Information paper for the IOTC Working Party on Methods 2019: **IOTC–2019–WPM10-INF03**

Prepared for the IOTC MSE Task Force (March 2019, Ispra, Italy)

Indian Ocean Yellowfin Tuna

MSE Update March 2019

Dale Kolody (email: dale.kolody@csiro.au)

Paavo Jumppanen

CSIRO, Australia

Summary

This working paper briefly describes key developments on the IOTC yellowfin operating model since the 2018 WPTT and WPM. Key points include:

- The yellowfin OM was updated in relation to the 2018 stock assessment. This was not a specific request from the WPTT/WPM, but was undertaken in recognition that a number of potentially important assumptions had changed (including data revisions), and that this might prove useful in the context of the 2019 yellowfin assessment review process.
- A fractional factorial design (OMgridY19.1) was used to set up a 144 model grid that can, in principle, quantify the main effects of 11 factors (two 3-level and nine 2-level factors). The full factorial grid would have required 4608 models. We attempted to repeat the convergence 3 times per configuration, however this could not be achieved for 13 cases (>20 failures each). Some (possibly all) convergence failures appeared to be associated with biomass too low to allow the observed catches to be removed from the population. Many of the results were most sensitive to the CL weighting assumptions, often resulting in bimodal distributions.
- A second fractional factorial grid (OMgridY19.2) was run with an intermediate CL weighting assumption (the assessment post-fit ESS raised to the power of 0.75, and capped at 100). 121 of 144 models converged (twice), and forms the basis of the results presented here.
- The alternative growth curve of Dortel et al. (2015) was added as a new reference case uncertainty dimension, as parameterized in the Fu et al (2018) assessment. This was not considered as a robustness test, because we could not identify any strong reason to expect it to be less reliable than the main assessment assumption (one possible criticism arises from the process used for filtering the tag data, which was not clear, and could bias the estimates)
- Tag mixing time assumptions were changed from 3 quarters to 4 and 8 quarters. Note that these mixing rates are only relevant to 25% of scenarios each, as the other 50% of scenarios have a negligible tag weighting.
- An additional regional scaling factor was admitted to the reference case OM (the most extreme with respect to the biomass ratio in the NW region to the other regions, from among the two alternatives considered as worthy of consideration in the recent assessment).
- Dome-shaped (double normal) selectivity was admitted as an alternate assumption to the usual logistical longline selectivity in gridB19.2, however it was omitted from projections pending further investigation.
- The new set of YFT conditioning results is better behaved than previous iterations, with key inferences generally showing a unimodal distribution with few outliers, and a central tendency slightly more pessimistic than the recent assessment. Post hoc model sampling on the basis of stock status characteristics (as applied in the two previous OM development iterations) appears unnecessary this time. We do not know the extent to which this new stability is attributable to: i) the revised data and assumptions in the 2018 assessment (possibly including the phasing issue in which it appears that the minimizer can be trapped fitting to seasonal CPUE patterns rather than annual trends), ii) the requirement for repeated convergence and/or iii) confounding of interactions in the fractional factorial design.
- MP evaluations are presented as a subset of gridOMref19.2, in which only the logistic selectivity was retained. While this modified grid is balanced (i.e. all factor levels

represented an equal number of times), we are not certain if there are unintended consequences in the fractional design.

- Example MP results are presented for the three tuning objectives defined in the 2018 TCMP. The results suggest that the rebuilding objectives (rebuild to BMSY by 2024, 2029 and 2034) are attainable, but only if TAC change constraints are relaxed from 15%, and catches in the short-medium term are considerable lower than recent catches.
- A subset of the requested robustness tests was completed, showing qualitatively unsurprising results (e.g. adverse model assumptions increase the risk of not meeting management objectives).

Summary points are presented for discussion and/or endorsement from the IOTC MSE task force with respect to requirements for i) the 2019 TCMP (primarily reference case requirements) and the 2019 WPTT/WPM (further development of robustness scenarios and candidate MPs). Decisions for the next iteration of the YFT MSE are documented in the 2019 MSE Task Force report.

Introduction

This paper represents a progress update on key technical elements of the IOTC yellowfin MSE project to obtain feedback in preparation for the 2019 IOTC TCMP, WPM and WPTT. The intended audience is already familiar with the scope of the work and technical jargon. Other interested parties should consult the more accessible project reports found in https://github.com/pjumppanen/niMSE-IO-BET-YFT/.

The decision was taken to update the yellowfin OM in relation to the newest assessment (Fu et al 2018). This was not a specific request from the WPTT/WPM, but was undertaken in recognition that a number of potentially important assumptions had changed since the last assessment (including data revisions), such that this exercise should prove useful both for improving confidence in the OM, and potentially informing the proposed 2019 yellowfin assessment review process.

OM ensembles examined in this iteration

Table 1 lists the OM configurations and rationale. Grid factor details are provided in Table 2, with elaboration of new options provided in the text below (other options are explained in earlier documents).

Table 1. Operating Model definitions (the numerical sequence does not necessarily follow a temporal progression).

OM Ensemble	**Rationale (factor abbreviations are defined in Table 2)**
OMgridY19.1	(not reported) This test grid identified that the OM ensemble was very sensitive to the 2 levels of CL weighting, resulting in very

	bimodal OM characteristics. This prompted inclusion of 3 levels for OMgridY19.2.
OMgridY19.2	144 models with 11 factors in a fractional factorial design that quantifies main effects only (2 way interactions are confounded) h70, h80, h90 M10, M08, M06 ess10, CLRW, CL75 t0001, t10 q0, q1 iH, iC i1, i3 gr1, gr2 iR1, iR2 SL, SD x4, x8
OMrobY19.2.CP5	OMgridY19.2 excluding models with catch penalty >0.00001 (66 models retained)
OMrefY19.2SL	The reference case OM for discussion purposes at the 2019 MSE Task Force meeting - OMgridY19.2 excluding models with dome-shaped selectivity and convergence failures (69 models retained)
OMrobY19.2opt	OMgridY19.2 excluding models with B/BMSY<1 - the intent was to test whether the MPs are likely to provide reasonable behaviour even if the current OM (and assessment) is pessimistically biased.

OMrobY19.2SL.over	A robustness scenario with consistent 10% overcatch for all fleets (catch is accurately reported) (conditioning is unchanged from OMrefY19.2)
OMrobY19.2SL.iuu10	A robustness scenario with consistent 10% *unreported* overcatch for all fleets (conditioning is unchanged from OMrefY19.2)
OMrobY19.2.qTrend3	A robustness scenario with a longline CPUE catchability trend of 3% per year in projections (conditioning is unchanged from OMrefY19.2)
OMrobY19.2.ICV3	A robustness scenario with a longline CPUE CV of 0.3 (aggregate annualized) auto-correlation = 0.5. (conditioning is unchanged from OMrefY19.2)
OMrobXXX.recShock (not presented this iteration)	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 OMrefB19.4)
OMrobXXX.impErrCV10 (not presented this iteration)	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 OMrefB19.4)
OMrobXXX.under (not presented this iteration)	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 OMrefB19.4)

OMrobBXXX.ICVxxx (not presented this iteration)	An exploratory scenario with a very small longline CPUE CV (aggregate annual =) to explore what might be achievable with a good abundance index

Table 2. Model specification abbreviations. Bold indicates the BET assessment assumption(s). Some abbreviations may relate to additional explorations that were either not completed, reported in earlier iterations, or pertain to BET.

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
	CPUE area-weighting factors
iR1	preferred assumption used in the assessment
	"pr_7994_m8" from Hoyle (2018)
iR2	An alternate ("pr_7594_m8" from Hoyle, 2018), among the 3 candidates proposed for the 2018 YFT assessment, in which Region 2 represents the lowest proportion of the total biomass
	mean Age-length relationship (growth curve)
gr1	Dortel et al. (2015)
gr2	**original from assessment (Fonteneau 2012)**
	Quarterly recruitment deviation penalty

sr4	σ_R = 0.4
sr6	**σ_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 estimates in the 2000s from several recent YFT assessments
	Natural mortality multiplier applied to the age-specific values in the reference case stock assessment.
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)
	(temperate series use iC)
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/SL	**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 recent selectivity deviations estimated
SD	Double normal 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 (sample-size specific, capped at 100, mean of ~53)
CL75	

ESS = One iteration of re-weighting; the output ESS from a reference case assessment specification raised to the power of 0.75, capped at 100, mean of ~ 22)

New assumptions in the March 2019 iteration

Catch-at-Length input Effective Sample Sizes

The assessment assumed uniform ESS = 5 (option ess5). One iteration of observation-specific reweighting was introduced in previous versions of the OM to add contrast (option CLRW = assessment post-fit ESS, capped at 100, mean across all samples = 53). Given the strong influence of the contrasting assumptions a third intermediate option was added this time (option CL75 = (assessment post-fit ESS)^0.75, capped at 100, mean across all samples = 22). The three are contrasted in Figure 1.

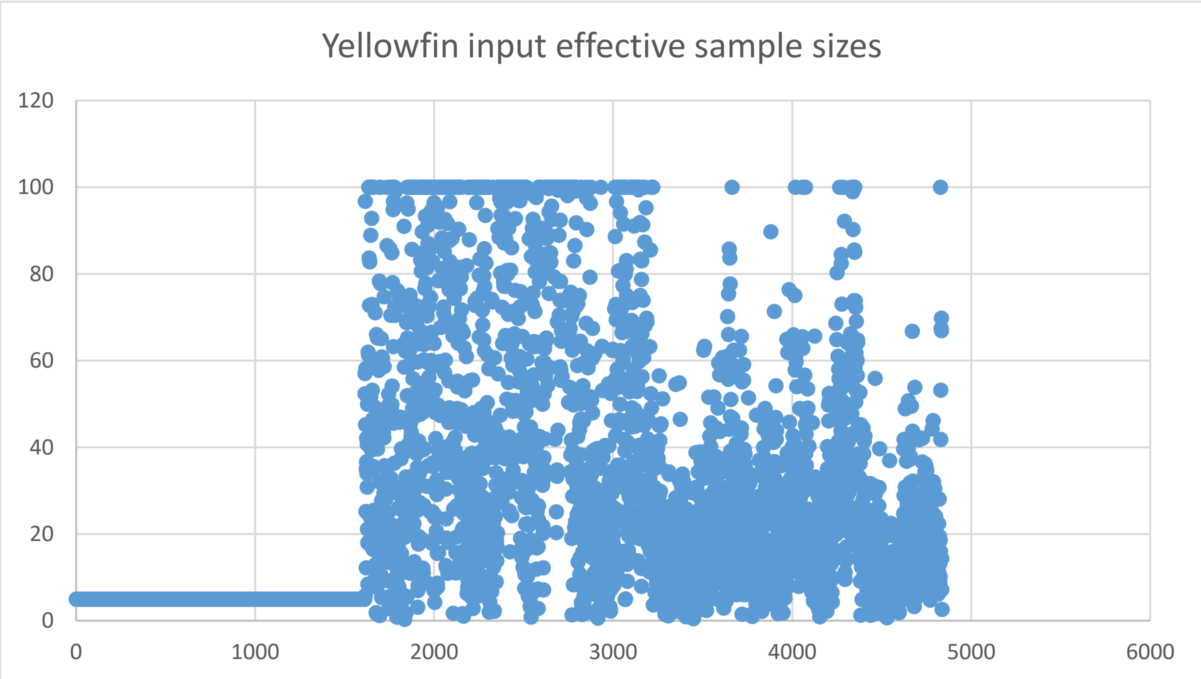


Figure 1. Size composition effective sample size assumptions used in OM conditioning. Three blocks of observations are strung together across all fisheries (ess5, CLRW, CL75).

Short-term and chronic tag loss assumptions

The approach used in the 2018 assessment was adopted.

Alternate growth curve

In the 2018 yellowfin assessment, the growth curve of Dortel et al. (2015) was recognized as a plausible alternative to the Fonteneau (2012) curve used in previous assessments. We are not aware of any strong argument as to why the Dortel curve should be considered inferior (though the method for filtering quality observations might be debatable), so both are proposed to be equally likely for the reference case OM.

Alternate CPUE assumptions

Targeting method

The assessment uses clustering in the temperate regions and HBF in the tropical regions in attempt to account for species targeting. The alternate proposal was to use clustering in both the tropical and temperate regions. As shown in Figure 2, the clustering approach is slightly more optimistic in western equatorial, but largely the same in the eastern equatorial.

Area-weighting (regional scaling)

Hoyle (2018) examined dozens of new candidate CPUE area-weighting factors, all of which corrected an earlier problem (ignoring the change in 5x5 degree surface area with changing latitude). For the reference case OM, we included the recommended weighting from the latest analysis (1979-94 period, model 8) and considered the two alternatives from the assessment (Figure 3). The difference among series appears trivial, except perhaps in the SE region. We retained only the 1975-1994 alternative weighting factors for OM testing, as it seems to provide the greatest contrast to the preferred assessment assumption.

Catchability trend

The 1 % per year (compounded annually) CPUE trend assumption clearly has the largest effect on the CPUE series among the factors examined (Figure 4).

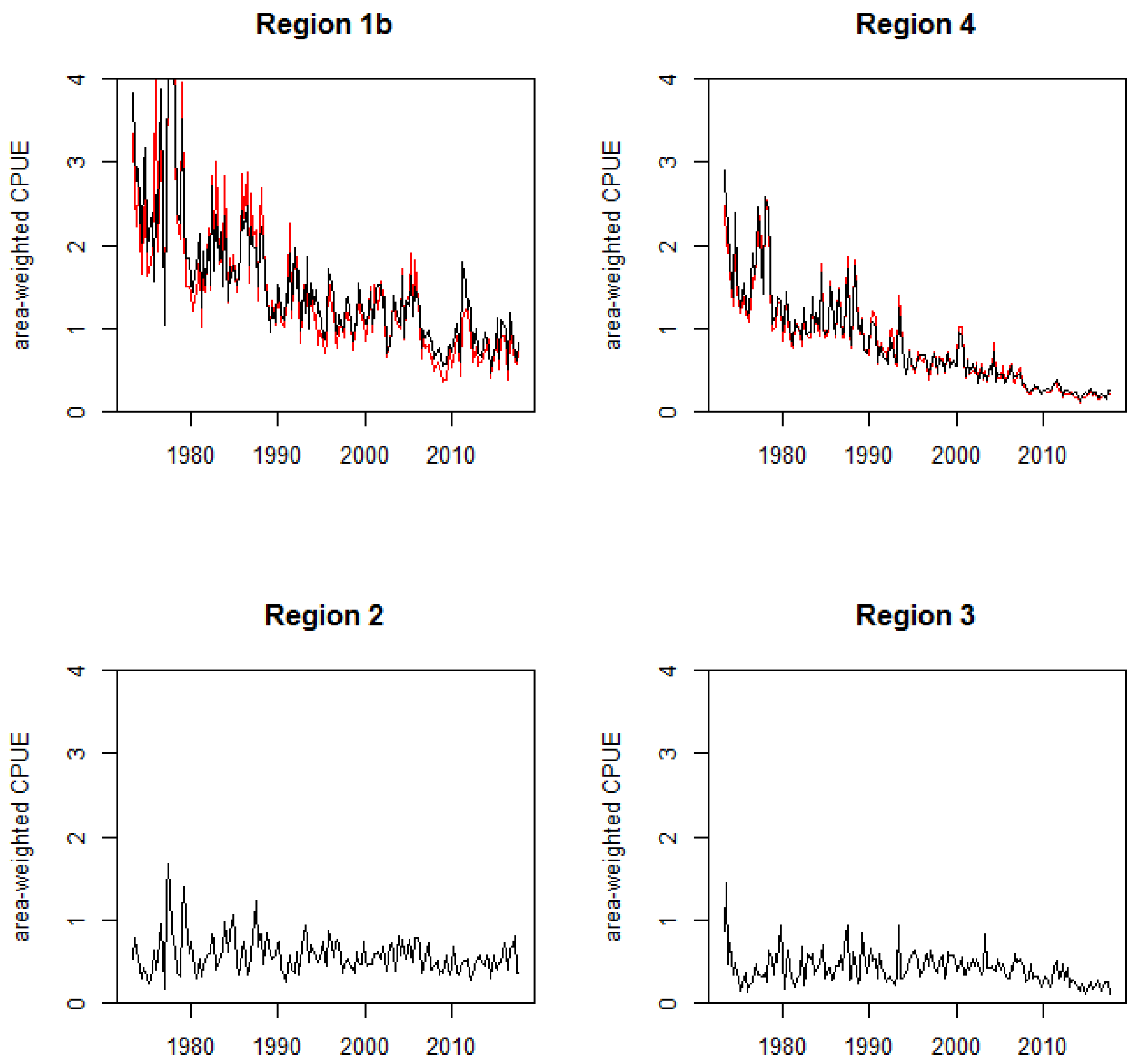


Figure 2. Standardized CPUE series used in the current iteration of the OM (Red=HBF, black=clustering).

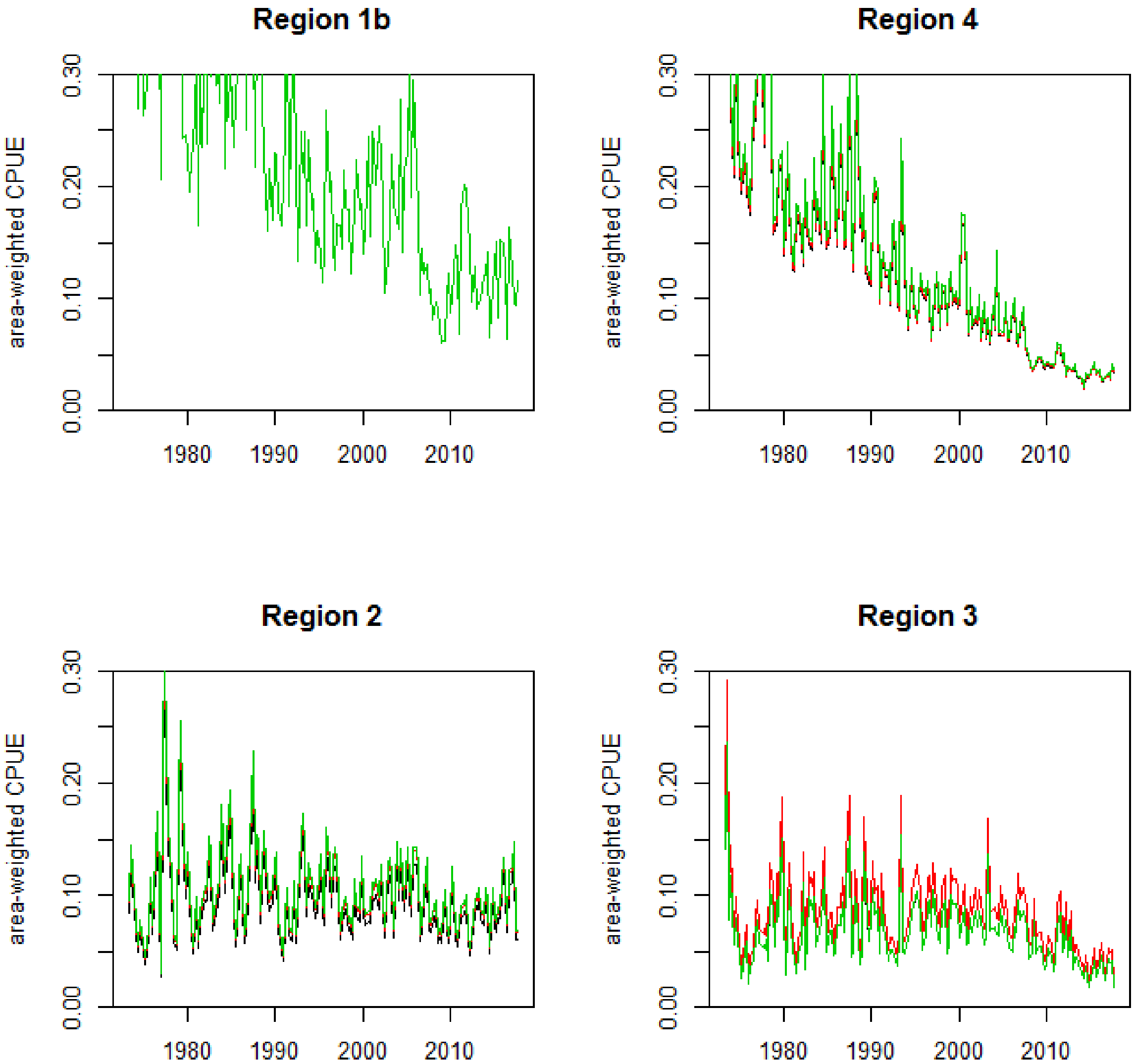


Figure 3. Comparison of the original assessment CPUE series with 3 different area weighting factors

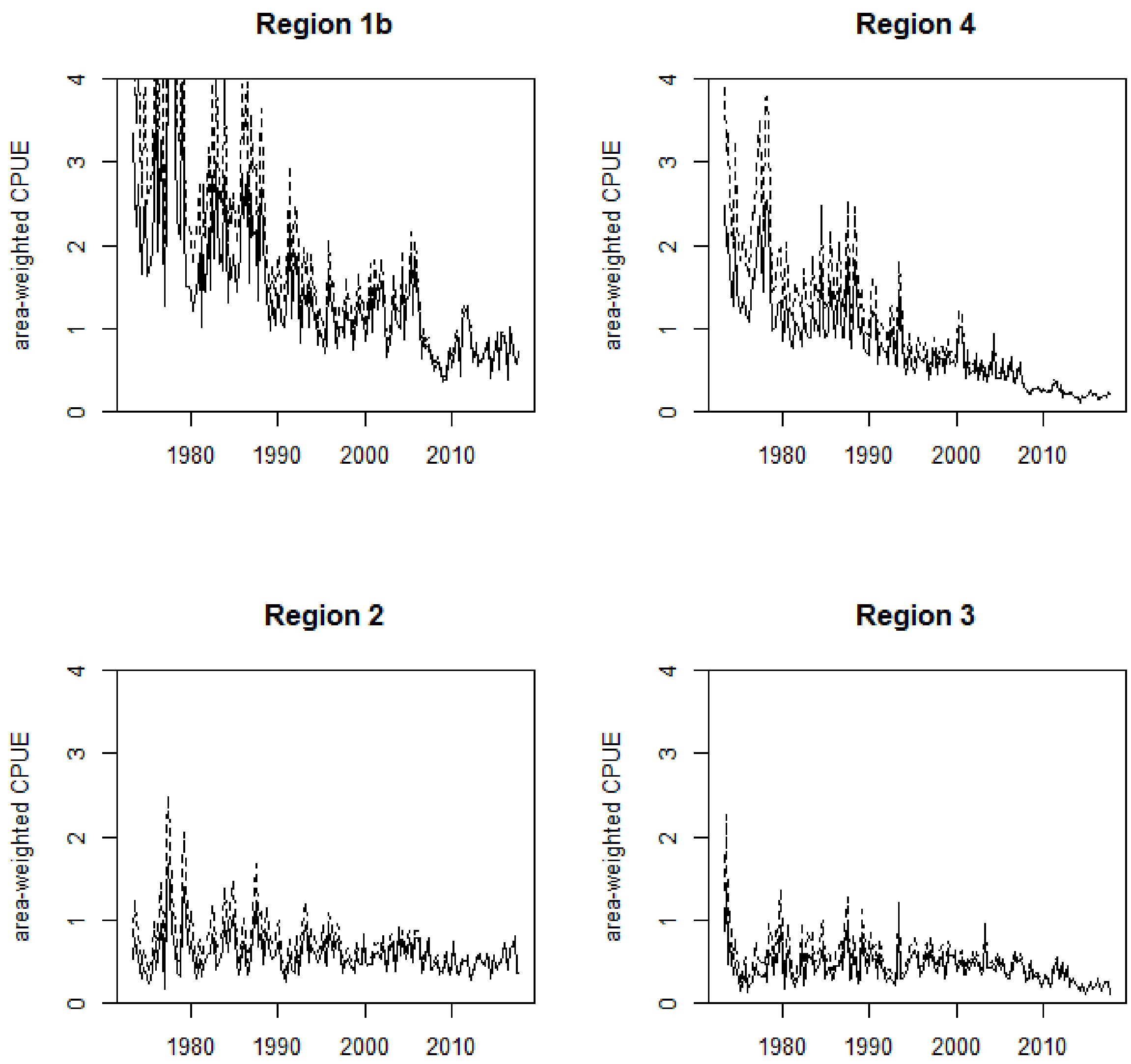


Figure 4. Comparison of CPUE series with catchability trend of 0 (solid) and 1% per year (broken).

Parameter estimation sensitivity to initial values

We know that the SS3 function minimization is sensitive to the initial parameter values. As discussed in the BET companion paper, we opted to repeat the function minimization until 2 independent convergences were met from jittered starting parameter values (or 10+ failures occurred). Convergence was not achieved in ~15% of the models in the YFT grids examined (and the failures appear to be related to the catch penalty problem).

Parameters on Bounds (and arbitrary priors)

In this iteration, there was not time to systematically address the problem of parameters on bounds. We note that this remains a concern, but it appears to be less an issue than the previous iteration. As with BET, it remains unclear how to automate a solution for this problem. i.e. If bounds are simply relaxed, implausible results can occur, and somewhat arbitrary priors may be required to prevent this from happening.

Fractional Factorial Experimental Design

This approach was recommended by the WPTT/WPM 2018, and appears to be very useful as discussed and explored in the BET companion paper. The 144 model fractional grid of OMgrid19.2 would be 6912 models in a full factorial. However, further we recognize that the *main-effects-only* design may under-represent the uncertainty, and should be investigated further.

Comparison of Yellowfin OM ensemble conditioning results

OMgridY19.2 - 144 model ensemble, fractional factorial main effects design (3 X 3 level and 6 X 2 level factors), of which 121 reached the convergence criteria (twice). The intent was a screening grid to quantify main factor effects on the conditioning, to prioritize the most important factors, potentially drop less important factors, and define a new fractional grid without aliasing of main effects or 2 way interactions. Due to time constraints, the second step was not conducted.

OMgridY19.2CP5 - OMgridY19.2 excluding models with catch penalty >0.00001 (66 models retained).

OMgridY19.2SL - The reference case OM for discussion purposes at the 2019 MSE Task Force meeting - OMgridY19.2 excluding models with dome-shaped selectivity and convergence failures (69 models retained).

Figure 5 - Figure 12 compare a number of grid summary diagnostics, partitioned by model assumption for each of the 3 grids above, from which we note:

- The YFT grid is much more stable than previous iterations, i.e. stock status characteristics are basically unimodal, and the problematic recruitment deviation trends are greatly reduced. Thus, the controversial bi-variate grid sampling approach employed in the last iteration appears unnecessary at this time.
- There is a bimodal distribution in the catch likelihood reported by SS3 (Figure 5). Many of the models that failed to converge had this penalty active at runtime. This means that F is approaching the user-defined limit (2.9) which seems like a reasonable grounds for doubting the plausibility of a model. i.e. the lower the population, the better the data fit, while the catch penalty is preventing the population from hitting an arbitrarily defined implausible value (F=2.9 is ~95% exploitation rate). It may be only a very small demographic subset that

is reaching this high F, i.e. it may only be the oldest fish in a particular quarter, in the most depleted region). Even though the catch penalty may seem small, it appears to be an indication of problematic combination of assumptions

- As might be expected, filtering on the basis of catch penalty > 0.00001 removes disproportionately more models at the very pessimistic end of the stock status characteristics. The factor options that are most susceptible to the problem include the logistic selectivity function, and relatively high weighting of CPUE relative to size composition (e.g. i1 and ess5).
- There is a strong conflict between the fit to the CPUE data and size composition, with the CPUE clearly supporting more pessimistic depletion levels (though the MSY estimates are not as clearly affected). In general, we would prefer to overweight the CPUE trend (because if you do not believe the relative abundance index, these assessments are probably almost meaningless). However, in the case of the YFT OMs, the alternative CL weighting was originally introduced to add variability to the bi-variate sampling grid. This might merit revisiting in the current context.
- The least sensitive assumption option appears to be the mixing period, but this may be misleading because this option is irrelevant for 50% of the results (those with down-weighted tags).

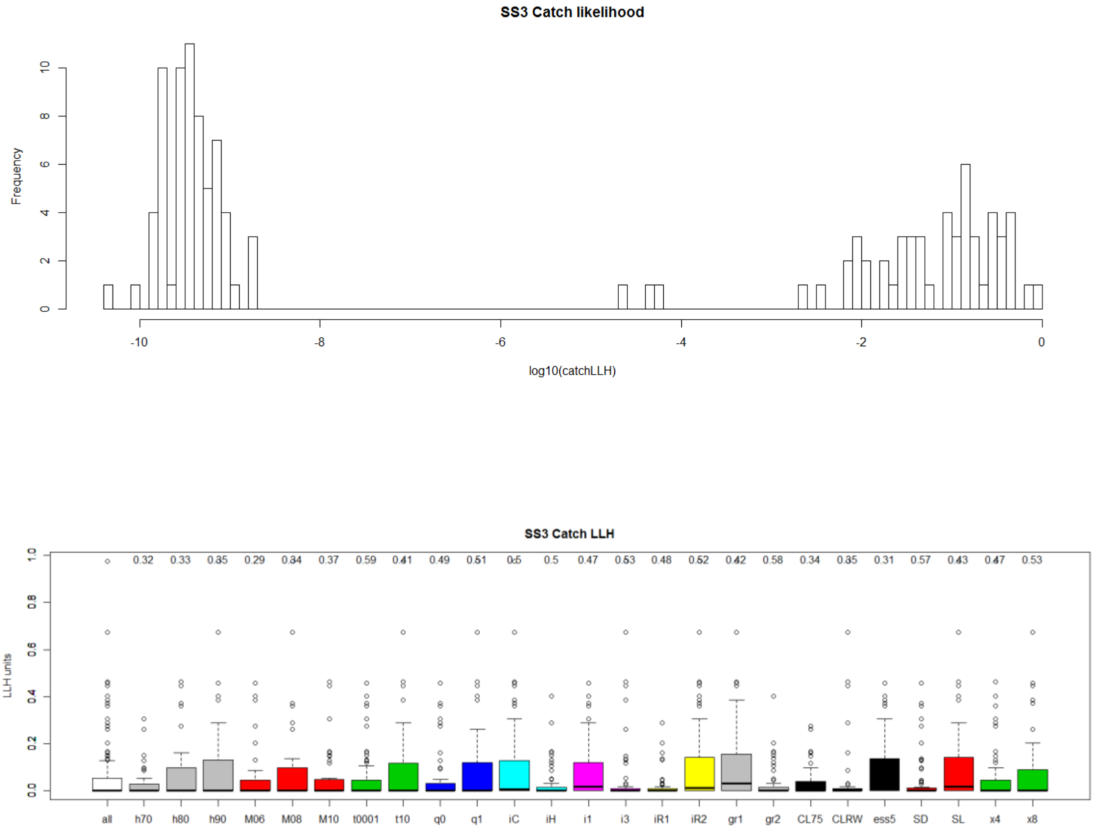
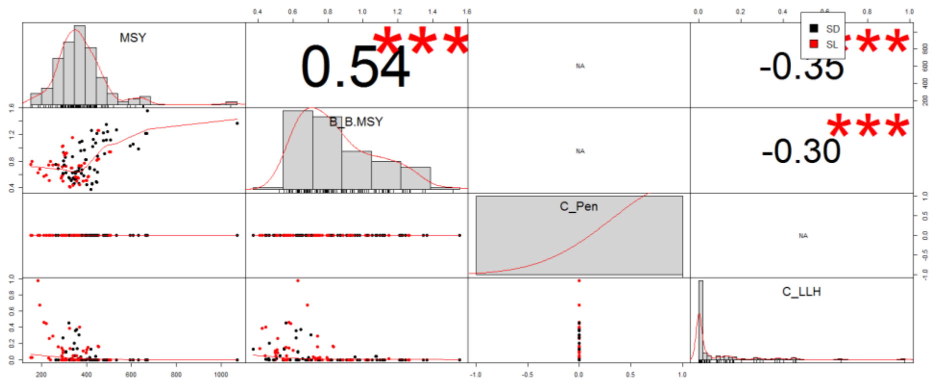
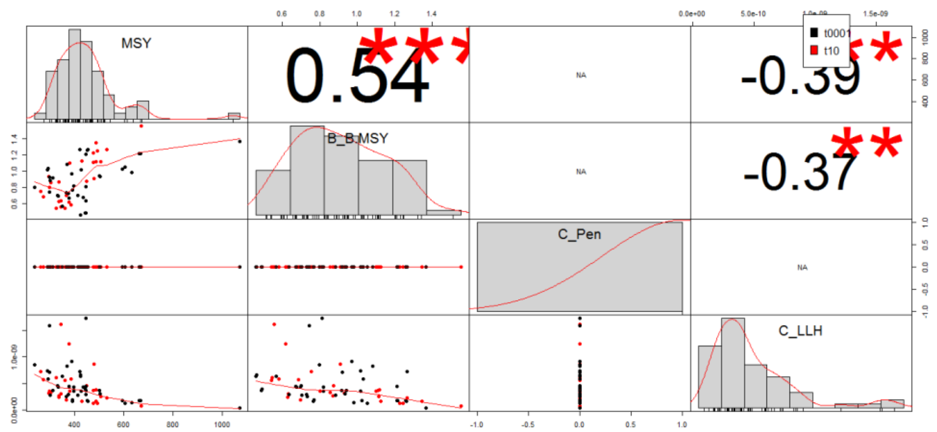


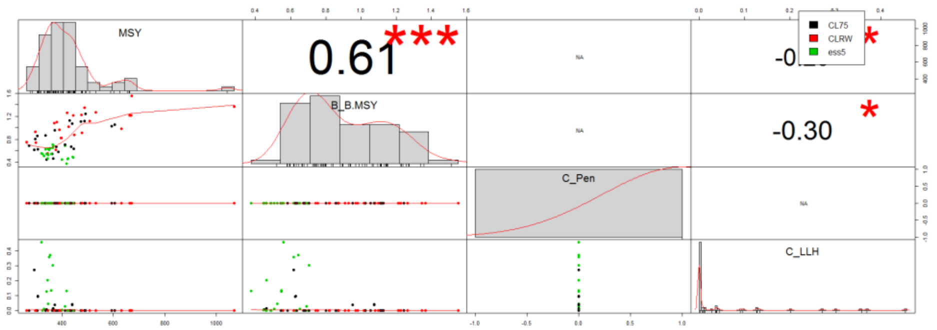
Figure 5. Catch likelihoods for OMrefY19.2 (partitioned by grid options in the bottom panel).



Y19.2



Y19.2.CP5



Y19.2.SL

Figure 6. Relationship among key stock status characteristics and the SS3 catch likelihood (C_LLH). C_Pen = crash penalty (0 in all cases examined)

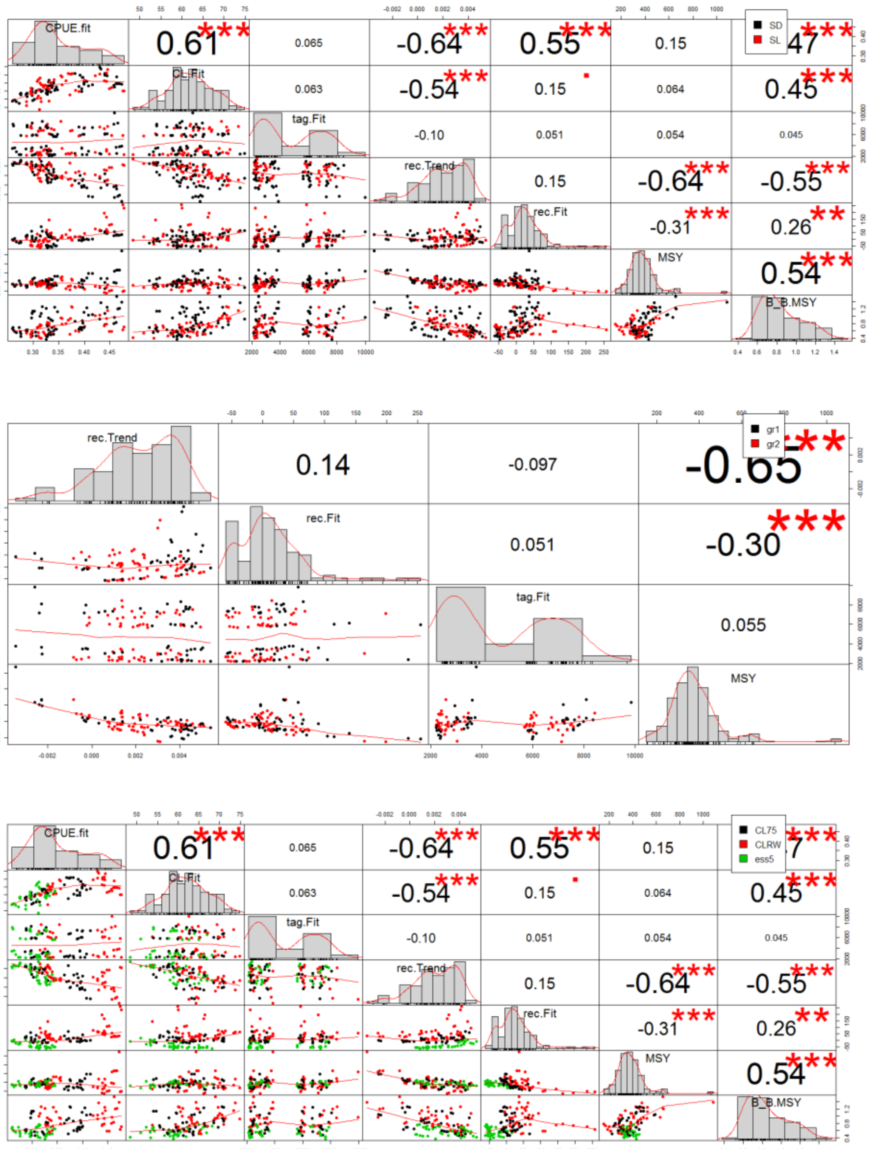
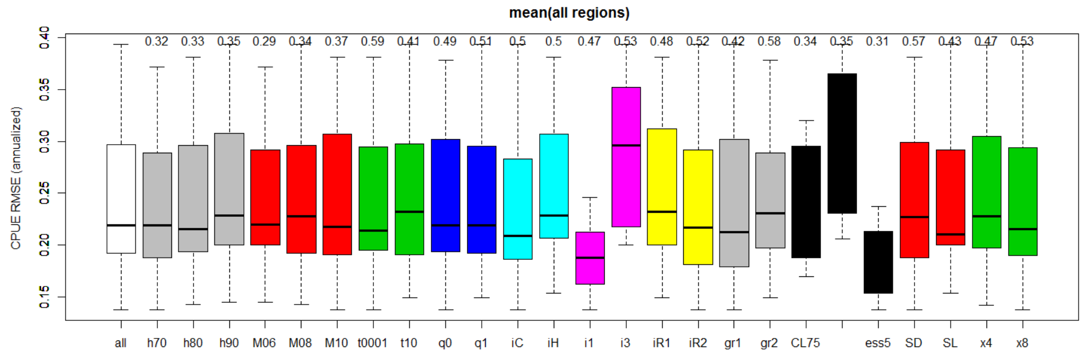
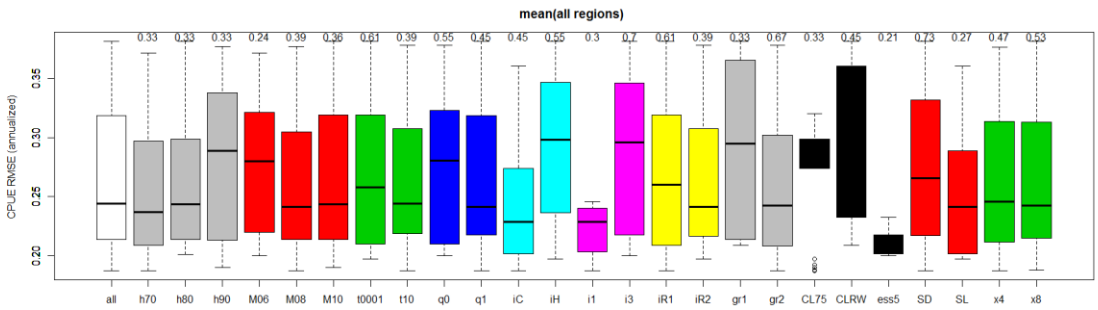


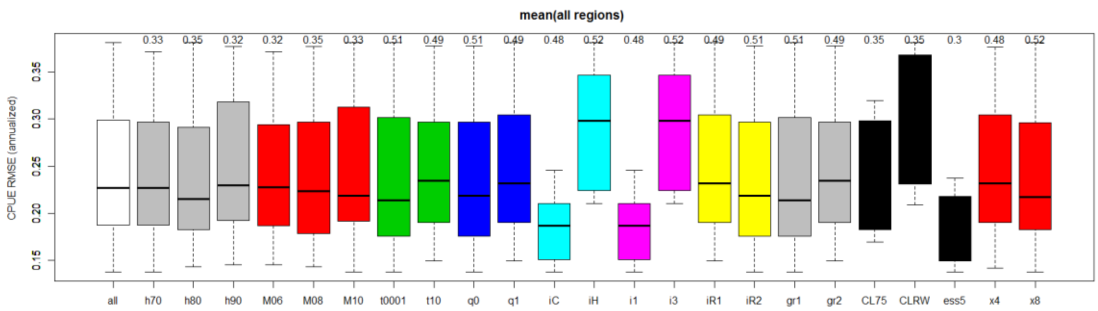
Figure 7. Multiway comparison of OMgridY19.2 model characteristics, partitioned by new or unexpectedly influential dimensions - longline selectivity (top), growth (middle), size composition weighting (bottom).



Y19.2

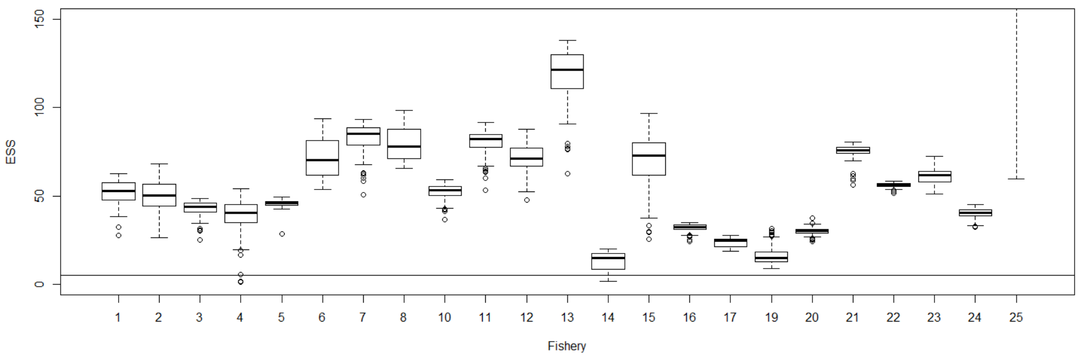


Y19.2.CP5



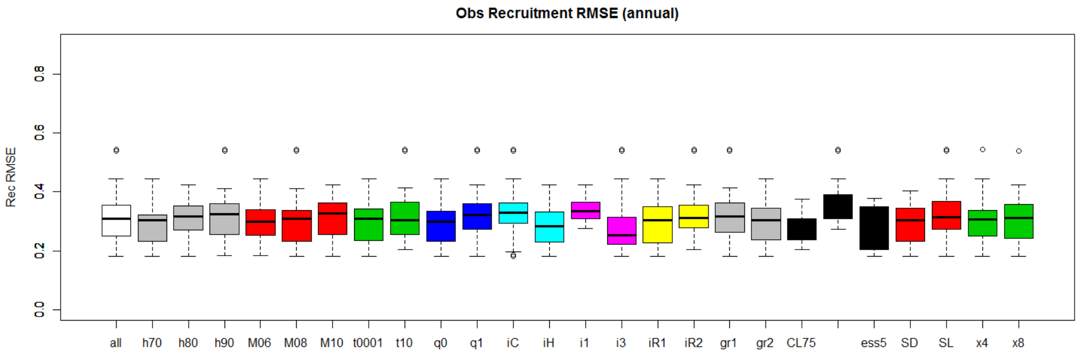
Y19.2.SL

Figure 8. Comparison of CPUE fit among BET OM ensembles.



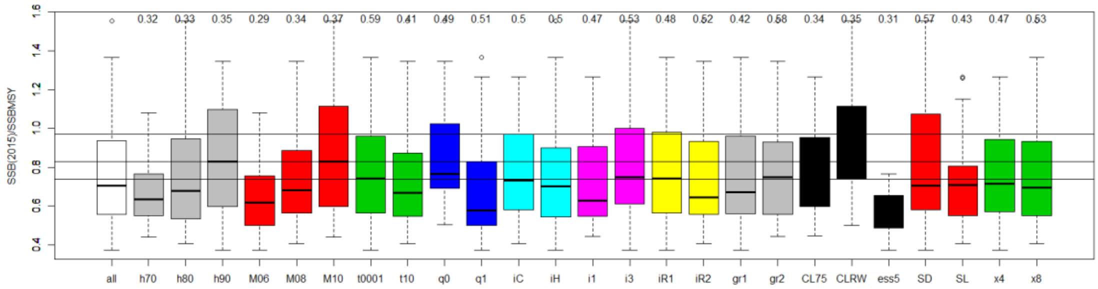
Y19.2

Figure 9. Comparison of CL fit among BET OM ensembles.

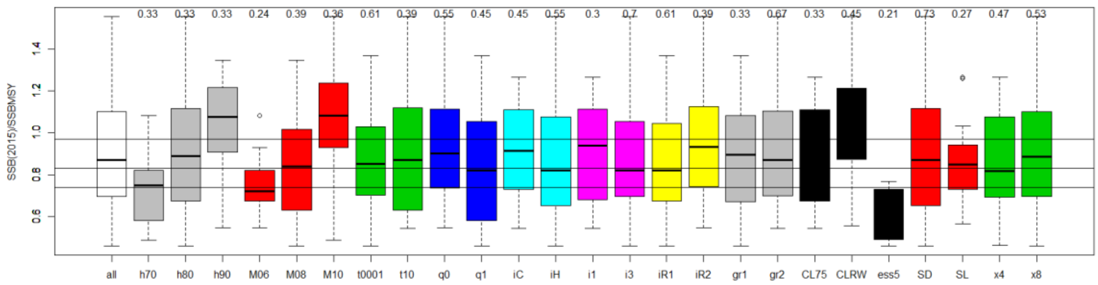


Y19.2

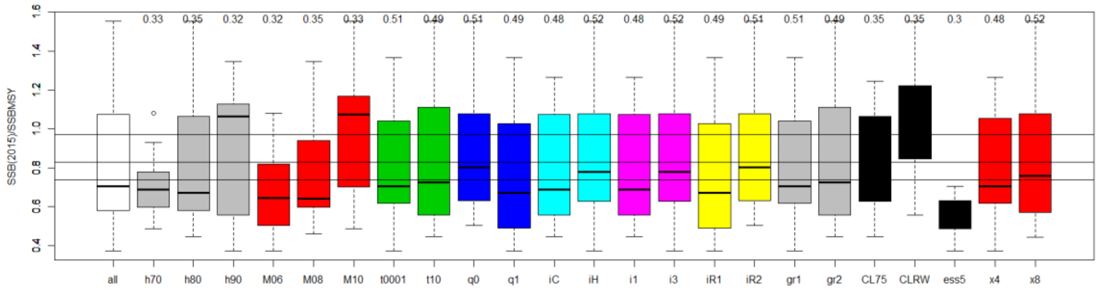
Figure 10. Comparison of Recruitment Deviation magnitude among BET OM ensembles.



Y19.2

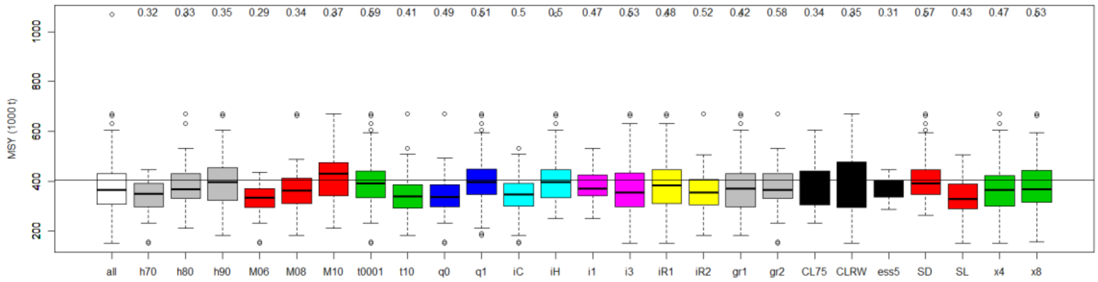


Y19.2.CP5

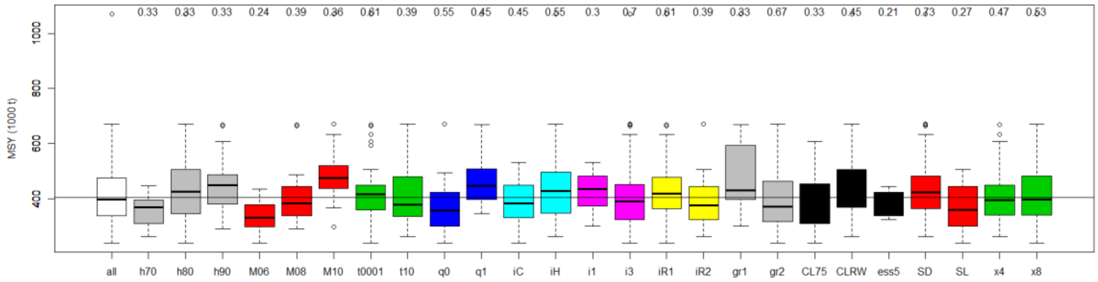


Y19.2CP5

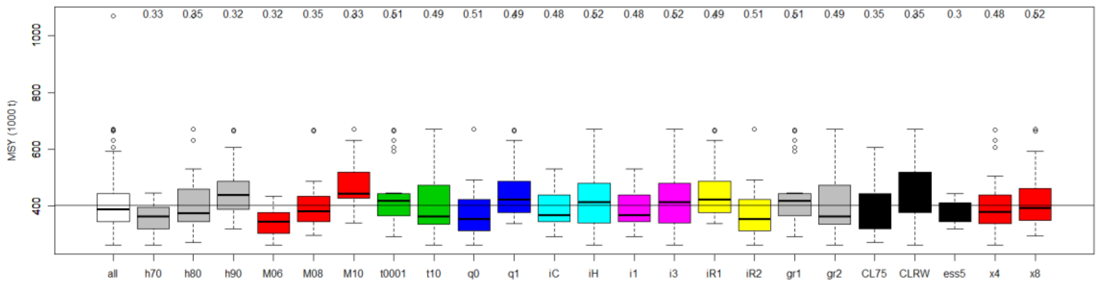
Figure 11. Comparison of estimated current depletion among BET OM ensembles.



Y19.2

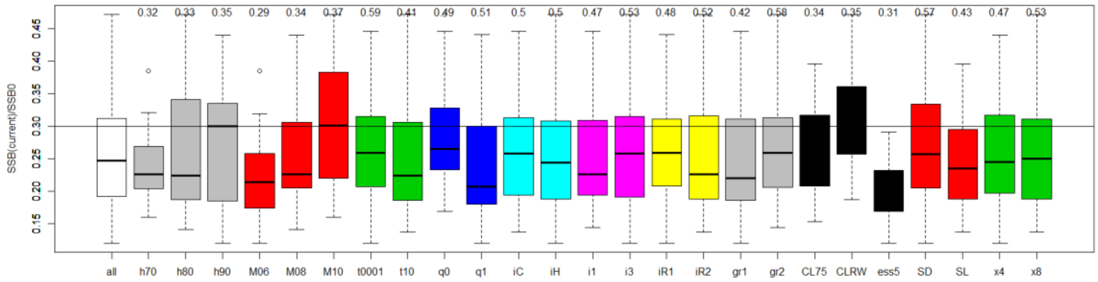


Y19.2.CP5

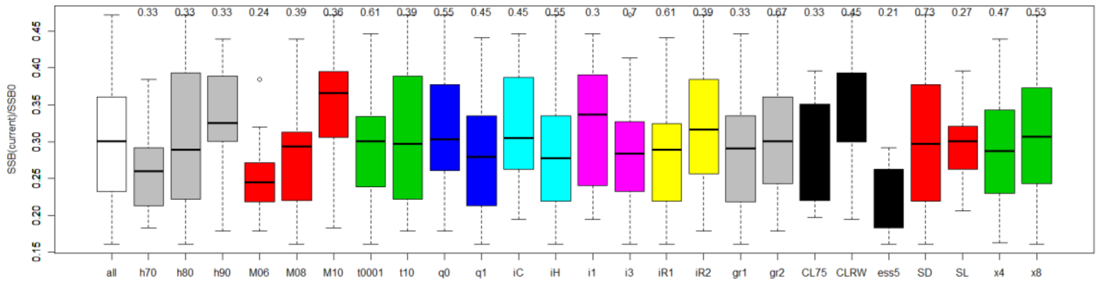


Y19.2.SL

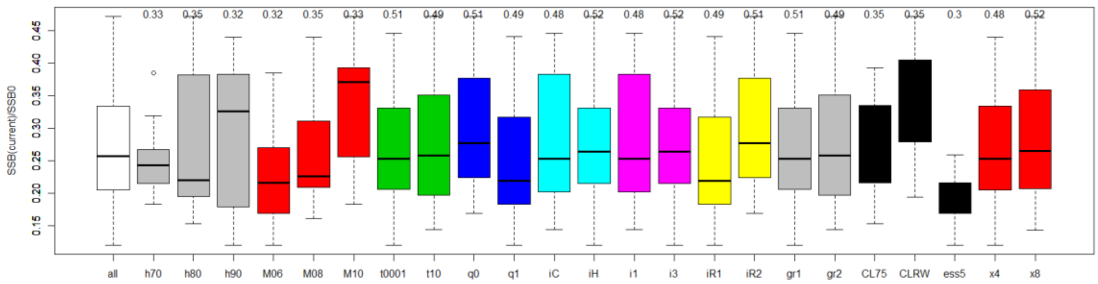
Figure 12. Comparison of estimated MSY among BET OM ensembles.



Y19.2



Y19.2.CP5



Y19.2.SL

Figure 13. Comparison of estimated depletion among YFT OM ensembles.

Revised Projection Assumptions

The reference case OM projection assumptions were updated as recommended by the WPTT and WPM in 2018 (attachment 1), including:

- The reference case annual CPUE CV = 0.2 and auto-correlation = 0.5. Note that this is not actually a change from the previous iteration (a problem was identified that turned out not to be a problem). These values are toward the lower end of the OMgrid19.2 ensemble (Figure 8), but near the upper end of the models with the extreme down-weighting of the size composition data (ess5, as in the 2018 stock assessment). It is thought to be unrealistically optimistic to expect the CPUE to be more accurate and precise than these values. However, it may be more appropriate to use a higher or model-specific CV, particularly if CL75 and CLRW options are retained. However, the CV = 0.30 robustness scenario results suggests that the difference is not very influential in the MPs tested (e.g. Figure 15 and Figure 17).
- Simulated MP implementation was changed to start in 2021, and the bridging catches for the intervening years were updated from the WPTT 2018 figures (constant 2017 catch).

Yellowfin OM MP evaluation results

Key MP evaluation summary plots for OMref19.2 are shown in Figure 14 and Figure 15, for the default candidate MPs discussed in previous iterations (M = Pella Tomlinson model, I = CPUE-based, and C = constant catch) and for the three rebuilding tuning objectives defined by the 2018 TCMP. From these we note:

- 10-25% of the OM results have a suspiciously high recruitment event at the end of the conditioning time series, which requires further investigation (i.e. recent recruitment estimates should have very low CVs imposed in SS3, with projection assumption error added from the OM projection code).
- The YFT MP evaluations are more pessimistic than the previous iteration.
- In general, catches need to be substantially reduced from current levels to attain any of the tuning objectives.
- The 2024 rebuilding objective could not be attained with a 25% TAC change constraint (but could with a 50% constraint).
- The 2029 and 2034 rebuilding objectives could not be attained with a 15% TAC change constraint (but could with a 25% constraint).
- The candidate MP results all show undesirable behaviour post-2030, in that they continue rebuilding substantially beyond the target.

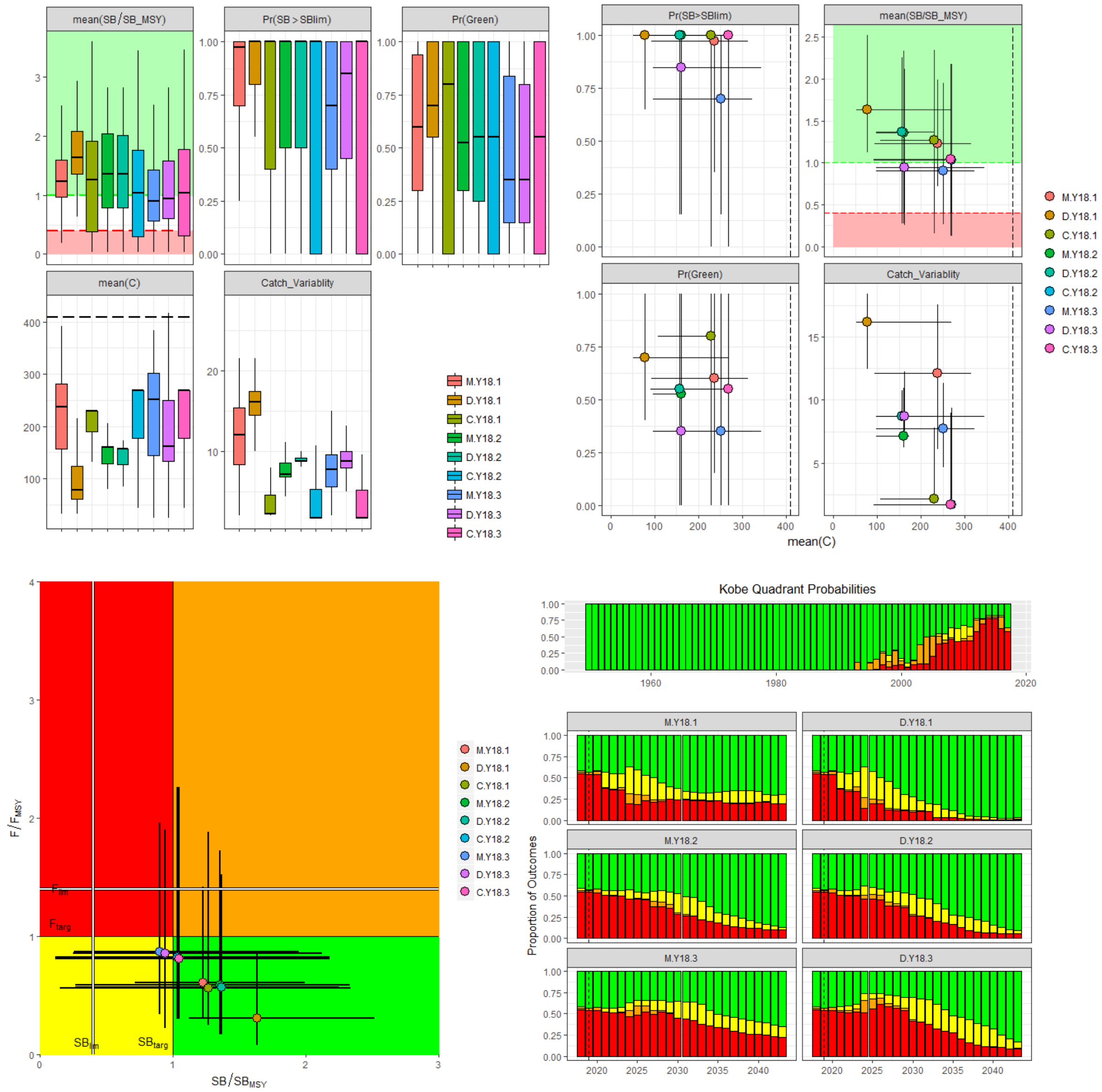


Figure 14. Reference case OMref19.2SL MP evaluation 20 year projection summary plots.

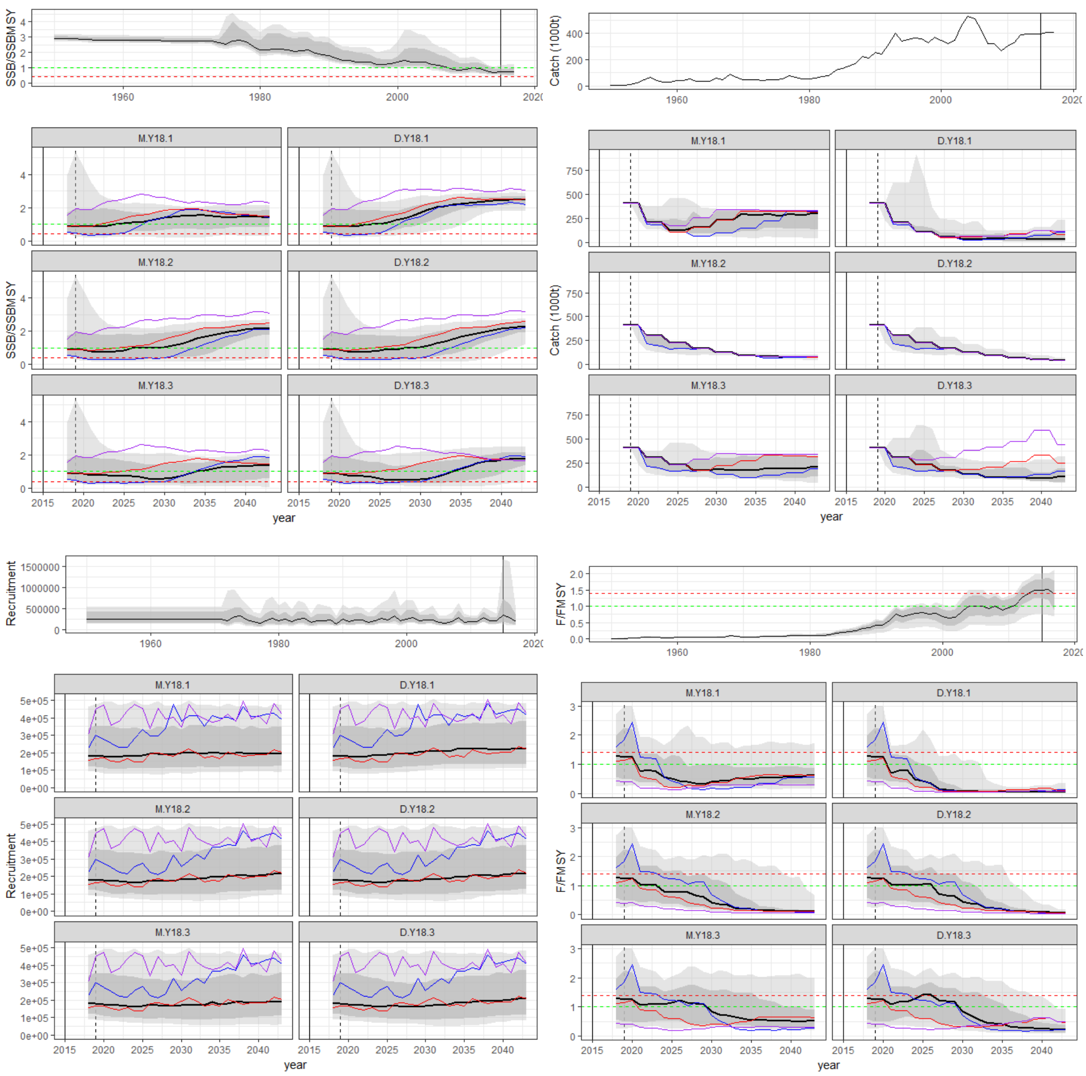


Figure 15. Reference case OMref19.2SL MP evaluation time series plots.

Robustness tests

We comment on 8 robustness tests below, only 3 of which were completed this iteration.

1) If the OM is biased to be too pessimistic, how will the tuned MPs perform? Figure 16 shows the subset of OMrefY19.2SL MP evaluations for models in which B(current)/B(MSY) >1, when managed by the MPs that are tuned to the full reference case OM. This has proved a useful demonstration in other situations. Ideally it would show that if the OM happens to be more pessimistic than the industry perception, the feedback-based MPs will still provide reasonable performance. Unfortunately, in this case, the MPs still prescribe substantial catch cuts for the optimistic OMs, and allow the biomass to rebuild well above the target. This suggests that the MPs do not have much

power to distinguish between optimistic and pessimistic situations, though we do not know if this is due to the limited signal to noise ratio (e.g. in the CPUE), or poor MP design.

2) Annual aggregated longline CPUE CV = 0.3 (auto-correlation = 0.5) MP evaluation results are shown in Figure 17.

3) 10% reported over-catch (projections only; reference case conditioning) MP evaluation results are shown in Figure 18.

4) 10% unreported over-catch (projections only; reference case conditioning) - this was not completed.

5) 3% LL catchability trend (projections only; reference case conditioning) - this was not completed.

6) Dome-shaped longline selectivity (noting potential for interaction with M and growth) - This option was included in the conditioning of OMgridY19.2 (e.g. figs Figure 8 - Figure 13), but merits further evaluation and has not been tested in projections.

7) Include free school PS CPUE in the conditioning with a 1% per year CPUE catchability trend. This has not been done, but we propose to look at this in a slightly more defensible way before the 2019 WPTT/WPM. Specifically, we suggest estimating a constant catchability trend for the PSFS CPUE series, while truncating the NW LL CPUE series at 2007 (start of piracy), and projecting the CPUE series forward in the MP with this catchability trend provided that i) constant PSFS catchability seems consistent with the LL CPUE up to 2007 and ii) the catchability-corrected PSFS CPUE series differs appreciably from the LL CPUE series post-2007.

8) The recruitment failure robustness test was not completed this iteration, as it was not mentioned in the WPTT and WPM reports, but we expect that this was an oversight and will be completed for the WPTT and WPM 2019.

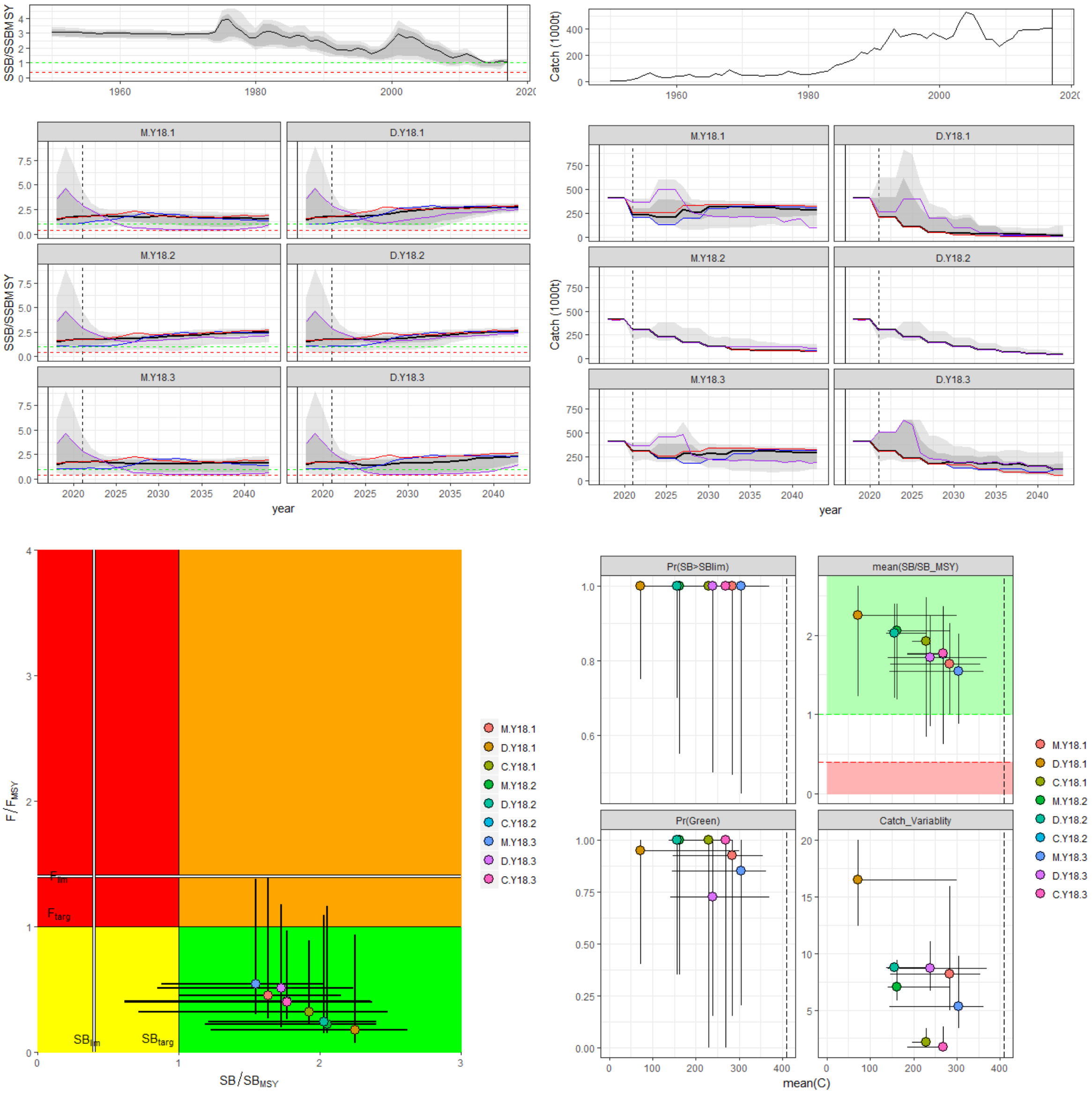


Figure 16. "Robustness" test results from OMrob19.2opt, highlighting the outcomes for the optimistic models in the reference case.

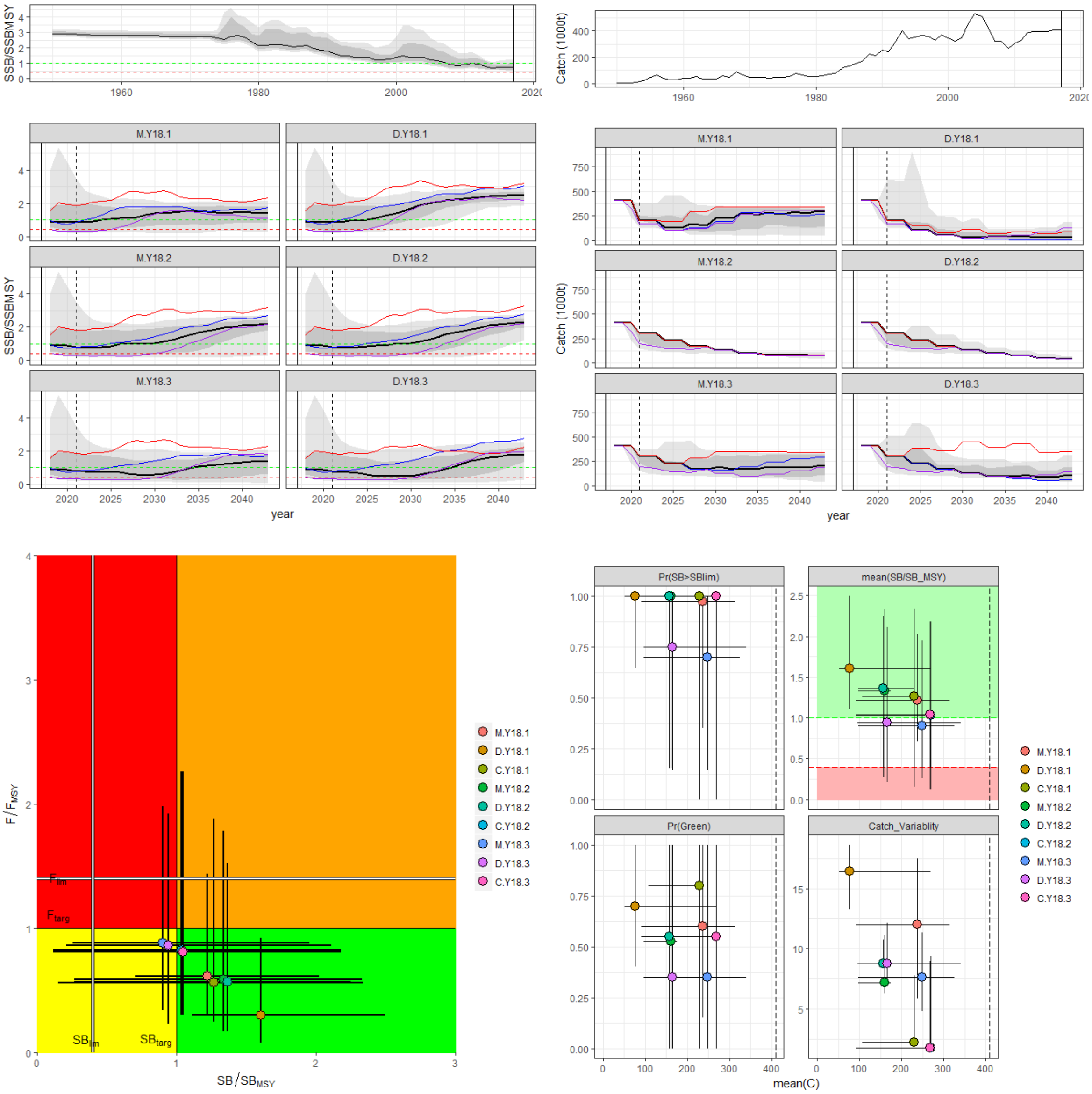


Figure 17.Robustness test results from OMrob19.2SLcv3, illustrating implications of annual longline CPUE CV = 0.3.

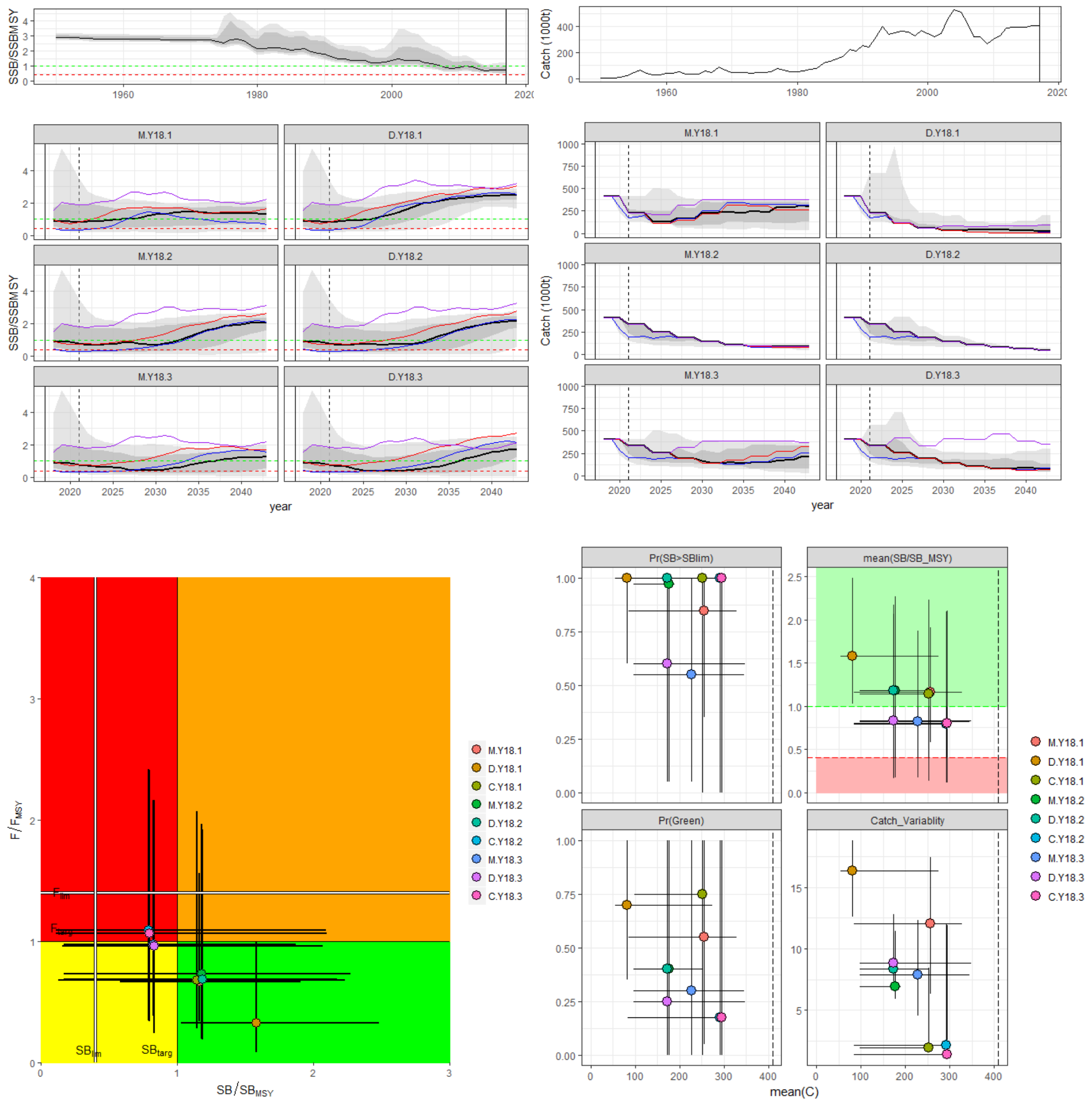


Figure 18.Robustness test results from OMrob19.2SLoc10, illustrating implications of 10% (reported) overcatch.

Key Discussion Points for the IOTC MSE Task Force:

1. Relatively high weighting (i.e. good fit) for the CPUE series (combined with logistic selectivity) tends to be associated with a high proportion of models that exhibit implausibly high F (resulting in a numerical penalty that constrains the function minimization). This represents a conflict with the general principle that relative abundance indices are generally the most informative data in an assessment and should be fit reasonably well as the highest priority.

2. We recommend absorbing the alternative growth curve into the reference case OM (and retaining the original), because the results seem to be moderately sensitive to this result and it seems to be as plausible as (and probably more thoroughly reviewed than) the default in the assessment.
3. We recommend retaining the alternative area-weighting scheme for longline CPUE indices (in addition to the original scheme), because the assessment results seem to be moderately sensitive to this option. Though this seems counter-intuitive given the seemingly small difference in values and might be an artefact of the fractional grid or convergence sensitivity to initial conditions.
4. We recommend retaining the double normal longline (CPUE) selectivity option as a robustness scenario at this time. However, further examination of plausibility is required, as there were unresolved bounds problems with this option (at least for BET), and the dynamics may be substantially altered.
5. We recommend proceeding with a fractional factorial design for the reference case, without aliasing of 2 way interactions. However, if the total grid size cannot be kept below ~200 models, we request the group to decide whether it is preferable to reduce the number of factors/levels or reduce the design to only include non-aliasing of main effects. Candidate grids will be presented during the MSE Task Force meeting.
6. We recommend continuing to seek 2-3 convergences for every model specification to reduce the effect of minimization artefacts.
7. Diagnostics for plausibility remains problematic with a large grid of complicated models. We look for recommendations for the group re:
    a. The catch likelihood term. Should we use this penalty as a plausibility filter?
    b. Can we impose intentionally informative priors on the selectivity for small fisheries (e.g. <1% of total catch)?
    c. Can we agree on priors and/or bounds for initial equilibrium fishing mortality and migration rates?
    d. Can we reduce the size composition sampling options?
8. Robustness test OMs that require reconditioning - do these need to be rerun with a full conditioning grid (with one dimension changed), subject to the same level of diagnostic evaluation as the reference case, and subject to MP evaluation?
9. Preparation for the 2019 TCMP:
    a. Agree OM reference set definition and contingencies.
    b. Do not use the bi-variate grid-sampling as in previous YFT iterations (unless the interactions of the new grid revert to the earlier unstable behaviour).
    c. Do the TCMP 2018 tuning levels cover a sufficiently desirable part of the management trade-off space, or should we define additional levels?
    d. Do not present any robustness scenario results to the 2019 TCMP. Possible exception would be to demonstrate that the MPs can provide reasonable performance even if the assessment/OM is currently pessimistically biased (though the example test was not very successful in this regard).
    e. Tune a range of MPs for behaviour that is more stable, and/or more responsive to new data.
10. Preparation for the WPTT/WPM 2019:
    a. Evaluate MP results for new tuning levels from 2019 TCMP. Attempt to refine MP performance if TCMP provides additional management objective insights, or evidence from other studies suggests that additional information can be extracted using different approaches.

    b. Conduct a more thorough investigation of the dome-shaped LL selectivity robustness scenario (noting bounds issues).
    c. Examine potential for using PS CPUE by estimating continuous trend, with LL CPUE from NW removed after 2007 (piracy start). i.e. This is more defensible than simply assuming a 1% per year trend as recommended by the 2018 WPTT.
    d. Review issues of simulating CPUE observation error in relation to spatial and temporal variability and potentially missing observations.
11. There is no funding identified for bigeye and yellowfin MSE scientific and technical support beyond Dec 2019.

Acknowledgements:

This work was jointly funded by the Global Environment Facility - Common Oceans - Areas Beyond National Jurisdiction Tuna Project with the technical cooperation of the Food and Agriculture Organization of the United Nations), and Australia's Commonwealth Scientific and Industrial Research Organization (CSIRO - Oceans & Atmosphere). The work does not necessarily reflect the views of the funding organizations.

References:

Dortel, E, Sardenne, F, Bousquet, N, Rivot, E, Million, J, Le Croizier, G, Chassot, E, 2015. An integrated Bayesian modeling approach for the growth of Indian Ocean yellowfin tuna. Fish Res. 163: 69-84. https://doi.org/10.1016/j.fishres.2014.07.006

Fu, D, Langley, A, Merino, G, Ijurco, AU, 2018. Preliminary Indian Ocean Yellowfin Tuna Stock Assessment 1950-2017 (Stock Synthesis). IOTC–2018–WPTT20–33.

Hoyle, SD, 2018. Indian Ocean tropical tuna regional scaling factors that allow for seasonality and cell areas. IOTC-2018-WPM09-13

Appendix 1. Extracts from the 2018 Methods and Tropical Tuna Working Party reports relevant to bigeye and yellowfin MSE

Methods Working Group (2018) draft report

1. BIGEYE TUNA AND YELLOWFIN TUNA MSE: UPDATE

1.1 Review of Operating Models based on WPM and SC feedback, including possible robustness tests

1. The WPM **NOTED** that two presentations were made in this section, summarizing three related working papers, IOTC–2018–WPM09–09 (BET OM definition), IOTC–2018–WPM09–10 (YFT OM definition) and IOTC–2018–WPM09–11 (BET and YFT MP Evaluations).
2. The WP **NOTED** that the MSE for both species is being pursued in the strict sense of MP in which the MP consists of simulation-tested combination of data collection, analysis methods and HCR (which makes this work different to the SKJ assessment, when no specification on data and analyses methods was made)
3. The WPM **NOTED** paper IOTC–2018–WPM09–09 which provided an update on the IOTC Bigeye Tuna MSE Operating Model Development. The following abstract was provided by the authors:

   "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. (See paper for full abstract):

4. The WPM **SUGGESTED** the following changes to the reference case OM grid:
   1. CPUE variability set to a level that would result in an annual CV of 0.2 (retaining auto-correlation of 0.5)
   2. Extend bridging catches, with first TAC in 2021
   3. Additional uncertainty dimensions:
      i. alternative growth function (noting the large effect on the recent WCPFC bigeye assessment). WPTT asked to review and specify the most appropriate alternative. This could be a robustness scenario.
      ii. alternative regional CPUE scaling factors
      iii. alternative historical catch series. Proposals were discussed, but the options were thought to either not represent a large change from the preferred series, or were difficult to justify as plausible.
5. The WPM **RECOMMENDED** exploring partially-confounded experimental design as a computationally tractable method for expanding the number of uncertainty dimensions and the main interactions (at the expense of losing higher order interactions). It should be adopted if if it is not found to have a significant reduction in full grid uncertainty.
6. The WPM **SUGGESTED** the following priorities for robustness scenarios:

1. Annual aggregated CPUE CV = 0.3 (auto-correlation = 0.5)
2. 10% reported over-catch (projections only; reference case conditioning)
3. 10% unreported over-catch (projections only; reference case conditioning)
4. 3% LL catchability trend (projections only; reference case conditioning)
5. Spatial Structure- possibly additional area around eastern INDONESIA, another in the Bay of Bengal Region and the area around Oman (other area stratification as is).
6. Non stationary M, linf and K in the projections.
7. Stock Structure (based on ongoing IO stock structure project).

7. The WPM **NOTED** that some of these robustness tests should be considered long-term ambitions, which would require more specific definitions and input from the secretariat and external parties, and would likely delay the current development timeline.

8. The WPM **NOTED** that some of the effects tested separately in the Robustness scenarios could eventually happen simultaneously and at least some scenarios should consider these effects in combination (e.g. catch misreporting and recruitment failure in the same simulation). However, it was further noted that an MP cannot be expected to handle every adverse situation and "exceptional circumstances" procedures are applicable in the worst cases

9. The WPM **NOTED** paper IOTC–2018–WPM09–10 which provided an update on the IOTC Yellowfin Tuna Operating Model Development. The following abstract was provided by the authors:

> *"This paper summarizes progress on the development of Operating Models (OMs) for IOTC yellowfin (YFT) tuna. MP evaluation updates for yellowfin and bigeye tunas 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).*
>
> *The latest version of the MSE software is publicly available from github, with a recently updated technical description and user manual (https://github.com/pjumppanen/niMSE-IO-BET-YFT/).The BET and YFT MSE projection software has undergone several changes in the past year, with a substantial rewrite to improve memory usage and parallel processing, which greatly improves MP evaluation speed. Most of these changes to the computational engine are not visible to the end user. (See paper for full abstract):"*

10. The WPM **NOTED** the high uncertainty and large number of implausible models in the uniformly weighted grid of the YFT Reference set OMs. It was recognised that the proposed approach of sampling the uniform grid with respect to the central tendency of the assessment was not ideal, but represented a pragmatic path forward.

11. The WPM **DISCUSSED** the alternative option of filtering plausible models in relation to habitat constraints as was used for albacore, and noted the following disadvantages in this case:
    1. It is not obvious that a meta-analysis of the productivity of 3 or 4 other YFT populations would provide more valuable insight about productivity than the arguments employed within the IOTC assessment process.
    2. The YFT MSY distribution forms a long-tailed continuum, unlike the disjointed polymodal distribution for ALB
    3. Unlike ALB, the YFT distribution also had many models that were implausibly unproductive (not only over-productive)

12. The WPM **SUGGESTED** the following changes to the YFT reference set OM grid, and expected that the WPTT would refine these recommendations, particularly with respect to insights from the new YFT assessment:

1. CPUE variability set to a level that would result in an annual CV of 0.2 (retaining auto-correlation of 0.5)
2. Extend bridging catches, with first TAC in 2021
3. Additional uncertainty dimensions:
   i. alternative growth function (noting the large effect on the recent WCPFC bigeye assessment). WPTT will be asked to review and specify the most appropriate alternative. This could be a robustness scenario.
   ii. alternative regional CPUE scaling factors
   iii. alternative historical catch series. Proposals were discussed, but the options were thought to either not represent a large change from the preferred series, or were difficult to justify as plausible.
   iv. It was noted that a new YFT catch data series will be discussed for the assessment at the WPTT, which is probably appropriate for the OM as well
4. Sample the OM grid using the bi-variate sampling approach (sampling with respect to the central tendency of MSE and SB(current)/SB(MSY), but with variance assumptions that are compatible with the distributional characteristics of the BET grid (for consistency)

13. The WPM **RECOMMENDED** exploring partially-confounded experimental design as a computationally tractable method for expanding the number of uncertainty dimensions and the main interactions (at the expense of losing higher order interactions). It should be adopted if if it is not found to have a significant reduction in full grid uncertainty.

14. The WP **SUGGESTED** the following priorities for robustness scenarios:
    1. Annual aggregated CPUE CV = 0.3 (auto-correlation = 0.5)
    2. 10% reported over-catch (projections only; reference case conditioning)
    3. 10% unreported over-catch (projections only; reference case conditioning)
    4. 3% LL catchability trend (projections only; reference case conditioning)
    5. dome-shaped longline selectivity (noting potential for interaction with M and growth)

15. The WPM **NOTED** that some of these robustness tests should be considered long-term ambitions, which would require more specific definitions and input from the secretariat and external parties, and would likely delay the current development timeline.

16. The WPM **NOTED** that some of the effects tested separately in the Robustness scenarios could eventually happen simultaneously and at least some scenarios should consider these effects in combination (e.g. catch misreporting and recruitment failure in the same simulation). However, it was further noted that an MP cannot be expected to handle every adverse situation and "exceptional circumstances" procedures are applicable in the worst cases

17. The WPM **NOTED** that alternative MP tuning levels should be adopted to add contrast to the results for the TCMP02.

Working Party on Tropical Tuna (2018) draft report

Bigeye

1. The WPTT **NOTED** paper IOTC–2018–WPM09–09 which provided an update on IOTC bigeye tuna operating model development, October 2018, including the following summary provided by the authors:

   "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."

2. The WPTT reviewed and **ENDORSED** the progress to date on MSE for bigeye tuna while recognizing the discussions held at TCMP and the advice of WPM, but **INDICATED** the need to consider some additional uncertainty dimensions in the bigeye tuna MSE workplan agreed by WPM.

3. In particular, WPTT **ENCOURAGED** that the MSE work consider the importance of an alternative growth curve for bigeye tuna. The WPTT **SUGGESTED** the growth curve estimated by Farley et. al. (2016) is based on a broader size range (up to 160cm+) and may have a more plausible Linf value (~178 cm) than the Eveson (2015) model currently used in the OM. Furthermore, the Farley et. al. (2016) growth curve is derived from samples from the eastern Indian Ocean so may provide additional information on growth from a different region. However, the WPTT acknowledged that the Farley et. al. (2016) growth function may not describe well the length-at-age for fish smaller than 70cm LJFL which is the size range of most of the tagged fish for which the model estimates age.

4. Therefore, the WPTT **SUGGESTED** either anchoring the growth curve to a plausible age at zero length, or preferably combining the data from the Farley et al. (2006) growth curve with the Eveson (2015) and fitting both Von-Bertalanffy Growth Function (VBGF) and multi-stanza growth models to determine the best model fit.

5. The WPTT expressed some concern in combining size at age data from different time periods to estimate a single growth curve due to the potential for temporal shifts in growth, but also **NOTED** that the inclusion of an additional growth curve was to capture a plausible range of uncertainty in growth.

6. The WPTT **NOTED** that there may be a need to revise the number of age classes used in the models when using a different growth curve due to shift in the distribution of size at age

Yellowfin

1.1 Update on Management Strategy Evaluation Progress

7. The WPTT **NOTED** paper IOTC–2018–WPM09–10, which provided an update on the development of the operating model for IOTC yellowfin tuna (October 2018).

8. The WPTT **NOTED** paper IOTC–2018–WPM09–11, which provided an update on IOTC bigeye and yellowfin management procedure evaluation progress (October 2018), including the following abstract provided by the authors:

   " This document presents MP evaluation results for bigeye and yellowfin tunas, using the new operating models (OMs) proposed in Kolody and Jumppanen (2018a, b) and the new tuning levels requested by TCMP (2018). The results of various robustness scenarios are included, at this point largely to help facilitate the discussion of their role in the MP development and selection process and how they should be presented to the TCMP."

9. The WPTT reviewed and **ENDORSED** the progress to date on MSE for yellowfin tuna while recognizing the discussions held at TCMP and the advice of WPM, but **INDICATED** the need to alter some of the assumptions used in the operating model grid and consider some additional uncertainty dimensions in the yellowfin tuna MSE workplan agreed by WPM.

10. The WPTT **NOTED** the need to modify the assumed time required to achieve mixing of tagged YFT with the untagged population to 4 quarters (from 3 quarters) based on decisions taken for the 2018 YFT stock assessment. Further, the WPTT **ENCOURAGED** that the MSE work consider the importance of also assuming the time needed for mixing of the tagged and untagged populations of 8 quarters for use in examining robustness of MPs to this assumption.

11. The WPTT **ENCOURAGED** that the MSE work consider the importance of alternative growth for yellowfin tuna based on the growth model estimated by Dortel (2014) for use in examining robustness of yellowfin MPs to alternative growth models.

12. The WPTT further **ENCOURAGED** that the MSE work also consider the importance of adding the Purse Seine Free School CPUE as documented in IOTC–2018–WPTT20–36_Rev1, assuming a 1% per year cumulative increase in catchability (q) for the time period, for use in examining robustness of yellowfin MPs.

13. The WPTT also **NOTED** that the decisions taken for the 2018 YFT assessment regarding short-term and chronic tag loss differed from the YFT Operating Model grid and **RECOMMENDED** that the 2018 YFT assessment assumptions be mimicked in the Operating Model grid.

14. The WPTT **NOTED** that the proposed new uncertainty dimensions would be evaluated with respect to plausibility and impact before deciding whether to assign them to the OM reference set or robustness trials. The informal MSE working group will review these decisions in March 2019.