IOTC Document Number
IOTC-2021-WPM12(MSE)-04

OCEANS AND ATMOSPHERE
www.csiro.au

Indian Ocean Bigeye Tuna Management Procedure Evaluation Update March 2021

Prepared for the Management Strategy Evaluation Task Force of the Indian Ocean Tuna Commission Working Party on Methods Meeting, March 2021

Dale Kolody, Paavo Jumppanen
CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart TAS 7000, Australia

Citation

Kolody, D, Jumppanen, P. 2021. Indian Ocean Bigeye Tuna Management Procedure Evaluation Update March 2021. Working Paper prepared for the Management Strategy Evaluation Task Force of the Indian Ocean Tuna Commission Working Party on Methods Meeting, March 2021. IOTC-2021-WPM12(MSE)-04.

Contents

Acknowledgments

This work was jointly funded by the Australian Department of Foreign Affairs and Trade and Australia's Commonwealth Scientific and Industrial Research Organization (CSIRO - Oceans & Atmosphere), administered by the United Nations Food and Agriculture Organization. Technical oversight and advice was provided by various IOTC Working Groups, and notably the participants of the IOTC MSE Task Force, including Toshihide Kitakado, Gorka Merino, Hilario Murua, Dan Fu and M. Shiham Adam. Operating model conditioning built upon the stock assessment work of Dan Fu, Adam Langley and others at the IOTC, with useful Stock Synthesis advice from Ian Taylor. The original R-based MSE code from phase 1 was adapted from the Atlantic Bluefin MSE work developed by Tom Carruthers (funded by the ICCAT GBYP project). Summary result graphics were adopted from the Fisheries Library in R (FLR) code provided by Iago Mosqueira. Thanks to Simon Hoyle and Dan Fu for various files and clarifications related to the 2019 bigeye tuna assessment and CPUE standardization.

Summary

This working paper briefly describes developments on the Indian Ocean Tuna Commission (IOTC) bigeye (BET) Management Procedure (MP) evaluation project, since the 2020 Working Party on Tropical Tunas (WPTT) and Working Party on Methods (WPM). Some concerns about the BET OMs were identified and discussed in 2020, but the issues were of a generic nature affecting multiple IOTC species, with potential solutions being explored by multiple parties (e.g. the yellowfin stock assessment and OM are affected to a greater and more urgent extent). No changes were proposed for the bigeye reference set OM grid presented in 2020, while one additional robustness test was added (5% reported overcatch, on top of 5% unreported overcatch during the projection period). This paper describes:

- further development of Management Procedures (MPs) based on the Pella-Tomlinson Random Effects (PTRE) model with joint process and observation error (implemented with TMB software). A new MP was added which involves fitting the PTRE model, but then calculating the TAC recommendation by conducting internal projections to identify the constant catch quota that will result in the depletion hitting a target level in a certain number of years. i.e. The approach parallels the manner in which the Commission appears to interpret the K2SM.
- A suite of nine MPs, spanning a range of structural forms, were tuned to the TCMP tuning objectives with the reference set OM, and are compared with the standard IOTC MSE graphical outputs.
- Three representatives of the tuned MPs were evaluated against the 6 Robustness tests identified by the WPTT and WPM.
- Additional tests were conducted to determine whether the bigeye MP evaluations were sensitive to some nuisance assumptions related to spatial refugia and the specific catch equation implementation within the OM. Unlike the Indian Ocean yellowfin situation, the evaluations appear to be robust.
- The report concludes with a list of issues for the Task Force to consider, including:
    - What to present to TCMP 2021 (we recommend two specific MPs at the two existing tuning levels).
    - Further OM development considerations, notably the use of plausibility diagnostics
    - Further MP development options
    - Timeline and funding support toward BET MP adoption

1 Introduction

This paper represents a progress update on key technical elements of the IOTC bigeye tuna (BET) Management Procedure (MP) evaluation project since WPM (2020) and WPTT (2020), with the relevant feedback summarized in Appendix A. No specific changes were proposed for the bigeye reference set OM described in Klody and Jumppanen (2020B), while one additional robustness test was requested (5% reported overcatch, on top of 5% unreported overcatch during the projection period).

A brief summary of the state of the BET reference set OM and robustness tests are provided in Appendix B. The objectives of this report are:

- Describe a new MP, that combines a Pella-Tomlinson Random Effects model with constant catch internal projections to attain a target depletion level, in a manner that resembles how the Commission appears to interpret the Kobe 2 Strategy Matrix.
- Provide the standard evaluation results for a suite of MPs for the reference set OM and Robustness tests.
- Test whether the BET reference set OM is sensitive to nuisance assumptions related to spatial refugia and the choice of the catch equation implementation.
- Highlight the feedback requirements from the Task Force

The target audience for this paper is already familiar with the scope of the work and technical jargon. Other interested parties may need to consult the history of project reports found in https://github.com/pjumppanen/niMSE-IO-BET-YFT/ (and the IOTC meeting report archive).

2 A new MP that resembles the IOTC usage of the Kobe 2 Strategy Matrix

Kolody and Jumppanen (2020A) described a joint process and observation error Pella-Tomlinson Random Effects (PTRE) model, implemented in TMB (Template Model Builder), that was numerically efficient, and appeared to provide better performance than the observation error only model that had been used in previous MP development iterations. The PTRE model has since undergone testing with YFT as well as BET, and several modifications were made to make the algorithm more robust to adverse situations (e.g. there is a repeated minimization from multiple starting points, and an automated (reproducible), iterative restriction of priors when there are minimization problems). The original MP involved fitting the PTRE model, and applying a standard "hockey-stick" type Harvest Control Rule (HCR), which calculates the next TAC as a function of the estimated stock status. The new MP also uses the PTRE model to make inferences about the state of the stock, but then recommends the TAC on the basis of simple internal constant catch projections.

The new MP (cartoon presented in Figure 1) involves

i) fit the PTRE model,

ii) solve for the constant catch quota that will allow the population to hit a pre-defined depletion level in a pre-defined period of time

We had no prior expectation that this MP would be better than the original hockey-stick type MP, but it may be attractive from the perspective of communication, because of the way that it mimics the Commission decision process of interpolating the K2SM provided by a stock assessment. Whether or not this is useful might depend on whether the managers can keep the distinction clear between the normal stock assessment process, and the internal workings of the MP.

Control parameters for the new MP include:

- The target depletion level. This was the tuning parameter used to hit the TCMP tuning objectives. Due to the biases inherent in the simplified PTRE model, relative to the OMs, this parameter might be very different from an a priori value that might seem intuitively desirable.
- Target period – how many years in the future the target depletion is expected to be achieved. For testing, this was arbitrarily set to 10 years, i.e. the MP always looks exactly 10 years into the future, regardless of the time window defined for tuning objectives.
- TAC change constraint (15% as requested by the TCMP)
- Nuisance parameters in the PETR model. Various fixed parameters and priors are required to ensure numerical stability in the PETR model under a wide range of test conditions. These are not conventional MP control parameters per se, but they could affect the MP evaluation results and hence need to be a pre-defined part of the MP.

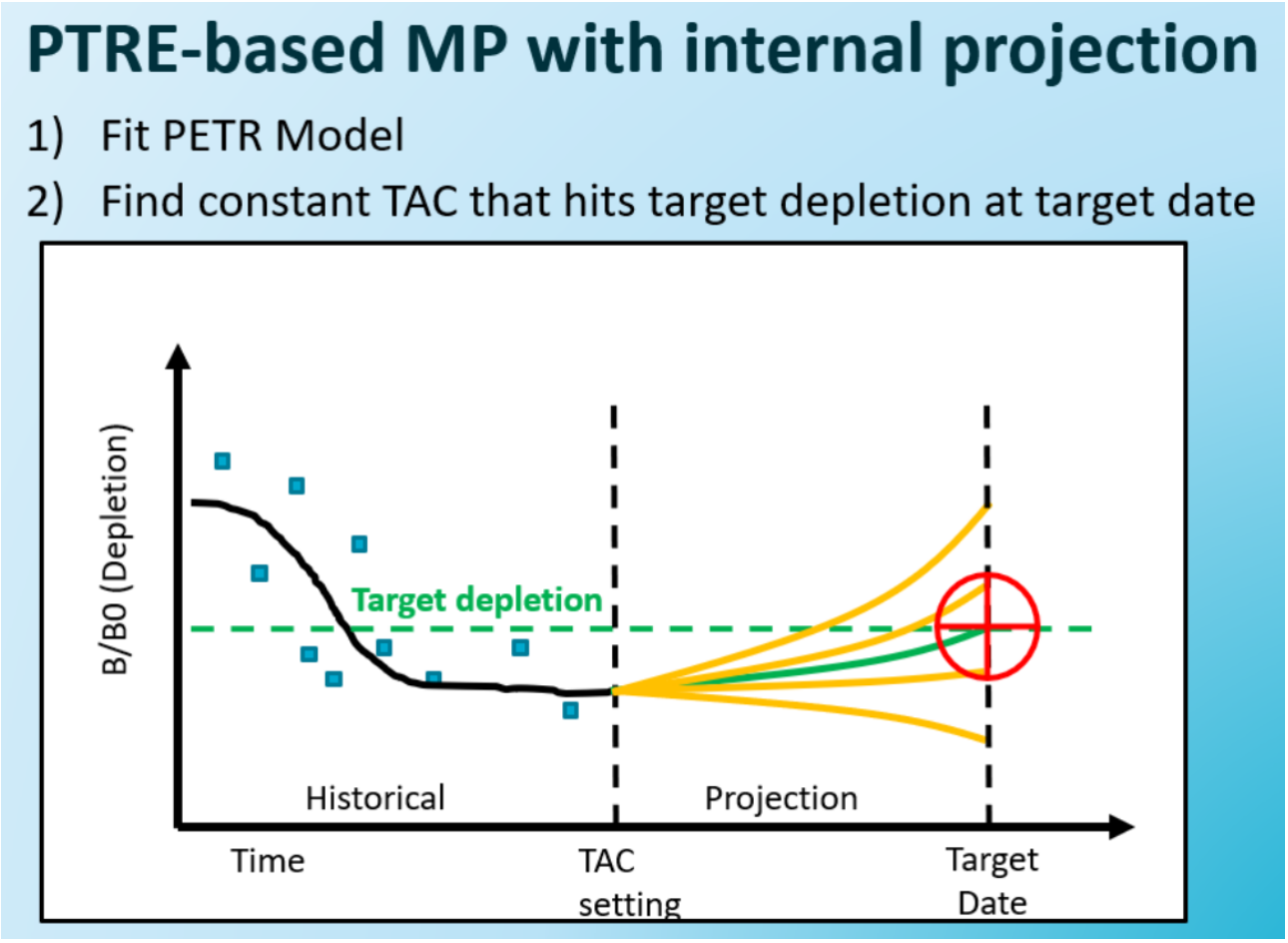

Figure 1. Cartoon representation of the new projection-based MP

3 Bigeye Reference set OM Management Procedure Evaluation results

Nine MPs were selected for testing (Table 1) against the reference set OM (Appendix B), including constant catch projections, data-based MPs, and model-based MPs. Results are presented for the two TCMP tuning objectives separately:

B2: Pr(Kobe green zone 2030:2034) = 0.6. The stock status is in the Kobe green quadrant over the period 2030-2034 exactly 60% of the time (averaged over all simulations).

B3: Pr(Kobe green zone 2030:2034) = 0.7. The stock status is in the Kobe green quadrant over the period 2030-2034 exactly 70% of the time (averaged over all simulations).

MP evaluation result labels include the MP name from Table 1, followed by the tuning objective (B2 or B3).

Most MPs have some proportion of realizations in which the TAC cannot be removed in full, particularly in the latter part of the projection period (e.g. evident with the constant catch MP in Figure 8). This is discussed in more detail in section 5. It is also notable that the variability among realizations tends to be higher in the latter part of the time series. To some extent this is probably a function of time and chance allowing realizations to diverge in different directions, but some component is also an inevitable consequence of the diverging catchability trends among scenarios. i.e. the CPUE indices are unbiased for 50% of realizations, and the other 50% have a bias that continues to increase over time.

B2 Tuning Objective results:

The standard format, 15 year, time-integrated performance plots are shown in Figure 2 - Figure 4 for tuning level B2, from which we note:

- All MP results are fairly similar in terms of expected biomass risk.
- All MPs are expected to take average catches that are slightly higher than current catches over the next 15 years.
- The constant catch, and data-based MPs have the lowest catch variability
- MP PT41F.t15.tmb appears to have the lowest biomass risk

The standard time series plots for the B2 tuning are shown in Figure 5 - Figure 8, from which we note

- The trajectories for all of the feedback-based MPs are similar, with a gradual rise and fall of biomass over the projection period on average.
- Catches are fairly steady, with a slight rising trend over time.
- The variability of the model-based MPs is higher than the CPUE-based MPs

- We would consider the PTRE hockeystick MP (PT41FM.t15.tmb) to be the best in terms of minimizing conservation risk, while the PTRE projection MP (PTBoB0Targ.t15) seems to be slightly more stable over time, without the rising catches and fishing mortality near the end of the projection period (e.g. Figure 6 - Figure 8).

B3 Tuning Objective results:

The results for the B3 tuning objective are qualitatively very similar to B2, except slightly more conservative. The standard format, 15 year, time-integrated performance plots are shown in Figure 9 - Figure 11 and the time series plots for the B3 tuning are shown in Figure 12 - Figure 15.

Table 1. Management Procedures included in this report.

Management Procedure	Characteristics
CCt	Constant Catch
IT5.t15g1.5*	CPUE-based MP with a 5 year recent trend window and responsiveness parameter of 1.5
IT7.t15g2	CPUE-based MP with a 7 year recent trend window and responsiveness parameter of 2
PT41F.t15	Pella-Tomlinson observation error model, with a hockey-stick HCR with bends at 40% and 10% of unfished biomass
PT41F.t15.tmb*	Pella-Tomlinson joint process and observation error (PTRE) model, with a hockey-stick HCR with bends at 40% and 10% of unfished biomass
PT41FM.t15.tmb	Pella-Tomlinson joint process and observation error (PTRE) model, with a hockey-stick HCR with bends at 40% and 10% depletion, where current depletion estimate is a conservative percentile from the depletion uncertainty estimates.
PT60F.t15	Pella-Tomlinson observation error model with a hockey-stick HCR, with a hockey-stick HCR with bends at 60% and 0% of unfished biomass
PT62F.t15	Pella-Tomlinson observation error model with a hockey-stick HCR, with a hockey-stick HCR with bends at 60% and 20% of unfished biomass
PTBoB0Targ.t15*	Pella-Tomlinson joint process and observation error (PTRE) model, which solves for the constant catch quota that will attain a pre-defined depletion target in 10 years.

* One of three representative MPs selected for robustness tests

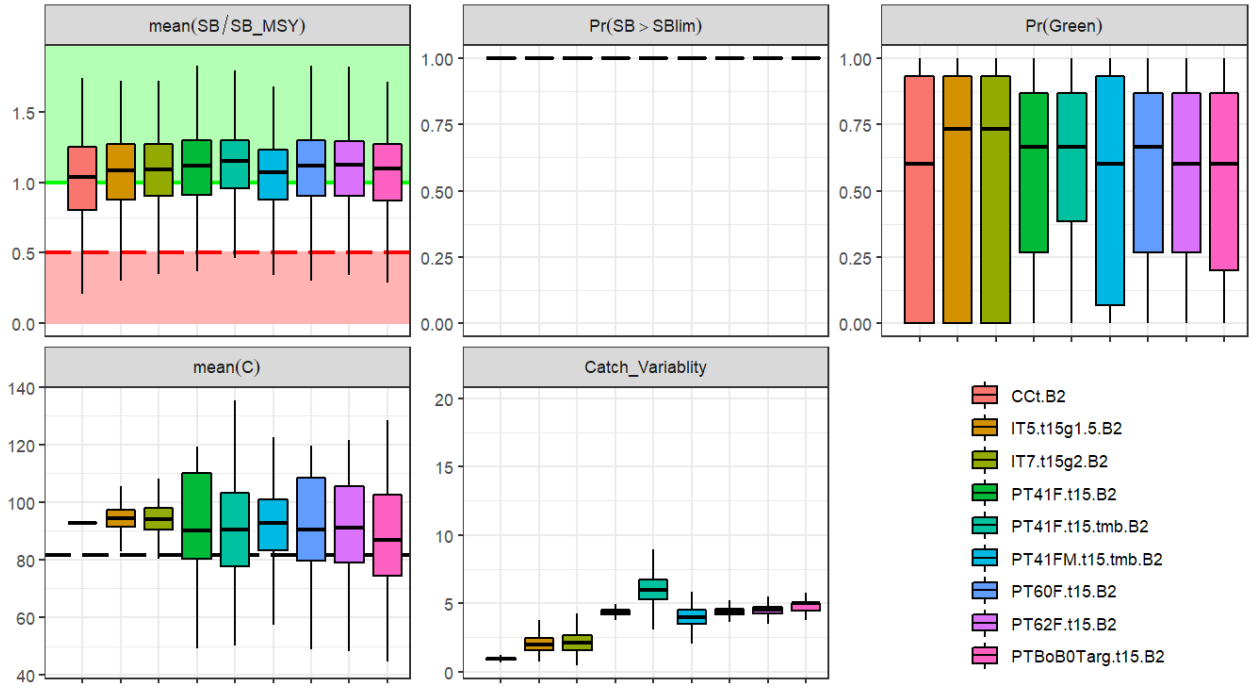

Figure 2. Boxplots comparing (B2 tuned) candidate MPs with respect to key performance measures averaged over the period 2021 -2035. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2020 catch.

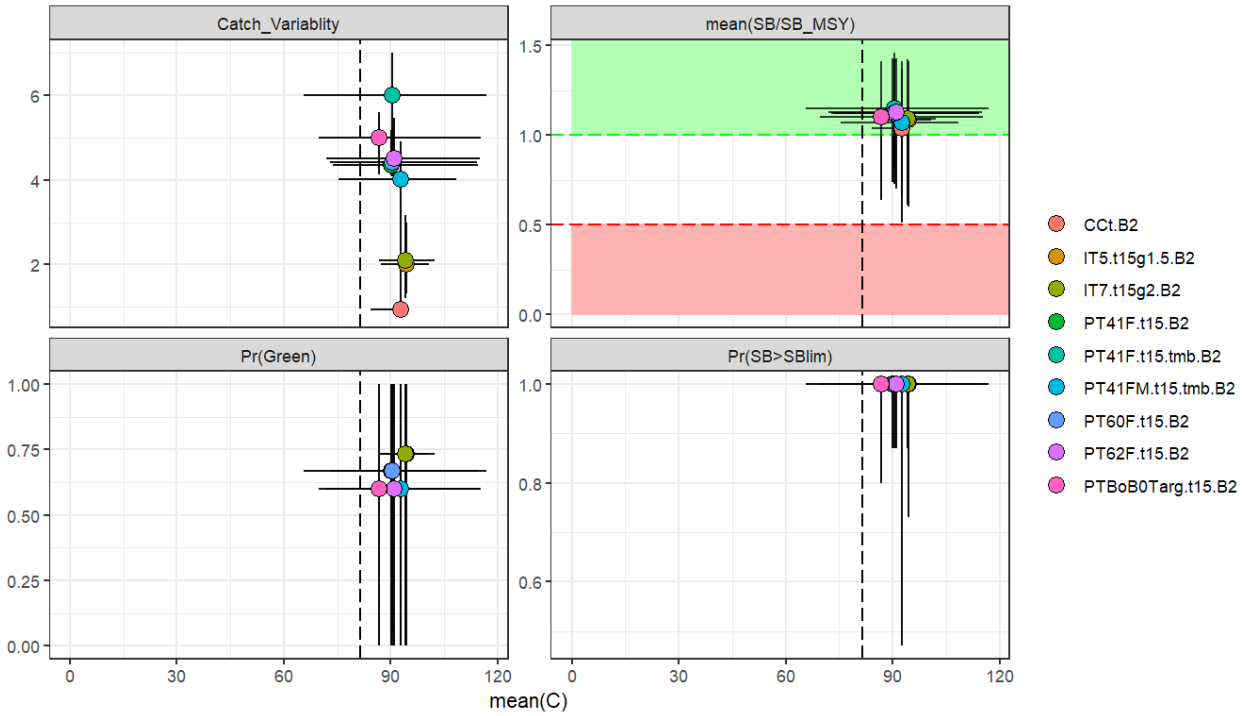

Figure 3. Trade-off plots comparing (B2 tuned) candidate MPs with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2021 - 2035. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2017 catch.

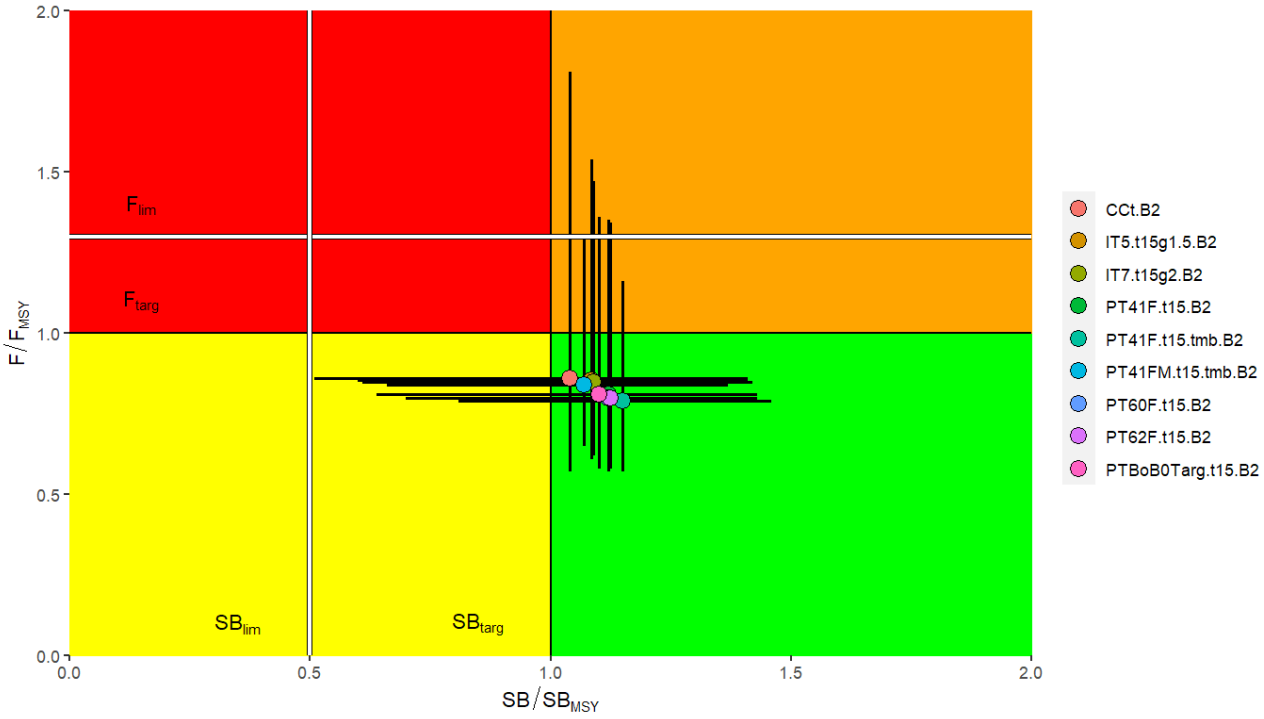

Figure 4. Kobe plot comparing (B2 tuned) candidate MPs on the basis of the expected 15 year average (2021-2035) performance. Circle is the median, lines represent 10th-90th percentiles.

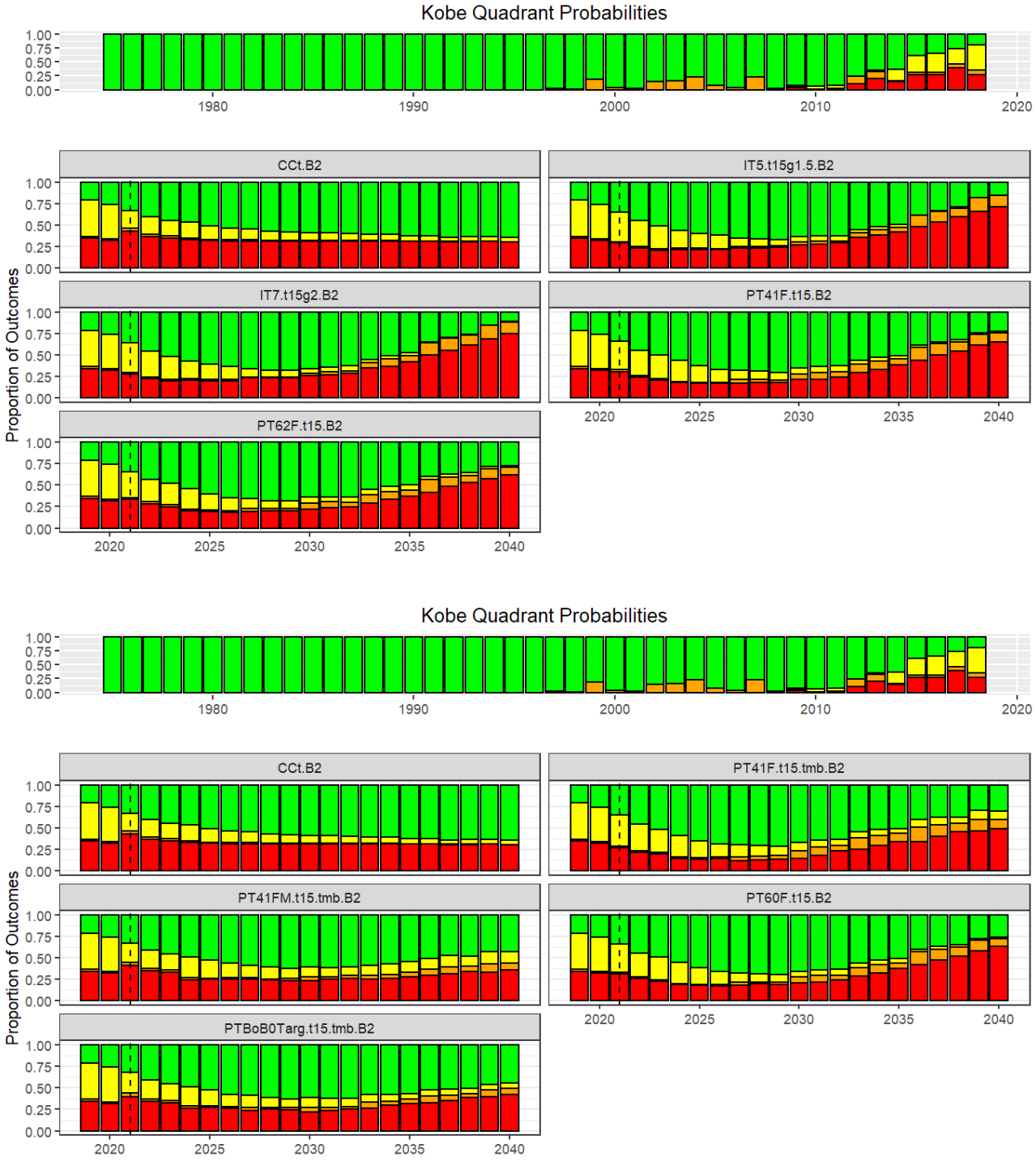

Figure 5. Proportion of simulations in each of the Kobe quadrants over time for each of the candidate (B2 tuned) MPs. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

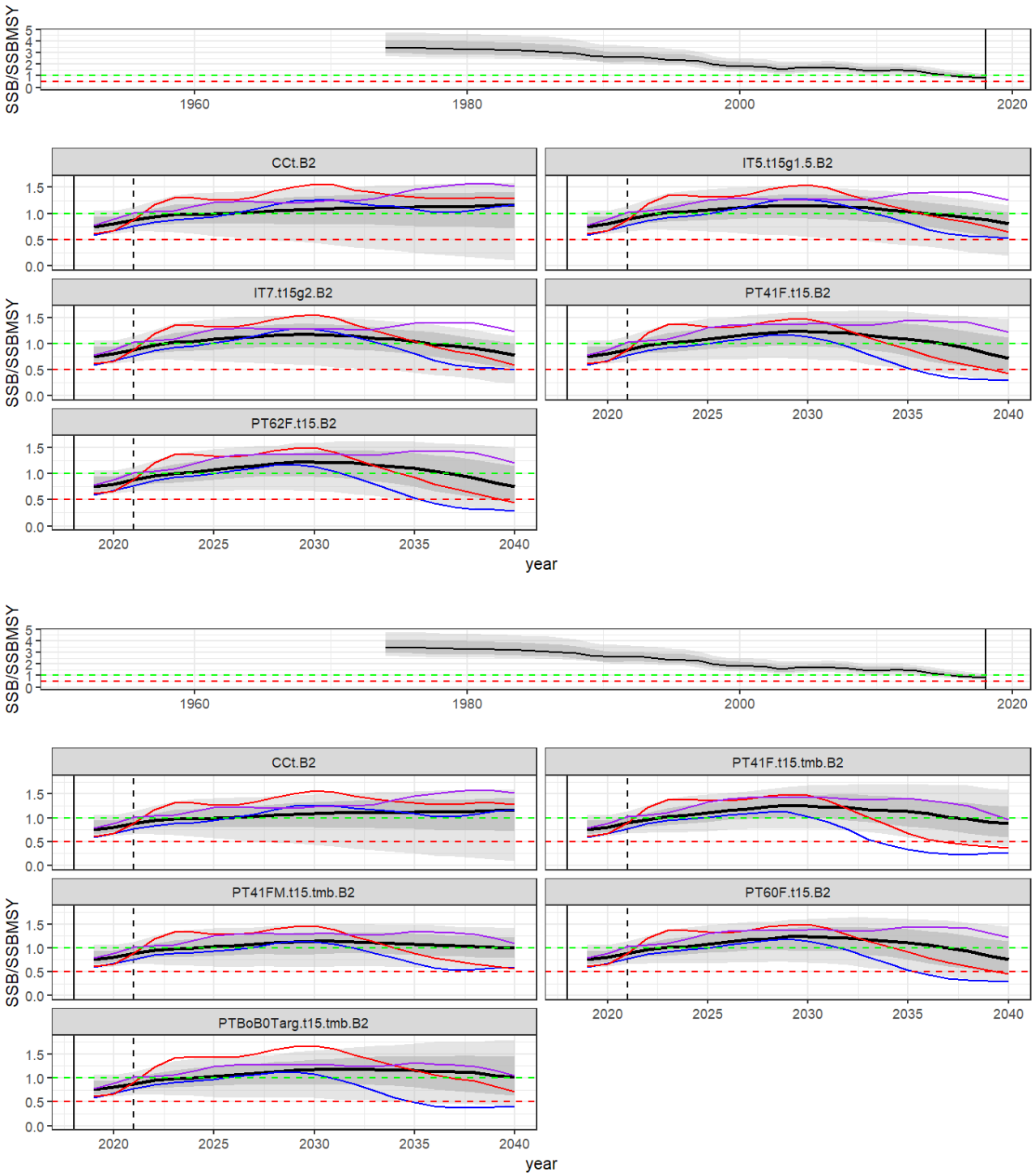

Figure 6. Time series of spawning stock size for the candidate (B2 tuned) MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

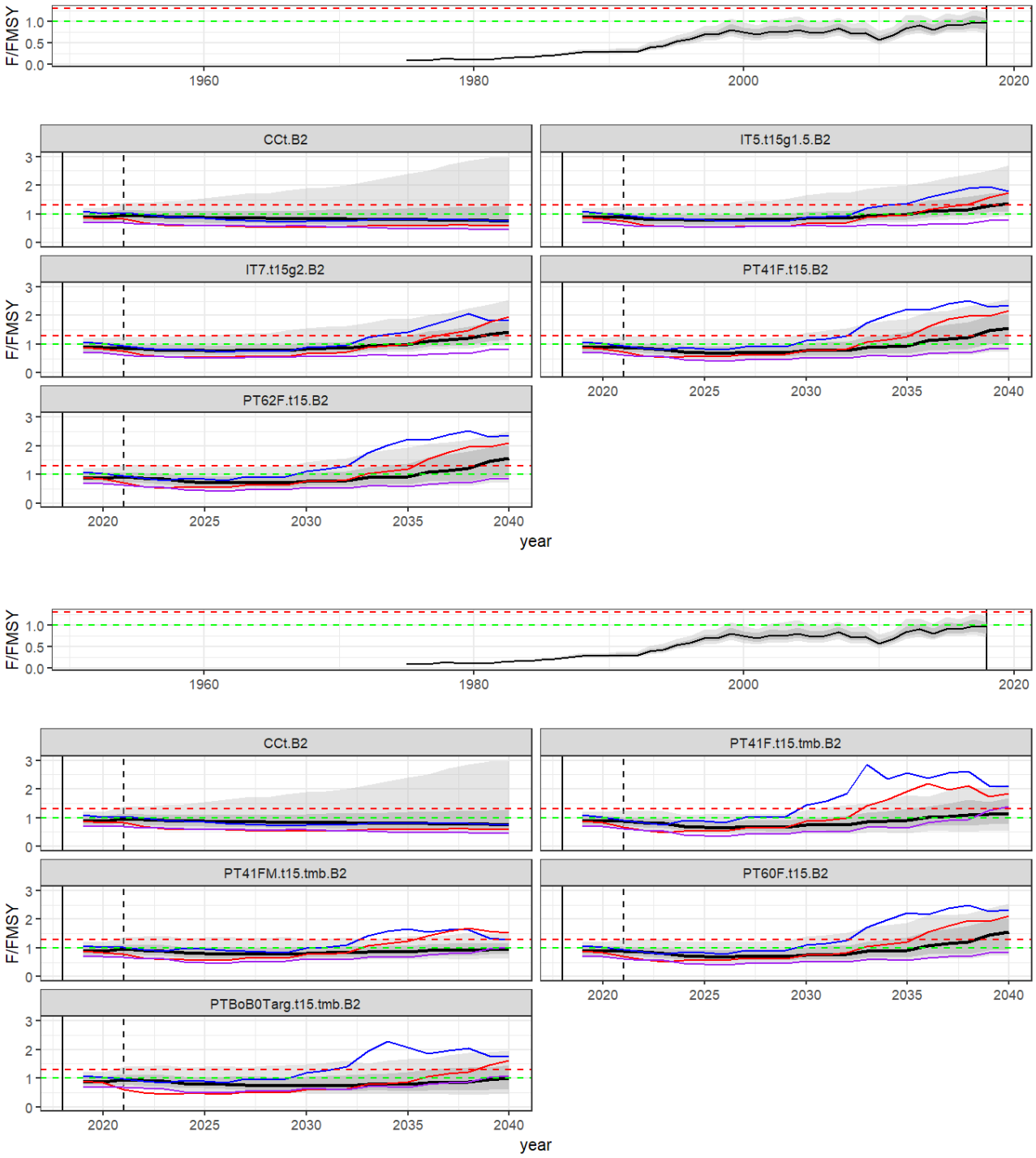

Figure 7. Time series of fishing intensity (Upper bound truncated at F = 3) for the candidate (B2 tuned) MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

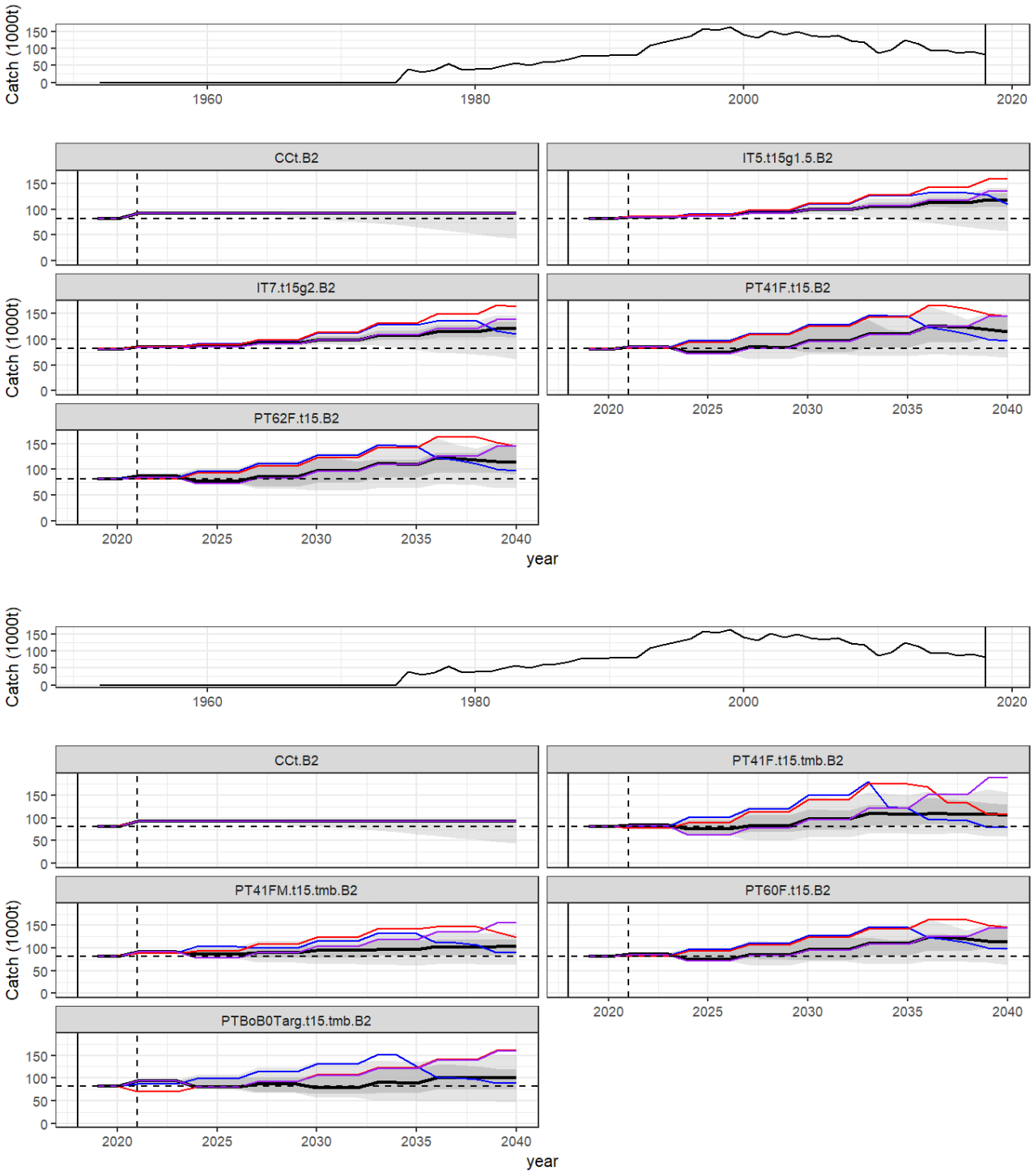

Figure 8. Time series of catch for the (B2 tuned) candidate MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2017) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

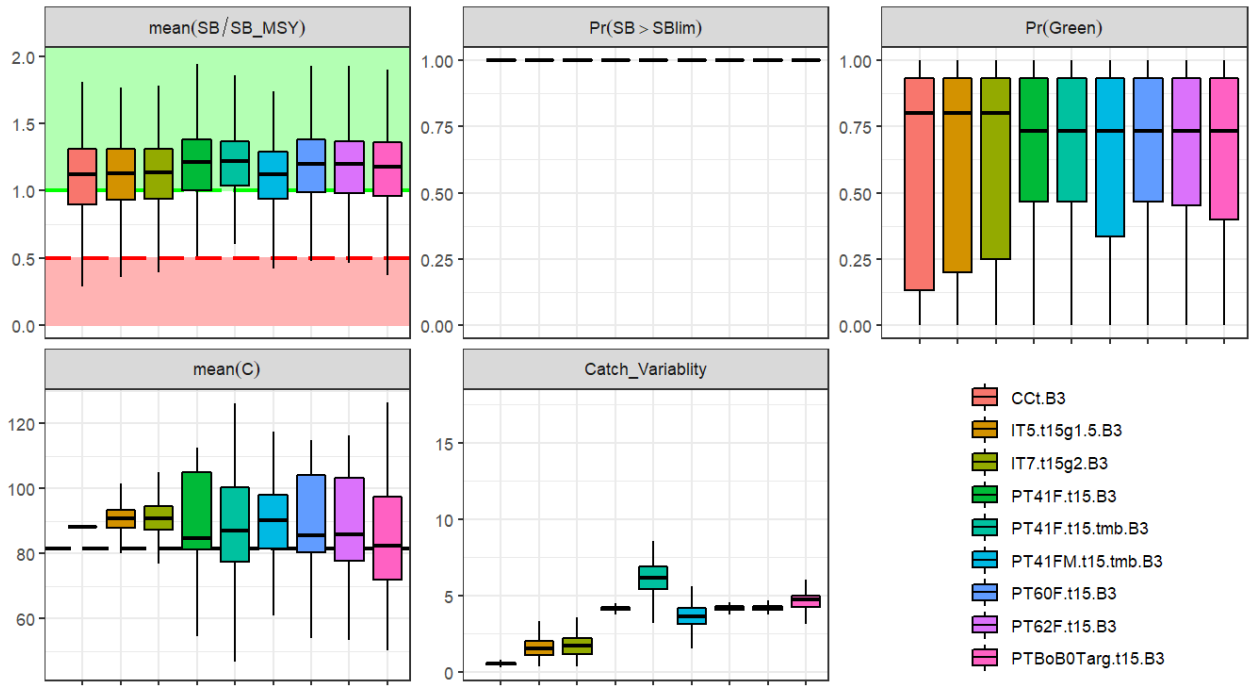

Figure 9. Boxplots comparing (B3 tuned) candidate MPs with respect to key performance measures averaged over the period 2021 -2035. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2020 catch.

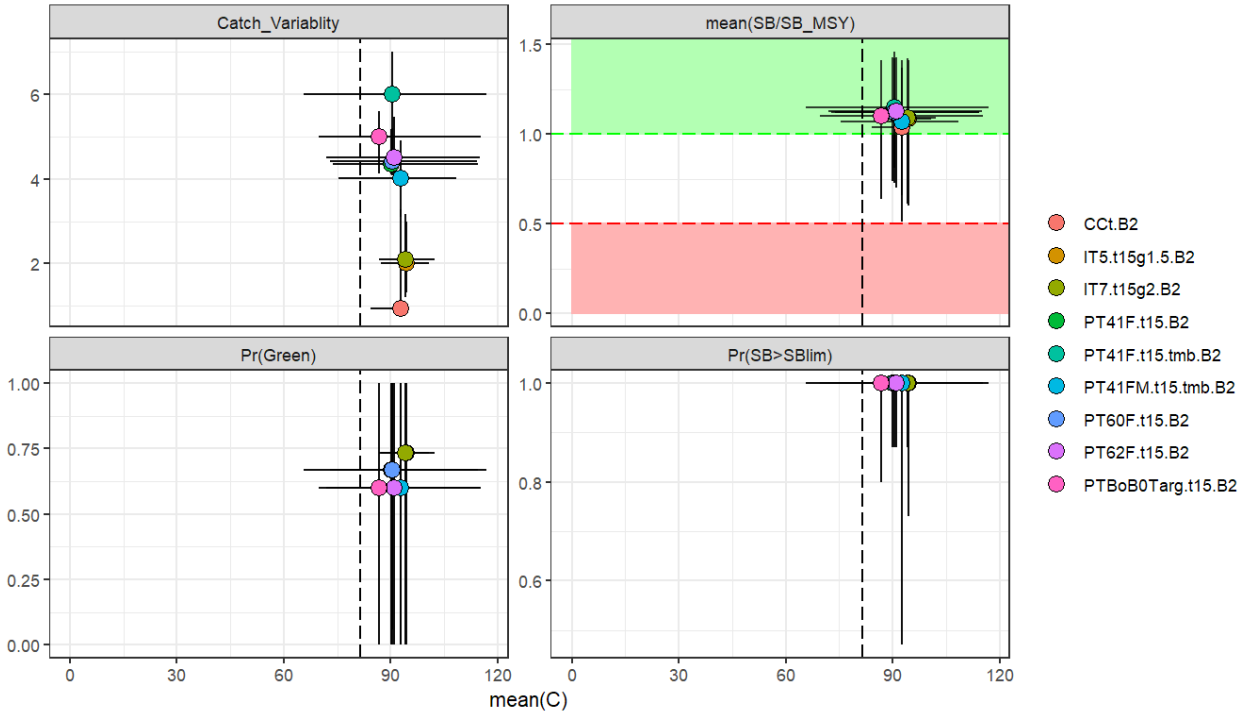

Figure 10. Trade-off plots comparing (B3 tuned) candidate MPs with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2021 - 2035. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2017 catch.

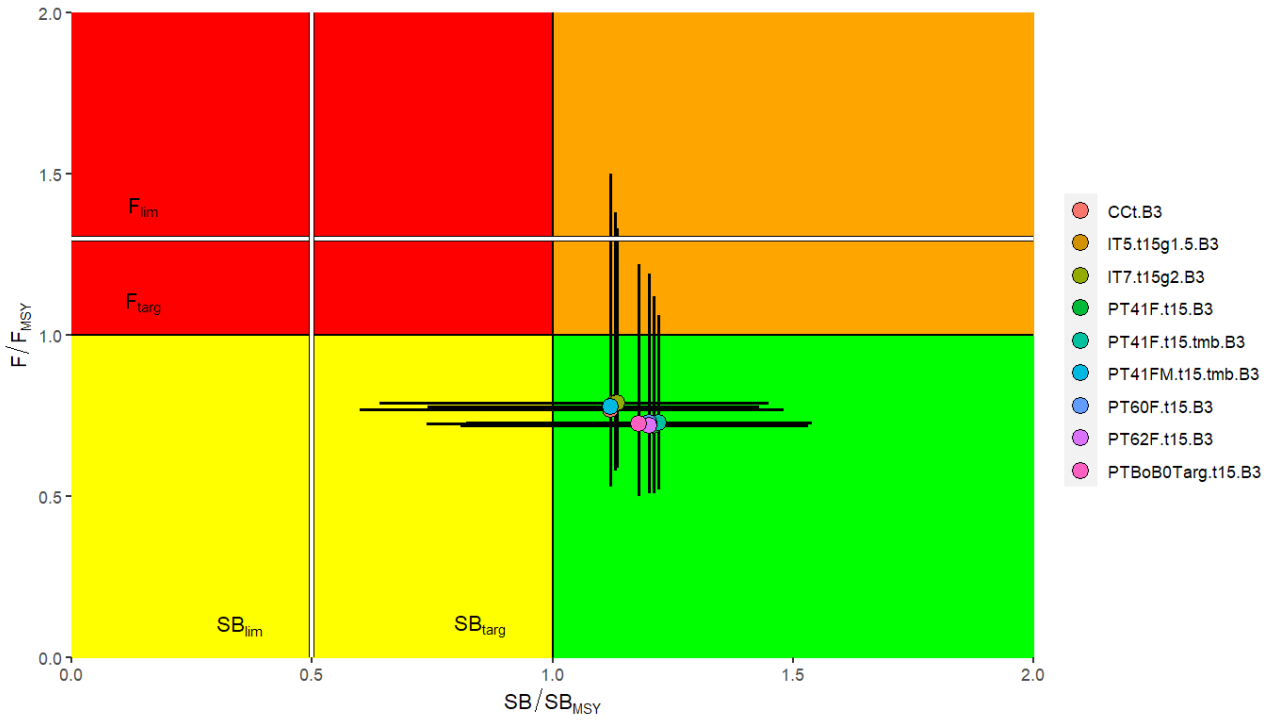

Figure 11. Kobe plot comparing (B3 tuned) candidate MPs on the basis of the expected 15 year average (2021-2035) performance. Circle is the median, lines represent 10th-90th percentiles.

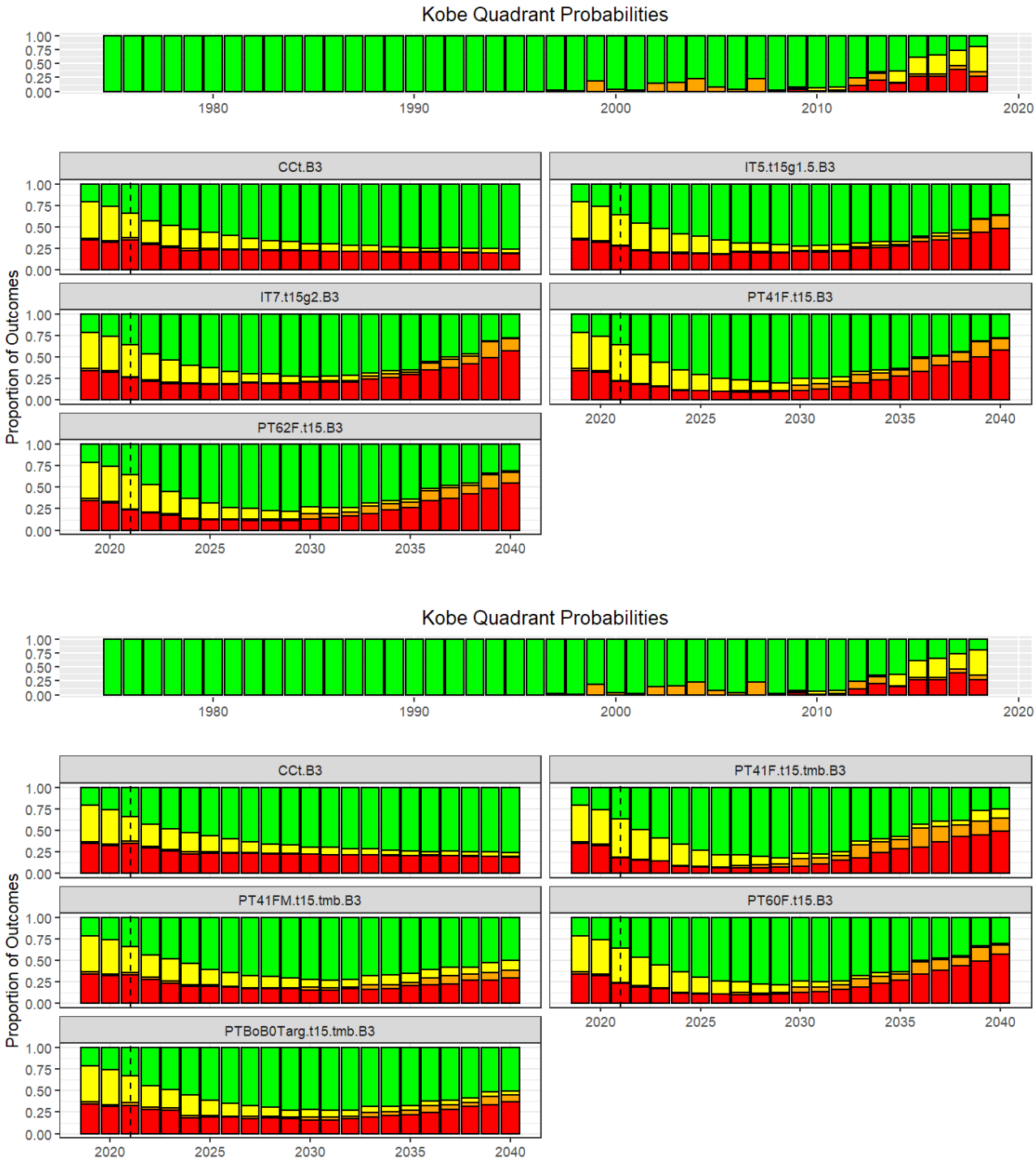

Figure 12. Proportion of simulations in each of the Kobe quadrants over time for each of the candidate (B3 tuned) MPs. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

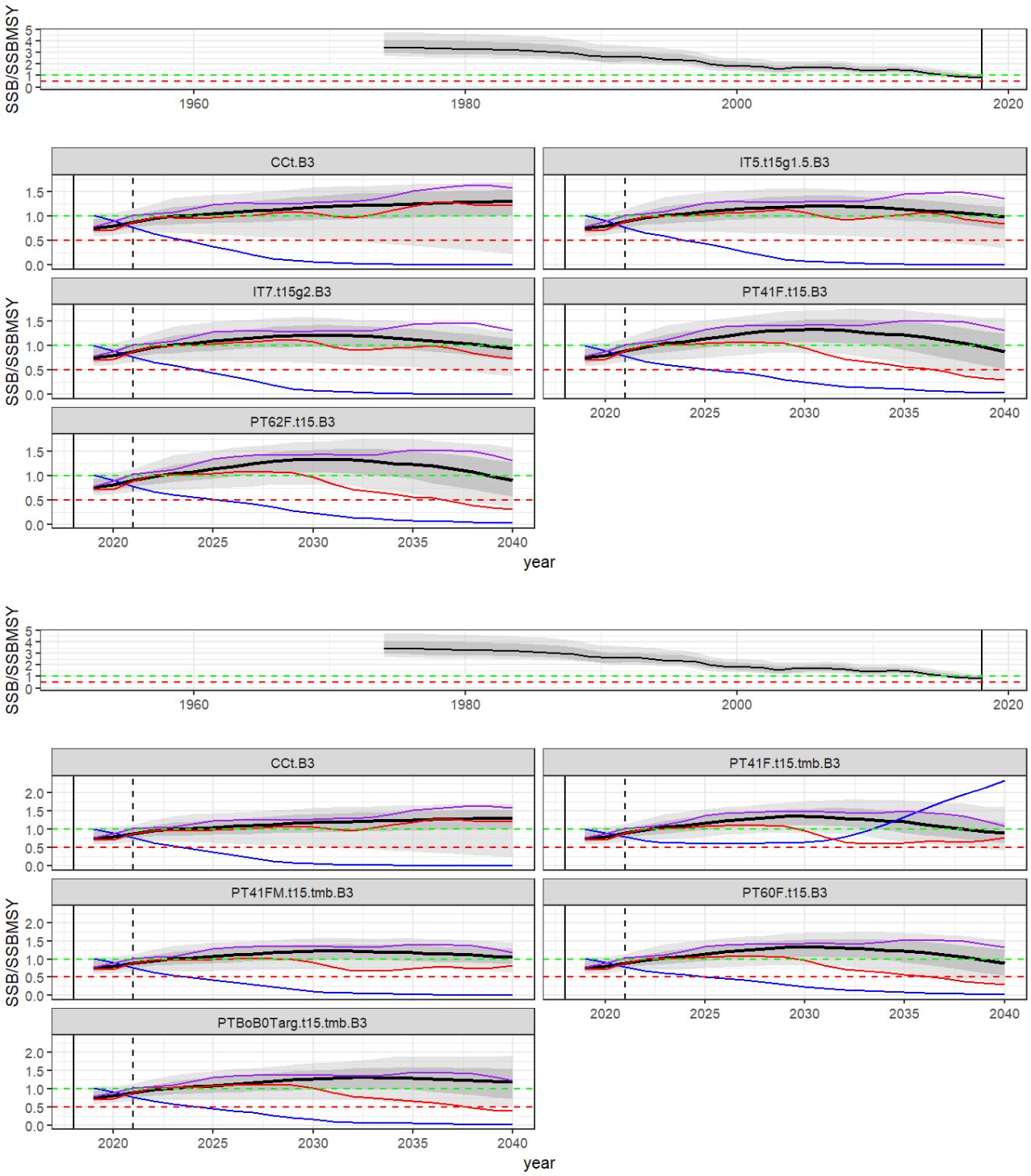

Figure 13. Time series of spawning stock size for the candidate (B3 tuned) MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

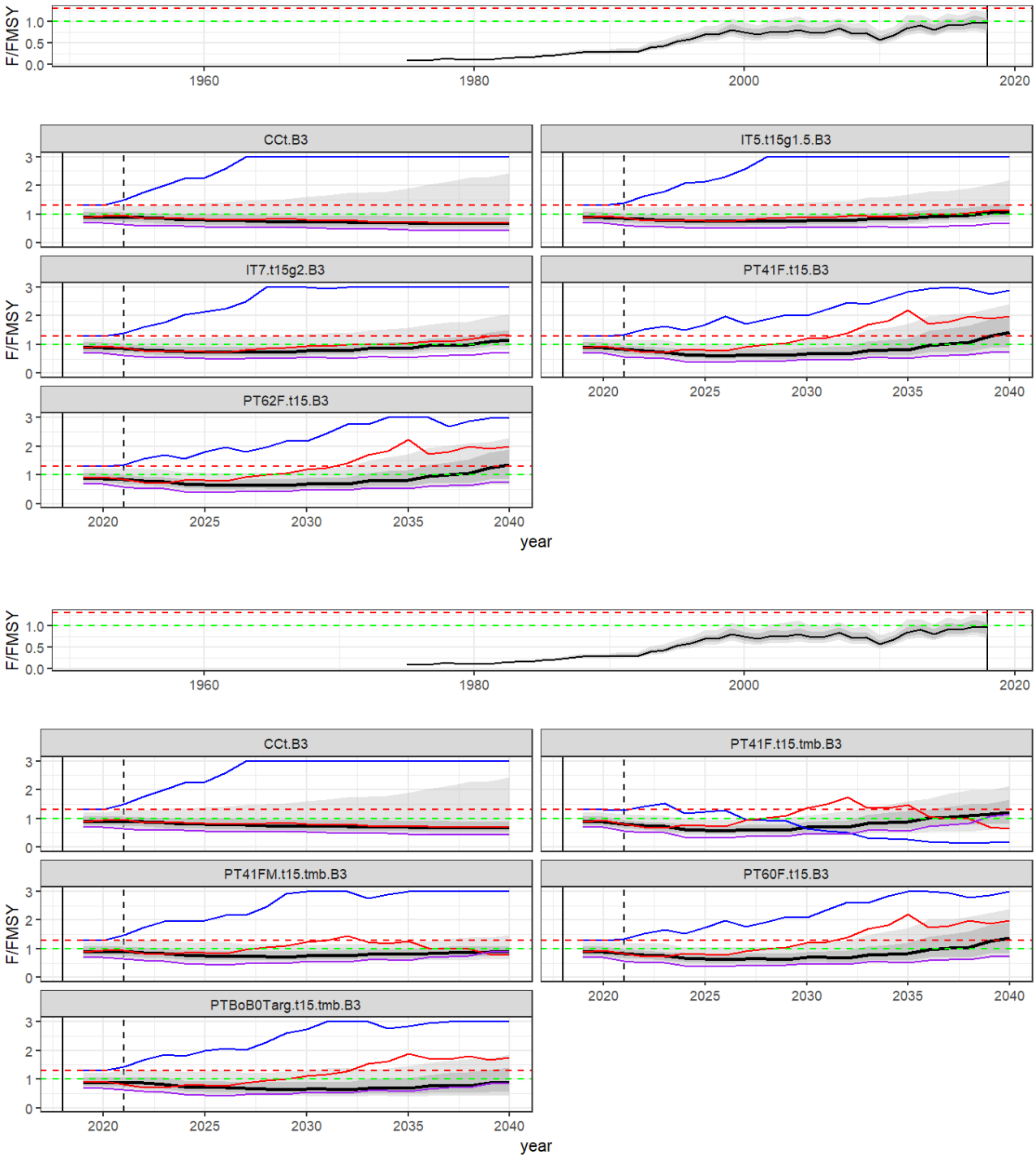

Figure 14. Time series of fishing intensity (Upper bound truncated at F = 3) for the candidate (B3 tuned) MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

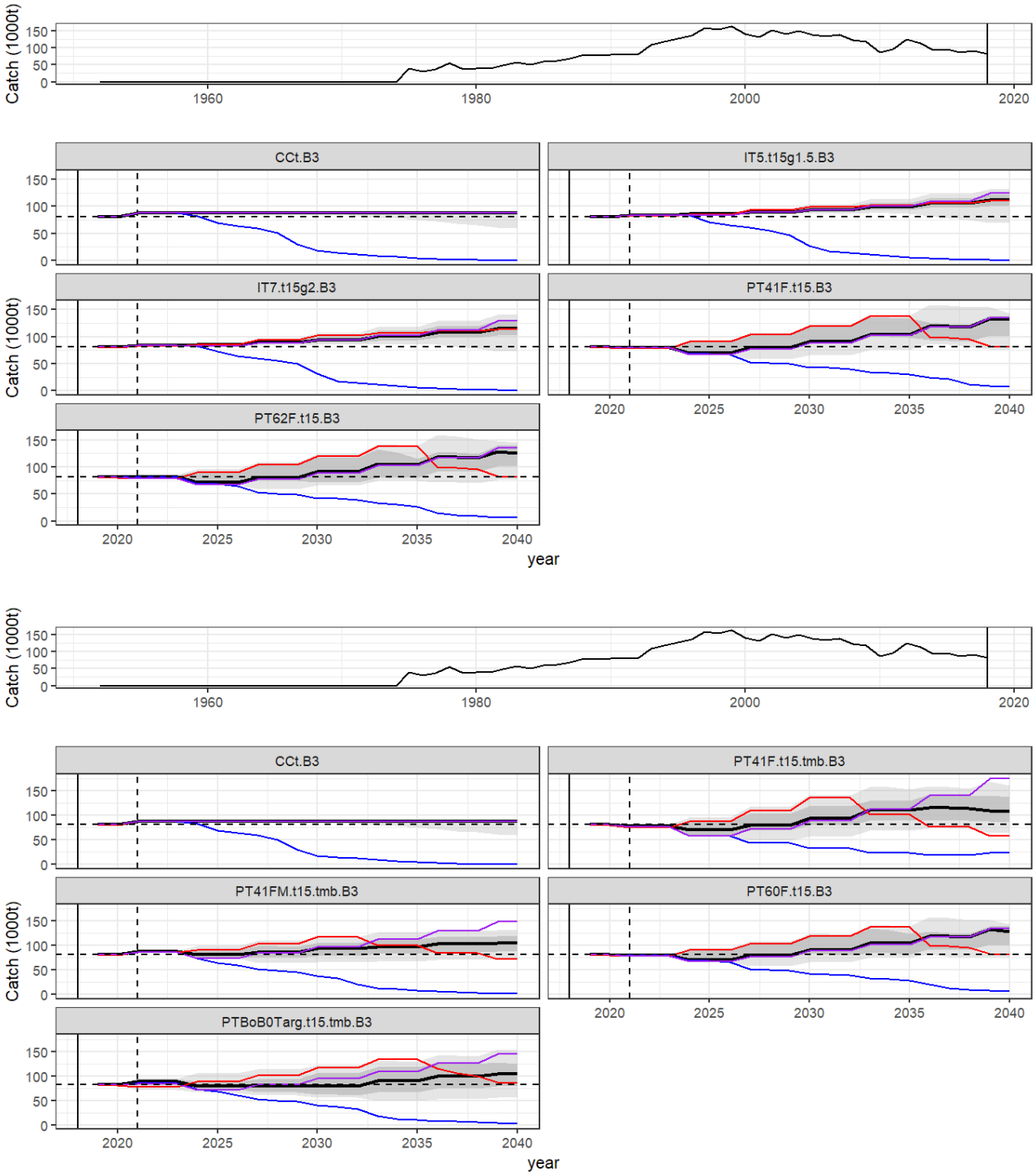

Figure 15. Time series of catch for the candidate (B3 tuned) MPs. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2017) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

4 Bigeye Robustness Test Management Procedure Evaluation results

The 6 robustness tests listed in Table 2 were conducted for a representative subset of 3 MPs (flagged in Table 1, and including PT41F.t15.tmb, which we consider to be the most promising), for both the B2 and B3 tuning objectives.

The 15 year aggregate performance graphics are shown for robustness tests 1-6 (and the reference set) from Table 2 are summarized in Figure 16 - Figure 33, with the corresponding time series plots shown in Figure 34 - Figure 58. None of the Robustness tests had a catastrophic effect on MP performance, and performance was usually degraded in a way that was qualitatively predictable:

- Elevated CPUE CV = 30% (ICV30) tended to have a trivial effect on median performance relative to the reference set OM (base), but increased variability.
- There was not much difference between any of the three 10% overcatch scenarios, and all represented a somewhat increased conservation risk relative to the reference set OM
- The recruitment shock scenario had an adverse impact on catch and increased conservation risk relative to the reference OM. The model-based MPs appeared to recover more effectively than the CPUE-based MP.
- The 3% per year CPUE catchability trend had the expected impact on the model-based MPs, i.e. elevated catch and increased biomass risk relative to the reference set.

Curiously, the time-aggregated results (Figure 16) suggest that the adverse effect of the 3% per year CPUE catchability trend appears much less evident for the CPUE-based MP than the model-based MPs. It is not obvious why the CPUE-based MP should be more robust to this problem, and the time series plots (e.g. Figure 46) do indicate the expected increase in overfishing in the last 5-10 years of the projection period.

Table 2. Operating Models discussed in text - a more comprehensive definition of the reference set OM is provided in Appendix B.

OM Ensemble	
OMrefB20.1	Reference set 500 realizations sampled with replacement from an equally-weighted 72 model fractional factorial grid described in Appendix B
1. ICV30	A robustness test in which the information content of the projected Longline CPUE is reduced (spatially-aggregated annual σ_I = 0.30, auto-correlation = 0.5)
2. 10% ROC	A robustness scenario in which every fishery has a 10% over-catch implementation error, with accurate catch reporting
3. 10% UCC	A robustness scenario in which every fishery has a 10% over-catch implementation error, that is not reported
4. 5% ROC, 5% UCC	A robustness scenario in which every fishery has a 10% over-catch implementation error, exactly half of which is reported.
5. 3% qTrend	A robustness scenario in which there is a 3% per year LL CPUE catchability trend starting in the projections (conditioning unchanged from the reference case)
6. Rec Shock	A robustness scenario with 8 consecutive quarters of poor recruitment (55% of expected values, similar to estimates for YFT in the early 2000s). (conditioning and sampling is unchanged from OMrefB20.1.500)
7. R	As the reference set, except that the original R-based Pope's approximation is used for the catch equations
8. Cpp EC2	As the reference set, except the "Effort Ceiling" for the Baranov equations is capped at 2.0 (i.e. the fishing mortality for each fleet can no more than double relative to the average of the values estimated in 2016 and 2017). (Reference set Effort Ceiling = 20)
9. Cpp EC2, uniform	As the reference set, except fish of all ages are uniformly redistributed among regions each quarter.

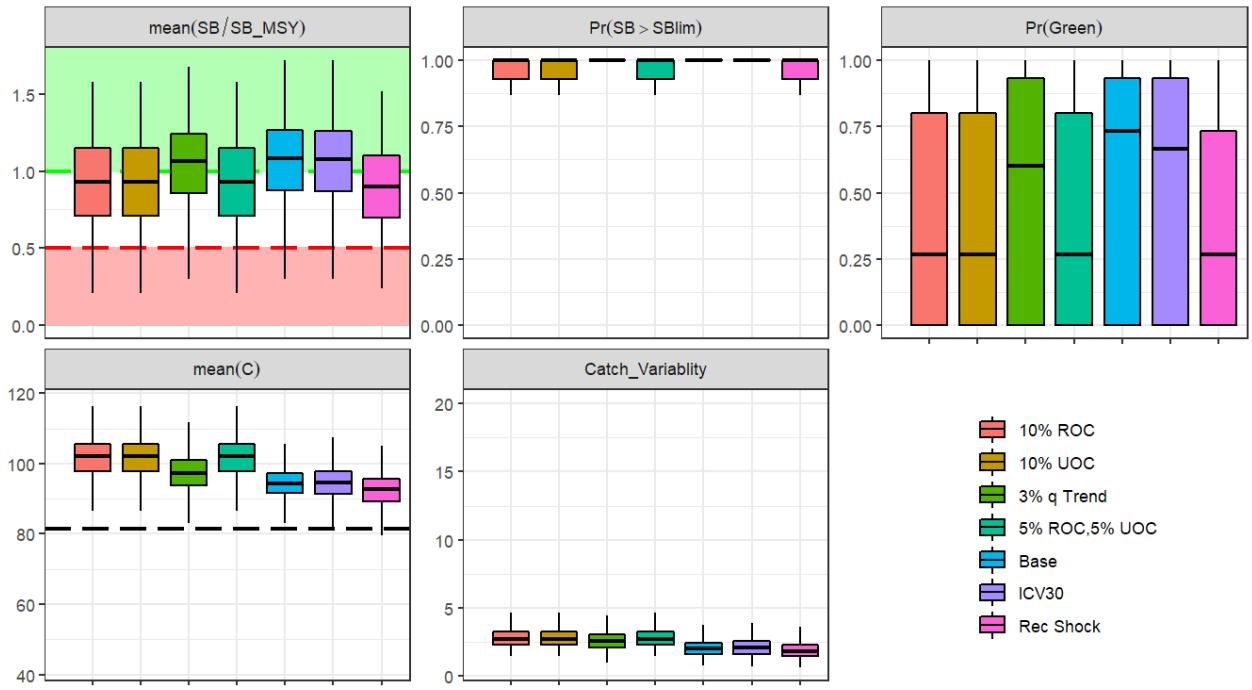

Figure 16. Boxplots comparing various robustness tests versus key performance measures averaged over the period 2021 -2035 for reference set MP IT5.t15g1.5.B2. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2020 catch.

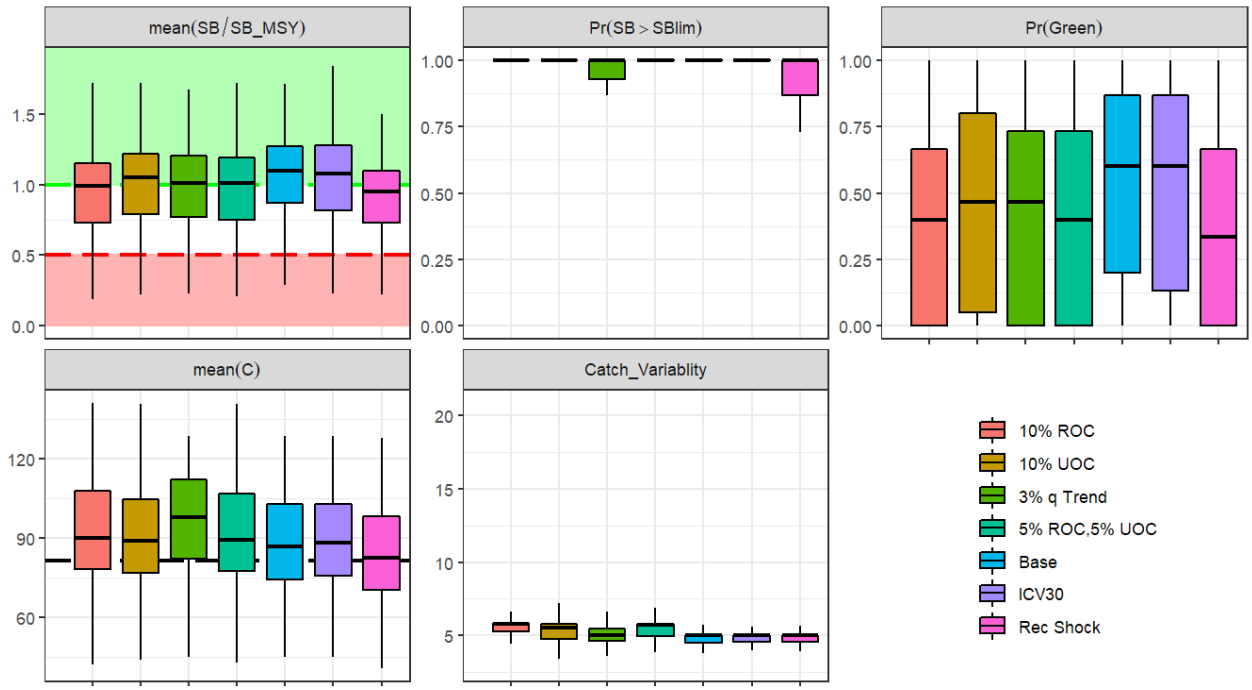

Figure 17. Boxplots comparing various robustness tests versus key performance measures averaged over the period 2021 -2035 for reference set MP PTBoB0Targ.t15.tmb.B2. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2020 catch.

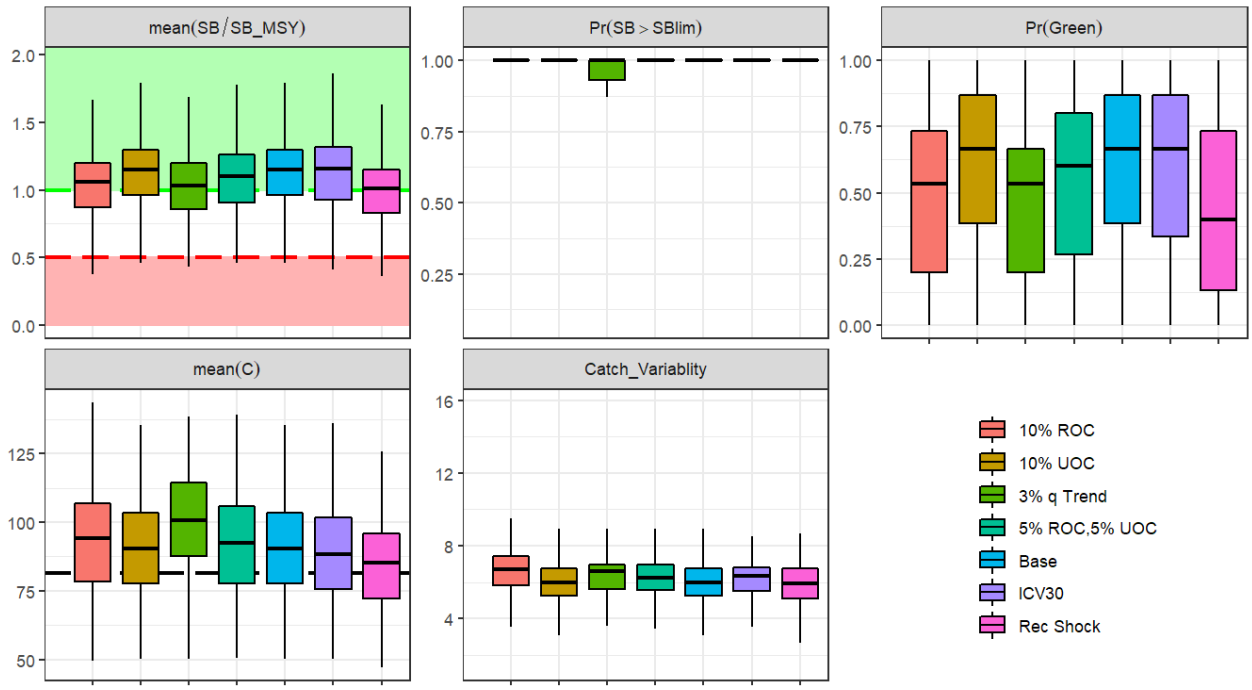

Figure 18. Boxplots comparing various robustness tests versus key performance measures averaged over the period 2021 -2035 for reference set MP PT41F.t15.tmb.B2. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2020 catch.

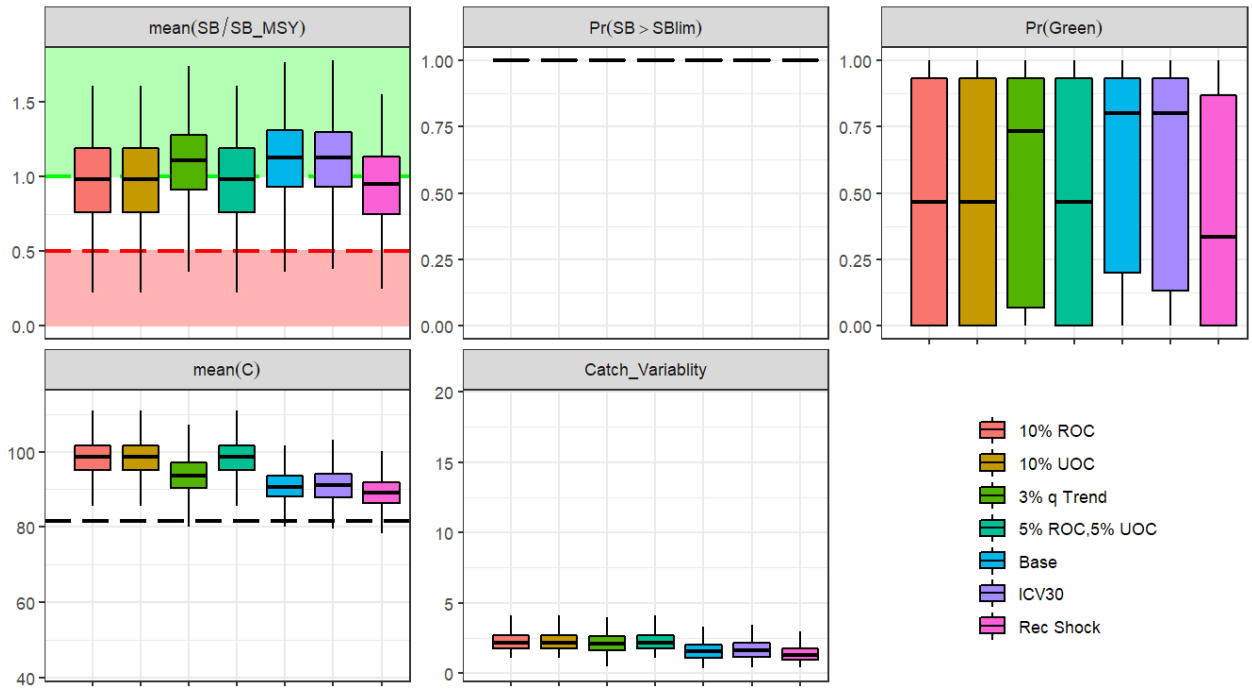

Figure 19. Boxplots comparing various robustness tests versus key performance measures averaged over the period 2021 -2035 for reference set MP IT5.t15g1.5.B3. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2020 catch.

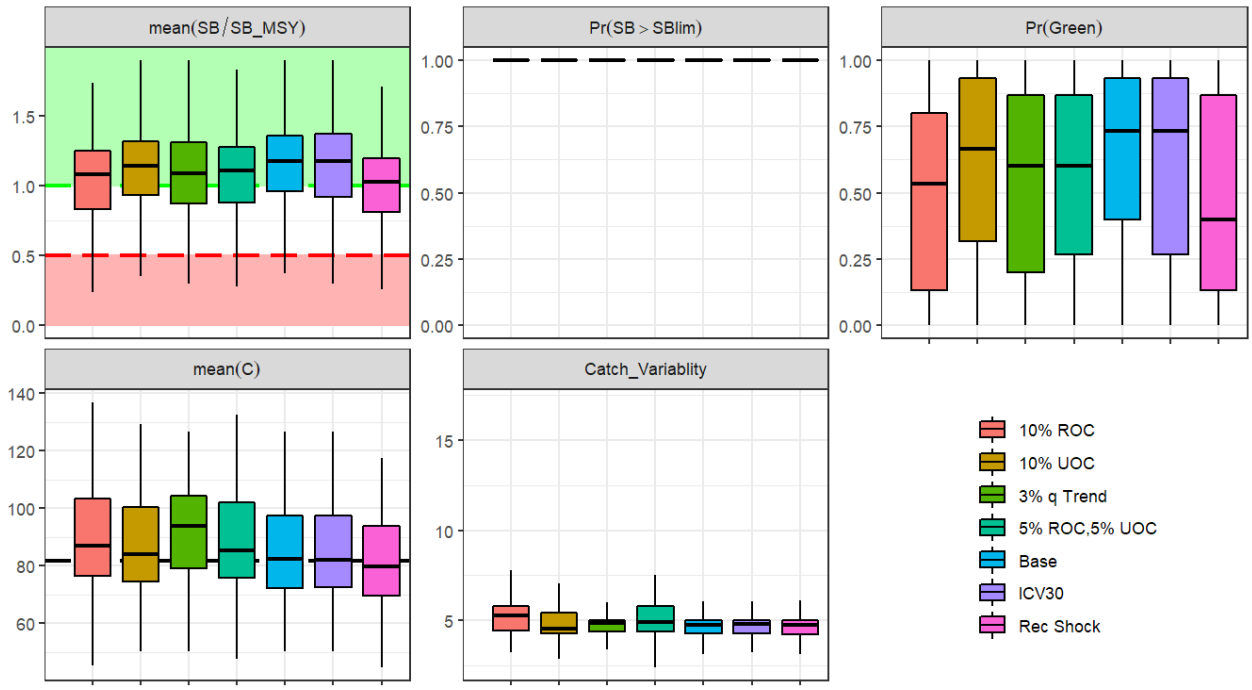

Figure 20. Boxplots comparing various robustness tests versus key performance measures averaged over the period 2021 -2035 for reference set MP PTBoB0Targ.t15.tmb.B3. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2020 catch.

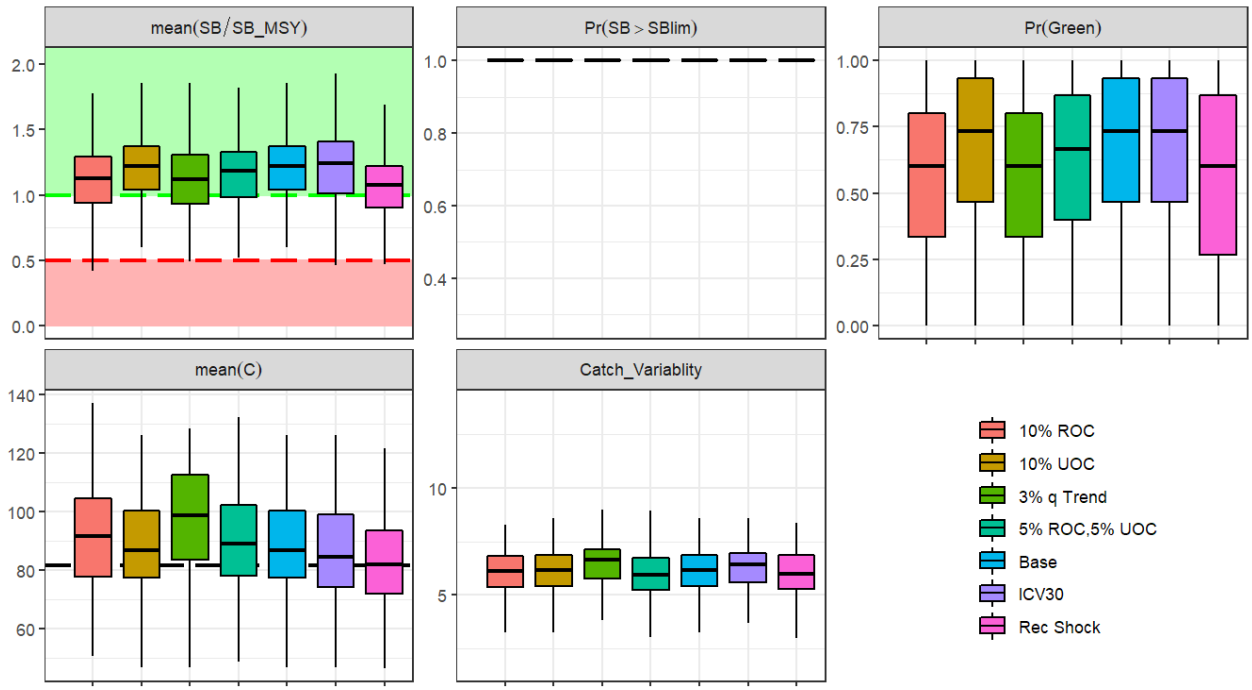

Figure 21. Boxplots comparing various robustness tests versus key performance measures averaged over the period 2021 -2035 for reference set MP PT41F.t15.tmb.B3. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2020 catch.

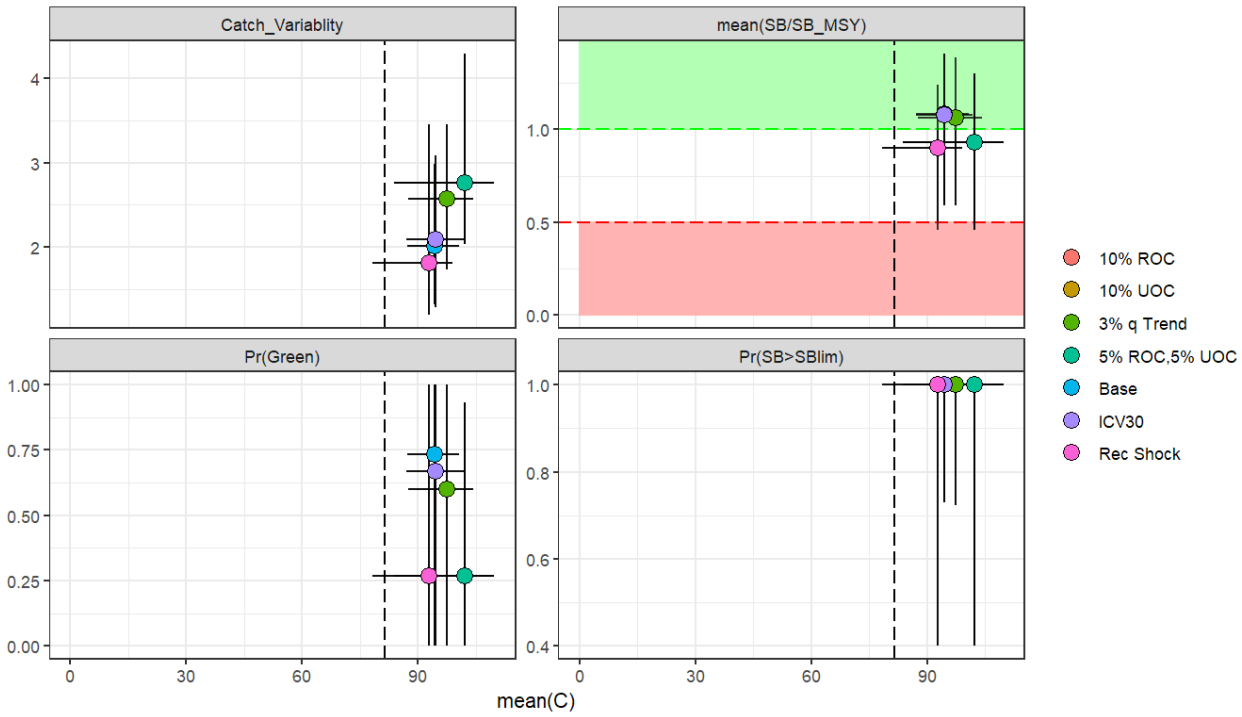

Figure 22. Trade-off plots comparing various robustness tests for reference set MP IT5.t15g1.5.B2 with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2021 - 2035. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2017 catch.

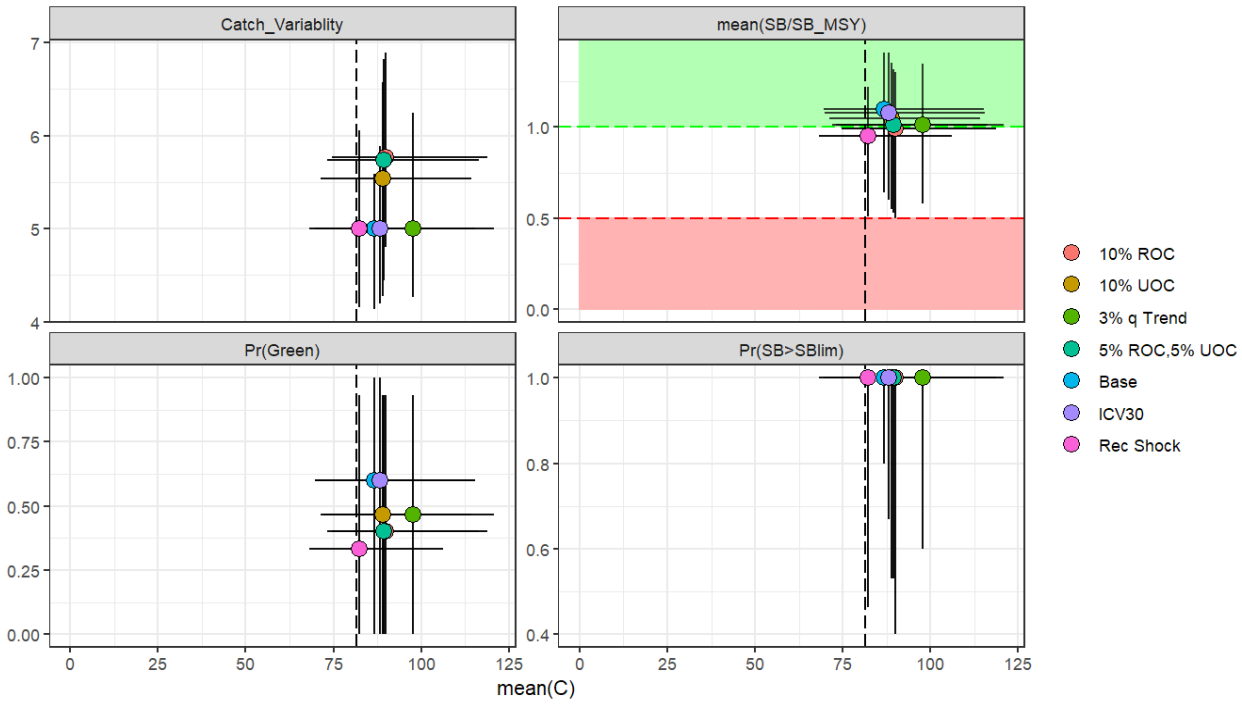

Figure 23. Trade-off plots comparing various robustness tests for reference set MP PTBoB0Targ.t15.tmb.B2 with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2021 - 2035. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2017 catch.

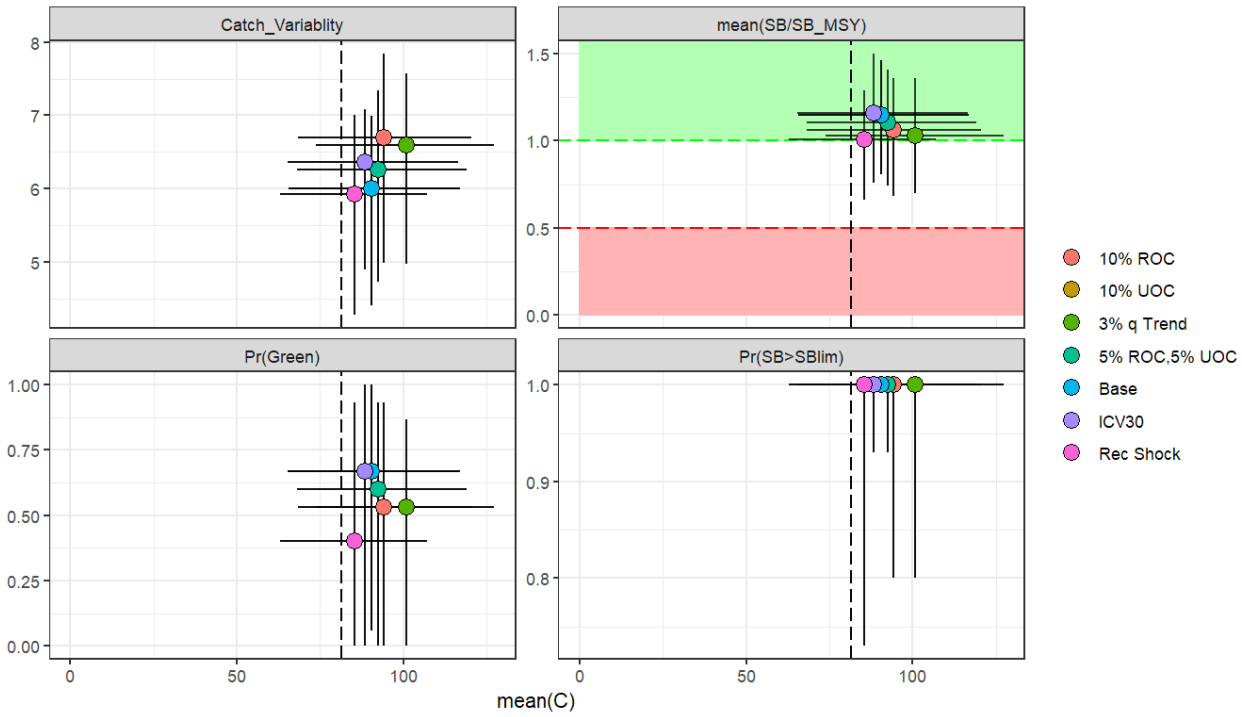

Figure 24. Trade-off plots comparing various robustness tests for reference set MP PT41F.t15.tmb.B2 with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2021 - 2035. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2017 catch.

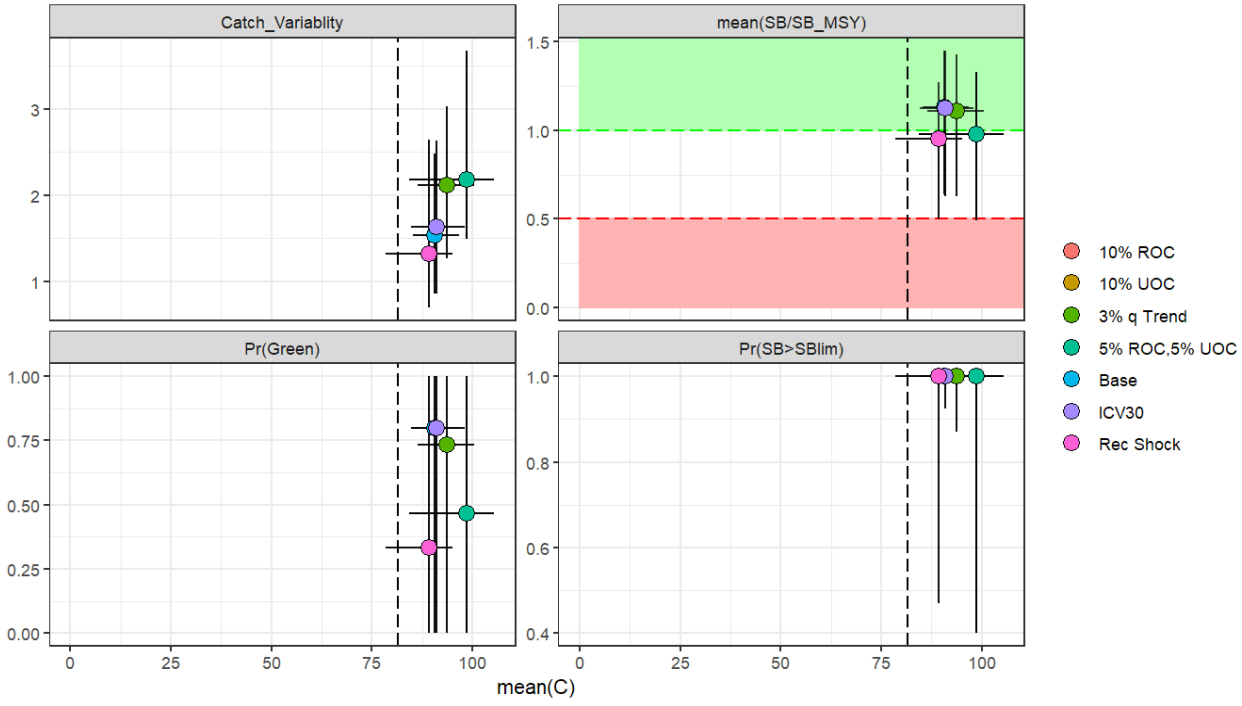

Figure 25. Trade-off plots comparing various robustness tests for reference set MP IT5.t15g1.5.B3 with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2021 - 2035. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2017 catch.

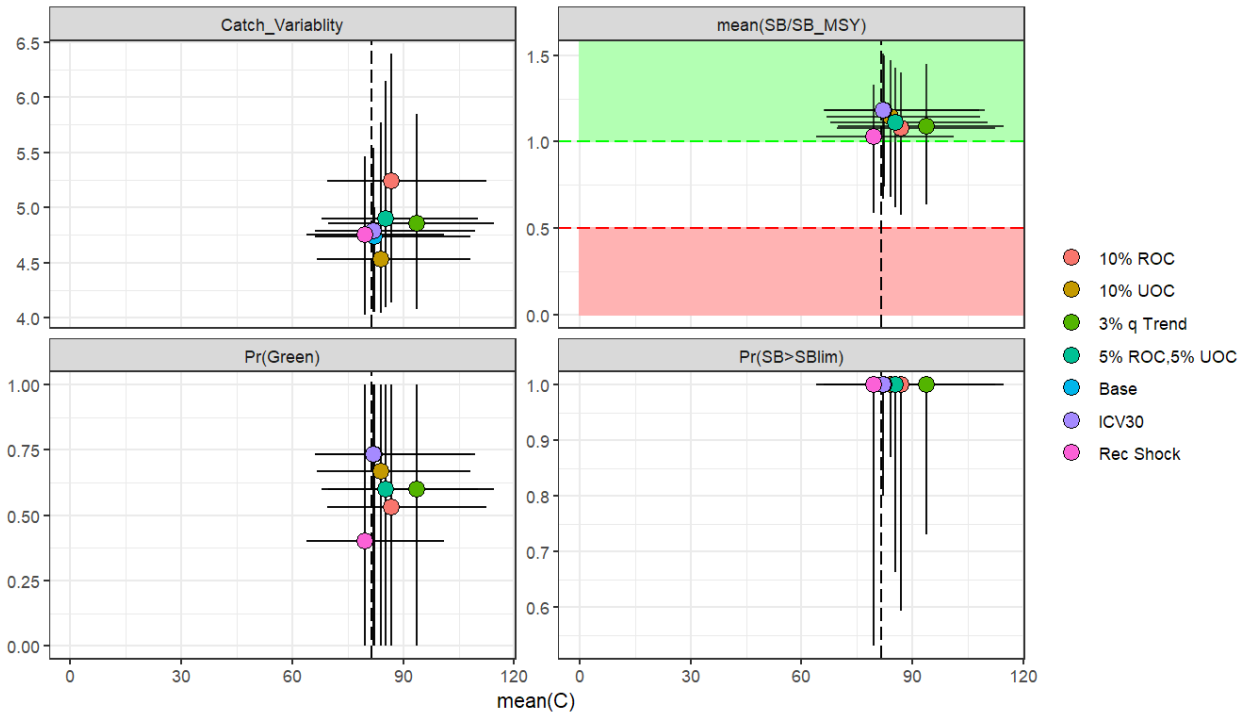

Figure 26. Trade-off plots comparing various robustness tests for reference set MP PTBoB0Targ.t15.tmb.B3 with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2021 - 2035. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2017 catch.

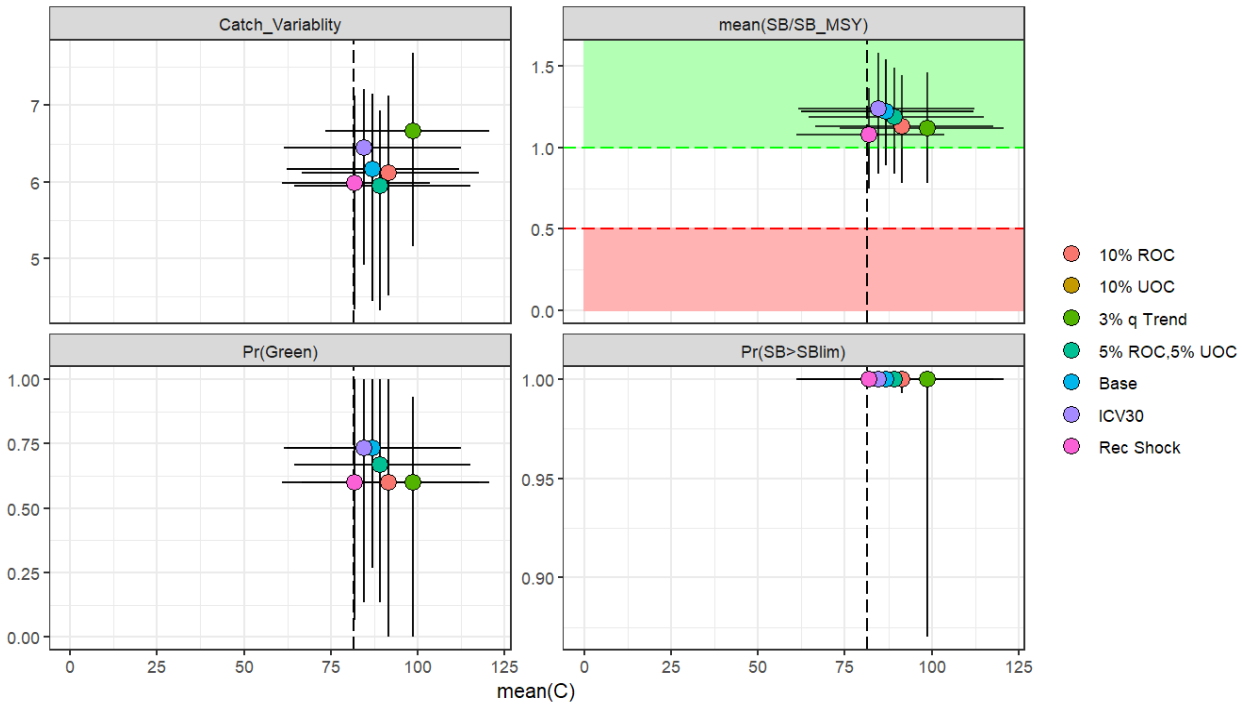

Figure 27. Trade-off plots comparing various robustness tests for reference set MP PT41F.t15.tmb.B3 with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2021 - 2035. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2017 catch.

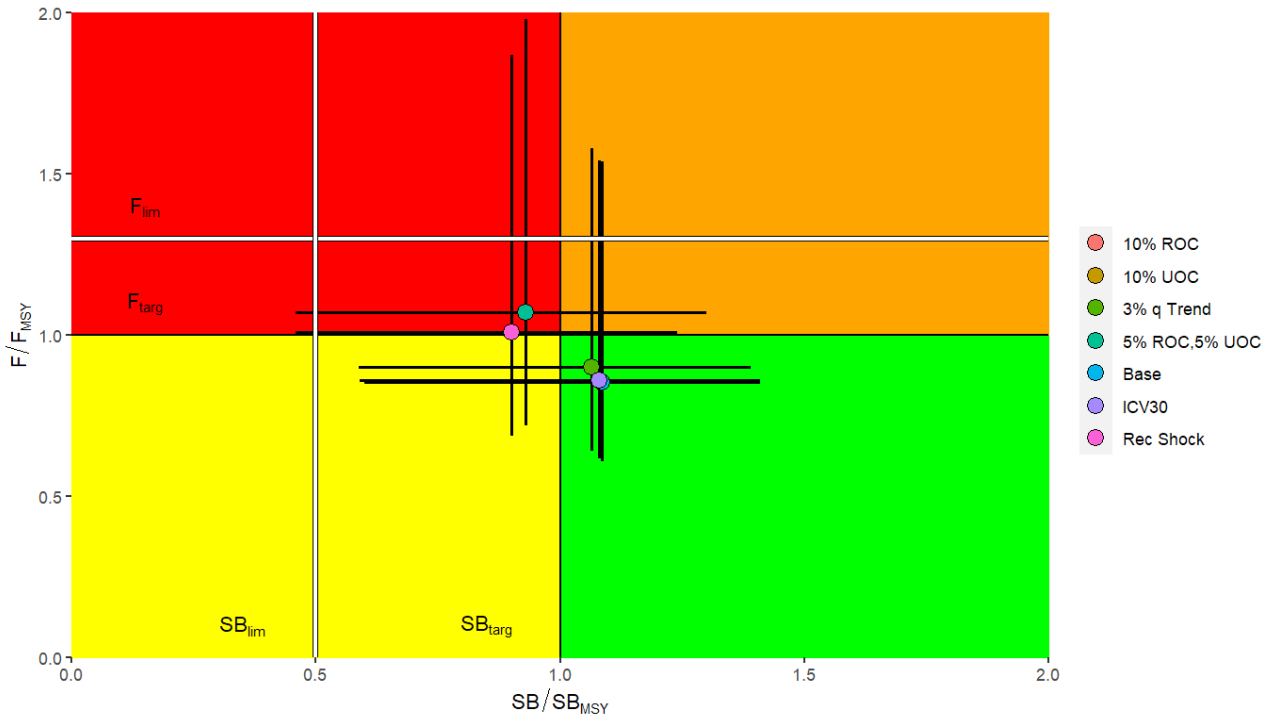

Figure 28. Kobe plot comparing various robustness tests for reference set MP IT5.t15g1.5.B2 on the basis of the expected 15 year average (2021-2035) performance. Circle is the median, lines represent 10th-90th percentiles.

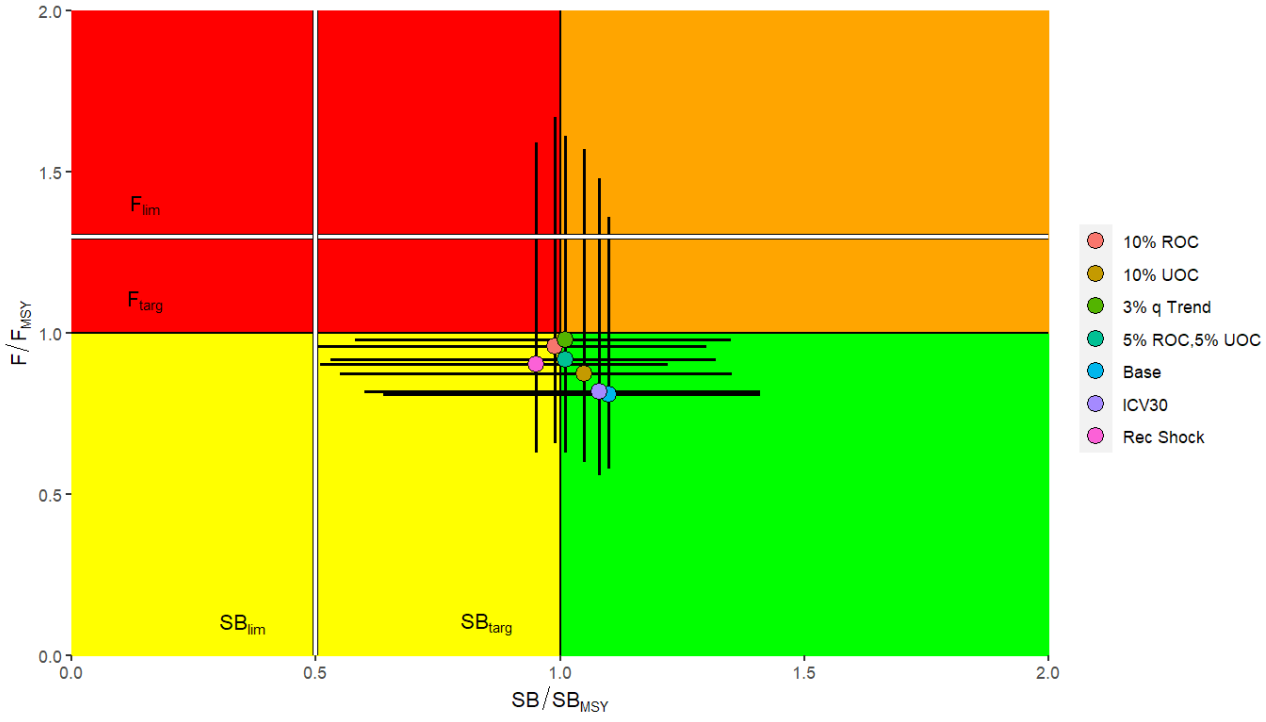

Figure 29. Kobe plot comparing various robustness tests for reference set PTBoB0Targ.t15.tmb.B2 on the basis of the expected 15 year average (2021-2035) performance. Circle is the median, lines represent 10th-90th percentiles.

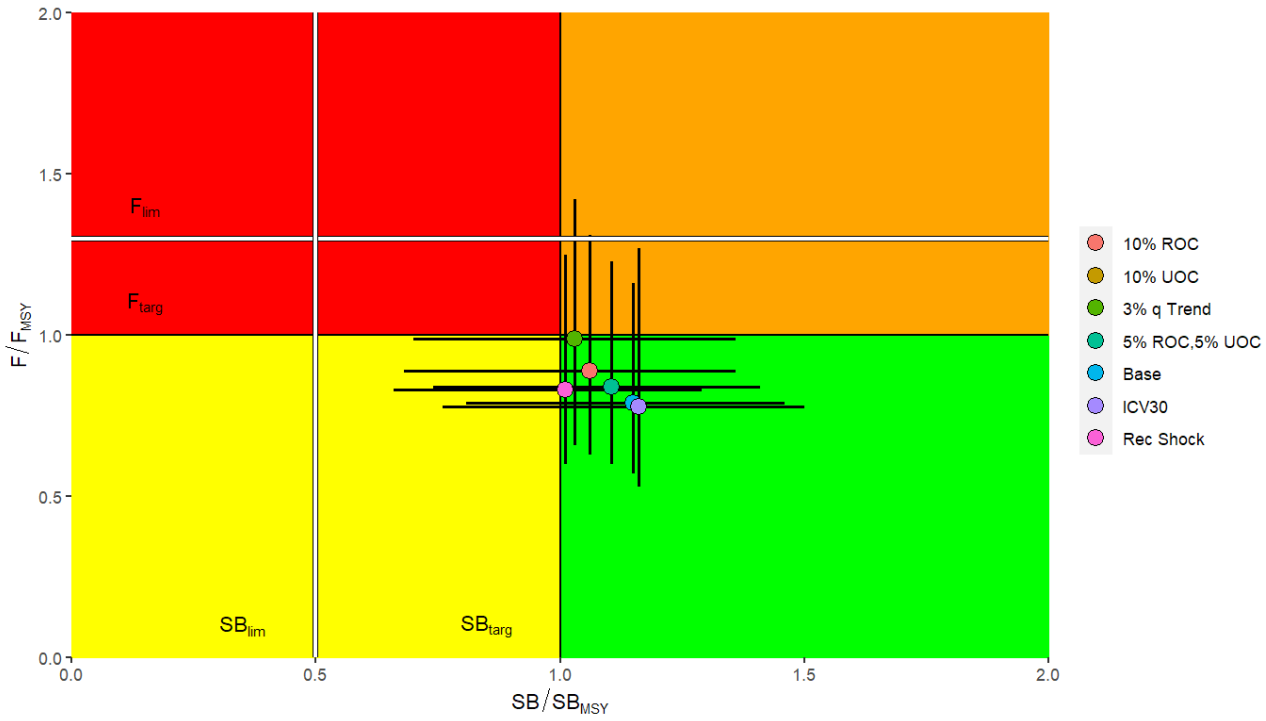

Figure 30. Kobe plot comparing various robustness tests for reference set MP PT41F.t15.tmb.B2 on the basis of the expected 15 year average (2021-2035) performance. Circle is the median, lines represent 10th-90th percentiles.

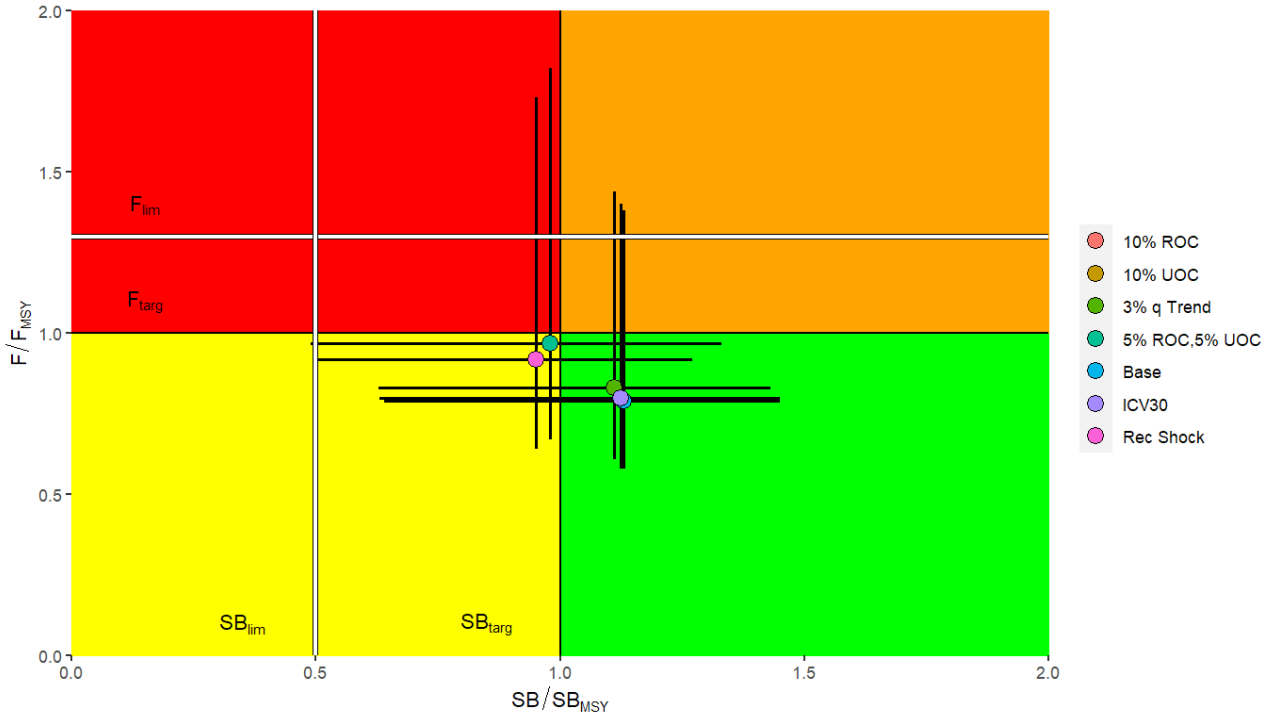

Figure 31. Kobe plot comparing various robustness tests for reference set MP IT5.t15g1.5.B3 on the basis of the expected 15 year average (2021-2035) performance. Circle is the median, lines represent 10th-90th percentiles.

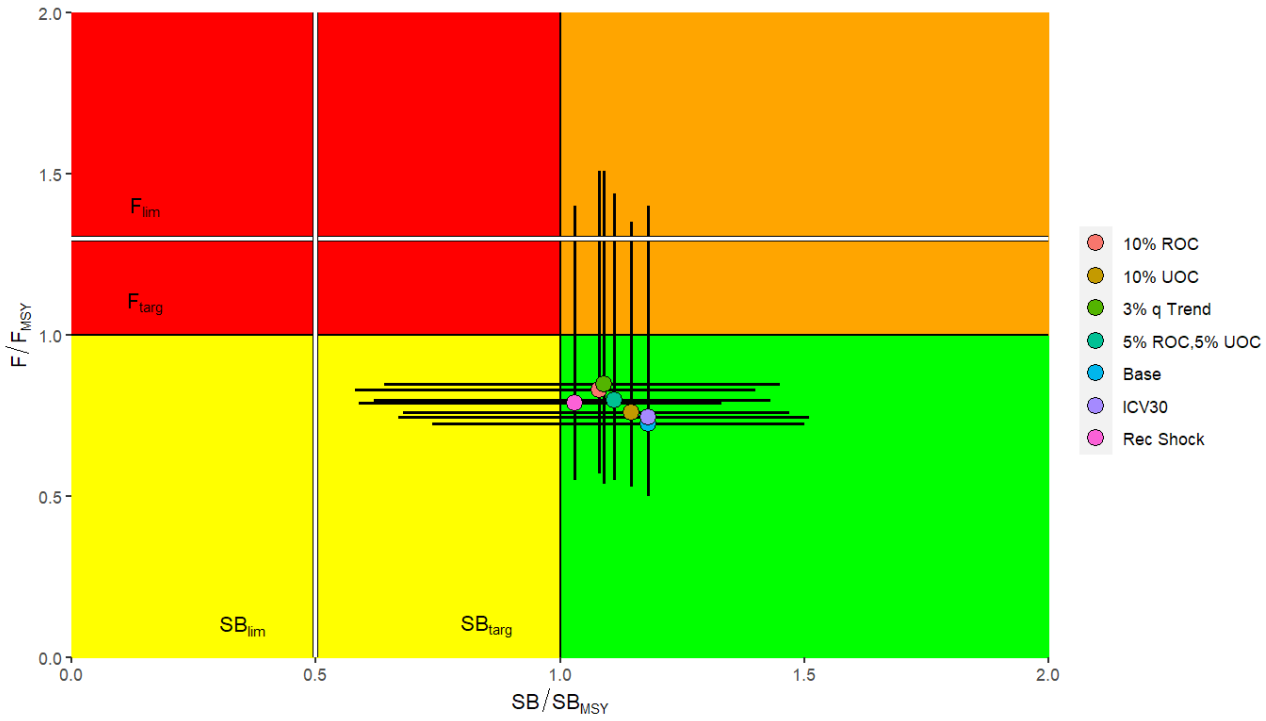

Figure 32. Kobe plot comparing various robustness tests for reference set MP PTBoB0Targ.t15.tmb.B3 on the basis of the expected 15 year average (2021-2035) performance. Circle is the median, lines represent 10th-90th percentiles.

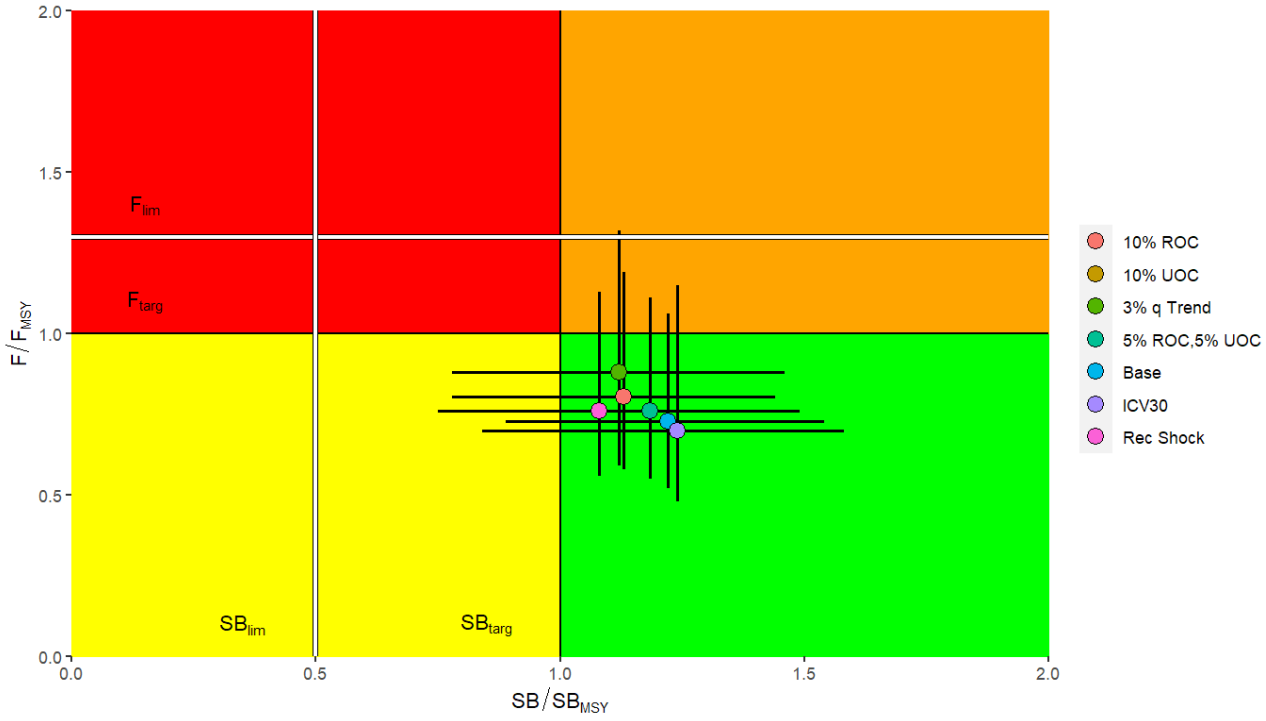

Figure 33. Kobe plot comparing various robustness tests for reference set MP PT41F.t15.tmb.B3 on the basis of the expected 15 year average (2021-2035) performance. Circle is the median, lines represent 10th-90th percentiles.

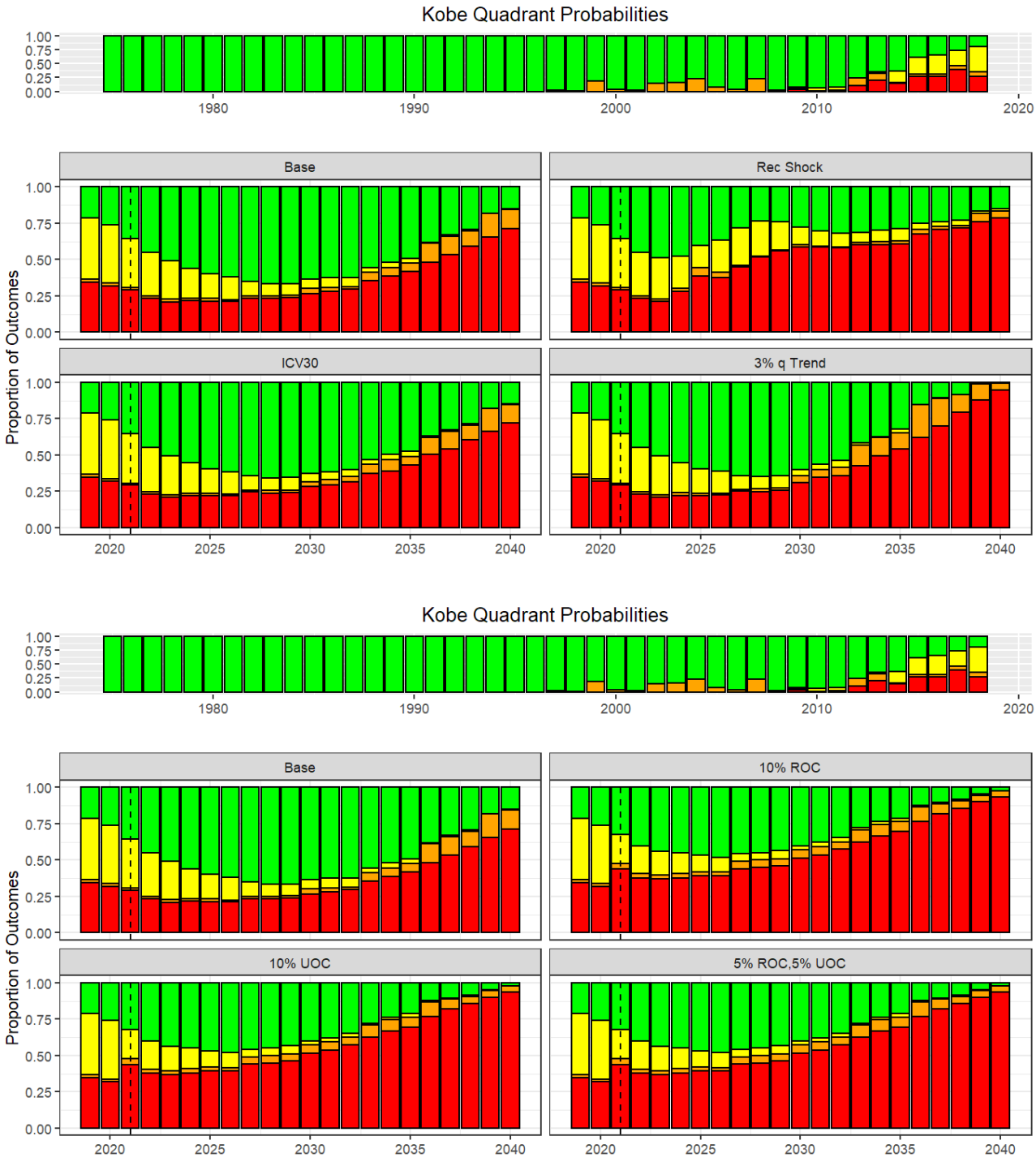

Figure 34. Proportion of simulations in each of the Kobe quadrants over time for various robustness tests and reference set MP IT5.t15g1.5.B2. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

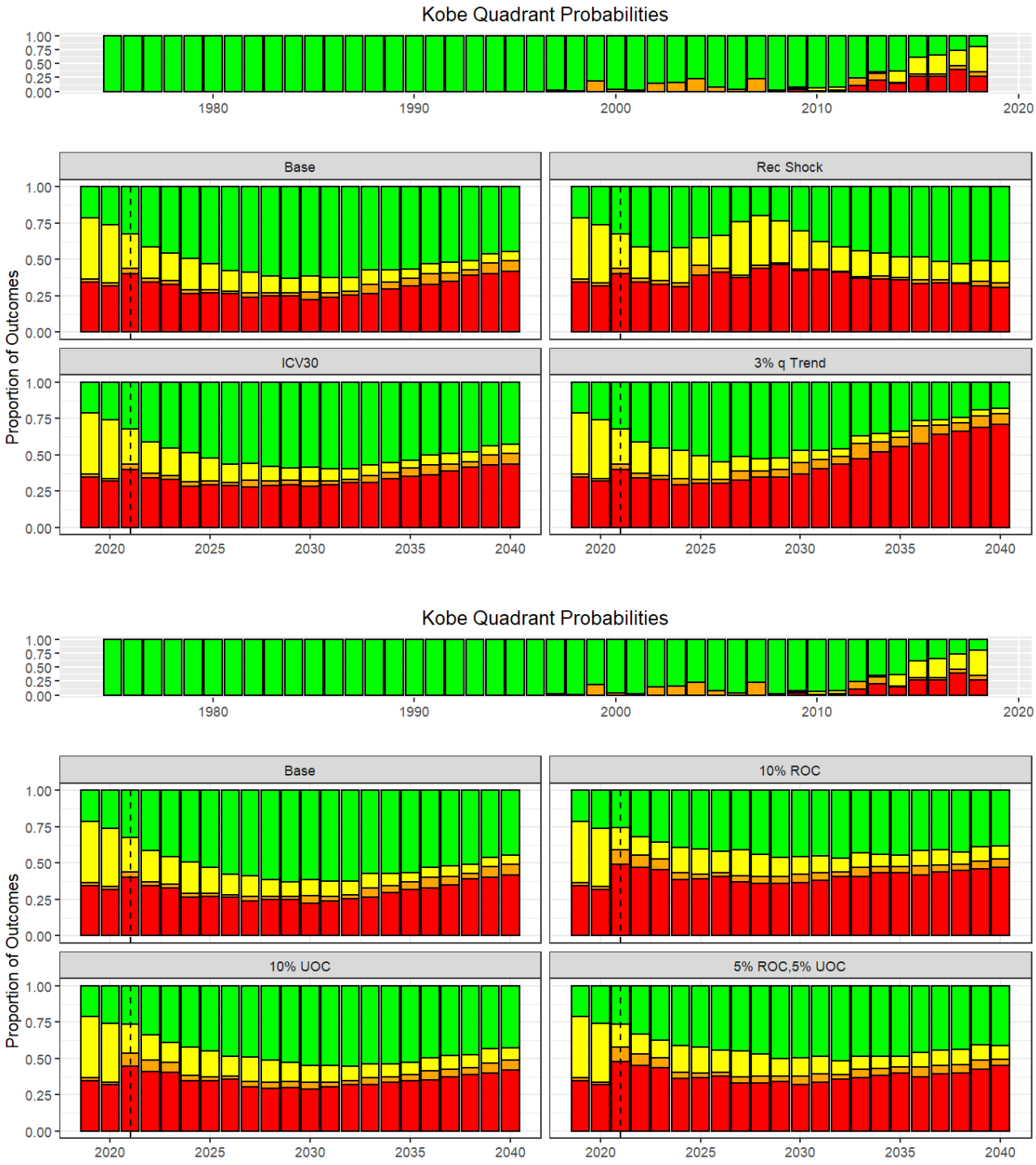

Figure 35. Proportion of simulations in each of the Kobe quadrants over time for various robustness tests and reference set MP PTBoB0Targ.t15.tmb.B2. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

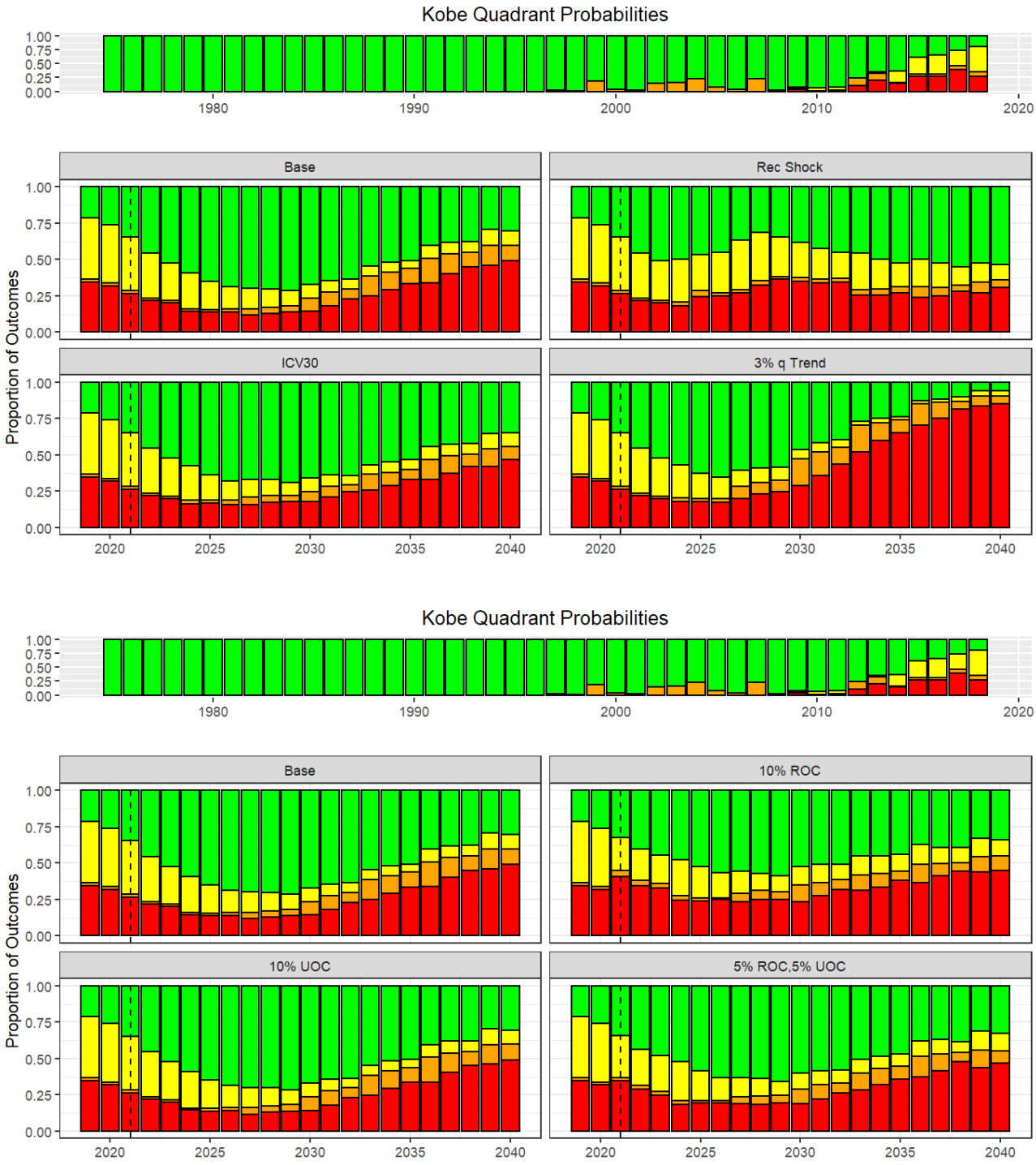

Figure 36. Proportion of simulations in each of the Kobe quadrants over time for various robustness tests and reference set MP PT41F.t15.tmb.B2 Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

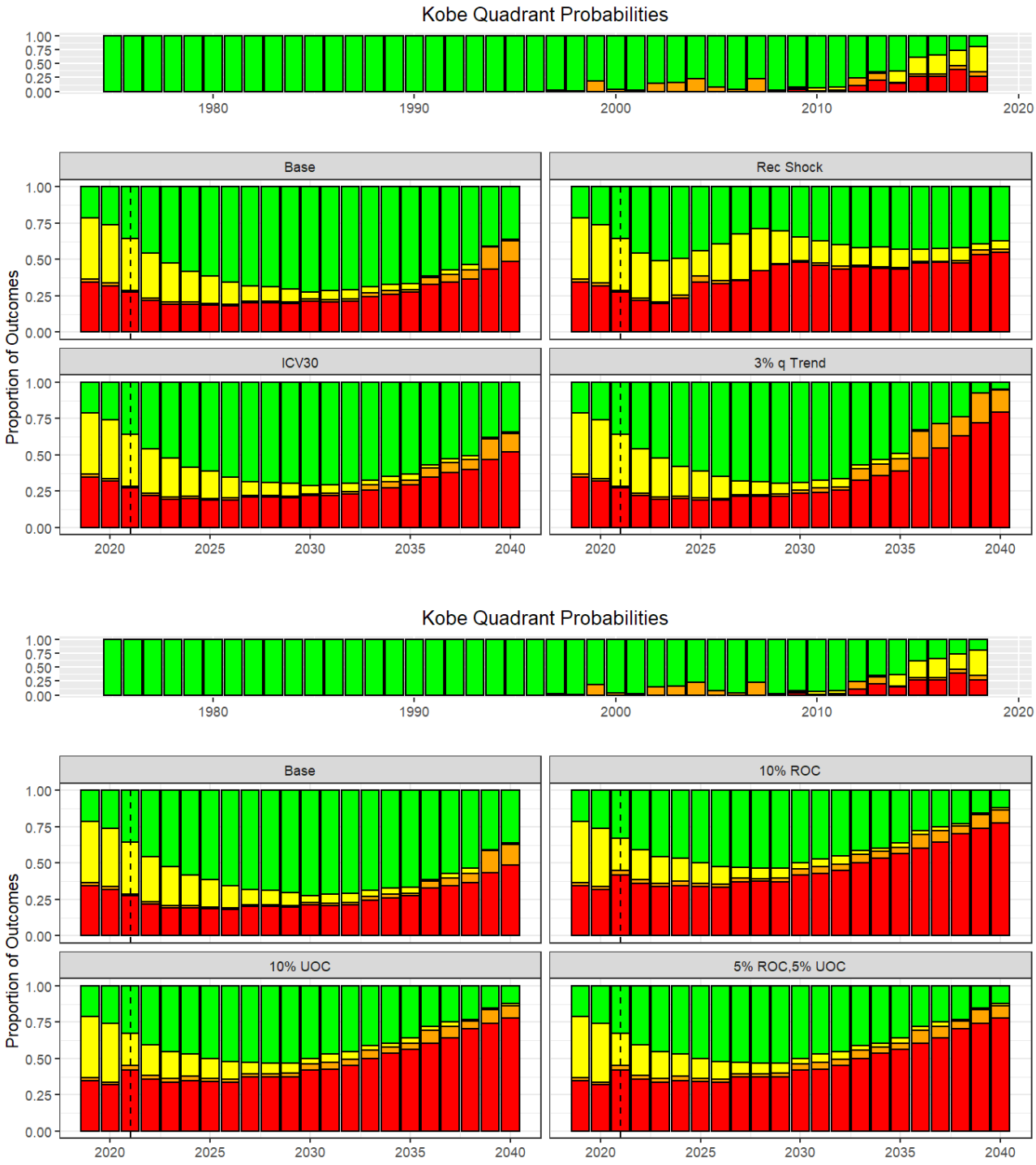

Figure 37. Proportion of simulations in each of the Kobe quadrants over time for various robustness tests and reference set MP IT5.t15g1.5.B3. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

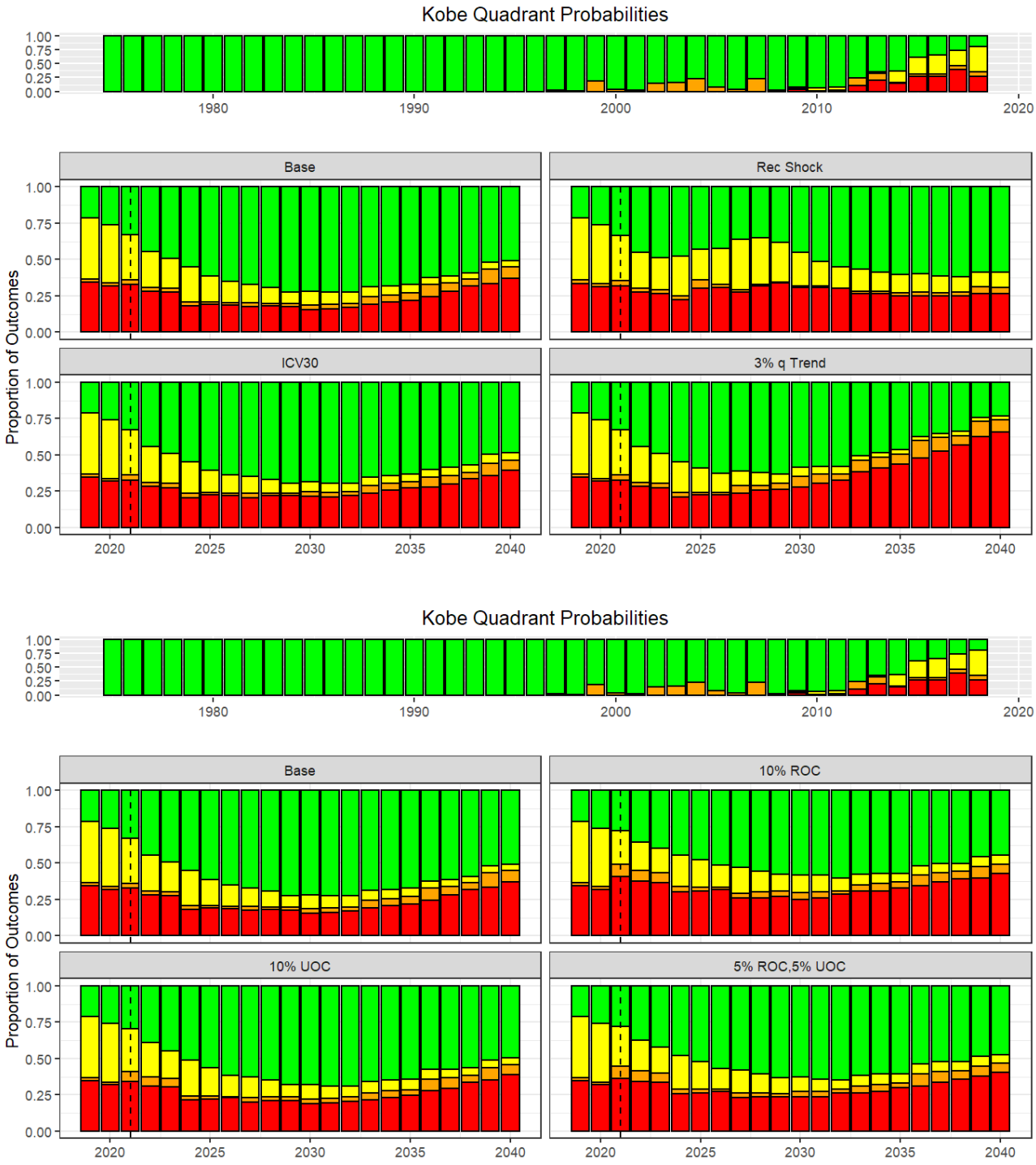

Figure 38. Proportion of simulations in each of the Kobe quadrants over time for various robustness tests and reference set MP PTBoB0Targ.t15.tmb.B3. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

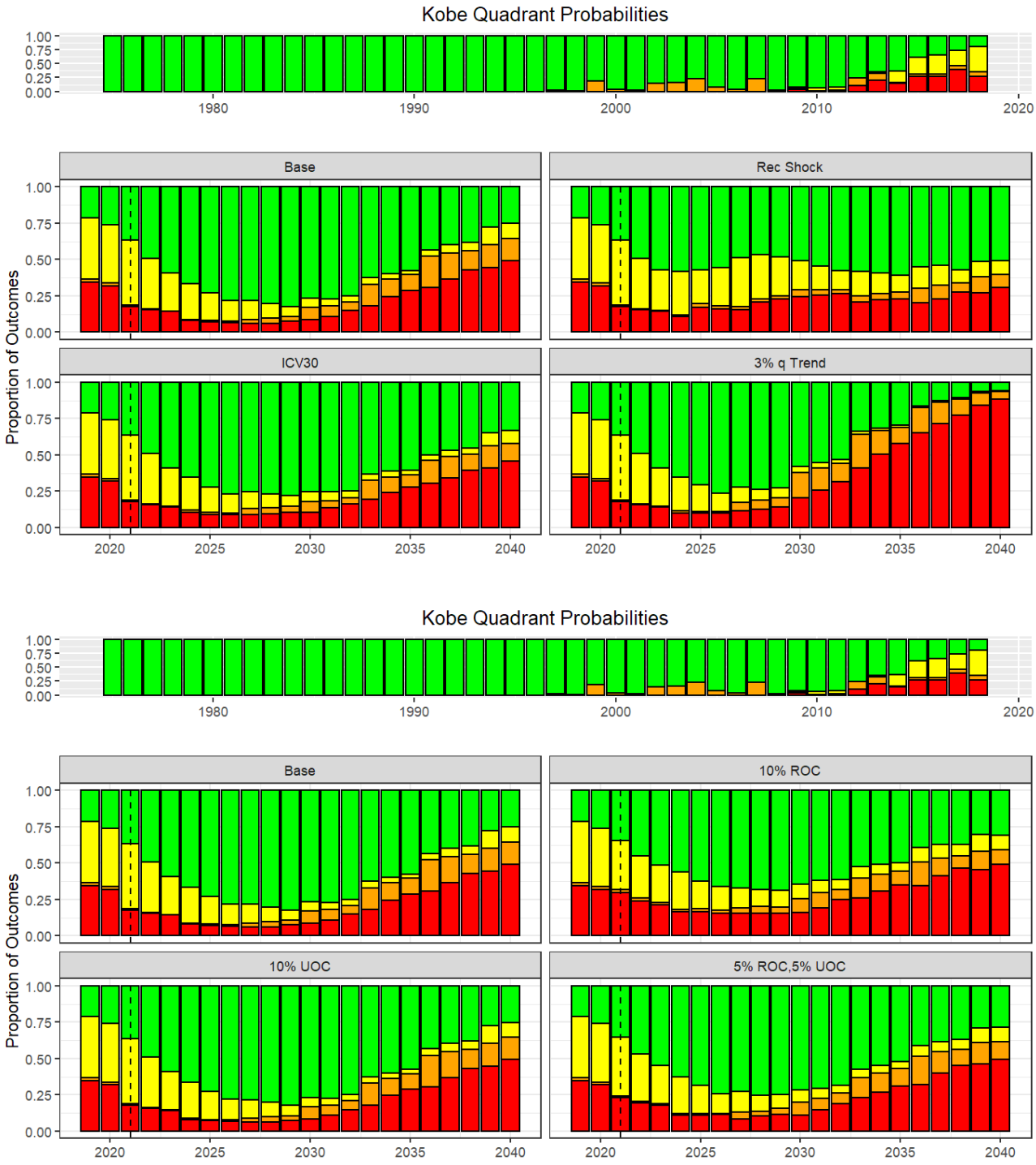

Figure 39. Proportion of simulations in each of the Kobe quadrants over time for various robustness tests and reference set MP PT41F.t15.tmb.B3. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

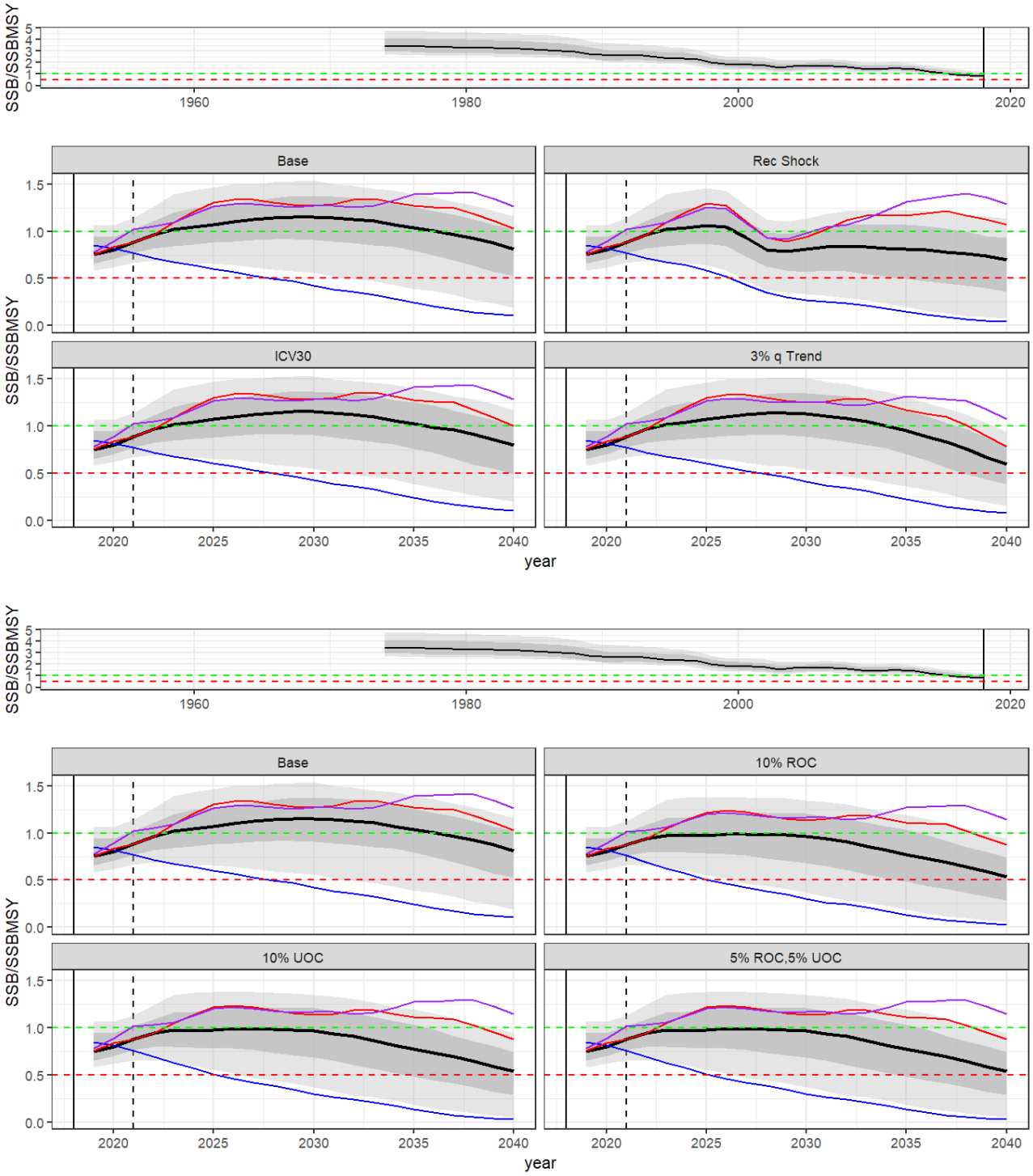

Figure 40. Time series of spawning stock size for various robustness tests and reference set MP IT5.t15g1.5.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

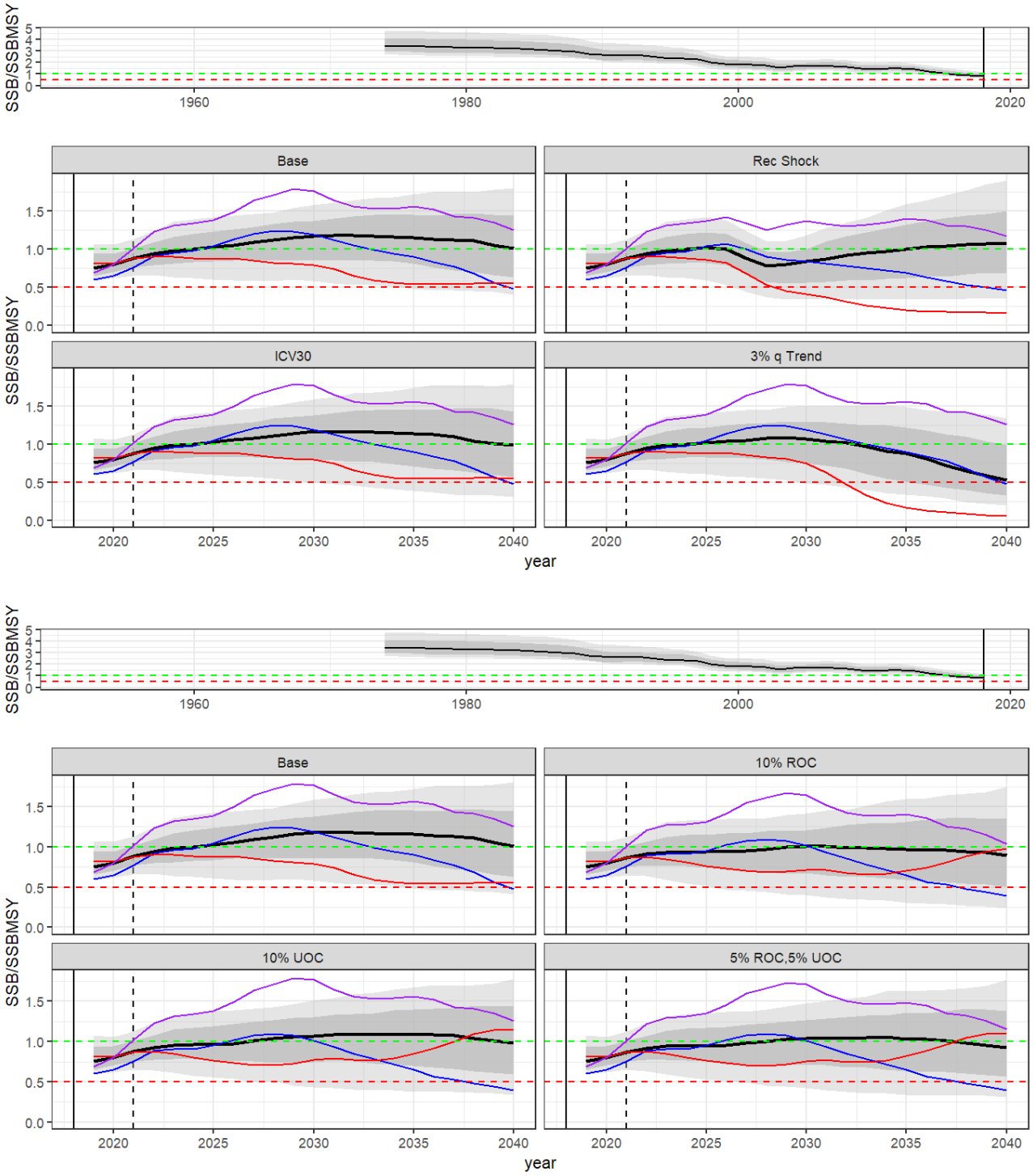

Figure 41. Time series of spawning stock size for various robustness tests and reference set MP PTBoB0Targ.t15.tmb.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

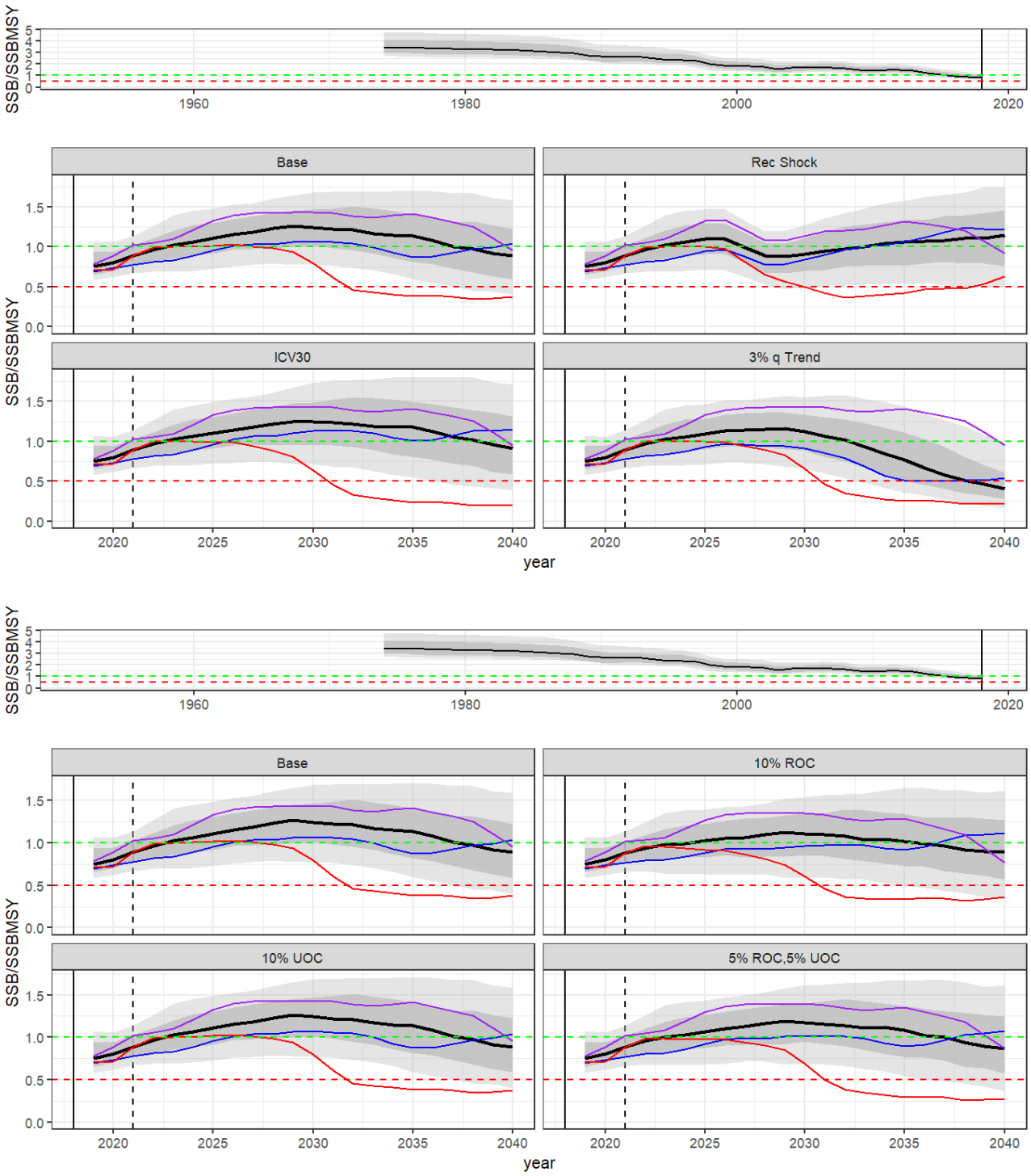

Figure 42. Time series of spawning stock size for various robustness tests and reference set MP PT41F.t15.tmb.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

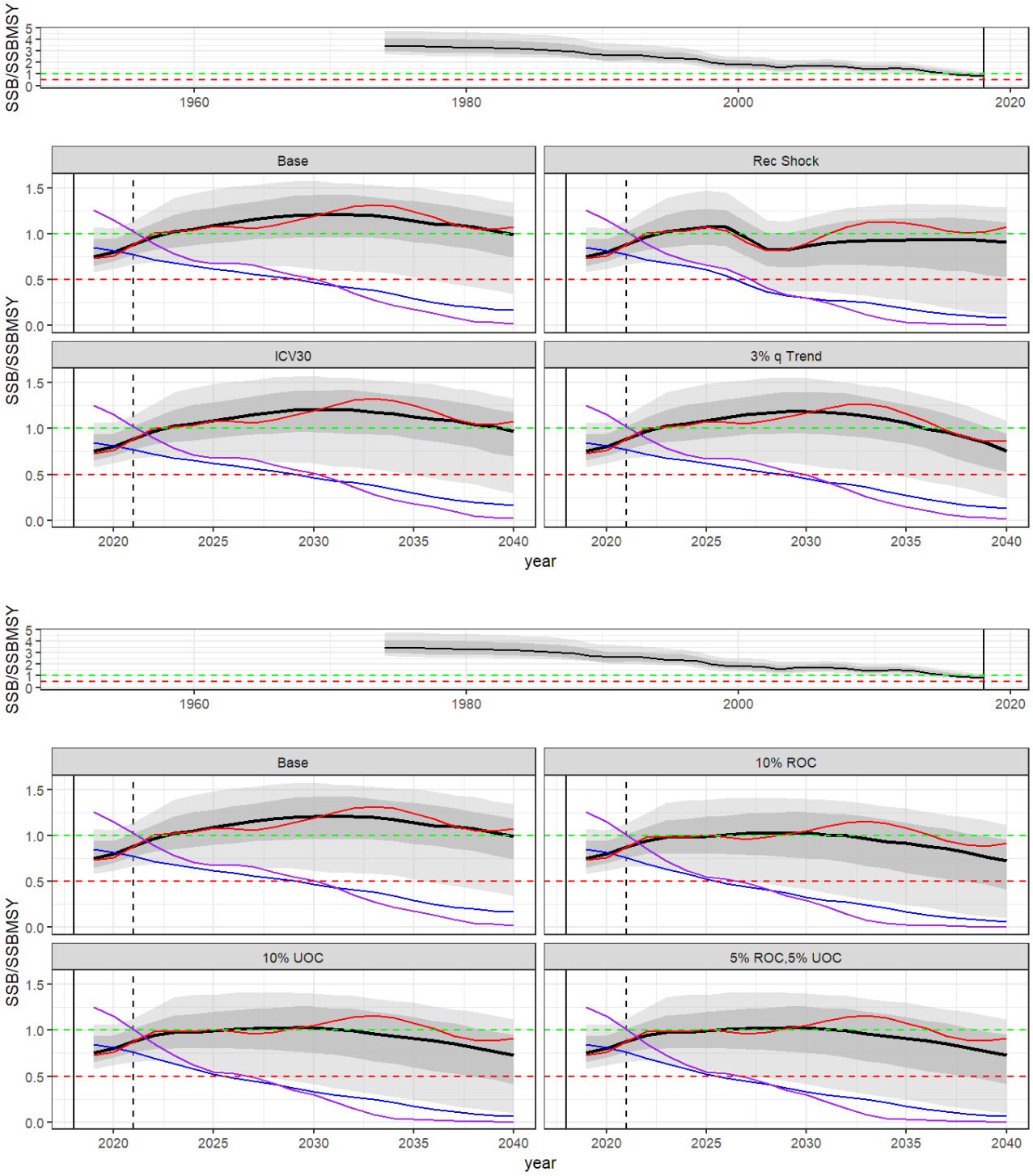

Figure 43. Time series of spawning stock size for various robustness tests and reference set MP IT5.t15g1.5.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

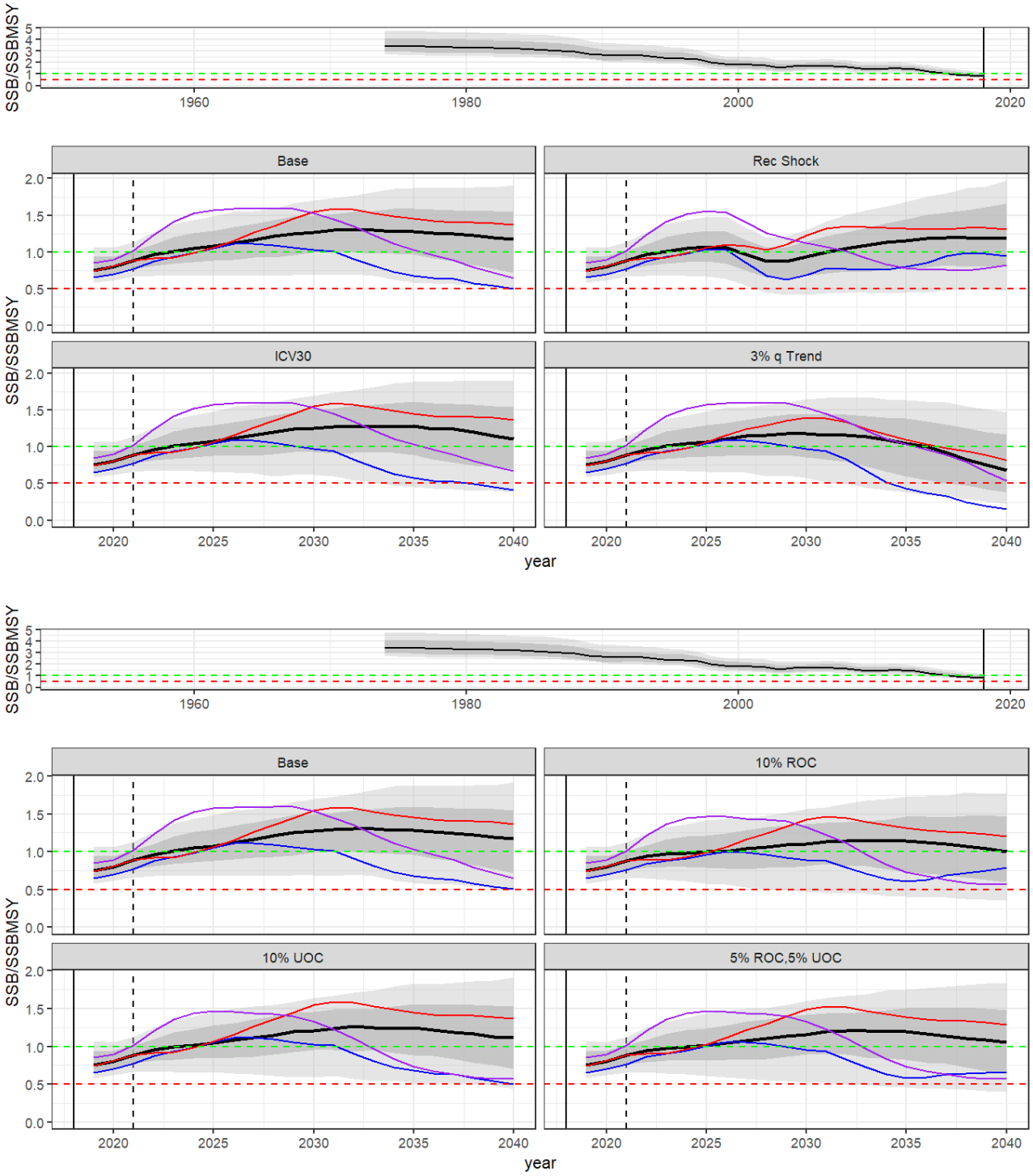

Figure 44. Time series of spawning stock size for various robustness tests and reference set MP PTBoB0Targ.t15.tmb.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

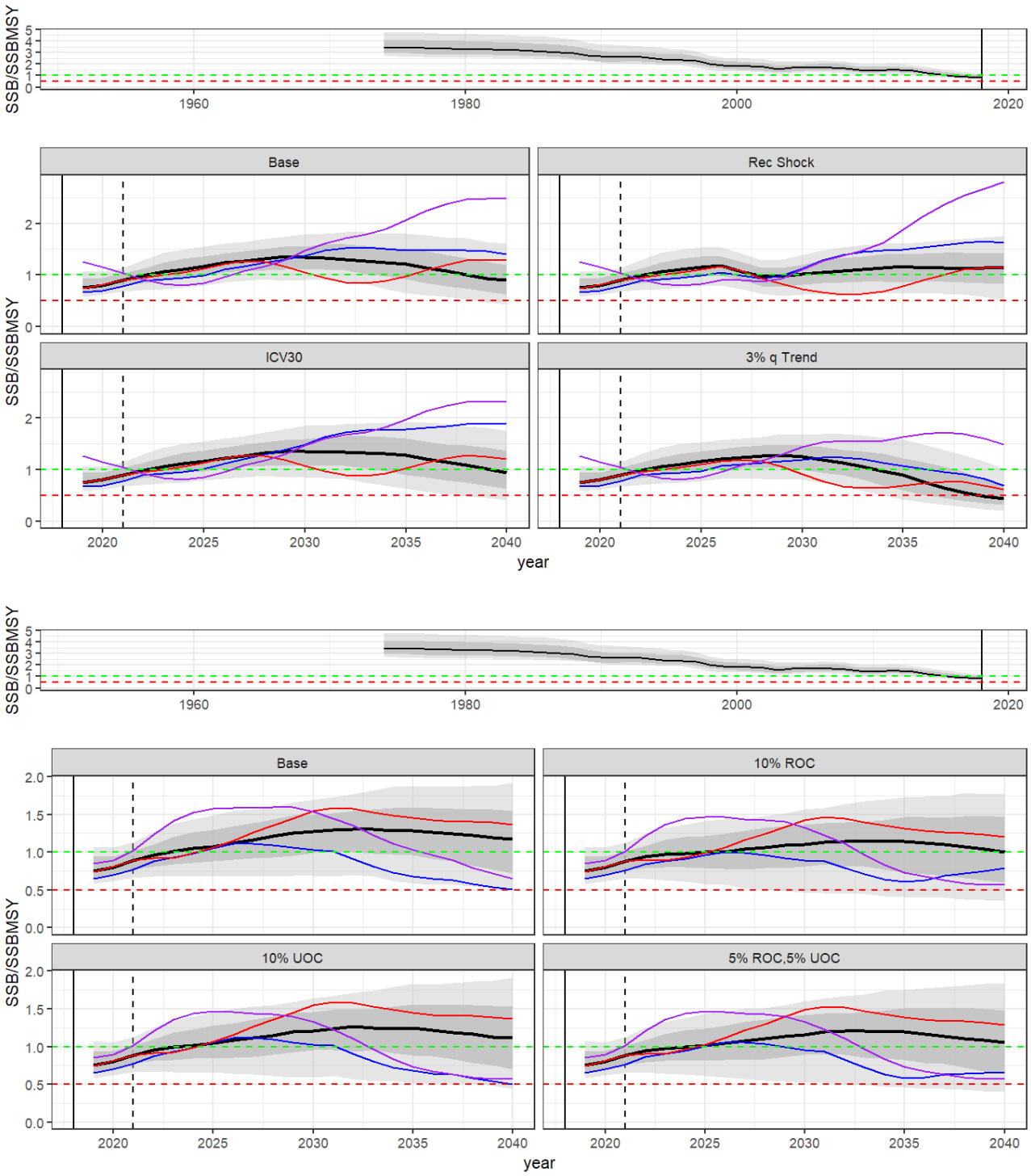

Figure 45. Time series of spawning stock size for various robustness tests and reference set MP PT41F.t15.tmb.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

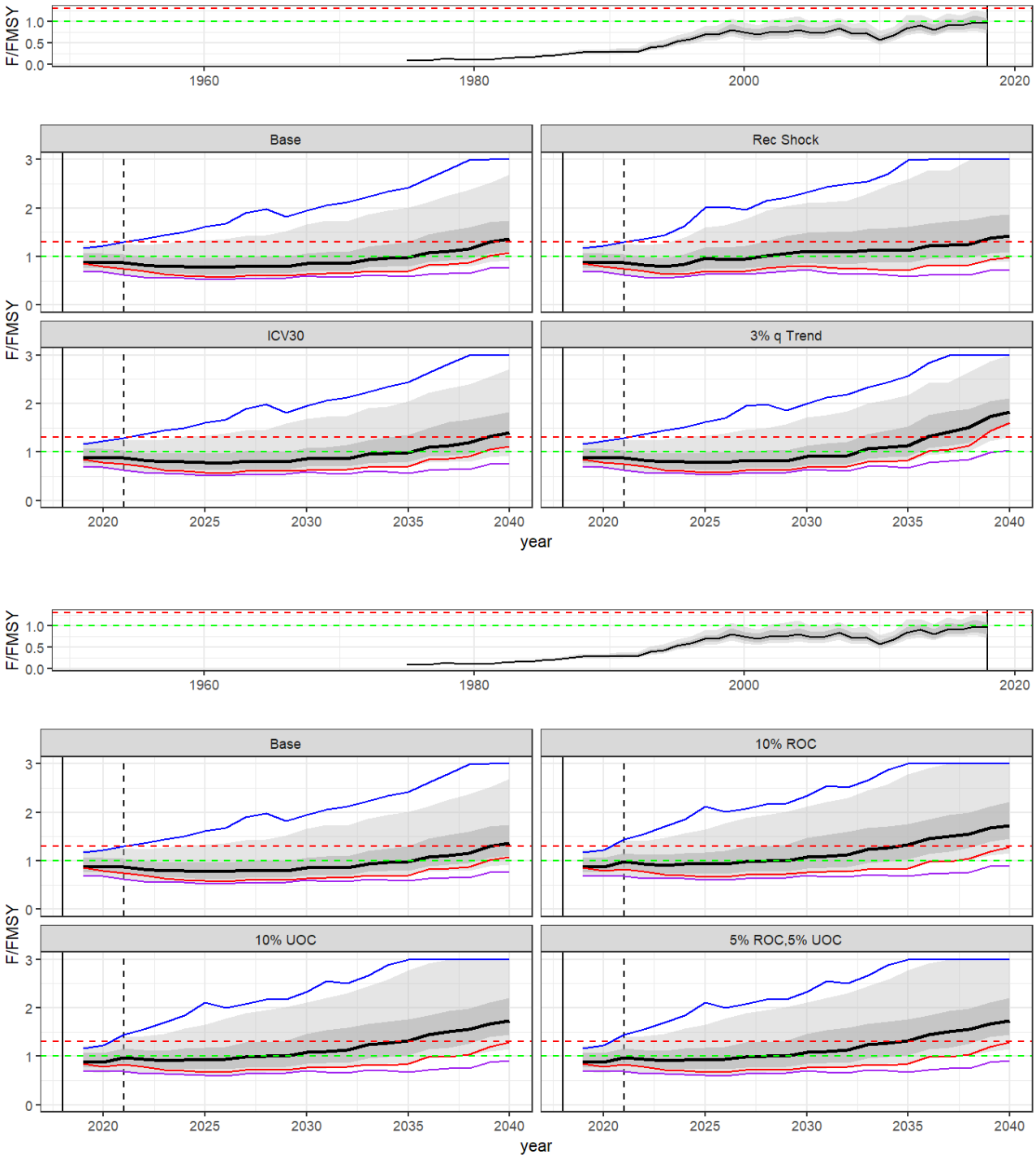

Figure 46. Time series of fishing intensity (Upper bound truncated at F = 3) for various robustness tests and reference set MP IT5.t15g1.5.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

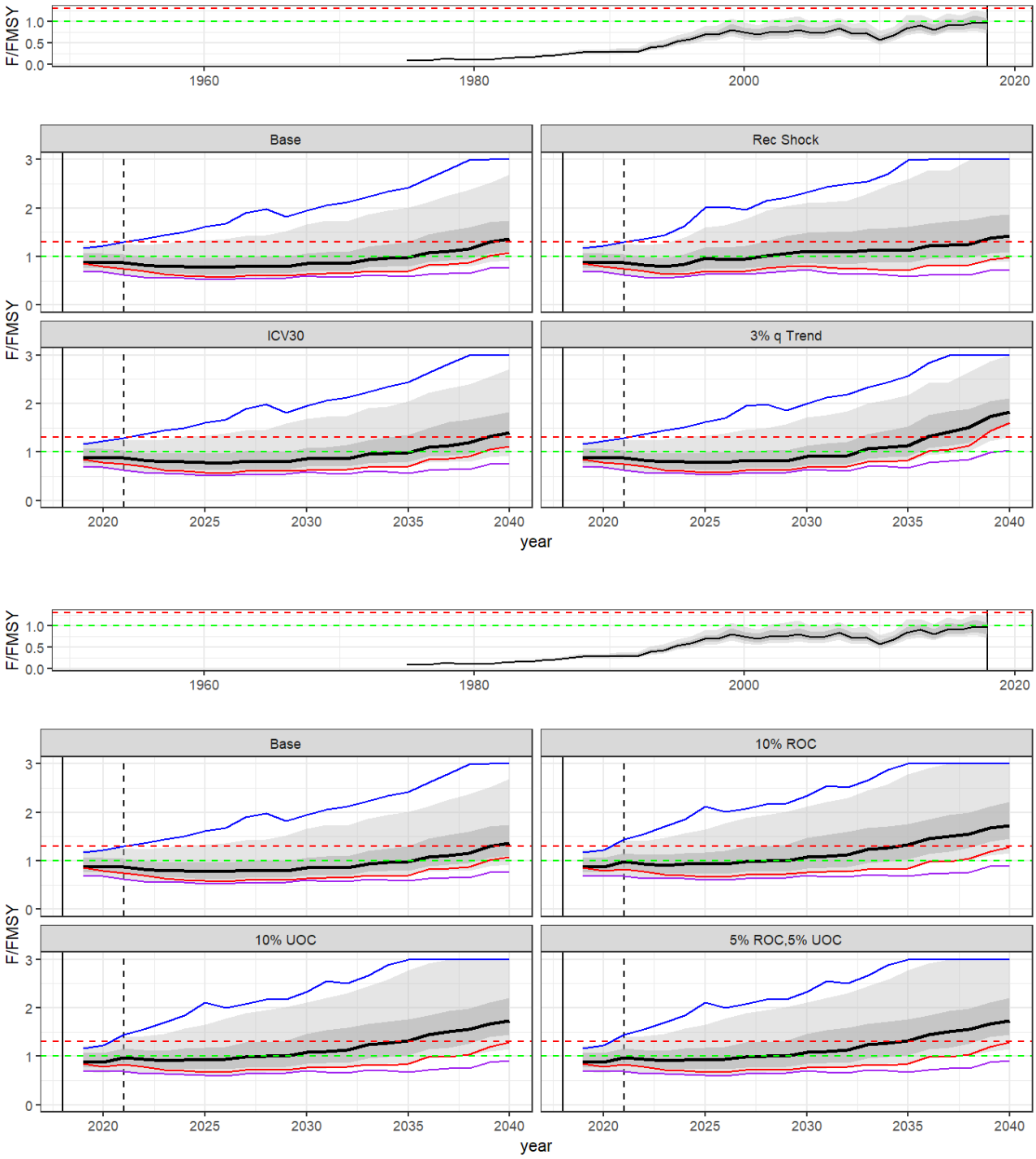

Figure 47. Time series of fishing intensity (Upper bound truncated at F = 3) for various robustness tests and reference set MP PTBoB0Targ.t15.tmb.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

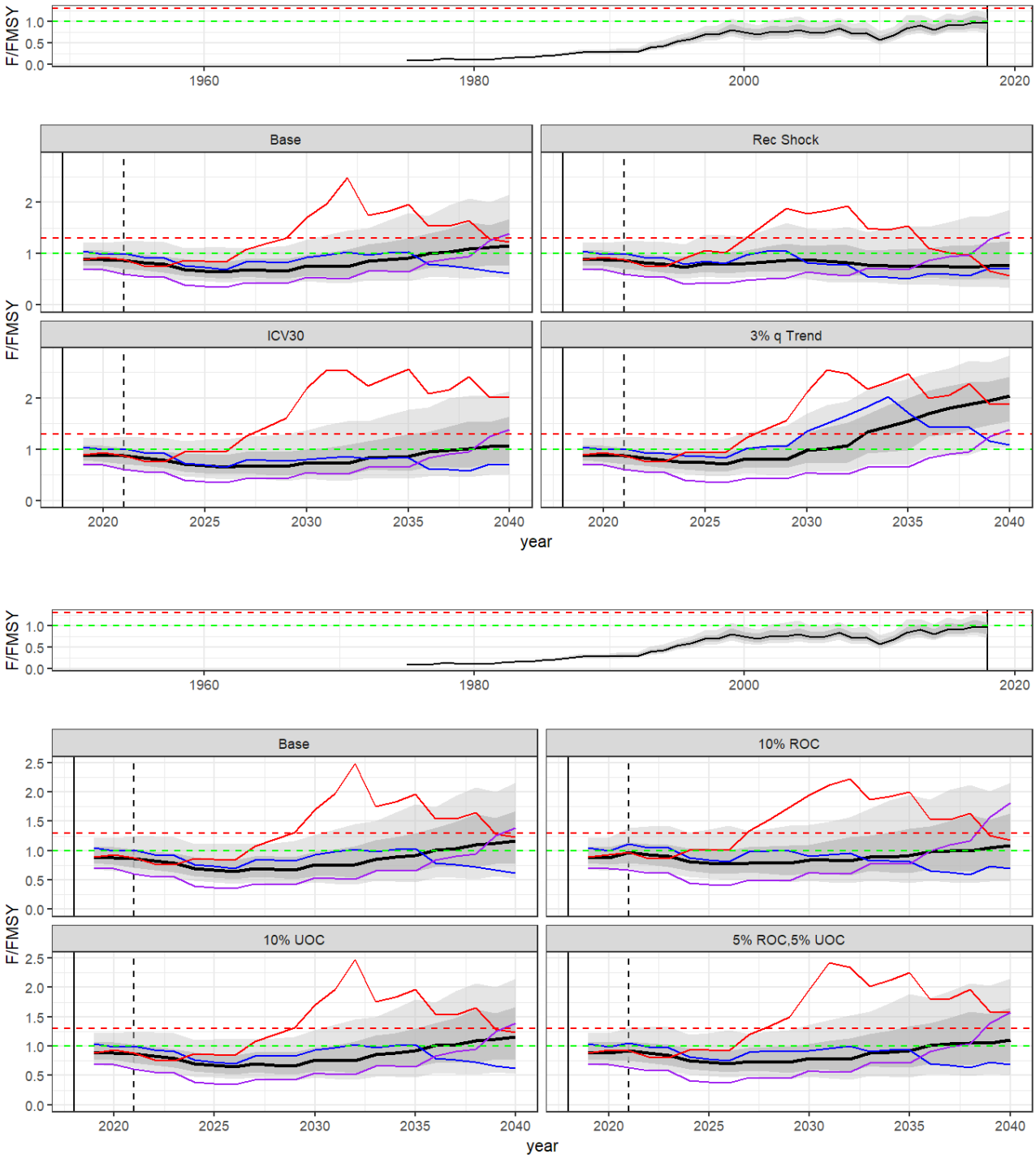

Figure 48. Time series of fishing intensity (Upper bound truncated at F = 3) for various robustness tests and reference set MP PT41F.t15.tmb.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

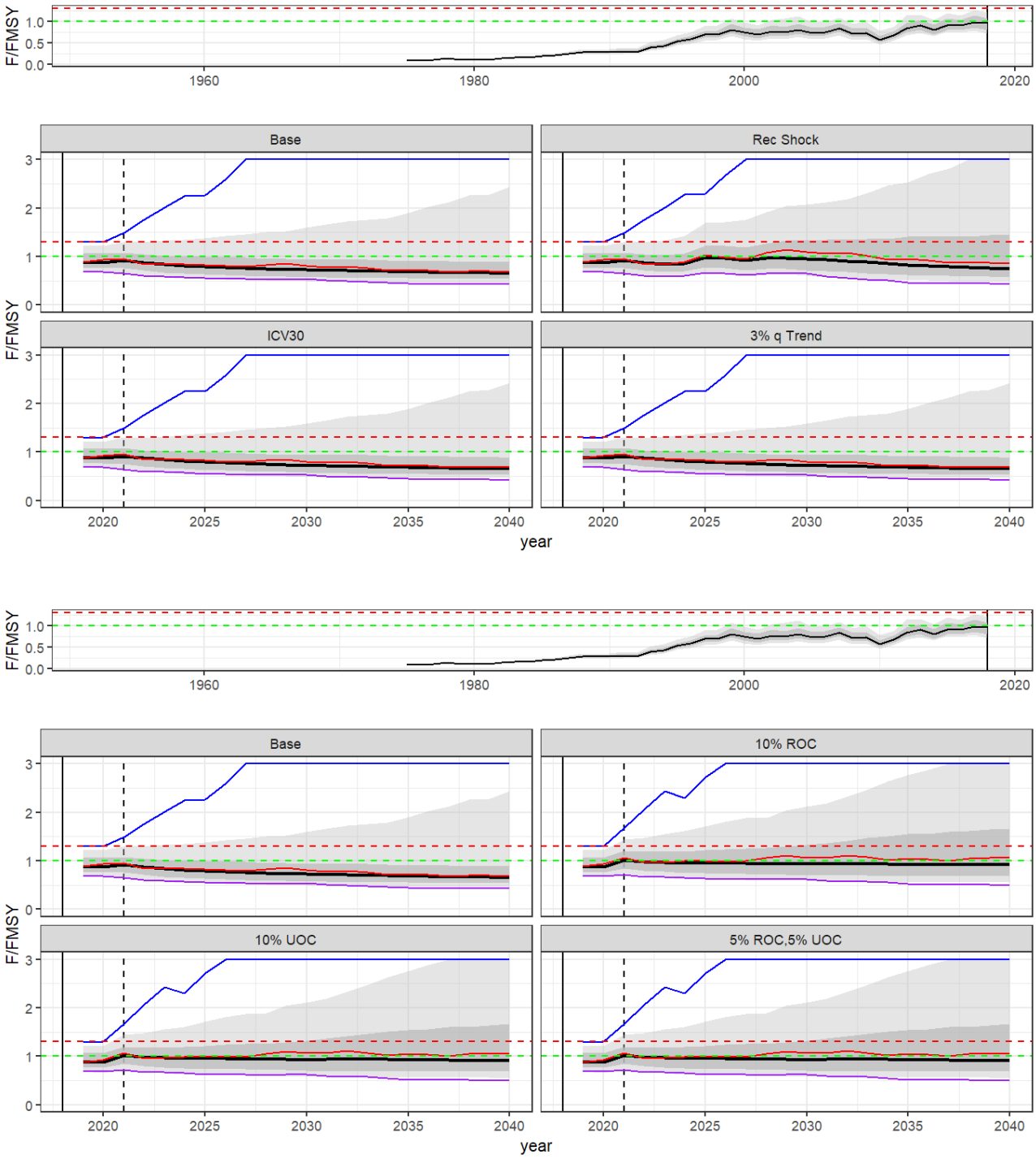

Figure 49. Time series of fishing intensity (Upper bound truncated at F = 3) for various robustness tests and reference set MP CCt.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

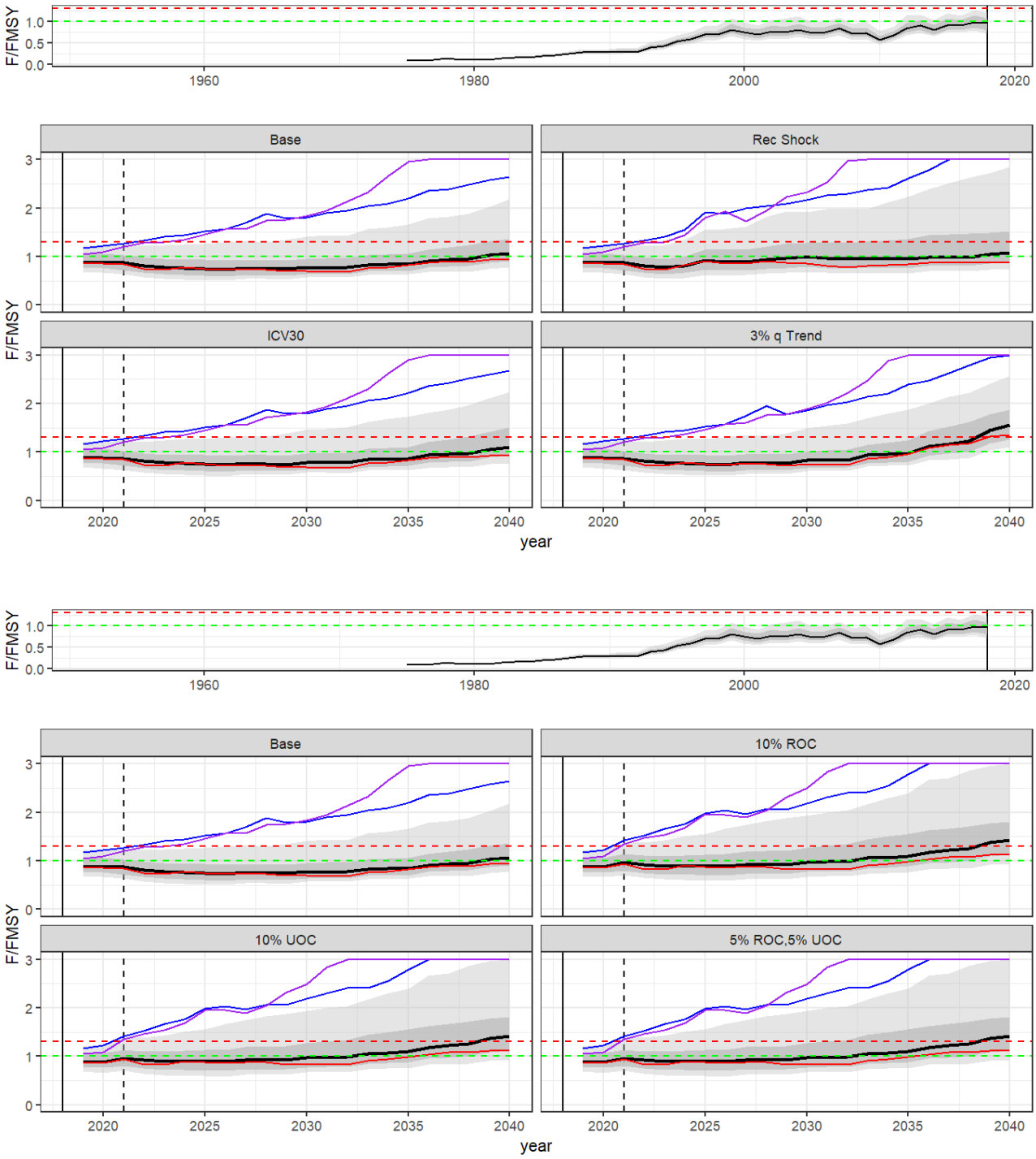

Figure 50. Time series of fishing intensity (Upper bound truncated at F = 3) for various robustness tests and reference set MP IT5.t15g1.5.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

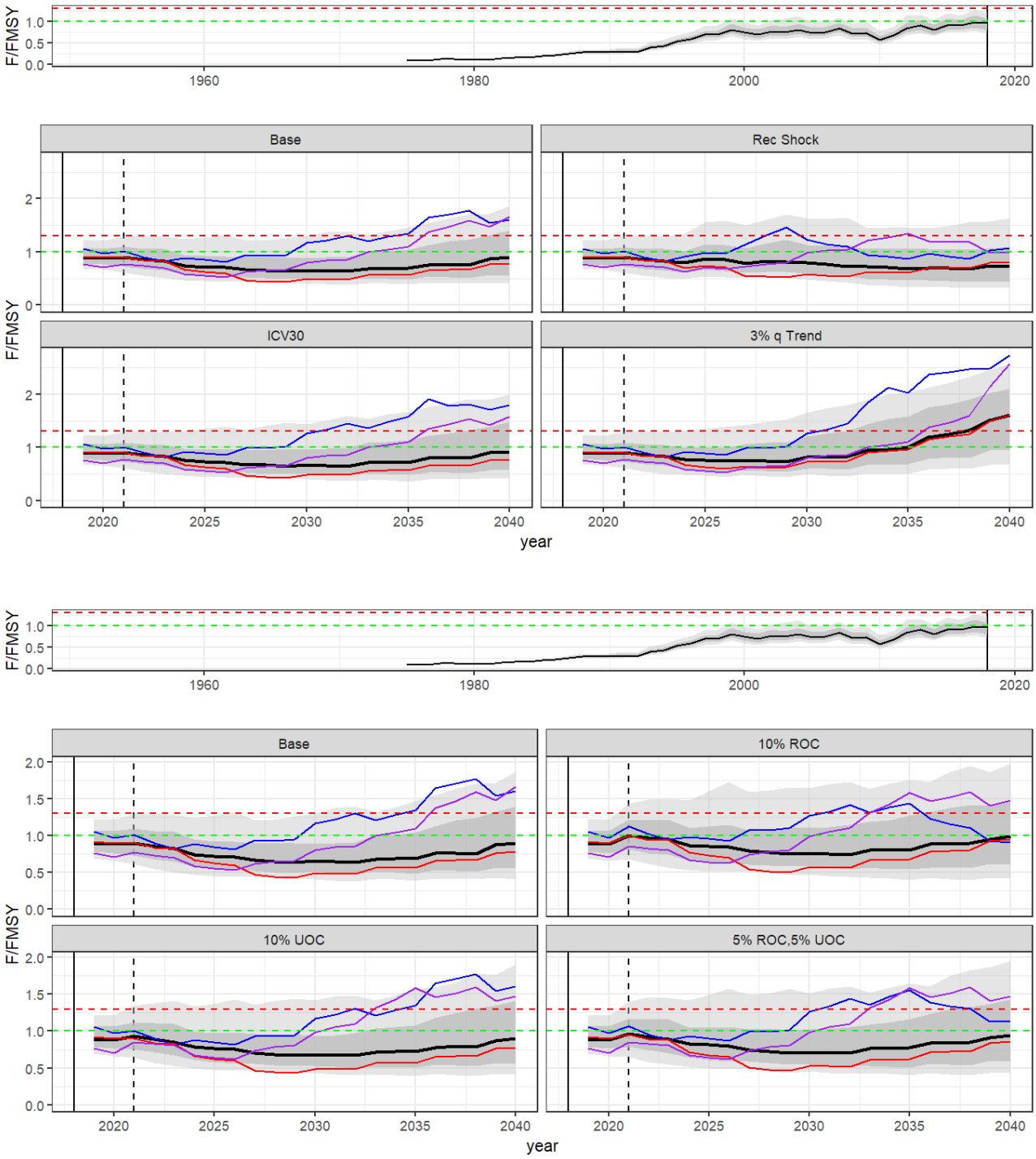

Figure 51. Time series of fishing intensity (Upper bound truncated at F = 3) for various robustness tests and reference set MP PTBoB0Targ.t15.tmb.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

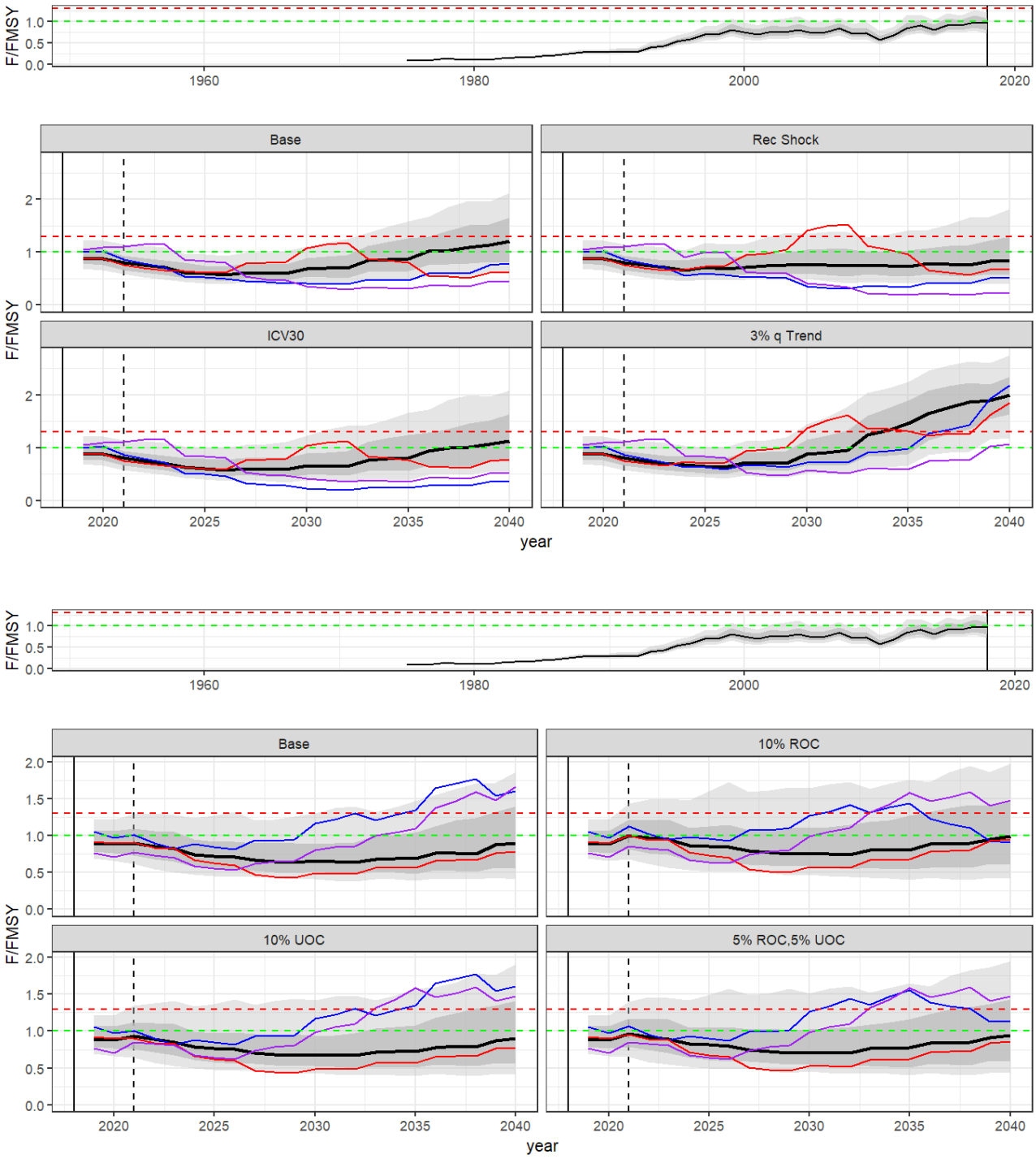

Figure 52. Time series of fishing intensity (Upper bound truncated at F = 3) for various robustness tests and reference set MP PT41F.t15.tmb.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

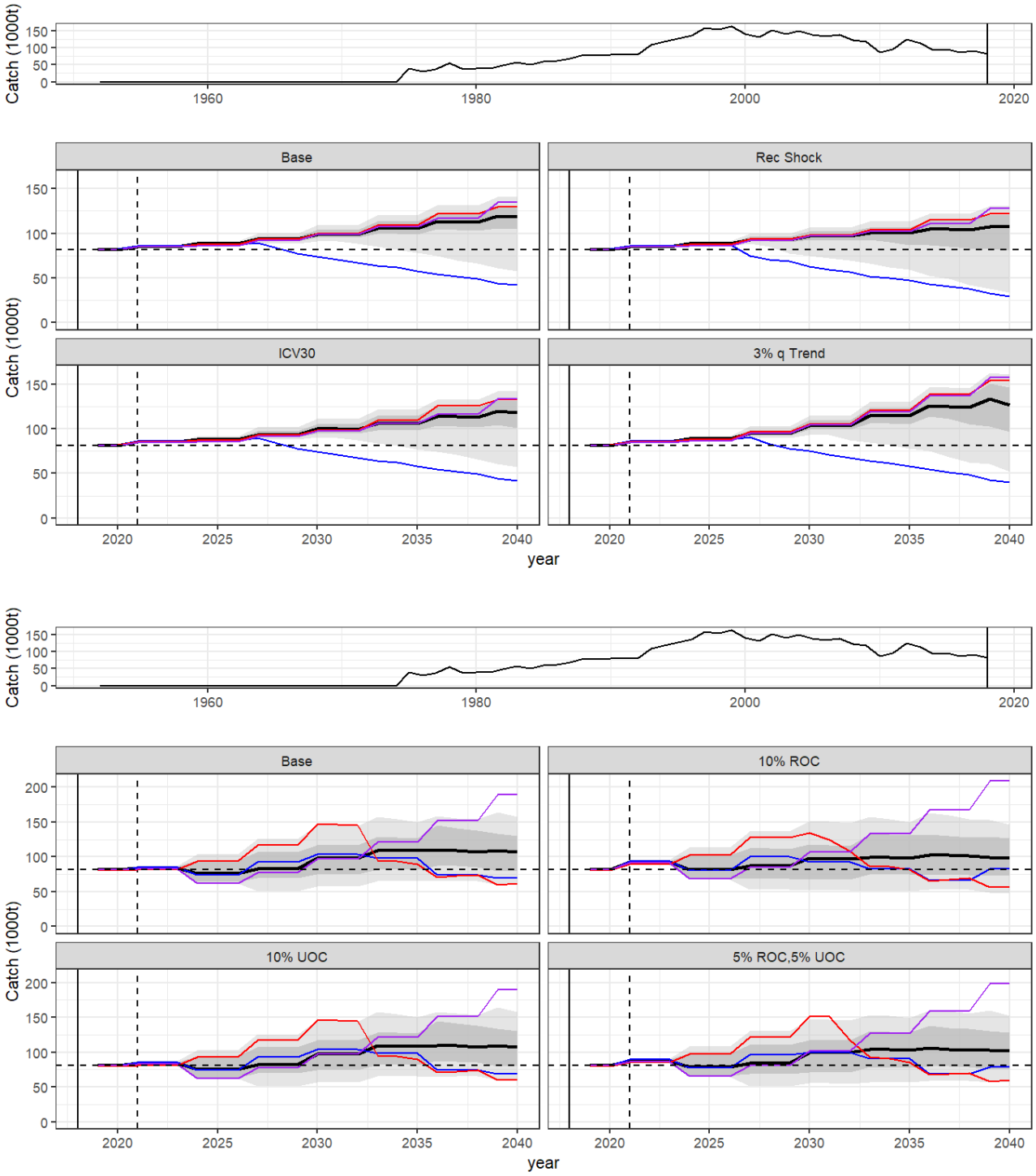

Figure 53. Time series of catch for various robustness tests and reference set MP IT5.t15g1.5.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2017) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

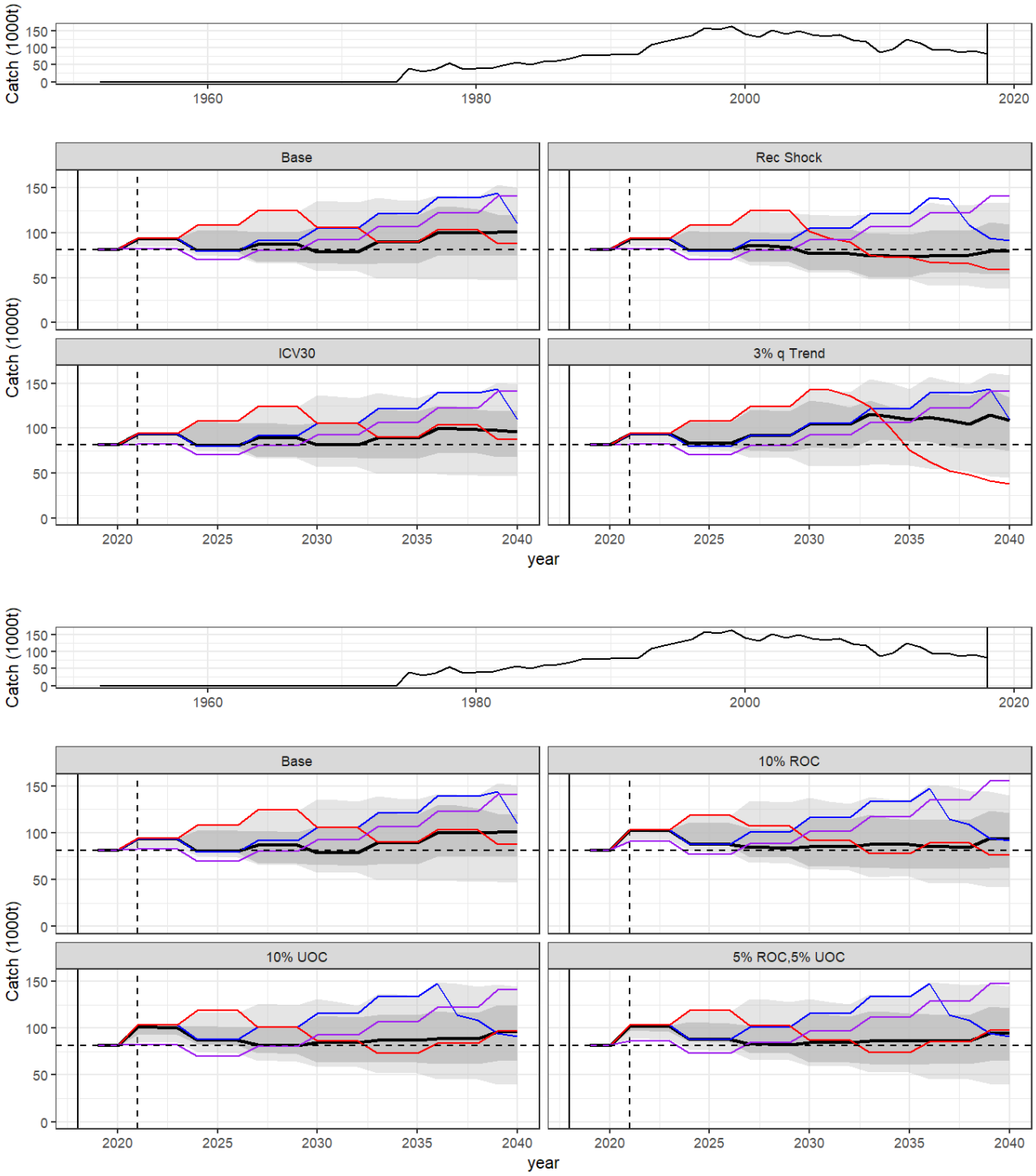

Figure 54. Time series of catch for various robustness tests and reference set MP PTBoB0Targ.t15.tmb.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2017) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

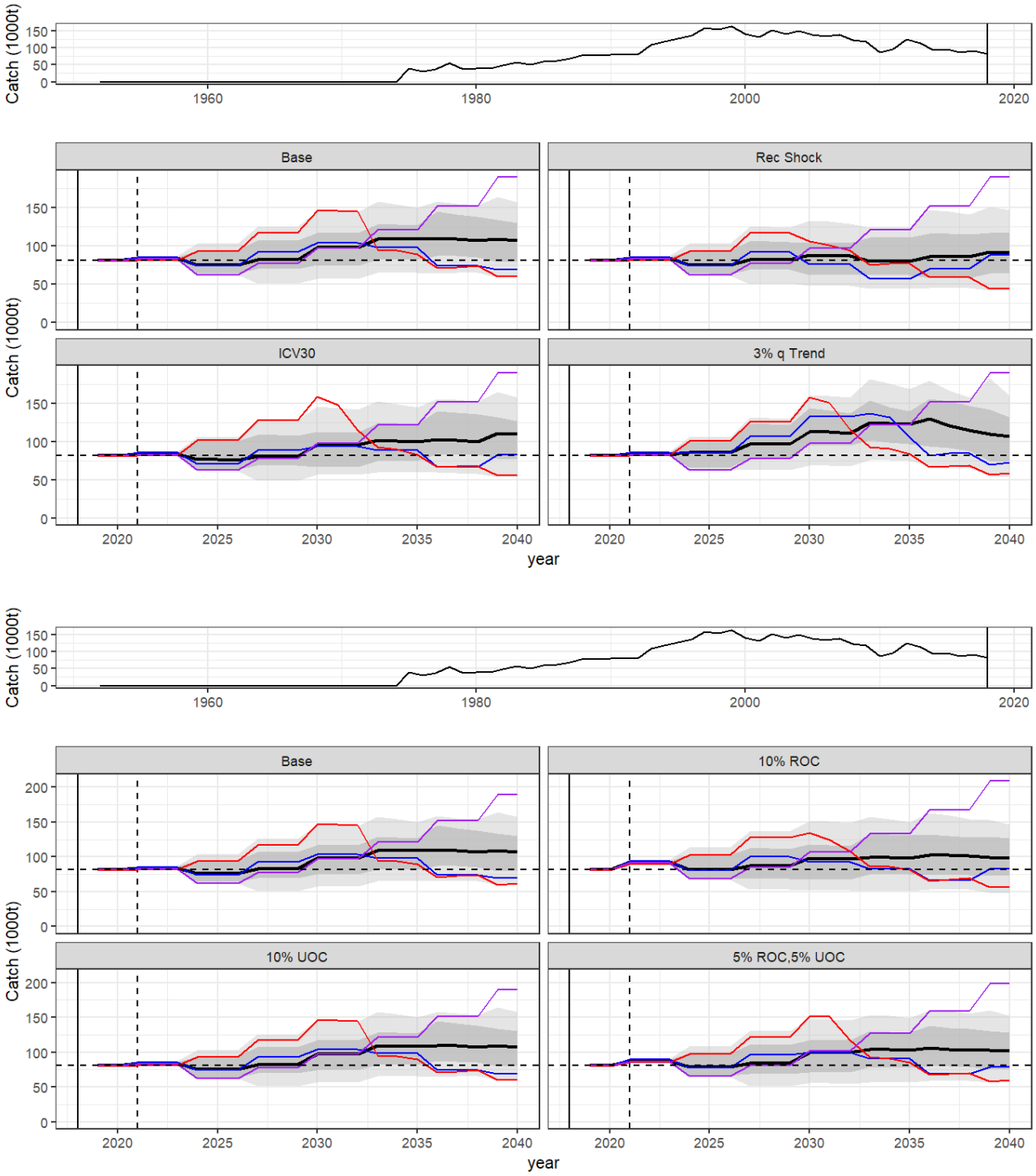

Figure 55. Time series of catch for various robustness tests and reference set MP PT41F.t15.tmb.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2017) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

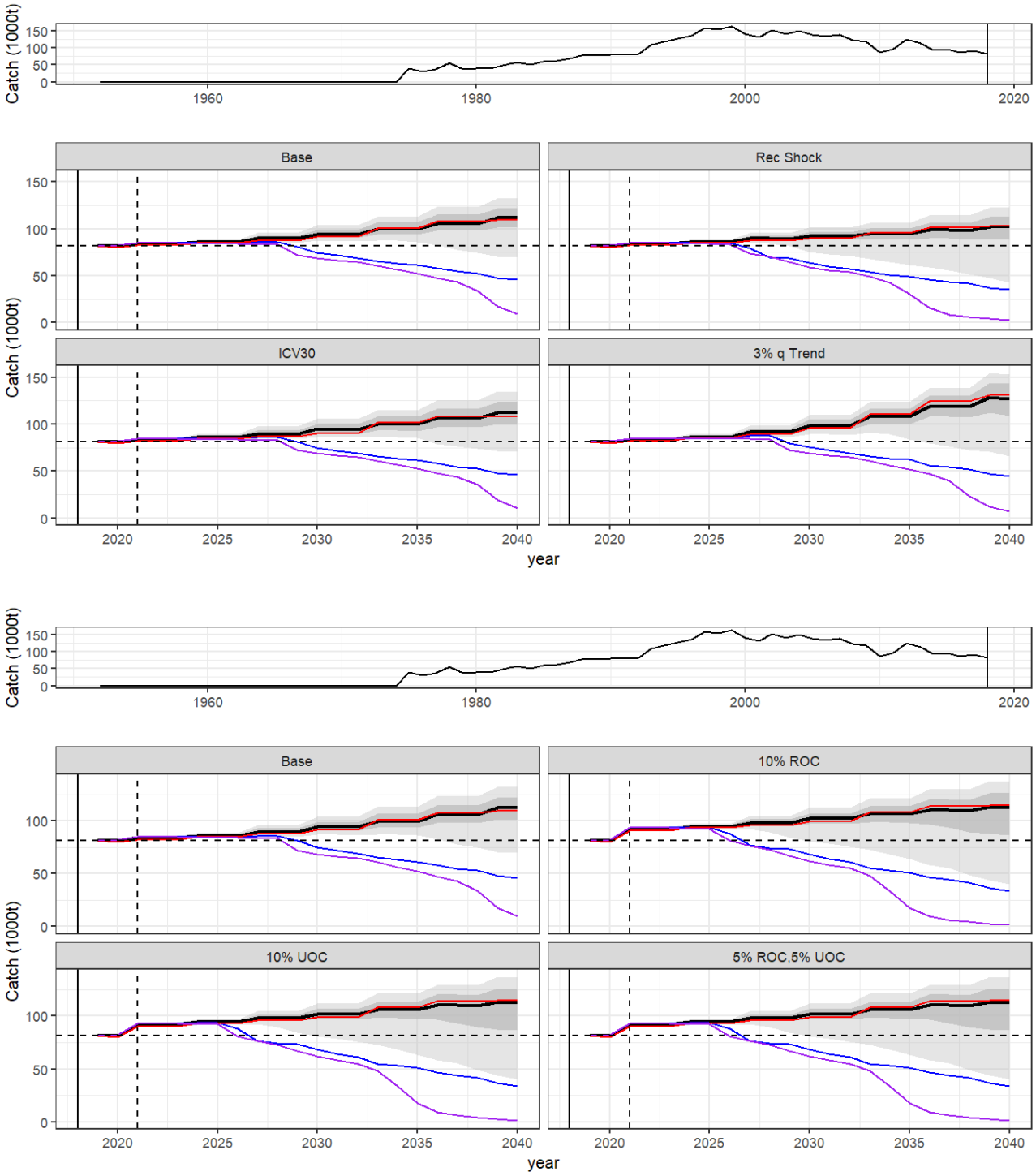

Figure 56. Time series of catch for various robustness tests and reference set MP IT5.t15g1.5.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2017) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

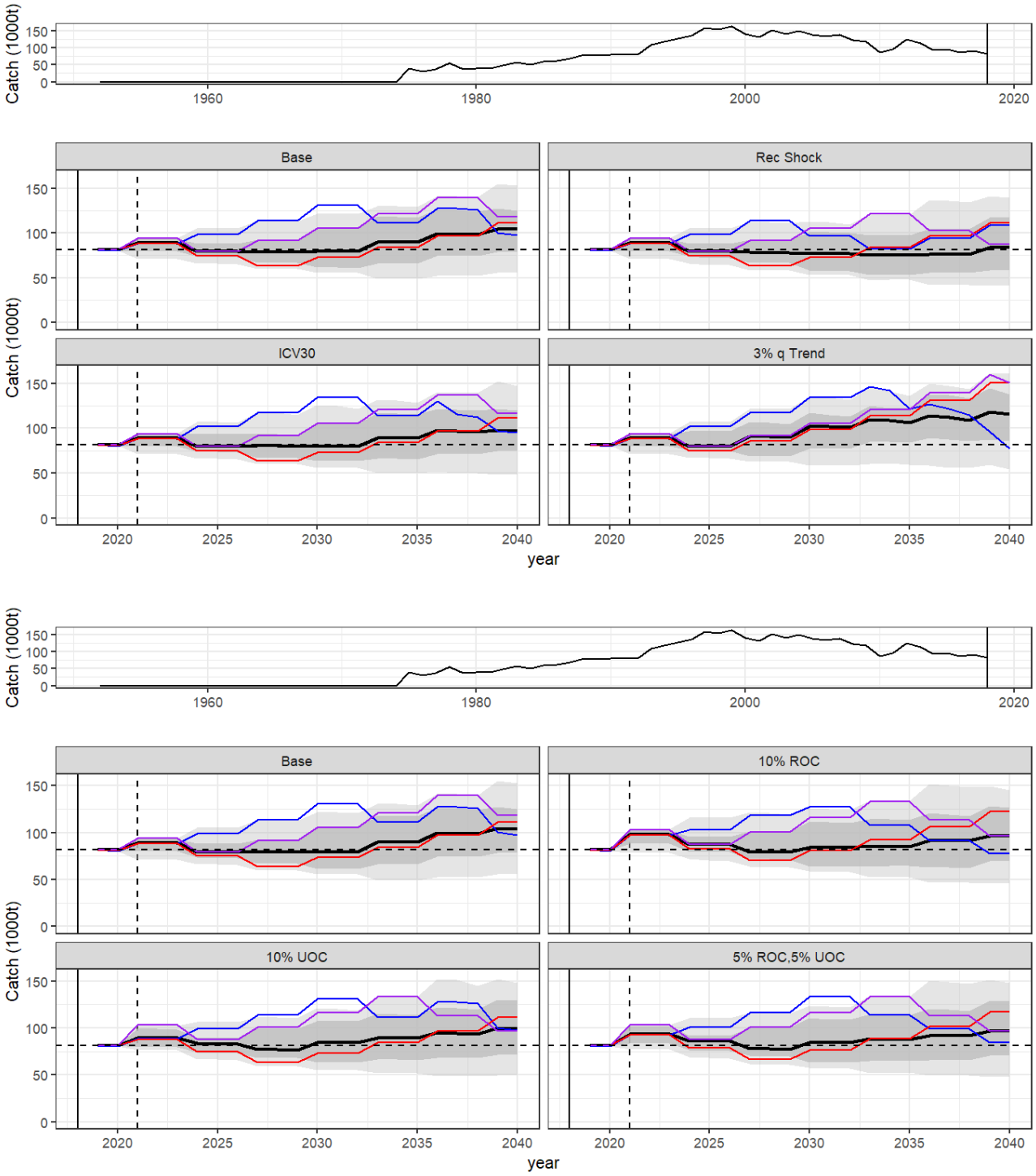

Figure 57. Time series of catch for various robustness tests and reference set MP PTBoB0Targ.t15.tmb.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2017) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

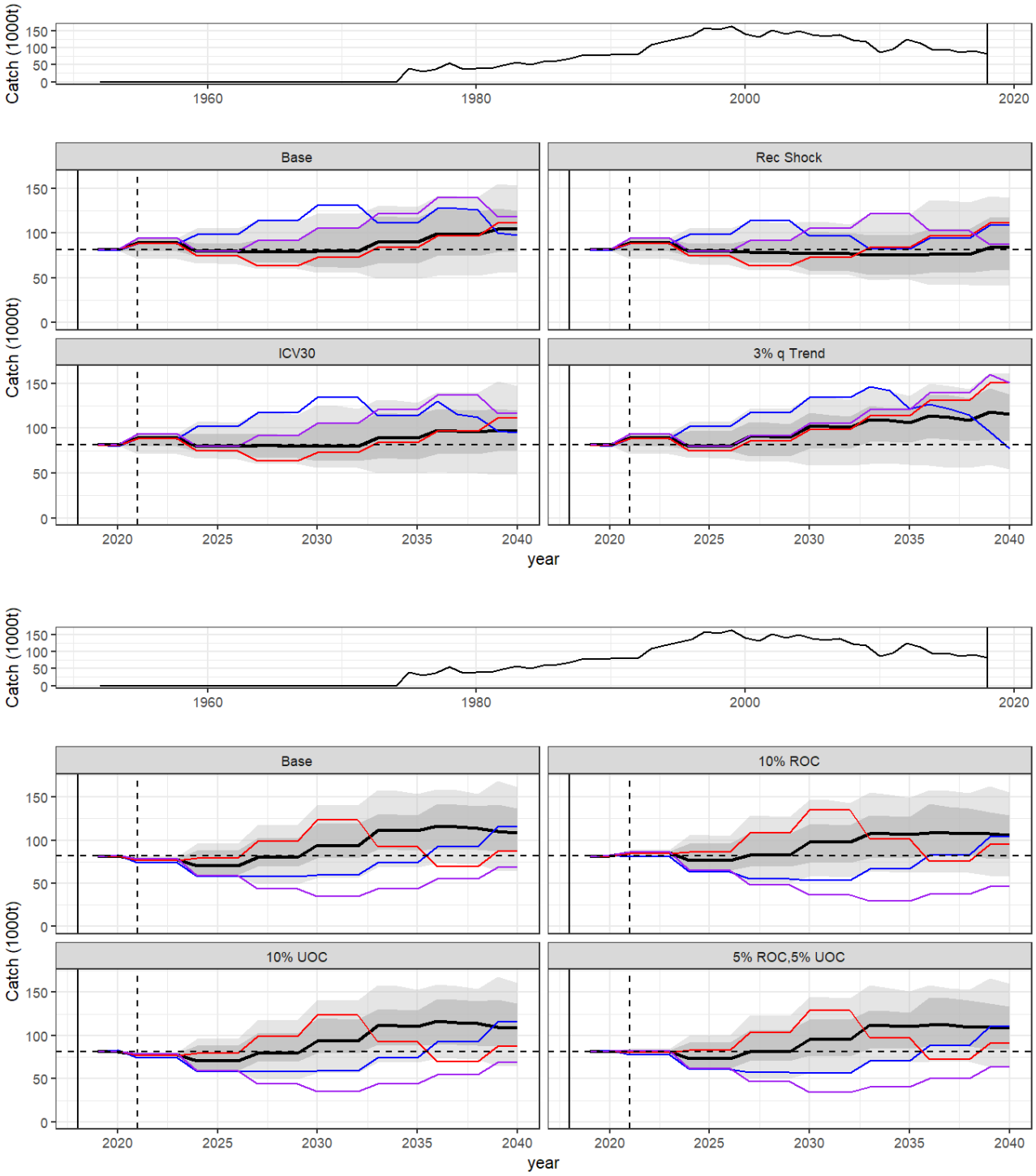

Figure 58. Time series of catch for various robustness tests and reference set MP PT41F.t15.tmb.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2017) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

5 Bigeye Management Procedure Evaluation sensitivity to nuisance spatial assumptions

As noted in WPTT (2020, included here in Appendix A), the issue of very high fishing mortality in some strata remains a concern for both the BET and YFT OMs (and stock assessments). Kolody et al. (2020) showed that the vast majority of BET OMs were flagged as having a catch penalty sufficient to indicate that the model could not extract the observed catch in at least one quarter/age/area strata, when the SS hybrid F implementation had a cap of F = 2.9 (95% exploitation rate). The problem largely disappeared when the cap was raised to F = 6.0 (exploitation rate ~99%), and changing the cap did not appear to have much effect on the stock status inferences. We do not necessarily consider that any exploitation rate in the 95-99% range is plausible, but examination of some individual models suggested that the problem tended to arise at a point decades in the past without any obvious effect on recent dynamics (and hence the initial state of the population used for MSE testing). We interpreted this to mean that the model often lacks the flexibility to represent the seasonal and/or interannual processes required to position the fish in exactly the right spot at all times, but that this probably does not have much effect on the overall productivity and initial state estimates required for the MP evaluations. Accordingly, we did not adopt the catch penalty as a plausibility filter in this iteration. But the shortage of fish problems identified for yellowfin (Kolody and Jumppanen 2021), suggests that this merits more consideration, as it could impact MP evaluation results.

Robustness tests 7-9 (Table 2) duplicate some of the simple analyses undertaken for yellowfin, to see if MP performance is sensitive to some of the finer details of the catch equation implementation and mixing of fish. These OMs have identical initial stock status and production dynamics to the "intermediate" reference set OM (OMgridY21.5 – 4 areas but no tags – labelled "Cpp EC20" here). From Figure 59 - Figure 74, we observe that:

- MP performance is not very sensitive to whether the Baranov catch equations use an effort ceiling of 20 ("Cpp EC20") or 2 ("Cpp EC2"). The effort ceiling is the maximum scalar that can be applied to the recent model-specific fishing mortality estimated for each fleet. i.e. a doubling of effective fishing effort seems at least possible, while an increase by a factor of 20 seems very unlikely for most fleets for a fully developed fishery (though the potential effect of fish aggregation complicates this assumption for some fleets).
- It does not make much difference whether the projection catch extraction is based on the Baranov's equations ("Cpp ECx") or Pope's approximation ("R"). Note that the difference is more complicated than the actual equation - the Pope approach simply solves each quarter independently for 25% of the TAC, while the Cpp approach solves for the whole year simultaneously.
- If fish of all ages are uniformly redistributed every quarter ("uniform mixing"), the model should be very similar to a spatially-aggregated model, i.e. each fleet can access almost all of the fish, and the effects of space/time refugia are minimized. This has a minor effect on the BET MP evaluations.

Figure 73 - Figure 74 illustrate the Catch/TAC ratio over time for these MP evaluations, and demonstrate that the TAC was readily extracted for most realizations over the period 2020-2035 from all models, while a maximum of ~25% of models have trouble extracting the full TAC in the 2035-2040 period for 3 of the 4 OMs. The exception OM is the uniform mixing scenario, in which there did not appear to be any problem removing the TAC at any time. Uniform mixing is not consistent with what the SS models estimate, but could potentially be more realistic if the models cannot represent the scale of the BET seasonal movements realistically.

The BET results are very different from YFT, and consistent with what we would expect for a healthy population, in which fisheries are not struggling to locate fish.

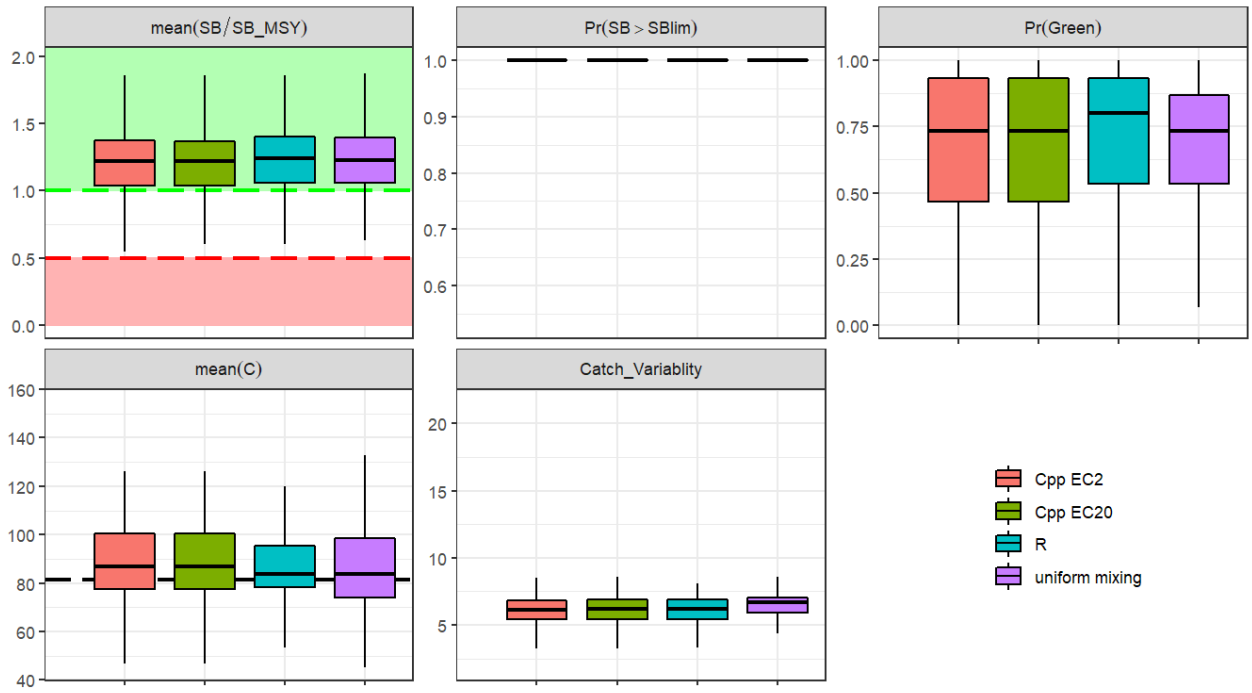

Figure 59. Boxplots comparing OM model implementations for MP PT41F.t15.tmb.B2 with respect to key performance measures averaged over the period 2021 -2035. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2020 catch.

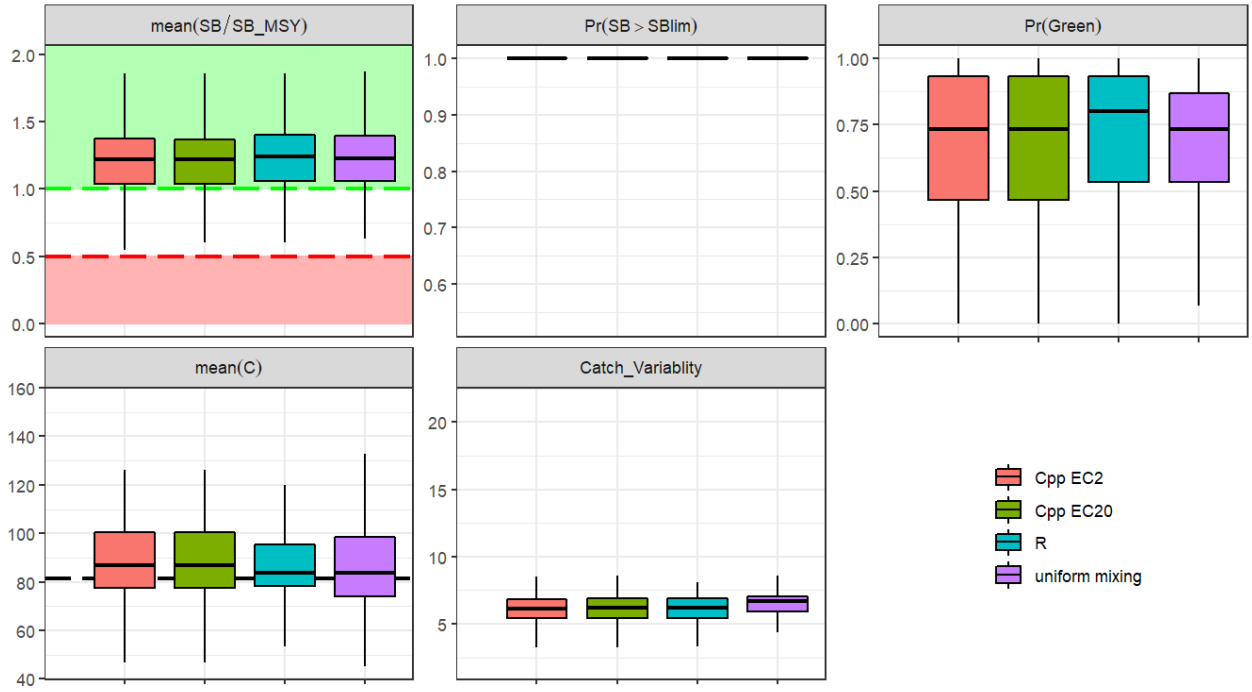

Figure 60. Boxplots comparing OM model implementations for MP PT41F.t15.tmb.B3 with respect to key performance measures averaged over the period 2021 -2035. Horizontal line is the median, boxes represent 25th - 75th percentiles, thin lines represent 10th - 90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The horizontal dashed black line is 2020 catch.

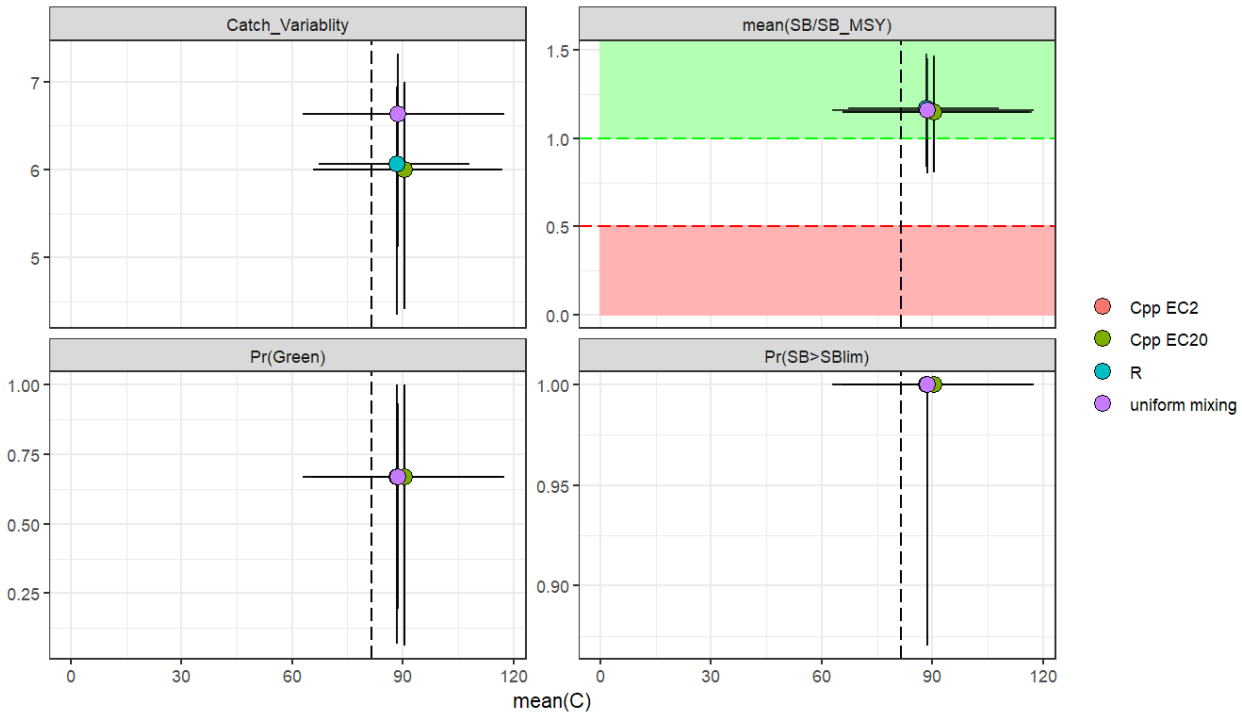

Figure 61. Trade-off plots comparing OM model implementations for MP PT41F.t15.tmb.B2 with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2021 - 2035. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2017 catch.

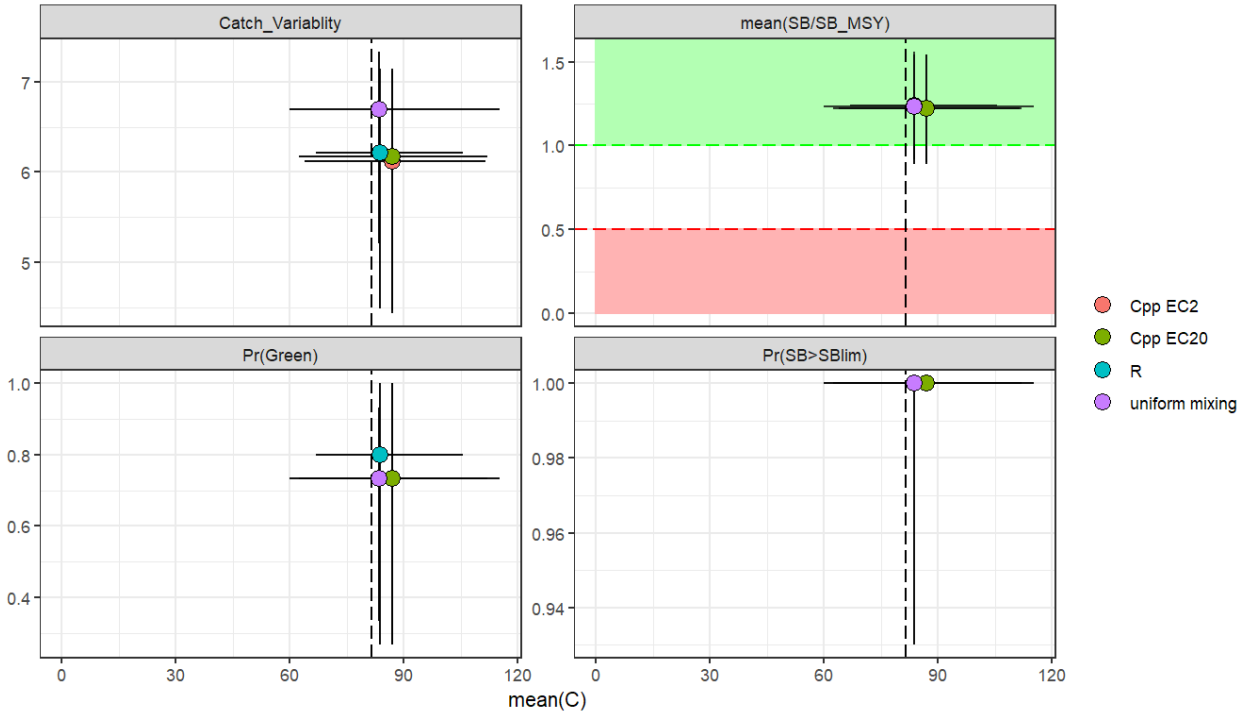

Figure 62. Trade-off plots comparing OM model implementations for MP PT41F.t15.tmb.B3 with respect to catch on the X-axis, and 4 other key performance measures on the Y-axis, each averaged over the period 2021 - 2035. Circle is the median, lines represent 10th-90th percentiles. Red and green horizontal lines represent the interim limit and target reference points for the mean SB/SBMSY performance measure. The dashed vertical black line is 2017 catch.

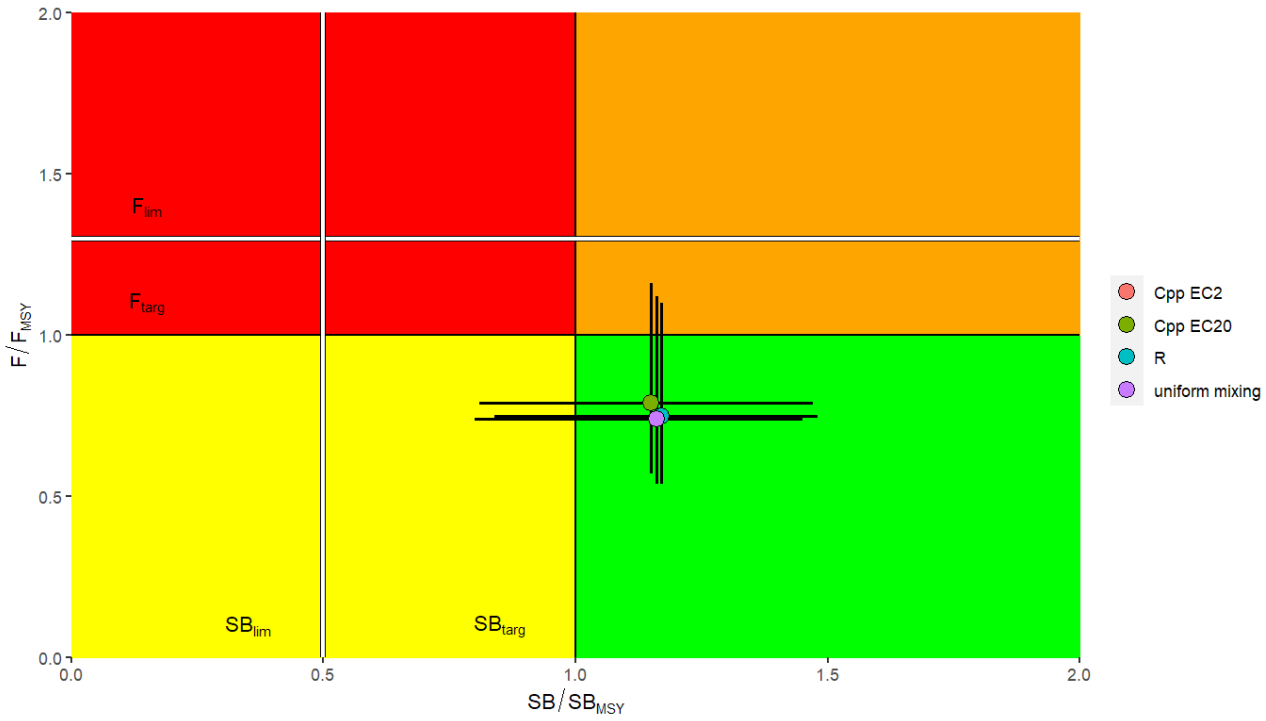

Figure 63. Kobe plot comparing OM model implementations for MP PT41F.t15.tmb.B2 on the basis of the expected 15 year average (2021-2035) performance. Circle is the median, lines represent 10th-90th percentiles.

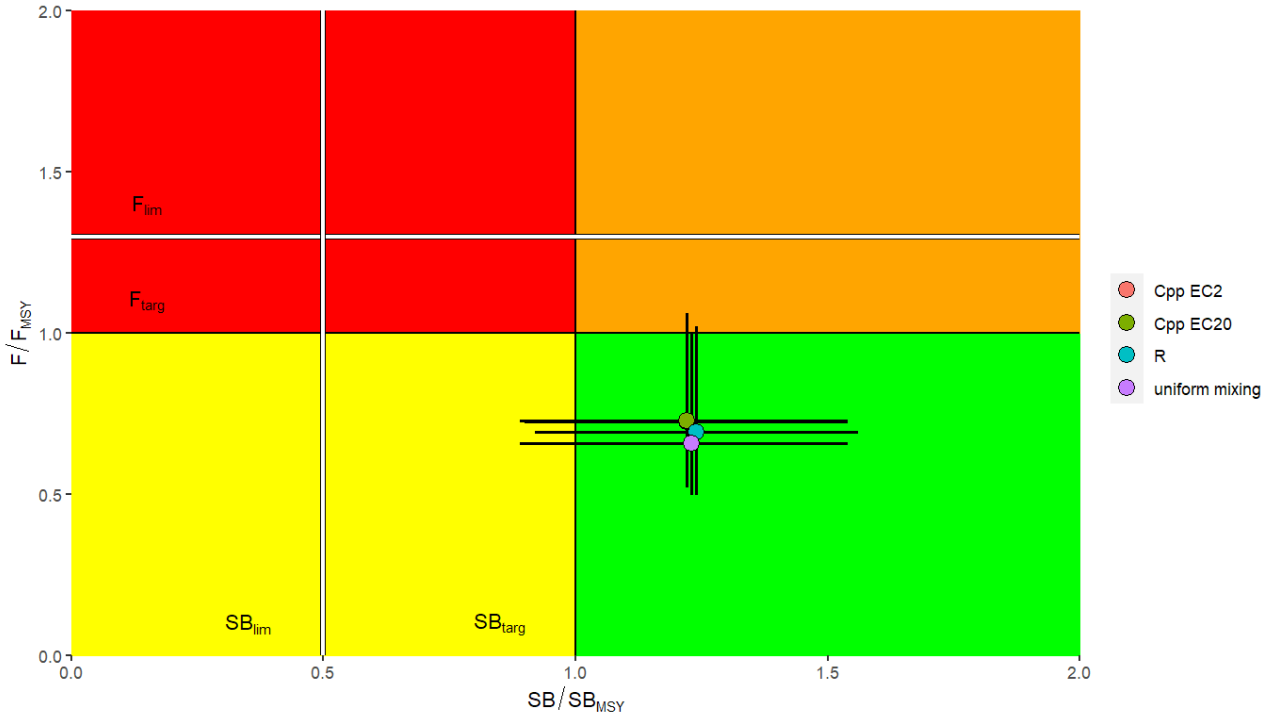

Figure 64. Kobe plot comparing OM model implementations for MP PT41F.t15.tmb.B3 on the basis of the expected 15 year average (2021-2035) performance. Circle is the median, lines represent 10th-90th percentiles.

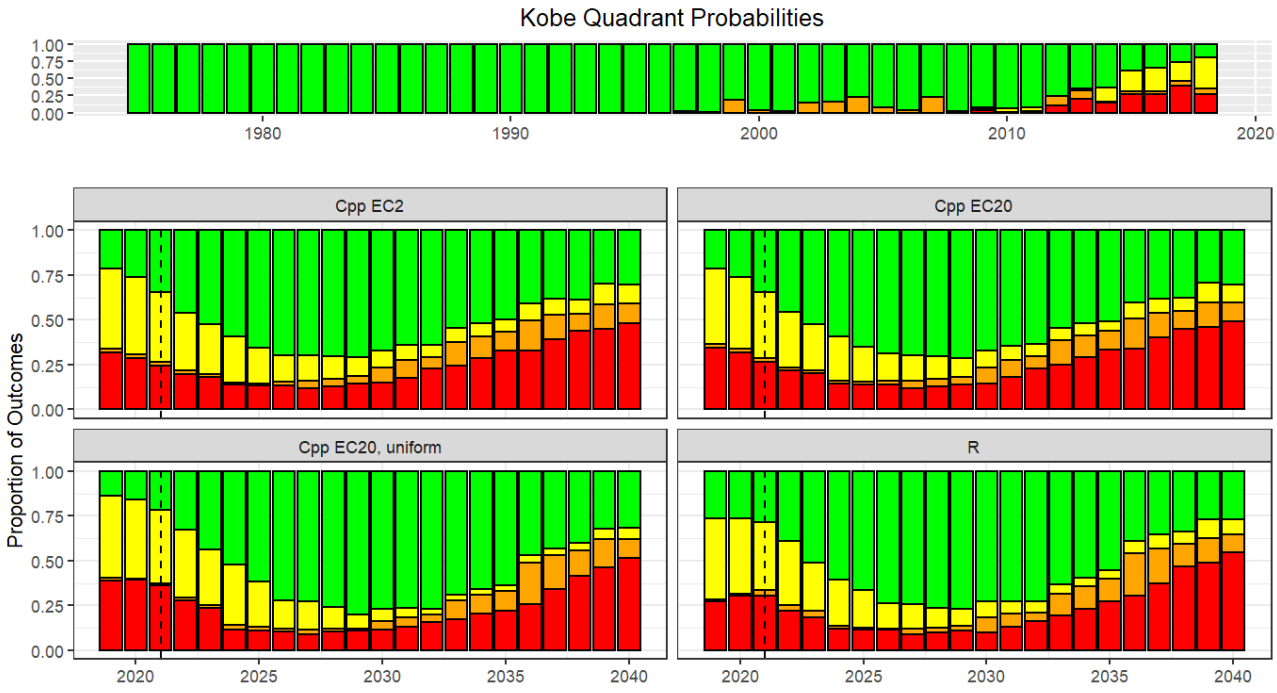

Figure 65. Proportion of simulations in each of the Kobe quadrants over time comparing OM model implementations for MP PT41F.t15.tmb.B2. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

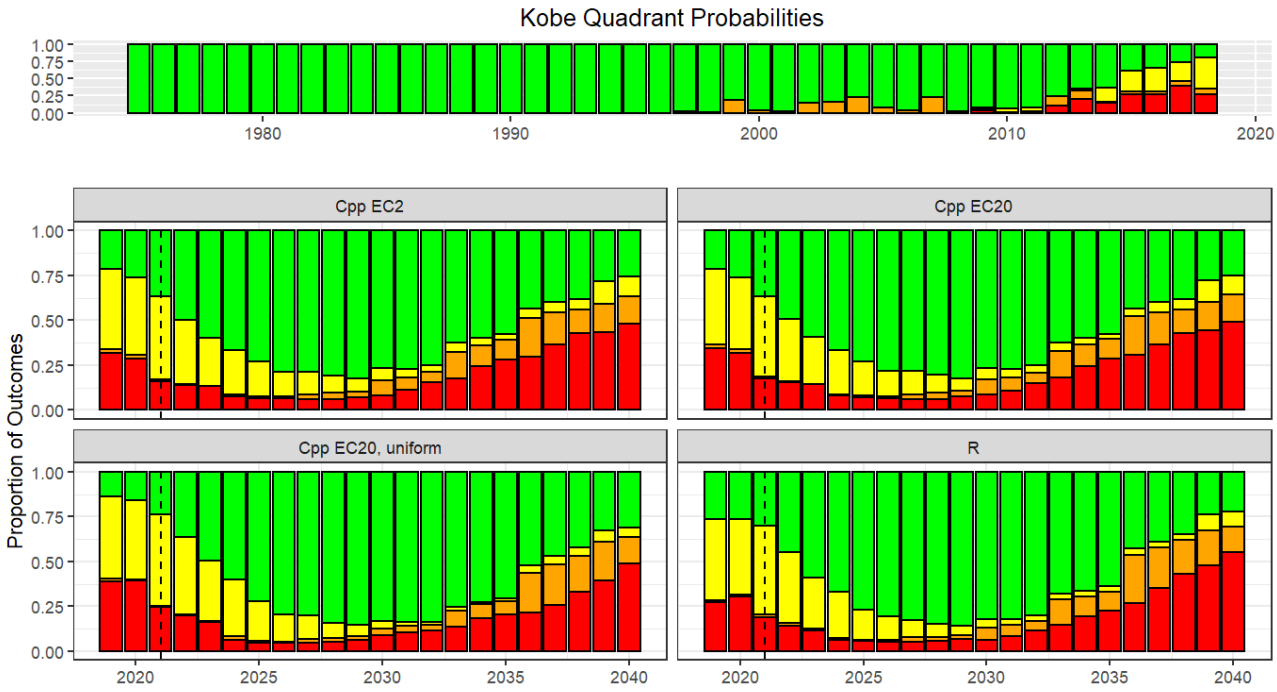

Figure 66. Proportion of simulations in each of the Kobe quadrants over time comparing OM model implementations for MP PT41F.t15.tmb.B3. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

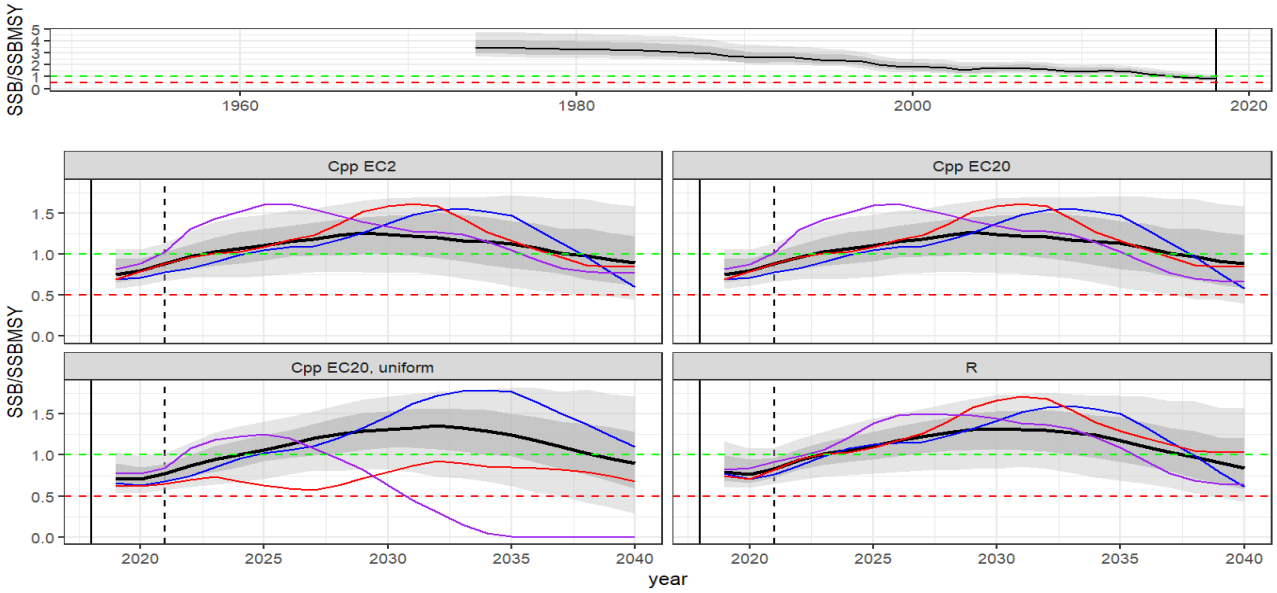

Figure 67. Time series of spawning stock size comparing OM model implementations for MP PT41F.t15.tmb.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

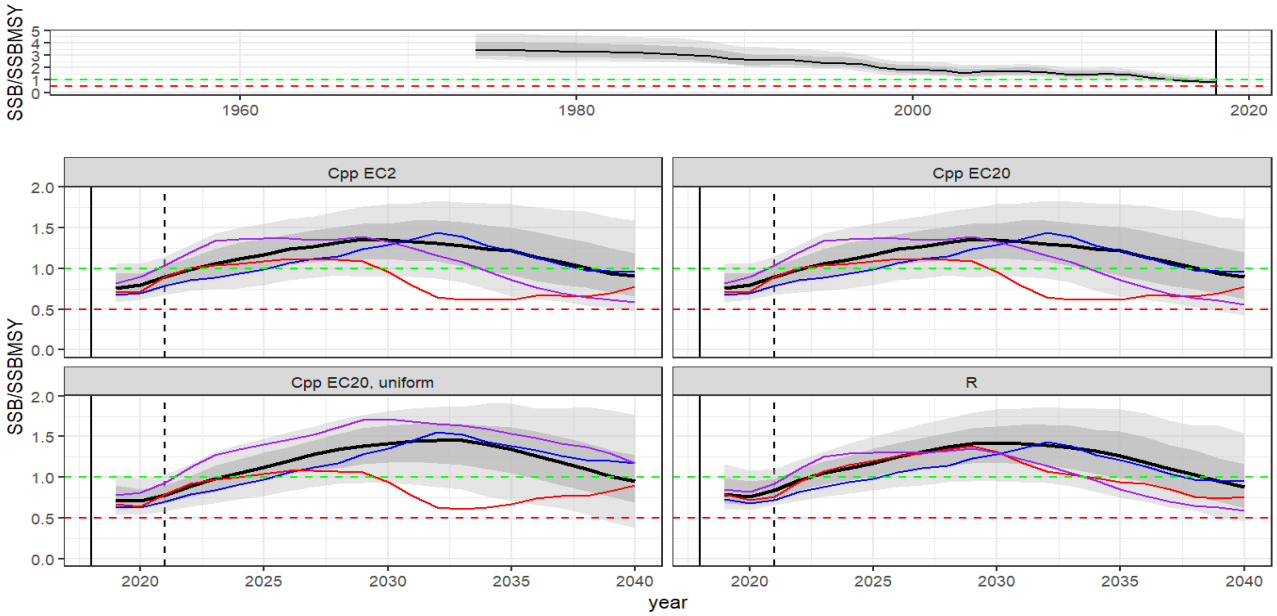

Figure 68. Time series of spawning stock size comparing OM model implementations for MP PT41F.t15.tmb.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

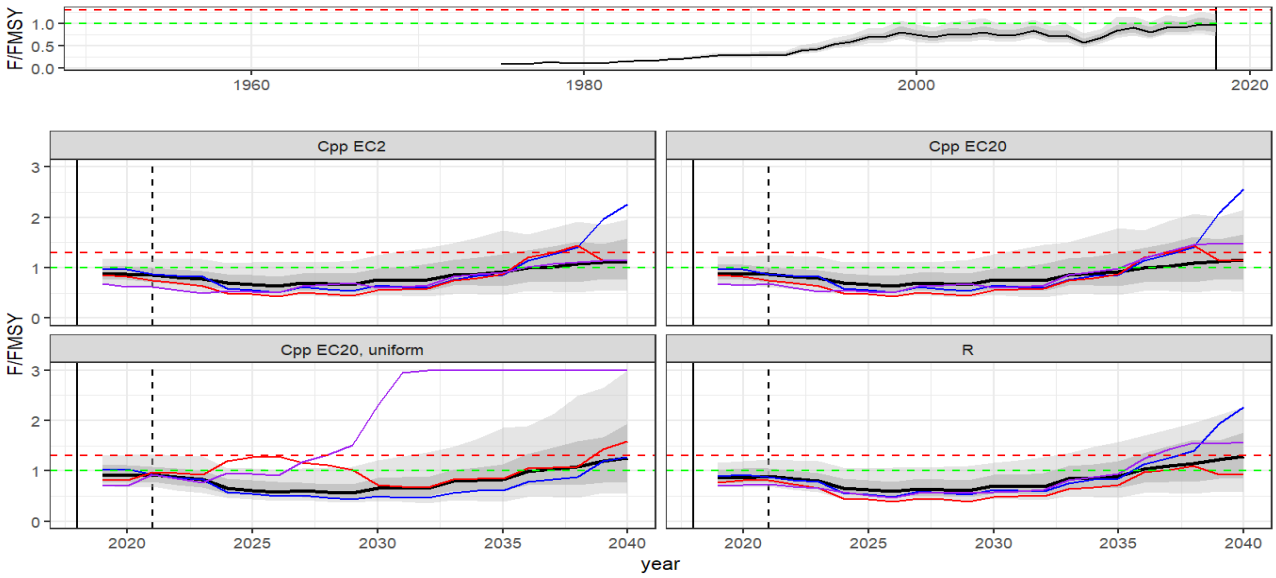

Figure 69. Time series of fishing intensity (Upper bound truncated at F = 3) comparing OM model implementations for MP PT41F.t15.tmb.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

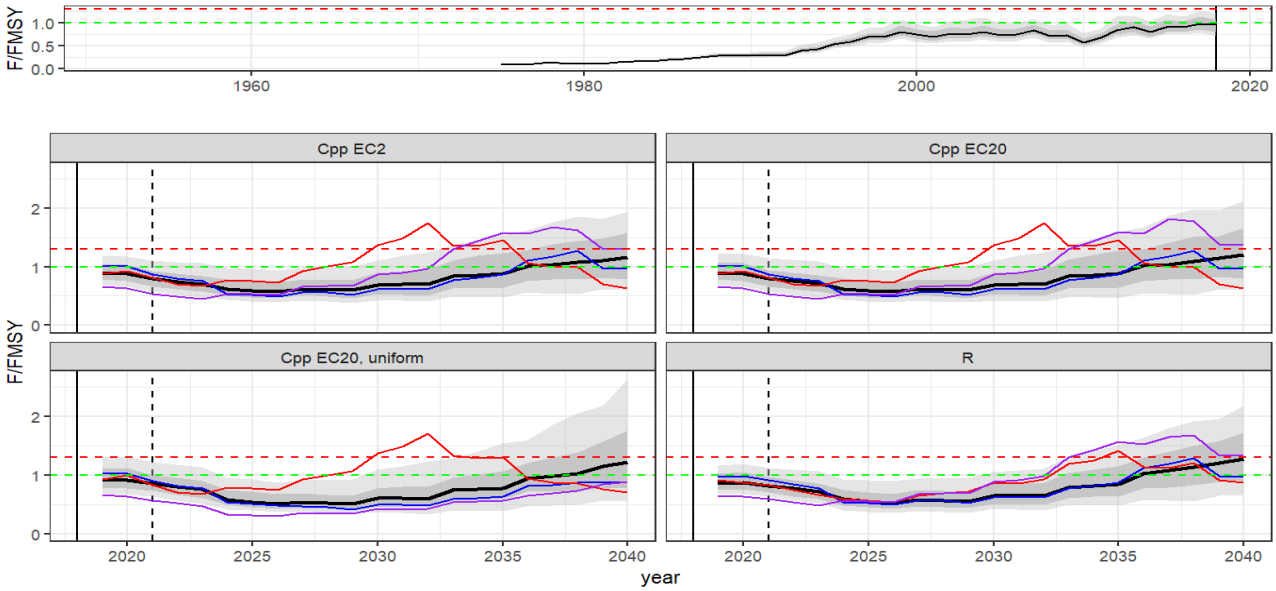

Figure 70. Time series of fishing intensity (Upper bound truncated at F = 3) comparing OM model implementations for MP PT41F.t15.tmb.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. Thick broken lines represent the interim target (green) and limit (red) reference points. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

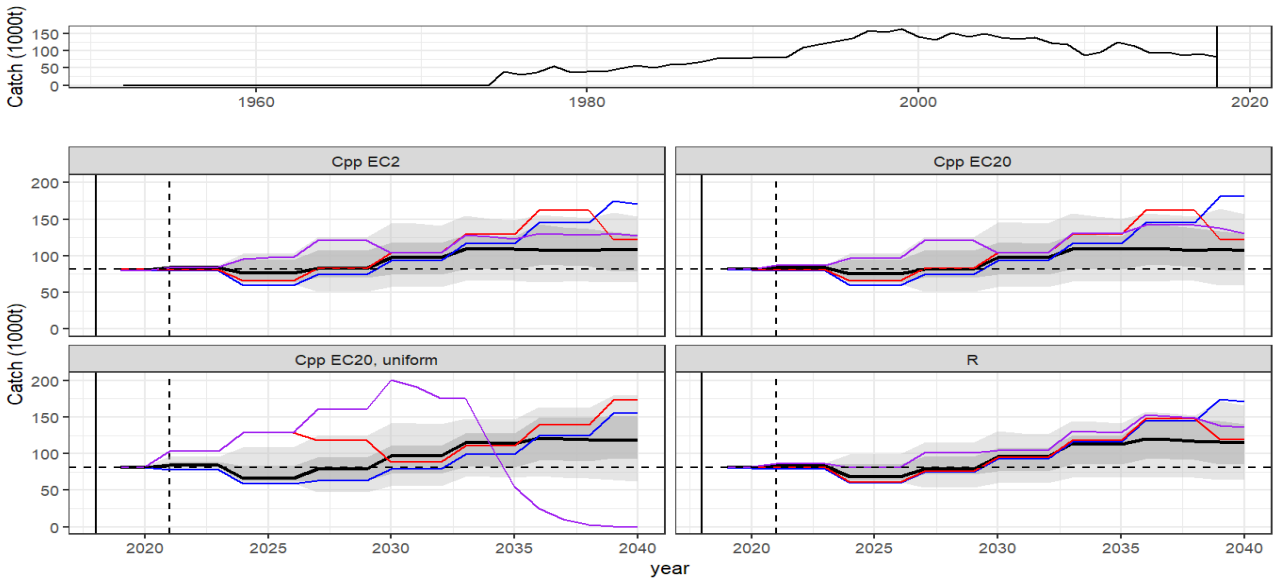

Figure 71. Time series of catch comparing OM model implementations for MP PT41F.t15.tmb.B2. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2017) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

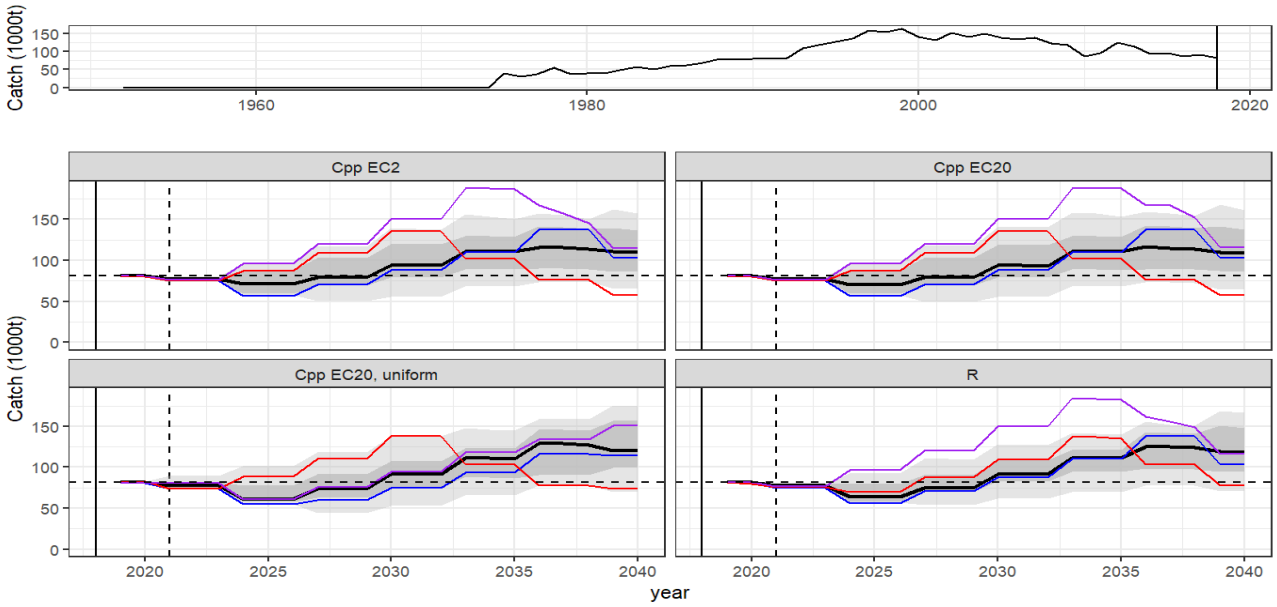

Figure 72. Time series of catch comparing OM model implementations for MP PT41F.t15.tmb.B3. The top panel represents the historical estimates from the reference case operating model, and lower plots represent the projection period. The solid vertical line represents the last year used in the historical conditioning. The broken vertical line represents the first year that the MP is applied. The median is represented by the bold black line, the dark shaded ribbon represents the 25th-75th percentiles, the light shaded ribbon represents the 10th-90th percentiles. The broken black horizontal line represents recent (2017) catch. The 3 thin coloured lines represent examples of individual realizations (the same OM scenarios across MPs and performance measures), to illustrate that individual variability greatly exceeds the median.

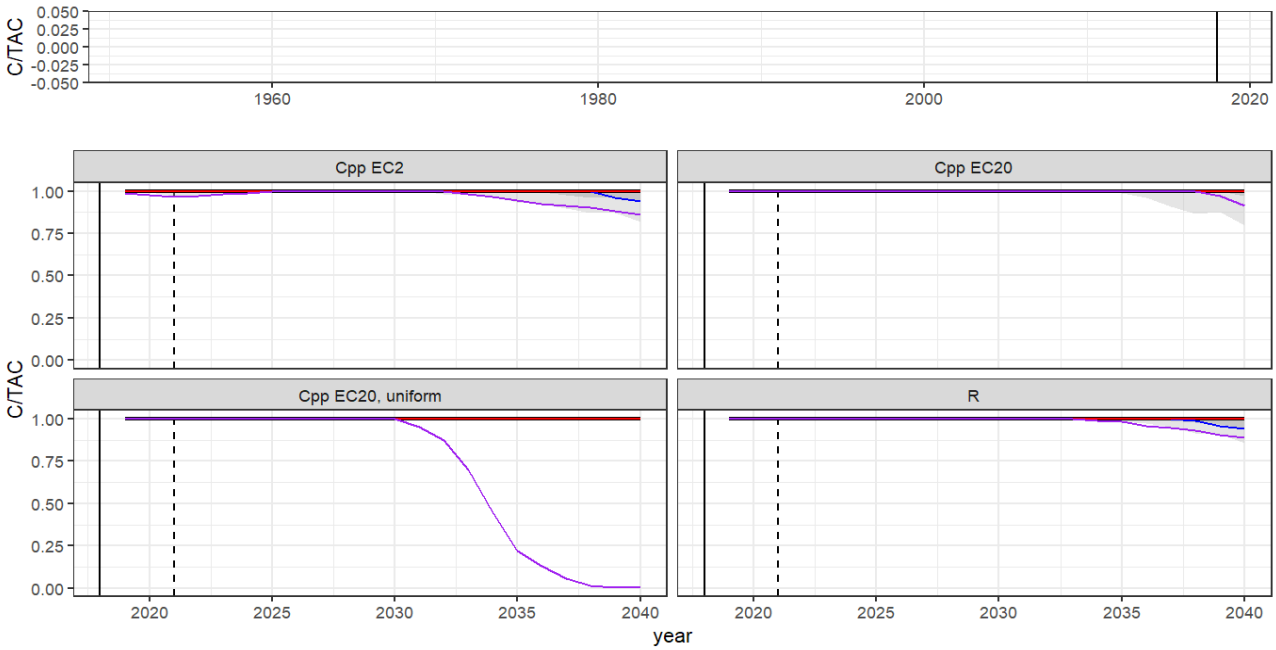

Figure 73. C/TAC over time comparing OM model implementations for MP PT41F.t15.tmb.B2. Historical estimates irrelevant and blank in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

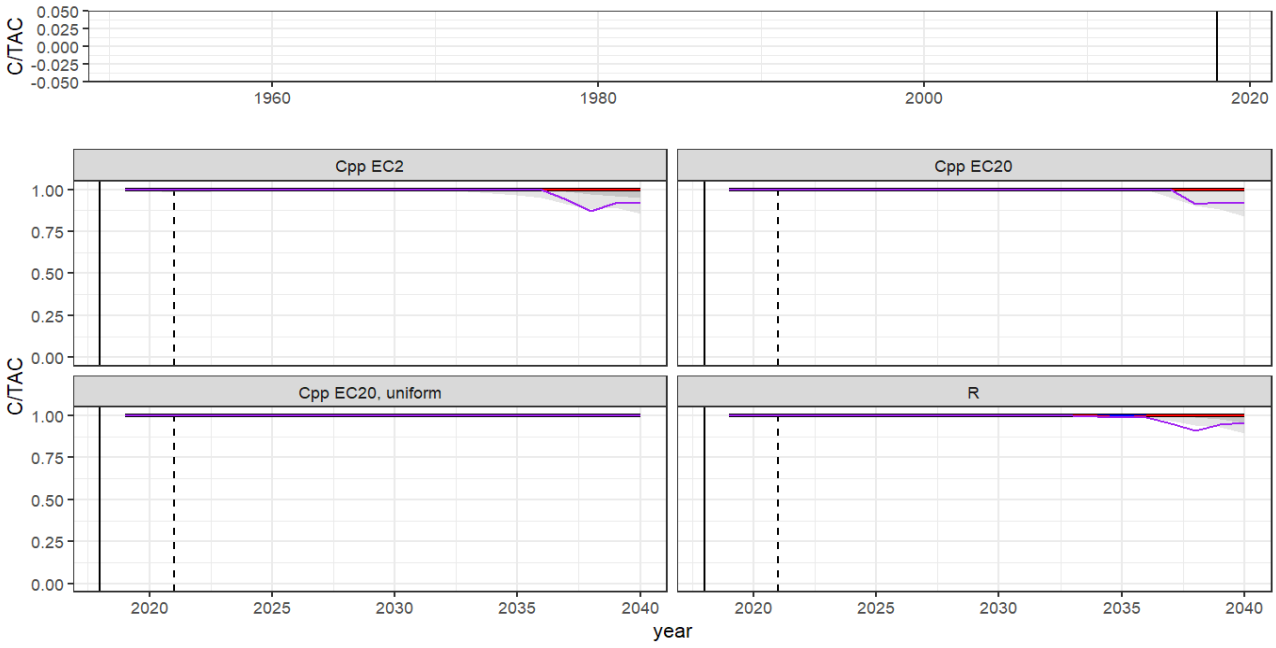

Figure 74. C/TAC over time comparing OM model implementations for MP PT41F.t15.tmb.B3. Historical estimates irrelevant and blank in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021).

6 Key Points for the IOTC MSE Task Force Consideration:

We welcome feedback on all elements of the MSE work, and suggest the following priority points for the IOTC MSE Task Force to consider:

1. At this time we are not aware of any obvious need to alter the BET OMs as proposed by the WPTT/WPM 2020 and as implemented in Appendix B. However, we expect that ongoing investigations for other IOTC species may prove relevant for BET as well, e.g. including:
    - Improved representation of standardized CPUE series uncertainty by the CPUE Working Group (e.g. potentially including alternative CPUE series based on alternative standardization approaches, regional scaling factors and/or new insight about the plausibility of catchability trend assumptions).
    - The role of model diagnostics for weighting models in the OM ensembles (and the allocation of models between reference set and robustness ensembles), will presumably continue to evolve, particularly in relation to the new YFT assessment.

2. The range of MPs tested for BET seem to suggest that the biggest determinant of MP performance will likely be the tuning objective selected by the TCMP, and both of B2 and B3 seem reasonable at this time. We would consider the PTRE hockeystick MP (PT41FM.t15.tmb) to be the best in terms of minimizing conservation risk, while the PTRE projection MP (PTBoB0Targ.t15) seems to be slightly more stable over time, without the rising catches and fishing mortality near the end of the projection period. We would suggest presenting the TCMP with comprehensive results for these 2 MPs X 2 tuning objectives. We would expect to tune a handful of additional MPs derived from these two, but with contrast in other control parameters.

3. Since there is seemingly no urgency or mechanism for an MP to be adopted by the Commission in 2021, there are still avenues that could be explored for improving MPs for BET (and other species), including:
    - incorporate size composition data as an index of incoming recruitment
    - systematic exploration of the interactions among control parameters
    - incorporation of PTRE uncertainty estimates within the Harvest Control Rule.

4. If the task force is satisfied with progress, it is probably be time to start finding the mechanism to move the Commission toward MP adoption. The current CSIRO project to provide BET and YFT MSE scientific and technical support will conclude in June 2021. The

Australian government has expressed interest in continuing to fund this work and a decision is expected in the near future.

References

Fu, D. 2019. Preliminary Indian Ocean bigeye tuna stock assessment 1950-2018 (Stock Synthesis). IOTC-2019-WPTT21-61

Kolody, D, Jumppanen, P. 2020A. A candidate Management Procedure based on a Joint Process and Observation Error Random Effects Production Model. IOTC–2020–WPM11–13

Kolody, D, Day, J, Jumppanen, P. 2020B. Indian Ocean Yellowfin Tuna Management Procedure Evaluation Update April 2020. Report prepared for the Indian Ocean Tuna Commission Informal Management Strategy Evaluation workshop 2020 (This document will be assigned a number in the IOTC archive at a future date).

Kolody, D, Jumppanen, P. 2021. Indian Ocean Yellowfin Tuna Management Procedure Evaluation Update March 2021. Working Paper prepared for the Management Strategy Evaluation Task Force of the Indian Ocean Tuna Commission Working Party on Methods Meeting, March 2021. IOTC-2021-WPM12(MSE)-03.

Methot, R.D., Wetzel, C.R., 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142: 86– 99.

WPM 2020. Report of the 11th Session of the IOTC Working Party on Methods. Microsoft Teams Online, 14 -15 October 2020. IOTC–2020–WPM11–R[E].

WPTT 2020. Report of the 21st Session of the IOTC Working Party on Tropical Tunas. Virtual Meeting, 19-23October2020. IOTC–2020–WPTT22(AS)–R[E]_Rev1.

CONTACT US

t 1300 363 400
+61 3 9545 2176
e enquiries@csiro.au
w www.csiro.au

AT CSIRO WE SHAPE THE FUTURE

We do this by using science to solve real issues. Our research makes a difference to industry, people and the planet.

As Australia's national science agency we've been pushing the edge of what's possible for over 85 years. Today we have more than 5,000 talented people working out of 50-plus centres in Australia and internationally. Our people work closely with industry and communities to leave a lasting legacy. Collectively, our innovation and excellence places us in the top ten applied research agencies in the world.

WE ASK, WE SEEK AND WE SOLVE

FOR FURTHER INFORMATION

CSIRO Oceans and Atmosphere
Dale Kolody
t +61 3 6232 5121
e dale.kolody@csiro.au
w https://www.csiro.au/en/Research/OandA

Appendix A. Extracts from the 2020 Methods and Tropical Tuna Working Party reports relevant to bigeye MSE Technical Workplan

The WPM and WPTT 2020 did not make any specific recommendations for updating the bigeye MSE reference set Operating Models for 2021. The WPTT 2020 noted:

152. The WPTT **NOTED** that some of the OMs appear to have unrealistically high fishing mortality in a small number of age/region/quarter strata, and it remains unclear whether this is i) a genuine problem that has serious effects on inferences from some models, ii) a genuine situation with a trivial effect on some models, or iii) an artefact of misleading labels in Stock Synthesis or r4ss outputs.

155. Similar to the discussion for yellowfin tuna, the WPTT **SUGGESTED** using an additional robustness scenario for bigeye that includes both 5% overcatch reported and 5% overcatch not reported.

Appendix B. Brief summary of the State of the IOTC Bigeye Tuna MSE Operating Models as of March 2021.

State of the IOTC Bigeye Operating Models for Management Procedure evaluation March 2021

Dale Kolody (dale.kolody@csiro.au)

Paavo Jumppanen

CSIRO, Australia

Introduction

This document provides a brief description of the most recent state of the Indian Ocean bigeye tuna Operating Models (OMs) used for Management Procedure (MP) evaluation, including the reference set *OMrefB.20.1,* and 6 robustness tests requested by the WPTT and WPM. The documentation for the latest version of the MSE software, technical documentation, and series of project reports is publicly available from github https://github.com/pjumppanen/niMSE-IO-BET-YFT/. The iterative and sometimes circuitous decision process undertaken by the IOTC technical working groups and analysts to reach the current state of the OM are not described here. These may be found in various IOTC working papers, information papers and meeting reports, along with various model results and diagnostics that were used to guide the OM development process.

The reference set OM, *OMrefB20.1,* is described in more detail in Kolody et al. (2020). It was endorsed by the WPTT (2020), and has not changed since then.

Conditioning Software

This version of the OM is an ensemble of models conditioned using the *Stock Synthesis* (SS) assessment software version SS3.24z (e.g. Methot and Wetzel 2013).

Projection Software

The projection software is custom built, available from https://github.com/pjumppanen/niMSE-IO-BET-YFT/. The population dynamics equations conform to fairly standard assumptions, and are fully documented in the technical reference (also on github).

OM structure

The various models in the OM ensemble are derived from the most recent bigeye stock assessment (Fu 2019). Key differences from the assessment include: i) aggregation of the 4 temperate season-specific CPUE series, into a single series (each independently renormalized by the respective series mean over a common period of non-missing observations), ii) recruitment deviations for the most recent 12 quarters were constrained to the stock recruitment relationship (though lognormal noise was introduced into the initial age structure for the MSE projections)m and iii) some parameter bounds were relaxed if the bound seemed likely to be influential to the model outcome. Key structural assumptions include:

- 4 regions (Figure 1) with age-dependent movement
- Quarterly dynamics (implemented with calendar quarters defined as SS model-years)
- 15 fisheries (Table 1)
- Beverton-Holt recruitment dynamics

- Parameter estimation objective function includes
    - Standardized longline CPUE (1 series per region)
        - Only the Hooks Between Floats option was used for the standardized tropical LL CPUE series (the cluster analysis of the previous OM iteration was not included because the CPUE group did not update that series)
    - Size composition data
    - Tags (excluded in some OM scenarios)
    - Penalties on recruitment deviations from stock recruit relationship and mean spatial distribution
    - Diffuse priors on all estimated parameters
- Estimated parameters:
    - Fishery selectivity (stationary, various functional forms, parameters shared among some fleets)
    - Longline catchability - regional scaling factors are used to scale relative density to relative abundance among regions
    - Virgin recruitment
    - Recruitment deviations from the Beverton-Holt stock-recruit relationship, and recruitment spatial partitioning among tropical regions (no spatial deviations over time).
    - Juvenile and adult movement rates
    - Initial fishing mortality
- Other fixed parameters and assumptions are either adopted as in the Fu (2019) assessment, or the grid structure described below.

Reference Set Grid OMgridB20.1

- Model structural and parameter uncertainty is introduced to the OM by combining the alternative assumption options listed in Table 2.
- Only the point estimates (maximum posterior density) for parameters and initial states are used in the OM.
- A fractional-factorial experimental design was used to select a subset of 72 models for fitting, which would allow the estimation of all main effects in the context of a GLM (the full factorial grid with all interactions would require 432 models).
- In recognition that the IOTC bigeye assessment model parameter estimates can be sensitive to initial starting conditions, minimization was repeated from randomly jittered starting conditions until either (i) successful minimization was achieved 3 times (maximum gradient of the objective function with respect to the estimated parameters <0.01) or (ii) 10 attempts at minimization were completed.

OM Reference Set *OMrefB20.1 (subset of OMgridB20.1)*

- Within an individual model configuration, the replicate with the lowest objective function value (from the jittered minimizations) was retained (initially). The best fit models were subsequently rejected from the reference grid if:
    - Minimization unsuccessful (max. grad. >0.01) – in this iteration, there were no failures following the repeated, jittered minimization process
    - The SS3 Catch Penalty (i.e. model struggles to remove the observed catch, which is assumed to be related to the pessimistic retrospective patterns). This potentially could indicate a serious problem, but was ignored in this iteration.
- All retained models were subject to a qualitative comparison of simple diagnostics to identify outlier behaviour or polymodal stock status inferences (no obvious problems were noted). The four most extreme models (highest and lowest depletion and productivity) were visually examined in more detail, without obvious evidence for blatant model failure (e.g. systematic lack of fit).
- Each SS model is assigned a plausibility weighting. To date, models have only been assigned a weighting of 0 or 1, such that all retained models are uniformly weighted. *OMrefB20.1* consists of 500 models randomly sampled (with replacement) from the grid of retained models.
- Key projection assumptions are summarized in Table 3.

References

Fu, D. 2019. Preliminary Indian Ocean bigeye tuna stock assessment 1950-2018 (Stock Synthesis). IOTC-2019-WPTT21-61

Kolody, D, Jumppanen, P, Day J. 2020. Indian Ocean Bigeye Tuna Management Procedure Evaluation Update March 2020. Report prepared for the Indian Ocean Tuna Commission Informal Management Strategy Evaluation workshop 2020. IOTC-2020-WPM11-11

Methot, R.D., Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research 142 (2013)* 86–99.

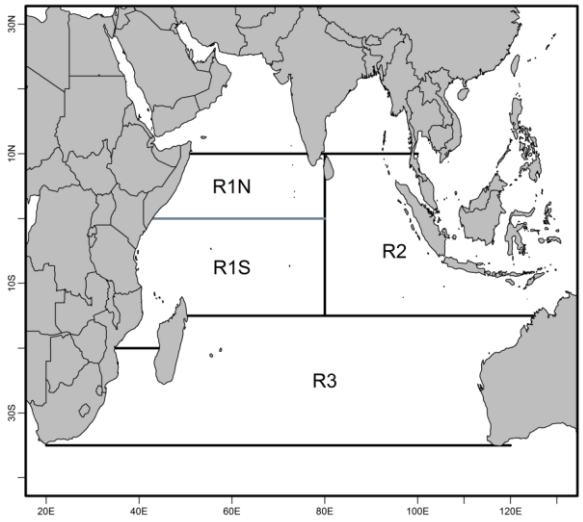

Figure 1. Spatial structure for the bigeye tuna OM (figure from Fu 2019).

Table 1. Fishery definitions in the BET 2016 assessment (from Fu 2019).

Code	Method	Region	Flag	Notes
FL2	Longline, fresh tuna fleets	2	All	
LL1N	Longline, distant water	1N	All	
LL1S	Longline, distant water	1S	All	
LL2	Longline, distant water	2	All	
LL3	Longline, distant water	3	All	
PSFS1N	Purse seine, free school	1N	All	
PSFS1S	Purse seine, free school	1S	All	
PSFS2	Purse seine, free school	2	All	
PSLS1N	Purse seine, associated sets	1N	All	
PSLS1S	Purse seine, associated sets	1S	All	
PSLS2	Purse seine, associated sets	2	All	
BB1	Baitboat and small-scale encircling gears (PSS, RN)	1N	All	Primarily catch from the Maldives baitboat fishery.
LINE2	Mixed gears (hand-line, gillnet/longline combination)	2	All	Gears grouped on the basis that primarily catch large bigeye.
OT1	Other (trolling, gillnet, unclassified)	1N	All	
OT2	Other (trolling, gillnet, unclassified)	2	All	

Table 2. Assumptions in OMrefB20.1 Stock Synthesis conditioning. Assumptions not listed are adopted from the Fu (2019) assessment (bold indicates the reference case assumptions in the assessment), or described in the text above.

Abbreviation	Definition
	Stock-recruit function (*h* = steepness)
h70	Beverton-Holt, *h* = 0.7
h80	**Beverton-Holt, *h* = 0.8**
h90	Beverton-Holt, *h* = 0.9
	Natural mortality multiplier relative to reference case M vector
M10	**1.0**
M08	0.8
M06	0.6
	Tag recapture data weighting (tag composition and negative binomial)
t0001	λ = 0.001
t01	λ = 0.1
t10	**λ = 1.0**
	Assumed longline CPUE catchability trend (compounded)
q0	**0% per annum**
q1	1% per annum
	longline CPUE Regional-scaling factors
iR1	preferred estimate from Hoyle (2018) – 7994_m8
iR2	alternate from Hoyle (2018) – 8000_m8
	Longline fishery selectivity
SL	**Stationary, logistic, shared among areas**
SD	Stationary, double-normal (potentially dome-shaped), shared among areas
	Size composition input Effective Sample Sizes (ESS)
ESS10	**ESS = 10, all fisheries**
CLRW	ESS = One iteration of re-weighting from reference case model, applied at the level of the individual observation, capped at 100.

Table 3. OM Projection assumptions in the bigeye reference set and robustness sets. Reference set values not listed are identical to the model-specific conditioning assumptions/estimates. Robustness case values are identical to the reference set except as noted in Table 4.

OM	Projection assumption	Value
OMrefB20.1	Reference set OM	
	Initial population error CV (a = age in quarters)	0.6exp(-0.1a)
	Recruitment deviation penalty Recruitment deviation lag(1) auto-correlation (these are annual values, but they are parameterized by the quarterly quarterly equivalents)	$max(\sigma_R$ = 0.42, SS estimate) $max(\rho_R$ = 0.21, SS estimate)
	CPUE observation error CPUE observation error lag(1) auto-correlation (implemented annually)	$max(\sigma_I$ = 0.2, SS estimate) $max(\rho_I$ = 0.5, SS estimate)
	Multinomial Catch-at-length sample size (all fisheries)	100
	Selectivity stationary for all fisheries	
	Quota Implementation error	CV = 0
	First MP quota year	2022
	Bridging catches 2019-2021	81 Kt (2018 Observed Catch)
	MP data lag (i.e. data up to and including 2020 would inform the 2023 quota)	2 years
	Quota allocation (average observed over)	2017-2018
	Number of stochastic realizations	500

Table 4. Robustness tests requested for BET by the WPTT (2020). Conditioning, and other assumptions not listed are identical to the reference set (Table 3).

1)	Increased Longline CPUE error variance $\sigma_I = 0.3$, $\rho_I = 0.5$
2)	10% overcatch, accurately reported
3)	10% overcatch, unreported
4)	10% overcatch, 5% reported, 5% not reported
5)	8 consecutive quarter recruitment shock (55% of average, near start of projections)
6)	3% per year LL catchability trend (not in SS conditioning; projections only)