Australian Government
Bureau of Rural Sciences

CSIRO

Exploration and initial evaluation of candidate management procedures for southern bluefin tuna

Richard Hillary
Fiona Giannini
Paige Eveson
Marinelle Basson
Campbell Davies
Belinda Barnes

Prepared for the CCSBT 3rd Operating Model and Management Procedure Technical Meeting, Seattle 21st to the 25th of June, 2010.

Table of Contents

Abstract 1
Introduction 2
Results 4
CMPs tuned to the reference grid 4
Robustness testing of the CMPs 6
CCMP 8
ASMP 8
BREM_1 8
BREM_2 9
Effect of the upq robustness grid across CMPs 9
Comparing more/less reactive CMPs in more extreme cases 10
Concluding remarks 10
Acknowledgements 12
References 12
Figures and Tables 13

Abstract

This paper is a summary of the initial evaluation of a set of candidate management procedures (CMPs) for southern bluefin tuna. A selection of the CMPs defined in document CCSBT-OMMP/1006/4 are tuned to meet the levels of performance specified by the Commission and compared against each other on the reference set and robustness grids. In addition, a simple tuned constant catch MP was included for reference purposes. General conclusions with regards to CMPs performance for the reference grid are that there is a basic trade-off of earlier reduction in catches, resulting in lower SSB risk sooner and earlier increase in catches versus smaller, or later, reduction in catches, little, or no increase in catches over the evaluation period, and higher SSB risk. General conclusions in terms of robustness of performance are that the more responsive CMPs are more likely to decrease short-term SSB risks, mitigate against serious further depletion for "pessimistic" robustness grids and give better average catch performance (without increasing SSB risk), for more "optimistic" robustness grids.

Introduction

Document CCBST-OMMP/1006/4 (Hillary et al., 2010) provides technical specifications and independent exploration of behaviour of the CMPs and performance, relative to historical data, of the model-based MPs. In this paper we deal with 3 CMPs that were considered worthwhile pursuing based on these "proof of concept" explorations:

1. **ASMP:** Aerial Survey Management Procedure – This is a "model-free" MP that uses the aerial survey data only, specifics defined in CCBST-OMMP/1006/4. In short, it takes a log-scale moving average of the aerial survey and compares this with an empirically defined "target" level (a reference mean level one would expect to see at the interim 20% of B_0 rebuilding target) of the aerial survey and adjusts the catch accordingly.
2. **BREM_1:** Biomass random effect model 1 – Model-based MP, specifics defined in CCBST-OMMP/1006/4. Using aerial survey and CPUE decomposes dynamics of adult biomass into recruitment effects and growth/decline effects. This version has a target relative biomass and adjusts a reference catch level to achieve it based on using the current-to-target relative biomass ratio and the recruitment and adult biomass growth/decline trends as well.
3. **BREM_2:** Biomass random effect model 2 – exact same model and information as **BREM_1** but adjusts the previous year's catch, rather than a reference catch level, using the information on the "distance" from the target relative biomass and the recruitment and biomass growth/decline information.

To compare the relative performance of the feed-back CMPs defined above with a constant catch approach we also tuned a constant catch MP (henceforth, **CCMP**) to all the levels and took it through full robustness testing as well. Each CMP on the reference grid *c1s1l13hsqrt* was tuned to tuning levels 1-6, defined as follows:

- **1 - reaching 20% of SSB_0 in 2035 with a probability of 0.6**
- 2 - reaching 20% of SSB_0 in 2035 with a probability of 0.7
- **3 - reaching 20% of SSB_0 in 2035 with a probability of 0.9**
- 4 - reaching 20% of SSB_0 in 2040 with a probability of 0.6
- **5 - reaching 20% of SSB_0 in 2040 with a probability of 0.7**
- 6 - reaching 20% of SSB_0 in 2040 with a probability of 0.9

In the Results section, the performance of the CMPs tuned to levels 1, 3 and 5 using the reference grid are presented.

Default operational constraints assumed were:

- MP begins in 2012 with a potential change in TAC every 3 years – option 'c': (a) start 2012 TAC every year, (b) start 2012, every 2 years, (c), start 2012 every 3 years, (d) start 2013 every three years, and (e) fixed constant catch.

- Maximum and minimum permissible changes in TAC were 3000t and 100t, respectively
- No lag was assumed between determining a TAC and its implementation

As an initial exploration of the impact of a delayed start and the lag effect, ASMP was tuned to levels 1-6 for TAC option 'd' (2013 start, every 3 years) and with a 1 year lag between TAC determination and implementation. In paper CCBST-OMMP/1006/4 there were 4 potential Pella-Tomlinson CMPs, but a decision was made *not* to take these CMPs forward for tuning and robustness testing. Guaranteeing convergence of the optimiser on the MLE proved a difficult task (as was the case with the Fox model in the previous MP work). Even with some basic semi-intelligent algorithms for setting the initial parameter estimates there are a significant number of clearly non-convergent parameter estimates that either hit the boundaries or end up in regions of parameter space with very low likelihood. These CMPs were able to tune in all cases but there are two issues which we believe justify their removal from further evaluation:

1. In the real world one simply would not proceed to set a TAC from a model that had clearly not converged on the MLE – one would attempt to diagnose the problem by looking at alternate starting values and doing analyses simply not replicable in the "blind" estimation framework we employ in the MP testing phase. Any model-based CMP with these kind of issues is borderline un-testable in the MSE framework as we are not simulating the real world application of the CMP.
2. "Tuning will take care of it" – as mentioned, the Pella-Tomlinson CMPs tuned without any problem, but even though the tuning algorithm takes care of the frequency of non-convergent solutions, this is not a solution. We explored estimating log(K) versus K in the tuning phase (even after rescaling the model to 1000s tonnes to begin with) to stabilise the appearance of false/boundary estimates and, while the frequency of such estimates decreased, this actively changed the tuning estimate. For the Pella-Tomlinson CMPs the primary part of the harvest control rule is essentially an F_{msy} strategy, but where the tuning parameter acts to re-scale the estimates of F_{msy} to achieve the relevant tuning target. The estimates of F_{msy} are proportional to r – a parameter estimated in the MP. This parameter is bounded above and the frequency with which the estimates hit the upper bound differs when estimating K or log(K) – when the MLE is found there is good consistency between the K or log(K) variants but for the log(K) option the optimiser doesn't go off track as often. Given these more frequent upper boundary estimates of r the average estimates of F_{msy} (across the grid and projection samples) are higher for the case where K is the parameter and not log(K). To achieve the same tuning goal this requires *larger* tuning parameters in the log*(K)* case, given the difference in the average F_{msy} estimates. The tuning parameter estimates lack robustness given the apparent estimation instability within the CMPs themselves.
3. We note that for the **BREM_1, 2** model-based CMPs no such convergence issues were encountered – the model is simple and quasi-linear on a log-scale – so we felt comfortable enough in taking these tuned model-based CMPs forward for full robustness testing.

Results

The results section is structured as follows: The performance of the CMPs tuned to levels 1, 3 and 5 using the reference grid is presented. These tuning levels were chosen as they do not overlap (levels 2 and 5 produce similar results across CMPs) and span the range of rebuilding constraints well, hence this was considered an efficient way to proceed; robustness testing is broken down into two general parts, the first of which looks at performance across all 20 robustness grids, and the second of which investigates more specifically each of the CMPs relative to their expected "troublesome" robustness grids and the comparative performance of all the CMPs at the extremes of the robustness scenarios.

CMPs tuned to the reference grid

Figures 1 and 2 shows the future performance of each of the CMPs relative to the 2 short-term checkpoint SSB rebuilding statistics: (i) $p(SSB_y > 0.1 * SSB_0)$, and (ii) $p(SSB_y > 2 * SSB_{2009})$. For tuning levels 1 and 5 clearly **BREM_1** and **ASMP** outperform **BREM_2** and **CCMP** with regards to both statistics. This is not the case for tuning level 3 where the performance is much closer but arguably with **CCMP** being the best. In the first instance this differential performance is a factor of the relative reactivity of the CMP: **BREM_1** and **ASMP** adapt a reference catch level based on "local" conditions in the CPUE and the aerial survey (only in the case of **ASMP**); **BREM_2** adapts the previous year's catch and obviously **CCMP** moves straight to a fixed catch level. Both **BREM_1** and **ASMP** act much quicker in cutting catches to start moving towards the tuning target, whereas **BREM_2** is much more gradual in the action it takes – this fast versus slow action time means the more reactive CMPs move to decrease the SSB risk faster than the others. The reason this does not follow with tuning level 3 is that the 90%ile part of the tuning target drives the performance of the CMPs much closer together: a dynamic CMP has some freedom in terms of the catch reduction scheme but this freedom entrains uncertainty into the population dynamics (a negative when tuning to such a high percentile) that a constant-catch policy does not, potentially offsetting the advantages of the dynamic CMPs. It should be noted that only the zero TAC option can meet the 0.6 and 0.7 but not the 0.9 (at 2022) and 0.6, 0.7 and 0.9 (at 2025) short-term checkpoint target probabilities detailed in the SFMWG 2010 report (Anon., 2010) – this is driven by the weak cohorts of the early 2000s moving through the spawning population.

Figures 3 to 6 show the SSB and catch worm plots for CMPs **CCMP**, **ASMP**, **BREM_1** and **BREM_2**, respectively, for tuning levels 1, 3 and 5 and across the full reference grid. Concentrating on the SSB worms first, for tuning levels 1 and 5 we see that **BREM_1** and **ASMP** are able to avoid the appearance of very low future SSB trajectories that appear for both **CCMP** and **BREM_2** – in fact for tuning level 5 one of the **CCMP** trajectories actually goes to zero**.** We note that these are semi-speculative conclusions based upon worm trajectories alone (they are a small set of realisations) but such behaviour is clearly observed in the various performance statistics detailed in Figures and 1 and 2 and later on in the detailed robustness trials. This is linked to the previous point made about more reactive CMPs mitigating such short-term risks better and avoiding such trajectories. Also, from inspection of the catch worm plots, while the short-term price of mitigating this risk of further depletion is lower catches in the short-term, the future increase in catches as the stock recovers is plain to see.

Figures 7 and 8 compare box-plot summaries across CMPs of the short-term (to 2022) average catch and the percentage change in TAC, when a change is allowed, respectively. Note this is **not** the same as the AAV (average annual variation in TAC from one year to the next) statistic as outputted from the projections – as it stands, that output statistic underestimates the true percentage change as it includes those years when a change in TAC was not permitted and an accumulation of zeroes decreases the true estimate. From Figure 7 we see that the more reactive CMPs (**ASMP** and **BREM_1**) give rise to lower average catch levels than the other 2, and from Figure 8 they make larger percentage changes as well. Note also that in all cases **BREM_2** outperforms **CCMP** in terms of median catch levels while maintaining better general short-term risk profiles – see Figures 1 and 2.; the point being, that also from a catch perspective the feedback CMPs perform better across the board than a simple fixed catch policy. Comparing **ASMP** to **BREM_1** in terms of catch performance measures we do see that the latter makes stronger percentage changes in the TAC, but has very similar average catch levels with noticeably less uncertain average catch levels. This added stability is a result of the model-based, integrated nature of **BREM_1** versus the model-free nature of **ASMP**: the model-based integrated CMP takes partially the same information but "smooths" it via the estimation procedure and population model, and utilises the CPUE data as well; the model-free CMP (**ASMP**) may be a moving average (which smooths variation in the aerial survey data somewhat) but is still more likely to "over-react" than **BREM_1** (this model-based CMP integrates both data sources and requires a consistent signal to react stronger and is constrained via the use of the population and estimation model as a biological "smoother"). Hence, the less variable the average catch levels and, to some degree, the better the short-term SSB risk statistics for **BREM_1**. This variability in the catch levels, as a result of the behaviour of the MP, propagates through the population dynamics, increasing uncertainty and risk.

Figures 9 and 10 summarise the key general points about the short-term trade-off between catch levels and SSB/CPUE relative increase. The reason for including the catch versus CPUE trade-off is to highlight that one should really concentrate on the trade-off of catch versus relative efficiency of capture, when considering the trade-off between catch and SSB recovery. It is clear that, over the short term, SSB recovery comes at the price of decreased average catches but we should also highlight that SSB recovery means CPUE recovery. Although average yields might be down, if catch is proportional to effort times exploitable biomass (which we assume is the case for the long-line fleet on the reference grid) then the trade-off, from an economic viewpoint, is less extreme than when simply thinking of catch versus SSB recovery, given less effort is required to take the lower TAC as the stock is now larger. It is noted that for at least one robustness grid (*omega75sqrt*) and perhaps also for the surface fishery, such a simple production function ($C=qEB$) does not apply and such SSB recovery will have a lesser impact economically. However, from a purely population-dynamic viewpoint, given the stock is low enough so that the stock-recruit relationship is effectively linear any percentage increase in SSB should yield the same percentage increase in mean recruitment. Given the selectivity of the surface fishery to the abundance of this component of the stock, such a positive effect will be felt there the fastest.

To initially explore the impact of lags in: i) the implementation of the TAC and ii) in the start of the MP we tuned the **ASMP** CMP to the 6 tuning levels with option 'd' (begin in 2013) and assuming a 1 year lag for implementation. Only for tuning level 3 (the strictest, 90%) was there any noticeable difference (with the reference catch level some 1000 tonnes lower than in the ‘c’/no lag base-case) – when the CMP was asked to rebuild to such a high percentile over the shorter-time frame the only option is to have a significant cut in average catches for tuning level 3. It is also worth noting that if one raises the maximum change in TAC to 5000t this difference in the reference catch becomes visibly smaller so there is a trade-off between decision and implementation time, the start of the MP, and the maximum reactivity of the CMP.

Robustness testing of the CMPs

Table 1 provides a comparison of the CMPs mean performance over all 20 robustness grids at the 3 levels each CMP was tuned to. The performance statistics considered were:

TunStat: (Primary tuning statistic) The probability of reaching 20% of SSB_0 in 2035 or 2040, depending on tuning level.
ST.stat.1: (Short-term statistic 1) The probability of reaching 10% of SSB_0 in 2022 or 2025, depending on tuning level.
ST.stat.2:(Short-term statistic 2) The probability of doubling the current SSB in 2022 or 2025, depending on tuning level.
AvCatch: (Average catch) The average catch (in tonnes) from MP start year until short term check point year, 2022 or 2025, depending on tuning level.
CvCatch: (CV catch) The coefficient of variation (ratio of standard deviation to mean) of the catch from MP start year until short term check point year, 2022 or 2025, depending on tuning level.
AvCPUE: (Average CPUE) The average CPUE from MP start year until short term check point year, 2022 or 2025, depending on tuning level.
CvCPUE:(CV CPUE) The coefficient of variation of the CPUE from MP start year until short term check point year, 2022 or 2025, depending on tuning level.

Looking at the performance of CMPs in this way does not show the detail of relative performance between robustness grids for a particular CMP and tuning, but does allow a first look at whether there is an obvious difference in overall performance between CMPs and the chosen tuning levels when tested with the full set of robustness grids. For tuning level 1, **BREM_1** tends to outperform the other CMPs in terms of both lower short-term risk (with the highest mean value for the short-term statistics) and with the highest average CPUE. This seems to be achieved through having the lowest average catch, and being more reactive which is evident in **BREM_1** having the highest variability in CPUE and catch. This trade-off between mean short-term statistics and CPUE, with average catch and variability in CPUE and catch is observed in the mean performance of the other CMPs at this tuning level. **ASMP**, after **BREM_1**, has the highest mean value for short-term statistics and high average CPUE, but again a lower average catch and high variability in catch and CPUE compared to **BREM_2** and **CCMP**. **BREM_2** actually results in SSB performance statistics lower than **CCMP**, with the lowest means for the short-term statistics given the highest average catch of 8589 t, which is 239 t greater than the constant catch for this tuning level.

Looking at the mean performance of CMPs at tuning level 5 provides much the same interpretation as that of tuning level 1. Tuning level 3, however, as the most extreme tuning level in terms of the restrictiveness required to get to the target biomass level in the shorter time-frame of 25 years with the highest probability of 0.9, means that relative performance between CMPs is different to that at the other two tuning levels. Though there is some change in relative performance between CMPs at this tuning level, the relationship between high mean short-term statistics and average CPUE relating to low average catch and high variability in catch and CPUE is still evident here. Though **BREM_1** does not perform the best in terms of the mean short-term statistics and average CPUE (that being achieved by **CCMP**), it still performs well relative to the other CMPs in these statistics at this tuning level.

Tables 2-5 provide the robustness performance statistics for individual robustness grids, for each of the four CMPs (at tuning level 1). This allows a comparison to be made of the performance of CMPs across all the robustness grids as already detailed in the MP documentation and listed below, if only at one tuning level, but such tables relating to other tuning levels can easily be provided. In terms of the detailed performance plots the graphics are those explained in the paper by Eveson (2003).

The robustness grids:

- **c1s1l2:** *LL1* overcatch scenario based on Case 2 of Market Report.
- **troll:** Include troll survey data.
- **tagmix:** Incomplete tag mixing: assume that season-1 *F*'s (*H*) (during which the surface fishery occurs) used in the tagging likelihood are 50% higher than the corresponding F's applied to the whole population.
- **recuncor:** Projected recruitment deviates uncorrelated to historical estimates from conditioning.
- **downwearlysize:** Downweight the initial size composition data for LL1 and LL4 (see Polacheck and Kolody, 2003, CCSBT-MP/0304/07).
- **regimeshift:** Regime shift: the stock-recruitment relationship changes in 1978. The two relationships share the same steepness parameter but two separate B_0 are estimated, one for each period.
- **aerdome**, **aerflat** Change selectivity of aerial survey (ages 2-4) throughout the series to [0.3,1,0.3] and [1,1,1] (instead of [0.5,1,1] assumed in the reference set). It was noted that it may be possible to reduce the options by closer inspection of the spotter data.
- **c0s1l1**, **c2s1l1**, **c3s1l1** Effects of overcatch on *CPUE*: S = 0%, 50% and 75%.
- **Laslett, STwin** Substitute alternative CPUE series by Laslett and ST-windows (the most extreme trends) to represent alternatives for changes in spatio-temporal distribution of fishing effort.
- **run3, run6** alternative glm model runs.
- **omega75** Omega value of 0.75 (CPUE non-linearity factor) or a higher value that is more supported by data (note that the value of that 0.75 has little support relative to the linear relationship).
- **highCPUECV:** In conditioning, increase lower bound of CV of *CPUE* to 0.30 (from 0.20 in base) and fix process error for aerial survey (tau_aerial) to 0.05. In projections use CV of CPUE = 0.30 and aerial CV=0.30.
- **highAerialCV** Increase CV of aerial survey to 0.50 while leaving CV of CPUE at 0.20. [*Note: does not require new grid]*.
- **upq**, **downq:** Step function change in catchability 30% up and 20% down between 2006 and 2007 unknown to the MP (to set up, change line 34 of sqrt.dat).

- **downupq:** Catchability goes down by 20% in 2007 and returns to normal in 5 years as fishermen adjust to new management regime (needs to be coded).
  Coding to be as for above, but with ramp back to "normal" in 5 years...
- **truncCPUE** Drop first 10 years of CPUE data.

CCMP

By its nature, the **CCMP** is unable to react to the different versions of 'reality' that the robustness grids assume. As there is no feedback between indicators of the state of the stock and the TAC set under the **CCMP**, there is a higher risk of further decline in spawning biomass, especially for robustness grids where the stock level is considerably lower than that of the reference grid. Figure 11 compares the performance of **CCMP** on robustness grids *omega75sqrt*, *STWinsqrt* and *upq* with the **CCMP** applied to the reference grid. These three robustness grids are considered as they represent the most problematic robustness grids in terms of low spawning biomass levels. The top half of Figure 11 mostly shows catch related statistics that generally don't change when comparing robustness grids due to the unreactive nature of the **CCMP**. For *omega75sqrt* there is some variation in catch levels which suggests that some of the replicates are reaching extremely low spawning biomass levels and so the TAC is not available to be taken from the stock. The bottom half of Figure 11, which contains mostly biomass/risk statistics, shows that even in 2032 there is still considerable risk that the stock will be below 2009 levels for the *omega75sqrt* and *STwinsqrt* scenarios, noting that the 2009 stock levels are lower than the reference grid levels. Though all of the CMPs are unable to reach target biomass levels under these low spawning biomass scenarios, **CCMP** represents the highest risk in terms of long-term stock levels.

ASMP

Figure 12 compares the performance of the **ASMP** on the reference grid with the **ASMP** applied primarily to robustness trials that make different assumptions about the aerial survey: *aerdome*, *aerflat, highAerialCV* (note this trial uses the reference grid but assumes an increase in CV for the aerial survey in projections) and *upq*. The inclusion of the *upq* trial is to see how a CMP that can react to population trends but is not influenced by CPUE interpretational complications might perform, relative to the other CMPs that are either unreactive or use the CPUE data. These plots show that **ASMP** is robust over these scenarios with very little change in the performance indicators as compared to those of the reference grid. For the *upq* scenario there are clearly some increases in short-term SSB depletion risk, decreases in longer-term SSB target performance but no observable difference in the consistency of the TAC and the biomass trend but an elevated catch-to-biomass ratio, which will become important when comparing **ASMP** with the BREM CMPs on this grid.

BREM_1

Figure 13 compares the performance of **BREM_1** on the reference grid to the key CPUE robustness grids: *downq, downupq, upq* and *highCPUECV*, although it is true that the aerial survey forms the second key index used in the BREM CMPs. Given the robustness of **ASMP** to the aerial survey-specific robustness grids, and the fact that the BREM CMPs use practically the same information in the actual harvest control rule (moving averages of the geometric mean recruitment ratio) whilst being more constrained than **ASMP** (integrated model that smoothes the signals), it is highly likely that the BREM CMPs will have the same robustness characteristics. Looking at Figure 13 in terms of catch performance statistics there is really very little difference across the key robustness grids, save the catch-to-biomass ratio for the *upq* grid is notably higher than all the others. In terms of biomass performance the *upq* and *highCPUECV* grids clearly separate from the others as relatively problematic for **BREM_1** to handle: SSB rebuilding relative to 2009, 1980 and MSY are all clearly worse than for the other grids, as are the short-term rebuilding and the minimum future biomass statistics. For the *highCPUECV* case, one could attribute this to the general lower starting state of the population (as the estimates of recent abundance are driven more by the tagging data than in the reference case) but this does not really follow completely from the plots and it does not really apply to the *upq* case either – the higher CPUE data driven by the upturn in *q* in 2006 and 2007 will be interpreted as a larger current exploitable stock driven by larger historic recruitments. Nor does it follow that their decreased SSB rebuilding performance is driven by higher estimates of catches – clearly from Figure 13 they are no higher than their counterparts who are performing better in terms of SSB. What seems to be the only remaining driver is the greater inconsistency between the biomass trends and the TAC set in Figure 13 – the strong step-change in *q* (*upq*) and the elevated uncertainty (*highCPUECV*) options seem to confuse the MP so that catches are not necessarily set too high but too often are not consistent with the trends in the real populations. See Figure 11 for a more extreme instance of this for **CCMP** and the *upq* robustness grid. Further evidence of this effect can be seen in the catch-to-biomass ratios in Figure 13: particularly for the *upq* grid these ratios are consistently higher than all the others, yet the yields are no higher on average even with this elevated harvest rate – the MP is tending to take higher catches from lower biomass levels given the signal confusion. Interestingly for the *downq* cases there is no such apparent signal confusion and reduced SSB rebuilding performance – quite the opposite presumably because the false trend is a negative one so the MP would interpret this as a stronger decline and act accordingly. There is no apparent decrease in biomass-to-TAC consistency perhaps driven by the fact that the *downq* effect is 20% while the *upq* one is 30%, but this is speculative.

BREM_2

Given the same data is used for **BREM_1** and **BREM_2** the same robustness grids were used to illustrate performance. All the points that applied to the performance of **BREM_1** and the *upq* and *highCPUECV* grids apply here also, as do those about the *downq* and *downupq* options – see Figure 14: Inconsistency between biomass trend and TACs set, with notably higher catch-to-biomass ratios for the *upq* case; poorer biomass rebuilding performance and short-term risk statistics. The only notably poorer performance observation (relative to **BREM_1**) is in the minimum future biomass relative to 2009 statistics – for **BREM_2** the median levels of this statistic for the *upq* and *highCPUECV* cases are further below the *downq, downupq* and reference cases and heavily skewed towards lower probabilities of maintaining minimum future biomass levels above the historical minimum of 2009.

Effect of the upq robustness grid across CMPs

Clearly, the *upq* robustness grid proved problematic to all the CMPs but to differing degrees – **CCMP** probably performed the worst across all grids but really not by much and only truly noticeably when looking at the occurrence of future SSB declines below 2009 levels and the frequency with which catch levels exceeded the true population size. We should, perhaps, break down the effect into two main factors: (i) the effect of an erroneous view of the population from conditioning on the observed CPUE (with higher *q* in 2006/2007), and (ii) the effect on CMPs that use this CPUE in some fashion to set future catches. Effect (i) seems to affect the whole suite of CMPs given they all display reduced SSB performance - even those which have nothing to do with CPUE. Effect (ii) seems to (perhaps obviously) affect the BREM CMPs more than most – **ASMP** now outperforms **BREM_1** in terms of the primary tuning target and is much closer in performance when looking at the short-term SSB statistics (Tables 3 and 4). Perhaps the clearest effect is in terms of the consistency of the TAC and the biomass trend and the instances of higher catch-to-biomass ratios – the BREM CMPs use of these adjusted CPUE data seems to result in not higher average catches but higher catches set on lower exploitable populations given the erroneous positive signals in the CPUE data not in line with what is happening in the actual exploitable population. There is still very little difference between the overall performance of say **ASMP** and **BREM_1** across the board, but this consistency issue is the one reasonably clear difference and potential advantage of a CMP not using CPUE.

Comparing more/less reactive CMPs in more extreme cases

Table 1 presented the mean performance of the CMPs over all robustness grids which can provide a good overall view but does not allow comparison to be made of how the CMPs respond to certain robustness trials. Of particular interest is how the CMPs perform when applied to the more extreme cases of robustness trials where particular assumptions have greatly changed the state and dynamics of the stock. To illustrate the comparison of CMPs on these extreme cases, two examples have been chosen from the set of robustness trials. Figure 15 compares the performance of the CMPs when applied to the robustness trial *omega75sqrt*. For all CMPs, *omega75sqrt* has the lowest probability of achieving the spawning biomass target for tuning level 1 when compared to the other robustness trials. Figure 16 compares the performance of CMPs when applied to the robustness trial *Laslettsqrt*. This trial provides the most optimistic view of the stock (apart from the troll robustness trial which does not allow for much comparison as the short-term and tuning statistics are 1 for all CMPs).

For *omega75sqrt*, the more reactive CMPs, **BREM_1** and **ASMP** are able to mitigate the short- and long-term risks to the spawning biomass more so than **CCMP** or **BREM_2,** as is evident in the spawning biomass related statistics in Figure 15. This is achieved in part by a lower average catch over the period (see Figure 15). For the more optimistic spawning biomass scenario, *Laslettsqrt*, Figure 16 shows all CMPs reaching comparable levels of spawning biomass in both the short- and long-term. The main difference in the CMPs when applied to this robustness trial is that the more reactive CMPs, **BREM_1** and **ASMP**, allow for higher average catches in the longer term. This suggests that **BREM_1** and **ASMP** are able to take advantage of a larger stock by increasing catch levels with little to no loss in spawning biomass rebuilding performance, further emphasising the earlier stronger cut/faster rebuilding/higher longer-term catches versus more gradual cuts/slow rebuilding/smaller and slower TAC increases trade-off already outlined on the CMPs tuned to the reference grid.

Concluding remarks

All of the CMPs detailed herein were able to tune to the 6 tuning levels, assuming that the MP started in 2012, with a TAC decision made every three years and with a minimum/maximum change of 100/3000 tonnes. The effect of a late start (2013) and a 1 year lag between TAC determination and TAC implementation was only significant at tuning level 3 – the most conservative of all the levels – and could be offset to some degree by raising the maximum change in TAC to 5000 tonnes. With respect to the reference grid, across all tuning levels there was a clear trade-off between short-term SSB rebuilding (the 2022/2025 statistics) and average short-term catch levels, which was expected given the current SSB depletion levels as demonstrated in the previous MP work. The more reactive CMPs that cut catches more strongly early in the period, whilst yielding lower short-term average catches, have better short-term rebuilding statistics, and in the longer-term (i.e between 2022/2025 and the target tuning years of 2035/2040) yield higher average catches, as a result of the stronger SSB recovery. The strictest tuning level, 3, narrowed this performance trade-off and in fact the constant catch MP (at a low level) performed the best in terms of short-term SSB rebuilding. The advantage of being able to adapt to trends in the population was negated by the inherent uncertainty due to changing catch levels that then propagate through the population – i.e. given a high target rebuilding percentile of 0.9 the best option is to set a low fixed catch which is indeed what happened. In terms of fixed catches versus feedback decision rules and general catch performance across all tuning levels, the less reactive but still dynamic CMP **BREM_2**, whilst maintaining better short-term rebuilding statistics (apart from the extreme case of tuning level 3), yielded higher average catches than the constant catch option – from either SSB-rebuilding or yield perspectives, in general, a tuned CMP will outperform a simple constant catch policy.

This apparent trade-off between reactivity and SSB recovery versus stable higher catch levels with higher short-term risk suggests, for the BREM models, perhaps a hybrid CMP that essentially consists of a weighted sum of the previous year's TAC and the "target" part of the **BREM_1** decision rule: $TAC_y = \omega TAC_{y-1} + (1-\omega)\ TAC_{BREM\_1}$. Such a CMP has already been coded but not tuned – if a middle ground between the **BREM_1** and **BREM_2** CMPs in a performance sense is considered useful then this can be achieved in the context of the current MP work program. Note also that such a hybrid harvest control rule (HCR) could be achieved using the aerial survey CMP as the basis: $TAC_y = \omega TAC_{y-1} + (1-\omega)\ TAC_{ASMP}$. For either case, given the work done already shows the clear trade-offs between one form over the other, we could work with the generic hybrid HCR detailed for both the BREM and ASMP type rules with a single weight parameter, ω, so that we can "slide" from the more to the less reactive forms.

In terms of performance across the robustness grids for tuning level 1, average performance across all 20 grids strengthened the observations made across tuning levels for the reference grids: the CMPs which are more responsive (**ASMP**, **BREM_1**) yielded lower average catch levels, better short-term SSB rebuilding statistics and higher average CPUE levels, whereas the less (**BREM_2**) and non-reactive (**CCMP**) CMPs on average yielded higher average TACs, worse short-term SSB rebuilding statistics and lower mean CPUE levels. More detailed inspection showed that, for the robustness grids where the spawning biomass level is lower than in the reference case, the more-reactive CMPs performed significantly better than either **BREM_2** or **CCMP** – these two CMPs (**CCMP** more so than **BREM_2**) led to a substantial number of stock trajectories going to zero and significantly higher risk of further SSB depletion relative to 2009. For the more optimistic robustness grids, SSB rebuilding performance was very similar. However, the more responsive MPs eventually out perform their less responsive/non-feedback counterparts as they are quicker to increase catch levels in response to the faster recovery of the stock. One interesting case was the relative performance of the CMPs on the *upq* robustness grid. All the CMPs' performance was affected (largely negatively) even though two of them do not use CPUE data – presumably driven by the falsely optimistic estimated state of the population caused by the apparent increase in CPUE in 2006 and 2007. The key observable performance difference between CMPs that use the CPUE (**BREM_1,2**) and those that can react but do not use it (**ASMP**) was in the consistency of the TAC set and the biomass level/trend and the occurrence of higher catch-to-biomass ratios, which in turn slightly increased the future probability of further SSB declines relative to 2009.

Acknowledgements

Many thanks to Jason Hartog for his invaluable assistance during the tuning phase of the work. This work was funded by the Department of Agriculture Fisheries and Forestry, the CSIRO "Wealth from Oceans Flagship" and the Australian Fisheries Management Authority.

References

Eveson, J.P. 2003. An update of the graphics used for evaluation the performance of candidate management procedures for southern bluefin tuna. CCSBT-ESC/0309/23.

Hillary, R.M., Eveson, J.P., Basson, M., and Davies, C. 2010. Technical specifications and proof of concept analyses for candidate management procedures for Southern bluefin tuna. *CCBST-OMMP/1006/4.*

Anonymous. 2010. Report of the Second Meeting of the Strategy and Fisheries Management Working Group Meeting.

Figures and Tables

Table 1: A summary of the mean robustness performance statistics (mean taken across all robustness grids) for each of the CMPs.

CMP	***ASMP***			***CCMP***			***BREM_1***			***BREM_2***		
Tuning level	*1*	*3*	*5*	*1*	*3*	*5*	*1*	*3*	*5*	*1*	*3*	*5*
AvTunStat	0.6	0.88	0.68	0.61	0.88	0.7	0.61	0.88	0.68	0.61	0.88	0.69
ST.stat1	0.45	0.6	0.6	0.41	0.65	0.54	0.49	0.61	0.66	0.39	0.55	0.53
ST.stat2	0.38	0.56	0.57	0.34	0.61	0.5	0.43	0.57	0.64	0.32	0.5	0.49
AvCatch	8005	4652	9352	8350	4240	9156	7572	4542	9022	8589	5617	9457
CvCatch	0.22	0.35	0.23	-	-	-	0.23	0.37	0.26	0.12	0.23	0.13
AvCPUE	0.67	0.73	0.73	0.65	0.76	0.7	0.69	0.74	0.75	0.64	0.71	0.7
CvCPUE	0.39	0.45	0.41	0.38	0.45	0.4	0.41	0.45	0.43	0.38	0.43	0.4

Table 2: The robustness performance statistics for CCMP at tuning level 1.

CCMP *tun.lev = 1*	*TunStat*	*ST.stat1*	*ST.stat2*	*AvCatch*	*CvCatch*	*AvCPUE*	*CvCPUE*
c1s1l13hsqrt	0.601	0.346	0.321	8353	-	0.59	0.38
c1s1l13hsqrt_downq	0.744	0.553	0.417	8353	-	0.56	0.39
c1s1l13hsqrt_downupq	0.734	0.524	0.400	8353	-	0.66	0.40
c1s1l13hsqrt_highCPUECV	0.401	0.182	0.203	8352	-	0.49	0.41
c1s1l13hsqrt_truncCPUE	0.752	0.525	0.365	8353	-	0.63	0.39
c1s1l13hsqrt_upq	0.387	0.170	0.217	8351	0.00	0.61	0.36
c1s1l13hsqrt_mixtag	0.605	0.347	0.354	8353	-	0.60	0.38
c1s1l13hsqrt_regime	0.621	0.385	0.319	8353	-	0.60	0.38
Laslettsqrt	0.898	0.778	0.450	8353	-	0.82	0.40
omega75sqrt	0.163	0.053	0.116	8294	0.02	0.42	0.30
STwinsqrt	0.261	0.090	0.127	8352	-	0.43	0.37
c0s1l1sqrt	0.636	0.453	0.323	8353	-	0.57	0.37
c1s1l2sqrt	0.665	0.412	0.365	8353	-	0.61	0.39
c1s1l13hsqrt_aerdome	0.609	0.360	0.332	8353	-	0.60	0.38
c1s1l13hsqrt_aerflat	0.598	0.353	0.315	8353	-	0.59	0.38
c1s1l13hsqrt_earlysize	0.510	0.299	0.265	8353	-	0.57	0.37
c1s1l13hsqrt_troll	1.000	1.000	1.000	8353	-	1.72	0.37
c2s1l1sqrt	0.588	0.336	0.256	8353	-	0.61	0.39
c3s1l1sqrt	0.576	0.363	0.192	8353	-	0.64	0.40
run3sqrt	0.802	0.647	0.438	8353	-	0.71	0.39
run6sqrt	0.590	0.341	0.282	8353	-	0.59	0.38

Table 3: The robustness performance statistics for ASMP at tuning level 1.

ASMP *tun.lev = 1*	*TunStat*	*ST.stat1*	*ST.stat2*	*AvCatch*	*CvCatch*	*AvCPUE*	*CvCPUE*
c1s1l13hsqrt	0.601	0.392	0.385	7765	0.22	0.62	0.40
c1s1l13hsqrt_downq	0.761	0.609	0.475	8106	0.22	0.58	0.41
c1s1l13hsqrt_downupq	0.742	0.591	0.453	8064	0.22	0.68	0.41
c1s1l13hsqrt_highCPUECV	0.263	0.167	0.189	8321	0.20	0.51	0.42
c1s1l13hsqrt_truncCPUE	0.758	0.596	0.422	8091	0.22	0.65	0.40
c1s1l13hsqrt_upq	0.399	0.200	0.281	7314	0.22	0.66	0.39
c1s1l13hsqrt_mixtag	0.610	0.400	0.418	7759	0.22	0.63	0.40
c1s1l13hsqrt_regime	0.618	0.430	0.378	7866	0.22	0.62	0.40
Laslettsqrt	0.903	0.844	0.486	8478	0.22	0.84	0.40
omega75sqrt	0.177	0.064	0.173	6400	0.24	0.47	0.30
STwinsqrt	0.257	0.096	0.174	7185	0.22	0.47	0.40
c0s1l1sqrt	0.650	0.527	0.379	7788	0.22	0.59	0.38
c1s1l2sqrt	0.677	0.481	0.429	7826	0.22	0.64	0.41
c1s1l13hsqrt_aerdome	0.604	0.398	0.383	7877	0.22	0.62	0.40
c1s1l13hsqrt_aerflat	0.590	0.398	0.375	7839	0.22	0.62	0.40
c1s1l13hsqrt_earlysize	0.508	0.347	0.310	7689	0.22	0.60	0.39
c1s1l13hsqrt_troll	1.000	1.000	1.000	11729	0.16	1.65	0.36
c2s1l1sqrt	0.581	0.369	0.319	7824	0.22	0.64	0.41
c3s1l1sqrt	0.563	0.405	0.237	7929	0.22	0.66	0.42
run3sqrt	0.812	0.723	0.486	8256	0.22	0.73	0.40
run6sqrt	0.577	0.403	0.337	7752	0.22	0.62	0.40

Table 4: The robustness performance statistics for BREM_1 at tuning level 1.

BREM_1 *tun.lev = 1*	*TunStat*	*ST.stat1*	*ST.stat2*	*AvCatch*	*CvCatch*	*AvCPUE*	*CvCPUE*
c1s1l13hsqrt	0.586	0.446	0.426	7258	0.23	0.63	0.41
c1s1l13hsqrt_downq	0.813	0.673	0.543	7131	0.23	0.60	0.42
c1s1l13hsqrt_downupq	0.717	0.621	0.495	7721	0.23	0.69	0.42
c1s1l13hsqrt_highCPUECV	0.448	0.263	0.317	6726	0.22	0.54	0.46
c1s1l13hsqrt_truncCPUE	0.759	0.650	0.483	7548	0.23	0.67	0.42
c1s1l13hsqrt_upq	0.308	0.207	0.298	7362	0.23	0.67	0.40
c1s1l13hsqrt_mixtag	0.602	0.445	0.469	7303	0.23	0.64	0.41
c1s1l13hsqrt_regime	0.608	0.481	0.438	7347	0.23	0.64	0.41
Laslettsqrt	0.843	0.840	0.483	8686	0.24	0.84	0.40
omega75sqrt	0.191	0.081	0.201	6149	0.23	0.47	0.31
STwinsqrt	0.307	0.137	0.235	6303	0.24	0.50	0.43
c0s1l1sqrt	0.665	0.575	0.421	7145	0.23	0.61	0.40
c1s1l2sqrt	0.669	0.521	0.480	7390	0.23	0.65	0.42
c1s1l13hsqrt_aerdome	0.610	0.454	0.444	7314	0.23	0.64	0.41
c1s1l13hsqrt_aerflat	0.600	0.451	0.427	7288	0.23	0.63	0.41
c1s1l13hsqrt_earlysize	0.517	0.392	0.350	7171	0.23	0.62	0.40
c1s1l13hsqrt_troll	1.000	1.000	1.000	12646	0.24	1.67	0.36
c2s1l1sqrt	0.563	0.422	0.360	7389	0.23	0.66	0.42
c3s1l1sqrt	0.539	0.439	0.275	7514	0.24	0.68	0.43
run3sqrt	0.781	0.739	0.508	8042	0.23	0.74	0.41
run6sqrt	0.579	0.441	0.392	7265	0.23	0.63	0.42

Table 5: The robustness performance statistics for BREM_2 at tuning level 1.

BREM_2 *tun.lev = 1*	*TunStat*	*ST.stat1*	*ST.stat2*	*AvCatch*	*CvCatch*	*AvCPUE*	*CvCPUE*
c1s1l13hsqrt	0.599	0.329	0.297	8509	0.12	0.59	0.38
c1s1l13hsqrt_downq	0.759	0.541	0.401	8608	0.11	0.56	0.39
c1s1l13hsqrt_downupq	0.728	0.507	0.375	8677	0.12	0.65	0.40
c1s1l13hsqrt_highCPUECV	0.397	0.158	0.181	8466	0.08	0.49	0.41
c1s1l13hsqrt_truncCPUE	0.752	0.508	0.345	8650	0.11	0.63	0.39
c1s1l13hsqrt_upq	0.393	0.155	0.196	8366	0.12	0.61	0.36
c1s1l13hsqrt_mixtag	0.616	0.328	0.328	8509	0.12	0.59	0.38
c1s1l13hsqrt_regime	0.624	0.361	0.294	8549	0.12	0.59	0.38
Laslettsqrt	0.880	0.771	0.417	8940	0.12	0.81	0.39
omega75sqrt	0.174	0.046	0.111	7889	0.14	0.42	0.29
STwinsqrt	0.274	0.079	0.118	8200	0.12	0.43	0.37
c0s1l1sqrt	0.644	0.448	0.306	8511	0.12	0.56	0.37
c1s1l2sqrt	0.660	0.393	0.342	8550	0.12	0.61	0.39
c1s1l13hsqrt_aerdome	0.615	0.345	0.311	8477	0.12	0.59	0.38
c1s1l13hsqrt_aerflat	0.602	0.338	0.298	8499	0.12	0.59	0.38
c1s1l13hsqrt_earlysize	0.519	0.286	0.243	8472	0.12	0.57	0.37
c1s1l13hsqrt_troll	1.000	1.000	1.000	10079	0.13	1.69	0.37
c2s1l1sqrt	0.579	0.317	0.237	8540	0.12	0.61	0.39
c3s1l1sqrt	0.572	0.348	0.176	8590	0.12	0.63	0.40
run3sqrt	0.808	0.633	0.419	8779	0.12	0.71	0.39
run6sqrt	0.588	0.330	0.262	8502	0.12	0.59	0.38

Figure 1: short-term tuning statistic 1 (p(B[y]>0.1B[0])) at tuning levels 1 (top left), 3 (top right), and 5 (bottom left) for each of the CMPs on the reference grid.

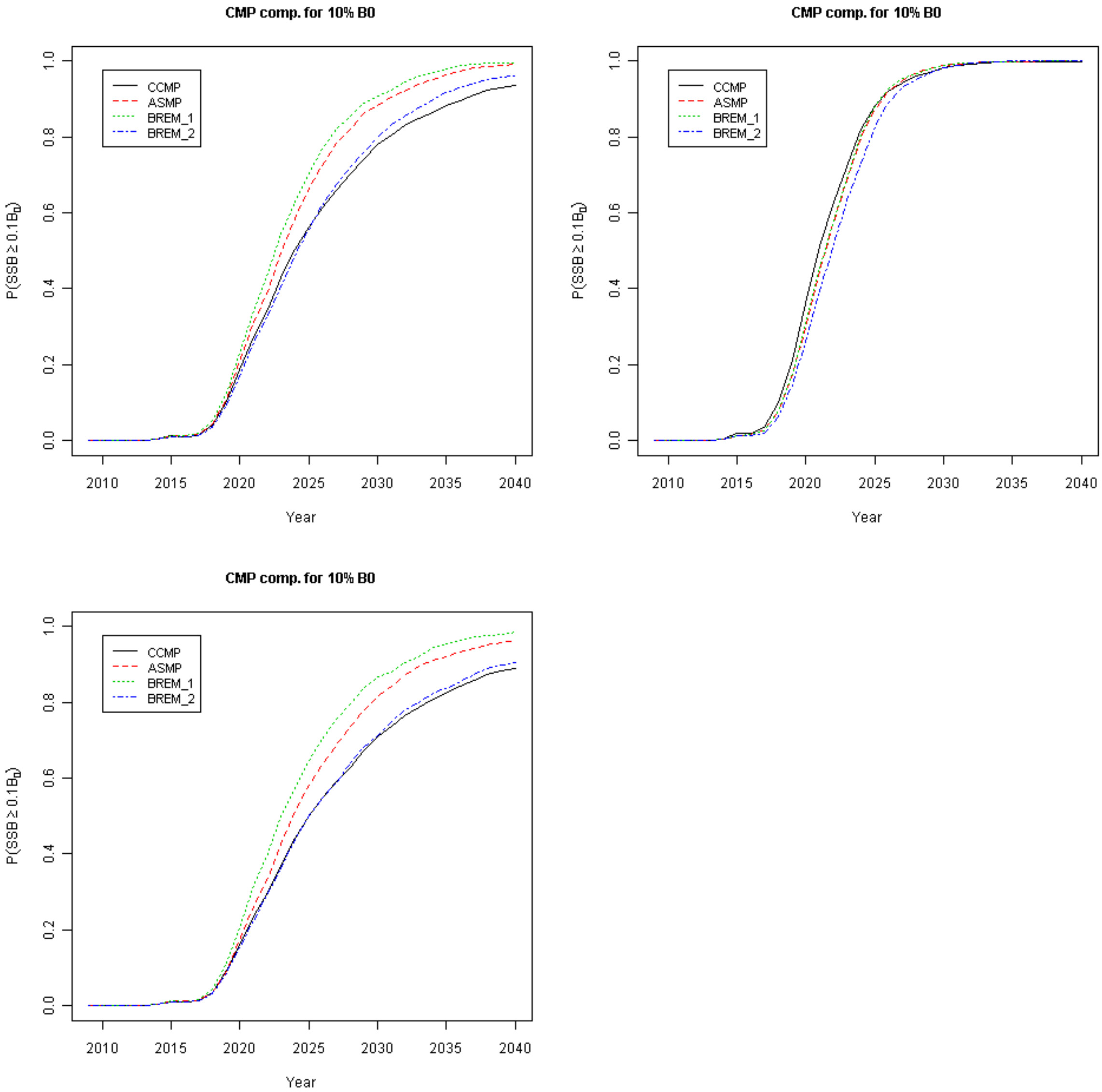


Figure 2: short-term tuning statistic 2 (p(B[y]>2*B[2009])) at tuning levels 1 (top left), 3 (top right), and 5 (bottom left) for each of the CMPs on the reference grid.

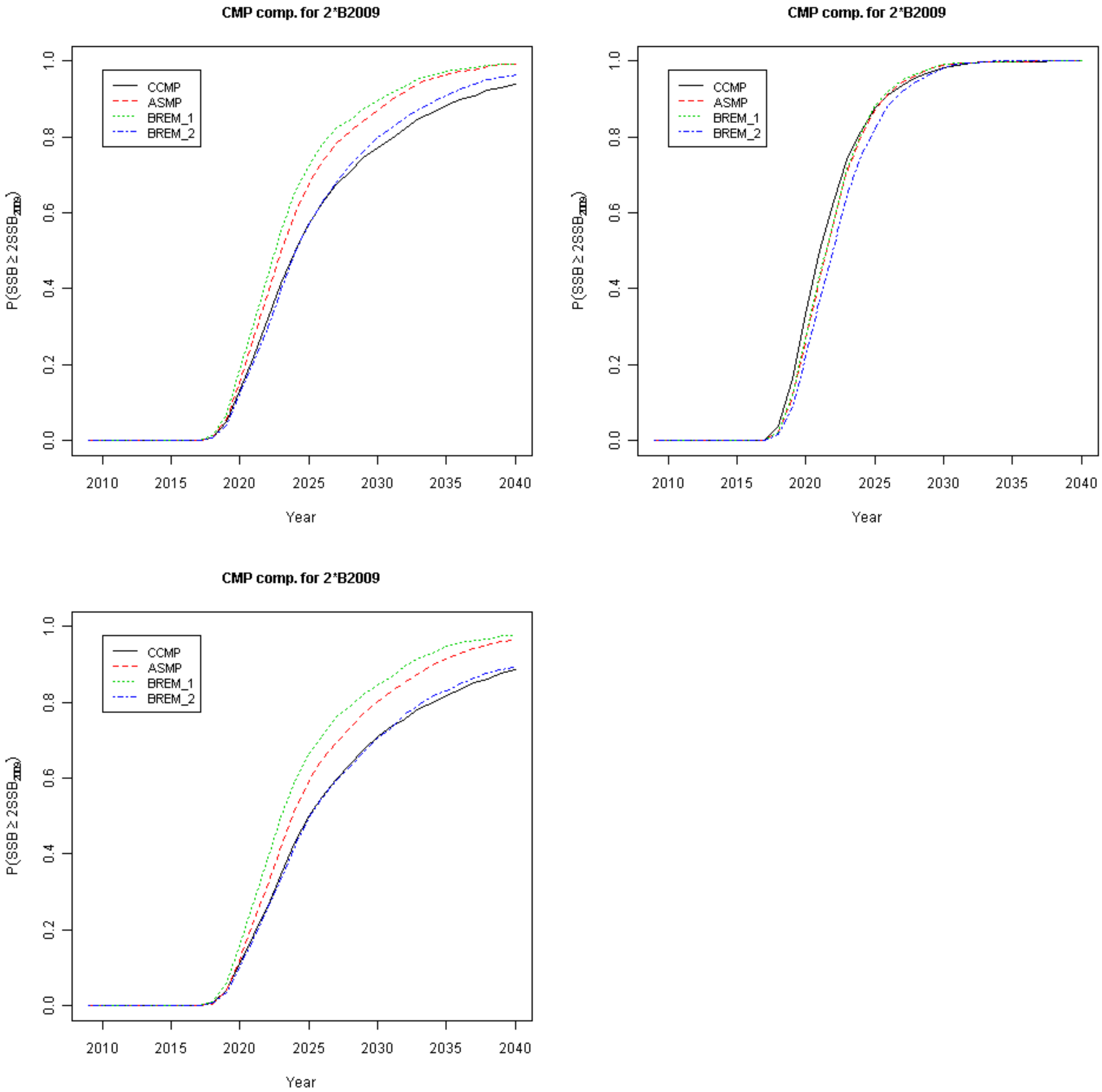


Figure 3: SSB and catch worm plots at tuning levels 1 (top left), 3 (top right), and 5 (bottom left) for CCMP.

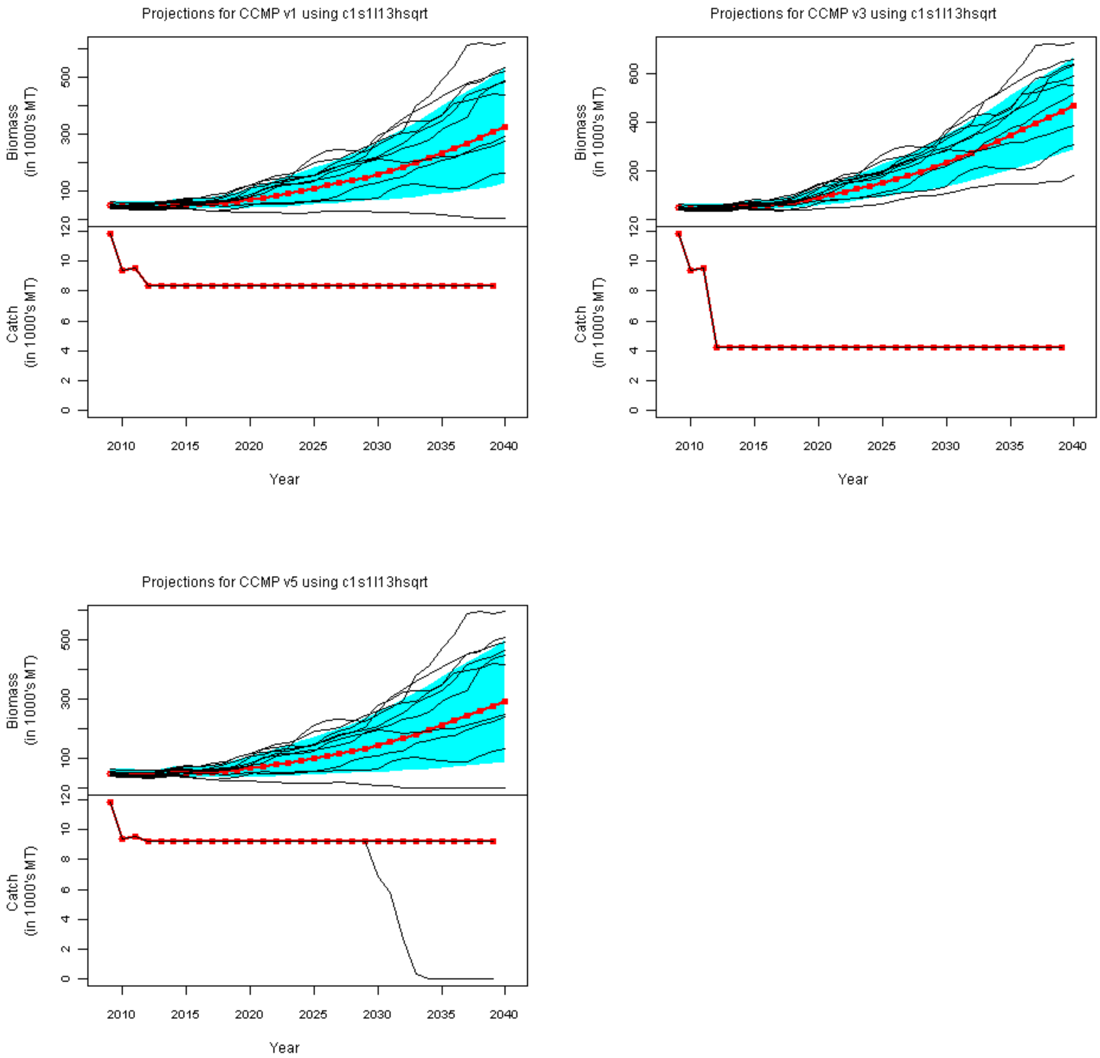


Figure 4: SSB and catch worm plots at tuning levels 1 (top left), 3 (top right), and 5 (bottom left) for ASMP.

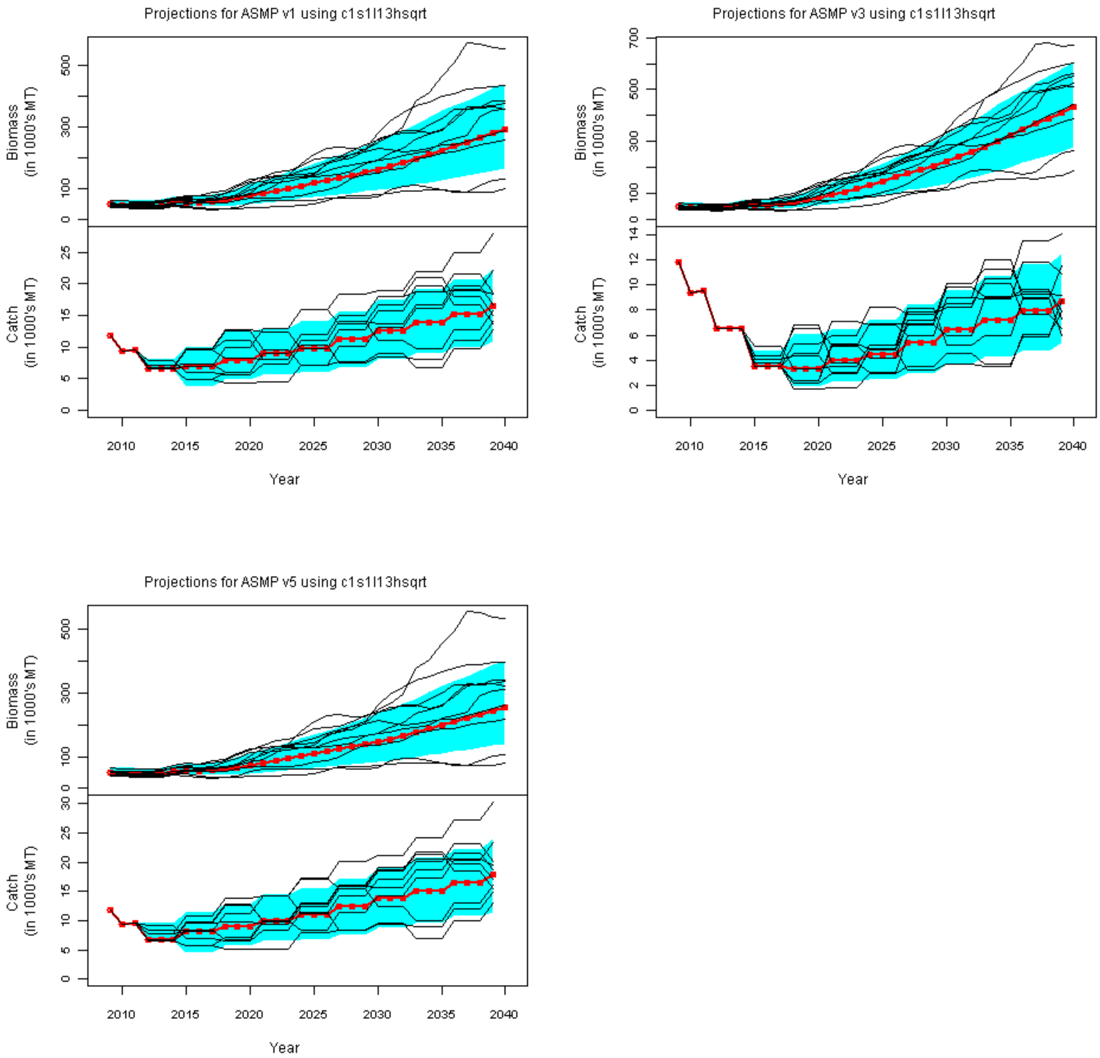


Figure 5: SSB and catch worm plots at tuning levels 1 (top left), 3 (top right), and 5 (bottom left) for BREM_1.

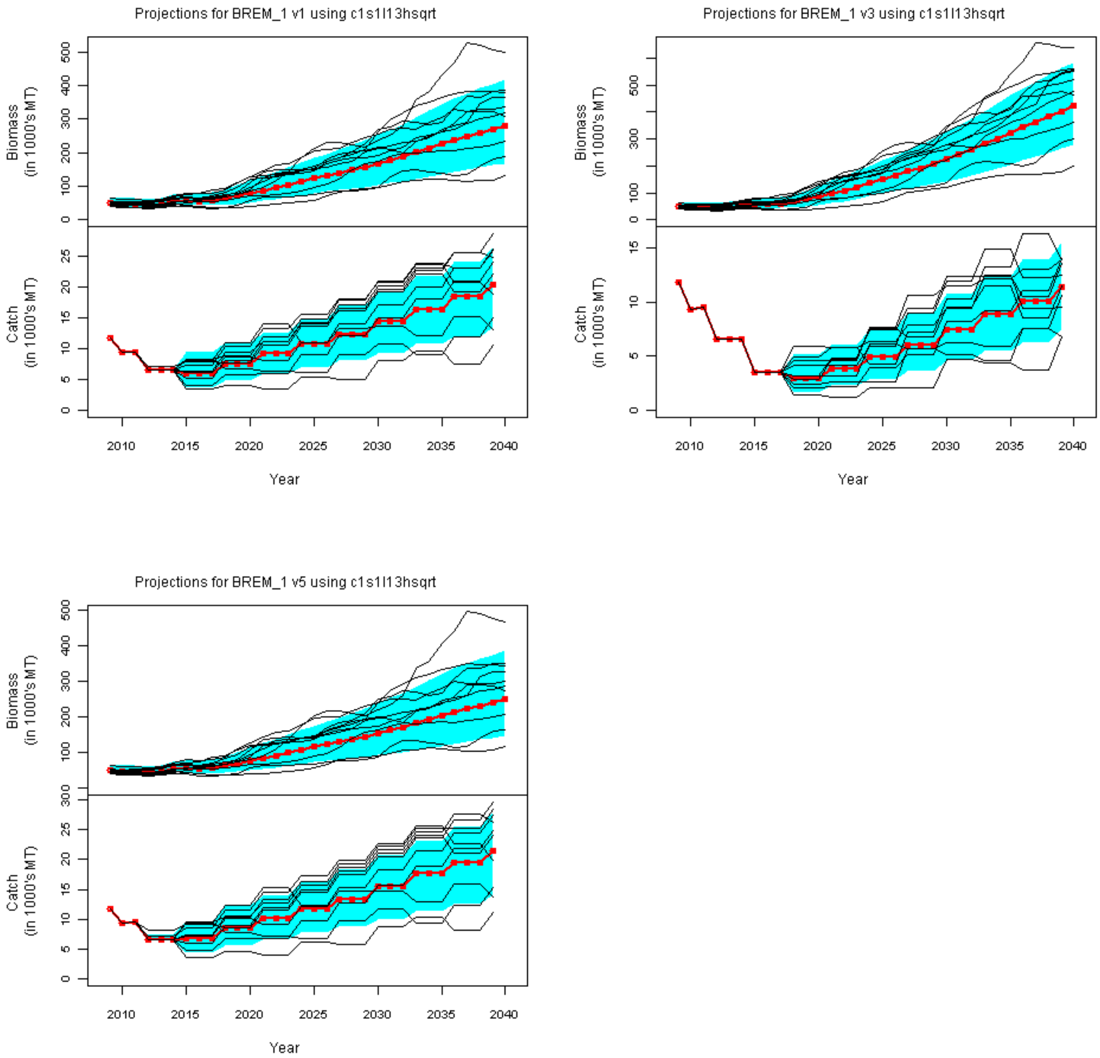


Figure 6: SSB and catch worm plots at tuning levels 1 (top left), 3 (top right), and 5 (bottom left) for BREM_2.

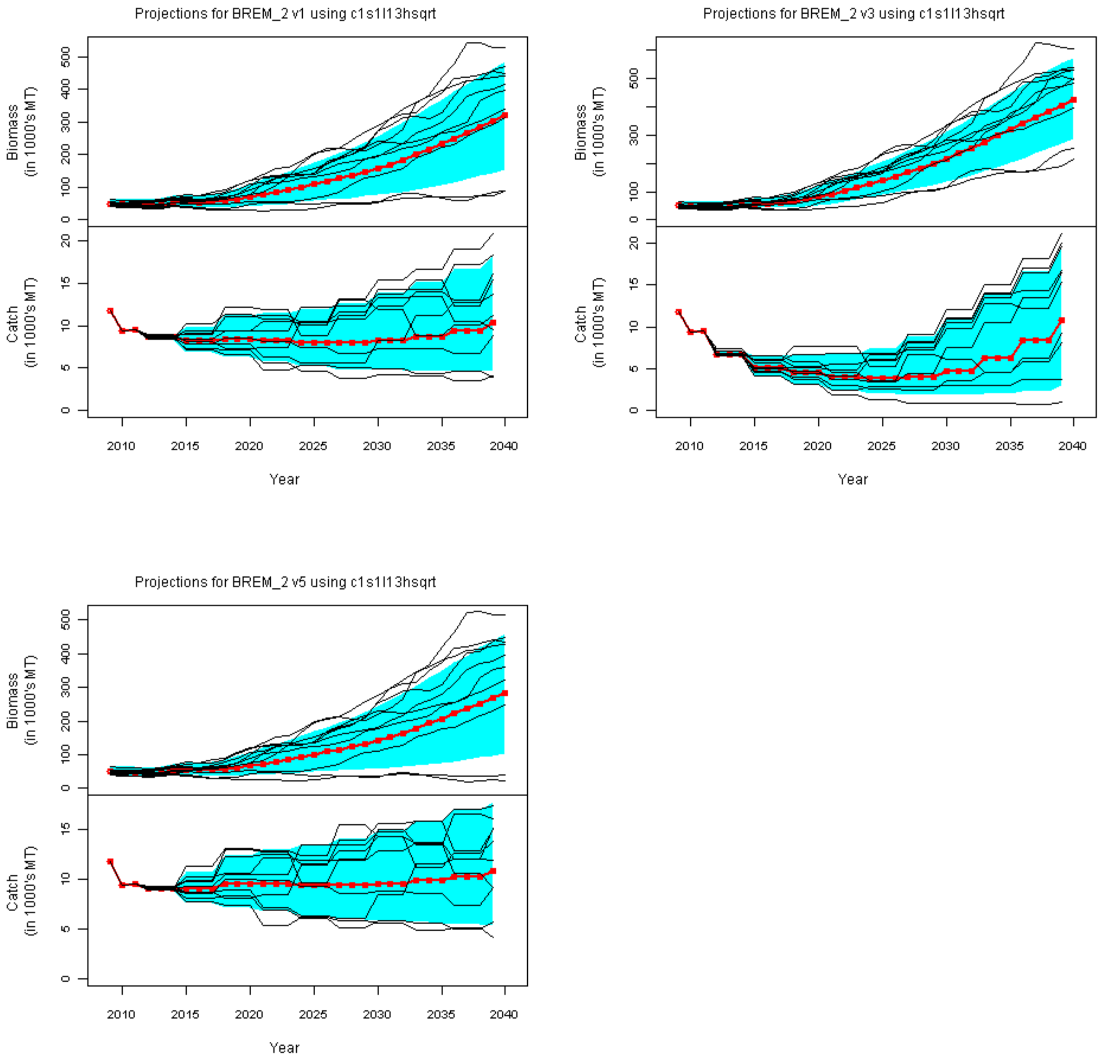


Figure 7: Boxplot summaries of average catch (from start of MP to year of the first short-term rebuilding checkpoint) at tuning levels 1 (top left), 3 (top right), and 5 (bottom left). The orange broken line is the current 9,449t TAC level.

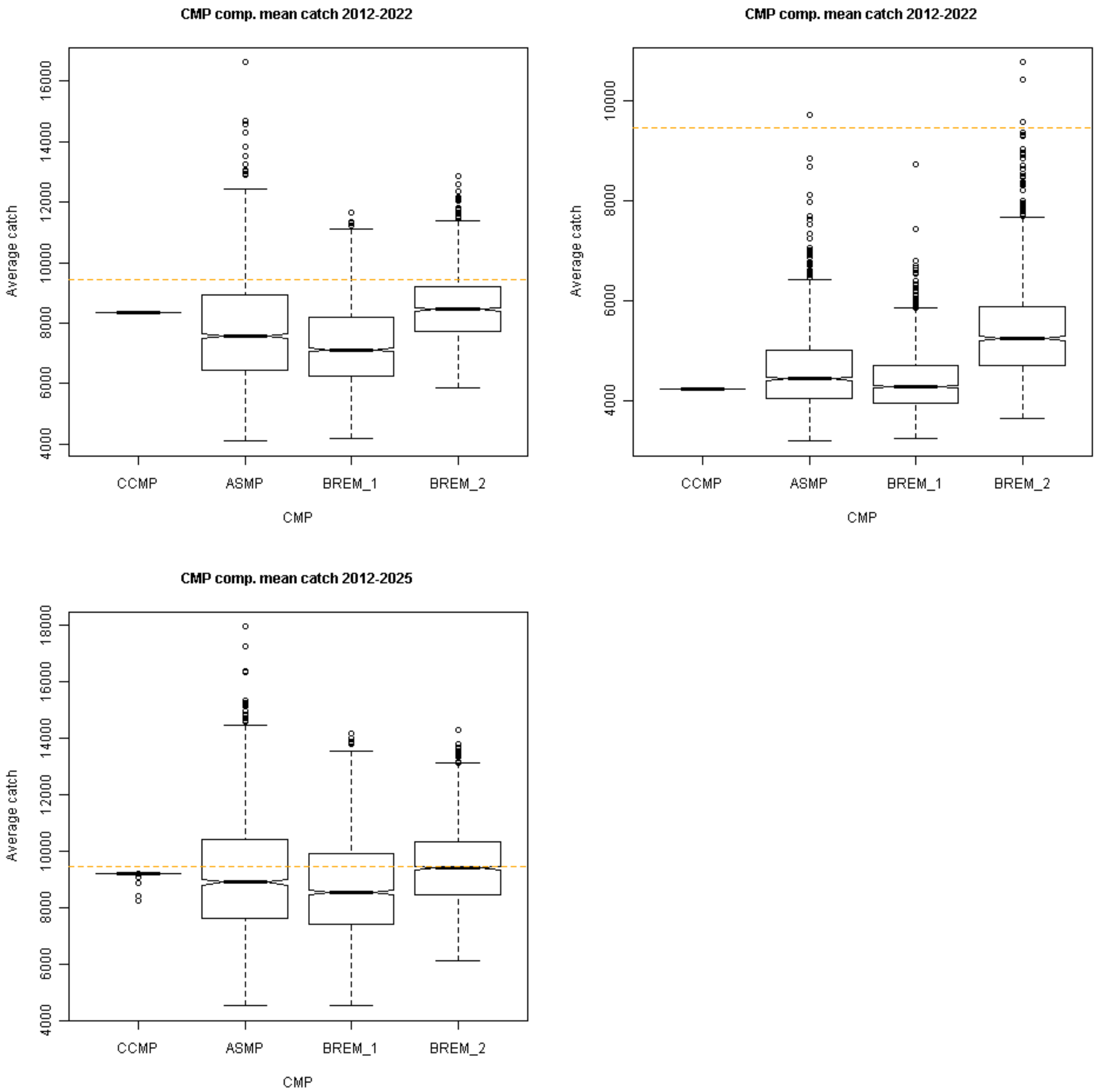


Figure 8: Boxplot summaries of the percentage change in TAC when a change was permitted at tuning levels 1 (top left), 3 (top right), and 5 (bottom left). Note this is *not* the same as the AAV statistic which in, this case, deflates the true percentage change by including the zero changes in the years when no change was allowed.

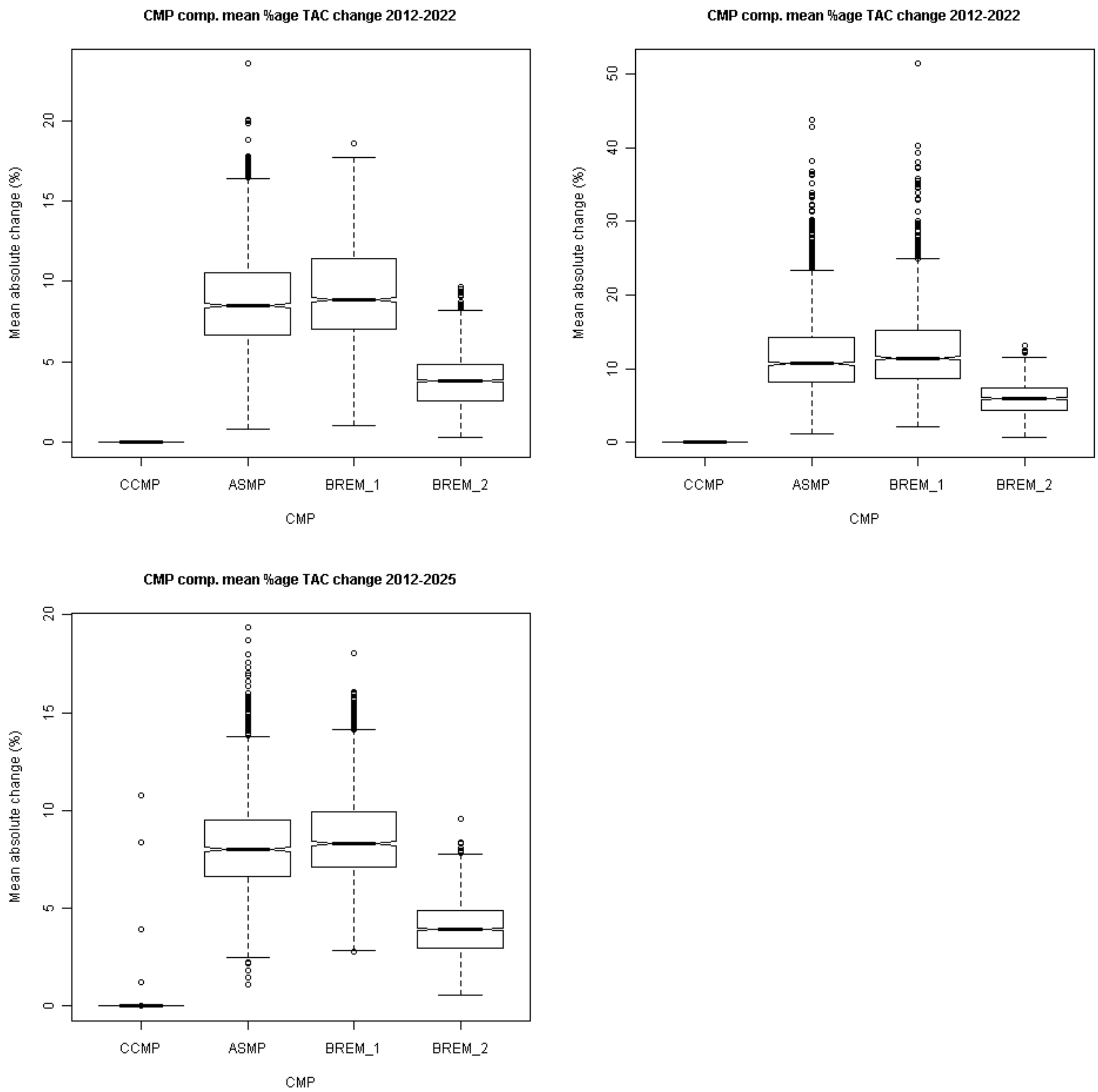


Figure 9: Catch versus SSB recovery trade-off plots at tuning levels 1 (top left), 3 (top right), and 5 (bottom left). On the *y*-axis is the 95% CI of the average catch (averaged over the period from the start of the MP to the year of the short-term statistics) while on the *x*-axis is the 95% CI of the the year-averaged SSB (averaged over the same time period) relative to 2009.

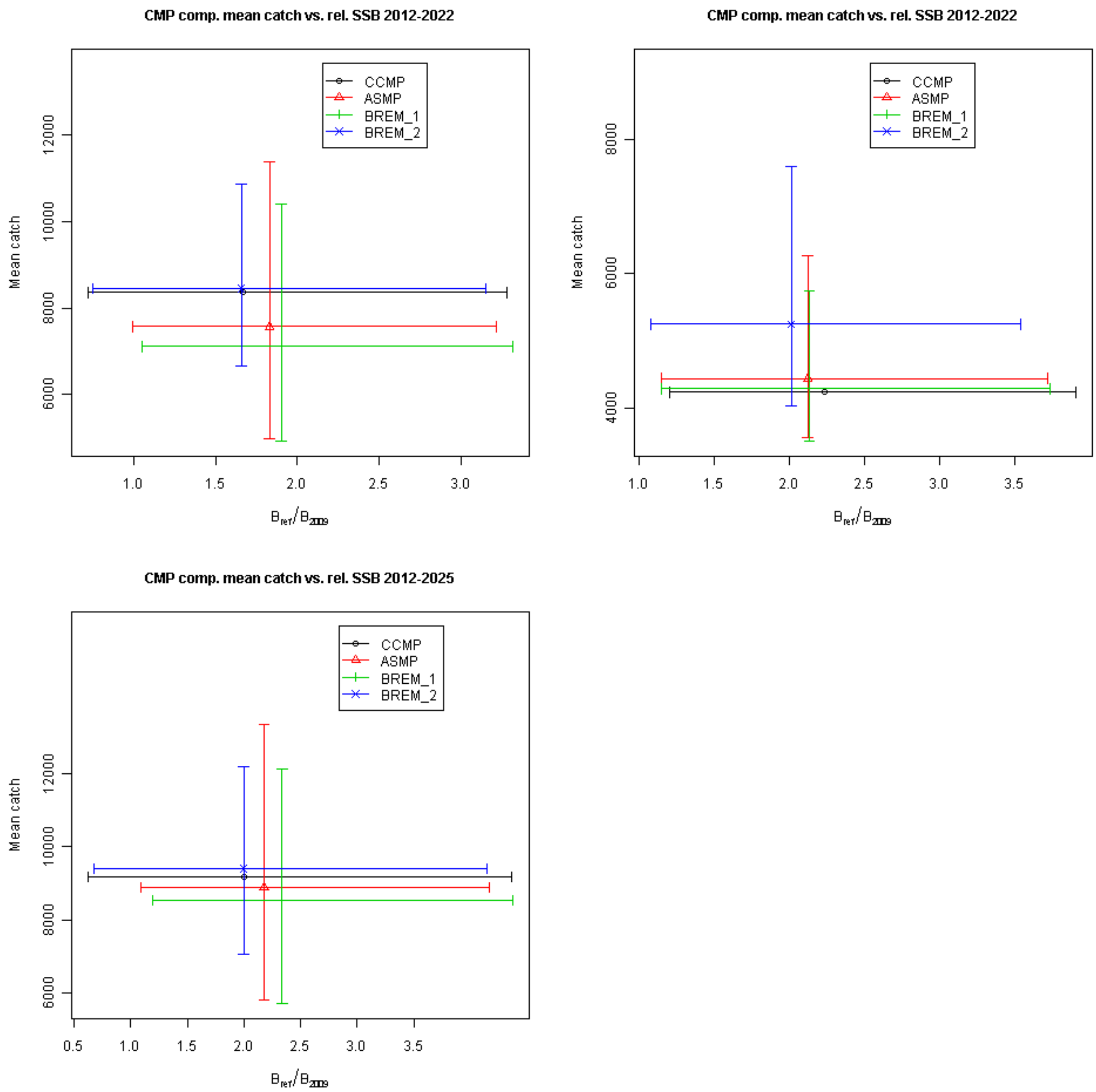


Figure 10: Catch versus CPUE recovery trade-off plots at tuning levels 1 (top left), 3 (top right), and 5 (bottom left). On the *y*-axis is the 95% CI of the average catch (averaged over the period from the start of the MP to the year of the short-term statistics) while on the *x*-axis is the 95% CI of the the year-averaged CPUE (averaged over the same time period) relative to 2009.

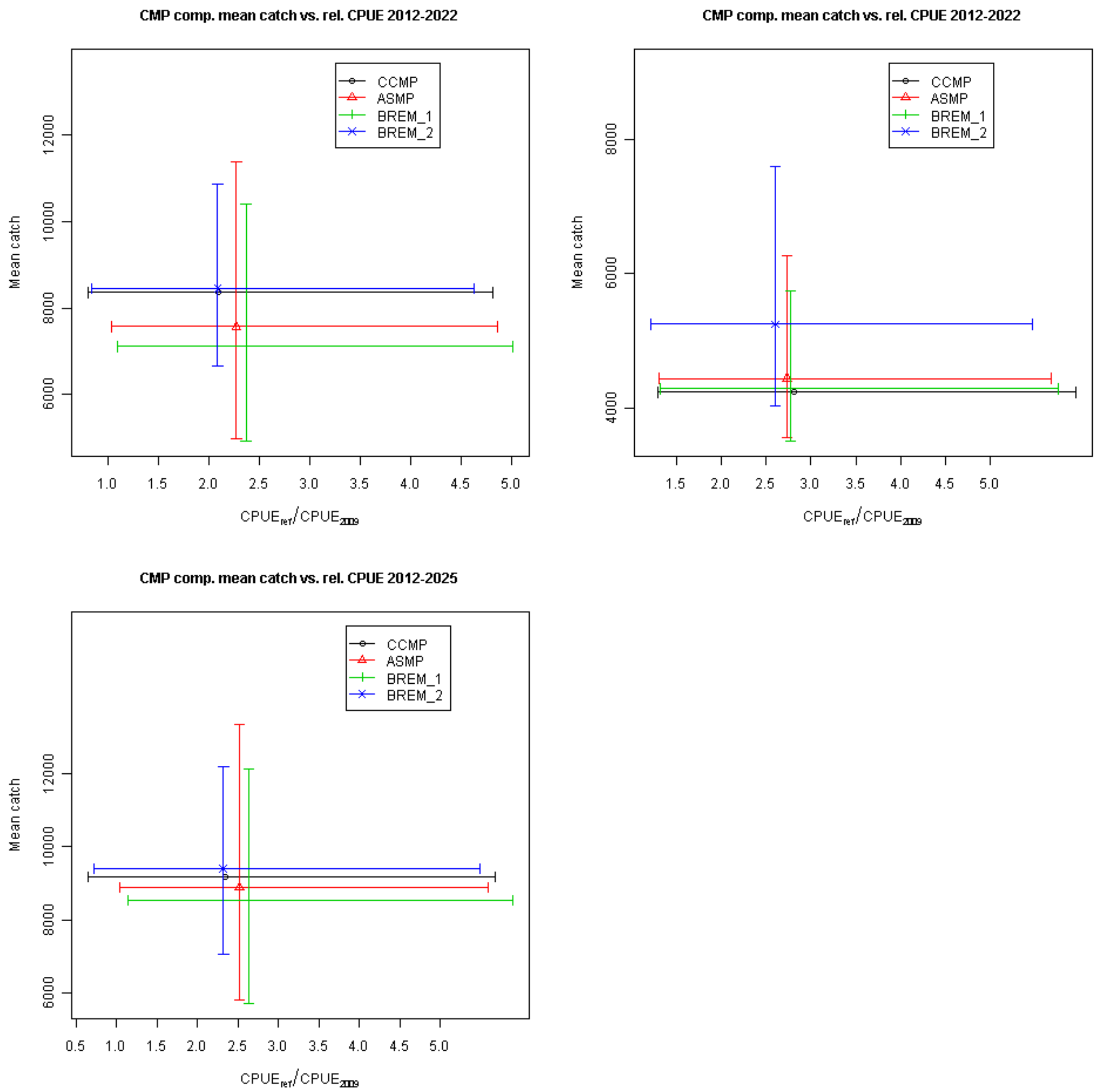
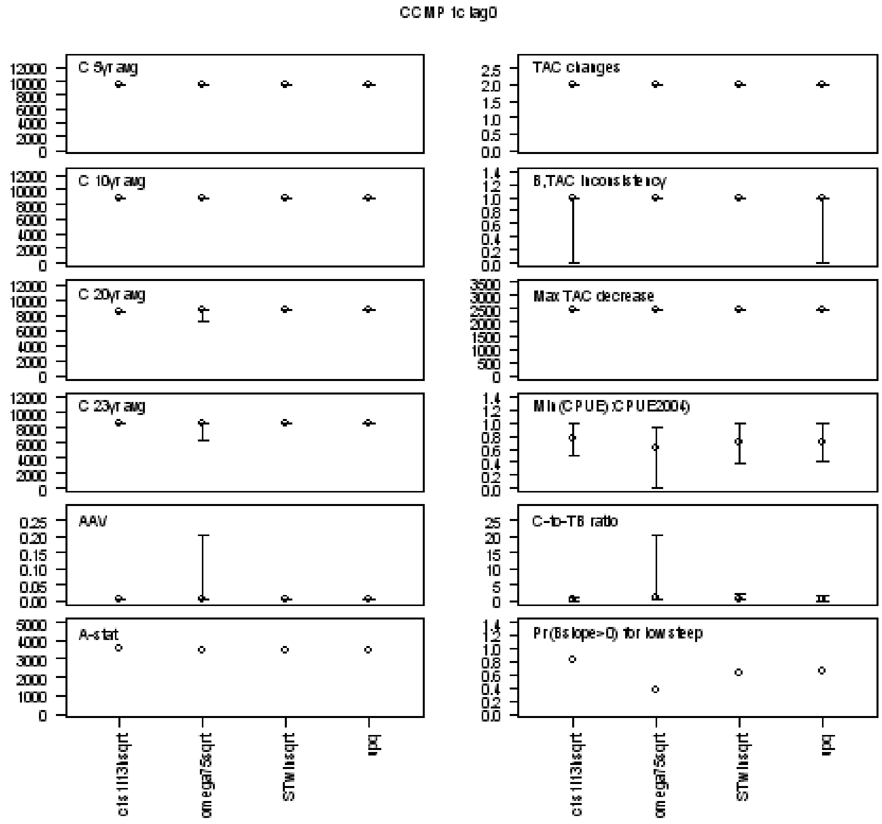


Figure 11: Comparison of robustness trials omega75sqrt, STWinsqrt, and upq with the reference grid c1s1l13hsqrt of CCMP using the set of agreed performance indicators



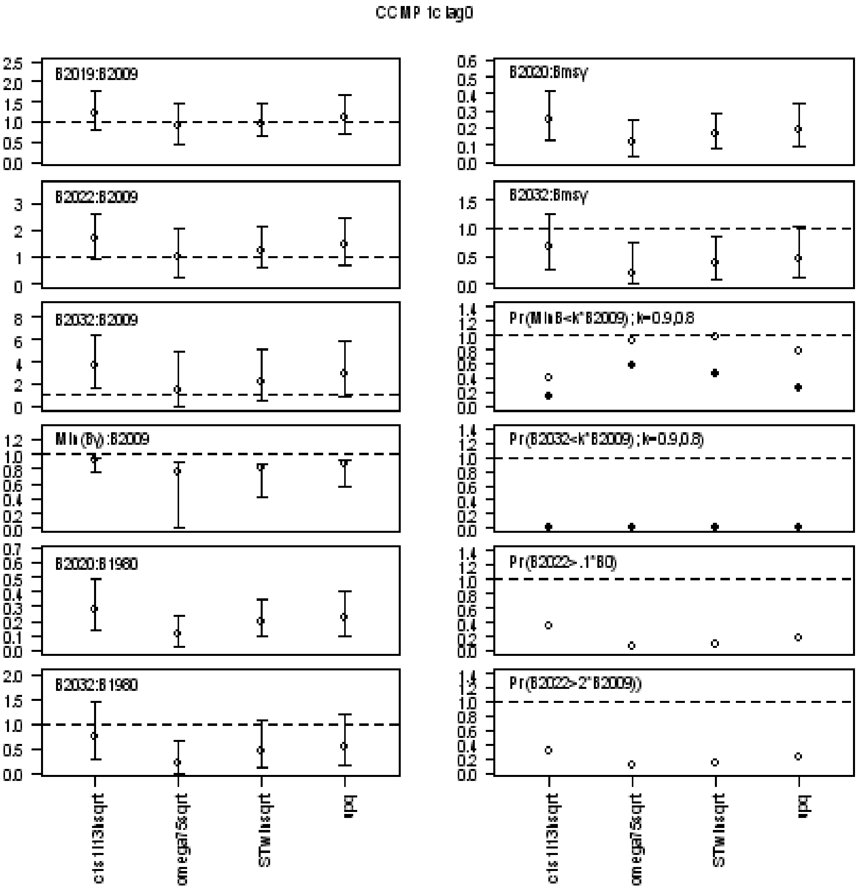
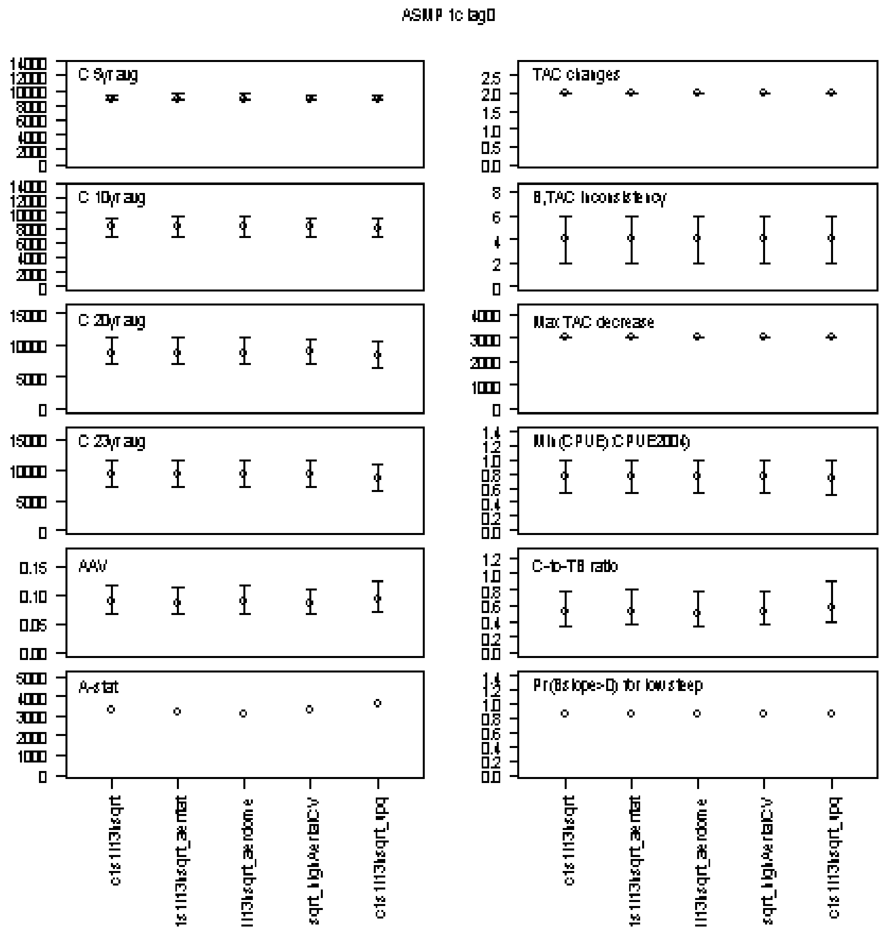

CCMP 1c lag0
B2019:B2009
B2020:Bmsy
B2022:B2009
B2032:Bmsy
B2032:B2009
Pr(MinB<k*B2009) ; k=0.9,0.8
Min(By):B2009
Pr(B2032<k*B2009) ; k=0.9,0.8)
B2020:B1980
Pr(B2022>.1*B0)
B2032:B1980
Pr(B2022>2*B2009))
c1s1l13lsqrt
omega75sqrt
STwlsqrt
ipq

Figure 12: Comparison of robustness trials aerdome, aerflat, and highAerialCV with the reference grid c1s1l13hsqrt of ASMP using the set of agreed performance indicators.



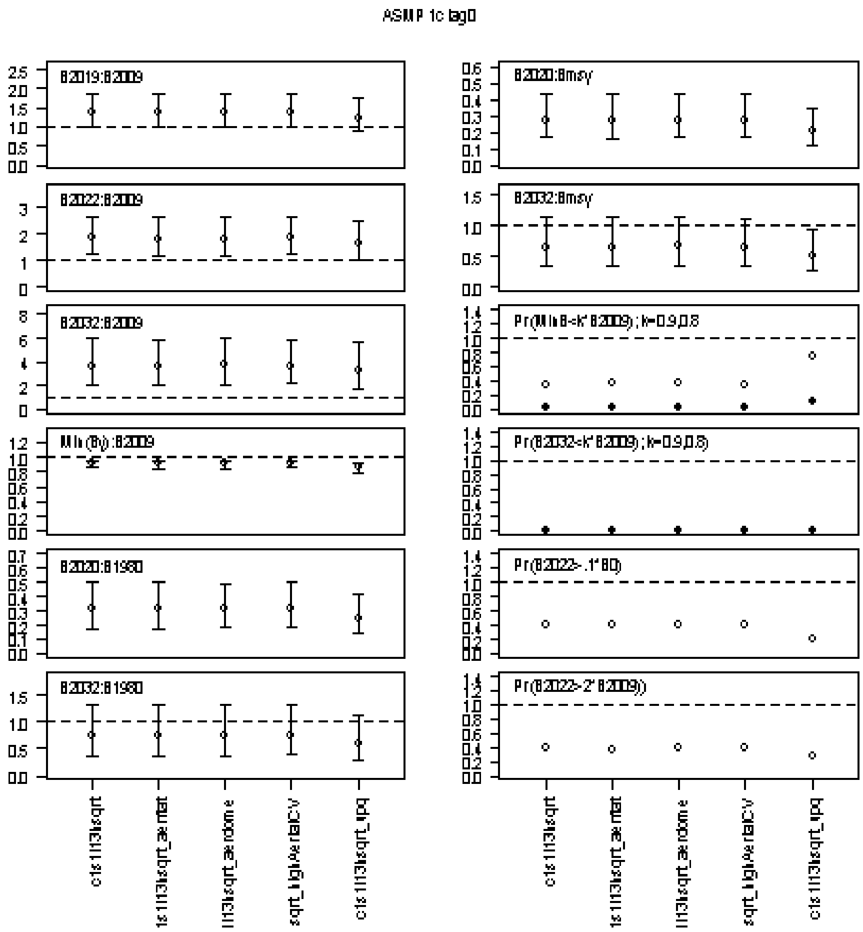
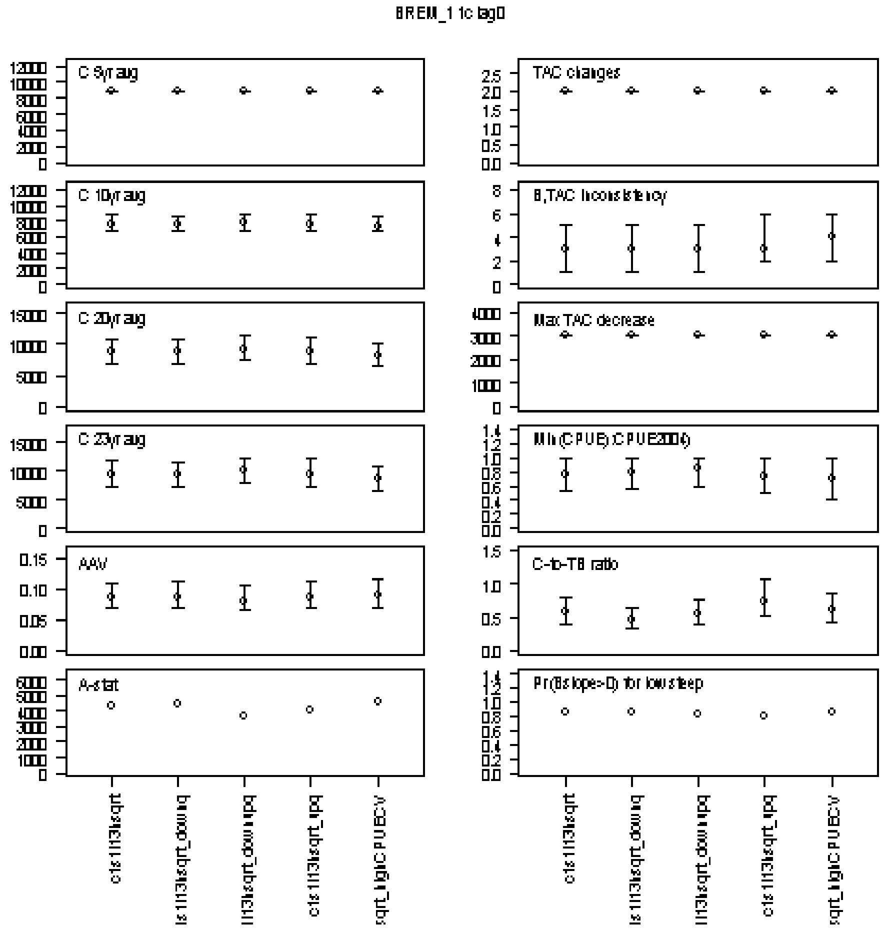

ASMP 1c lag0
B2019:B2009
B2020:Bmsy
B2022:B2009
B2032:Bmsy
B2032:B2009
Pr(MinB<k*B2009); k=0.9,0.8
Min(Bt):B2009
Pr(B2032<k*B2009); k=0.9,0.8)
B2020:B1980
Pr(B2022>.1*B0)
B2032:B1980
Pr(B2022>2*B2009))

Figure 13: Comparison of robustness trials on downq, downupq, upq and highCPUECV with the reference grid c1s1l1shsqrt of BREM_1 using the set of agreed performance indicators.



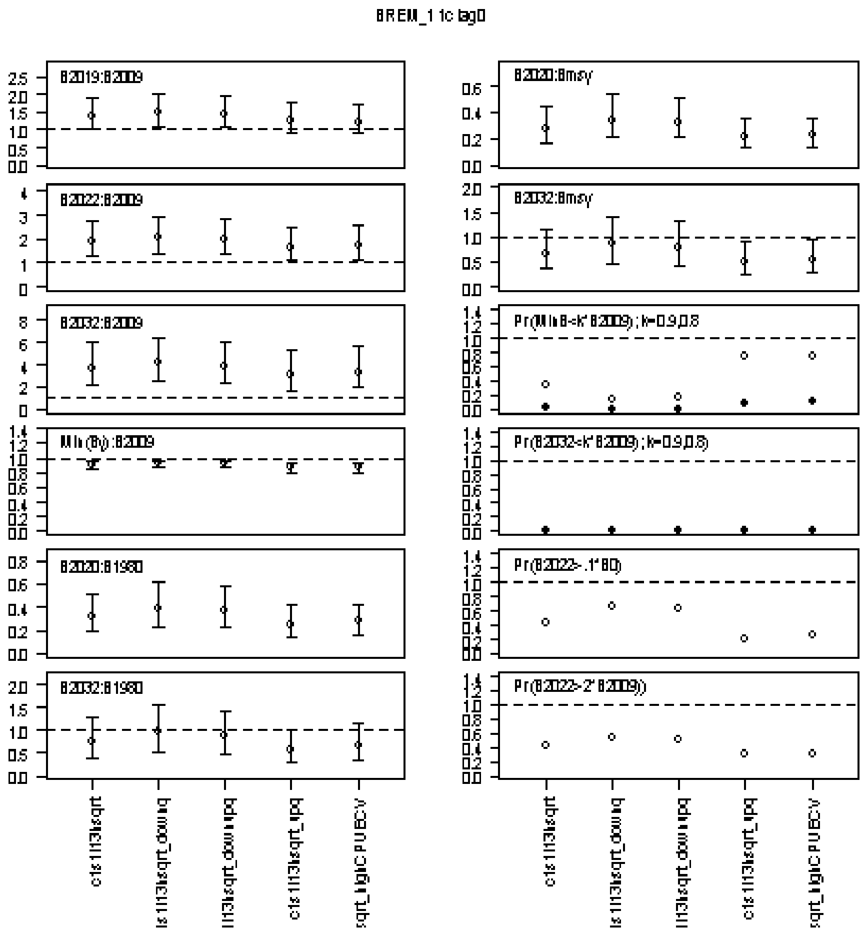
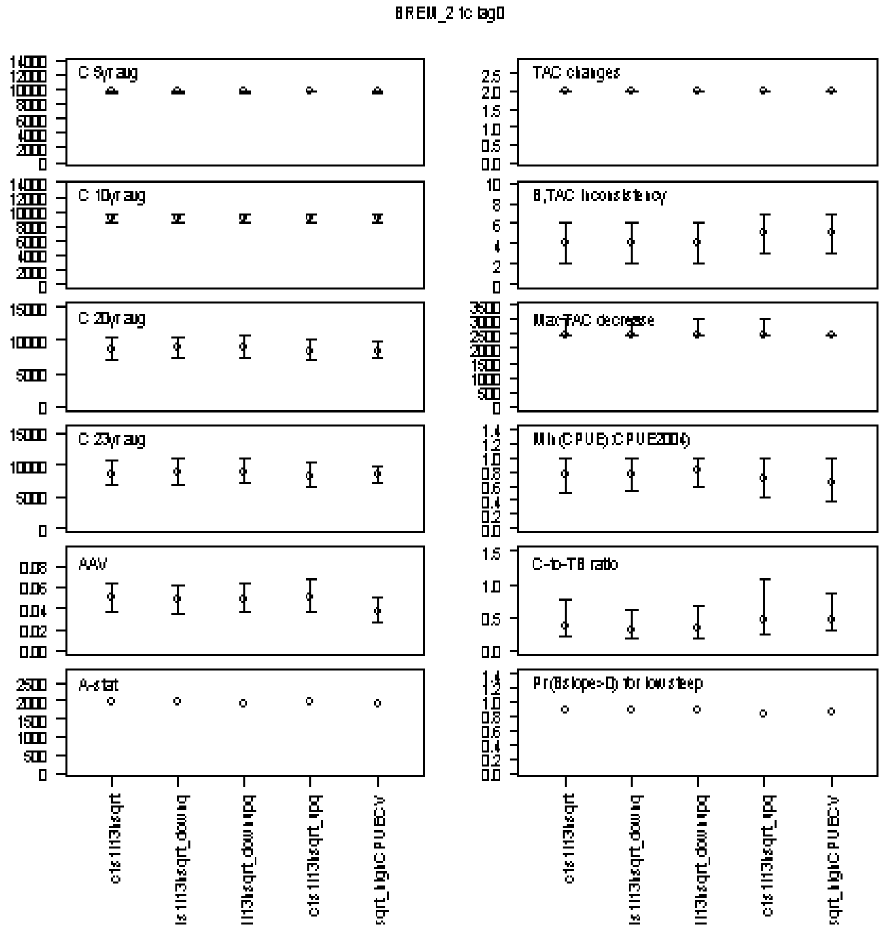

BREM_1 fc lag0
B2019:B2009
B2020:Bmsy
B2022:B2009
B2032:Bmsy
B2032:B2009
Pr(MinB<k*B2009); k=0.9,0.8
Min(By):B2009
Pr(B2032<k*B2009); k=0.9,0.8
B2020:B1980
Pr(B2022>.1*B0)
B2032:B1980
Pr(B2022>2*B2009))

Figure 14: Comparison of robustness trials on downq, downupq, upq and highCPUECV with the reference grid c1s1l1shsqrt of BREM_2 using the set of agreed performance indicators.



BREM_2 tc lag0

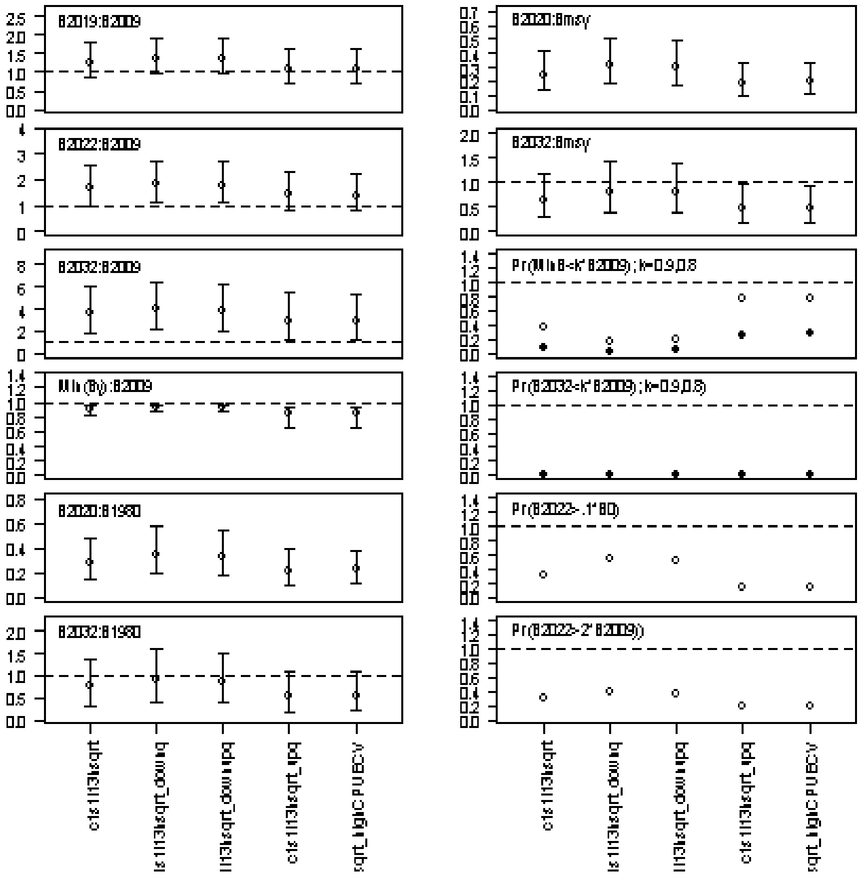
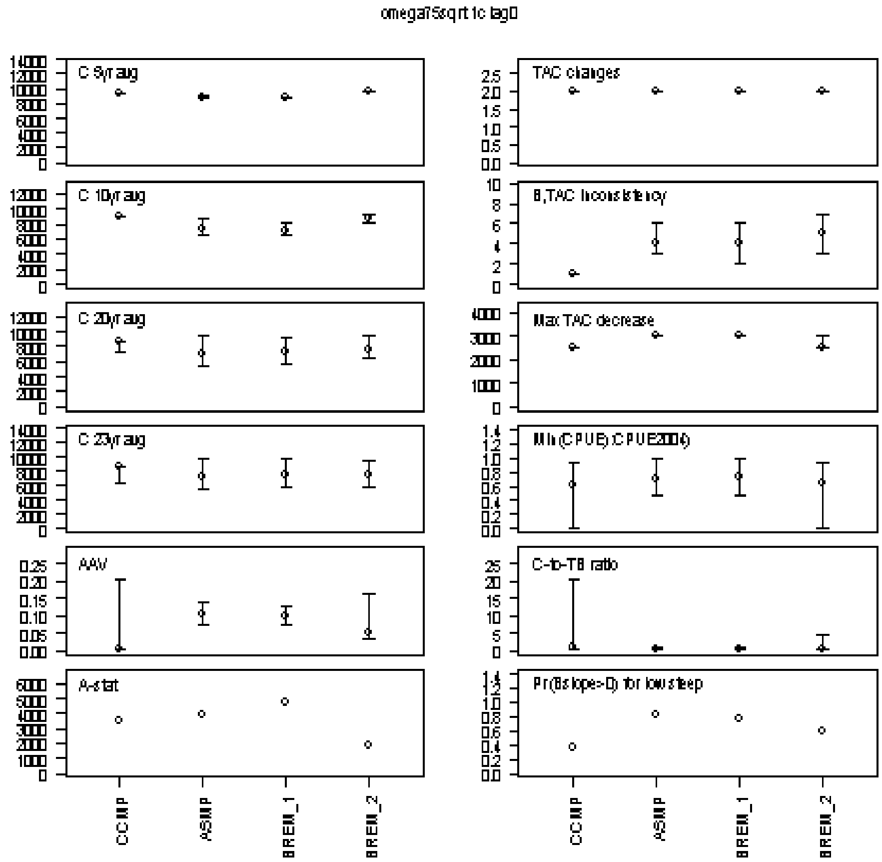


Figure 15: Comparison of the performance of CMPs on robustness trial omega75sqrt using the set of agreed performance indicators.



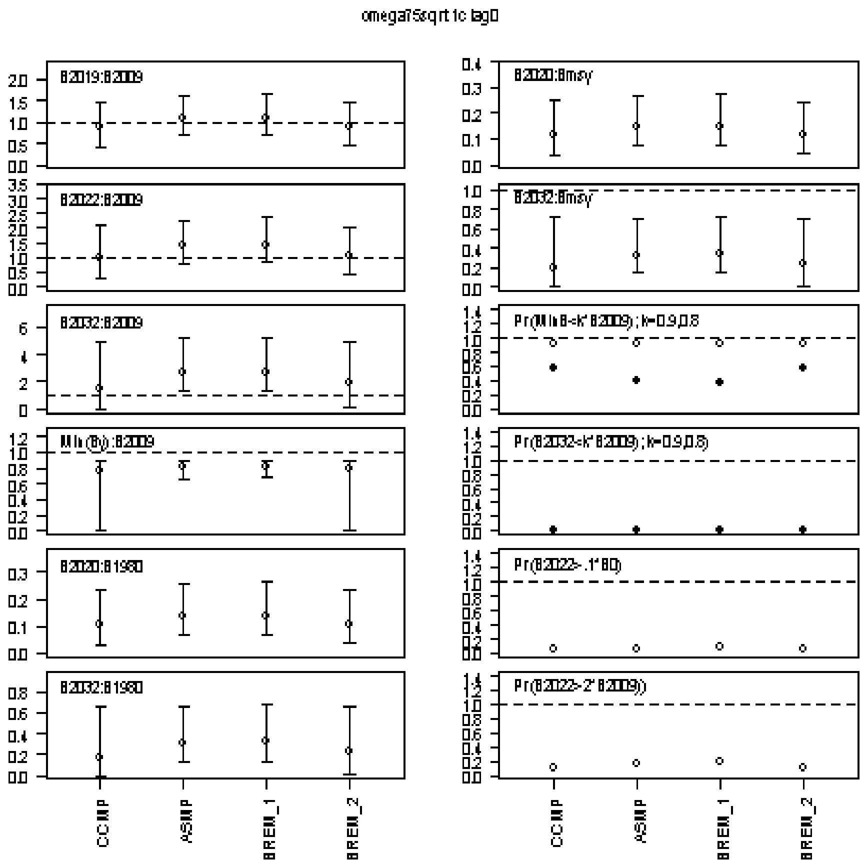
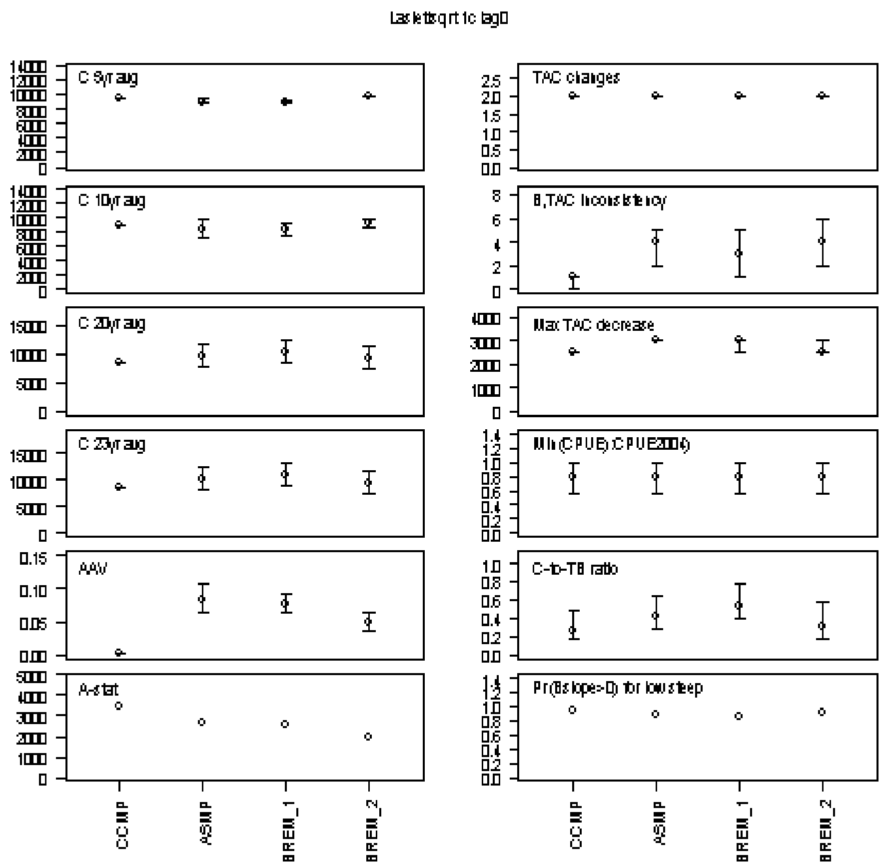
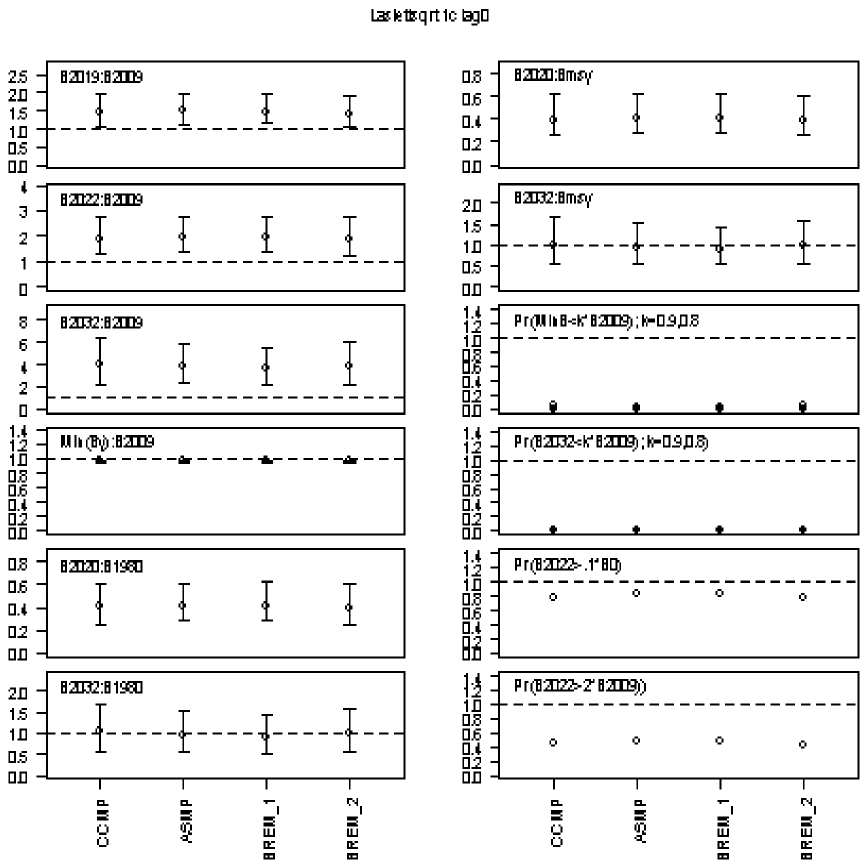

omega75sqrt1c lag0
B2019:B2009
B2022:B2009
B2032:B2009
Min(By):B2009
B2020:B1980
B2032:B1980
B2020:Bmsy
B2032:Bmsy
Pr(MinB<k*B2009); k=0.9,0.8
Pr(B2032<k*B2009); k=0.9,0.8)
Pr(B2022>.1*B0)
Pr(B2022>2*B2009))
CCMP
ASMP
BREM_1
BREM_2

Figure 16: Comparison of the performance of CMPs on robustness trial Laslettsqrt using the set of agreed performance indicators.




B2019:B2009
B2020:Bmsy
B2022:B2009
B2032:Bmsy
B2032:B2009
Pr(MinB<k*B2009); k=0.9,0.8
Min(By):B2009
Pr(B2032<k*B2009); k=0.9,0.8)
B2020:B1980
Pr(B2022>.1*B0)
B2032:B1980
Pr(B2022>2*B2009))
CCMP
ASMP
BREM_1
BREM_2