUE robustness grouping. As with the first suite of MPs evaluated in *CCSBT-OMMP/1006/05*, the **upq** robustness trial results in the worst SSB performance with the conceptually similar **updownq** also resulting in poorer SSB performance. While the **updownq** trial is conceptually "worse" (with a 50% increase in catchability rather than 30% for **upq** and a longer "ramp" down time) than the **upq** scenario, the reason the **upq** trial is more problematic is because it has an effect on the conditioning of the OM, while the **updownq** trial involves the reference grid with the effect only having an impact on the projections. The **upq** scenario forces the OM to interpret the exploitable biomass in 2006 and 2007 to be lower than in the reference case – it has to do this to maintain the same CPUE with the up-shift in q, given the $CPUE = q \times B$ relationship. This decreases the stock abundance just prior to the projection period resulting in poorer performance of the MPs. For the **updownq** case something different occurs. Unknown to the MPs, the elevated trend in the CPUE data between 2009 and 2013 is not related to abundance increases but an increased catchability ($CPUE_{09\text{-}13} = (1.5 \times q) \times B_{09\text{-}13}$), but the MPs interpret it as increased abundance and set higher catches (see Figure 8). This results in noticeably poorer short-term SSB rebuilding statistics, and a generally lower level of rebuilding performance in the medium to longer term. Performance on the **downq** and **downupq** trials is almost the opposite of the **upq** trial, but for similar reasons. In the conditioning phase, the reduced catchability results in higher estimates of the exploitable biomass in 2006 and 2007, and so the stock is projected into the future at a higher level than the reference case, resulting in better short-term SSB rebuilding statistics and lower exploitation rates (as evidenced by the implied effort levels – see Figure 8).

With regards to MP performance on the pessimistic robustness trials, the worst performance is observed for the **omega75** trial – all the MPs struggle to prevent further SSB declines and have poor short-term rebuilding statistics (see Figure 9). The reason is two-fold: (i) the starting state of the stock for this grid is very low, as the historical decline in CPUE reflects an even stronger decrease in the actual exploitable biomass, and (ii) in the future in particular further declines are not well detected by the MPs (which assume a linear production function in terms of biomass) and overly-optimistic TACs are set as a result. The **STwin** trial is marginally less problematic for the MPs than **omega75** – the starting state of the stock is much lower than in the reference case but the MPs at least have the correct interpretation of the CPUE and future declines are lesser in magnitude, with the MPs having better short-term rebuilding statistics. For the **lowR** trial, even with 4 years of future recruitment 50% lower than the stock-recruitment relationship would predict, there was much less of an impact than for the **omega75** and **STwin** trials. The median levels of minimum future SSB versus that in 2009 (Figure 9) were about the same as the reference case, but with lower 10th percentiles. Short-term rebuilding statistics were slightly worse, as were medium-term rebuilding levels, but by 2032 the SSB levels were about the same as in the reference case.

For the optimistic trials, the performance of the MPs for the **troll** robustness trial was much the same as for the previous round of MPs. There is a low probability of future SSB decline, with strong medium to longer term rebuilding and all the short-term rebuilding monitoring statistics are at or close to 1 (Figure 10). In terms of catches, average catches are above current levels at all future timeframes, AAV statistics are consistently low (never > 10%), with lower average implied effort but with a stronger propensity to increase exploitation rates than the reference case in the later periods (2025-2039) (Figure 10). For the **Laslett** trial there is also little chance of further SSB declines, with the short-term rebuilding statistics above the 0.7 (> $0.1xB_0$) and 0.6 (>$2xB_{2009}$) probability levels, respectively, with similar levels of medium to long-term rebuilding as the reference case. In terms of catch performance, medium to long-term catches are significantly above recent levels though lower than in the **troll** trial, AAV statistics are good being almost always below 10% and the implied effort is at similar levels as the reference case but with a slightly stronger propensity to increase the implied effort in the later years, even for the more constrained and less reactive MPs (Figure 10).

With regards to the unreported catch trials, the SSB performance is very similar, with marginally better rebuilding statistics relative to the reference case for the **c1s1l2** trial (market review option 2 – more unreported catch and larger stock size), and marginally worse rebuilding relative to the reference case for the **c2s1l1** and **c3s1l1** trials – the ones which assume 50% and 75% of the unreported catch contribute to CPUE, respectively. This might seem a little counterintuitive but the reason is the relative flexibility of the MP versus the OM. As has been seen in previous meetings, the OM has problems in fitting to CPUE series with increasing levels (i.e. above 25%) of unreported catch contributing to CPUE, as the higher recruitments required to fit such data are not supported by the other available observations. The BREM estimation model does not contain these other data apart from the scientific aerial survey, but has more freedom than the OM does to fit these higher CPUE levels. As a result, the MP will have a more optimistic view of the stock than the OM and sets slightly larger TAC levels than in the reference case (Figure 11). In general, there is little significant difference in terms of both SSB and catch performance across the set of unreported catch alternatives relative to the reference case.

For the trials which concentrate on structural uncertainties within the population model (**mixtag** and **regime**) there were fewer significant differences when comparing with the reference case. For the tag-mixing trial the starting stock size is larger, given the tag-based exploitation rates are assumed to be lower when viewed relative to the "whole" stock. This gives marginally better minimum future SSB and rebuilding performance. For the B_0 regime-shift there are even less observable differences relative to the reference case – especially when we factor in that the two alternatives have different steepness grids (three for the regime-shift versus five for the current reference grid). We do see better short-term performance on the regime-shift trial with regards to the interim SSB rebuilding statistics – see Figure 12 – relative to both the reference case and the tag-mixing trial. The reason for this is that the second regime (1978-present) estimate of B_0 is actually *smaller* than the single estimate, albeit with a higher CV, in the reference grid – median (and 90% CI) is 810,461 (533,288-1,227,250) tonnes versus 980,691 (755,144-1,217,360) tonnes for the reference case. The absolute estimates of recent SSB are much closer and even after factoring in the steepness differential (3 versus 5 grid options) this difference makes $0.1xB_0$ effectively "closer" to current biomass for the regime-shift case. The difference between the regime-shift and tag-mixing trials is much closer for the $2xB_{2009}$ statistic as this removes this B_0 effect. There are no obvious differences in catch performance, relative to the reference case, for either of these two structural robustness trials (Figure 12).

Discussion

In the previous round of MP testing, there was a clear trade-off between more reactive MPs (i.e. **S1**, **S2**, **S5**) making early cuts with better short-term rebuilding statistics and average catch levels, and less reactive MPs (**S3**, **S4**) making gentler cuts early on with worse short-term rebuilding statistics and lower average catches. A concern raised at the OMMP (2010) meeting was that, for the more reactive MPs, this increased catch performance came from more rapidly increasing catch levels in the later (2025-2040) period, possibly increasing the exploitation rate in this period unduly. To try and deal with these three points, a set of MP alternatives based around a single flexible harvest control rule were developed, simulated and evaluated (i.e. **S1-S4)**.

What became clear was that the original two-way trade off (average catch versus short-term SSB rebuilding risk) became more of a three-way trade-off, with the added dimension being that one had to significantly increase the smoothness of the TAC transition from year to year to avoid the observed strong increases in TAC in the latter part of the testing timeframe (2025-2040). An implied effort statistic (as a proxy for exploitation rate) and the percentage change in this statistic from 2025 to 2040 were found to be useful in exploring this extra trade-off dimension. Certainly for tuning option 5 (70% chance that SSB will be above $0.2B_0$ by 2040), *all* MP alternatives acted to increase the exploitation rate in the latter period in response to the strong rebuilding signals in the data. Those which are more reactive did this more obviously, given better initial rebuilding and also the greater flexibility to act on more positive signals. One key point to make is that, regardless of how much or how little this effect was observed, the final levels of exploitation rate were *significantly* lower than those observed at the start in 2013. SSB levels (both medians and lower 10^{th} percentiles) were still increasing by 2039, suggesting there is little risk of any of the MP alternatives resulting in unsustainable levels of catch even in the 2035-2040 timeframe. An important corollary of this observed SSB growth even in 2040 is that we are also unlikely to reach the interim rebuilding target with the SSB "flattening out" at 20% of B_0. Beyond this we cannot say given the scope of the MP evaluation. It is clear, however, that having more reactive MPs that better mitigate short-term SSB depletion risk do not suggest an additional risk in the long-term of potentially unsustainable catch increases post-recovery towards the interim SSB rebuilding level.

In terms of the performance of the MPs on the key robustness trials, as with the previous round of trials, more reactive MPs are more able to prevent the chances of further SSB declines on the more pessimistic grid options. In terms of the more optimistic grid options, there were less obvious differences between the different alternatives, with similar SSB and catch performance when viewed over the whole testing timeframe. In terms of the CPUE grouping of robustness trials, those relating to increases in catchability in the period after the 2006 change in operation caused the most problems, with again the more reactive MPs better equipped to handle these more pessimistic robustness trials. In terms of performance on the unreported catch alternatives, there was little if any observable difference in terms of catch, with a marginal decrease in SSB performance for the unreported catch trials which attribute more of the unreported catch to CPUE. This is because the MPs are able to interpret this signal in the historical data more positively than the OM can, given it is not supported in the other available data not used by the MPs themselves; as a result they set higher catches than in the reference case, yielding marginally worse SSB performance. For the trials focussing on structural uncertainty within the actual population model in the OM (tag mixing and B_0 regime-shift) there was little difference observed in either catch or SSB performance. The tag mixing was marginally more positive in terms of short-term rebuilding given the higher estimates of recent stock size; the regime-shift trial had better statistics relating to rebuilding relative to 10% of B_0 because the estimates of B_0 post-regime shift were in fact lower than in the reference case, with similar recent (2009) biomass levels, thus making that interim level in fact closer than before.

What was observed was that the influence of the TAC memory effect was not apparently linear in nature; we did not see a steady transition in terms of rebuilding behaviour, AAV and average catch levels as we increased the memory effect from 25% (**S1**) to 50% (**S2**) to 75% (**S3**). The observed "grouping" of MP alternatives (**S1**, **S2, S5** versus **S3, S4**) suggested more of a threshold dynamic, whereby at some point between the 50% and 75% TAC memory weighting levels the "switch" from reactive to smooth occurred. This effect could, to some degree, be confounded with the increasing asymmetry defined for high TAC memory MP alternatives but it is not an obvious linkage – the asymmetry permits stronger reactions below target levels (to guard against the memory-induced inaction relative to negative signals) and stronger rebuilding and associated potential catch increases, but then penalises the strength of the TAC increases permitted later on. To have the same asymmetry for all alternatives would most likely negatively impact the rebuilding performance of the high TAC memory alternatives. This is interesting because it does not present us with a simple trade-off decision. There is apparently no intermediate level we can obtain – we must choose which behavioural trait we want and select accordingly. For example, to avoid stronger initial cuts and future potential increases in TACs, but potentially increase the risk to the stock in the short-term, a less reactive MP (i.e. **S3** or **S4**) may be selected. In contrast, given the current estimated low stock biomass, a more reactive MP such as **S1** or **S2**, with the better short-term SSB risk mitigation performance, may be the preferred option.

Acknowledgements

Many thanks to Jason Hartog and Ann Preece for their help in both undertaking the work and writing this paper.

References

Anonymous. 2010. Report of the Third Operating Model and Management Procedure Technical Meeting, Seattle, USA.

Hillary, R., Giannini, F., Basson, M., Eveson, P., Davies, C. and Barnes, B. 2010. Exploration and initial evaluation of candidate management procedures for southern bluefin tuna. CCSBT-OMMP/1006/5.

Hillary, R., Eveson, P., Basson, M., and Davies, C. 2010. Updated technical specifications and performance analyses for the brem suite of candidate management procedures. CCSBT-ESC/1009/10.

Tables

Table 1: Performance summary for the 5 MP alternatives on the reference grid for tuning option 5 (70% chance that SSB will be above 0.2B_0 by 2040), TAC option d (3000 t maximum change), and with a 1 year lag.

Statistic	**S1**	**S2**	**S3**	**S4**	**S5**
$p(B_{2025}>0.1B_0)$	0.74	0.71	0.62	0.65	0.73
$p(B_{2025}>2xB_{2009})$	0.69	0.66	0.58	0.59	0.68
$Mean(TAC_{2013\text{-}2025})$	7,402	7,609	8,235	8,077	7,548
$CV(TAC_{2013\text{-}2025})$	0.32	0.28	0.18	0.31	0.32
$Mean(TAC_{2013\text{-}2039})$	12,612	12,538	12,084	12,230	12,611
$CV(TAC_{2013\text{-}2039})$	0.49	0.47	0.39	0.43	0.47
$Mean(CPUE_{2013\text{-}2025})$	0.75	0.74	0.71	0.71	0.75
$CV(CPUE_{2013\text{-}2025})$	0.44	0.44	0.43	0.43	0.44

Table 2: Key robustness trials included in the work presented in this paper.

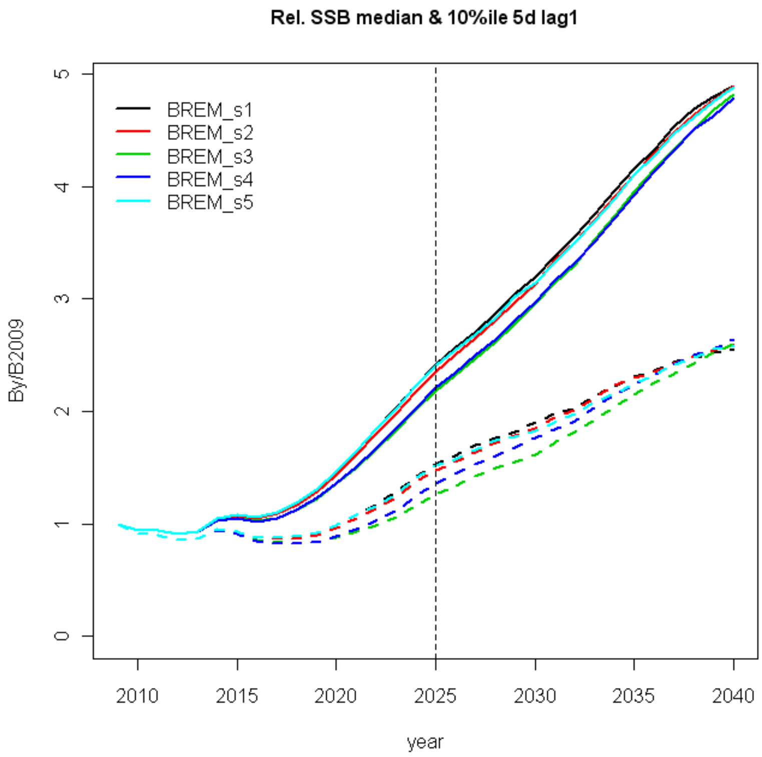
Grouping	Robustness trial	Explanation
CPUE	upq updownq downq downupq	*q* 30% ↑ 2006 and 2007 *q* 30% ↑ 2009-2013 *q* 20% ↓ 2006 and 2007 *q* 20% ↓ 2007-2011
Over-catch	c0s1l1 c1s1l2 c2s1l1 c3s1l1	No over-catch to CPUE Case 2 over-catch from JMR 50% over-catch to CPUE 75% over-catch to CPUE
Pessimistic	omega75 STwin lowR	Hyper-stable CPUE to biomass Lowest of suite of CPUE series Recruitment ↓ 50% 2009-2012
Optimistic	troll Laslett	Inclusion of trolling survey Highest of suite of CPUE series
Structural	mixtag regime	GAB tags fraction of total pop^n New B_0 (1978-2008) estimate

Table 3: Performance summary for the 5 MP alternatives averaged across all the robustness trials detailed in the paper (see results section) for tuning option 5 (70% chance that SSB will be above 0.2B_0 by 2040), TAC option d (3000 t maximum change), and with a 1 year lag.

Statistic	**S1**	**S2**	**S3**	**S4**	**S5**
$p(B_{2025}>0.1B_0)$	0.72	0.69	0.62	0.65	0.72
$p(B_{2025}>2xB_{2009})$	0.71	0.68	0.59	0.63	0.71
$Mean(TAC_{2013\text{-}2025})$	7,845	8,032	8,639	8,406	7,939
$CV(TAC_{2013\text{-}2025})$	0.32	0.28	0.19	0.33	0.31
$Mean(TAC_{2013\text{-}2039})$	12,953	12,841	12,471	12,525	12,880
$CV(TAC_{2013\text{-}2039})$	0.48	0.46	0.39	0.44	0.46
$Mean(CPUE_{2013\text{-}2025})$	0.82	0.81	0.78	0.79	0.82
$CV(CPUE_{2013\text{-}2025})$	0.42	0.42	0.41	0.42	0.42

Figures

Figure 1: Median (full line) and lower 10th percentile (dotted line) future SSB for tuning options 5d (top) and 2d (bottom) for each of the 5 MP alternatives with the vertical dotted line being the associated interim check-point year.



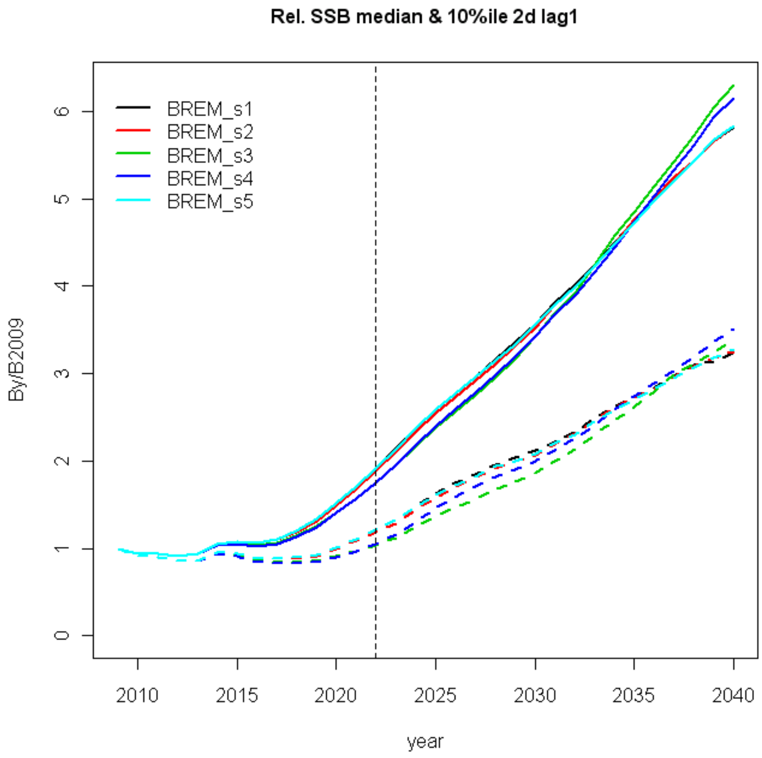
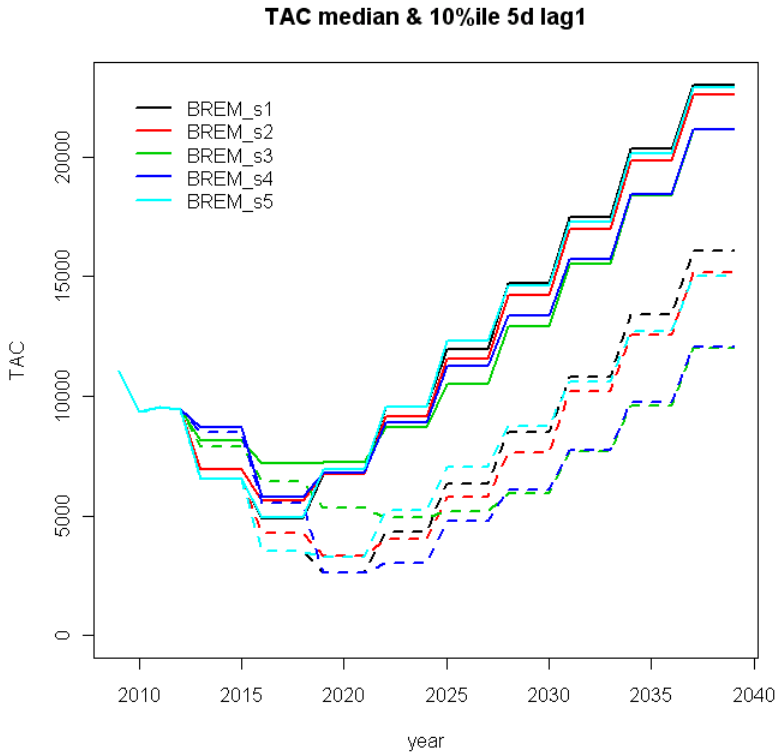


Figure 2: Median (full line) and lower 10th percentile (dotted line) future TAC for tuning options 5d (top) and 2d (bottom) for each of the 5 MP alternatives.



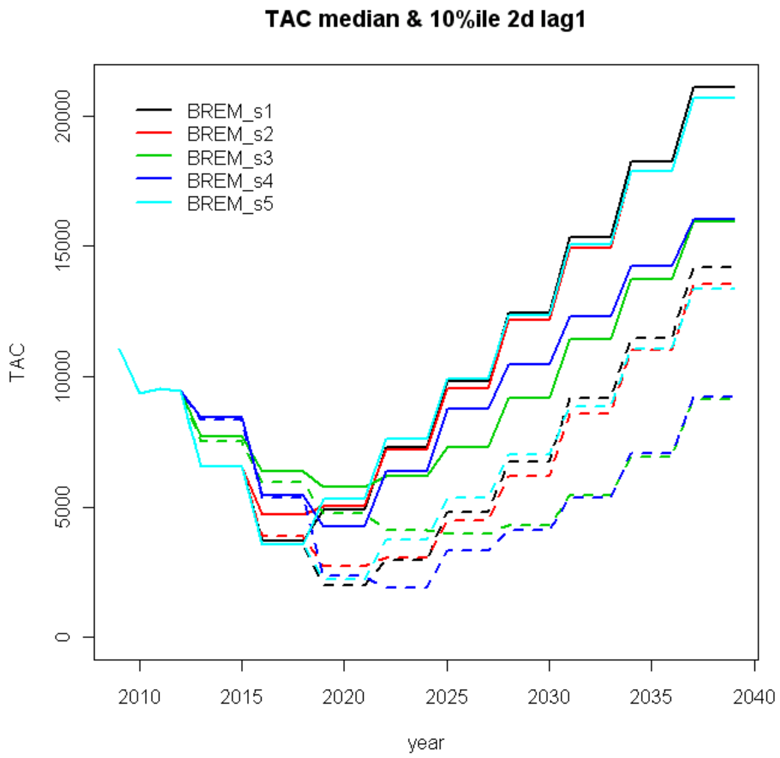
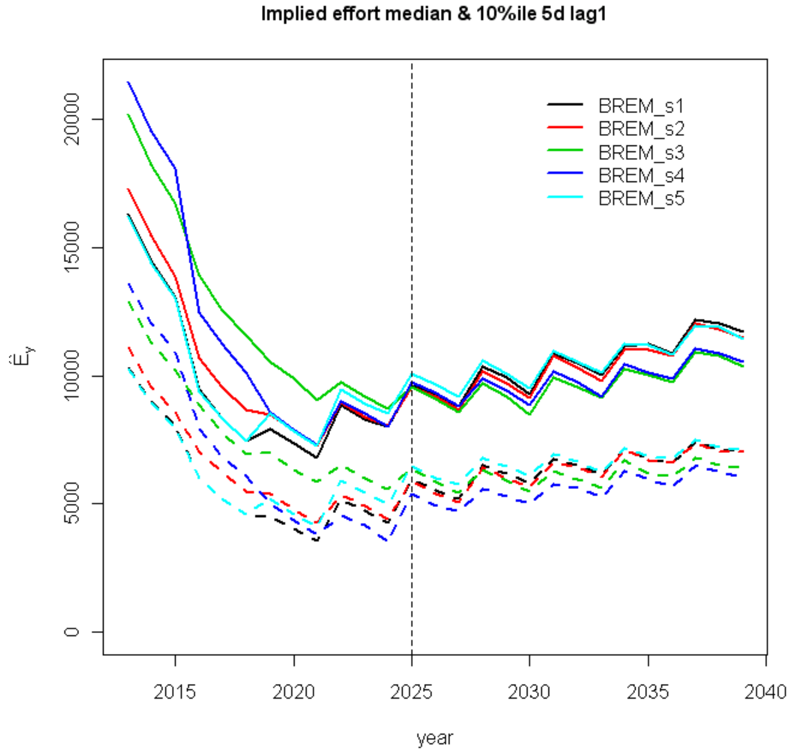


Figure 3: Median (full line) and lower 10th percentile (dotted line) future implied effort for tuning options 5d (top) and 2d (bottom) for each of the 5 MP alternatives with the vertical dotted line being the associated interim check-point year.



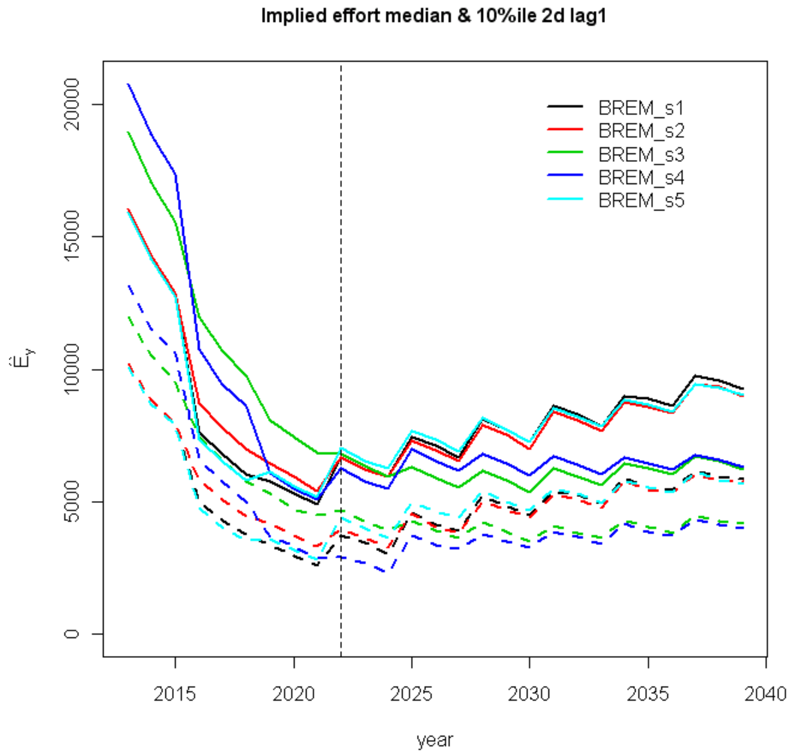
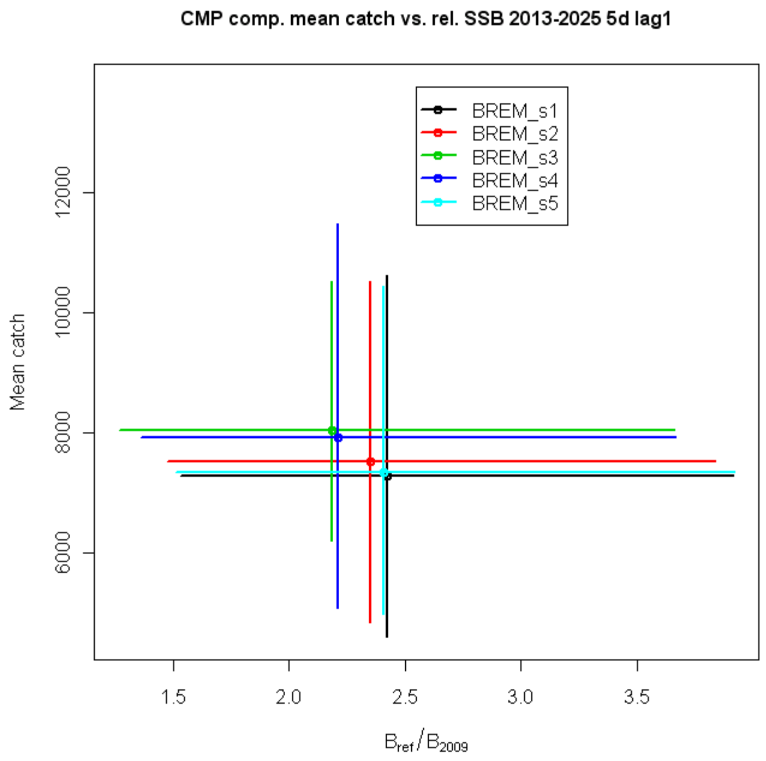


Figure 4: Interim-time-frame year-averaged catch versus average SSB (relative to 2009) trade-off for tuning options 5d (top) and 2d (bottom) for each of the 5 MP alternatives with the vertical dotted line being the associated interim check-point year. Circles are medians with whiskers representing the 80% confidence interval.



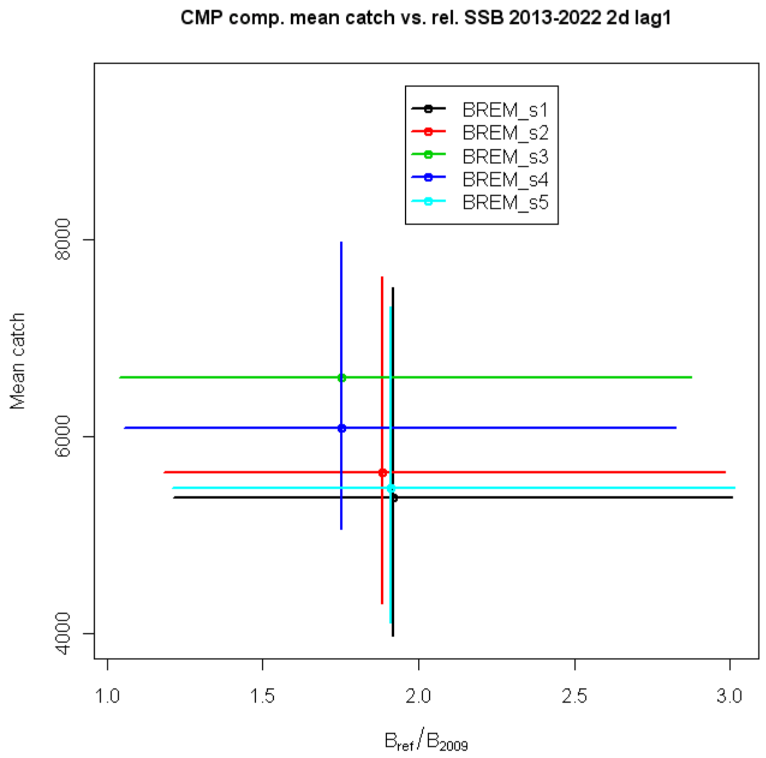
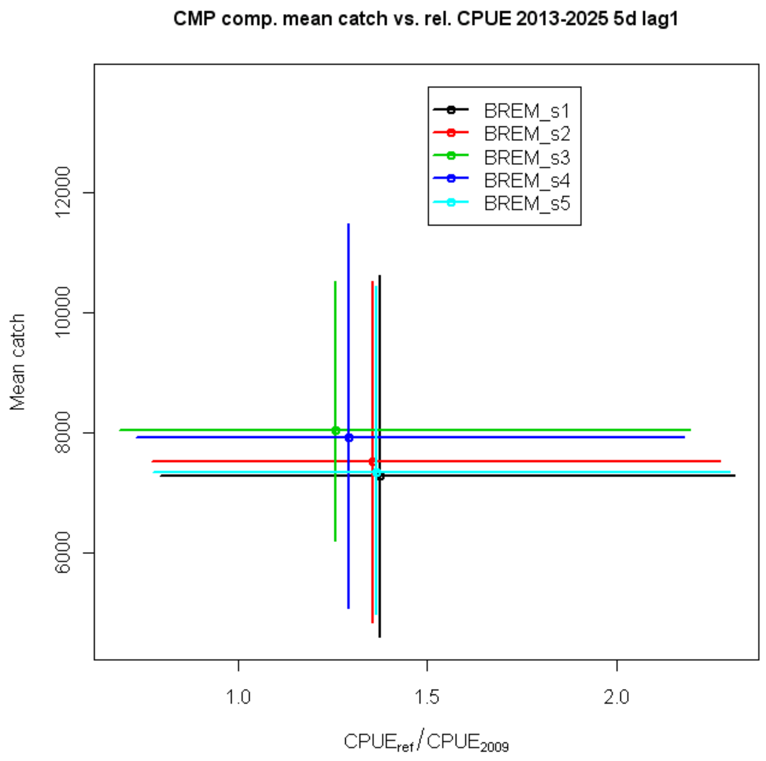


Figure 5: Interim-time-frame year-averaged catch versus average CPUE (relative to 2009) trade-off for tuning options 5d (top) and 2d (bottom) for each of the 5 MP alternatives with the vertical dotted line being the associated interim check-point year. Circles are medians with whiskers representing the 80% confidence interval.



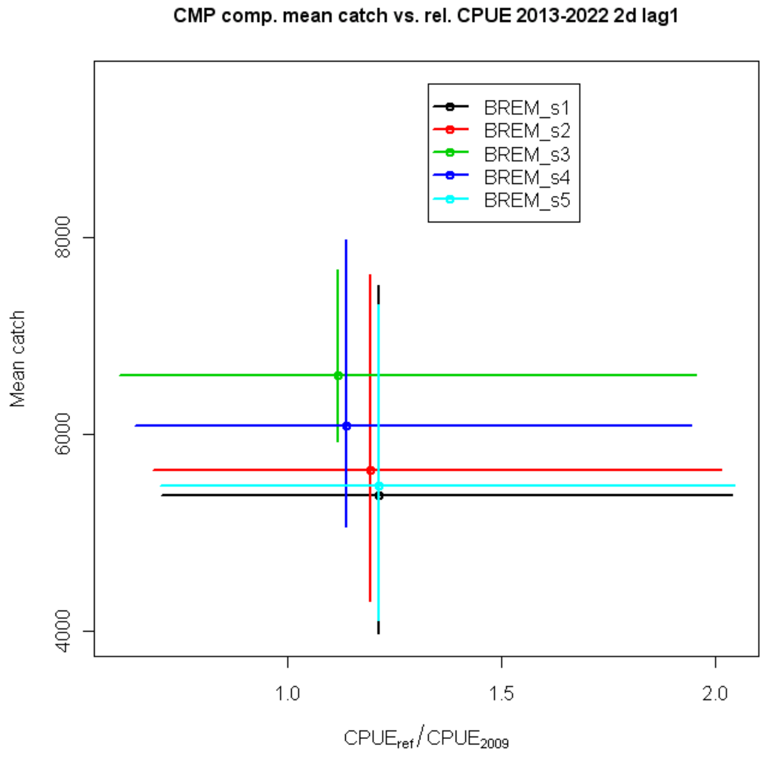
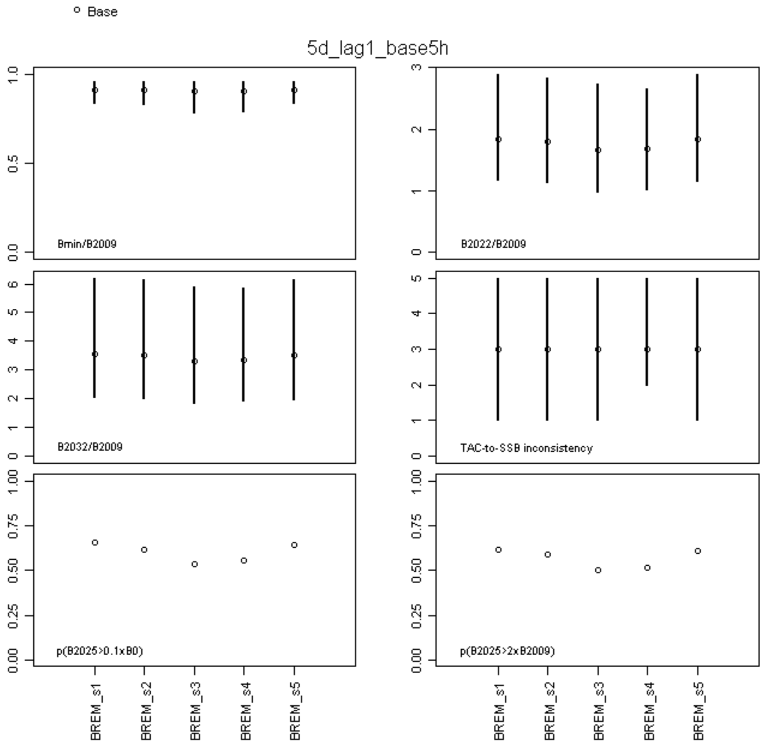


Figure 6: SSB performance summary for the 5 MP alternatives on the reference grid for tuning options 5d (top) and 2d (bottom) respectively.



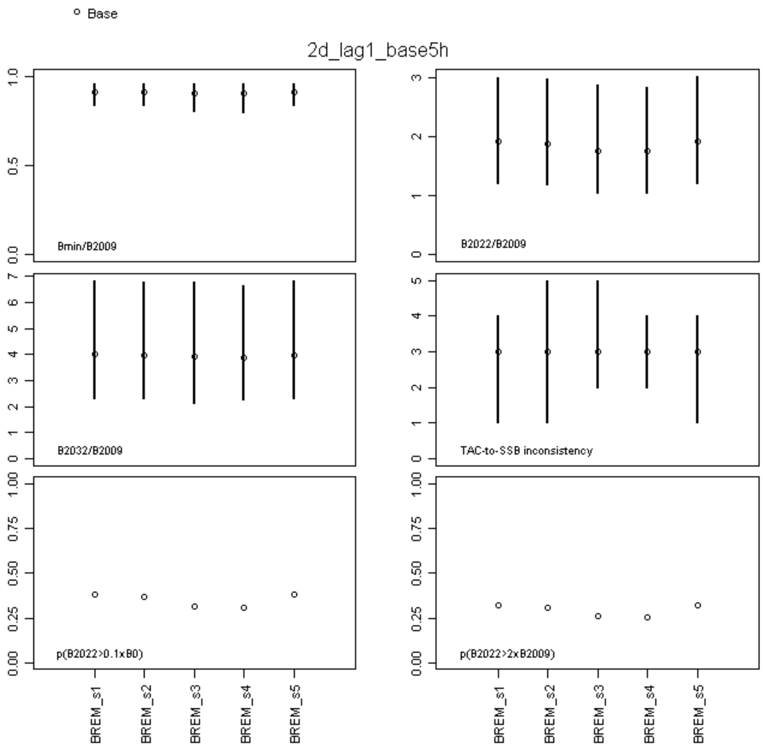
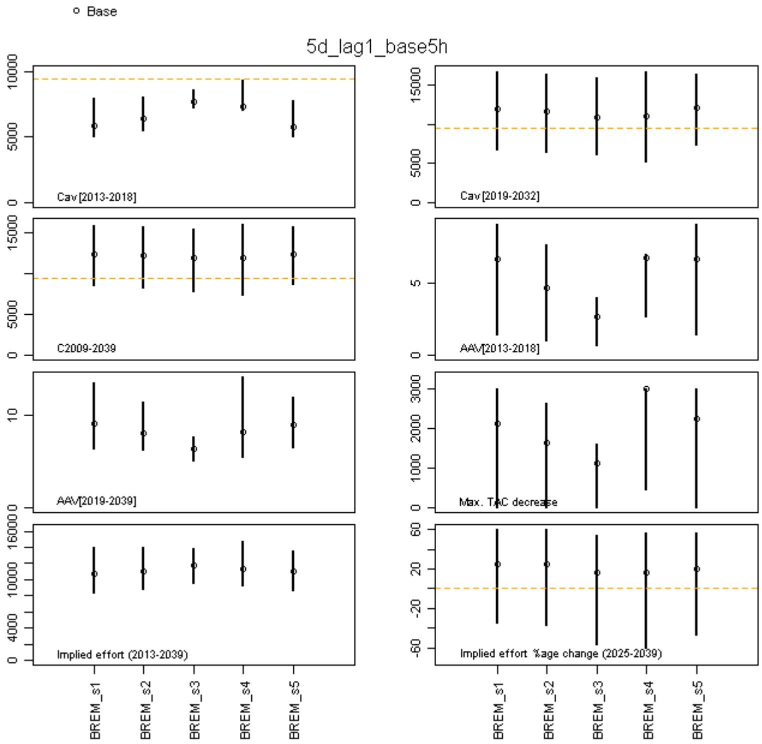


Figure 7: Catch performance summary for the 5 MP alternatives on the reference grid for tuning options 5d (top) and 2d (bottom) respectively. The orange line in the average catch frames represents the current 9,449t TAC.



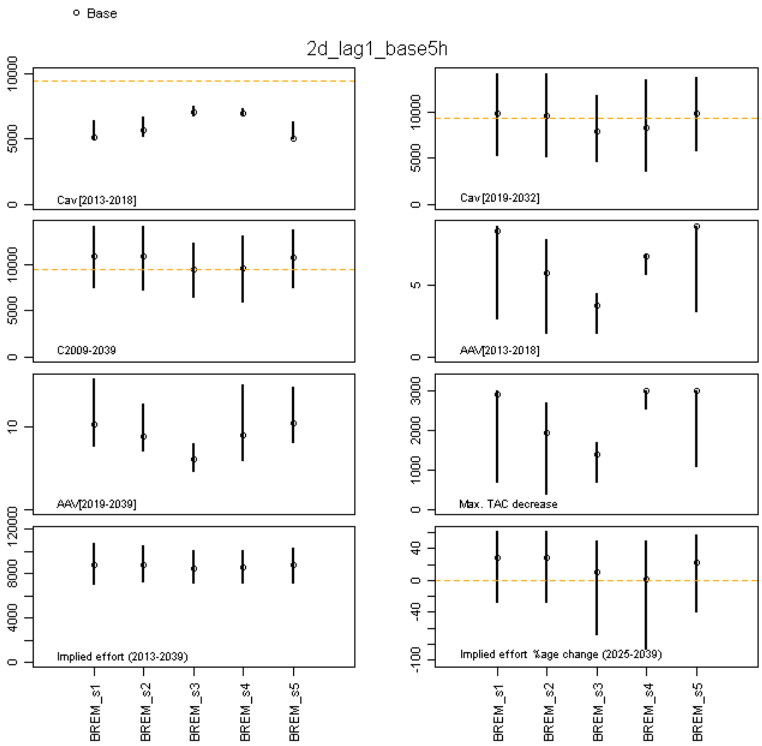


Figure 8: SSB (top) and catch (bottom) performance of the 5 MP variants on the robustness grids focussing on CPUE. The orange line in the average catch frames represents the current 9,449 t TAC.

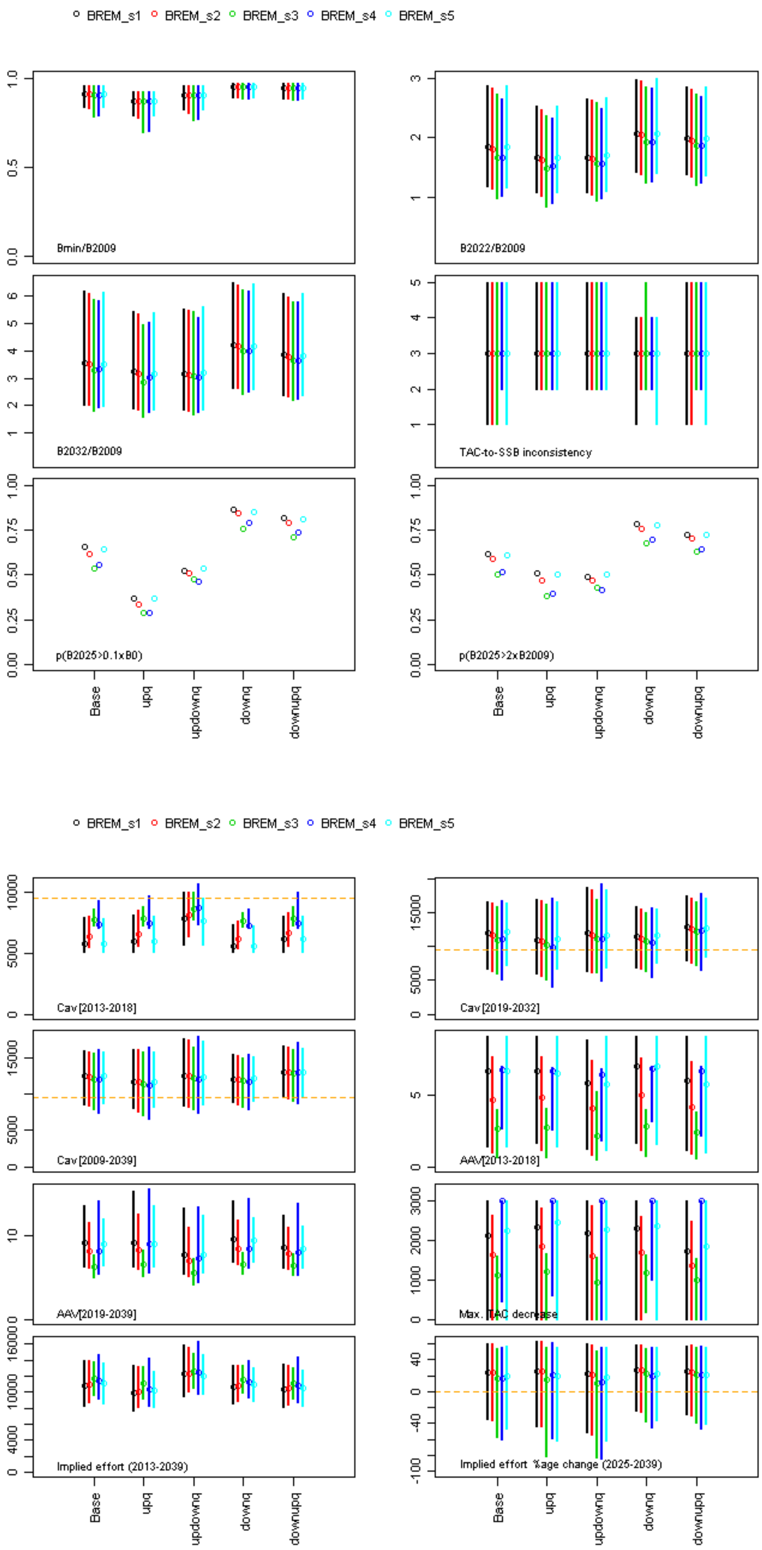
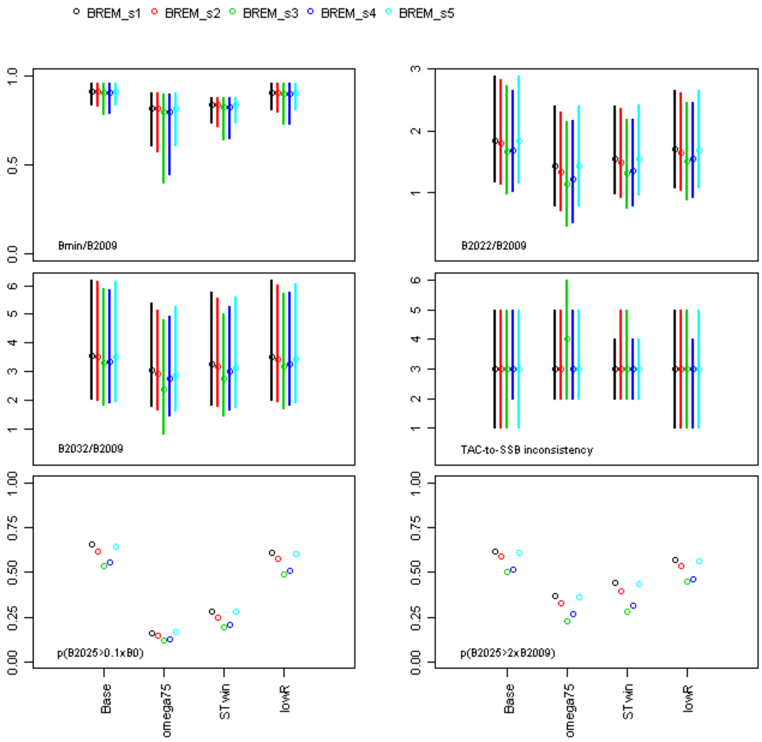


Figure 9: SSB (top) and catch (bottom) performance of the 5 MP variants on the more pessimistic robustness grids. The orange line in the average catch frames represents the current 9,449 t TAC.



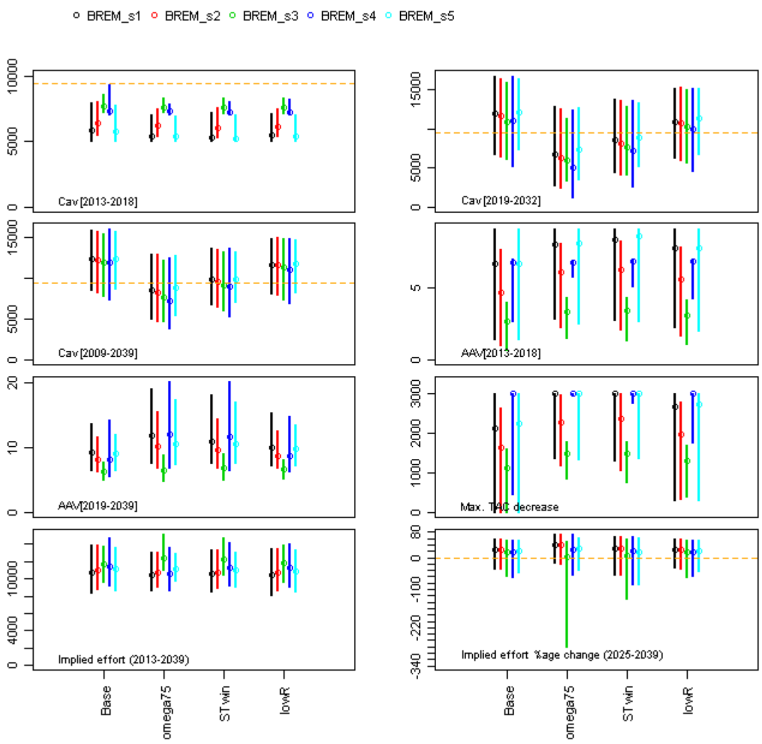
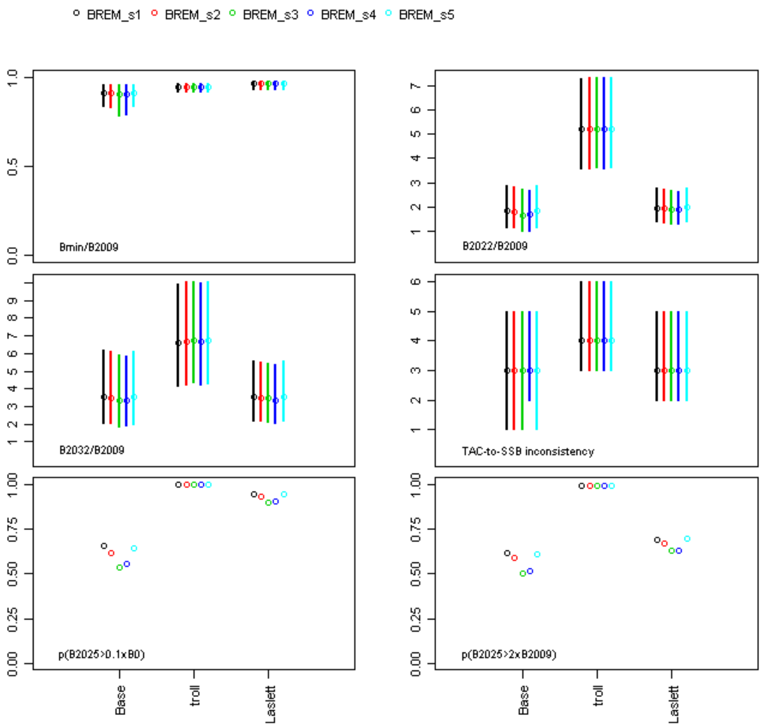


Figure 10: SSB (top) and catch (bottom) performance of the 5 MP variants on the more optimistic robustness grids. The orange line in the average catch frames represents the current 9,449 t TAC.



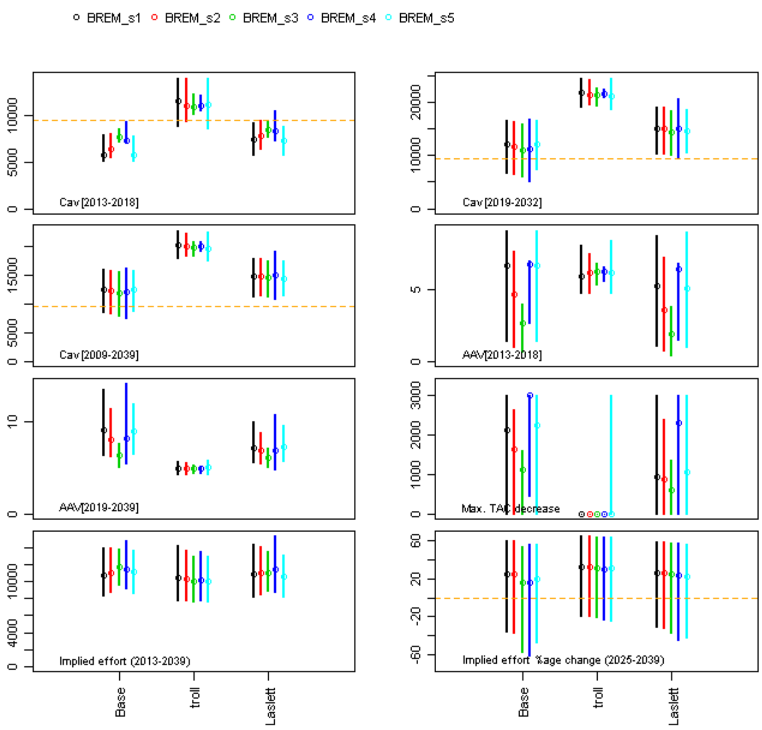
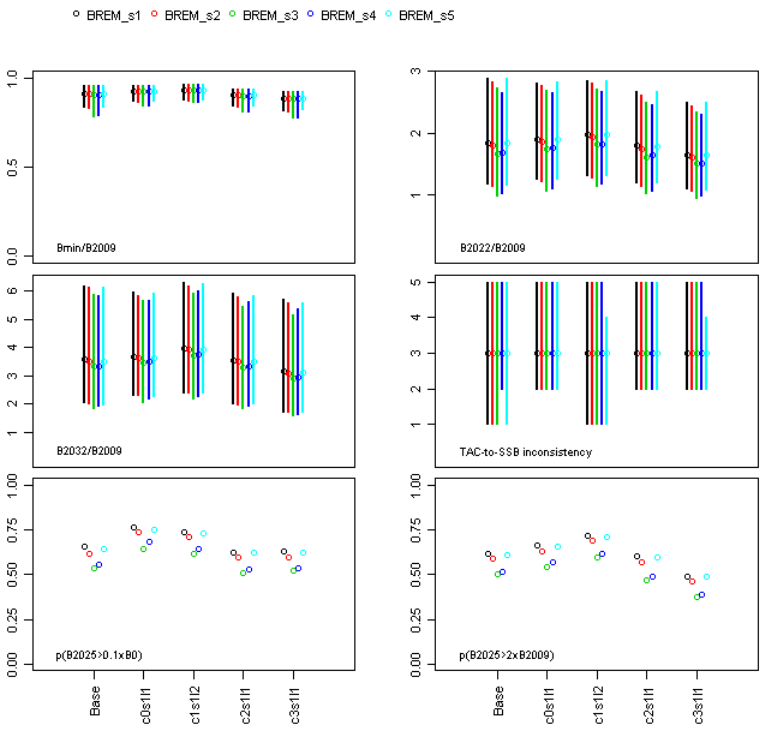


Figure 11: SSB (top) and catch (bottom) performance of the 5 MP variants on the robustness grids focussing on different over-catch alternatives. The orange line in the average catch frames represents the current 9,449 t TAC.



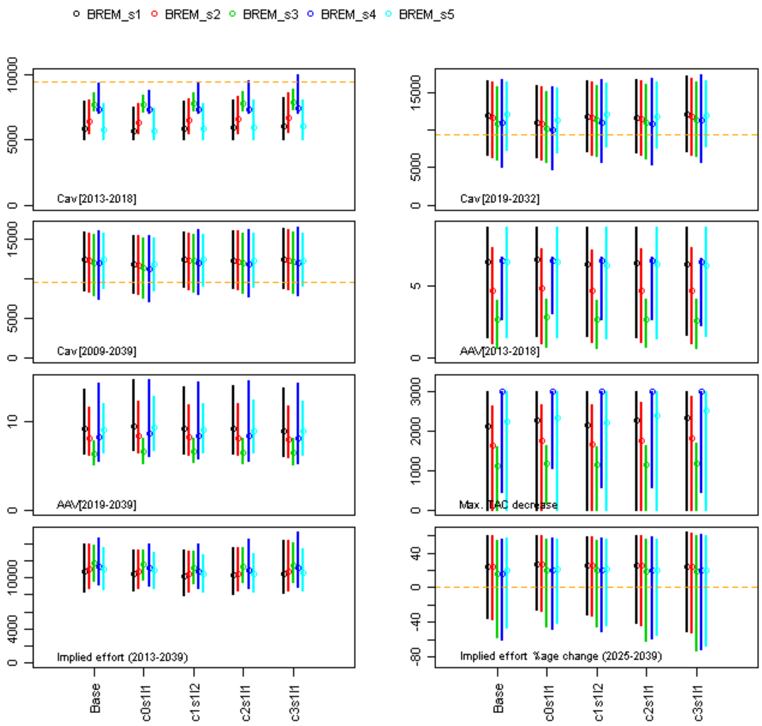


Figure 12: SSB (top) and catch (bottom) performance of the 5 MP variants on the robustness grids focussing structural issues within the OM – the tag mixing and B_0 regime-shift options. The orange line in the average catch frames represents the current 9,449 t TAC.