OCEANS AND ATMOSPHERE
www.csiro.au

Update on IOTC Bigeye Tuna MSE Operating Model Development October 2018

Dale Kolody and Paavo Jumppanen

Prepared for the Indian Ocean Tuna Commission Working Party on Methods and Working Party on Tropical Tunas, Oct-Nov 2018

Acknowledgments

This work was jointly funded by the Common Oceans - Areas Beyond National Jurisdiction Tuna Project (administered by FAO), and Australia's Commonwealth Scientific and Industrial Research Organization (CSIRO - Oceans & Atmosphere).

Contents

1 Summary 1

2 Introduction 3

3 Bigeye Reference Case OM Conditioning 4

3.1 Relationship between the stock assessment and Operating Model 4

3.2 BET OMgrid18.2 CPUE characteristics 11

3.3 Seasonal and spatial dynamics 12

3.4 BET OMgrid18.2 Convergence Diagnostics 12

3.5 Summary characteristics for bigeye reference case Operating Model OMrefB18.5 18

3.6 Numerical considerations of high fishing mortality 29

3.7 Revisiting BET OM projection variance assumptions 31

3.8 Reference set OMref18.5 projection dynamics 35

4 BET Robustness scenarios 37

5 Discussion 38

References 41

1 Summary

This paper summarizes progress on the development of Operating Models (OMs) for IOTC bigeye (BET) tuna. Additional background detail on recent software developments is provided in the yellowfin (YFT) companion paper (Kolody and Jumppanen 2018f). MP evaluation updates for BET and YFT are described in Kolody and Jumppanen (2018a). This paper builds on the work presented and reviewed at the IOTC informal MSE Working Group in March 2018 (Kolody and Jumppanen 2018d,e), and represents the first time that the formal IOTC WPTT and WPM have the opportunity to review the substantial BET OM developments since the phase 1 work was completed in 2016.

The new BET reference case OM is structured around the 2016 BET assessment, notably including new longline CPUE analyses, and spatial disaggregation to facilitate a more appropriate inclusion of the tagging data. Progress on phase 2 BET MSE began with a "mechanical" update of the reference case Operating Model (OM) to address the 2016 IOTC WPTT/WPM requests. The original proposal (OMrefB18.1) was composed of an ensemble of 108 stock assessment models, representing uncertainty in 5 dimensions in an equally-weighted design. For compatibility with the OM software, a modification to the CPUE interpretation was adopted - instead of using 4 temperate CPUE series with independent q by season, the temperate series were merged by independently renormalizing each seasonal series and assuming a shared q (the effect on stock status inferences was negligible for the reference case assessment). The informal MWG recommended going forward with OMrefB18.2. OMrefB18.2 is derived from an equally balanced grid of 432 models:

- 3 X Beverton-Holt stock recruit relationship steepness
- 3 X Natural mortality vectors
- 3 X tag likelihood weighting
- 2 X CPUE standardization method
- 2 X CPUE catchability trend
- 2 X CPUE CV assumptions
- 2 X effective sample size weightings

The models in this grid were retained or rejected after consideration of several criteria, including:

- reliable convergence of the function minimizer
- informative parameters not on bounds. This included one iteration of relaxing problematic bounds inherited from the default assessment and refitting models
- inspection of the quality of fit indicators for CPUE and size composition data

The final reference case ensemble, OMref18.5, consists of 252 uniformly weighted model specifications.

The central tendency of OMrefB18.5 is similar to the assessment with respect to MSY and SB/SB(MSY) estimates, and slightly more pessimistic in terms of depletion, with substantial variance for all three quantities. The diagnostics for the quality of fit to the data are not as informative as we might have hoped for evaluating model plausibility. Recruitment deviation trends were not an obvious problem (unlike YFT).

Additional analyses are presented to consider how the quarterly auto-correlated recruitment variability assumptions assumed in projections relate to the annual values that are commonly discussed, and no urgent need for change was identified. The effects of high F assumptions on MP evaluation were examined and appear to be unimportant for the BET situation.

Projection assumptions for the proposed reference case OM, OMref18.5, included:

- Initial states (with added error) and most parameters are defined by the SS specifications
- quarterly CPUE CV = 0.2, auto-correlation = 0.5
- quarterly recruitment CV = 0.6, autocorrelation = 0.5 (corresponds to annual CV = 0.42, annual auto-correlation = 0.21)
- first TAC implemented in 2019; bridging catches 2016:2018 = 87Kt (2016 level)
- catch implementation error CV = 0
- stationary selectivity (at the terminal estimated values)

The following robustness scenario OMs were defined for BET:

- Recruitment failure (55% of expected) for the first 8 quarters of MP application (similar magnitude to the YFT robustness scenario defined from the recruitment time series estimated for the early 2000s in the assessment)
- TAC implementation error - TAC ignored for 10 years, then restrictive (i.e. simulates apparent lack of incentive for longline fleets to catch more BET at present)
- TAC implementation error - consistent 10% over-catch, all fisheries (reported without error)
- TAC implementation error - 40% CV applied equally to all fisheries (overall CV >10%)
- 3% per year longline CPUE catchability trend in the projections (conditioning assumptions as in the reference case)

The robustness test implications for MP performance are presented in Kolody and Jumppanen (2018a).

2 Introduction

The Indian Ocean Tuna Commission has committed to a path of using Management Strategy Evaluation (MSE) to meet its obligations for adopting the precautionary approach. IOTC Resolution 12/01 *"On the implementation of the precautionary approach"* identifies the need for fishery reference points and harvest strategies that will help to maintain the stock status at a level that is consistent with the reference points. Resolution 13/10 "*On interim target and limit reference points and a decision framework*" identified interim reference points and elaborated on the need to formulate management measures relative to the reference points, using MSE to evaluate harvest strategies in recognition of the various sources of uncertainty in the system. Resolution 15/10 supersedes 13/10 with a renewed mandate for the Scientific Committee to evaluate the performance of harvest control rules with respect to the species-specific interim target and limit reference points, no later than 10 years following the adoption of the reference points, for consideration of the Commission and their eventual adoption. A species-specific workplan was re-affirmed at the 2017 Commission Meeting, outlining the steps required to adopt simulation-tested Management Procedures for the highest priority species (IOTC 2017). Recognizing the iterative nature of the MSE process, the workplan identified 2019 as the earliest possible date for MP adoption.

This paper describes i) the assumptions used for conditioning the current version of the reference case OM, ii) the process used to evaluate and reject or retain the models within the OM ensemble, iii) general characteristics of the final ensemble, iv) additional considerations for projections, and v) some robustness scenario OMs (which were used to test MPs in Kolody and Jumppanen 2018a). Considerations for the next iteration of the MSE process are discussed.

This paper assumes familiarity with fairly technical subject matter. More detailed explanations can be found in Kolody and Jumppanen (2016), Jumppanen and Kolody (2018) and various progress reports produced since the last YFT MSE update to the WPTT and WPM (Kolody and Jumppanen 2018a,b,c,d,e,f).

3 Bigeye Reference Case OM Conditioning

3.1 Relationship between the stock assessment and Operating Model

The intention has been to maintain a close relationship between the stock assessment modelling and the conditioning of OMs. The two processes are analogous in several respects, i.e. similar population dynamics models are fit to the same data, subject to the same concerns about model formulation and assumption violations, etc. The scientific process has been evolving rapidly in recent years. While the objectives of the two processes are different, it would be difficult to justify the two initiatives evolving in completely different directions. Accordingly, the bigeye assessment of Langley (2016) provides the core of the OM conditioning process. Key features of the assessment and OM include:

- Implementation with Stock Synthesis 3.24z software
- 4 regions (Figure 1)
- Quarterly dynamics, including recruitment and movement, using a configuration with calendar seasons defined as model years
- 15 fisheries (Table 1)
- Beverton-Holt recruitment dynamics
- Parameter estimation objective function includes
    - Standardized longline CPUE (Region 1A and 1B share one series, R2 has one series, and R3 estimates seasonal catchability by splitting the fishery and CPUE by season)
    - Size composition data
    - Tags (excluded in some OM scenarios)
    - Recruitment penalties on deviations from stock recruit relationship and mean spatial distribution
- 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, such that 1A, 1B, 2 share catchability and catchabilities are estimated independently for the 4 seasonal fisheries in region 3
    - Virgin recruitment

- Recruitment deviations from the Beverton-Holt stock-recruit relationship, recruitment spatial partitioning among tropical regions (1 and 2) and deviations from the mean spatial distribution. (check for BET)
- Juvenile and adult movement rates
- Initial fishing mortality

- Unlike the most recent YFT assessment, the BET assessment and management advice was based on an equally-weighted grid of 6 models in two dimensions:
  - three levels of stock-recruit steepness (h = 0.7, 0.8, 0.9)
  - two tag weighting assumptions (λ = 1.0, 0.1)

The various models, model ensembles and individual assumptions are summarized in Table 2 and Table 3. The latest version of the BET reference case OM is structured around the 2016 BET assessment, including new CPUE analyses, and spatial disaggregation to facilitate a more appropriate inclusion of the tagging data. Progress on phase 2 BET MSE began with a "mechanical" update of the reference case Operating Model (OM) to address the 2016 IOTC WPTT/WPM requests. The 2016 reference case proposed by the WPTT/WPM was composed of an ensemble of 108 stock assessment models, representing uncertainty in 5 dimensions in an equally-weighted design. That grid was considered to understate the uncertainty associated with the CPUE standardization approach and relative weighting of the different data sources. An expanded grid of 432 models was proposed to the WPM informal MSE group - OMgridB18.2:

- 3 X Beverton-Holt stock recruit relationship steepness
- 3 X Natural mortality vectors
- 3 X Tag likelihood weighting
- 2 X CPUE standardization method
- 2 X CPUE catchability trend
- 2 X CPUE CV assumptions
- 2 X Effective sample size (length composition) weightings

The following sections discuss additional modifications to the basic OM, the evolution to OMrefB18.5 and projection assumptions.

Table 1. Fishery definitions in the BET 2016 assessment (note that the order does not correspond to the fishery numbering in the assessment files).

Fishery	Time period		
	2012-14	2014	2015
FL2	10,147	13,383	16,153
LL1N	10,416	7,552	5,764
LL1S	16,486	15,868	16,316
LL2	6,394	7,568	4,704
LL3	3,824	4,833	4,886
PSFS1N	1,073	888	2,013
PSFS1S	3,097	3,957	6,753
PSFS2	0	0	200
PSLS1N	5,418	6,958	7,233
PSLS1S	6,359	7,915	8,627
PSLS2	293	159	0
BB1	5,083	6,188	5,717
LINE2	6,395	9,001	8,132
OT1	1,931	2,468	2,492
OT2	3,553	4,313	4,050
Total	80,469	91,052	93,040

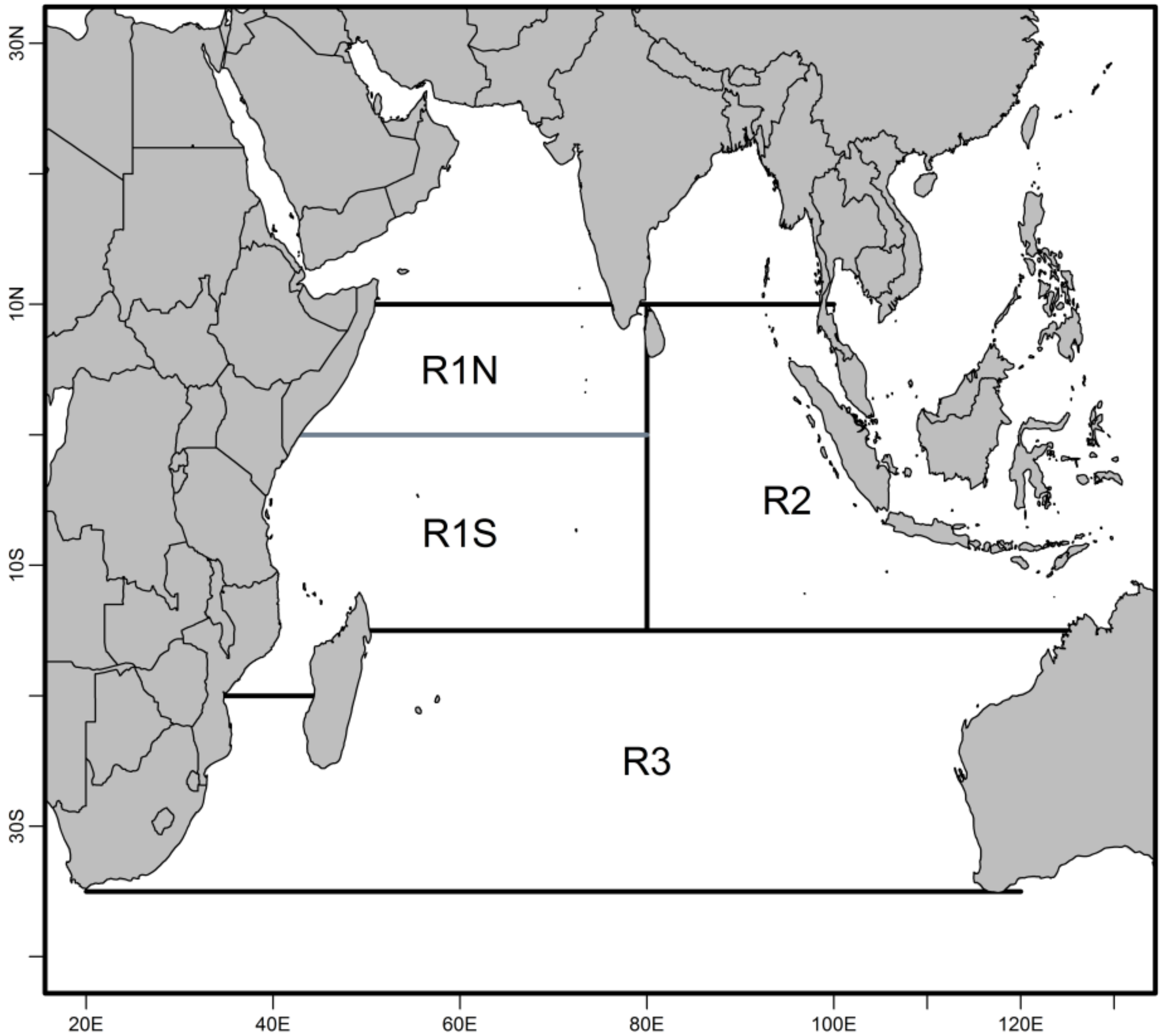


Figure 1. Spatial structure for bigeye tuna assessment and all OMs discussed in this report (figure from Langley 2016).

Table 2. Operating Model definitions. The OMs are listed in the order discussed in the text, reflecting the sequence of development.

Model Name	Definition (assumption abbreviations are defined in Table 3)
OMgridB18.1	Reference case OM as proposed by the WPM and WPTT in 2016. Consists of an ensemble of 108 models, derived from the assessment, with uncertainty in 5 dimensions h70, h80, h90 M10, M08, M06 t0001, t01, t10 q0, q1 iH, iC
OMgridB18.2	A grid consisting of an ensemble of 432 models: OMgridB18.1 with additional uncertainty in the weighting assumptions for the CPUE and size composition data: h70, h80, h90 M10, M08, M06 t0001, t01, t10 q0, q1 iH, i10H, iC, i10C ess10, CLRW
OMrefB18.5	OMgridB18.2 filtered for convergence with the max gradient < 0.01, and bounds constraints relaxed on ~80 models (results in 252 models)
OMrobB18.5.recShock	A robustness scenario with 8 consecutive quarters of poor recruitment (55% of expected values, similar to estimates for YFT in the early 2000s). (conditioning is unchanged from OMrefB18.5)

OMrobB18.5.impErrCV10	A robustness scenario in which each fishery has a 40% catch implementation error CV (independent by year and fishery). This corresponds to an annual aggregate CV >10%. (conditioning is unchanged from OMrefB18.5)
OMrobB18.5.under	A robustness scenario in which TACs are ignored for 10 years (fishing mortality constant at current levels) before the TAC is taken without error (conditioning is unchanged from OMrefB18.5)
OMrobB18.5.over	A robustness scenario with consistent 10% overcatch for all fleets (catch is accurately reported) (conditioning is unchanged from OMrefB18.5)
OMrobB18.5.qTrend3	A robustness scenario with a longline CPUE catchability trend of 3% per year in projections (conditioning is unchanged from OMrefB18.5)

Table 3. Model specification abbreviations. Bold indicates the BET assessment assumption(s). Some abbreviations may relate to additional explorations that were not completed, not reported, or pertain to YFT.

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**
Rh70	Ricker, *h* = 0.7
Rh80	Ricker, *h* = 0.8
Rh90	Ricker, *h* = 0.9
	Recruitment deviation penalty
sr4	$\sigma_R = 0.4$
sr6	**$\sigma_R = 0.6$**

sr8	σ_R = 0.8
	Future recruit failure
r55	3 years of poor recruitment (2019-2022); mean dev = -0.55, consistent with YFT assessment
	Natural mortality multiplier relative to SA-base
M10	**1.0**
M08	0.8
M06	0.6
	Tag recapture data weighting (tag composition and negative binomial)
t00	λ = 0
t0001	λ = 0.0001
t001	λ = 0.01
t01	**λ = 0.1**
t10	**λ = 1.0**
t15	λ = 1.5
	Assumed longline CPUE catchability trend (compounded)
q0	**0% per annum**
q1	1% per annum
q3	3% per annum
q5	5% per annum
	Tropical CPUE standardization method (error assumption for all series)
iH	**Hooks Between Floats (quarterly σ_{CPUE} = 0.2)**
i10H	Hooks Between Floats (quarterly σ_{CPUE} = 0.1)
iC	Cluster analysis (quarterly σ_{CPUE} = 0.2)

i10C	Cluster analysis (quarterly σ_{CPUE} = 0.1)
	Tag mixing period
x3	3 quarters
x4	**4 quarters**
x8	8 quarters
	Longline selectivity (in conditioning)
SS	**Stationary, logistic, shared among areas**
S4	LL selectivity independent among areas
NS	Temporal variability estimated in 10 year blocks
ST	Logistic selectivity trend estimated over time
Sdev	15 years of selectivity deviations estimated (XXX-XXX)[JP(H1]
Sspl	Cubic spline function (to admit possibility of dome-shape)
	Size composition input Effective Sample Sizes (ESS)
ESS2	ESS = 2, all fisheries
ESS5	ESS = 5, all fisheries
ESS10	**ESS = 10, all fisheries**
CLRW	ESS = One iteration of re-weighting; the output ESS from a reference case assessment specification (mean over time by fishery, capped at 100)

3.2 BET OMgrid18.2 CPUE characteristics

The alternative CPUE series adopted for the BET OM are presented in Kolody and Jumppanen (2018d). For compatibility with the OM software, a modification to the BET CPUE interpretation was adopted. The assessment assumed 4 temperate CPUE series with independent q by season, i.e. recognizing that there is a confounding between seasonal movements and catchability that the model structure cannot fully resolve. In the OM, the temperate series were merged by independently renormalizing each seasonal series, stitching them together into a single quarterly series and assuming a shared q. The effect on the point estimates of stock status on the reference case assessment were negligible.

3.3 Seasonal and spatial dynamics

Most of the IOTC discussion and management advice traditionally relate to annual summaries, and ignores the spatial and seasonal processes that the assessment models attempt to represent. This represents a potentially useful avenue for evaluating model plausibility that has been largely ignored to date. This iteration we provide information on both seasonal and annual variability for CPUE and recruitment. However, we still have not spent much effort evaluating spatial processes.

3.4 BET OMgridB18.2 Convergence Diagnostics

OMgridB18.2 consists of 432 SS model configurations. Additional time to evaluate model diagnostics has yielded further insight that was not available in time for the 2018 informal MSE working group (Kolody and Jumppanen 2018d).

Numerical convergence is usually evaluated on the basis of the maximum gradient (of the parameters with respect to the objective function) at the solution, and/or whether the inverse Hessian matrix (and delta-method uncertainty) can be calculated (the Hessian calculations were not conducted in the previous iteration due to time constraints). Figure 2 illustrates the relationship between the two convergence measures. The inverse Hessian calculations failed in 19 of 432 cases; only one failure was associated with a maximum gradient < 0.1. However, the Hessian calculation was also successful for a few models with a very large maximum gradient. The maximum gradient criterion forms a continuous distribution over the range of interest. We have considered two arbitrary minimum gradient criteria in the past: 0.1 and 0.01. These gradient thresholds (combined with a successful Hessian calculation) result in ensembles with 346 and 290 models respectively (further restricting the threshold to 0.001 results in 231 models). Visually comparing some stock status summaries suggests that the aggregate ensemble behaviour is not obviously sensitive to the convergence criteria (Figure 3, Figure 4). Accordingly, we opted to maintain the middle convergence criterion (0.01) to be consistent with the historical approach. Given that the Hessian calculation was taking an order of magnitude longer than the minimization (at least in some YFT scenarios), and was generally a less strict convergence criterion than the maximum gradient, we propose to stop doing the Hessian calculations in future iterations if time is an issue).

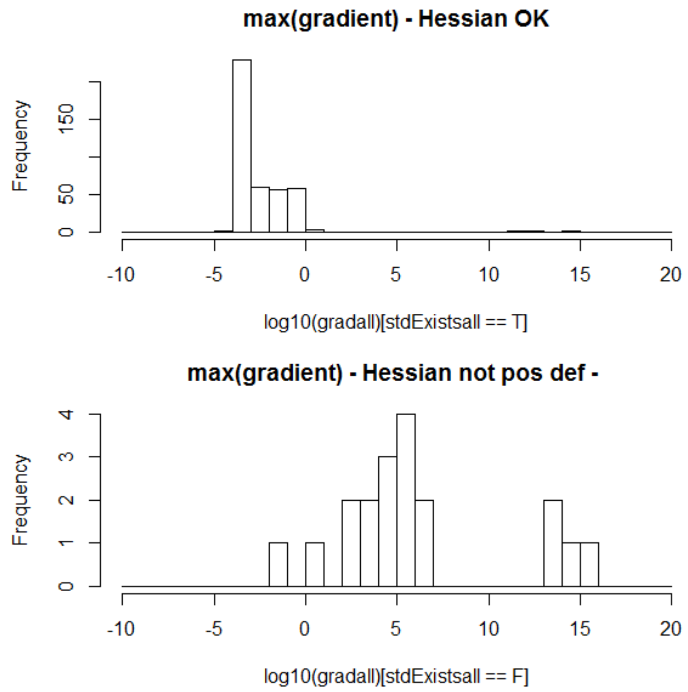


Figure 2. Grid OM 18.2 convergence diagnostics.

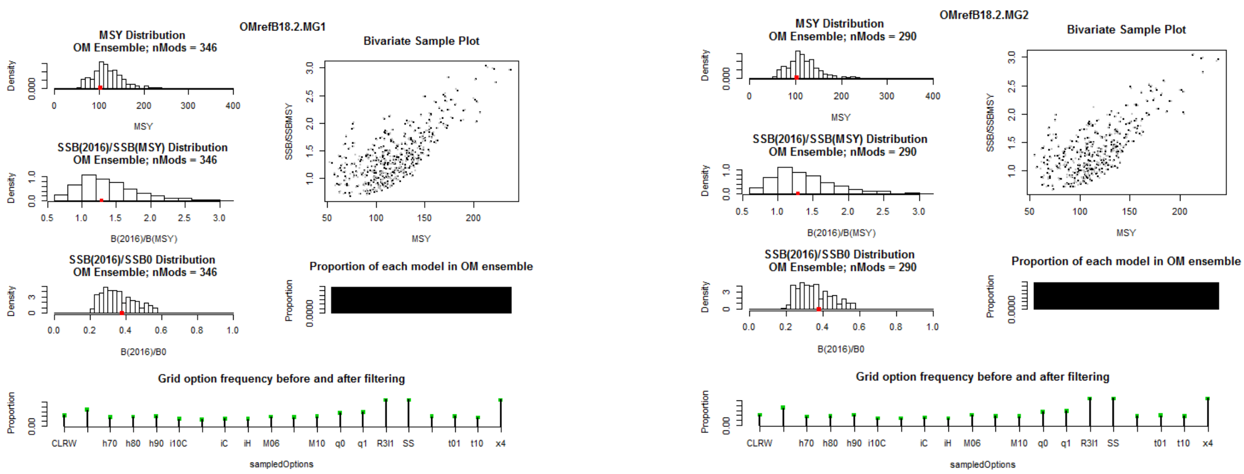


Figure 3. Summary characteristics of BET model ensembles assuming maximum gradient < 0.1 (left) and < 0.01 (right).

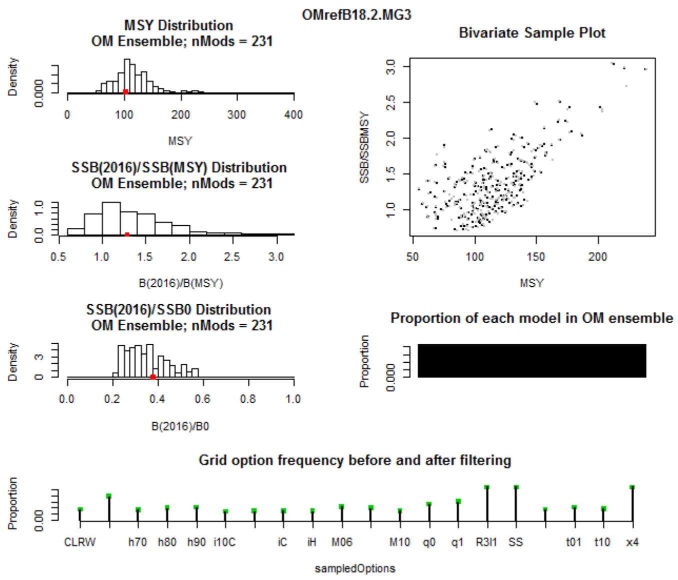


Figure 4. Summary characteristics of ensemble assuming maximum gradient < 0.001.

Parameter convergence to bounds represents another warning about model plausibility - SS reports parameters near bounds, and uses smooth penalties to ensure that bounds are (probably) not reached. In OMgridB18.2, the majority of models have at least one parameter near the bounds, and some have up to 5 (Figure 5). These bound exceptions are not closely associated with the numerical convergence failures. The questionable parameters fall into 3 categories: fishery selectivity functions, initial fishing mortality, and movement parameters. In the majority of cases, the questionable parameter approached a lower bound of zero, which represents a valid solution and the warning can be ignored (e.g. some fish are simply not vulnerable to a particular fishery, some fish may not move, or the initial fishing mortality may have been negligible). A few models hit an upper bound on movement rates. We have doubts about the model ability to estimate movement, but at this time we have no reason to argue that a particular movement rate estimate is unreasonable, as long as it is physically possible (i.e. 0-100%).

Out biggest concern about parameter bounds relates to the upper bound for some of the selectivity functions, notably the parameters for the oldest ages of fishery PSFS selectivity, and to a lesser extent BB1 selectivity (affecting 30% and 1% of the OMgridB18.2 respectively). This raises the issue of whether the default bounds inherited from the assessment are defensible values, and whether the model inferences are sensitive to the bounds?

As a test case, we selected a couple models with upper bound warnings for the PSFS (including the most extreme) and BB fishery selectivity constraints, and refit the models with the bounds raised by an arbitrary, but large amount, such that they are not restrictive (noting that the priors were already very uninformative). The bounds have a noticeable effect on the specific selectivity estimates (Figure 6 and Figure 7), but there is nothing to indicate that the unbounded selectivity functions are any less plausible than the bounded functions. Furthermore, there is a non-trivial effect on stock status measures (e.g. MSY changed by 30% as shown in Figure 8). This indicates that bounds problems cannot simply be ignored. Similarly, it is likely that bounds problems are associated with specific model formulations and assumption interactions, such that deleting the offending models may bias the OM ensemble. We opted to rerun the elements of the grid with the questionable parameter bounds relaxed. The final ensemble OMref18.5 is the result of OMgridB18.2, with about 80 models rerun with expanded selectivity bounds, and subsequently filtered for convergence criteria. No additional high parameter bound problems were identified.

We recognize that the issue with the bounds merits further consideration. We note that the reference case assessment that we checked did not report any bounds violations, though it is not known whether this was by design or good fortune. Perhaps some bounds are justifiable. Perhaps some of the models initially rejected with failed convergence would not have failed if the bounds were reset before the minimization was attempted. It is not necessarily an easy process to automate an approach to dealing with bounds problems with 100s of models and 100s of parameters. Sometimes relaxing one bound will simply force the model hit another bound. A similar issue arises in relation to priors - it remains to be checked whether informative priors are influential and defensible.

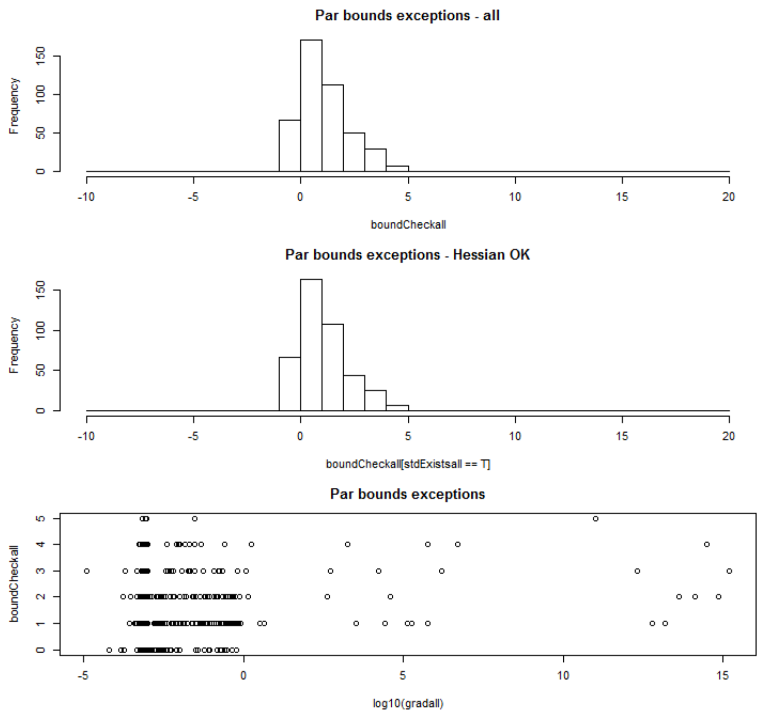


Figure 5. OMgrid18.2 parameter bound exceptions.

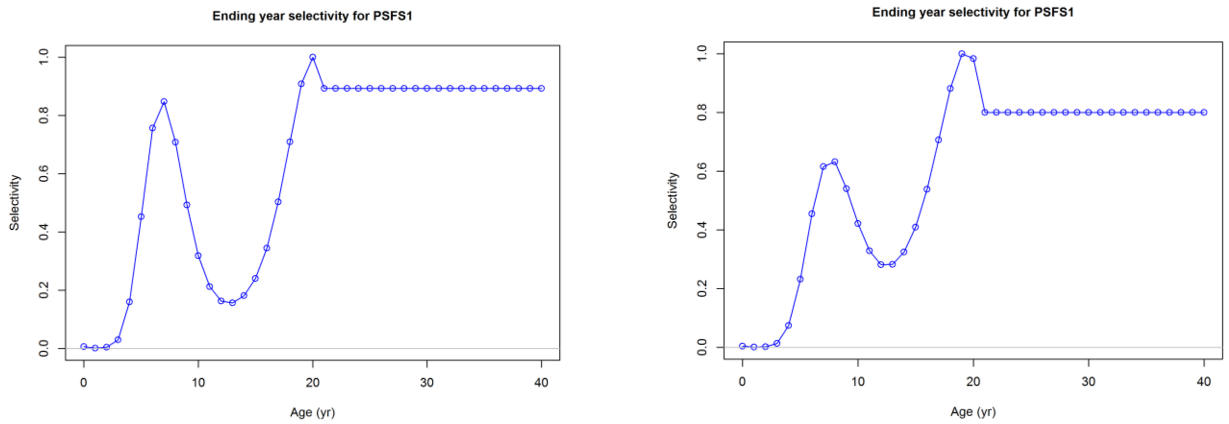


Figure 6. Comparison of model "R3I1_h80_M10_t0001_q1_x4_i10H_SS_CLRW" PSFS fishery selectivity with the default bounds constraints (left) and unconstrained bounds (right).

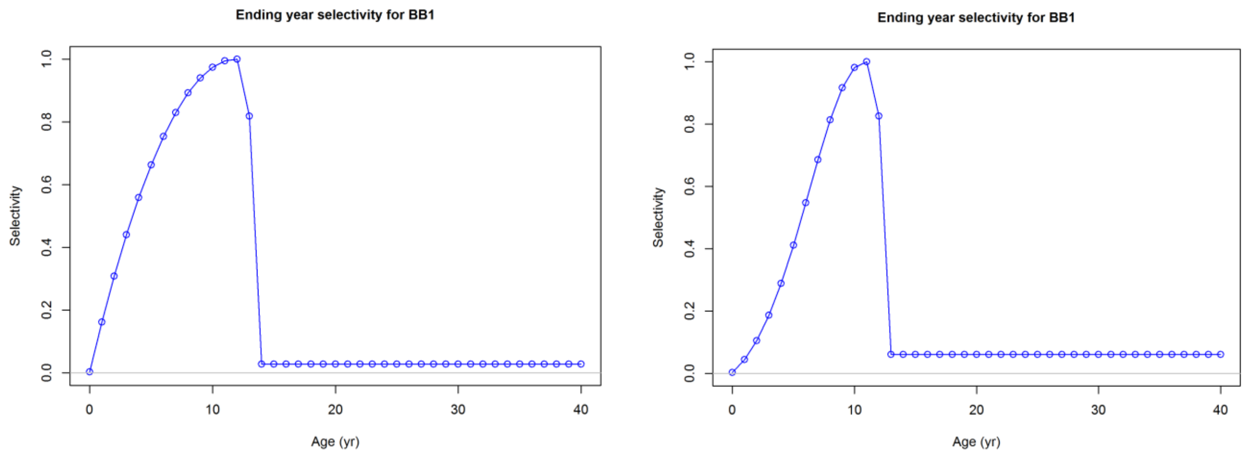


Figure 7. Comparison of model "R3I1_h70_M06_t01_q1_x4_i10C_SS_ess10" PSFS fishery selectivity with the default bounds constraints (left) and unconstrained bounds (right).

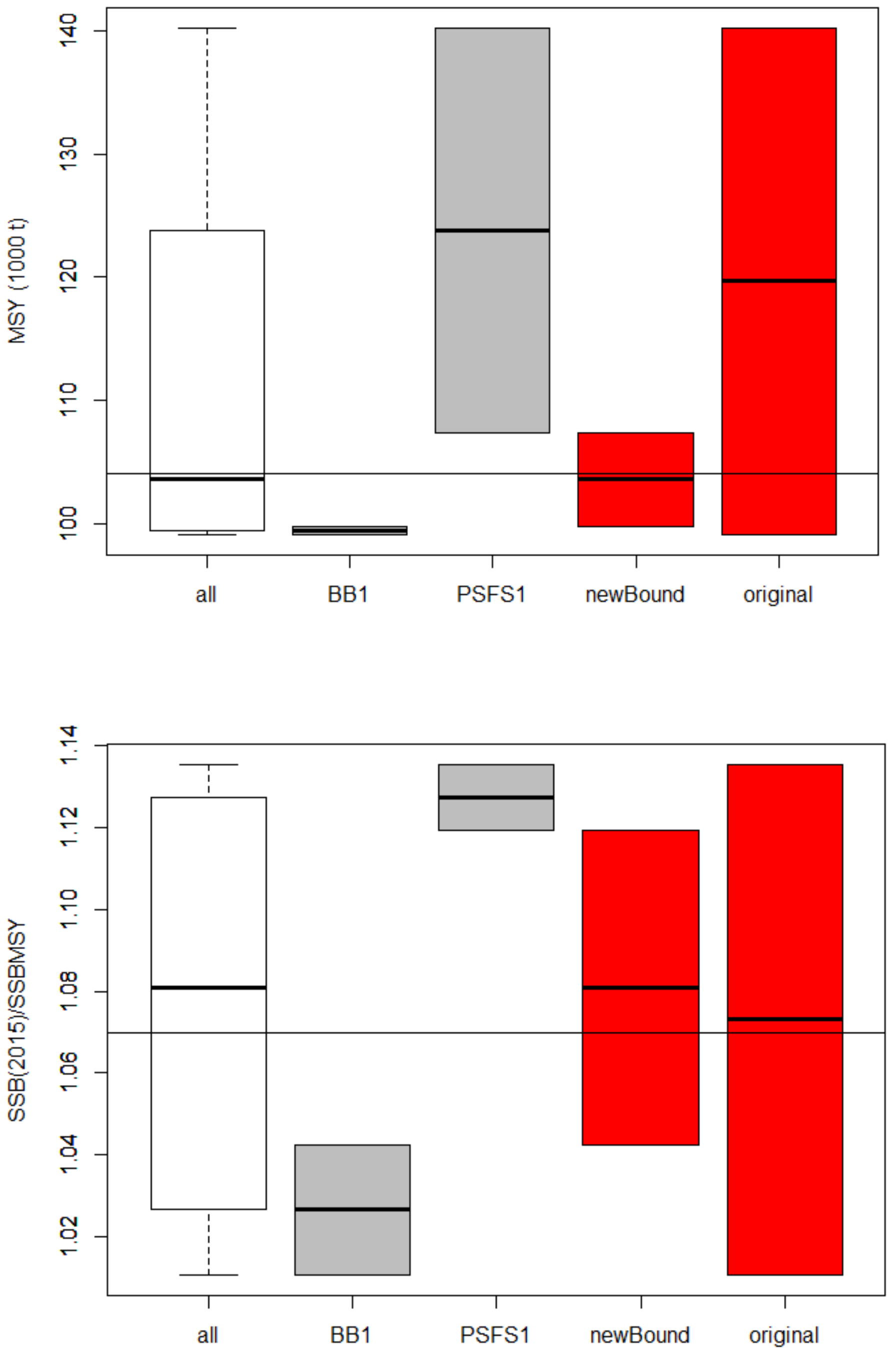


Figure 8. MSY (top) and depletion (bottom) estimates from two models with restrictive selectivity bounds for fisheries BB1 and PSFS1, and for the same models with the bounds removed. Models are aggregated in terms of the bound in question (grey boxes) and whether or not the bound was relaxed (newBound) or the original bound (original) in the assessment (red boxes).

3.5 Summary characteristics for bigeye reference case Operating Model OMrefB18.5

OMrefB18.5 consists of 252 equally-weighted models; 177 models from the full grid failed to meet the convergence criteria, and 2 models were removed because the quality of fit to the CL data were extreme outliers relative to the remainder.

Fit to the CPUE data are summarized in Figure 9 - Figure 13. Focussing on the inter-annual variability, it is notable that the fit to the CPUE series is very good (median of the mean among areas <0.15). The fits vary by region, but even the worst region (3) is very good (median of the means <0.2).

Tag likelihoods are summarized in Figure 15 - as would be expected, the tag weighting determines the quality of fit to the tags (and likelihood comparisons between weightings are not meaningful).

The (spatially and seasonally summed) recruitment variability is fairly consistent among assumptions - not surprisingly, greater weight to the tagging and size composition data introduces higher recruitment variability.

The distribution of key stock status descriptors are roughly centred on the assessment point estimates (Figure 17). There is a substantial degree of variability (e.g. MSY range is roughly half to double the assessment point estimate). The individual assumption levels are represented fairly evenly in the final ensemble relative to the original grid (though it is conceivable that some interactions were more prone to convergence failure and this is not discernible in the figures).

Key stock status indicators are further summarized in Figure 18, partitioned by assessment assumption. As would be expected, more pessimistic results tend to be associated with lower steepness, lower M and a 1% catchability trend increase. Higher weighting to the CPUE and one iteration of CL reweighting (almost always an increase in CL weighting) also tend to be pessimistic. The tropical HBF CPUE standardization (as used in the assessment) tends to be more pessimistic than the CPUE series derived from the cluster analysis.

As a consequence of the more detailed plausibility screening, OMrefB18.5 appears to be slightly more pessimistic, and the central tendency is slightly more consistent with the assessment, than the uniform grid that was examined at the informal WPM in March 2018.

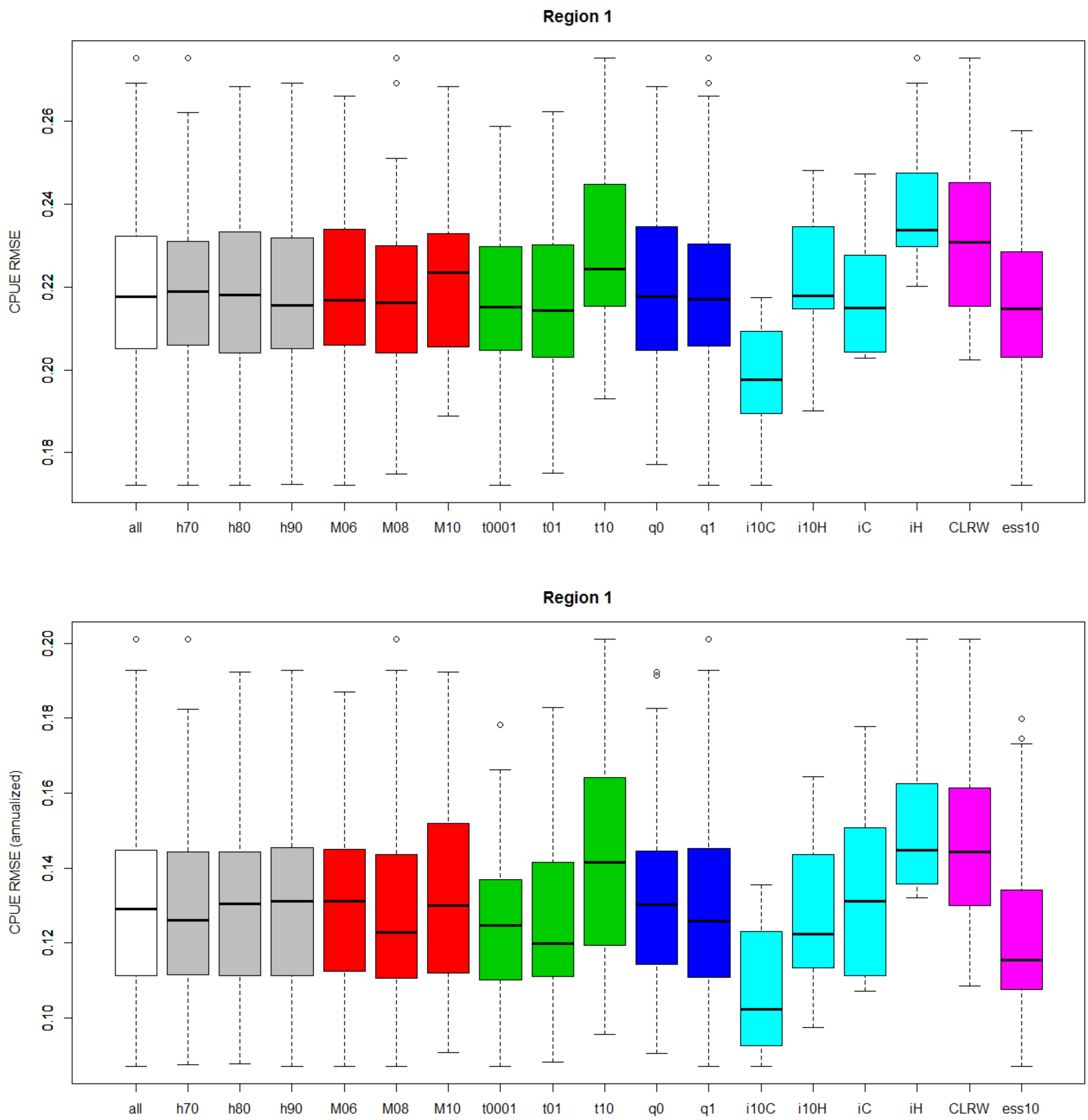
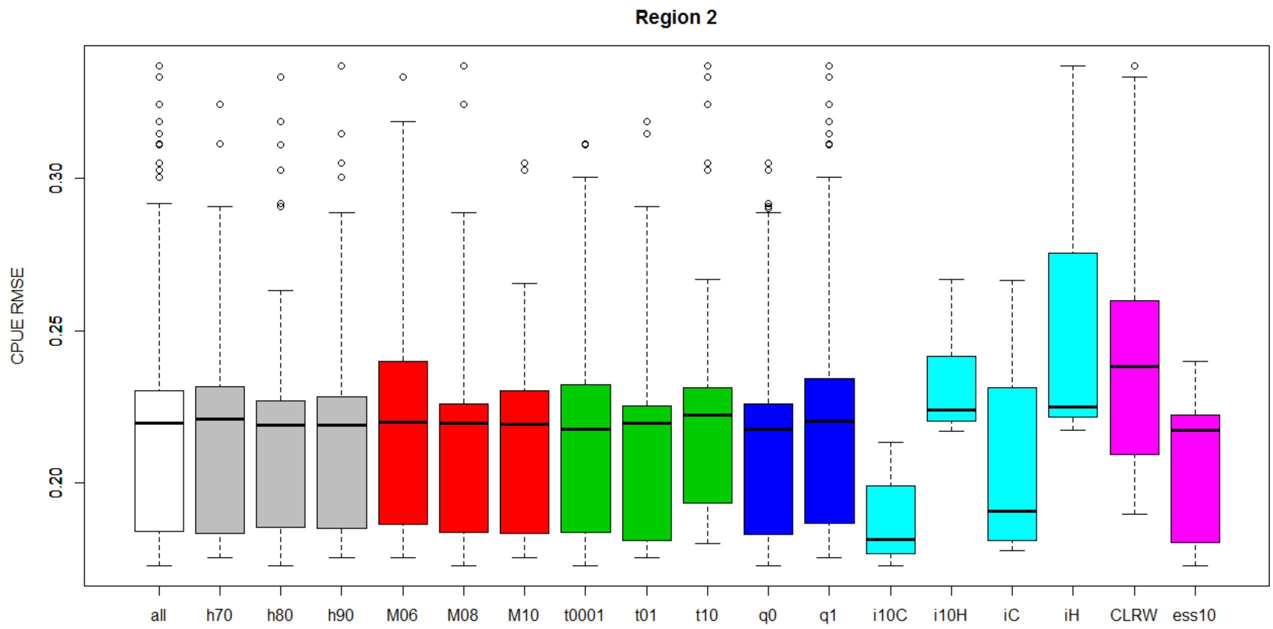


Figure 9. OMrefB18.5 quality of fit (RMSE) for the CPUE series in region 1 by quarter (top) and year (bottom).



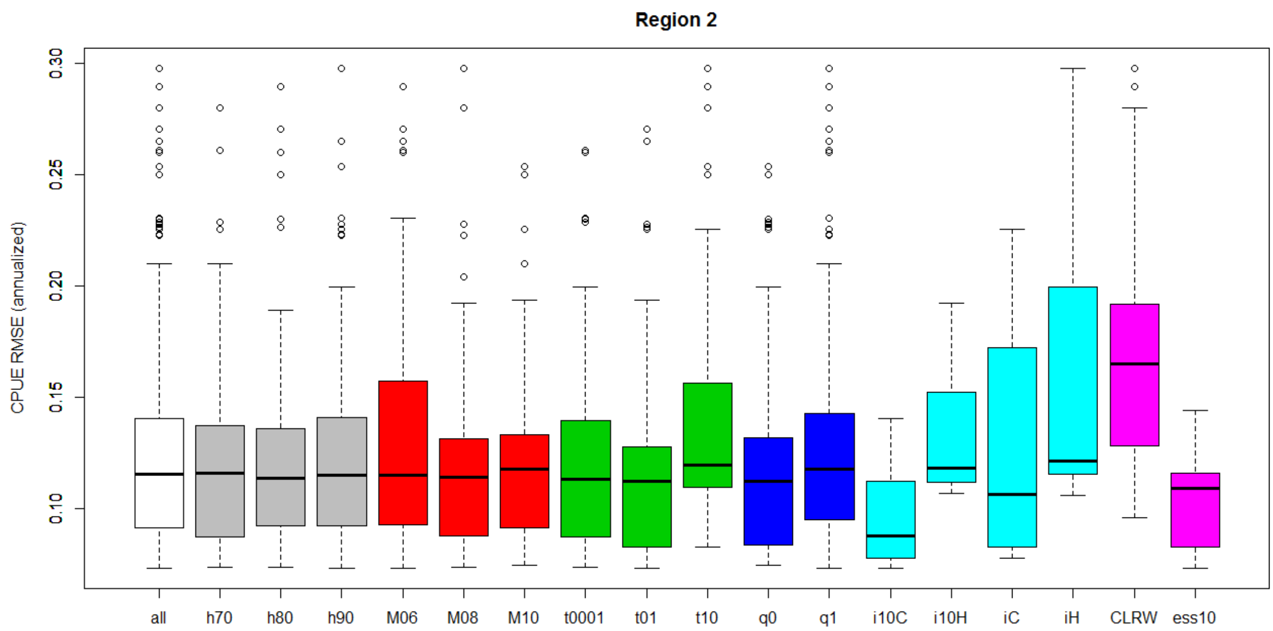
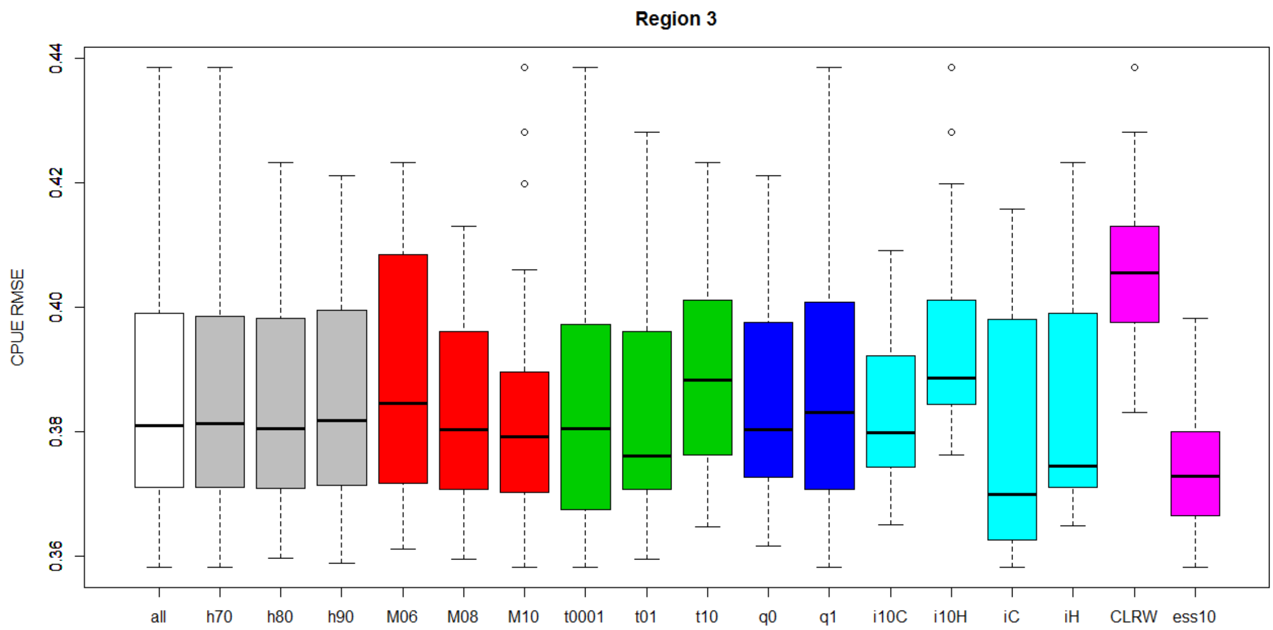


Figure 10. OMrefB18.5 Quality of fit (RMSE) for the CPUE series in region 2 by quarter (top) and year (bottom).



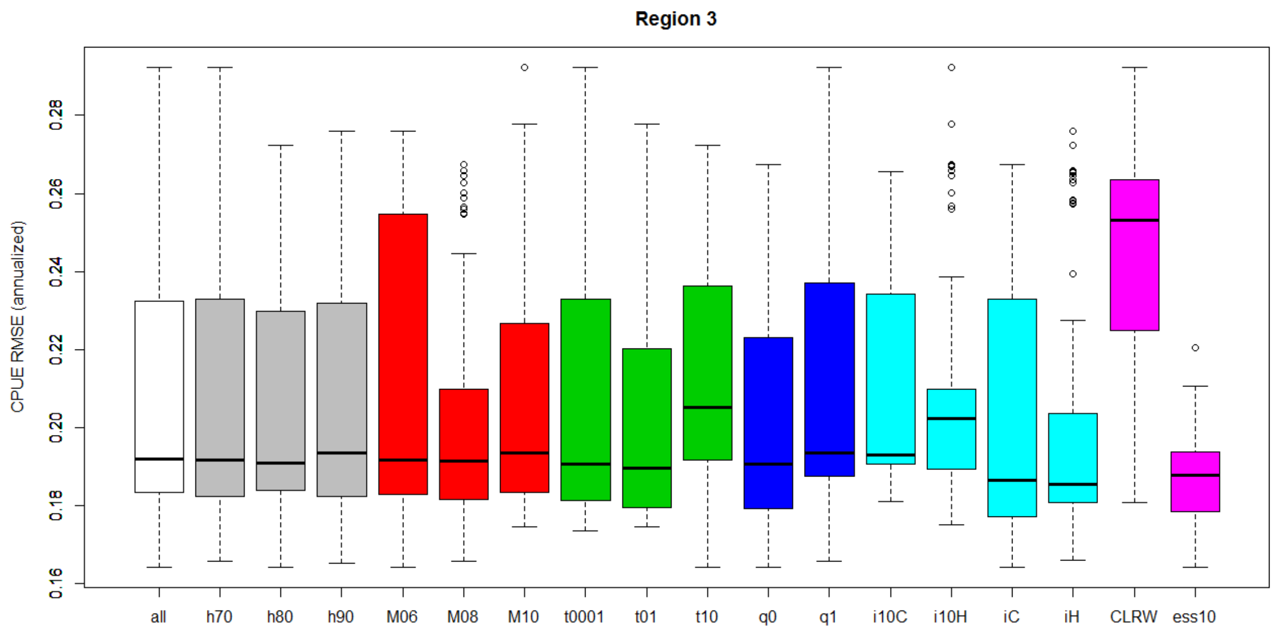
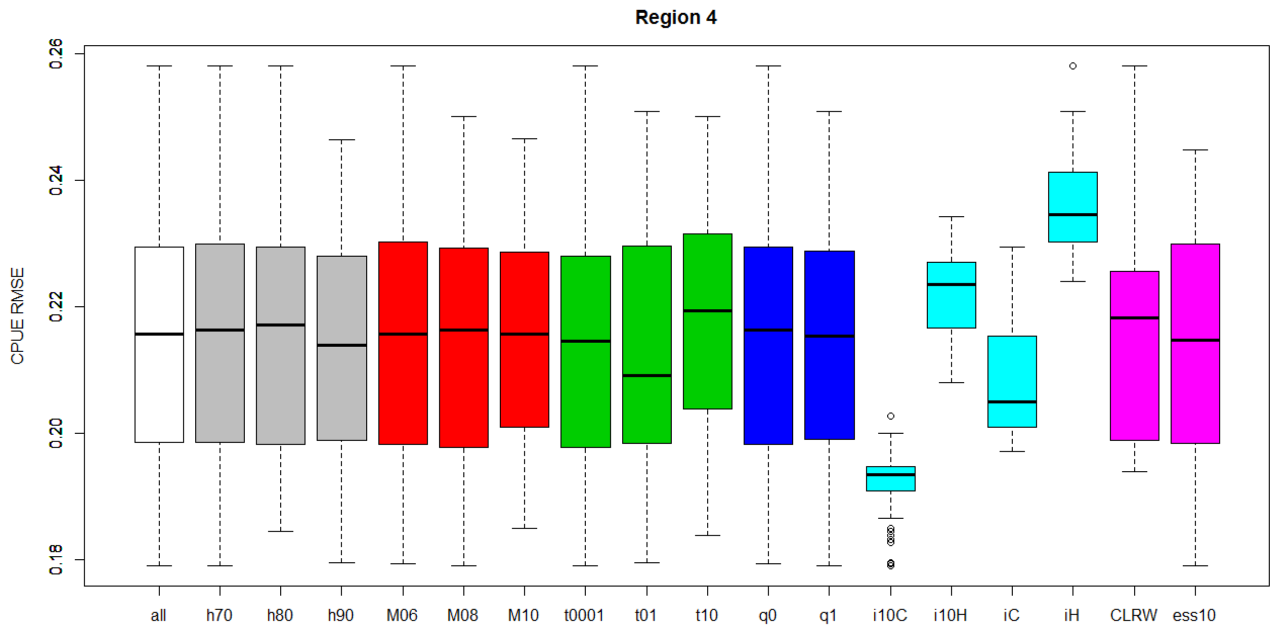


Figure 11. OMrefB18.5 Quality of fit (RMSE) for the CPUE series in region 3 by quarter (top) and year (bottom).



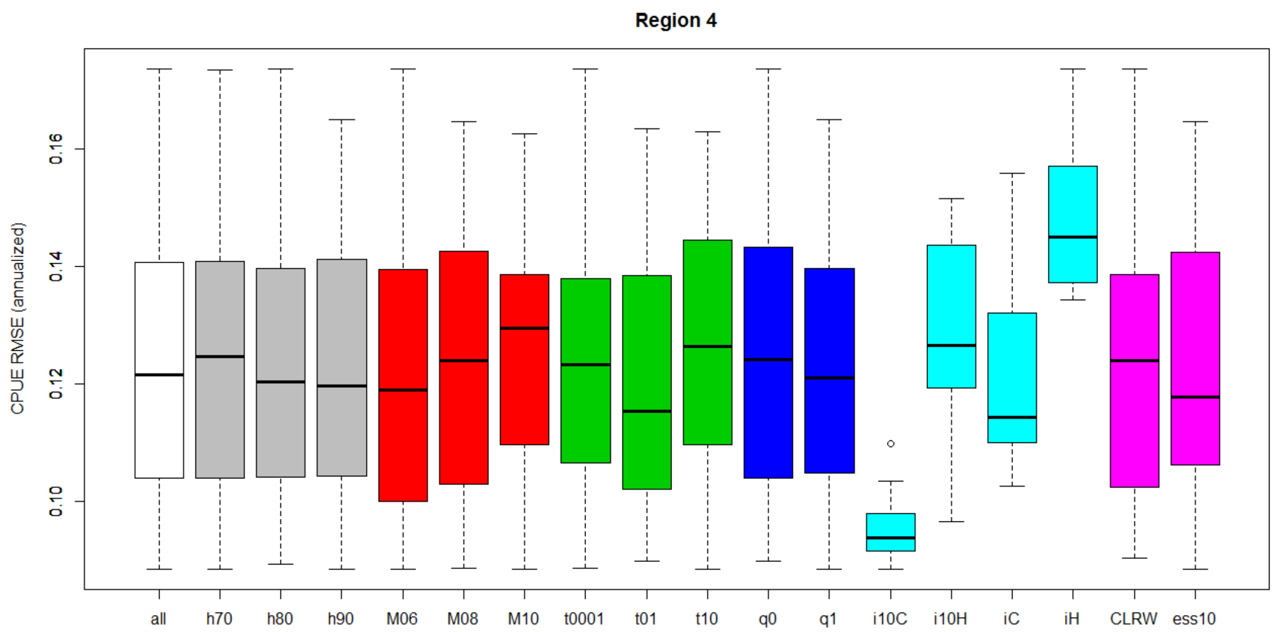
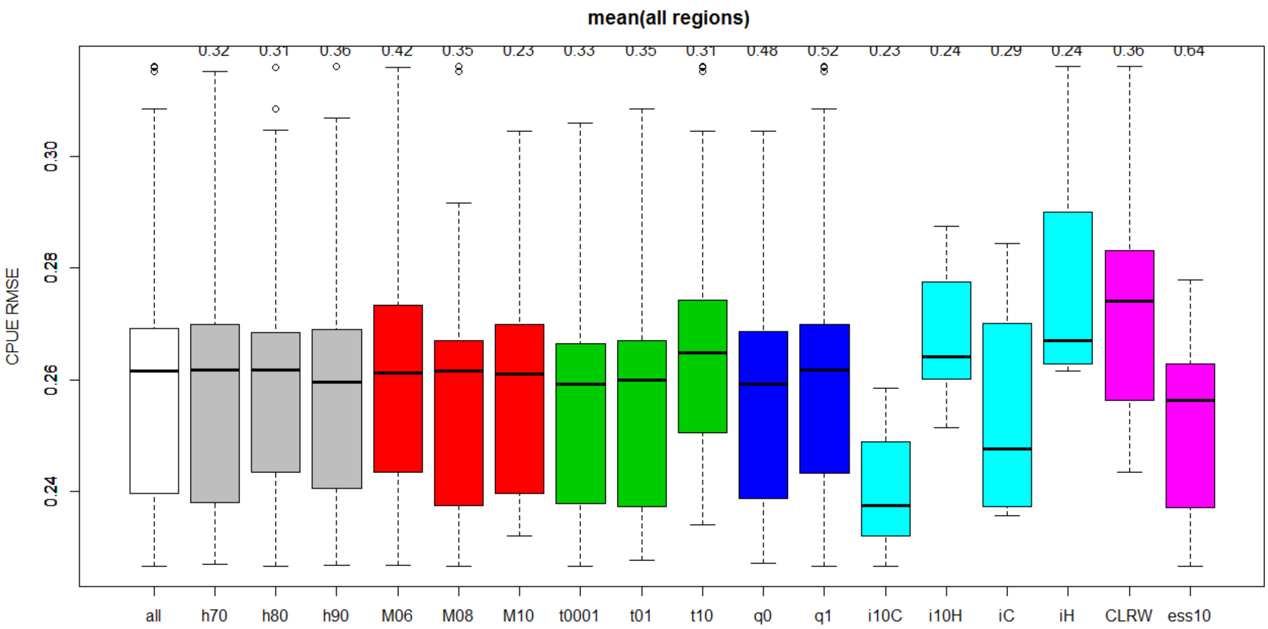


Figure 12. OMrefB18.5 Quality of fit (RMSE) for the CPUE series in region 4 by quarter (top) and year (bottom).



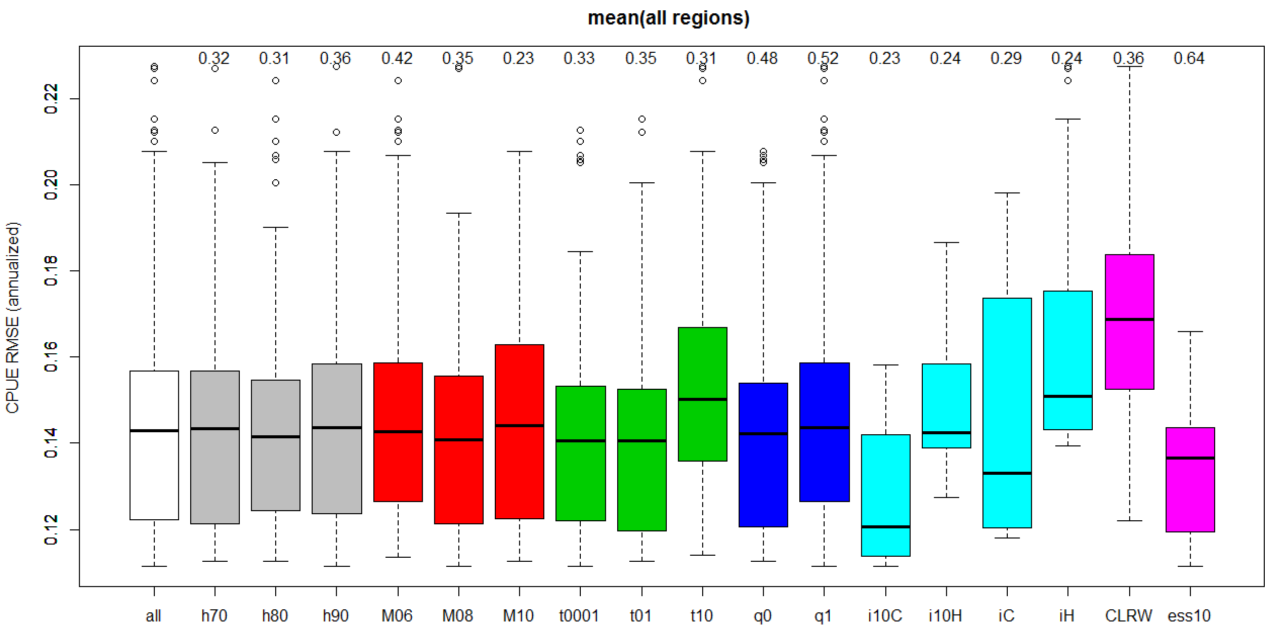


Figure 13. OMrefB18.5 Quality of fit (RMSE) for the mean of all CPUE series by quarter (top) and year (bottom).

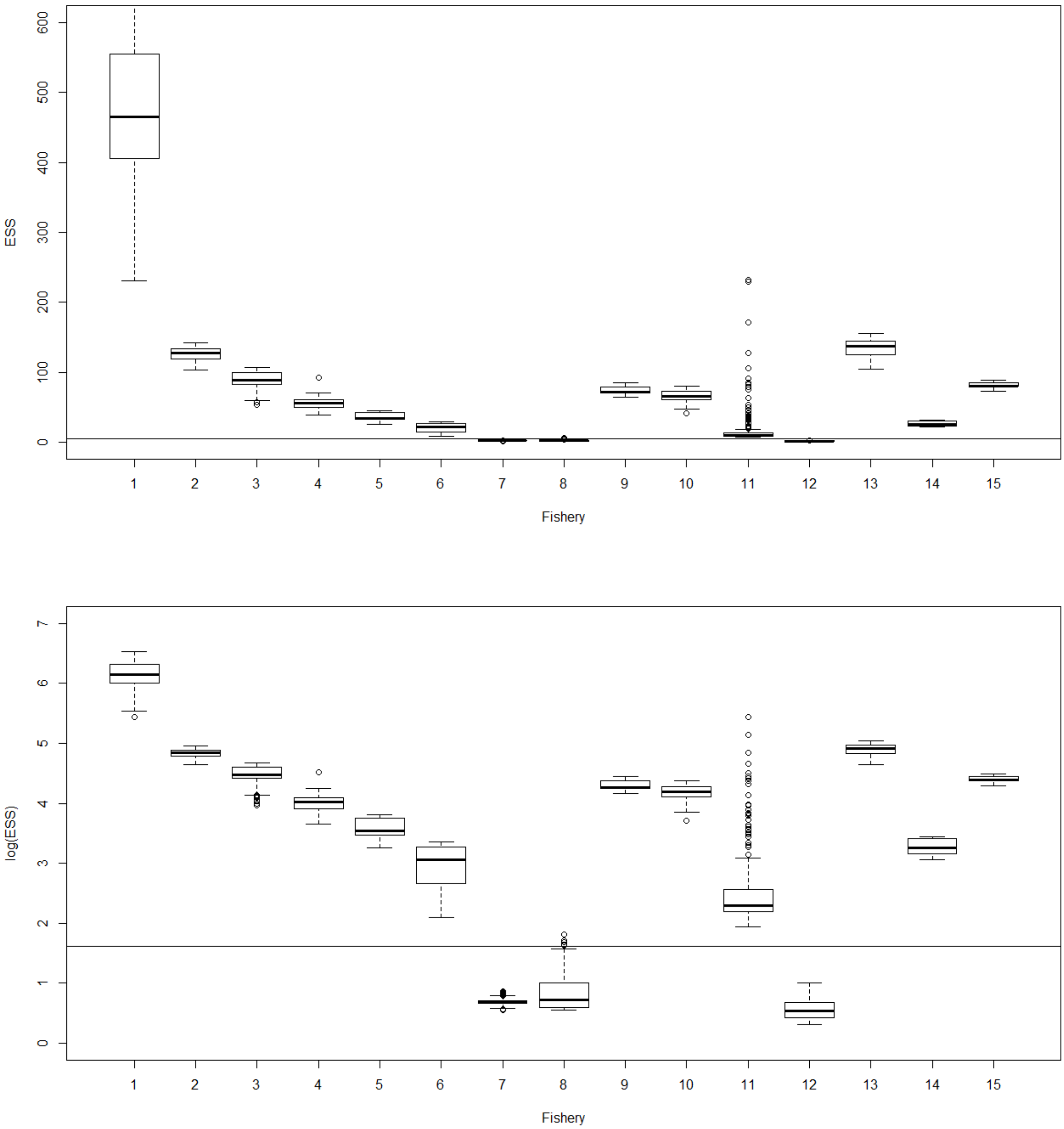


Figure 14. OMrefB18.5 quality of fit (post-fit Effective Sample Size) for the size composition data by fishery (all models combined)

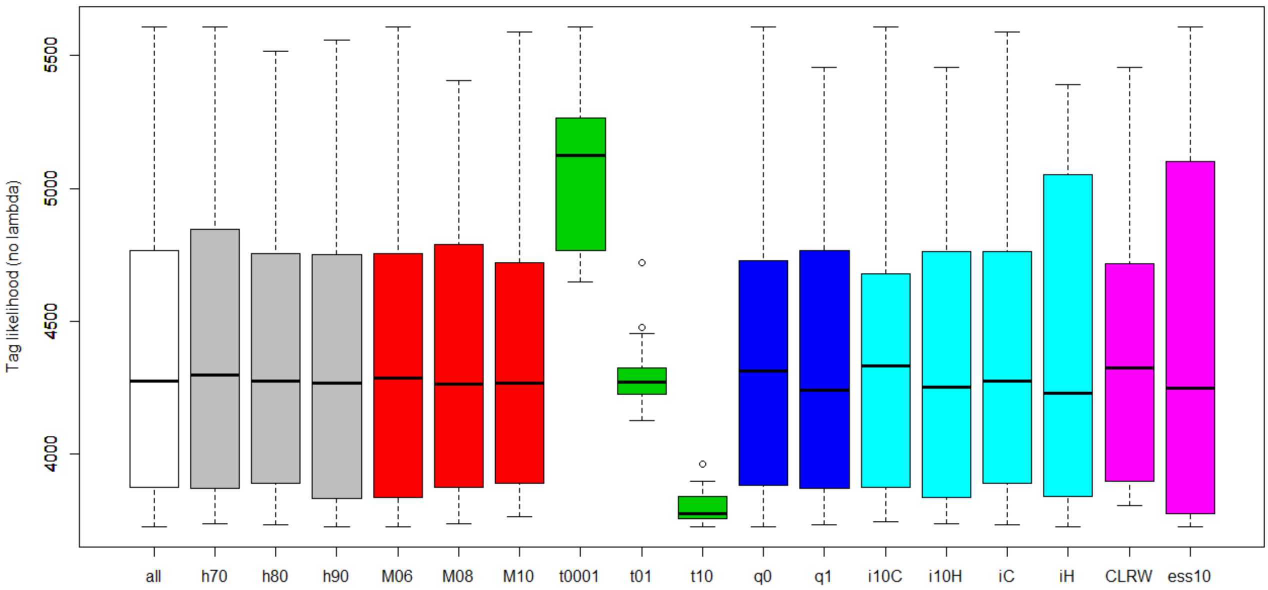


Figure 15. OMref18.5 Tag likelihood summaries marginalized over assumption levels.

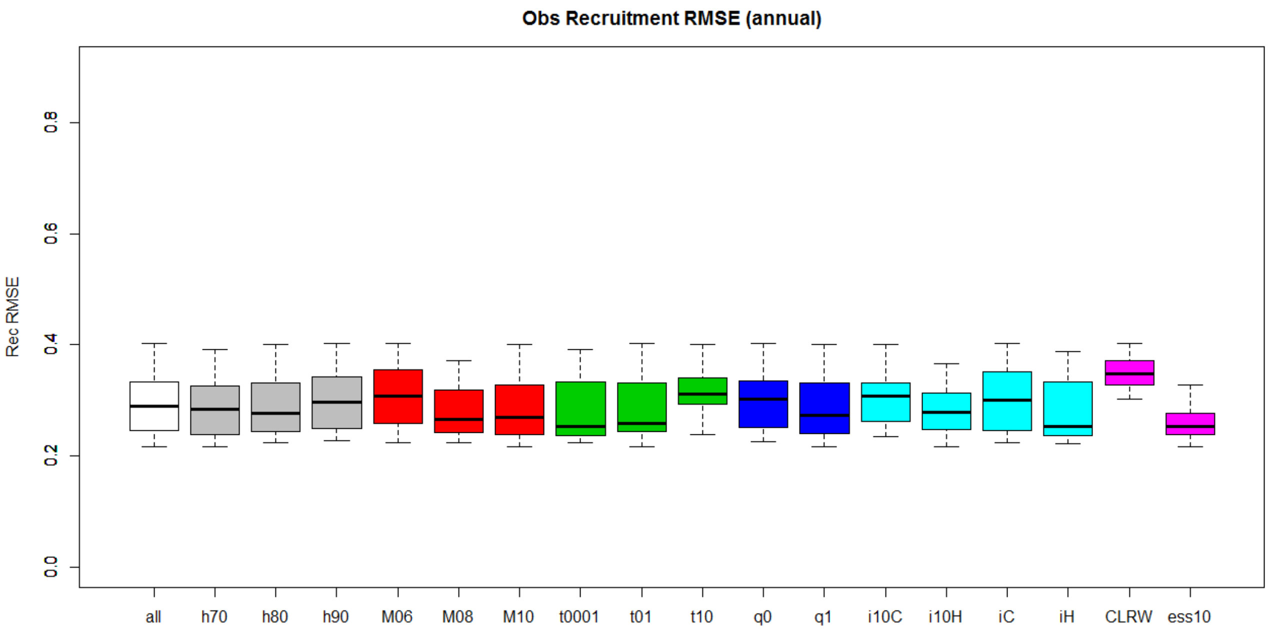


Figure 16. OMrefB18.5 recruitment RMSE (deviations aggregated across regions and seasons).

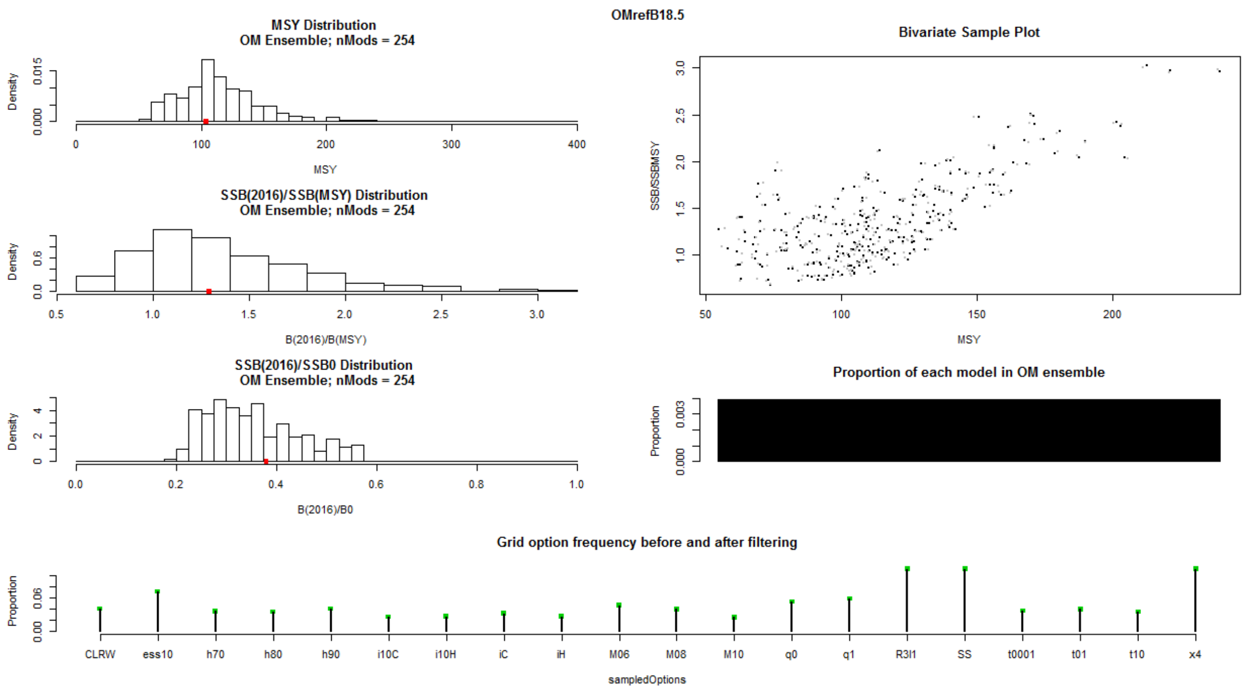


Figure 17. Summary characteristics of the proposed BET Operating Model OMrefB18.5.

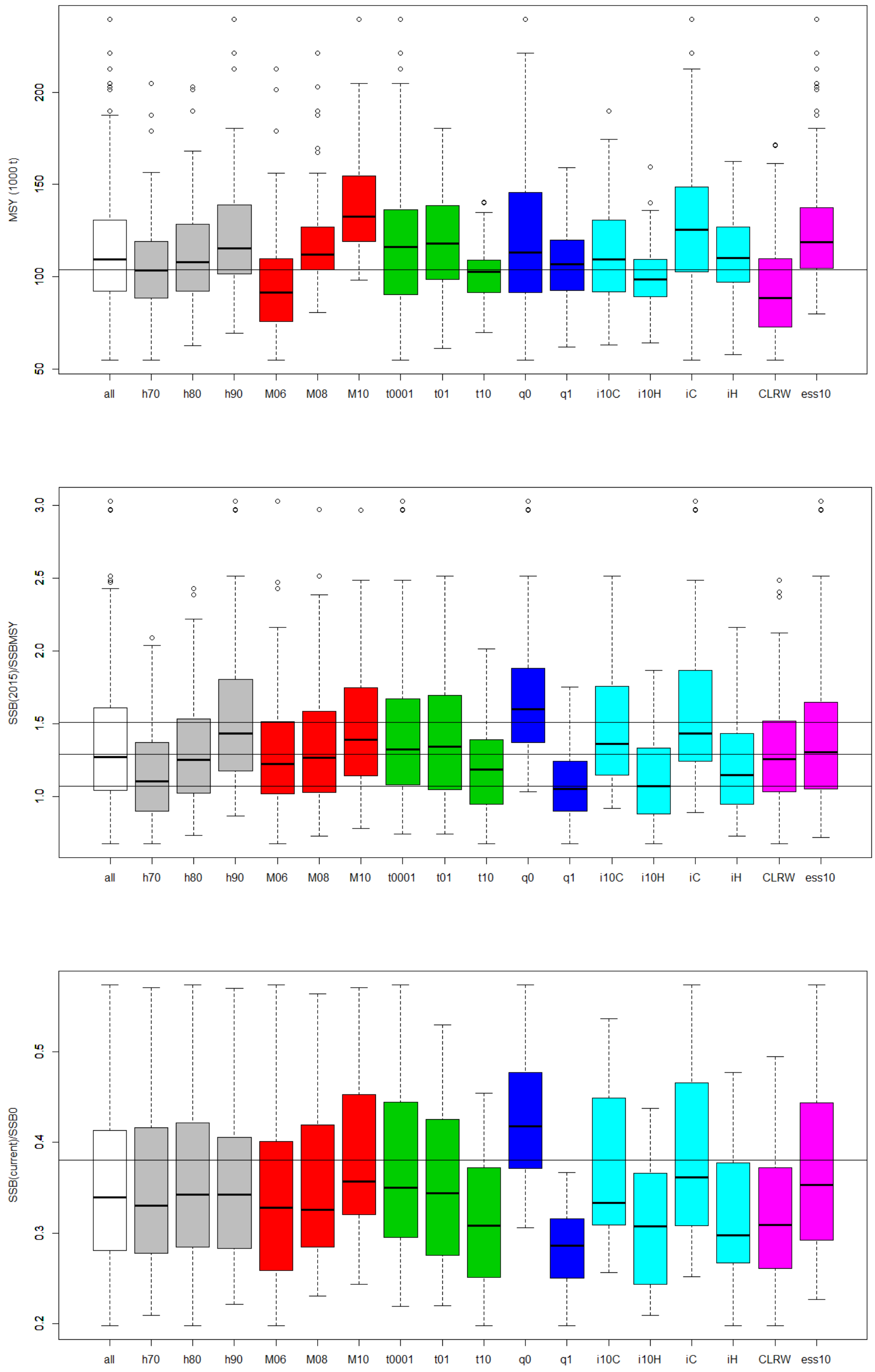


Figure 18. OMrefB18.5 key stock status inferences marginalized over stock status assumptions

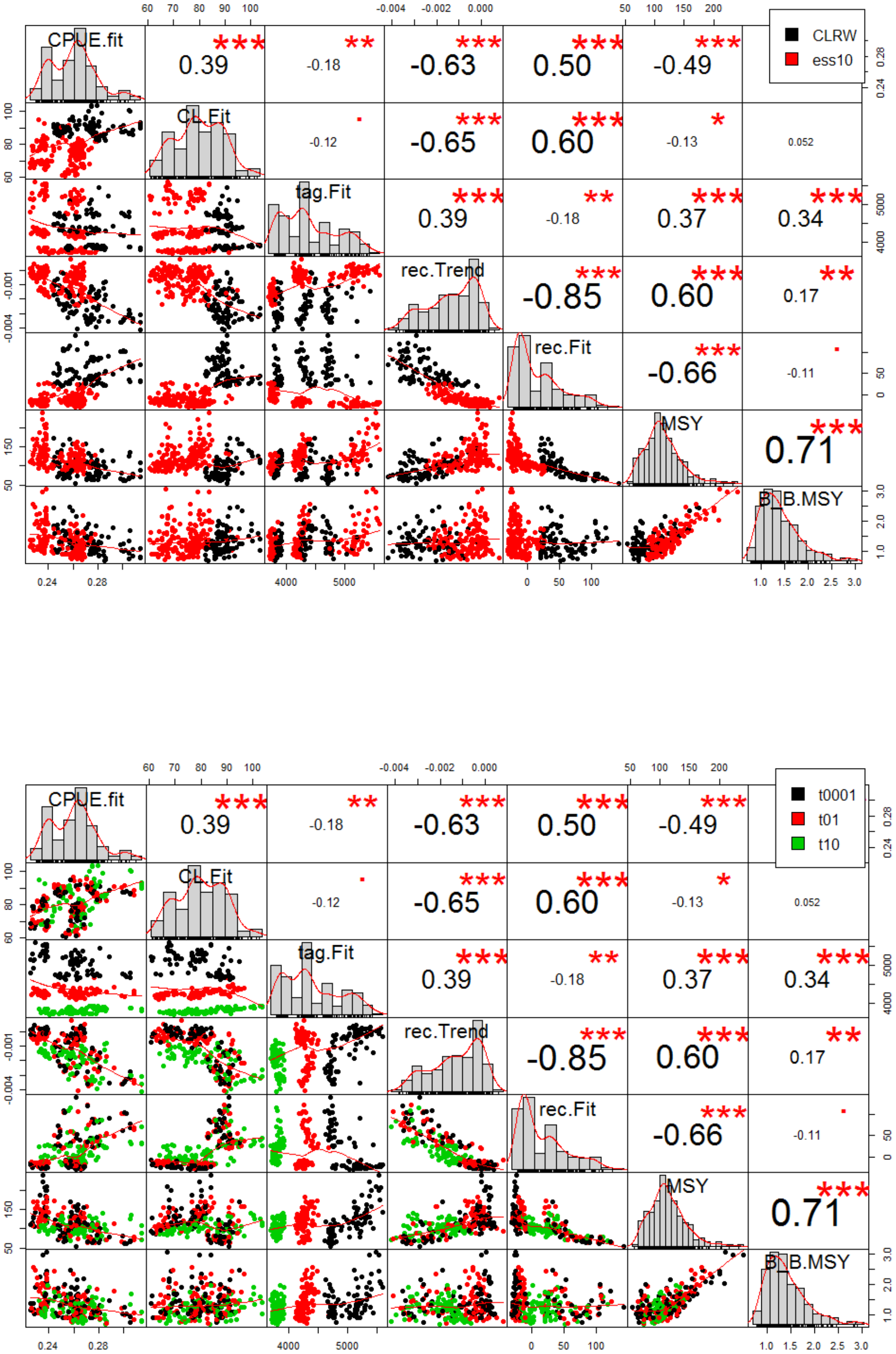


Figure 19. Operating Model OMrefB18.5 relationships among various quality of fit and stock status summary indices, partitioned by catch-at-length weighting assumption (top) and tag weighting assumption (bottom).

3.6 Numerical considerations of high fishing mortality

Figure 20 shows that there is very little difference among MP evaluation results for 3 different high F assumptions, when tested against the most aggressive of the recently defined BET tuning objectives (Pr(Green Kobe 2030:2034) = 0.5):

- ".cpp" - the default C++ sub-routine - assumes that Baranov F fishing mortality over 20 is possible, essentially driving any vulnerable component of the population to zero.
- ".C80" - the maximum F for the most highly selected age group of each fishery is constrained to 1.61 (again using the C++ sub-routine). This would be an 80% depletion (in an individual time-step) if a single fishery was operating (the depletion may be much higher given that there are multiple fisheries).
- ".R" - the high F constraint for the Pope's approximation in the original R sub-routine is more complicated (described in the user manual) and deviates systematically from the Baranov solution as F increases.

There is an additional difference in the two implementations in that the R sub-routine attempts to extract exactly one quarter of the annual quota independently in each quarter. Failure to extract the partial quota in the first quarter is not compensated for by extracting more in subsequent quarters. In contrast, the C++ sub-routine solves for the total quota removal across four seasons simultaneously. The C++ option is preferable in the sense that a shortfall in one quarter can be made up by a surplus in other quarters from new growth and recruitment. However neither option is likely a realistic reflection of how the fisheries would react to extreme depletion (when vessels would likely quit fishing, change targeting or move).

At this time, we are using the default C++ sub-routine for all MSE projections. In addition to being theoretically more attractive, it is also faster by a factor of around 2 (the overall MSE framework speed is still constrained by R code and the interface with C++).

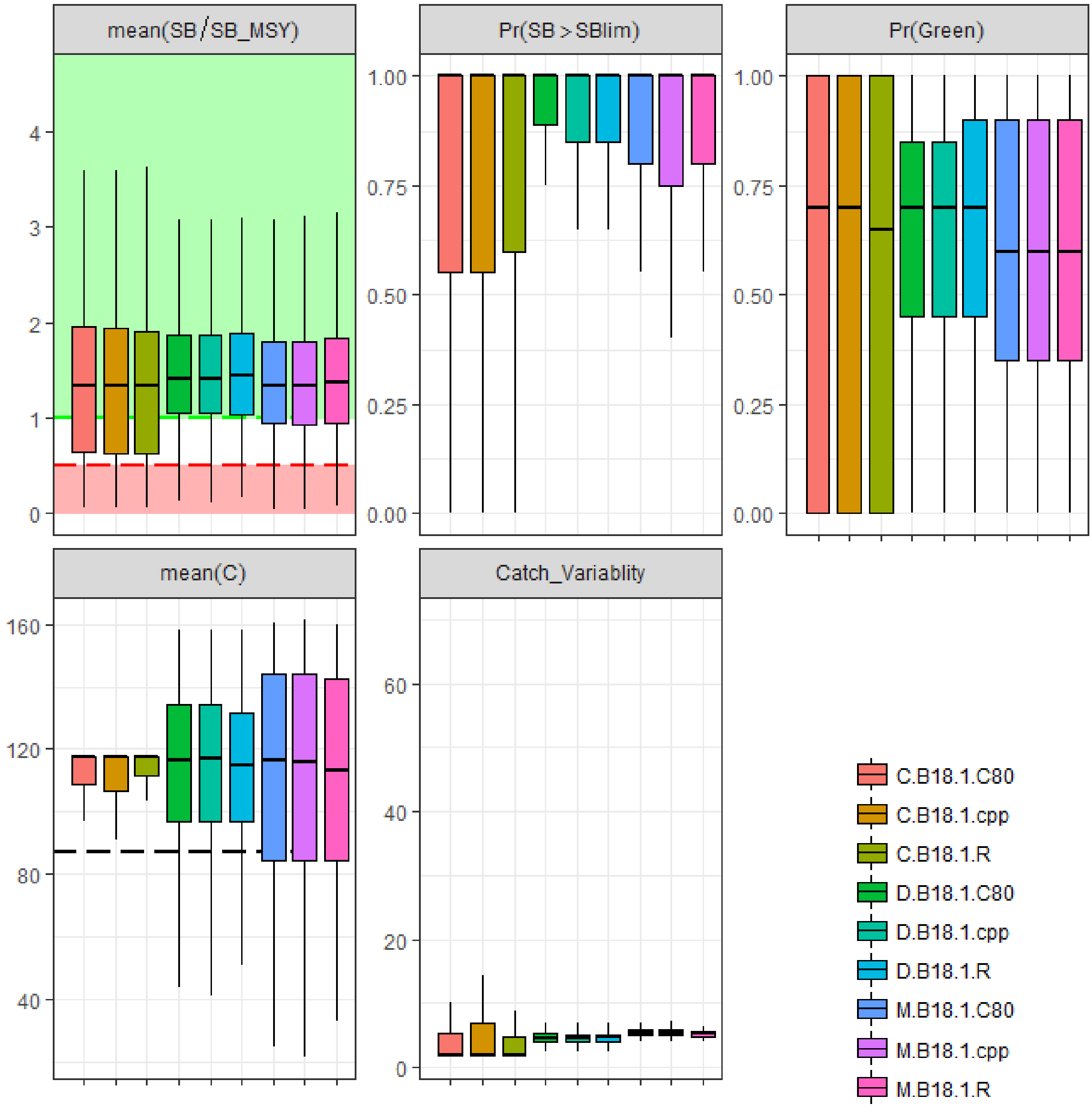


Figure 20. Comparison of 3 MPs for the most aggressive tuning criteria defined by the TCMP in 2018 (TB18.1 - see Kolody and Jumppanen 2018a for details), each assuming 3 different approaches for the numerical constraints on the high F scenarios.

3.7 Revisiting BET OM projection variance assumptions

Key projection assumptions for the proposed reference case OM, OMref18.5, included:

- Most parameters are defined by the SS estimates and conditioning assumptions
- Initial numbers-at-age states have additional error added (with variance declining with age)
- annual CPUE CV = 0.2, auto-correlation = 0.5
- quarterly recruitment CV = 0.6, autocorrelation = 0.5 (corresponds to annual CV = 0.42, annual auto-correlation = 0.21)
- first TAC implemented in 2019; bridging catches 2016:2018 = 87Kt (2016 level)
- catch implementation error CV = 0
- stationary selectivity (at the terminal estimated values)

The CPUE and recruitment variance assumptions are further considered below.

CPUE

The fit to the CPUE data as summarized in Figure 9 - Figure 13 suggest that the quality of fit in the conditioned models is very good. It is better than we would have reason to expect for the purposes of implementing an MP based on commercial CPUE, and better than we have assumed in the projections to date (annual, spatially-aggregated CV = 0.2). The issue of simulating a realistic CV going forward requires further consideration with respect to capturing the variability among years and areas, including auto-correlation, and the recognition that there is a high frequency of missing observations that is likely to persist going forward (only 10 years of the whole assessment period had CPUE observations for all 4 regions and 4 seasons simultaneously).

Recruitment

The BET and YFT OMs were initially parameterised with independent quarterly recruitment (σ_R = 0.6, auto-correlation ρ = 0.5). The σ_R value was selected to be consistent with the assessment assumptions, while ρ was arbitrarily chosen to be "big enough to matter, but not overwhelming".

Concerns with the initial assumptions include:

- Sensitivity to the σ_R assumption in the assessment (and OM projections) has not been tested for bigeye (though has been for yellowfin - see Kolody and Jumppanen 2018f).
- Variability among the conditioned OMs was not considered.
- If the projection time series are summed over 4 season years, the OM projection assumptions corresponds to an annual σ_R = 0.42 and ρ = 0.22, which was not directly compared with the assessment inferences.
- The interaction among quarterly stochastic error, annual stochastic error and deterministic seasonal effects was never explicitly examined. i.e. The current assessment structure assumes, σ_R = 0.6, with independent quarterly deviations, but if most of this variability is

due to a consistent seasonal pattern, this represents a much simpler management problem than interannual variability.

To check the appropriateness of the adopted values in more detail, we calculated the quarterly and annual recruitment CV from the output recruitment deviations from the reference case BET assessment (tagLambda1), with and without a simple linear model estimating fixed seasonal effects, and the corresponding auto-correlation (Table 4). The quarterly and annual recruitment deviation series are shown in Figure 21. It appears that:

- Seasonal effects are statistically significant, but of small magnitude in the recruitment time series (i.e. most of the variability is attributed to stochastic noise). As a consequence, the statistical characteristics of the annual recruitment time series remain about the same irrespective of the seasonal effects. So seasonality can effectively be ignored.
- When aggregated at an annual level, the quarterly recruitment assumptions in the OM projections to date result in a higher CV and lower auto-correlation than the assessment outputs. (Through simulation trial and error, rather than clever mathematics), we find that annual σ_R = 0.3, ρ = 0.34 is achieved (approximately) with quarterly σ_R = 0.37 and ρ = 0.65.

The annual σ_R from the assessment (σ_R = 0.3) is near the median ($\sigma_R \approx 0.35$) of the ISSF (2011) meta-analysis of 14 tuna populations. The annual σ_R = 0.42 assumed in the OM projections is around the 79th percentile of the ISSF (2011) analysis, and similar to the highest level observed in OMrefB18.5 (Figure 16). Figure 22 shows the difference in simulated time series using the OM assumption and values inferred from the assessment. between We would suggest that there is probably no urgent need to change the current OM recruitment variance assumptions. The current values are well within the range of the ISSF meta-analysis, and if they are high relative to the assessment, this is consistent with the desire for robustness in the MSE process. These assumptions could be scenario-specific, but presumably there would still need to be a minimum lower bound.

Table 4. Comparison of quarterly and annual recruitment characteristics from the 2016 assessment (assuming quarterly deviations are independent), from the 2016 assessment with estimated seasonal effects, and from the OM projection assumptions.

Rec Dev series	Summary period	SD	auto-correlation
Assessment	quarter	0.670	0.137
Fixed Seasonal Effects	quarter	0.523	0.227
Original OM projection assumption	quarter	0.6	0.5
Assessment	annual	0.306	0.344
Fixed Seasonal Effects	annual	0.292	0.340
Original OM projection assumption	annual	0.43	0.21

Figure 21. Comparison of quarterly and annual recruitment deviations from the BET assessment (black) and the equivalent series with annual and fixed seasonal effects estimated (red).

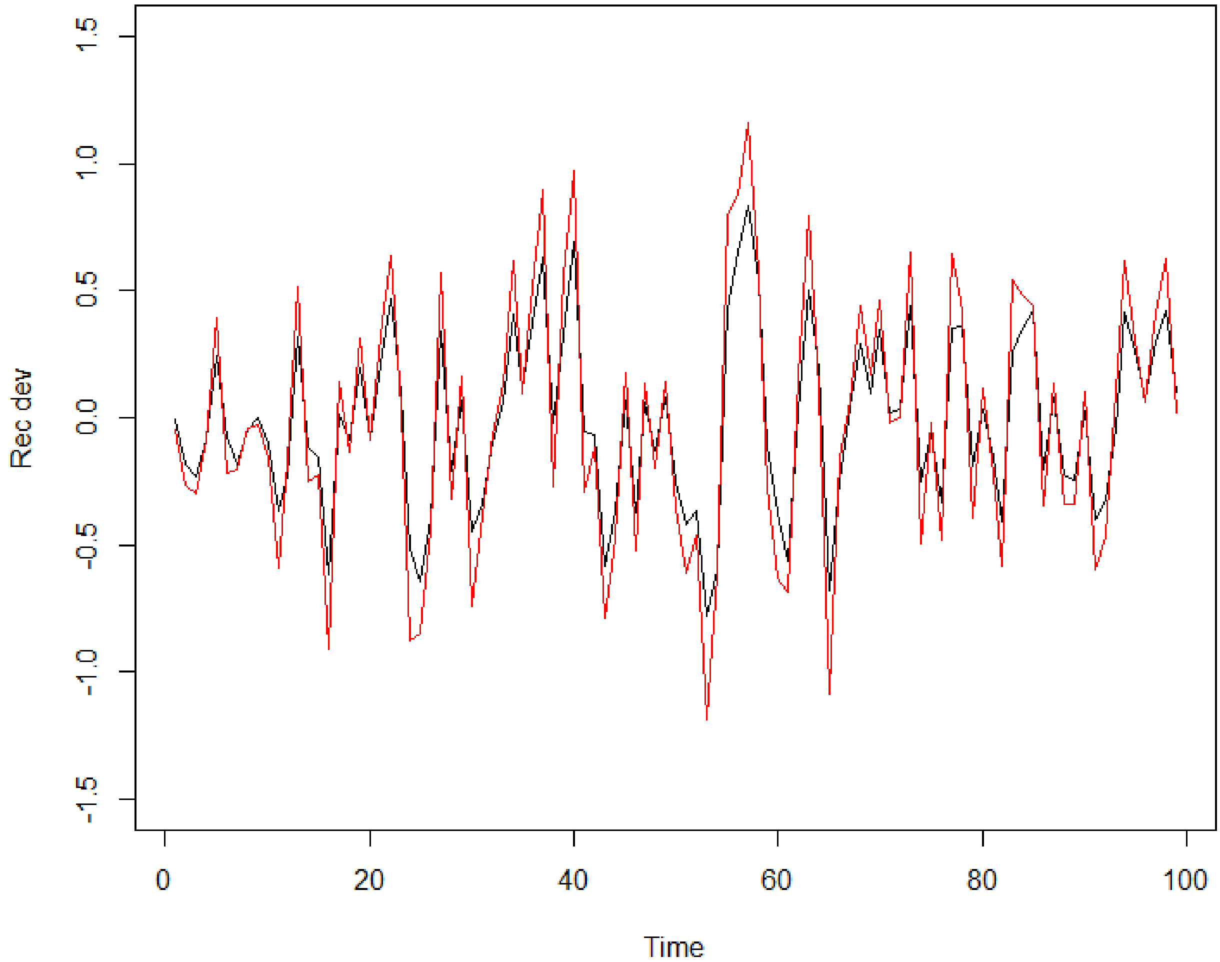


Figure 22. Simulated 100 year annual recruitment deviation time series, comparing the bigeye 2016 assessment variance and auto-correlation characteristics (black) with the current OM projection assumptions (red).

3.8 Reference set OMref18.5 projection dynamics

When projected forward with simple constant catch management, the basic projection dynamics of OMrefB18.5 appear to be consistent with general perceptions of current stock status (Figure 23). In the absence of fishing, SSB rapidly rebuilds to 2030, and continues to increase slowly beyond that. Current catches are estimated to be sustainable for >75% of scenarios, with modest SSB increases predicted over the next several years.

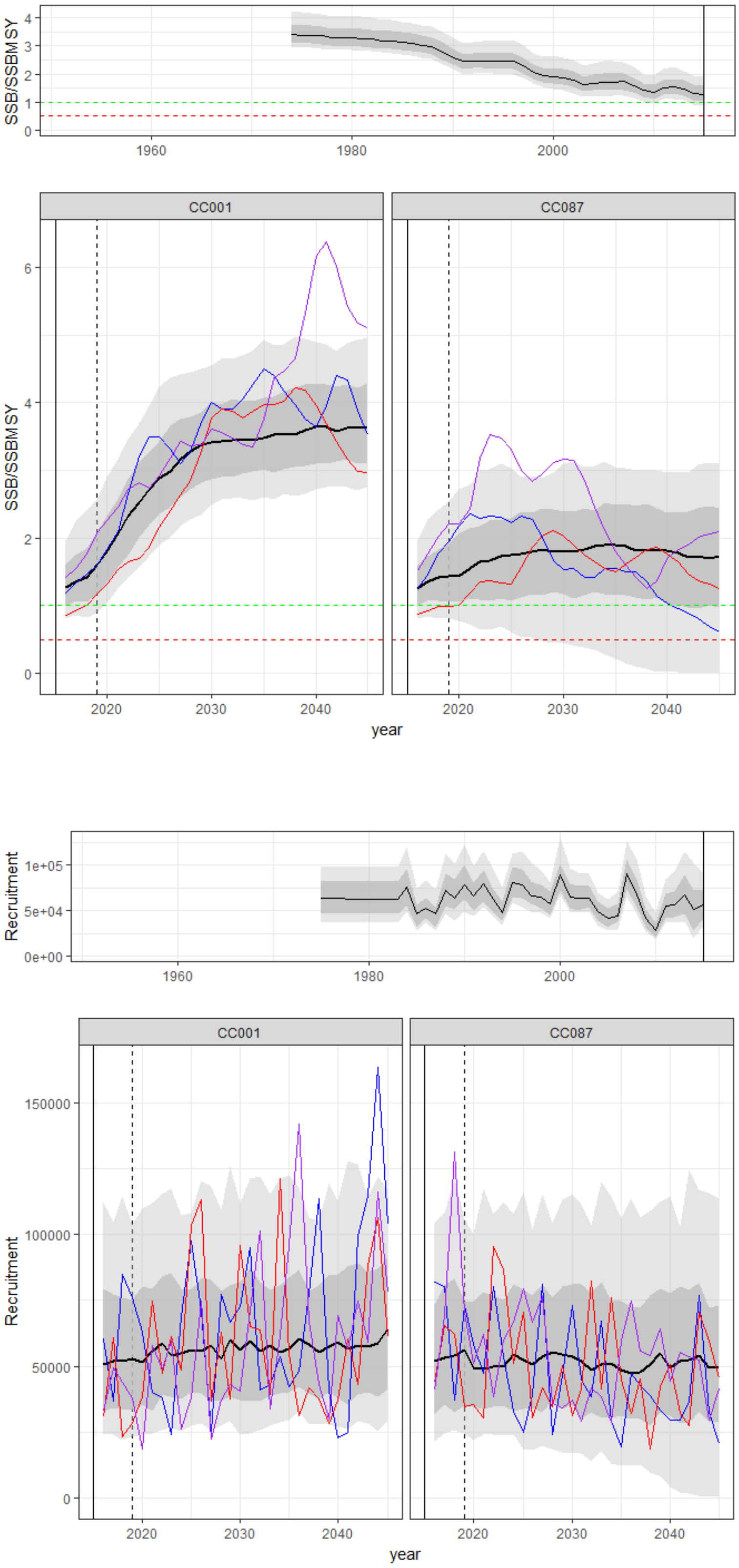


Figure 23. OMrefB18.5 spawning biomass (top) and recruitment (bottom) dynamics assuming zero future fishing (left) or constant current catch at 87 Kt (right).

4 BET Robustness scenarios

There are various robustness scenarios that could be defined for BET, and a proper discussion has not yet been undertaken (this has not been a high priority to date, but many of the issues would be similar to YFT). Five options were explored:

- OMrobB18.5.recShock - Given that YFT is estimated to have had a period of poor recruitment in the early 2000s, how would BET MP performance be affected if there were 8 consecutive quarters of poor recruitment (55% of expected values)? (Figure 24)
- Are the MPs robust to implementation error? Catch data are reported accurately in all 3 cases.
    - *OMrobB18.5.impErrCV10* - Each fishery has a 40% catch implementation error CV (independent by year and fishery). This corresponds to an annual aggregate CV >10%.
    - *OMrobB18.5.under* - Given that the BET fishery is not currently under management constraints, and is not over-fishing the stock, what will happen if high quota recommendations from an MP are ignored for 10 years (fishing mortality constant at current levels) before the MP becomes restrictive?
    - *OMrobB18.5.over* - What happens if there is a consistent 10% overcatch?
- *OMrobB18.5.qTrend3* - What happens if the longline CPUE catchability trend is 3% per year going forward (but remains as in the reference scenario for conditioning)?

The consequences for MP performance are presented in the companion paper Kolody and Jumppanen (2018a). These OMs are trivial to define and test because they involve using the reference case OM with changes to the projection specifications only. Other robustness tests that require modification to the code and/or reconditioning should be considered more carefully.

We provide some discussion points for future consideration of BET robustness scenarios below.

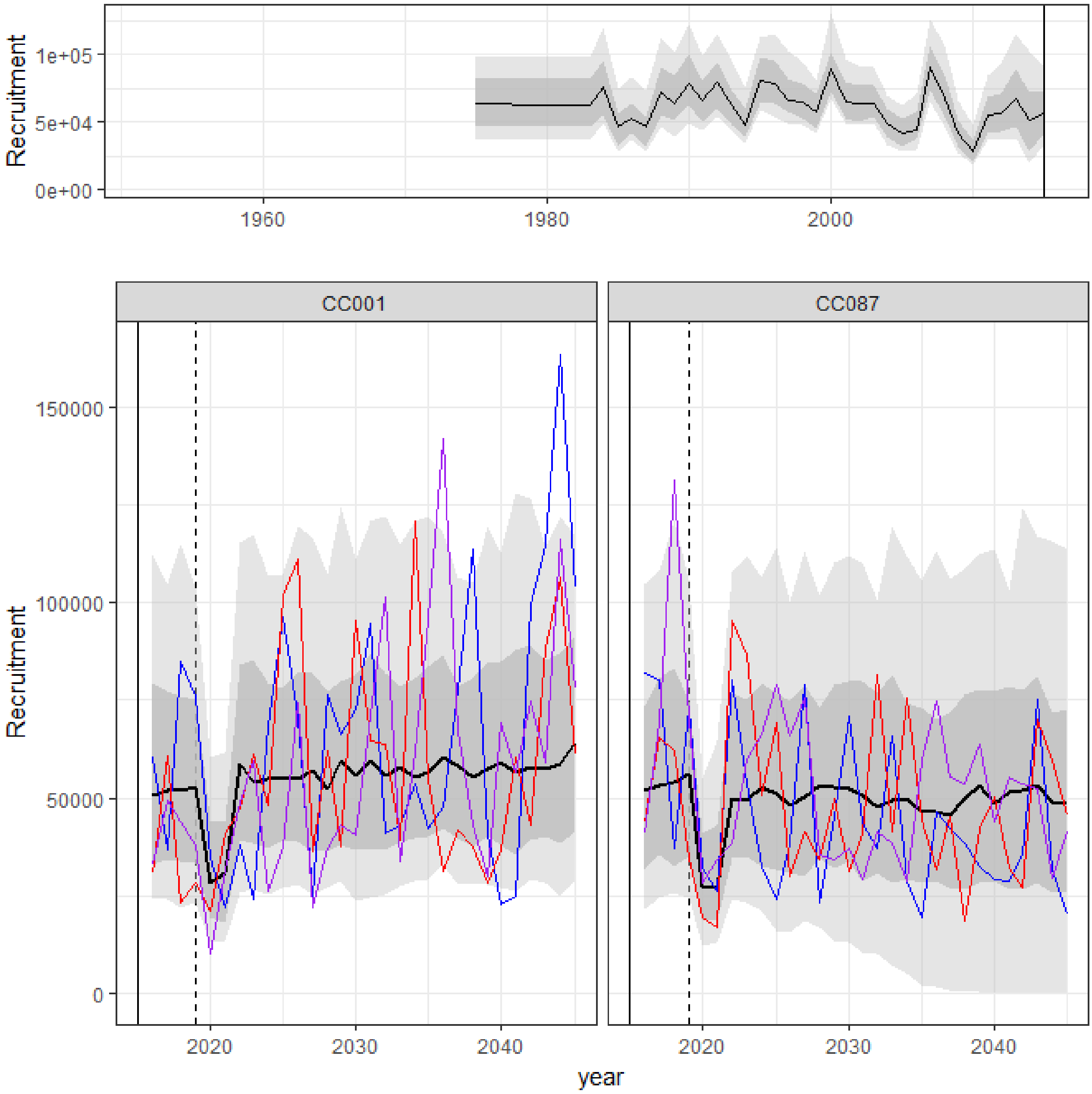


Figure 24. Bigeye recruitment time series for the robustness scenario OMrobB18.5.recShock.

5 Discussion

In most fisheries MSEs there are lingering concerns about the sufficiency of OM conditioning assumptions, process for plausibility screening, and weighting models, etc . However, at this time, we are not aware of obvious reasons for not proceeding with OMrefB18.5. We consider the following priority points for feedback/endorsement for the phase 2 BET MSE to move forward:

BET reference case OM

- Is there any objection to replacing 4 seasonal temperate CPUE series with a single aggregate?
    - The stock status inferences were almost identical, but the OM implementation with 4 series would require further development work. For the current MSE purposes, it is not clear that effort spent describing uncertainties in seasonal dynamics will be helpful.
- Are there any dimensions in the reference case conditioning grid that should be added or removed?
- Are there any alternative recommendations for dealing with potentially influential parameter bounds and priors?
- Are there additional model diagnostics that should be examined, presented and/or applied for defining plausibility of the grid? Seasonal and spatial issues have not been considered in much detail to date.
- Should uniform model-weighting be retained?
    - Alternatives include likelihood-based weighting (only possible within subsets of models that use the same data series), expert opinion weighting, or sampling in relation to the "collective wisdom of the assessment" as proposed for YFT (Kolody and Jumppanen 2018f).
    - There is nothing to indicate that the process of reducing the model grid down to 252 models has substantially biased the stock status inferences from the original 432 model grid.
- Should there be further refinement of the projection assumptions,
    - e.g. CPUE and recruitment variability could be linked directly to individual assessment specification outputs. If this is deemed necessary, it would be prudent to retain minimum levels, to ensure that the OM scenarios are not unrealistically easy to manage.

BET robustness tests:

What are the priorities for the BET robustness scenarios, and should they be presented to the TCMP in 2019?

- The robustness tests examined to date are easy to implement and test with simple modifications to the reference set OM.
- Additional robustness scenarios that require modifications to the conditioning and/or projection code, should be considered and specified carefully. i.e. Do they represent genuine concerns coming from the stock assessment process? Can they be meaningfully

quantified? Do they need to be tested as a full dimension within the reference case grid, or can they be defined by a representative subset of dimensions?

- In the interest of clear communication, its worth considering which robustness tests should be presented to the TCMP. Unless the tests identify a specific plausible concern, or they offer additional information that will be useful in helping to select among MPs, it may not be worth presenting them.

References

IOTC 2017. Report of the 21st Session of the Indian Ocean Tuna Commission Yogyakarta, Indonesia, 22–26 May 2017. IOTC–2017–S21–R[E].

ISSF 2011. Report of the 2011 ISSF Stock Assessment Workshop Rome, Italy, March14-17, 2011. ISSF Technical Report 2011-02. International Seafood Sustainability Foundation, Washington, D.C., USA.

Jumppanen, P, Kolody, D. 2018. User manual for IOTC Yellowfin and Bigeye Tuna MSE Software. https://github.com/pjumppanen/niMSE-IO-BET-YFT/.

Kolody, D, Jumppanen, P. 2018a. IOTC Bigeye and Yellowfin Management Procedure Evaluation Progress October 2018. IOTC–2018–WPM09–11.

Kolody, D., Jumppanen, P. 2018b. IOTC Bigeye Tuna Management Procedure Evaluation Update May 2018. IOTC-2018-TCMP02-info-10.

Kolody, D., Jumppanen, P. 2018c. IOTC Yellowfin Tuna Management Procedure Evaluation Update May 2018. IOTC-2018-TCMP02-info-11.

Kolody, D, Jumppanen, P. 2018d. Update on IOTC Bigeye Tuna Management Procedure Evaluation March 2018. IOTC–2018–WPTT20–INFO1.

Kolody, D, Jumppanen, P. 2018e. Update on IOTC Yellowfin Tuna Management Procedure Evaluation March 2018. IOTC–2018–WPTT20–INFO2.

Kolody, D, Jumppanen, P. 2018f. Update on IOTC Yellowfin Tuna MSE Operating Model Development October 2018. IOTC–2018–WPM09–10.

Langley, A. 2016. Stock assessment of bigeye tuna in the Indian Ocean for 2016 — model development and evaluation. IOTC-2016-WPTT18-20.

TCMP 2018. Report of the 2nd IOTC Technical Committee on Management Procedures. Bangkok, Thailand, 18–19 May 2018. IOTC–2018–TCMP01–R[E].

CONTACT US

t 1300 363 400

+61 3 9545 2176

e csiroenquiries@csiro.au

w www.csiro.au

AT CSIRO, WE DO THE EXTRAORDINARY EVERY DAY

We innovate for tomorrow and help improve today – for our customers, all Australians and the world.

Our innovations contribute billions of dollars to the Australian economy every year. As the largest patent holder in the nation, our vast wealth of intellectual property has led to more than 150 spin-off companies.

With more than 5,000 experts and a burning desire to get things done, we are Australia's catalyst for innovation.

CSIRO. WE IMAGINE. WE COLLABORATE. WE INNOVATE.

FOR FURTHER INFORMATION

Oceans & Atmosphere

Dale Kolody
t +61 6 6232 5121

e Dale.Kolody@csiro.au

w www.csiro.au