CSIRO
MARINE RESEARCH

FURTHER EXPLORATION AND EVALUATION OF THE FXR_01 CANDIDATE MANAGEMENT PROCEDURE RULE UNDER THE NEW REFERENCE AND ROBUSTNESS SETS

M.Basson, D. Kolody, J.Hartog, P.Eveson, T.Polacheck.

Prepared for the CCSBT 4th Management Procedure Workshop Meeting
16-21 May 2005, Canberra, Australia

Table of Contents

List of Figures

Abstract

The FXR_01 candidate decision rule and several variants were evaluated using the new reference set, robustness sets and tuning levels defined at the MP Technical meeting (Seattle, February 2005). Discussions at (and following) the third MP workshop on the behaviour of decision rules provided some guidance for characteristics that Industry and/or the Commission favoured. We used this guidance to develop and explore variants for the FXR_01 decision rule. Results for those variants that resulted in improved performance in terms of catch stability and risk are presented.

1. Introduction

The decision rule FXR_01 (Polacheck et.al. 2004, i.e. CCSBT_MP/0404/04 and its appendix) was tuned and evaluated with respect to the new reference set of SBT operating models. For a number of reasons, alternative formulations of the FXR decision rule warranted investigation.

Since the last round of MP development, experience with similar production model fitting procedures in an automated context revealed that the minimization was unstable and prone to frequent failures. Despite the fact that the MPs were not demonstrating obviously poor behaviour, this turned out to be a non-trivial problem with FXR in this round of MP testing as well (and was never really checked thoroughly in preceding rounds). An automated grid-search minimization was invoked to reduce this problem (described below).

Feedback from the third MP workshop and Industry provided some guidance on desirable performance of a decision rule. For example, early increases in catches were considered not to be desirable. We therefore considered several additional constraints on the changes in TACs as variations of the FXR_01 rule.

Concerns about multiple years of very low SBT recruitment starting around 2000 encouraged exploration of MP modifications within the age-aggregated Fox model framework that could potentially accommodate this age-structured problem. The hope is that such modifications might reduce the frequency with which meta-rules for dealing with exceptional circumstances need to be invoked.

After exploring FXR-based rules that take the above issues into consideration, we put forward 3 rules for wider scrutiny among members of the CCSBT MP community (SGF_303, CGF_01 and CGF_42 from Table 1, but note that we discuss SGF_306 as a proxy for CGF_42) at this fourth MP Workshop meeting. The selection was based primarily on the basis of median biomass tuning and low recruitment robustness tests. Electronic results including all tuning and robustness test trials will be available at the May CCSBT-MP meeting for consideration.

2. Methods

2.1 Review of the FXR_01 decision rule

This decision rule is based on the Fox production model which is fitted to the longline (LL1) CPUE biomass and total catch biomass. A 'preliminary' TAC (called pTAC) is calculated as the estimated *MSY* value times the ratio (B_y / B_{MSY}) times δ, where δ depends on the estimated *r* value (so that the TAC will be larger when the productivity of the stock is estimated to be high). The FXR_01 rule can be described mathematically by:

$$pTAC_{y+1} = \delta.MSY\left(\frac{B_y}{B_{MSY}}\right) \qquad (1)$$
$$\delta = r / r^*$$

where r^* is a tuning parameter of the decision rule and r is the estimated value of the parameter of the Fox model. Note that the above formulation is, of course, identical to $\delta.F_{MSY}.B_y$. The reason we refer to 'pTAC' is because subsequent to that calculation, the TAC is adjusted for the constraints on the maximum change in TAC from year to year (as determined at the MP Technical meeting (Seattle, February 2005)) and any other constraints imposed as variations (described below).

If $pTAC_{y+1} - TAC_y < -maxChange$ then $TAC_{y+1} = TAC_y - maxChange$

If $pTAC_{y+1} - TAC_y > maxChange$ then $TAC_{y+1} = TAC_y + maxChange$

2.2 Fox Minimization Reliability

The Fox model essentially estimates two parameters (*r* and *K*). We checked the convergence of the initial set-up in ADMB where both parameters were being estimated simultaneously. We found the default set-up to be poor for Cfull2, and often either failed to converge or seemingly successfully converged on a local minimum with implausible dynamics (e.g. fishing mortality really low and biomass essentially constant). We modified the implementation to a grid search over *r*, with *K* estimated for each *r* on the grid. The final phase of the minimization was initiated with both parameters free, starting from the best *r* value (though not *K*, purely due to the work-around manner in which the grid search was implemented in ADMB). Often this did not result in improved minimization over the grid estimates. This implementation seemed to eliminate the majority of dubious results. For the purpose of testing (i.e. to allow for a manageable run time), the grid resolution was coarse (*r* values of 0.05-0.95 at an interval of 0.05), with a corresponding pTAC resolution of around 2000 *t*) . If adopted, the grid resolution can, of course, be increased.

Even using the grid search method, the Fox model occasionally converged to minima that were arbitrarily close to a situation where the biomass approaches 0. This presumably relates to a fundamental limitation of the Fox model to describe SBT dynamics in some circumstances. The value of *pTAC* approaches 0 in these cases.

We also looked to see whether there were wide (or wild) fluctuations in *r* between estimation years (i.e. every 3 years or every 5 years), but this did not seem to be a problem.

Note that for convenience of managing the large numbers of runs, outputs and .tpl files, we renamed new versions that use the grid implementation to SGF or CGF instead of FXR (as indicated in Table 1).

2.3 Variations on FXR_01

Several potential variations of the original FXR_01 rule were explored, related to:

- the form of the pre-multiplier of the TAC (i.e. δ in equation 1 above)
- additional constraints on increases in TAC in the early years
- adjustment to TAC based on a LL1 recruitment index
- mandatory initial *pTAC* drop

We discuss these in sequence.

Note in the following that y is the year in which the MP calculation is conducted, the TAC is applied in y+1, using data up to y-2.

2.3.1 TAC proportional to r or r^2

During the early phases of evaluation, i.e. on the old 'reference set', we found that FXR, which uses a pre-multiplier of (r/r^*), performed better than the version which did not incorporate *r* into the pre-multiplier. Including *r* in the pre-multiplier means a factor of r^2 in the pTAC, given that $F_{MSY}=r/\ln K$ (see above) is also a function of the *r*-parameter. Preliminary evaluations with the new reference set, however, suggested that the r^2 factor may not be ideal. We therefore considered the following variation:

$$pTAC_{y+1} = \delta.MSY\left(\frac{B_y}{B_{MSY}}\right) \qquad (2)$$

where δ is a pre-multiplier, or tuning parameter, of the rule and NOT a function of *r*.

2.3.2 Constraints on increases in TAC in early years

We note that the original FXR_01 rule had no constraints on increases in the TAC other than the limits to maximum changes in TAC as used in the past. Feedback from the MP workshop and Industry suggested that rules which increased catches rapidly, particularly in the early years were not very desirable or ideal. Additionally, it is expected that constraints on early TAC increases would avoid some of the poor MP behaviour that was observed in previous iterations of the MP process, in which sensitivity to the operating model specifications caused some undesirable TAC recommendations in the early years (e.g. Kolody and Hartog 2004). We therefore considered additional constraints along the following lines (pTAC is the calculated TAC from equation 1 or 2 above):

- **Ceiling on TAC until 2015**: if(y<2015) TAC_y=min($pTAC_y$, TAC_{2004})
- **TAC increase constraint until 2015**: let the maximum increase in TAC be 2500 (instead of 5000) for 3-year blocks of TAC, or 4000 (instead of 8000) for 5 year blocks of TAC.

Table 1 below shows which rules contained which of the above constraints. All except SGF_101, which is the grid-optimized version of the original FXR_01 rule, have the ceiling on TAC until 2015.

2.3.3 Recruitment Feedback

In the light of concerns about recent poor recruitment, we also explored the performance of the FXR rule with additional modification to pTAC based on the mean proportion of age 4 in the LL1 CPUE. The proportion of age 4 can be thought of as a type of 'recruitment index' (although the signal is obscured due to changes in selectivity and overall abundance). Based on the range of values of this index under the default FXR rule and the reference operating model set (Cfull2), discrete cut-off points were chosen and the following reductions in pTAC applied (for the 3 year TAC setting scenario; 2500 is replaced by 4000 in the 5 year scenarios) :

- If recIndex(y-4:y-2)<0.1 drop pTAC by 2500
- If recIndex(y-4:y-2)<0.05 drop pTAC by 2500
- For y > 2009: If recIndex(y-7:y-5)<0.1 drop pTAC by 2500
- For y > 2009: If recIndex(y-7:y-5)<0.05 drop pTAC by 2500

where recIndex(y-4:y-2) is the average of the recruitment index over the years indicated.

All of the potential TAC drops above are applied concurrently, so that if recIndex(y-4:y-2)<0.05 and recIndex(y-7:y-5)<0.05, the drop in pTAC is 10000. However, the original constraints on TAC changes are always applied last, so that the actual TAC decrease can never exceed the specified maximum allowed change. This implementation is referred to as the *discrete* version of recruitment feedback.

We later also implemented a *continuous* version of recruitment feedback (CFG_01). The intent was to present a rule in which TAC changes are not highly sensitive to small changes in the recIndex. Instead of step function drops in TAC related to discrete recruitment thresholds, there is a linear drop in TAC as the recruitment index drops below the threshold:

- If recIndex(y-4:y-2) < 0.125 TAC = pTAC - 10*(0.125- recIndex(y-4:y-2))*maxChange;
- If (y > 2009) & (recIndex(y-5:y-7)<0.125)) TAC = pTAC - 10*(0.125 - recIndex(y-7:y-5))*maxChange

Both the potential TAC modifications are applied concurrently, but as in the discrete case, the original constraints on TAC changes are always applied last so that the actual TAC decrease can never exceed the maximum allowed change. The *continuous* version was only applied to what, at the time, appeared to be the best of the discrete recruitment feedback MPs (i.e. otherwise identical to SGF_304). Two alternative recruitment feedback slopes and reference points were also examined (not shown), but did not seem to be much different to CGF_01. CGF_42 from Table 1 was a late addition that we did not run in time to include in the document but expect will be similar in performance to SGF_306. This is confirmed at the end of the document.

2.3.4 'Initial Drop' - a time-dependent, externally-imposed pTAC modification

A time-dependent, externally-imposed pTAC modification was explored that can be used to influence the pTAC calculation for an arbitrary number of years, to accommodate our current perceptions about the status of the stock if we don't think that the next couple years of data will be sufficient to appropriately influence the Fox model stock status estimates. As tested here, this consisted of:

- if(y<2015) drop pTAC by 2500

This was only invoked in one rule that we present here (SGF_305).

TABLE 1. Modifications to FXR_01

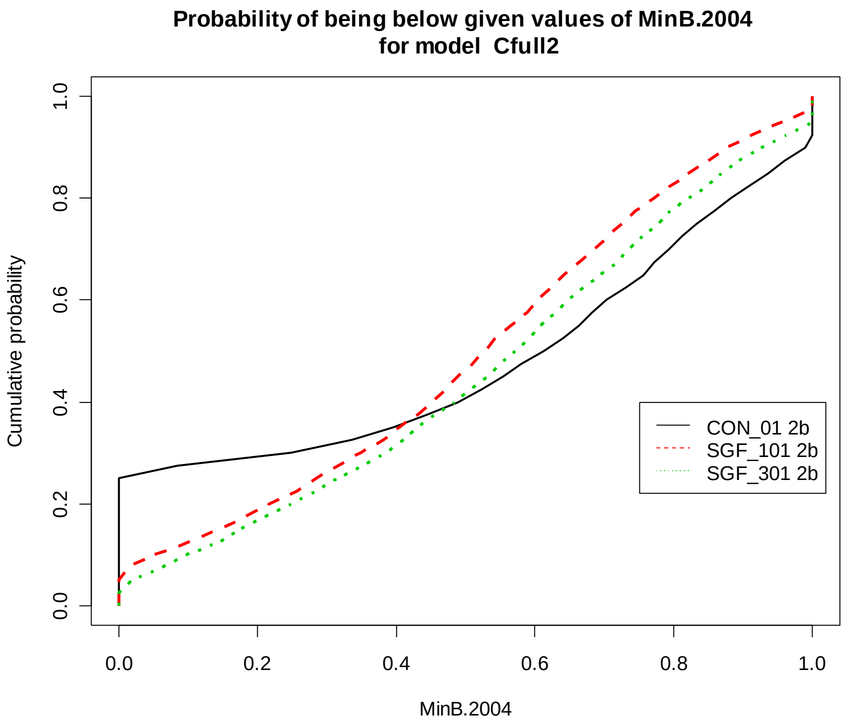
	pTAC proportional to r or r^2	2004 TAC ceiling until 2015[1]	TAC increase constraint until 2015(footnote1) (proportion of maxChange)	Poor Recruitment Feedback	Initial TAC Drop (proportion of maxChange)
FXR_01	r^2				
SGF_101 (Grid search version of FXR_01)	r^2				
SGF_301	r				
SGF_302	r	Yes			
SGF_303	r	Yes	0.5		
SGF_304	r	Yes	0.5	Discrete	
SGF_305	r	Yes	0.5	Discrete	0.5
SGF_306 (proxy for CGF_42)	r	Yes		Discrete	
CGF_01	r	Yes	0.5	Continuous	
CGF_42 (results not shown in full)	r	Yes		Continuous	

3. Results

3.1 r versus r^2 in the pre-multiplier

A comparison between SGF_101 (equivalent to FXR_01) and SGF_301 which is based on *r* instead of r^2, shows that SGF_301 has lower risk in terms of the minimum biomass (MinB/2004,Figure 1), and slightly lower risk in terms of B2022/B2004 (not shown). The trade-off between the two is in lower catches (SGF_301) in the short term but higher catches in the medium to longer term (Figure 2). Also, the lower 10^{th} %-ile of biomass doesn't drop as low for SGF_301 as it does for SGF_101. For biomass in 2032, however, SGF_101 performs slightly better. (We note that results in terms of medians and 10^{th} & 90^{th} percentiles for SGF_101 which uses the grid implementation of minimisation were almost identical to the original FXR_01 under the new reference set.)

[1] The constraint applies for the first 2 or 3 TAC-changes for 3-year blocks of TAC depending on whether the first change occurs in 2006 or 2008. For 5-year blocks of TAC starting in 2008, the constraint only applies to the first change.



1

Figure 1. Cumulative probability plot of MinB/B2004 for SGF_301(r) and SGF_101 (r^2).

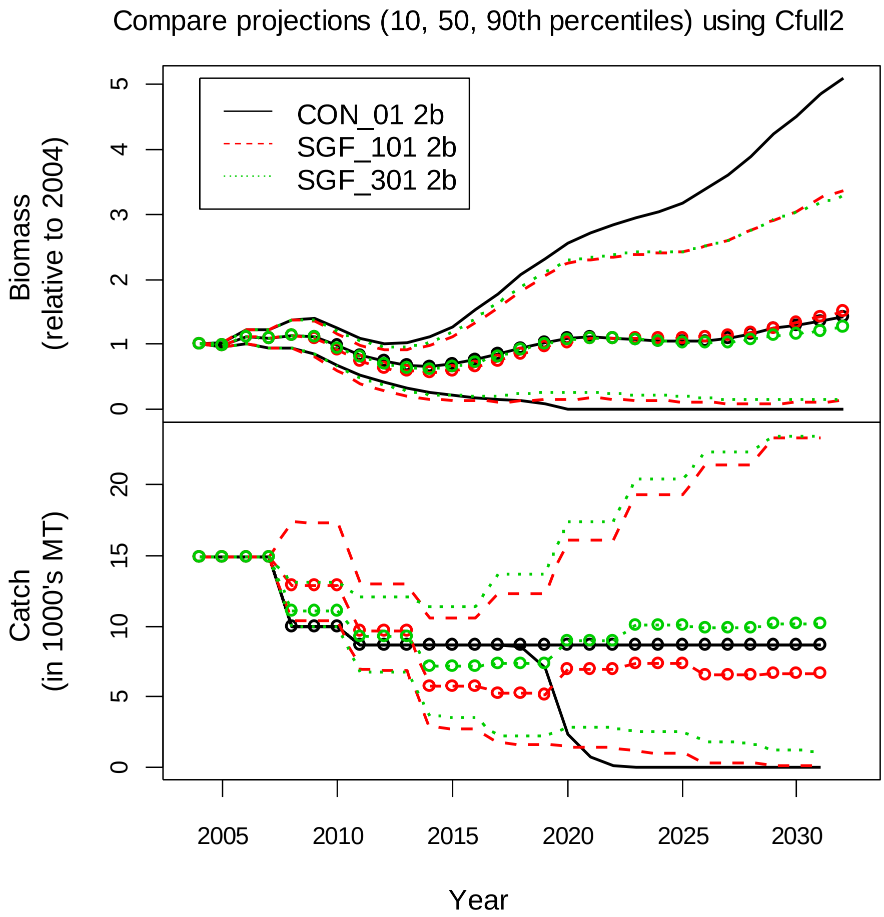


Figure 2. Biomass and catch performance for constant catch, and SGF_101 and SGF_301 rules tuned to 1.1

3.2 Ceiling on TAC until 2015

A comparison of SGF_301 and SGF_302 are almost identical in terms of performance and risk under the 2b tuning option. This is the case for 1b and 3b tunings too. The only difference between them is that SGF_302 has a ceiling on TAC until 2015, and this only affects a small number of trajectories. Therefore, performance in terms of medians and 10^{th} & 90^{th} percentiles is quite similar between the two rules (Figure 3). We have therefore kept the ceiling on the TAC in the subsequent rules noting that this was driven by feedback from industry regarding catch increases in the short-medium term. (We consider it inappropriate to prevent any increase in TAC in the long term). We would expect more of a difference between the two rules in the 'optimistic' robustness trials.

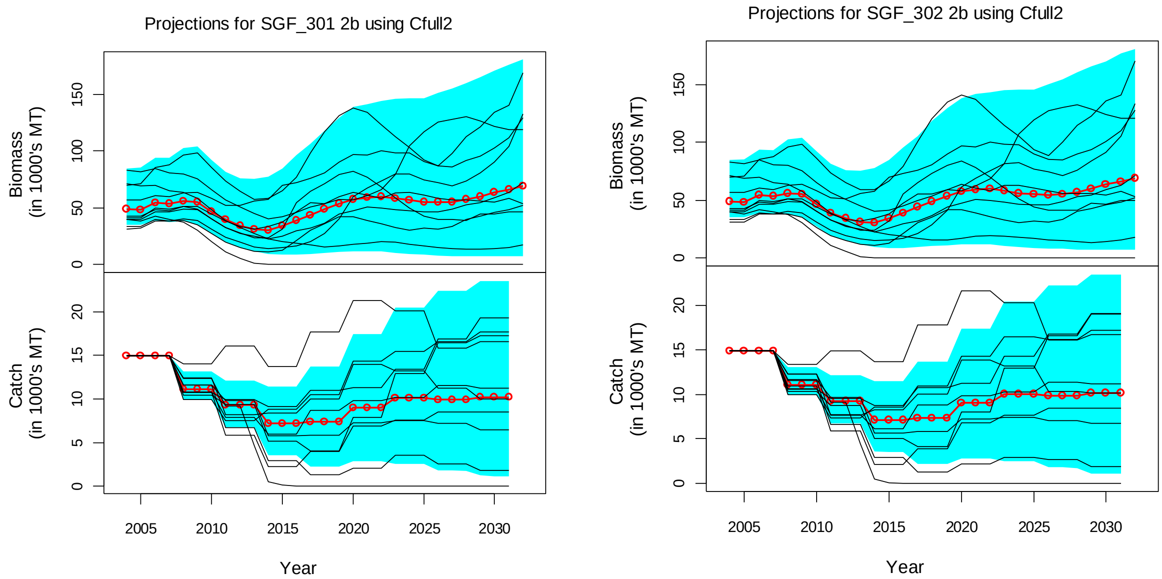


Figure 3. Biomass and catch trajectories for SGF_301 and SGF_302 which has a ceiling on TAC until 2015.

3.3 TAC increase constraint until 2015

The effect of a further constraint on the maximum increase in TAC allowed until 2015 is not discernable in the summary plots such as Figure 4 (SGF_302 without and SGF_303 with increase constraint). This is because there are very few trajectories which in fact want to increase the TAC by a large amount based on this decision rule. Although one would again expect to see more of a difference between the rules in the 'optimistic' robustness trials, the differences are still very small (e.g. <50t difference in 90^{th} percentile of average 10 year catch, and <100t in 90^{th} %-ile of average 20 year catch) and not discernable on a figure.

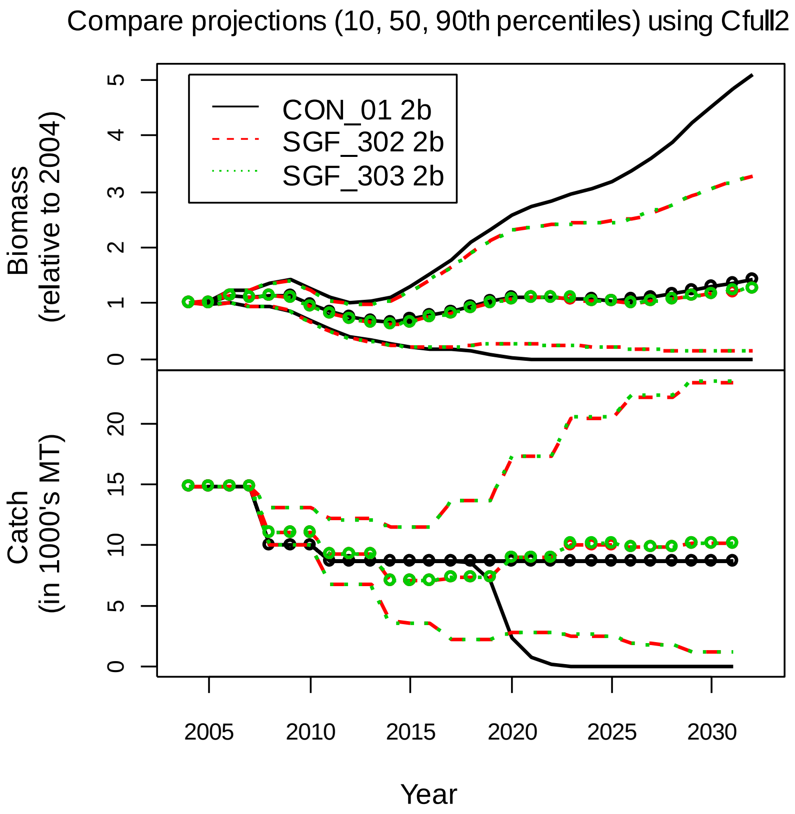


Figure 4. Biomass and catch trajectories for SGF_302 (without TAC increase constraint) and SGF_303 (with TAC increase constraint) tuned to the 1.1.

3.4 Recruitment feedback

In terms of median projected biomass there is not a great deal of difference between MPs with recruitment feedback (SGF_304) and similar rules without (SGF_303) (Figure 5). This is not surprising as the recruitment feedback would be expected to come into play and affect the scenarios that were not recovering (i.e. affect the lower percentiles and not the mean or median). However, note the narrower inter-percentile range for SGF_304, which indicates that the feedback is helping to mediate the risk for the most pessimistic scenarios. Although median SSB in the long term is slightly less for SGF_304 than SGF_303, the main differences lie in the risk to biomass and in trade-off between catch and biomass. Risk to biomass in terms of MinB/B2004 is lower for SGF_304 (Figure 6). The trade-off clearly lies in the lower catches in the short term, though in the longer term, SGF_304 achieves higher catches (see Figure 10 below). Results of robustness trials are discussed below.

The differences among rules that use recruitment feedback but have different constraints on TACs prior to 2015, are even smaller (Figure 7 and Figure 8). SGF_304 (upward TAC change constraint) is slightly more risky than the CGF_01 (continuous version of the recruitment feedback plus upward TAC change constraint) and SGF_306 (no upward TAC change constraint). Summary performance of the latter two are almost indistinguishable, and the continuous version (CGF_01) would presumably be preferable in an actual implementation from several perspectives (e.g. smooth rather then a step-function change in TAC thus preventing the magnitude of the TAC change to be highly sensitive to small differences in data).

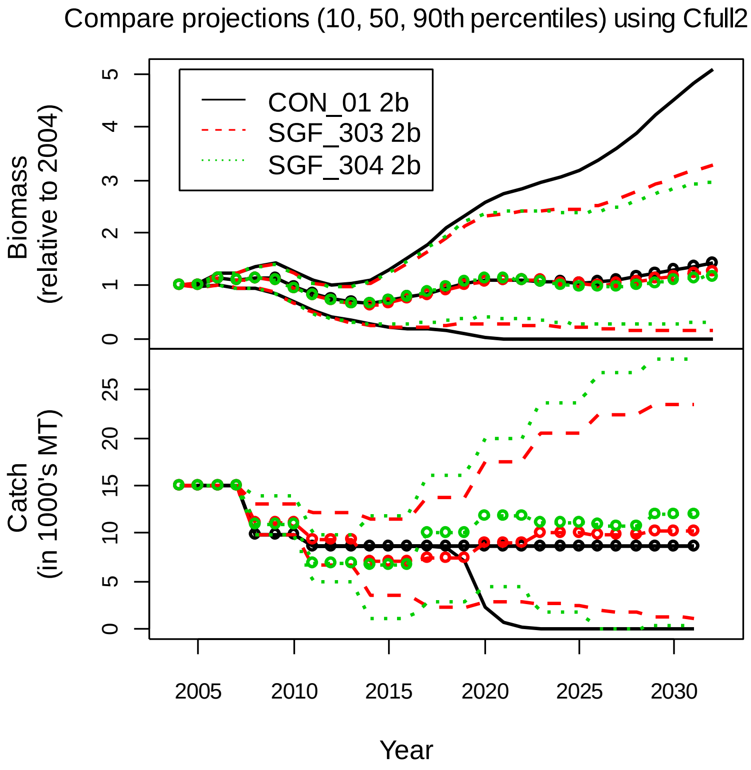


Figure 5. Comparison of SGF_303 (without recruitment index) and SGF_304 (with recruitment index) in terms of projected catch and biomass.

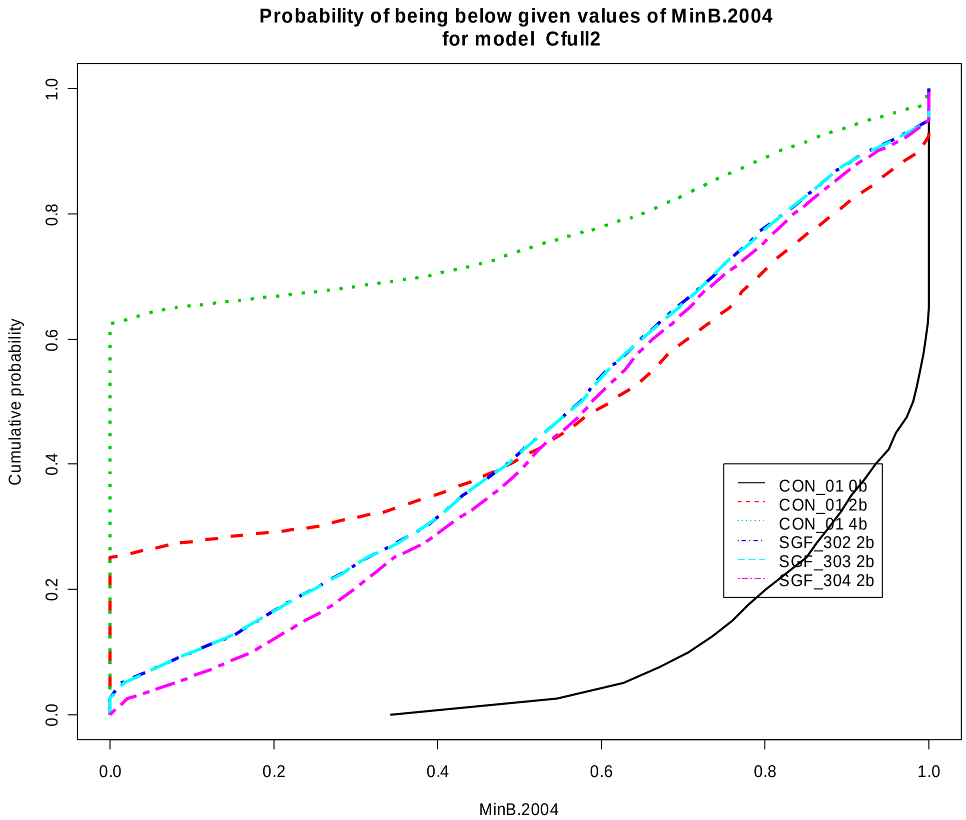


Figure 6. Cumulative probability plot for MinB/B2004. For Constant catch: 0b= no catch, 2b is tuned to 1.1, 4b=current catch. SGF_304 has recruitment feedback; the others in this figure do not.

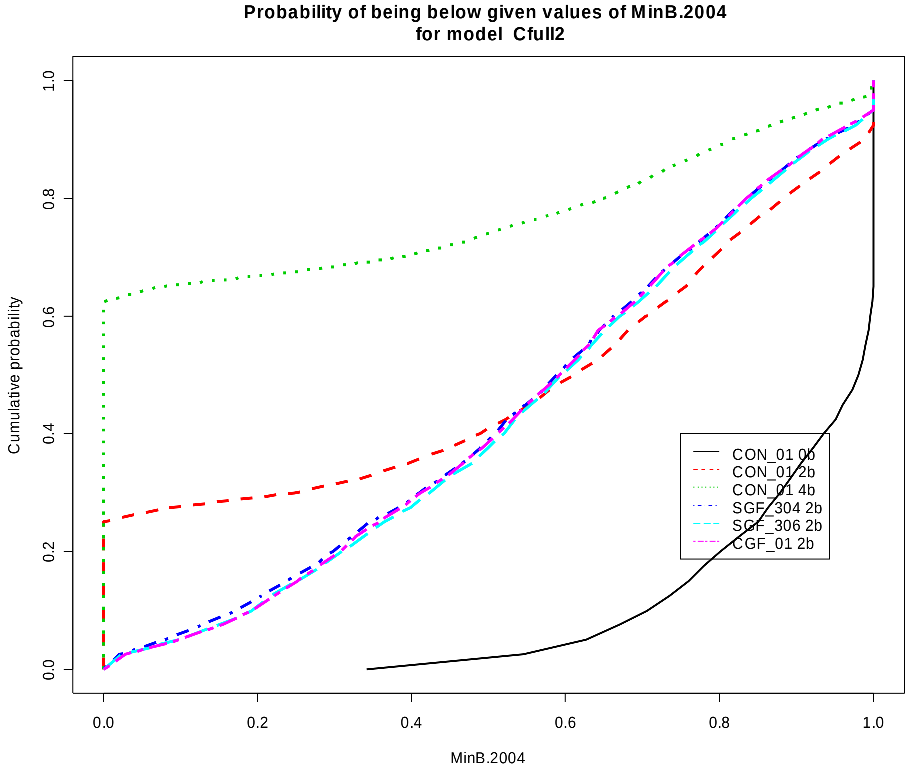


Figure 7. Cumulative probability plot for MinB/B2004 with rules using recruitment feedback. SGF_306 does NOT have the TAC increase constraint (See Table 1).

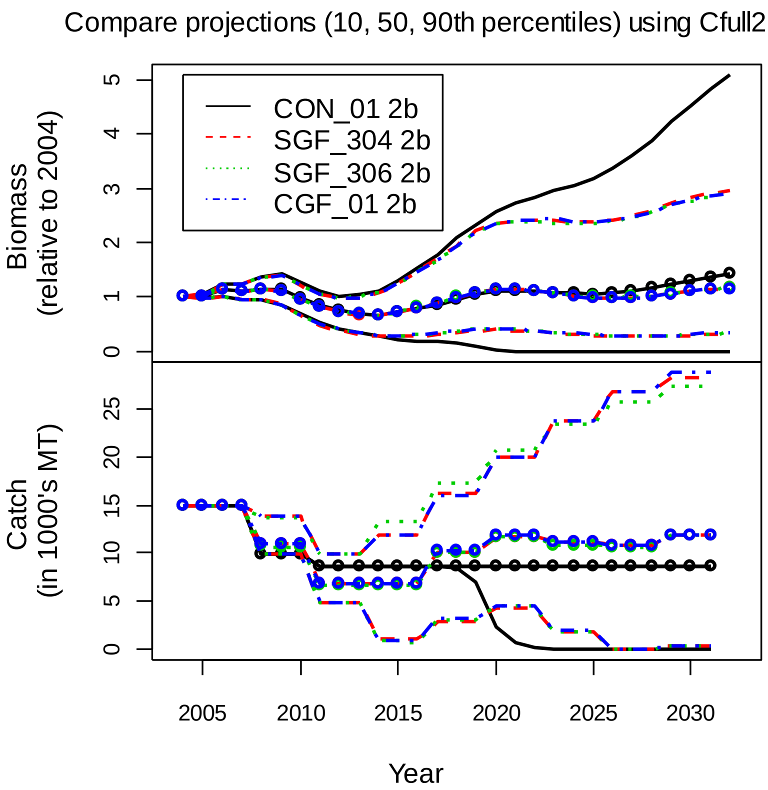


Figure 8. Comparison of biomass and catch trajectories for 3 rules and Con_01. SGF_306 does NOT have the TAC increase constraint (See Table 1).

3.5 Drop in TAC until 2015

Rule SGF_305 is based on SGF_304 but with the addition of a forced drop in the calculated TAC until 2015 (see text above and Table1). This rule did not perform as well as SGF_304 in our view because it allowed biomass in the longer term to decline and remain below the tuning level (Figure 9). This perhaps somewhat counter-intuitive longer term performance is a consequence of having adopting the median SSB as value to tune to (see discussion below). This poor behaviour is also evident in the trade-off plots for the short medium and long term (Figure 10). However, we do note from other trials (not shown) that the imposed initial drop can have a very similar effect to the recruitment feedback, but in this case, it seems counter-productive to add them both simultaneously.

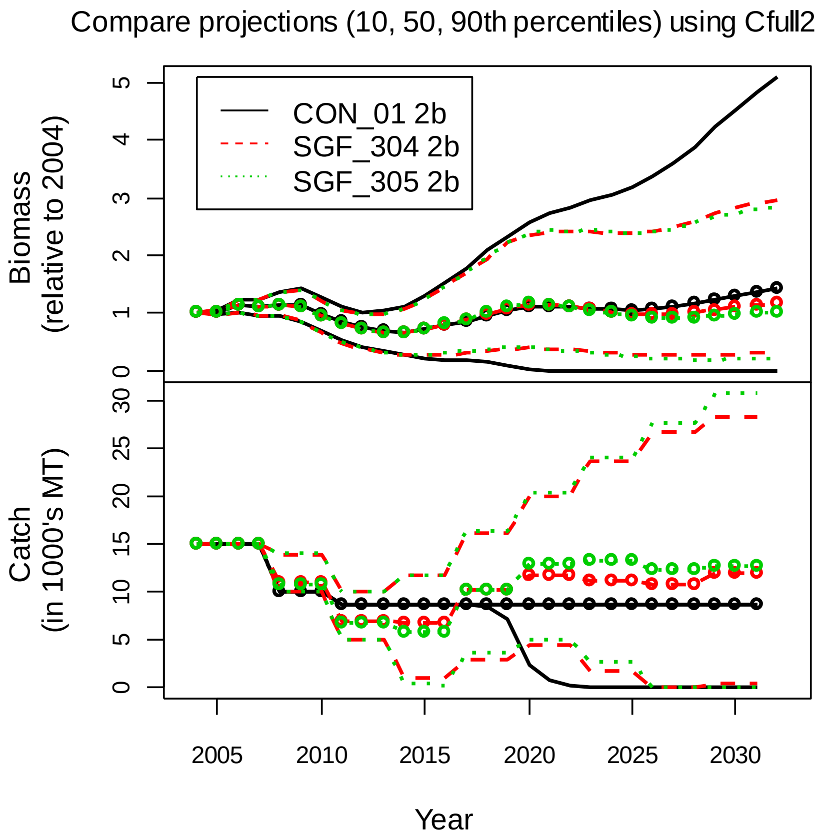
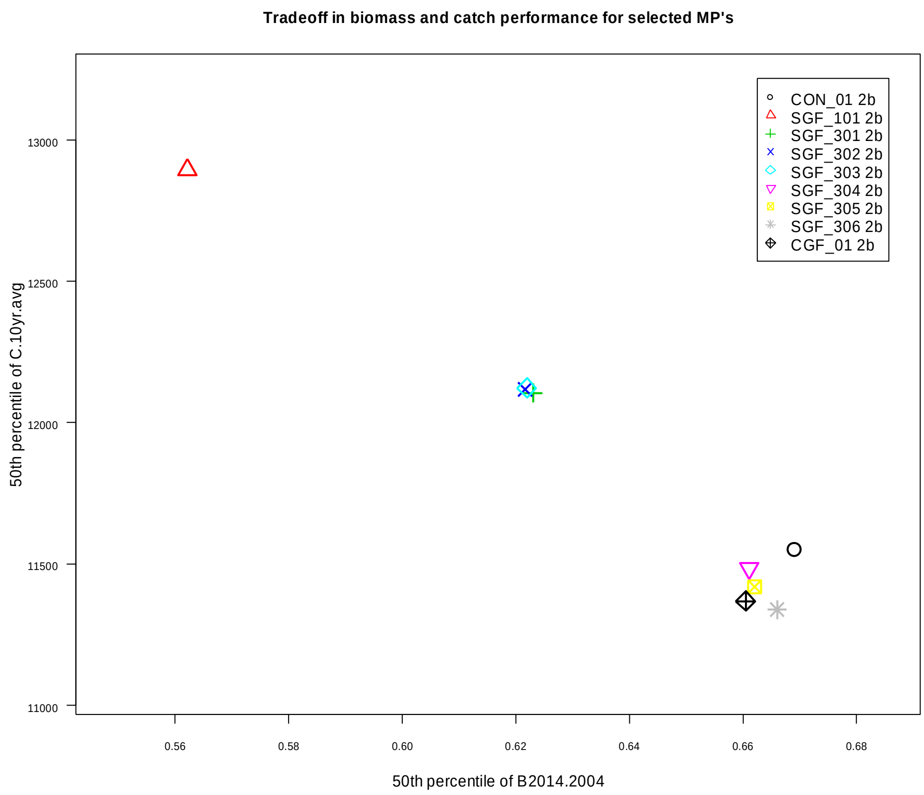


Figure 9. Comparison of rules with (SGF_305) and without (SGF_304) the initial TAC drop (2500 t to 2015).

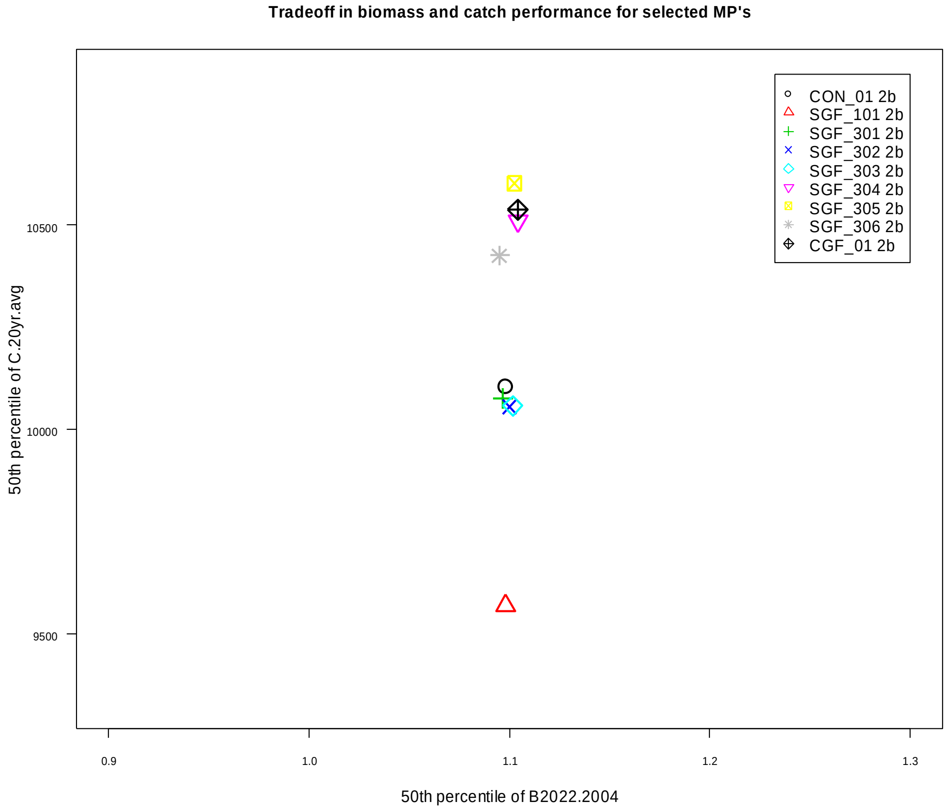
Trade-off figures also show the relative performance of the suite of rules we considered, with respect to median biomass and catch. The three trade-off panels (Figure 10) show how the short term vs long term performance switches between rules. For example, SGF_101 (red triangle) has high catch, low biomass performance in the short term, but then shows LOW catch, high biomass performance in the long term. It also shows the similarity in performance between SGF_301,302 and 303 (rules without recruitment feedback) and similarity between SGF_304, 306 and CGF_01 (rules WITH recruitment feedback) as already pointed out. SGF_305 has the lowest biomass in the long term; median B2032 is around the B2004 level suggesting a decline from the tuning level of 1.1 after 2022. We consider this not to be desirable behaviour. Although median long term biomass, B2032/B2004, is lower for the rules with recruitment feedback, we note that these rules show better performance than those without recruitment feedback in terms of the lower 10th percentile of B2032/B2004 (Figure 11).

A comparison of the performance statistics (Figure 12) also shows just how similar the rules are. There are two broad groupings seen in the results; those rules with recruitment feedback (SGF_304,305,306 and CGF_01) and those without (SGF_101,301,302,303). The recruitment feedback rules will favour decreasing TACs by the maximum amount allowed, while those without recruitment feedback have a little more flexibility in the magnitude of TAC reductions. Average catches are higher in the long term for the recruitment feedback rules, but are lower in the short to mid-term. The Biomass/TAC inconsistency is lower for the recruitment feedback rules.

Note that the AAV statistic becomes very large when the TAC is dropped to zero (or close to zero) and then increased after that, because the denominator is C_y+1e^{-6}. This needs to be considered when comparing performance of rules, and unfortunately makes this statistic difficult to interpret. This is particularly relevant with the current reference set in which stock sizes decline to very low levels in a substantial fraction of the scenarios, and where reduction to low catch levels would be appropriate.



(Figure 10a – see below for caption)



(Figure 10b – see below for caption)

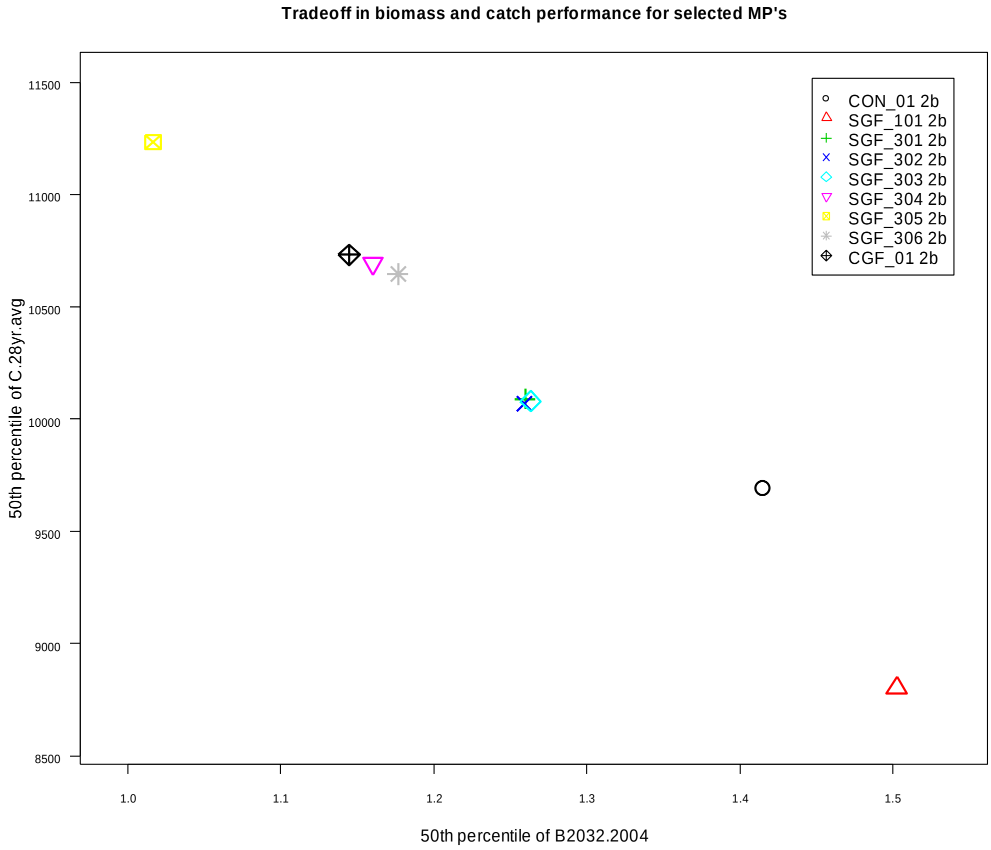


Figure 10. Three trade-off figures for the short (top panel), medium (middle panel) and long (bottom panel) term for median biomass and catch. The middle panel reflects tuning to 1.1

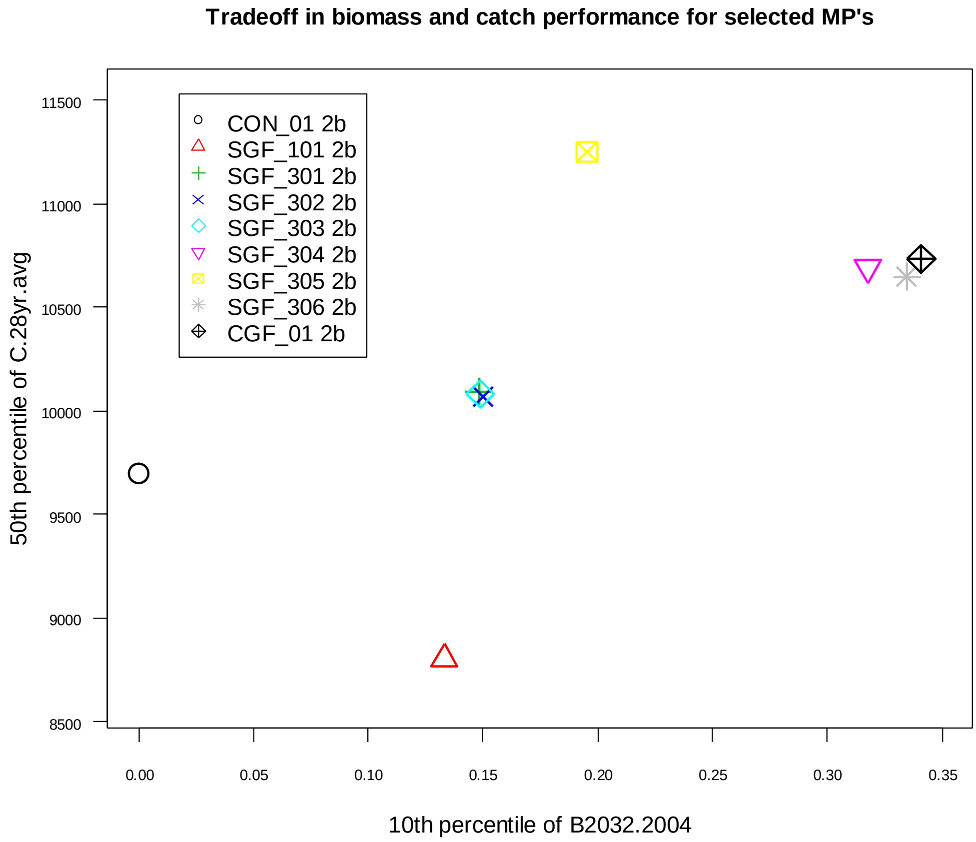


Figure 11. Trade-off in terms of the 10th percentile of B2032/B2004 and median catch for all rules.

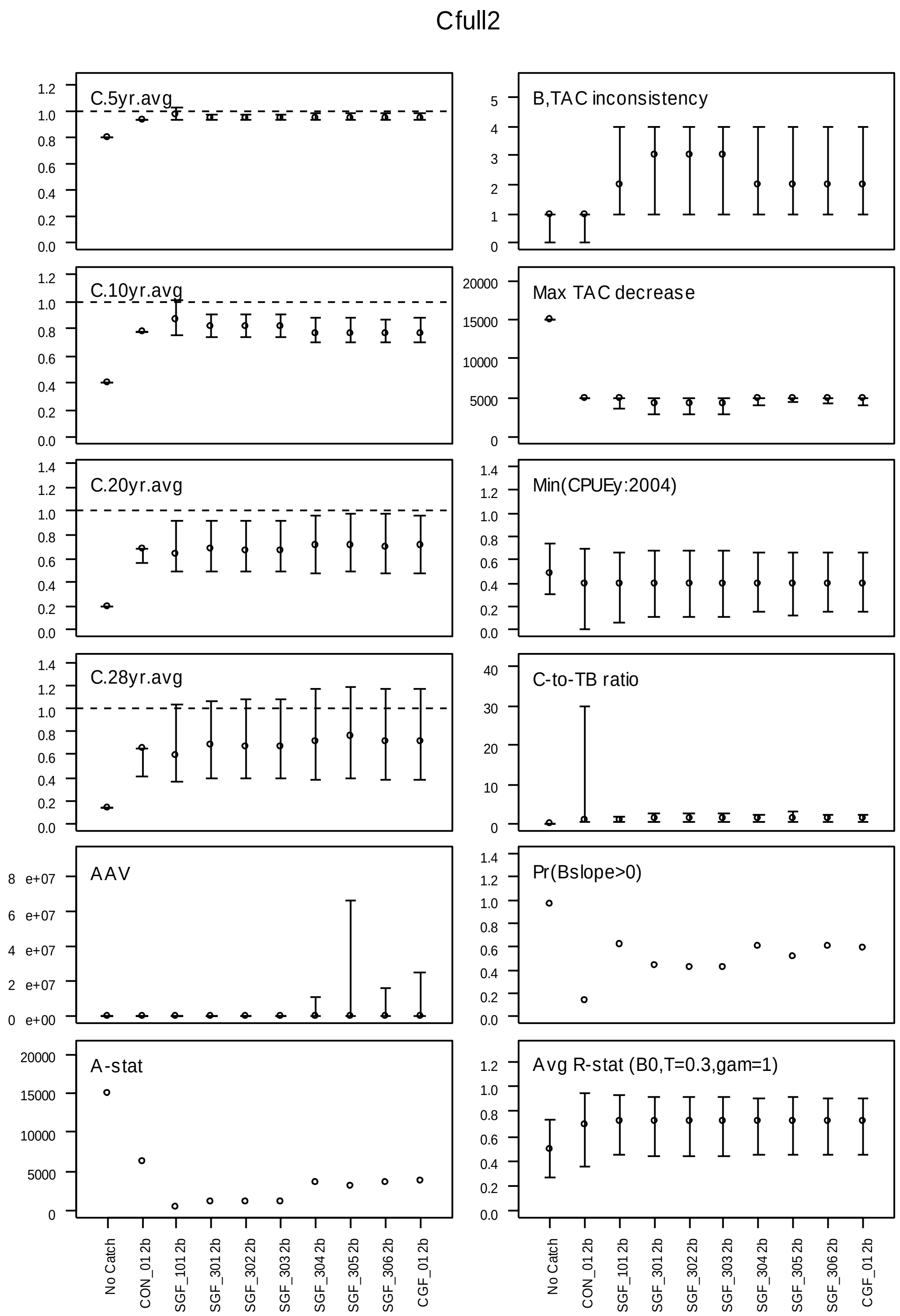


Figure 12 Catch performance measures for all MPs in Table 1.

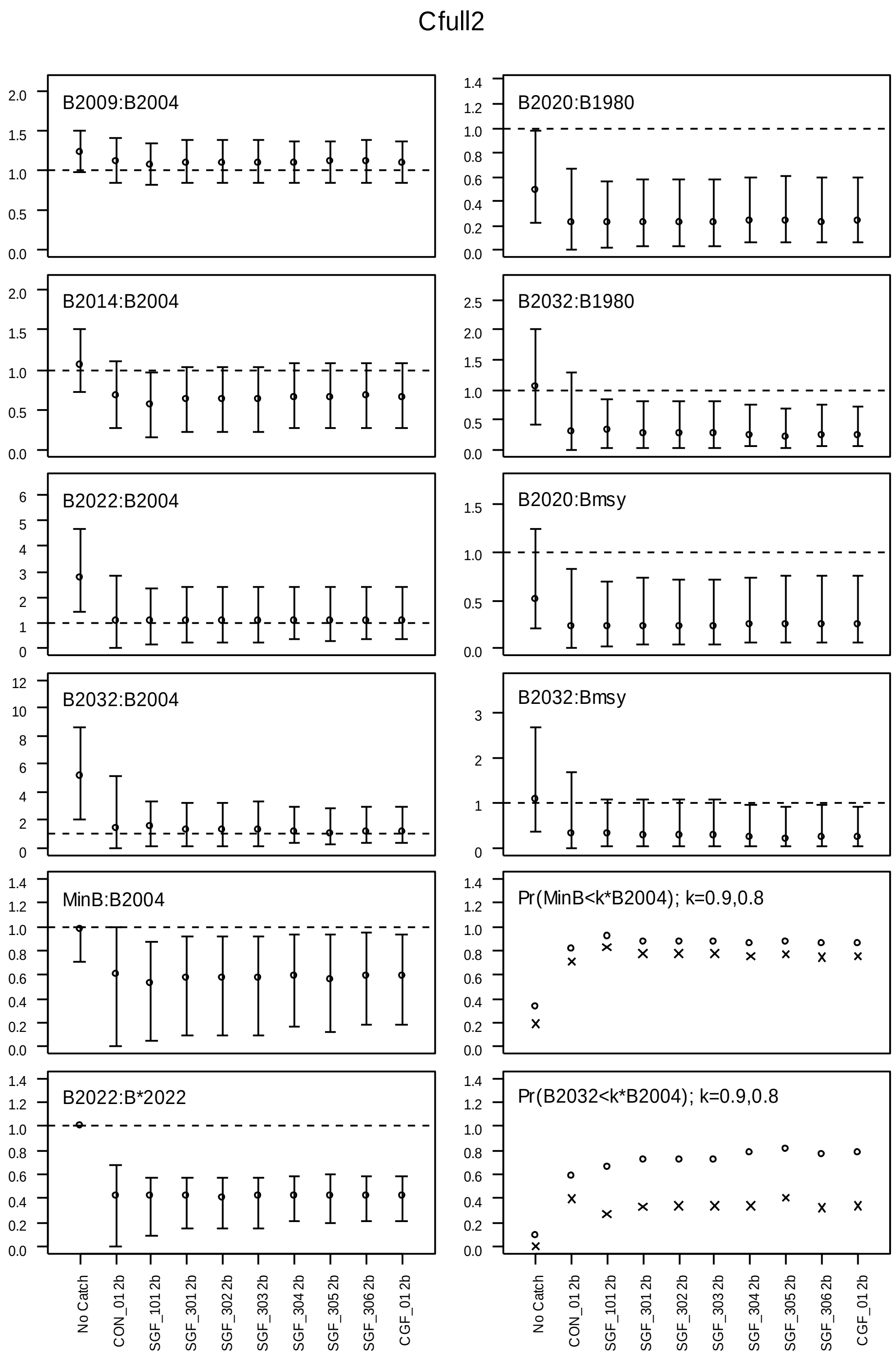


Figure 13 Biomass performance measures for all MPs in Table 1.

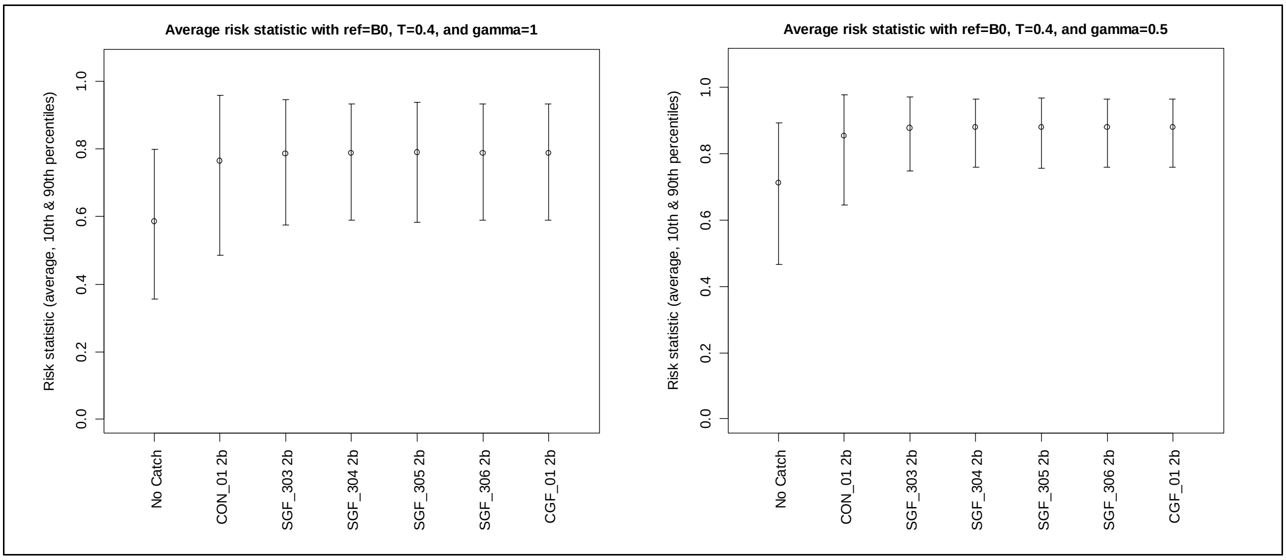
3.6 Comments on the risk statistic and its parameters

The new performance measure proposed at the 3rd meeting of the MP Workshop is the so-called "Average risk statistic" (for SSB). Three quantities need to be specified: the measure of biomass (e.g. B_0, B1980, B_{msy}), the threshold, T (a proportion of the measure of biomass) and gamma, the power to which the difference between the observed ratio and the threshold is raised. The notion was that a gamma value less than 1 would increase the contribution of situations where the biomass is well below the threshold, compared to when it is only slightly below the threshold. The Technical meeting of the MP Workshop (February 2005) encouraged members to explore different settings / parameter values for the risk statistic.

In the context of the MP evaluations, there are two potential uses for a statistic of this kind. One use is to quantify risk for comparison between different tuning levels (and the no-catch scenario which should give the lowest risk). A second potential use is to distinguish between the performance of different rules (tuned to the same level). With regard to the first use, we consider that B_0 is preferable to B_{msy} as a measure, because of our concerns about the lack of fit of the recruitment function, and the fact that B_{msy} is a function of the assumed stock-recruit relationship and steepness. We also consider that values such as 0.2 – 0.4 for T are reasonable in the light of the work of e.g. Mace (1994). We note that for a longer-lived, late maturing, species such as SBT, and given our priors on steepness, values of 0.3 – 0.4 may be more appropriate than 0.2 if the statistic is to be used for more than just a comparison between rules. B1980 would also be an appropriate measure given that the Commission's current management objective is framed in terms of B1980 and discussion in Polacheck (2003) with respect to the 1980 biomass as being an appropriate empirically based rebuilding target (e.g. the large set suite of indicators and assessment results that suggest that the basic population and habitat dynamics of the stock were being disrupted around this point). For the B1980 target, the appropriate threshold in this case would be 1.

A choice of gamma is not immediately apparent, and we explored the effect of gamma on the risk statistic for a two levels of the threshold wrt B0. Figure 14 shows that the higher threshold value (0.4 - upper 2 panels) compresses the risk bars (10^{th} – 90^{th} percentile) and increases the average statistic. Lower values of gamma compresses the interpercentile bars even more (compare left and right hand panels) and also increases the average (NB average, not median plotted in these figures). The fact there is no objective way of choosing gamma and the somewhat arbitrary basis of selecting both a target and threshold makes the interpretation of the statistic as an absolute measure of risk problematic.

Figure 15 also illustrates that the value of gamma does not affect the pattern of risk over different tuning levels.



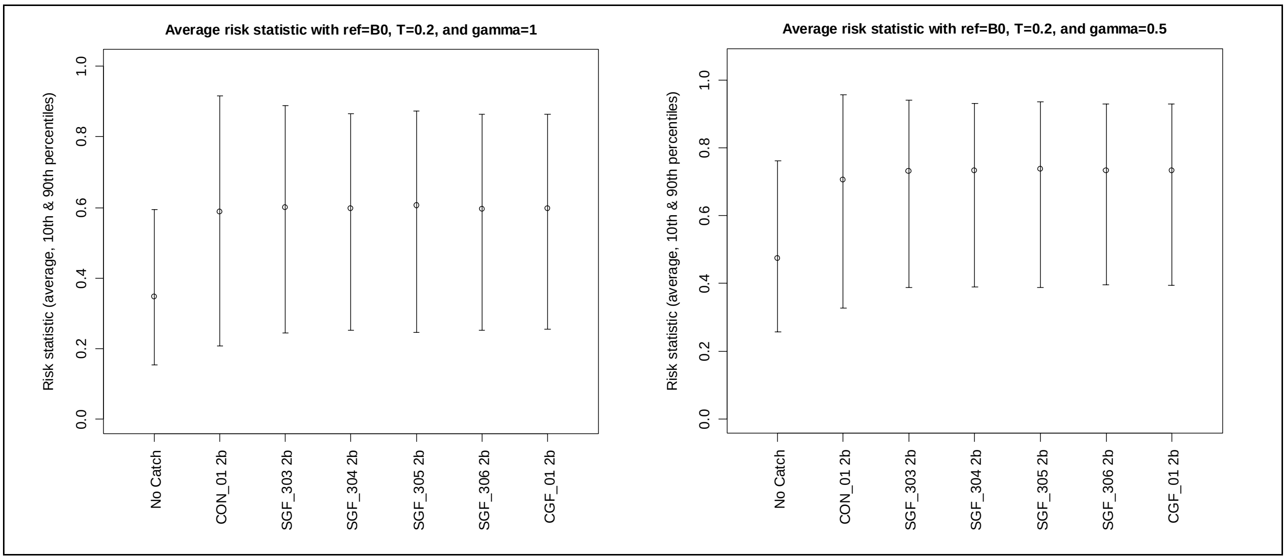


Figure 14 Values of the risk statistic for several rules and different values of the threshold, T, and the power parameter, gamma.

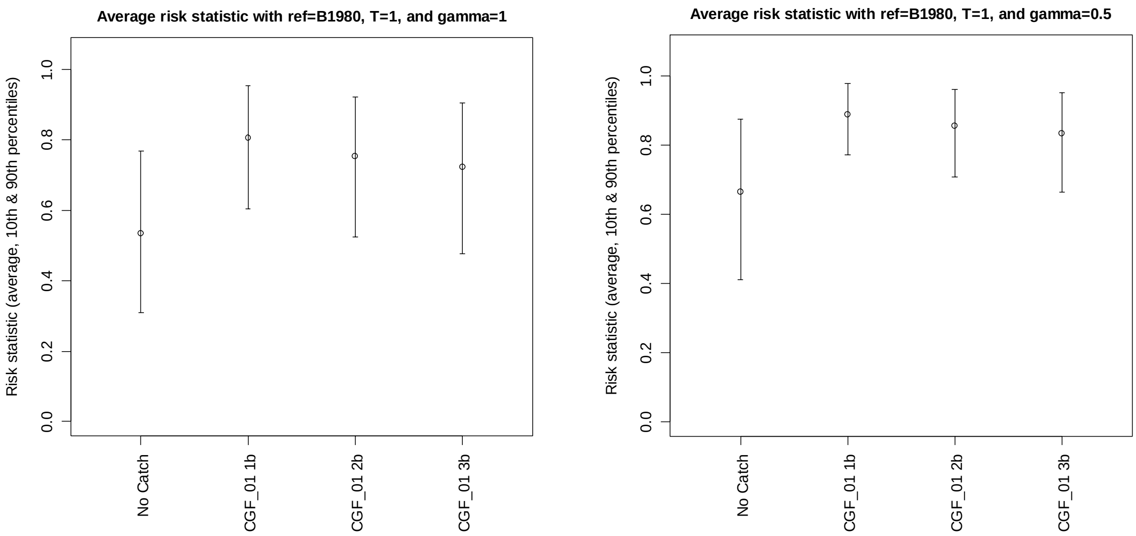


Figure 15. Risk statistic for no catch scenario and one rule tuned to the three levels (1b, 2b and 3b); left hand panel has gamma=1, right hand panel gamma=0.5.

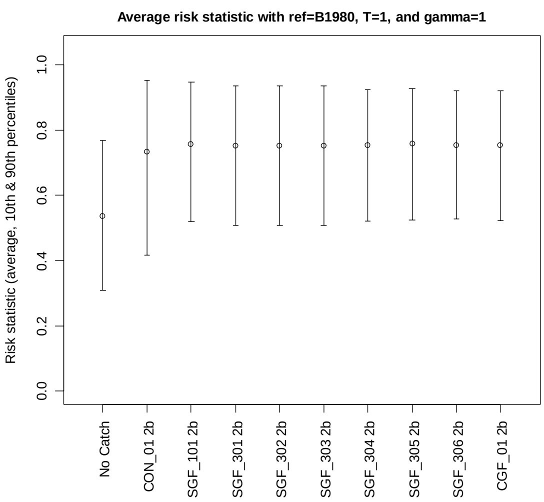


Figure 16. Risk statistic in terms of B1980 for all the rules in Table 1, and the 'no catch' scenario (2b tuning).

It is, however, informative to note that in terms of B1980, even the no catch scenario had a reasonably high average risk statistic (Figure 16). The fact that constant catch ‘rule’ (CON_01, tuned to 1.1 as the other rules in the figure) has a lower average and 10th percentile is, however, potentially misleading in the light of the risk measured in other ways, e.g. the probability of falling below B2004 (MinB.2004, e.g. Figure 1,Figure 7).

With respect to the second potential use of the statistic, the above figures show very little difference between the rules and are not at all useful for distinguishing between rules. This may be in part a function of the operating model and the current stock status. Thus, based on the fact that the stock is in a “high” risk situation, the MPs have minimal ability to ameliorate this at least within the constraints of tuning to 1.1 biomass level (e.g. for “reasonable” targets and thresholds many scenarios would still be judged to be in a risky state even if they achieved the tuning objective). We found the cumulative probability plots with respect to a range of biomass performance measures to be more informative. The statistic only becomes useful as a distinguishing measure when the threshold is chosen to be very low so that the upper percentile, and to a lesser extent the average, shows distinctions between performance (e.g. the risk that biomass falls below 5% of B0). Figure 17 illustrates that defined in this way, the constant catch CON_01 ‘rule’ has higher risk than the feedback rules, and those which have recruitment feedback (SGF_304,305,306, CGF_01) have lower risk than those which do not. The higher risk of SGF_305 (mentioned above) is also apparent in this figure. We consider that these measures of risk in terms of extremely low biomass relative to unexploited (or other very low biomass measures, e.g. 15% of B1980), are only appropriate for distinguishing between the performance of different rules, and are inappropriate for use as absolute measures of risk. The reason for this is that the risk of falling below 5% of B0 does not reflect the potentially substantial risk associated with biomass being between, for example, 20% and 5% of B0.

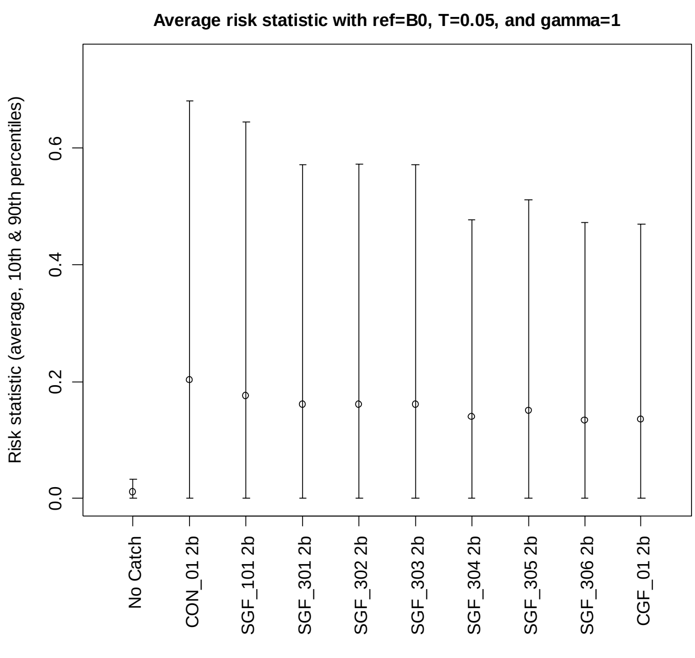


Figure 17. Average risk statistic for biomass falling below 5% of B0, showing difference between rules in the 90th percentile and the average.

3.7 Robustness trials

One of the drivers for exploring the incorporation of a recruitment index into the decision rule was the robustness of the rule to poor recruitment. A comparison of SGF_306 (recruitment index) with SGF_302 (no index), and of CGF_01 (recruitment index) with SGF_303 (no index) under the robustness trial "lowR4" (i.e. a sequence of 4 years of low recruitment) shows that the inclusion of the recruitment index in the rule does make it more robust in terms of some of the biomass performance measures (Figure 18). In terms of some of the catch performance measures SGF_306 and CGF_01 also perform better and at least not noticeably worse than SGF_302 and SGF_303 (Figure 19). (Note: a full set of figures for all the robustness trials can be made available)

It is also informative to look at 'wormplots' for some of the robustness trials. For example, wormplots for the poor recruitment robustness trials compared to the reference set (Cfull2) show that rules which use the recruitment index are more robust to this test than those which do not use the index (Figure 20).

A similar comparison for the robustness trials which assume no autocorrelation in recruitment (Cfull2_noAC and Cfull2_noAC_tripleR) illustrate an interesting point (Figure 21). In this case, the rule without recruitment index rebuilds biomass more under the robustness trials (which are essentially more optimistic than the reference set) than the rule with recruitment index. This appears to be as a result of the choice of tuning measure (not the value, but the fact that it is relative biomass in the year 2022). Essentially, those rules which cut catches by more early on have to increase catches later on in order not to overshoot the tuning level. In the somewhat more optimistic 'robustness' trials, the rule then also increases catches by a relatively large amount implying much better catch performance but similar (rather than better) biomass performance in the medium to long term. (Recall that the rules are not re-tuned for the robustness trials).

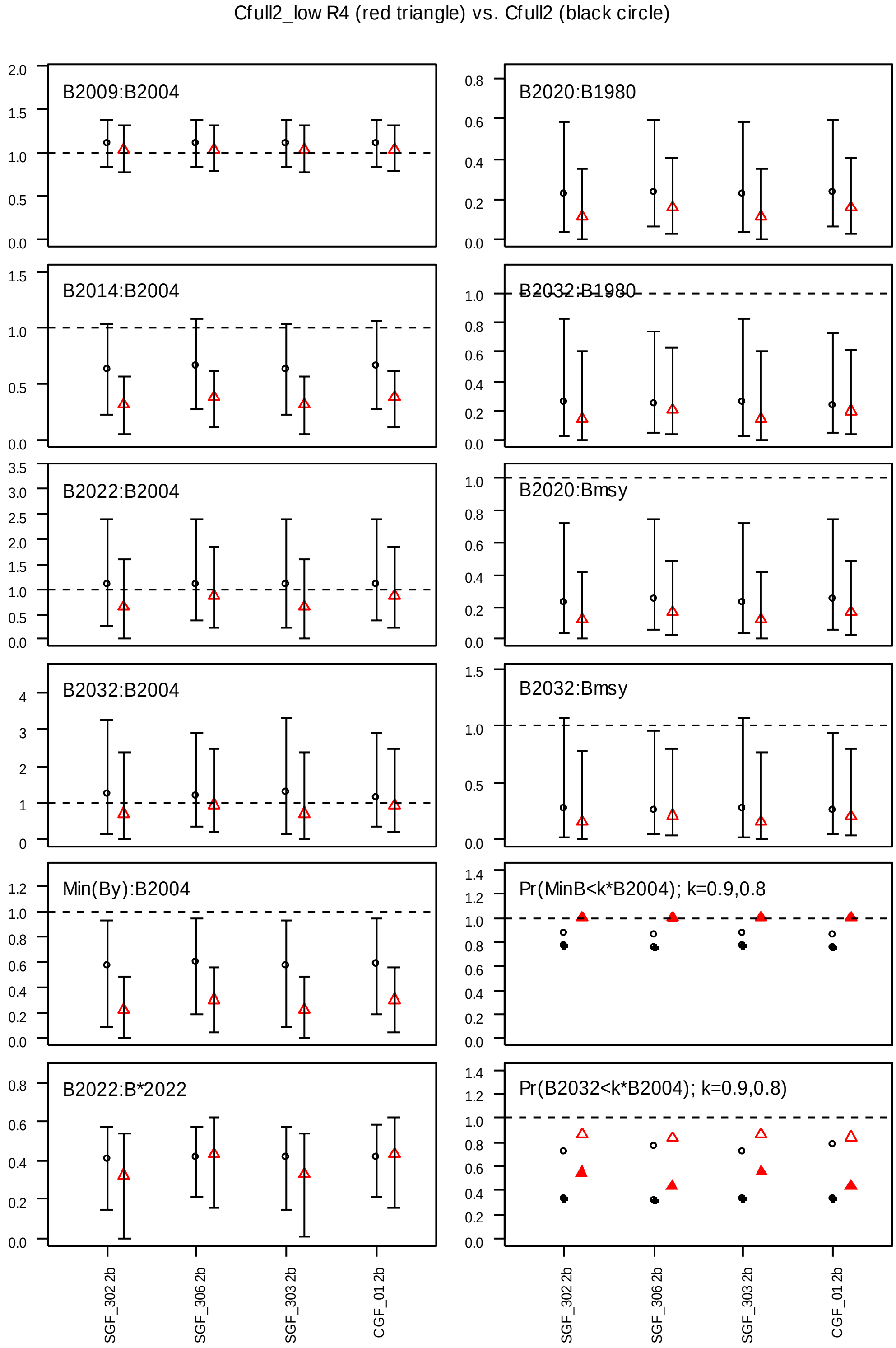


Figure 18. Biomass performance measures for robustness trial "lowR4". SGF_306 and CGF_01 includes recIndex feedback.

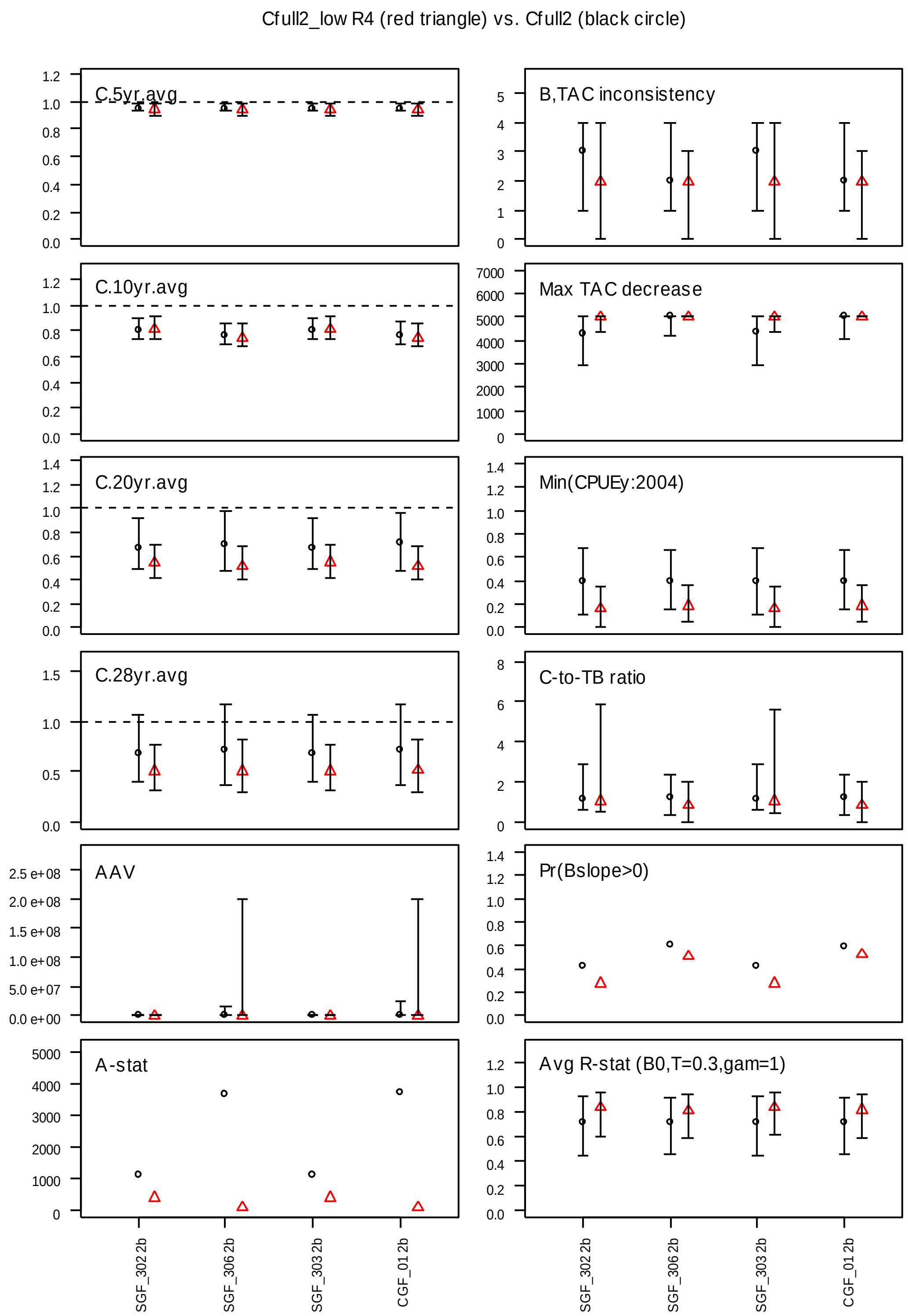
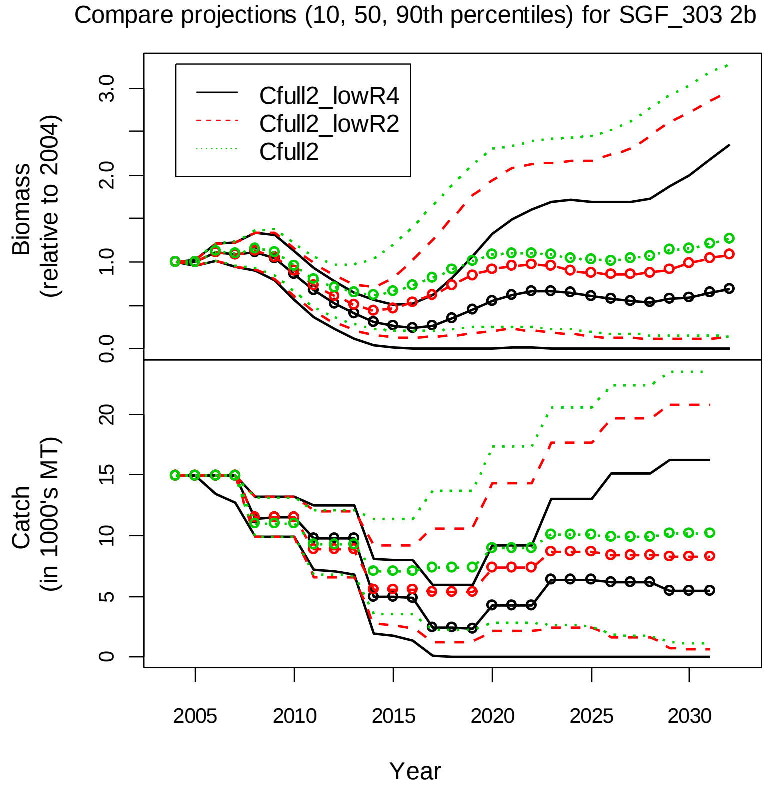


Figure 19. Catch performance measures for robustness trial "lowR4". SGF_306 and CGF_01 includes recIndex feedback.



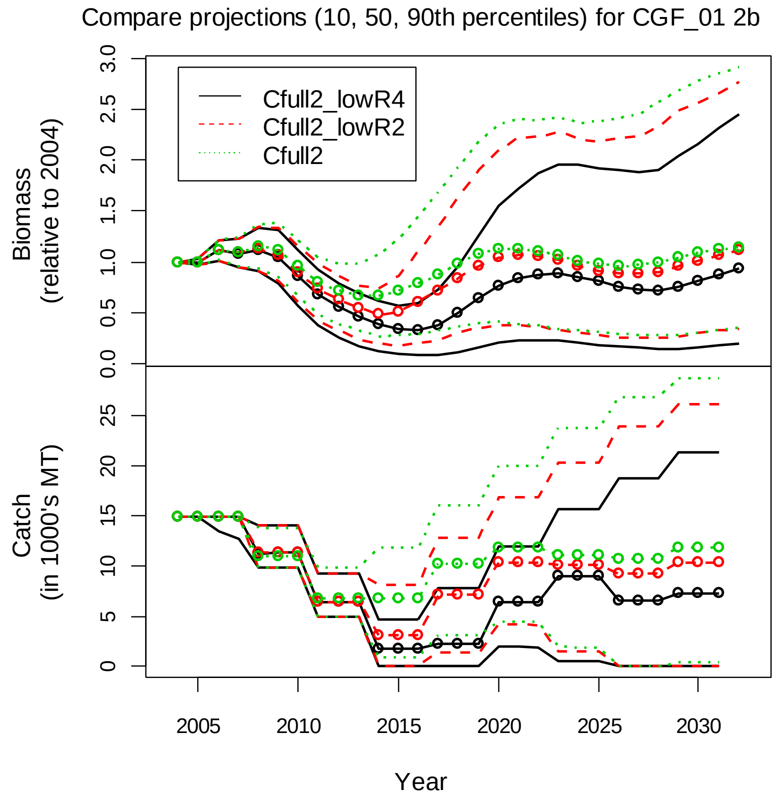
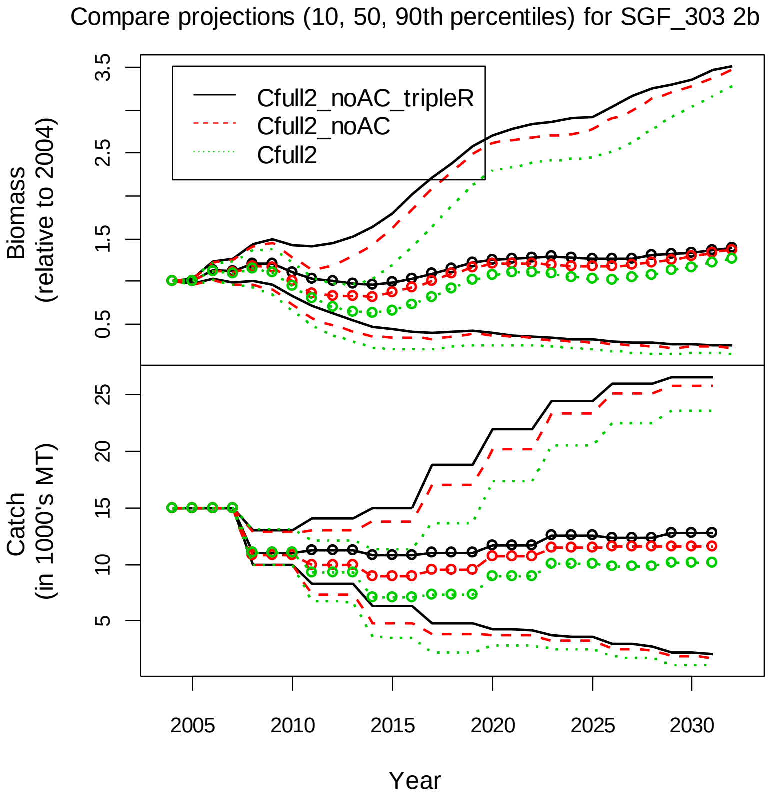


Figure 20. Wormplots of biomass and catch for 2 robustness trials assuming poor recruitment and the reference set for SGF_303 (no recIndex; upper panels) and CGF_01 (with recIndex; lower panels).



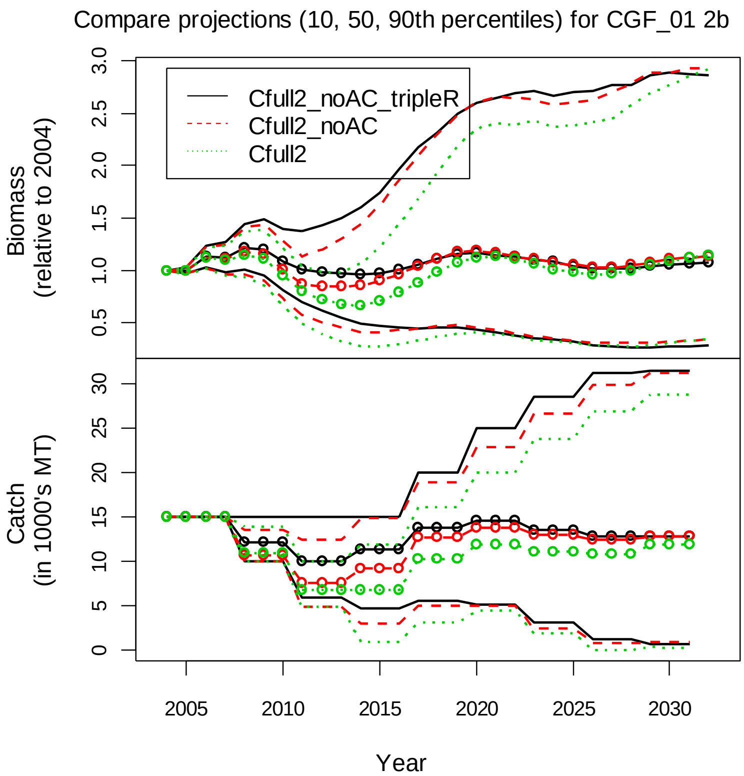


Figure 21. Wormplots of biomass and catch for the reference set and 2 robustness trials which assume no autocorrelation in recruitment for SGF_303 (no recIndex; upper panels) and CGF_01 (with recIndex; lower panels).

3.7 Footnote regarding continuous vs discrete recruitment index

As noted at the start, we did not have time to complete runs for CGF_42 until the end, but given the great similarity in performance between the rules with continuous and discrete recruitment indices, we discussed SGF_306 (discrete recIndex) as a proxy for CGF_42 (continuous recIndex). Figure 22 illustrates that they are almost identical, and we have already noted that we prefer the continuous version (CGF_42) because the resulting TAC is less likely to be sensitive to noisy data. (All results will be available in electronic form at the meeting.)

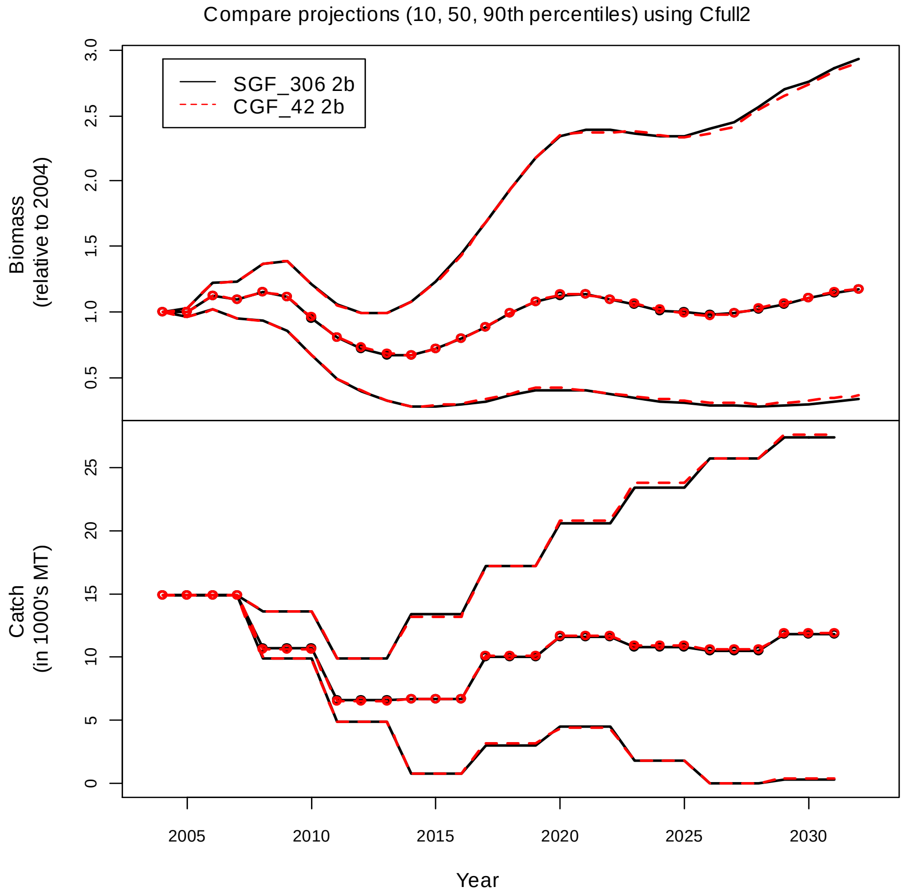


Figure 22. Comparison of continuous (CGF_42) and discrete (SGF_306) recruitment index versions in terms of biomass and catch trajectories (results are almost identical).

4. Discussion

Results of further exploration of the FXR rule and variations thereof were presented above. Some explorations were less successful than others in terms of the performance measures and some have not been presented. Of the rules in Table 1, we consider the following as our "shortlist" of rules: SFG_303, CGF_01 and CGF_42. Electronic results for all the above rules are available for consideration at the meeting. A reminder of their assumptions is repeated here:

	pTAC proportional to r or r^2	2004 TAC ceiling until 2015	TAC increase constraint until 2015 (see footnote 1 above) (proportion of maxChange)	Poor Recruitment Feedback	Initial TAC Drop (proportion of maxChange)
SGF_303	r	Yes	0.5		
CGF_01	r	Yes	0.5	Continuous	
CGF_42 (based on proxy SGF_306)	r	Yes		Continuous SGF_306: discrete	

We note that the potential extensions to the FXR rule as described in this paper were presented in a simple stepwise fashion that ignores a number of potential interactions among rule features. A number of arbitrarily chosen parameters were selected to represent each feature, with limited exploration of alternatives. Undoubtedly, if a unique performance criteria could be defined, MP performance could be improved. However, we do not expect that the improvement would be large if the definition was primarily conservation based, given the overall similarity in performance among the rules examined here.

We found that rule performance was highly dependent on the early TAC trajectory and found different means of reducing catch quickly within the MPs tended to produce rather similar results. Not surprisingly, large initial quota cuts are associated with lower biomass risk, faster recovery, and potentially higher total catches due to increased biomass in the long run. Given the high (>~40%) probability that the stock size is going to decrease to around 50% of 2004 levels before 2020 (at the 1.1 tuning target, even for the most conservative rules), we tended to emphasize the minimization of short-term biological risk in formulating rules (e.g. see Figure 6, and note that there is about a 50% chance of at least some biomass decline even in the absence of fishing). We would be surprised if equivalent levels of biological risk reduction could be attained without the large initial quota cuts. But if the commission is prepared to increase the biological risk in favour of reducing the magnitude of the initial cuts, any of these rules could easily be adjusted to move in that direction. We do, however, caution that substantial changes to parameters and tunings may lead to counter-intuitive results that need to be tested.

Some of the rules (e.g. SGF_305) illustrated the consequence of having adopting the median SSB as value to tune to. Essentially, those rules which cut catches by more early on have to increase catches later on in order not to overshoot the tuning level, and this then leads to relatively higher catches after the tuning year of 2022 and less rebuilding than other rules. The same effect was seen in robustness trials for the more optimistic scenarios (e.g. noAC and noAC_tripleR) by comparing performance of rules which did / did not include the recruitment index.

A simple last minute extension to the CGF_01 model (TCF_01, with a time dependent δ parameter (see equation 2) that declines toward 0.9 of Fox-F_{MSY} suggests that post-2022 behaviour, and hence some of the side-effects of tuning, can be substantially modified without impacting the pre-2022 dynamics (Figure 23).

We selected the rules SGF_303 and CGF_01 (discussion of CGF_42 follows) for further consideration by the wider CCSBT-SC/MP community for the following reasons (emphasizing the 2b tuning results):
1. TAC proportional to r (rather than r^2) seems to perform better particularly with respect to biological risk, as indicated in the cumulative probability plots.

2. The short-medium term ceiling on TAC has a minimal effect on median performance in terms of either catch or biomass, but is recognized as a desirable property by industry and the Commission. Although very few of the MP catch trajectories explored here actually invoked the ceiling constraint, it does ensure that unusual data will not cause the MP to take action that is clearly contrary to the general perceptions about the status of the SBT stock.

3. The short-medium term constraint on the magnitude of TAC increases had a minimal effect on catch and biomass trajectories, but it should add the benefit of stabilizing TAC oscillations when data are noisy, and hence reduce the risk of industry investment volatility and over-capitalization. The summary statistics do not strongly demonstrate that this is in fact happening, but given the perception of the highly depleted status of the stock, we would not expect that large TAC increases could be justified in the short-medium term anyway.

4. Recruitment index feedback somewhat reduced the short-medium term biological risk, and probably demonstrated its greatest advantage in the robustness tests to sustained poor recruitment (lowR2 and lowR4 trials). Presumably, this effect would be even greater if multiple bouts of low recruitment were to occur in the medium term (but no robustness test was designed specifically for this situation). The continuous version of recruitment feedback (CGF_01) is presumably preferable because it is likely to reduce the effect of noisy data on TAC changes.

5. These two rules represent the two characteristic trade-off aggregations among the rules presented here.

We have also “shortlisted” CGF_42 (not shown in full, but based on its discrete recruitment equivalent SGF_306) because it does not include the asymmetric constraint on the magnitude of change in TAC (maximum increase is half that of the maximum decrease) which may be considered undesirable by industry or the Commission. In any case, as noted in point 3 above, the short-medium term constraint on the magnitude of TAC increases had a minimal effect on overall performance.

Obvious disadvantages of these candidate MPs include:
1. Large TAC cuts at the first two opportunities can be expected if the reference set is a reasonable representation of reality. If recent recruitment is not as pessimistic as the reference set, we would still expect a near maximal cut at the first opportunity.

2. Biomass rebuilding beyond 2022 tends to slow down or even reverse somewhat in many of the realizations, particularly for CGF_01. We were not overly concerned about this, because it is not expected that the current MP, without any changes, will have a lifespan that reaches 2022, and because we consider this to be a side-effect of tuning to SSB in 2022. It can also be considered a short-sighted feature of the basic FXR parameterization, and we note below, that it does not need to be a permanent problem.

Finally, we note that in all cases explored, the modest 2022 rebuilding objective implied by a 1.1 tuning resulted in Fox model TAC recommendations that tend to prescribe fishing mortality higher than the Fox-model F_{MSY} (ignoring change constraints and recruitment feedback), because δ in equation 2 is greater than1. In so far as the Fox model represents SBT dynamics, it seems inevitable that the stock never will recover to even MSY levels if the 1.1 tuning is adopted for a long term strategy. This is borne out by the cumulative probability of B2032 being below B_{MSY} (from the operating model, rather than the fitted Fox model) which is around 90% for the rules explored here and for the no-catch scenario is around 45%. We've already commented on the effect of a time dependent tuning parameter which can substantially modify the dynamics post 2022 without impacting on the pre-2022 dynamics (Figure 23).

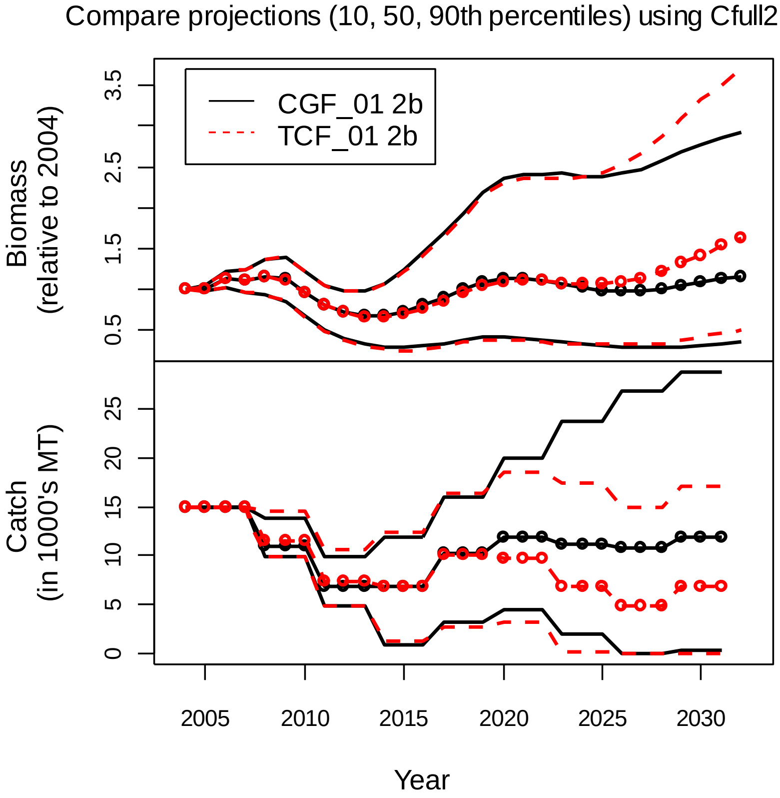


Figure 23. Illustration that concerns about post-2022 rebuilding performance can probably be addressed without seriously altering pre-2022 performance.

5. References

Kolody,D and J. Hartog. 2004. Examples of management procedure behaviour changes in response to operating model updating. CCSBT-ESC/0409/25.

Mace, P.M. 1994. Relationships between common biological reference points used as thresholds and targets of fisheries management strategies. Can.J.Fish.Aquat.Sci. 51:110-122.

Polacheck, T. 2003. Issues related to setting rebuilding objectives for southern bluefin tuna. CCSBT-ESC/0309/30.

Polacheck, T, P. Eveson, J. Hartog, M. Basson, and D. Kolody. 2004. Comparison of the performance of tuned candidate management procedures for southern bluefin tuna based on the final trial specifications and testing procedures. CCSBT-MP/0404/04 and Appendix.