IOTC–2018–WPM09–INFO2

OCEANS AND ATMOSPHERE
www.csiro.au

Update on IOTC Yellowfin Tuna Management Procedure Evaluation March 2018

Dale Kolody and Paavo Jumppanen

This paper was prepared for the Indian Ocean Tuna Commission Working Party on Methods informal MSE Meeting, Lisbon, Mar 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

This paper summarizes progress on phase 2 of the development of Operating Models (OMs) and evaluation of candidate Management Procedures (MPs) for IOTC yellowfin (YFT) tuna (a separate paper describes parallel progress for bigeye tuna). This report summarizes progress toward the cumulative list of development goals from WPTT and WPM 2016 and 2017.

The main focus has been the development of a new YFT reference case OM. The OM proposed by the IOTC working parties in 2016 (in the absence of results to review) was based on a number of independently reasonable assumptions, but many of the models had implausible dynamics due to assumption interactions (most evident in non-stationary recruitment) and very different inferences from the assessment. Models which excluded the tags tended to be very optimistic, and the CPUE and size composition data were not very informative about model plausibility. Tags were very helpful in constraining the model to a plausible parameter space, but there are reasons to doubt the tag inferences (i.e. questionable tag mixing assumptions). We propose a new approach for the OM. It involves i) expanding the grid of models with additional uncertainty dimensions, and ii) sampling the expanded grid (with replacement) to create an OM that has central tendencies for SB/SB(MSY) and MSY that are consistent with the assessment, but with CVs that are inflated by an arbitrary factor to be determined by the broader IOTC MSE community (results are shown for a factor of 3 inflation, CV ~ 13% for both quantities). Additionally, two dimensions were sampled to be equally representative - inclusion/exclusion of tags, and CPUE catchability trends of 0 and 1% per annum.

The proposed new reference case OM (OMref17.2) consists of a subsample from an (unbalanced) grid of 720 models (derived from the 2016 assessment) and including the following options:

- 3 X Beverton-Holt stock recruit relationship steepness
- 3 X Natural mortality vectors
- 3 X tag likelihood weighting
- 2 X tag mixing period
- 2 X CPUE standardization method
- 2 X CPUE catchability trend
- 2 X CPUE observation error assumptions
- 2 X catch-at-length sample size assumptions
- 3 X recruitment variance assumptions (σ_R)

OMref17.2 consists of a subset of ~330 models with a weighting factor for each, such that the MSE evaluations presented here are a random subset (200 replicates) with many models that are sampled only once, some that are sampled several times, and many that are not sampled at all. Key projection assumptions include:

- Initial states (with added error) and most parameters defined by the SS specifications
- temporal variability in selectivity for all fleets
- CPUE CV = 0.3
- annual recruitment CV = 0.6
- first TAC implemented in 2019; bridging catches 2016:2018 = 407Kt (2015 level)
- catch implementation error CV = 0.1

Results are presented for four MPs, evaluated with OMref17.2, at the two tuning levels requested by the TCMP for YFT:

- Y1: Pr(mean(B(2019:2039))/BMSY = 1.0) = 0.5
- Y2: Pr(mean(B(2024))/BMSY = 1.0) = 0.5

MP behaviour under the two tuning options is very similar, with Y2 being slightly more conservative. However, the MP testing has revealed a non-trivial issue in the OM (the risk was flagged at least two years ago, but the implications were not critical at the time). In a large number (e.g. 50 - 75%) of testing simulations, the MP is not able to take the full TAC because of numerical limits on fishing mortality (which will affect different regions to different degrees). This may be qualitatively realistic in the sense that some fleets will stop fishing when they cannot achieve economically viable catch rates. However, it is probably not quantitatively realistic, particularly for fleets that have the option to move to other regions. As a consequence, the effectiveness of the feedback control of the MPs is compromised (i.e. the TAC often has little relationship to the catch), and the MP tuning may result in HCR control parameters that do not make intuitive sense outside of the aggregate OM. e.g. The PT4010 MP (production model + "hockey stick" HCR) is tuned using the maximum catch parameter (for SB > 0.4SB0) and the result is ~3X MSY. Thus MP performance is inextricably linked to assumptions about how the fishery dynamics operate at extremely high fishing mortality rates, and the uncomfortable expectation that failure to extract the full TAC will be a very common event (and different fleets will be short-changed to different degrees). Additional diagnostics will need to be added to the simulation software to identify when this is an issue.

We seek guidance and/or endorsement from the MWG informal MSE working group (and BET/YFT MSE project steering committee) for the path forward. We consider the point above to be the highest priority issue (for which we hope to have options to discuss at the meeting), while the following points were proposed before the problem above became evident:

- YFT reference case OM (for TCMP 2018 and WPTT/WPM 2018 presentations):
  - OM to consist of OMref17.2 - a subset of 720 models sampled (with replacement) to attain the central tendency of SB/SBMSY and MSY matching the assessment, with ~13% CV on each.
  - Replicates to be increased from 200 to 2000
- Graphics are to be updated with the FLR package designed for the TCMP needs

- YFT robustness case OMs:
    - Not to be presented to TCMP 2018.
    - To seek guidance from the informal MWP MSE group on priorities.

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 (included in Attachment 1). Recognizing the iterative nature of the MSE process, the workplan identifies 2019 as the earliest probable date for MP adoption.

3 Summary of YFT MSE Progress toward the requests from the IOTC Working Parties

MSE for Bigeye and yellowfin tunas has been pursued in parallel, with the first phase of the scientific and technical work described in Kolody and Jumppanen (2016). A second phase project has commenced to support progress from Sep2017 to Dec2018. This second phase project is responsible for reporting progress to the IOTC subsidiary bodies (including the TCMP, WPM and WPTT), and implementing feedback to ensure that the MSE meets the scientific and technical needs of the IOTC community. Kolody and Jumppanen (2017) provided an update of the initial progress that had been made since the commencement of phase 2, including:

1) A "mechanical" update to the YFT reference case OM in line with the feedback from the 2016 IOTC technical working parties, and presentation of diagnostics for evaluating plausibility. The reference case is intended to encompass the main assessment uncertainties, and will usually provide the main descriptor of expected MP performance (MPs will be tuned to perform with respect to the reference set).

- The OM dropped the environmentally-linked movement from the assessment, since it has a minimal influence on the stock status inferences and adds another level of complexity to projections.
- It was recognized that the OM grid as defined included a large number of models that are very optimistic, e.g. MSY estimates considerably higher than the assessment. There was a strong (negative) correlation between MSY and the trend in recruitment deviates. Higher MSY models tended to explain a larger portion of the declining abundance trend to recruitment effects (rather than fishery depletion). The trade-off is achieved without an obvious detrimental fit to the CPUE or size composition data. However, the higher MSY models can also be considered implausible, as the recruitment deviations suggest a systematic lack of fit to the stock-recruit relationship. If steepness is estimated, these specifications are compatible with very low steepness and very low productivity, which is also implausible.
- Full weighting of the tagging data tended to constrain the models to a more plausible space (presumably by constraining absolute abundance). However, there is reluctance to constrain all OMs to options that include full tag data weighting because there are doubts about tag mixing assumptions being adequately met, and tags support dubiously low M estimates.
- Means of formulating the models to produce plausible results without the tags were briefly explored, and likely options include: increasing the weighting on the CPUE data (particularly region 2), and introducing temporal variability to the longline selectivity.

2) Exploration of potential YFT robustness OMs in line with the feedback from the 2016 IOTC technical working parties. Robustness cases generally include less likely, but potentially

troublesome dynamics, and may be used to identify MPs that are more robust to particular challenges (after they are tuned for the reference set). The term is also often used to refer to an untested scenario, which may subsequently be rejected, elevated to the reference set, or retained as a robustness scenario.

- Two attempts were made to formulate OM robustness scenarios that admit a potential tendency for longline fisheries to shift toward targeting younger individuals over time: i) estimating selectivity in 10 year blocks, and ii) estimating changes in selectivity as a monotonic function of time). Neither option resulted in a management situation that was substantially different from the OM-ref stationary selectivity assumption, and hence may not meet the expectations for robustness trials.
- Up-weighting the tagging data (tag $\lambda = 1.5$), results in similar, but slightly more pessimistic OM than the 2016 assessment tag weighting assumption ($\lambda = 1.0$). It is not clear that the $\lambda = 1.5$ robustness scenario adds a fundamentally different challenge for the MP than the $\lambda = 1.0$ option. However, it does emphasize the importance of the tag-weighting assumptions and the need to ensure that MP performance against pessimistic scenarios is explicitly considered (whether in reference or robustness scenarios).
- There was a request to consider inclusion of CPUE data prior to 1972. However, this was presumably aimed at BET, because the YFT assessment currently uses the CPUE data from 1972.

3) Presentation of some candidate MP results that meet the initial tuning objectives identified in TCMP (2017). Preliminary results were presented to illustrate how the tuning concept works, however, the results were not intended to be taken seriously, because of the concerns about the plausibility of the YFT OM.

WPM (2017) provided the following guidance for the next iteration (***bold** comments indicate progress):

> *51. The WPM AGREED on the general specification of the reference case OM, but RECOGNISED the need for further work to identify and eliminate implausible models (notably the very high MSY scenarios). The "habitat approach" (Arrizabalaga et al) was proposed as one option.*
>
> ***This is the main focus of the current paper.**
>
> *52. The WPM NOTED there were similar issues with some extremely high MSY values estimated in the skipjack assessment. This was also influenced by the tagging data and was overcome by excluding some of the data from the small-scale tagging programmes. The yellowfin tuna assessment only included the RTTP tagging data, however, if enough data exist for the species from the small-scale tagging programmes then this might also be investigated.*
>
> *** Adding the small-scale tagging data to the RTTP data is potentially a non-trivial task, involving considerable data processing (e.g. tag age assignments and differential**

tag recovery proportion estimates based on tag seeding experiments, fleet behaviour and landing ports) and hampered by the loss of tagging staff from the IOTC secretariat. The authors have been advised that the secretariat is investigating options for re-analyzing the tagging data, in which case this point may be revisited in a future iteration.

53. The WPM DISCUSSED the use of alternative catch history scenarios for a robustness OM, however, no specific proposals were made.

*__This will not be addressed unless/until specific proposals are made.__

WPTT (2017) provided the following guidance for the next iteration of the yellowfin MSE:

233. The WPTT ***AGREED*** *on the general specification of the reference case OM as defined by the WPTT and WPM in 2016. Noting that it was difficult to specify explicit new scenarios outside of the context of a recent assessment, the following scenarios were suggested for further consideration in the OM robustness tests (with potential inclusion in the OM reference set, subject to review by WPM):*

- *Ricker stock recruitment curve.*

  *__This has not yet been done__
- *Recruitment shock (sustained poor recruitment consistent with the worst outcomes in the historical record).*

  *__This option has been added to the simulator, but not run.__
- *Alternative options for growth (among those considered plausible in recent YFT growth analyses).*

  * **Reconditioning would require a re-analysis of the tagging data, and will be deferred unless/until the secretariat progresses this. Temporal variability in biology for future projections has been added to the simulator, but not parameterized or tested.**
- *Alternative selectivity (e.g. dome-shaped vs: asymptotic, and region-specific).*

  *__This has not yet been done__
- *Alternative catchability increase scenarios (e.g. 3 or 5%).*

  *__This has not yet been done__
- *Explore options for temporal variability in biological parameters (e.g. natural mortality, growth, recruitment and migration) in relation to climate change. It was noted that these sorts of effects might not be important over the time-scale which an MP might be expected to operate without a thorough review (e.g., 5-10 years), and if they are important, they might undermine a lot of the stationary dynamics assumptions that underpin the modern fisheries assessment and management paradigm.*

  ***Temporal variability in biology for future projections has been added to the simulator, but this has not been parameterized or tested.**

*234. The WPTT **SUGGESTED** using a partially confounded design to increase the number of dimensions that could be included in the reference OM.*

***This was not undertaken as an explicit top down experimental design approach, however, the original balanced design approach was abandoned as described in the following.**

Two additional errors in the YFT MSE were identified and corrected since the WPTT and WPM 2017:

- The CPUE catchability trend scenarios used for OM conditioning are now implemented as multiplicative errors (rather than additive) - the new interpretation is consistent with the original intent (and is more optimistic).
- There was a bug between the last year of the assessment and the first year of the projections that greatly reduced fishing effort for three quarters (the fixed version is more pessimistic).

The expectation remains that MSE results for the TCMP will be provided in the standard format agreed by the WPM as implemented in FLR (Iago Mosqueira, pers. comm.).

4 Relationship between the stock assessment and Operating Models

As detailed in Kolody and Jumppanen (2016), 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. It would be difficult to justify the two initiatives evolving in different directions from the same scientific process. Accordingly, the yellowfin assessment of Langley (2016) provides the core of the OM conditioning process. Key features of the assessment and OM include:

- Parameter estimation with Stock Synthesis 3.24z software
- 4 regions (Figure 1)
- Quarterly dynamics, including recruitment and movement (implemented with calendar quarters as SS-model-years)
- 25 fisheries (21 with some temporal variation handled as independent fisheries)
- Parameter estimation objective function includes
    - Total catch
    - Standardized longline CPUE (one series per region)
    - 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 (various functional forms, parameters shared among some fleets)
    - Longline catchability (in aggregate - regional scaling factors are used to scale relative density to relative abundance among regions)
    - Virgin recruitment
    - Recruitment deviations from the Beverton-Holt stock-recruit relationship, recruitment spatial partitioning among tropical regions (1 and 4) and deviations from the mean spatial distribution.
    - Juvenile and adult movement rates

OM conditioning has an increased emphasis on uncertainty quantification and projections required to develop robust feedback-based MPs through the MSE process. The reference set OM is an ensemble of assessment models that includes several alternative plausible assumptions. The approach to uncertainty quantification adopted here is similar to that used in the CCSBT, in which the emphasis is on model structural uncertainty (including parameters about which the data are expected to be uninformative), and stochastic recruitment uncertainty (and observation error) in the projections. The Maximum Posterior Density Estimates (best point estimates) for the individual models are collated, with the expectation that this source of uncertainty will generally be greater than the parameter estimation uncertainty conditional on any individual model. Once an adequate OM has been defined, it should not need to be updated with the frequency expected for the traditional stock assessment process. Unless new evidence emerges to indicate that the uncertainty encompassed by the OM no longer captures reality, we would hope that an MP would remain valid for something on the order of 5-10 years (i.e. until the next thorough MP review scheduled as part of the adoption process).

Robustness OMs are generally considered less likely than the reference set, but they are defined to represent plausible, troublesome situations, that may help identify pathological MP behaviour in particular circumstances, and assist in choosing among MPs that are otherwise equivalent. An MP cannot be expected to be robust to every imaginable outcome (attempting to do so would likely result in an extremely conservative MP and considerable lost economic opportunity). Carl Walters famously uses the term "vampires in the basement" to describe serious and unanticipated events which undermine ecological models. Because these types of events are unavoidable, a normal part of the MP approach involves regular oversight (e.g. simple analyses to determine if "exceptional circumstances" have arisen which render the MP inappropriate, at least temporarily), and a scheduled review period, at which point a detailed assessment should determine if the MP testing remains valid, and whether there have been other changes in circumstance, e.g. changing Commission objectives, new assessment tools, etc.

For the purposes of this paper, we refer to a number of individual models, OM ensembles, and option abbreviations as defined in Table 1 and Table 2.

Figure 1. Spatial structure for yellowfin tuna assessment and all OMs discussed in this report (figure from Langley 2015).

Table 1. Model and ensemble definitions. In some cases these represent candidates for discussion - not all have been created or tested.

Model Name	Definition (assumption abbreviations are defined in Table 2)
SA-base	The base case assessment from Langley (2016). h80, M10, t10, q0, iH, x3, SS
OM-SA	The single OM specification that most closely resembles SA-base (identical except for no environmental movement link). h80, M10, t10, q0, iH, x3, SS
OM-ref1	Reference case OM as proposed by the WPM and WPTT in 2016, reviewed in WPM and WPTT 2017. Consists of an ensemble of 216 models, each differing from OM-SA-analogue in 1-6 assumptions. Undefined options as in OM-SA. Includes dubiously high productivity in many cases h70, h80, h90 M10, M08, M06 t00, t01, t10 q0, q1 iH, iC x3, x8
OM-rob-selTrend	A robustness OM consisting of 36 models, designed to look at the implications of temporal variability in selectivity, potentially resulting in a shifted preference toward younger ages. Undefined options as in OM-SA. Reviewed in WPM and WPTT 2017, and not found to add any new challenge for the MP. M10, M08, M06 t01, t10 x3, x8 SS, NS, ST
OM-rob-tagWt	A robustness OM consisting of 36 models, designed to look at the implications of tag-weighing λ options, notably the recommendation of λ = 1.5. Undefined options as in OM-SA. Reviewed in WPM and WPTT 2017, and λ = 1.5 was found to be very similar to λ = 1.0. M10, M08, M06 t00, t0001, t001, t01, t10, t15

	x3, x8
OM-rob-MSY	Robustness OM consisting of 48 models that compare different options for achieving plausible productivity (MSY) without the tagging data. h70, h80, h90 M10, M08, M06 t00 iH, i10H, iC, i10C SS, Sdev ESS5, ESSrW
OM-rob-LLsel	A robustness OM consisting of xxx models to look at some alternative LL selectivity options. M10, M08, M06 t00, t10 x3, x8 SS, S4, SSdev, Sspl
OM-rob-Ricker	A robustness OM consisting of xxx models to compare the functional form of the stock-recruit relationship (Beverton-Holt and Ricker). h70, h80, h90, Rh70, Rh80, Rh90 M10, M08, M06 t10 x3, x8
OM-rob-qTrend	A robustness OM consisting of xxx models to compare the impact of high longline catchability trends. h70, h80, h90, Rh70, Rh80, Rh90 M10, M08, M06 t10 x3, x8 q0,q1,q3,q5
OM-ref17.2grid	Unweighted combination 693 models from OM-17.2Tag and OM-17.2noTag (models with poor convergence removed).
OM-ref17.2	Revised reference case OM, proposed by the authors for feedback at the 2018 IOTC informal MSE meeting. It consists of ~300 SS

	specifications, randomly sampled from OM grids OM-17.2Tag and OM-17.2noTag. Sampling of the two grids is conducted to achieve a balanced combination of the following options as discussed in the text (representation of the other grid options is generally not proportional to the original grid assumptions): t10, t0001 q0, q1
OM-17.2Tag	Balanced OM grid ensemble of 288 models, with full weighting on the tagging data. This balanced grid is subsampled along with OM-r1noTag to form OM-ref2, as described in the text. Undefined options are as in OM-SA. h70, h80, h90 M10, M08, M06 t10 q0, q1 iH, iH10, iC, iC10 x3, x8 ess5, CLRW
OM-17.2noTag	Balanced OM grid ensemble of 432 models, excluding the tagging data (tags highly down-weighted λ=0.0001), subsampled along with OM-r1Tag to form OM-ref2, as described in the text. Undefined options are as in OM-SA. h70, h80, h90 M10, M08, M06 t0001 q0, q1 iH, iH10, iC, iC10 sr4, sr6, sr8 ess5, CLRW

Table 2. Model specification abbreviations. Bold indicates the assessment base case assumption. Some abbreviations may relate to explorations that are not reported.

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	σ_R = 0.4
sr6	**σ_R = 0.6**
sr8	σ_R = 0.8
	Future recruit failure
r55	3 years of poor recruitment (2019-21 proposed by the authors); deviation of -0.55 (consistent with SA-base estimates in the early 2000s), applied on top of the usual random deviate)
	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 = 0.001
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 (all CPUE error assumption)
iH	**Hooks Between Floats (σ_{CPUE} = 0.3)**
i10H	Hooks Between Floats (σ_{CPUE} = 0.1)
iC	Cluster analysis (σ_{CPUE} = 0.3)
i10C	Cluster analysis (σ_{CPUE} = 0.1)
	Tag mixing period
x3	**3 quarters**
x8	8 quarters

	Longline selectivity
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)
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**
CLRW	ESS = One iteration of re-weighting; the output ESS from OM-SA (mean over time for each fishery independently), capped at 100

5 OM-ref17.2: Revised YFT reference set OM

5.1 A proposed approach and specification for the YFT reference OM

The path for proposing a new reference set OM involved a circuitous exploration of many model assumptions and their interactions, in a number of grids, some of which are defined in Table 1 and Table 2. The main discussion below focuses on the proposed path forward, while some intermediate steps and analytical blind alleys are summarized for the record, in the following section. Our approach attempts to address two concerns:

1) Is the reference case OM suitably diverse and plausible to develop robust MPs?

2) Is the reference case OM suitably consistent with the current perception of the stock status within the IOTC community?

It is unfortunate that the OM conditioning is being undertaken somewhat independently of the assessment process. Ideally the two processes would happen jointly and inform one another in a consistent fashion. The stock assessment (Langley 2015, 2016) involved fitting a small number of models, mostly varying from one another with respect to a single assumption at a time. The process progressed until a "best" model was identified subject to the available time constraints. In the eyes of the assessment analyst and broader WPTT participants, it was considered sufficient with respect to a number of quantitative and qualitative diagnostics to formulate the management advice (despite some concerns). In contrast, the grid-based approach pursued for the OM results in hundreds (or even thousands) of models, including interactions among many assumptions. The OM models are difficult to inspect in detail, and difficult to communicate to a large group, but we would expect many of the individual models to be at least as valid and informative as the reference case assessment that was adopted The collective ensemble is aimed at providing a realistic representation of uncertainty, rather than best point estimates. For pragmatic reasons discussed below, and in the next section, we have opted to give the assessment a preferential position for guiding the structuring of the OM.

Unfortunately, fisheries data are generally not as informative as we would like, and assessment inferences are often very sensitive to seemingly minor and somewhat arbitrary assumptions. A large grid of models with alternative assumptions ensures considerable diversity in the sources of uncertainty, which helps to ensure that MPs should be robust under a range of circumstances. But many models that are not likely to be relevant to the real yellowfin situation also arise, and may provide misleading insights about expected MP performance.

The majority of OM-refY17.1 models (defined by the WPTT and WPM in 2016 without the benefit of actually seeing fitted model results) were of dubious plausibility. Notably, the models without tags tended to be very optimistic in terms of productivity. It would be possible to simply reject most or all of the models without full tag weighting. However, this is an unattractive solution for two reasons: i) there are concerns about tag assumption violations and the associated inferences, and ii) this would result in a much smaller set of models than the original 216. Furthermore, the CPUE and size composition data are not grossly incompatible with the tags, so we would not necessarily expect the tags to be required to obtain plausible models.

To address this first issue, we explored a number of additional assumptions and expanded the candidate reference case OM to include results from two balanced grids. Grid OMref17.2noTag (432 models) is intended to identify plausible models that are not dependent on the tags, and OMref17.2Tag (288 models) adds more variability to the models that include tags (see Table 1 and Table 2). This combination of 720 models was successful in both respects. However, the grid still needs to be refined because of a large number of implausible models. We propose the following procedure for obtaining the yellowfin OM reference model, OMrefY17.2:

1) Combine grids OMref17.2Tag and OMref17.2noTag. Removing models with dubious convergence (on the basis of the maximum gradient), results in an unbalanced combination of 693 models, OMref17.2grid. Some of the characteristics relevant to the filtering discussion are shown in Figure 2 (note the huge range in MSY estimates).

2) Sample (with replacement) the model set from OMref17.2grid to obtain a distribution that is consistent with key measures of central tendency from the assessment, though with arbitrarily inflated variance. As discussed in the following section, the criteria that we chose are:

- SSB(2016)/SSB(MSY) - This stock status indicator provides the basis around which IOTC yellowfin resolutions and MP tuning criteria have been crafted. The central tendency of various OM grids (filtered for plausibility) were usually not far off of the assessment estimates, so there seemed to be no reason not to maintain a high degree of consistency with assessment. However, this stock status indicator was not very useful for identifying implausible models.
- MSY - This was chosen as the main indicator of model plausibility because many of the models had production dynamics that seemed grossly incompatible with perceptions of the stock.
- Bivariate sampling (with replacement) was undertaken to obtain (approximately) lognormal distributions with a CV 3X the level estimated in the assessment (inverse Hessian SD estimates). This ensures that the OM is far more uncertain than the assessment, both in terms of the net stock status variance, but also the structural diversity represented. The sampling was set up to admit correlation between MSY and SSBY/SSBMSY, but the correlation was set to 0 to maximize the number of models represented in the final ensemble.

3) The sampling was stratified so that some dimensions of the final OM attain arbitrarily specified frequencies (using a balanced design). We opted for

- 50% of models with and 50% without tags (options t10, t0001)
- 50% of models with and without longline CPUE catchability trend (options q0 and q1)

This results in OMref17.2, which consists of (approximately) 300 models as shown in Figure 3. The mean and mode of the SSB/SSBMSY and MSY distributions were always within 1.5% and 2.5% of the SA-ref value (with 2000 samples). It is evident (and to be expected) that some models are sampled far more frequently than others (and about half are not sampled at all). Due to the coarseness of OM-ref17.2grid, the marginal distributions are somewhat polymodal. Note that the sampled distributions should have identical numbers of models with and without tags (and q0, q1). Sampling to obtain a desired distribution is not perfect, because the grid of candidate model characteristics is too coarse to smoothly cover all strata (and interactions among some variables may not result in plausible models).

Overall, we consider the approach to be reasonably successful. But we recognize that this may seem like a subjective and somewhat backward process for deriving the reference case OM. We would normally expect that the OM models provide useful inferences about stock status, not that we would be selecting models on the basis of our preconceived perceptions of stock status. But subjective decisions are an inevitable part of the modelling process, whether or not they are recognized. Our approach has some resemblance to the approaches suggested by Martell et al (2008), in which they propose re-framing the stock assessment question so that management quantities of interest (including MSY) are defined as leading variables in the assessment models, with priors. We note the following points for OMref17.2:

- Relative to OMref1 (the 2016 proposal), the expanded model options in OMref17.2grid add additional diversity and smooth out the OM stock status distributions. Notably, the ensemble identifies several plausible models that do not depend on the tagging data, and hence should introduce more variability to challenge the MPs. These models were all reasonably consistent with the CPUE and size composition data, but required further filtering/sampling (for numerical convergence and plausibility of dynamics).
- The bivariate OM sampling approach is a transparent admission that we are relying on the central tendencies of MSY and B/BMSY from the stock assessment process as explicit criteria for defining the OM. This seems to imply that one assessment model (despite some recognized shortcomings) provides more assessment insight than the hundreds of models explored for the OM. However, we would express the situation differently - both the assessment and the OM exploration indicated that the data are not as informative as we would hope, and were largely consistent with a large range of inferences. The OM did not provide obvious evidence for rejecting the point estimates of the assessment (and uncertainty in the assessment is always admitted to be problematic). By adopting key assessment inferences as an anchor, the proposed OM recognizes the collective wisdom of the IOTC WPTT assessment community (for better or worse), including their deliberations and subjective perceptions (e.g. that the yellowfin population is probably near full exploitation, and recent catches were probably near MSY).
- Explicitly sampling with respect to SSB/SSBMSY maintains a level of consistency with the assessment reference points, and tuning objectives defined for MP performance evaluation and eventual selection.
- MSY-based sampling addresses one of the most obvious sources of model implausibility in the OM grids (unrealistic MSY and the related issue of production dynamics that are not consistent with standard assumptions of tuna recruitment compensation, and/or stationarity in the stock recruit relationship).
- OMref17.2 admits far more uncertainty than the assessment, both in terms of the magnitude of the stock status variance, and the structural diversity introduced through alternative assumptions. By defining the uncertainty relative to the assessment it provides a convenient framework for communication and reproducibility. e.g. If the WPTT/WPM agrees that a certain CV is appropriate, it can be reproduced despite other changes in the

OM grid that might be requested in parallel, and which could skew the central tendency in unexpected ways. It remains a topic for broader discussion as to whether we have "enough" (or too much) uncertainty within OMref17.2, but this can be easily adjusted using the current approach. By coincidence, the OMref17.2 CVs were very similar to the reported BET assessment CVs (which were derived from a small grid of models).

- The sampling approach allows a limited number of dimensions of the grid to be sampled in pre-specified proportions. However, this is not perfect, and can only be achieved with a relatively small subset of dimensions (because of the potentially incompatible interactions among some assumptions). We proposed that inclusion/exclusion of tags and CPUE catchability trends are the most important priorities for equal weighting. Steepness might be considered another priority, but this distribution was relatively evenly represented already (in aggregate, not necessarily with respect to the tag, LL catchability or other dimensions in the grid).

Additional characteristics of OMref17.2 are shown in Figure 4 - Figure 9.

Figure 10 shows the projected dynamics of OMref17.2 when the fishery is shut down. The central tendency of the projected biomass recovers to the levels from the early history of the fishery, while the variance is substantially higher. Qualitatively, this is to be expected because the early history represents different interpretations of a single historical realization, while the future represents 200 different realizations. Figure 11 shows the projected dynamics of OMref17.2 with constant current catch projections. The assessment K2MSM reports P(SB2018 < SBMSY) = 88% and P(B2025 < BMSY) = 100% (Table 3), while the OMref17.2 indicates 50% <P(SB2018 < SBMSY) < 75% for both dates. Qualitatively, this is consistent with what one would expect if the OM has higher variability (as intended). Projections for both the assessment and the OM run into numerical limits in this case, due to the very high exploitation rates required to sustain these catches.

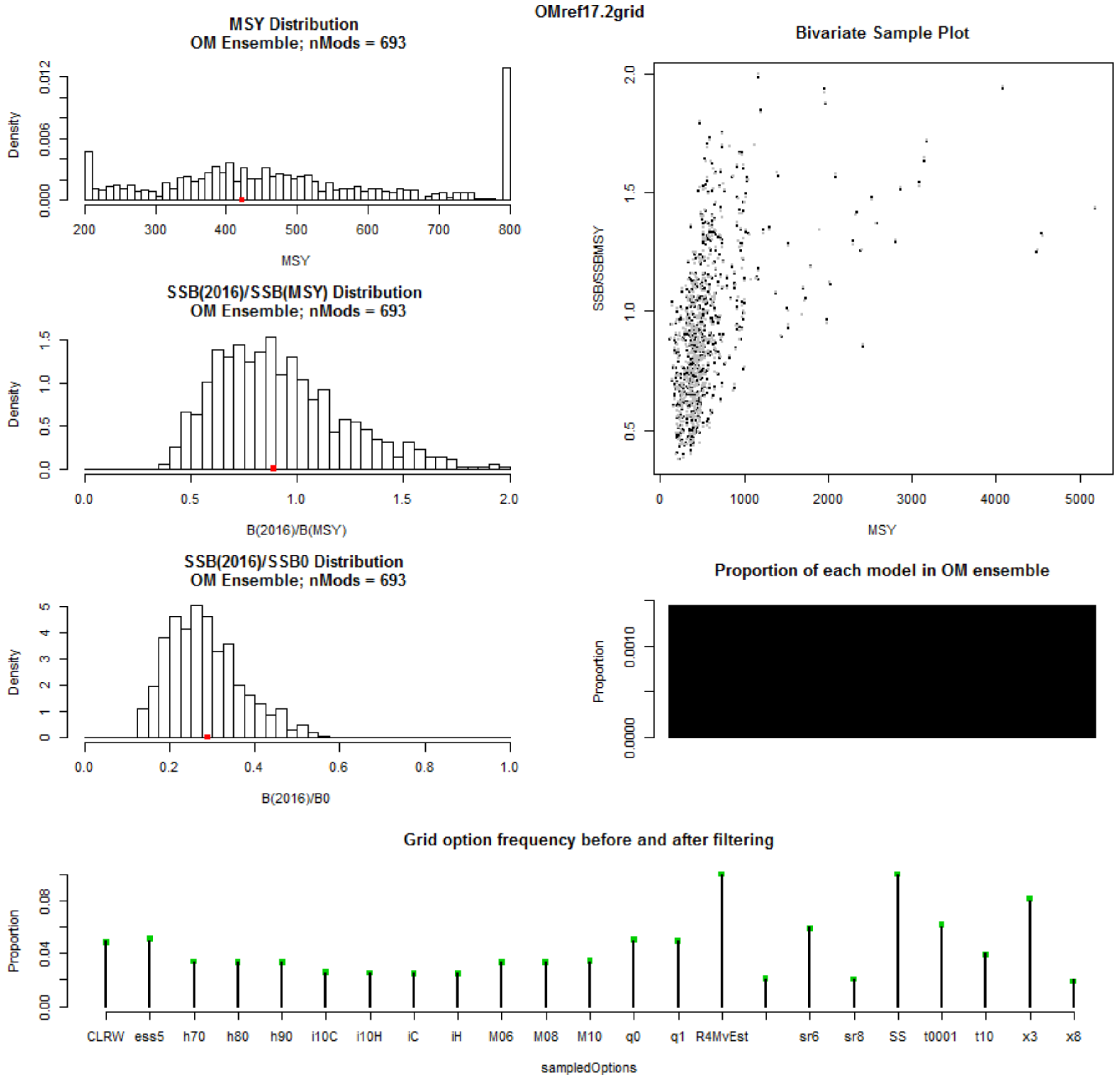


Figure 2. Characteristics of OMref17.2grid, the unweighted grid from which OMrefY17.2 is sampled. Red points indicate the point estimates from the 2016 assessment. The middle right panel indicates the relative frequency of the models sampled (i.e. uniform sampling in this case). The bottom panel indicates the relative proportion of the individual assumptions in the ensemble (green points) relative to the original grid (black lines), i.e. identical except for removal of the small number of models with numerical convergence problems. MSY values < 200 Kt and > 800Kt are aggregated at the bounds of the histogram.

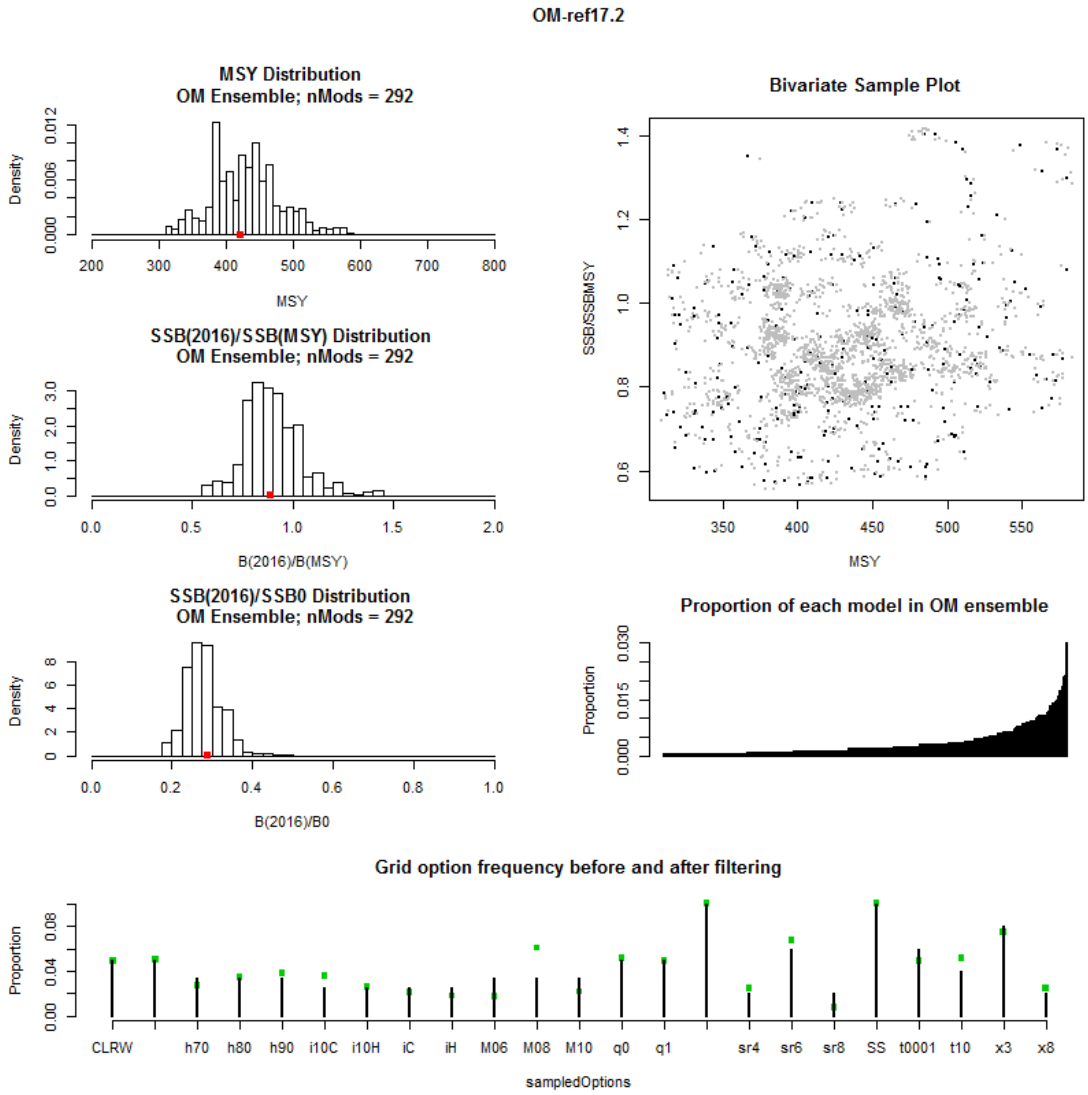


Figure 3. Characteristics of OMrefY17.2, the proposed reference set OM using the bivariate sampling. Red points indicate the point estimates from the 2016 assessment. The top right panel indicates the relationship between MSY and SSBY/SSBMSY (grey points are jitters to emphasize repeat sampling frequency). The middle right panel indicates the relative frequency of the models sampled. The bottom panel indicates the relative proportion of the individual assumptions in the ensemble (green points) relative to the original grid (black lines).

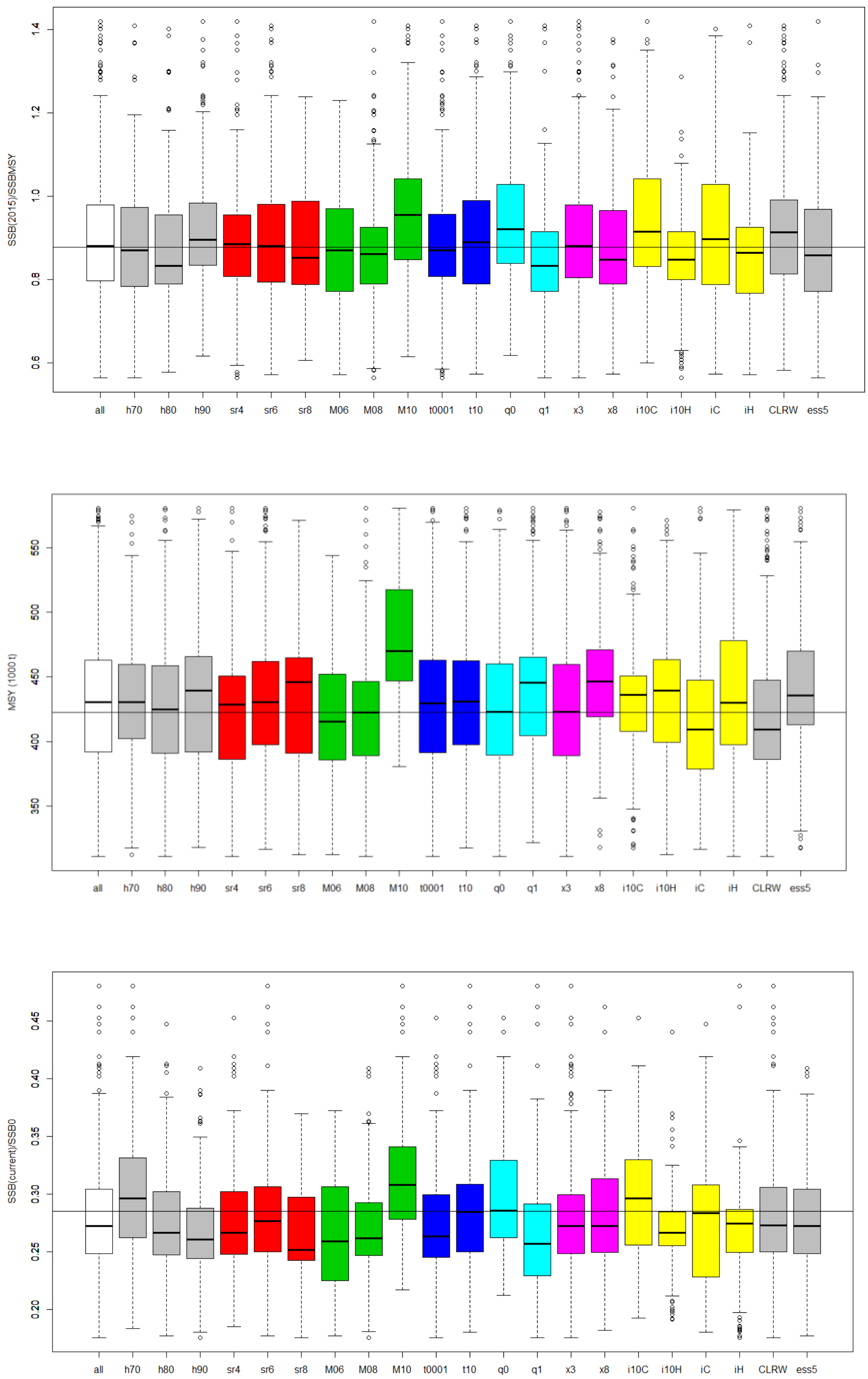


Figure 4. OM-ref17.2 stock status reference points, partitioned by assumptions.

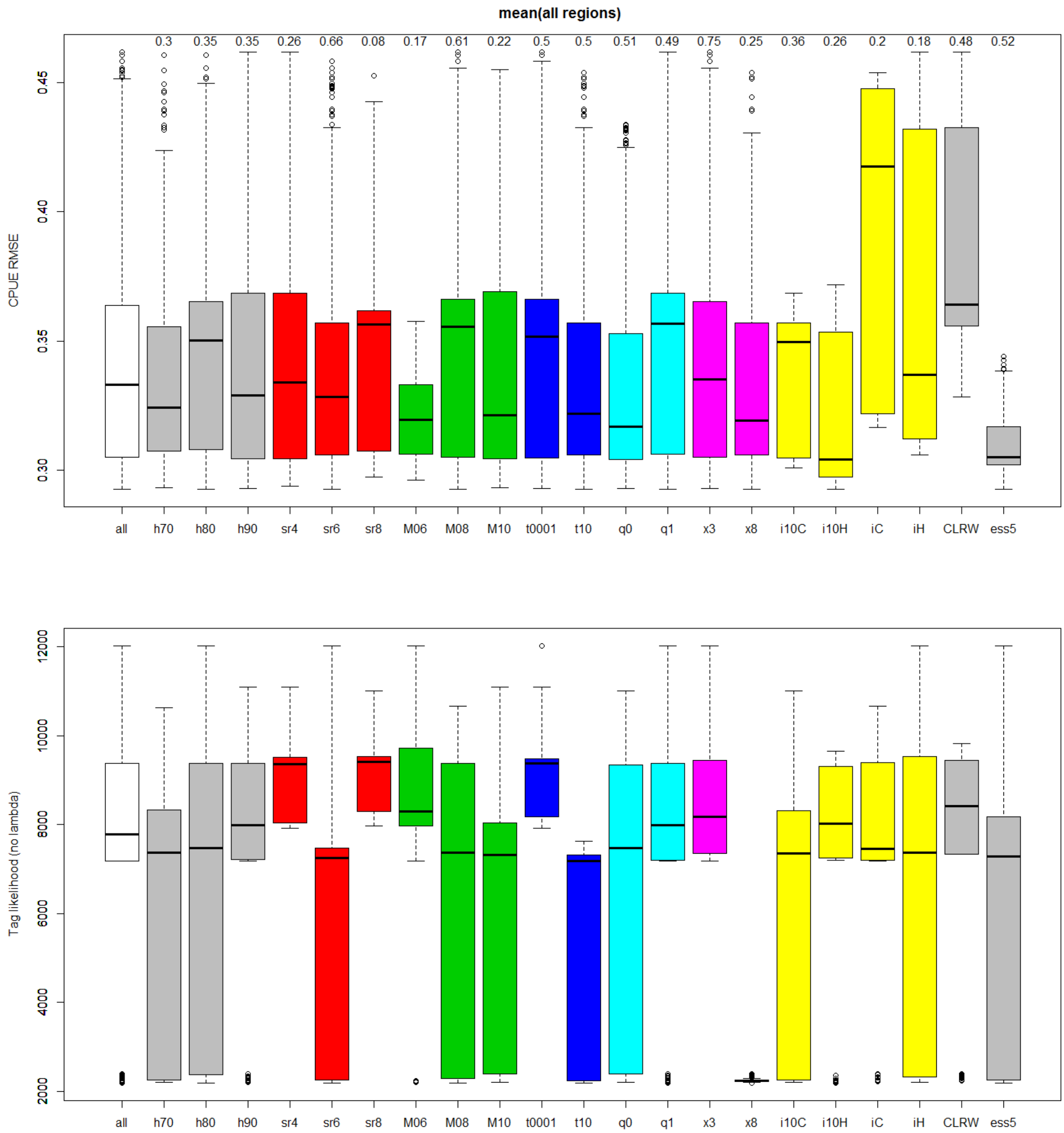


Figure 5. OM-ref17.2 quality of fit to the CPUE (top panel, RMSE, mean over regions) and tag (bottom panel, likelihoods before λ weighting). Note that many of these summary statistics are not easy to compare because there are different data in the models, i.e. HBF- or cluster-based CPUE, and the short/long tag mixing periods.

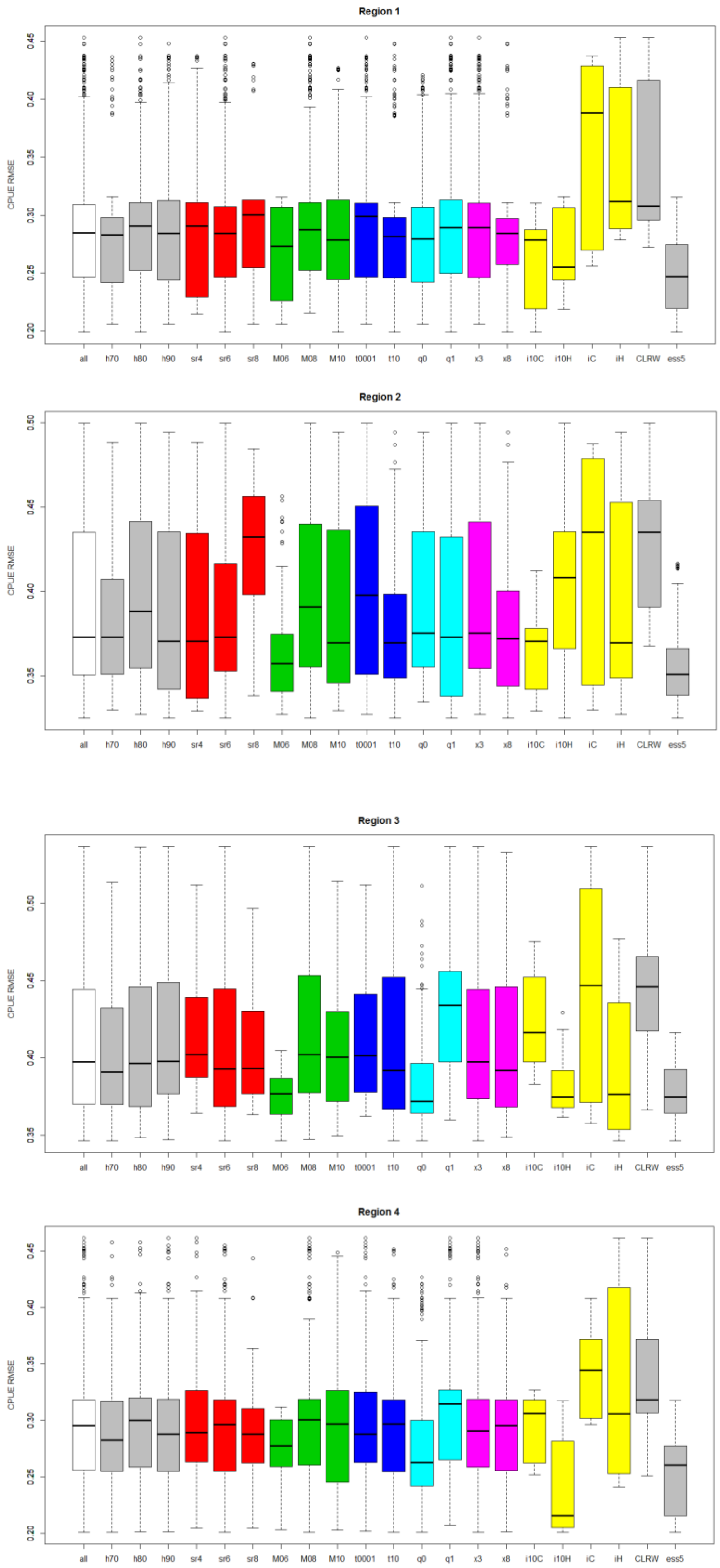


Figure 6. OM-ref17.2 fit to the CPUE (RMSE) by region, partitioned by assumption. Note that two fundamentally different CPUE series are used in the tropical regions.

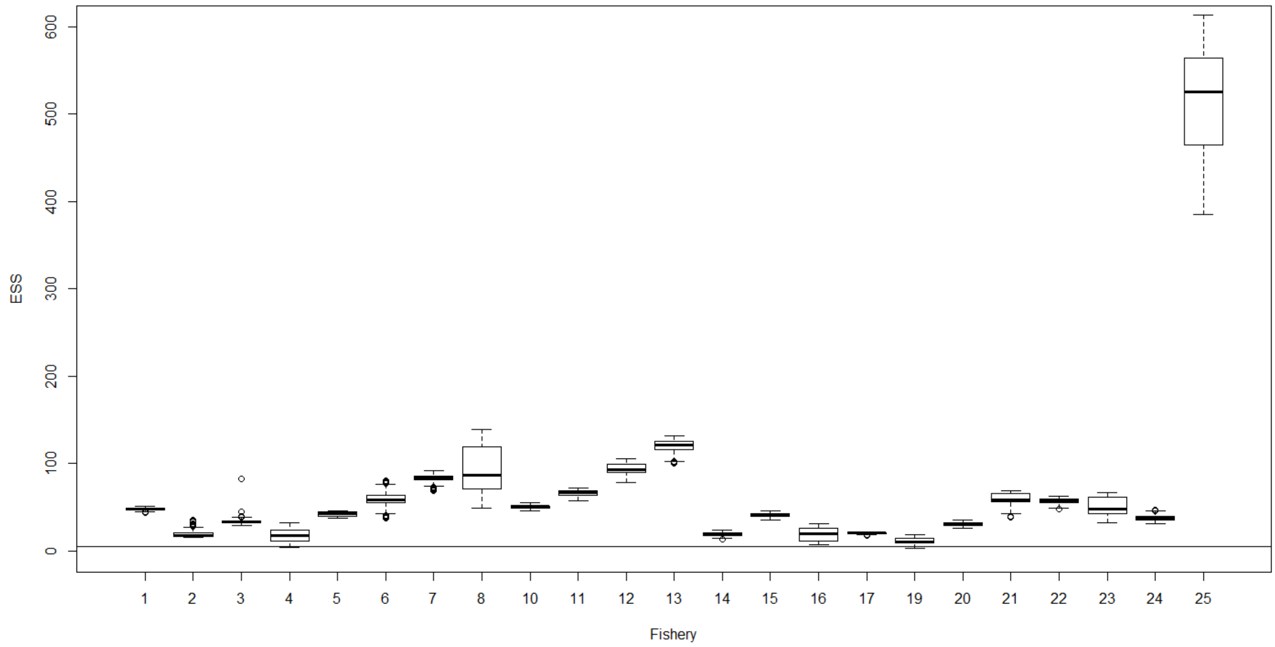


Figure 7. OM-ref17.2 quality of fit to the CPUE series (post-fit effective sample size), partitioned by fishery, but pooled over all model assumptions. The reference line (5) is the assumption in the assessment.

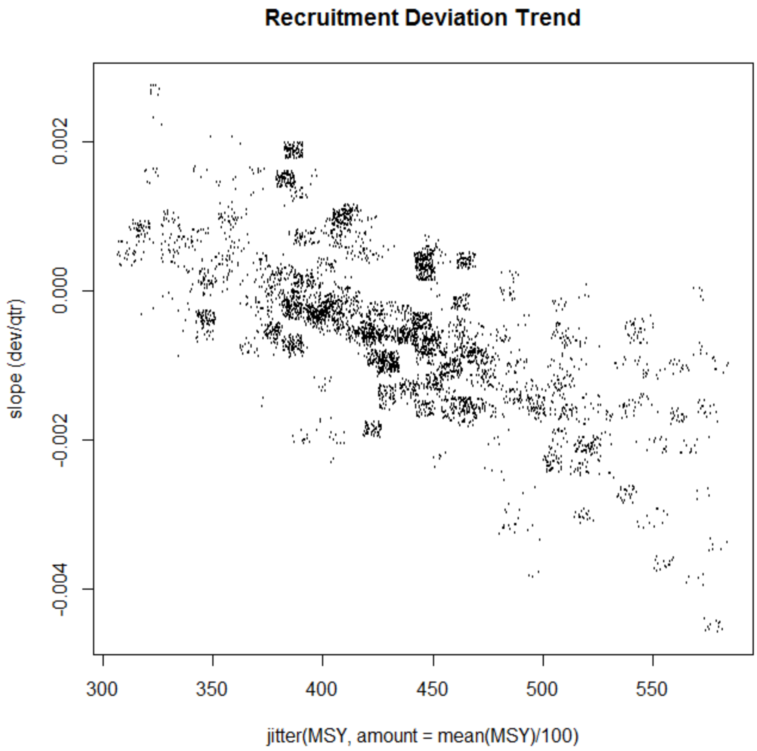


Figure 8. OM-ref17.2 relationship between MSY and the recruitment deviation trend (jittered).

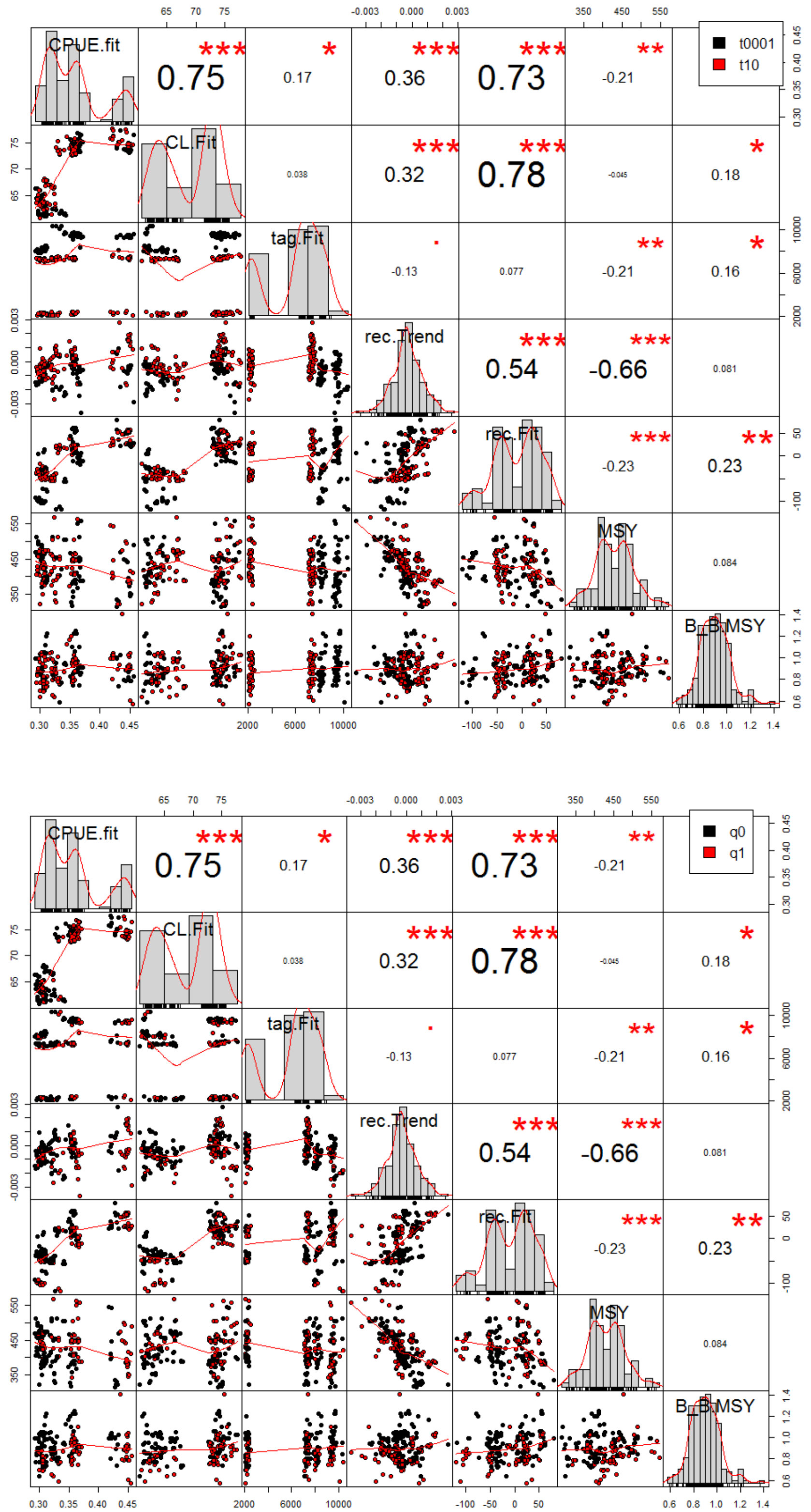


Figure 9. OMref17-2 correlations among (jittered subsample) stock status and quality of fit indicators. Points are partitioned by inclusion or exclusion of tags (top panel) or LL catchability assumption (bottom).

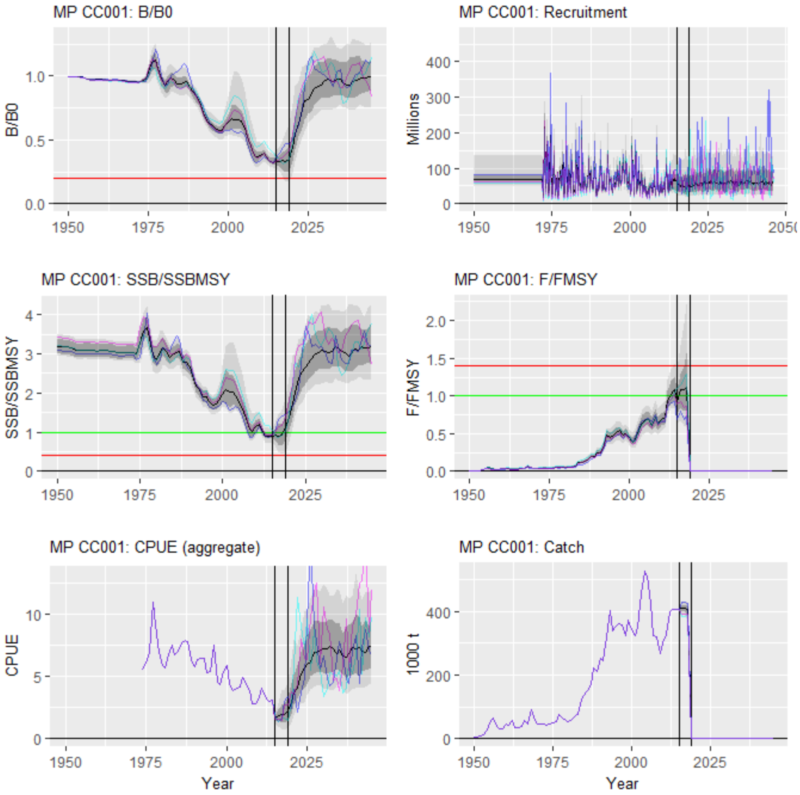


Figure 10. OMref17.2 projections with a fishery moratorium (starting in 2019).

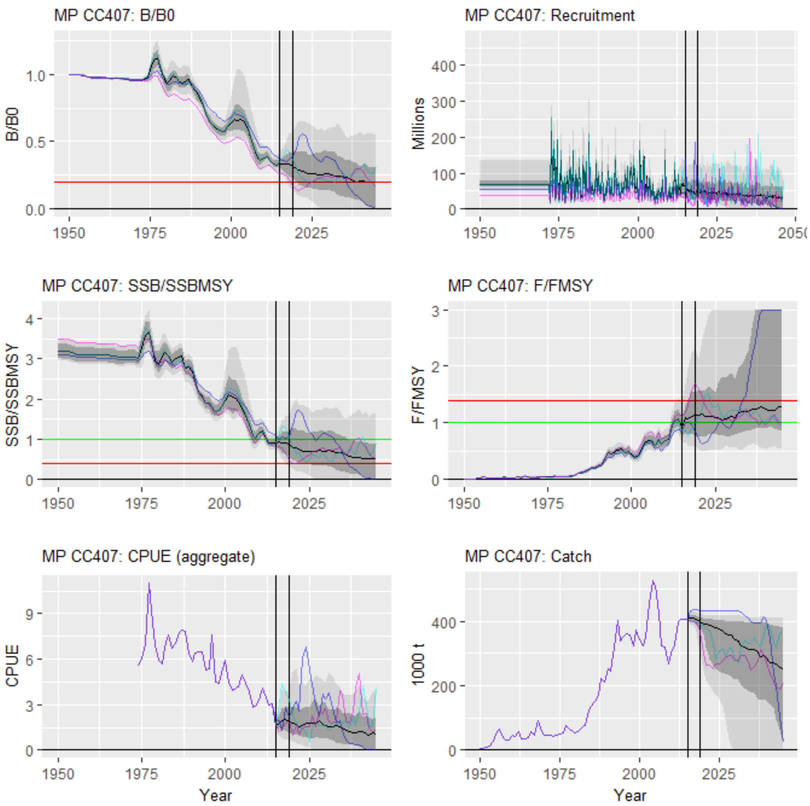


Figure 11. OMref17.2 projections with constant current catch (407 Kt).

Table 3. Summary of projections from the 2016 YFT assessment (from the WPTT 2016 report).

IOTC–2016–WPTT18–R[E]

TABLE 2. Yellowfin tuna: Stock synthesis assessment Kobe II Strategy Matrix. Probability (percentage) of violating the MSY-based target (top) and limit (bottom) reference points for constant catch projections (relative to the catch level from 2015 (407,575t), -30%, - 25%, ± 20%, -15%,± 10%, -5%), projected for 3 and 10 years), projected for 3 and 10 years.

Reference point and projection timeframe	Alternative catch projections (relative to the catch level from 2015) and probability (%) of violating MSY-based target reference points ($B_{targ} = B_{MSY}$; $F_{targ} = F_{MSY}$)								
	70% (285,302t)	75% (305,680t)	80% (326,059t)	85% (346,438t)	90% (366,816t)	95% (387,195t)	100% (407,574t)	110% (448,331t)	120% (489,089t)
$B_{2018} < B_{MSY}$	53	61	67	77	80	88	88	97	99
$F_{2018} > F_{MSY}$	2	7	23	47	65	73	100	100	100
$B_{2025} < B_{MSY}$	6	n.a.	20	37	60	100	100	100	100
$F_{2025} > F_{MSY}$	0	n.a.	10	40	57	100	100	100	100
Reference point and projection timeframe	**Alternative catch projections (relative to the catch level from 2015) and probability (%) of violating MSY-based limit reference points ($B_{lim} = 0.4\ B_{MSY}$; $F_{Lim} = 1.4\ F_{MSY}$)**								
	70% (285,302t)	75% (305,680t)	80% (326,059t)	85% (346,438t)	90% (366,816t)	95% (387,195t)	100% (407,574t)	110% (448,331t)	120% (489,089t)
$B_{2018} < B_{Lim}$	2	1	2	4	6	6	12	21	38
$F_{2018} > F_{Lim}$	0	0	1	10	32	52	100	100	100
$B_{2025} < B_{Lim}$	0	n.a.	1	7	30	>30*	>30*	>30*	>30*
$F_{2025} > F_{Lim}$	0	n.a.	0	11	53	>30*	>30*	>30*	>30*

* *At least one fishery not able to take the catch due to absence of vulnerable fish in the projection period. The probability levels are not well determined, but likely progressively exceed 30% as the catch level increases beyond 90%.*

5.2 Reference Case OM exploratory back-story

The approach described above evolved as a consequence of considering a number of other model options and selection approaches. We describe some of them briefly for the record below.

Quality of agreement between model predictions and data

Within the yellowfin model space explored to date, there is not much obvious variability in terms of the fit to the size composition data and CPUE series among models. Evidently, these data are not sufficiently informative to constrain the model parameters to a plausible space. This is admittedly a general qualitative conclusion based on visual inspections of quality of fit summaries. It is difficult to make this argument in terms of rigorous statistical principles, because we recognize that these models are not being formulated in a way that makes them strictly compatible with statistical theory (e.g. penalized likelihoods are being used rather than true likelihoods, different data are used in different models, distribution sample sizes are artificially manipulated).

In contrast, the tags tend to constrain the models in a way that is consistent with our prior expectations (e.g. the perception that the stock is probably near fully exploited and recent catches are probably near MSY), such that the tags did not support very high MSY scenarios in OM-ref1.

The tags provide information about absolute biomass which the other data do not. However, there are also non-trivial concerns about the tagging data:

- tag mixing is clearly not fast, and may never be achieved at the scale of the assessment spatial structure, hence substantial biases from tag estimators can be expected.
- tagging data tend to prefer M estimates that are lower than what are generally assumed to be likely for YFT (this may be a result of poor mixing, but could also suggest that perceptions of M are biased)
- As currently implemented in SS3, variability in tag reporting rates (notably the unknown reporting rate subsequent to the tag seeding experiments), cannot be properly represented.

Given the concerns about the tagging data, it is not clear how the fit to the tagging data should be considered as a plausibility criterion. Not surprisingly, the fit is better with higher tag weighting, and we consider these models to be plausible. But in OM-ref17.2 we considered it important to also include plausible models that were not dependent on the tags. This was found to be achievable in various ways, including increased weighting on the CPUE, iterative reweighting of the CL data, longline temporal variability in selectivity and/or decreased σ_R. Two balanced grids were combined in OMref-17.2grid, to ensure that a substantial pool of candidate models was available, with and without tags. However OMref-17.2Tag and OMref-17.2noTag were not equally balanced across all assumptions.

Habitat Constraints

WPM (2017) suggested that arguments about habitat constraints may be useful to identify and eliminate OM scenarios that are implausible on the basis of inconsistency with other assessments for the same species. While a variation of this approach was used for Indian Ocean albacore, we are reluctant to apply it to Indian Ocean yellowfin for a number of reasons:

- There are only three other yellowfin populations against which to compare, and there is no *a priori* reason to expect that the assessments for those other species are more accurate than the IO assessment.
- There is already some sharing of information and ideas across tRFMOs (e.g. M, steepness) that limits the independence of assessments and may result in self re-enforcing arguments.
- This approach would probably be effective for removing a handful of outlier results, but it is unlikely to provide sufficient subtlety for filtering/weighting models across a continuum from plausible to implausible, which seems to be present for IO yellowfin.

Given these concerns, we would prefer to identify model selection criteria from within the Indian Ocean yellowfin system, if possible.

Recruitment time series plausibility and MSY

The relationship between MSY and recruitment deviation trends represents an axis of model performance that appears to smoothly span the continuum in both quantities from what we consider to be plausible to implausible (e.g. Figure 12, bottom left panel). The trade-off represents a common problem in age-structured assessments - an abundance pattern (usually decline) can often be largely explained by either the recruitment time series or fishery depletion (and a mixture of the two). With truly stationary selectivity and good age sampling (and/or or reliable tag estimators and M), it may be possible to disentangle the two effects in principle. Unfortunately, with length-based sampling as an indirect proxy for ages, and a plausible case for non-stationary selectivity (i.e. diverse temporal and spatial patterns in fishing, changing targeting and gear configurations) and/or biased CL sampling, disentanglement may not be possible.

Very high MSY is associated with a negative recruitment deviation trend (as observed in OM-ref1), while Figure 12 shows that this relationship can also be extended in the other direction, with very low MSY weakly associated with positive recruitment deviation trends. The relationship between these quantities and the magnitude of the recruitment deviations is less clear.

While a trend in recruitment deviations is certainly possible (e.g. due to environmental changes), a large trend is not very compatible with the stationary production dynamics paradigm that prevails in most single species stock assessment and fisheries management theory. And these trends can arise as modelling artefacts. In some of the cases observed here, the trend could be interpreted as evidence of a systematic lack of fit to the stock recruit relationship. Figure 13 and Figure 14 show an example high-MSY model, that has a declining recruitment deviation trend. Estimating steepness removes the problem of the recruitment deviation trend, but introduces a different problem, in that the estimated value of steepness (h = 0.21) is also considered implausible for tunas (and the corresponding MSY estimate is dubiously low because it is less than half of the catches obtained over the past 15 years). We know from simulations that steepness estimates can be poor, even when models are structured more or less correctly for the simulator.

On the basis of the preceding arguments, we explored different options for weighting a grid of OMs to produce a reference set of predominantly plausible models, including:

1) Weight each model according to the consistency of the MPD recruitment deviation RMSE with the assumed recruitment CV (and the expected CV distribution determined from simulations).
2) Weight each model inversely to the magnitude of the recruitment deviation trends. The model weighting factor for the recruitment deviation trend was calculated (through simulations) as the probability of observing the deviation trend given the length of the observed time series, $\sigma_R = 0.6$, and auto-correlation rho = (0, 0.25, 0.5). The MDP recruitment auto-correlation (calculated quarterly) in SA-ref was 0.23. This approach is of course very simplistic, but with 0 < rho < 0.25, the sampled OM MSY distributions appear roughly log-normally distributed, with a mode near the SA-ref value and thinned tails.
3) Weight the models directly according to the MSY plausibility as defined in the assessment.

Option 1 was unsuccessful at eliminating many models with implausible MSY. Option 2 was reasonably successful at removing the very high MSY models, but not very successful at removing the very low MSY models. Option 3 was the most successful. This is not surprising - if MSY is considered the most important indicator of model plausibility, it makes sense to openly admit that it is the criterion that should be used to filter models. However, even option 3 was not very satisfying for generating the OM in the sense that the filtered grid is the product of the frequency of the models in the unweighted grid and the weighting of the plausibility criteria. Since different grids have very different initial frequencies (e.g. depending on whether or not tags are included), the filtered distributions are also potentially very different (and subject to the biases and whims of the analyst, and counter-intuitive model interactions). Thus using any filtering criteria in this manner has the potential to result in very different stock status perceptions in the OM.

These explorations of model uncertainty and sensitivity lead us to conclude that the key inferences from the stock assessment could be embraced as a loose scaffold around which to structure OM uncertainty. Rather than weighting a pre-existing grid on the basis of plausibility criteria, we can reverse the process and sample the grid to recreate a pre-defined distribution.

Since SSB/SSBMSY is the critical reference point around which YFT rebuilding resolutions have been proposed, and MP tuning objectives set, we propose that this should be the assessment characteristic around which the OM is structured. If we only sample OM-ref17.2grid to match the SSB/SSBMSY distribution of the assessment (MPD and CV in log-space), we get the results shown in Figure 15. This is clearly not very satisfactory because it fails to address the problem of models that are implausible with respect to MSY.

If we sample OM-ref17.2grid so that the MSY distribution matches the MSY of the assessment, we get the results shown in Figure 16. In this case, MSY is very consistent with the assessment, but this results in a SSB/SSBMSY distribution that is very broad and somewhat less optimistic than the assessment (mean = 0.81 compared to 0.89 MPD for the assessment). A similar approach was attempted by weighting with respect to an expected distribution of recruitment deviation trends (Figure 17), but it was not sufficiently restrictive to remove the implausibly low MSY results.

Recognizing that i) MSY appears to be a convenient and essential (though not sufficient) criterion for evaluating model plausibility, and ii) it is easy and desirable to maintain consistency with the assessment with respect to the central tendency of the SSB/SSBMSY reference point, we arrived at the bivariate sampling approach adopted in the preceding section. However, we recognize that the specification of variances remains tricky and somewhat arbitrary. In the interest of robustness, we want to challenge the MPs to conditions that are considerably more difficult than a single assessment specification. For OM-ref17.2, we assumed that the MSY and SSBY/SSBMSY CVs are 3X higher than the assessment estimates (plus the uncertainty in the dynamics is fundamentally more diverse, because of the structural differences in the ensemble of models). We would hope that MP selection would be reasonably robust to the uncertainty assumptions when tuned to median performance (e.g. "attain target on average"), but would expect them to be considerably more sensitive to higher percentile tuning objectives (e.g. "high probability of P(Kobe green)". Figure 18 and Figure 19 show alternative OMs in which the MSY and BY/BMSY CVs are equal to and double the assessment estimates (OM-ref17.2 CVs are tripled).

Sampling approach

Sampling the OM grids to achieve a desired result was an approximate process. The observed distribution (univariate or bivariate) from the original OM grids was stratified into bins (of usually 1.0 standard deviations width). Each bin was sampled assuming that each model within the bin was equally likely. e.g. in the univariate case, 34% of the OM realizations are randomly sampled (in log-space) from the grid in which the MSY is within 0 and +1 σ_{MSY}, and 14% of the OM realizations are sampled from within +1 and +2 σ_{MSY}, etc. Sampling was truncated at +/- 3 σ. In principle the sampling can be made as accurate as desired, by decreasing the bin widths. However, in practice, the coarseness of the original grid limits the effectiveness of the sampling. The sampling accuracy is further degraded if one attempts to impose additional stratification (e.g. imposing a requirement for 50/50 split of models with and without tags is equivalent to halving the number of models in the grid from which to sample).

The CV of SSBY/SSBMSY from the assessment is not output by SS3, and was calculated on the basis of the reported SDs for SSBY and SSBMSY which are reported. We assumed a multivariate normal distribution with correlation of ρ = 0.59 (an approximation calculated from the empirical relationship observed for the MPD estimates from an ensemble of models that was obtained from a subset of models with univariate sampling for MSY only). The bivariate sampling procedure can also admit the correlation between MSY and SSBY/SSBMSY, but this was assumed to be 0.

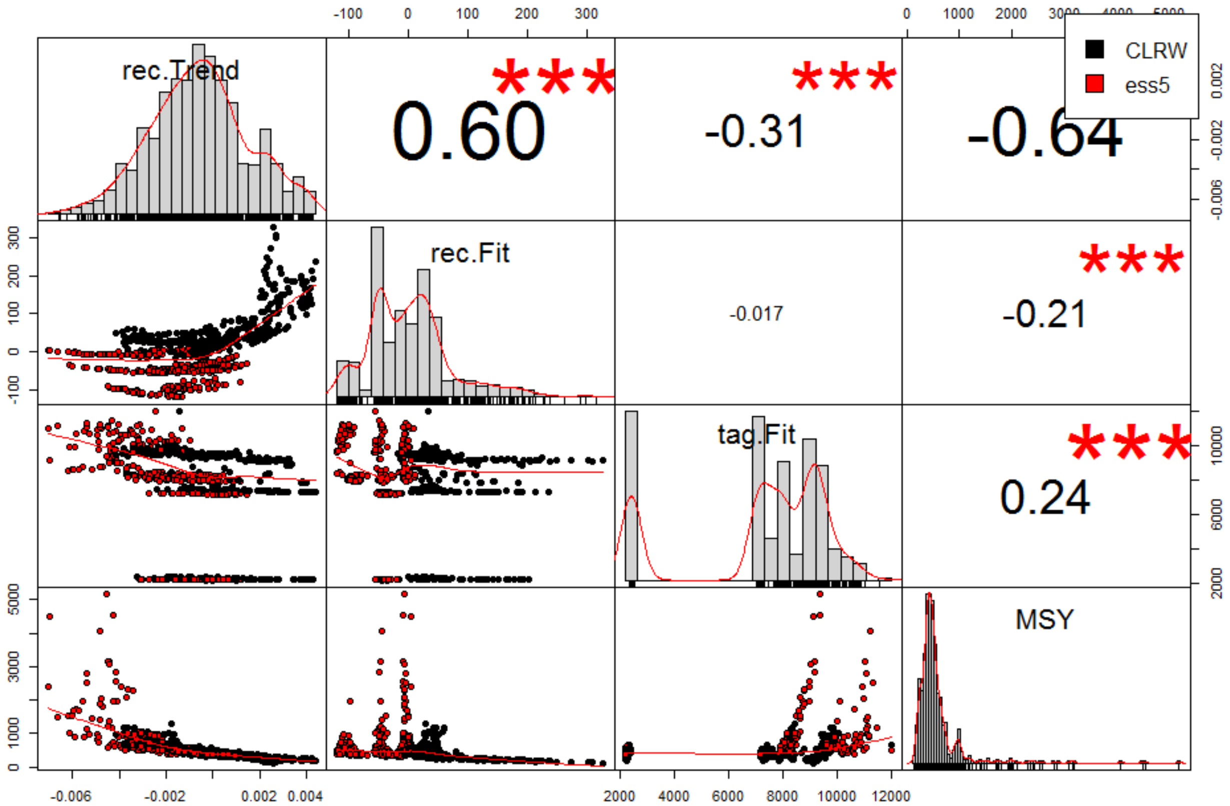


Figure 12. Relationships among MSY, recruitment RMSE, and recruitment deviation trend in the combined grid OM-refTag and OM-ref17.2noTag.

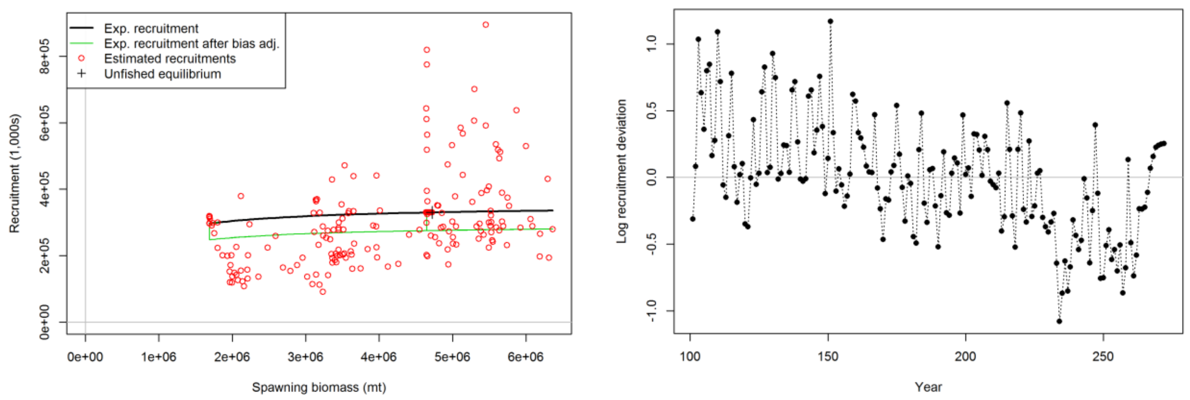


Figure 13. Stock and recruitment for one of the high MSY models from OM-ref1 illustrating the declining trend in recruitment deviations (steepness fixed at h = 0.8, MSY ~800K t).

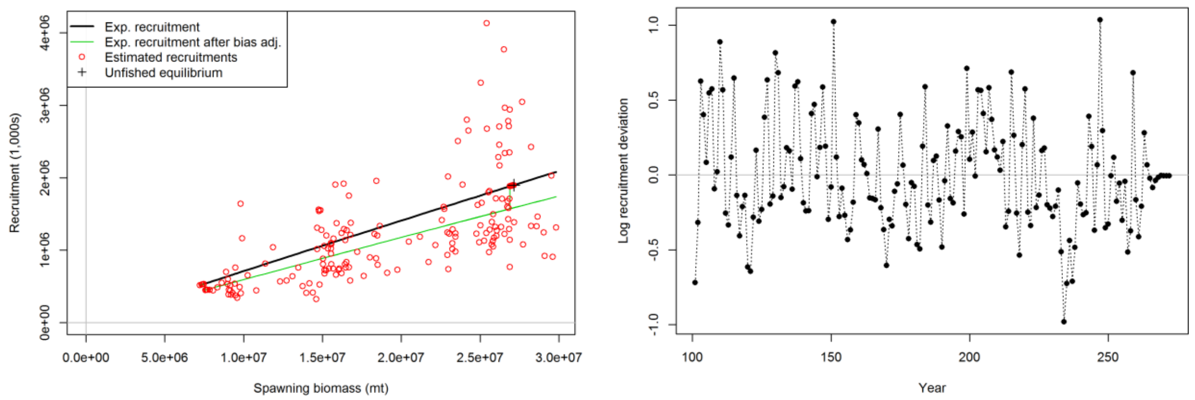


Figure 14. Stock and recruitment for the same model from Figure 13, except with steepness estimated (h=0.21, MSY ~100K t).

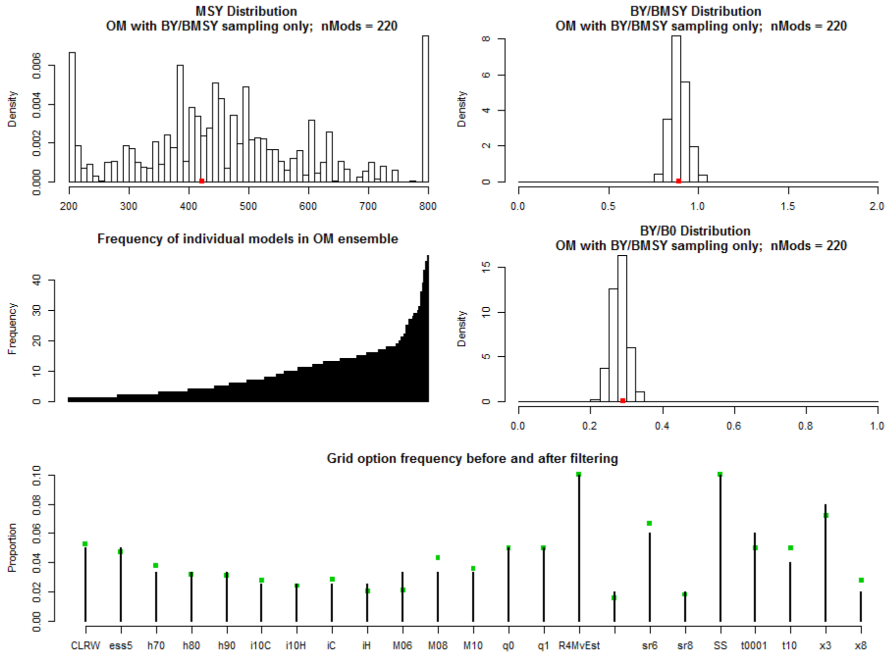


Figure 15. A potential OM based on only sampling OM-ref17.2grid with respect to SSB/SSBMSY.

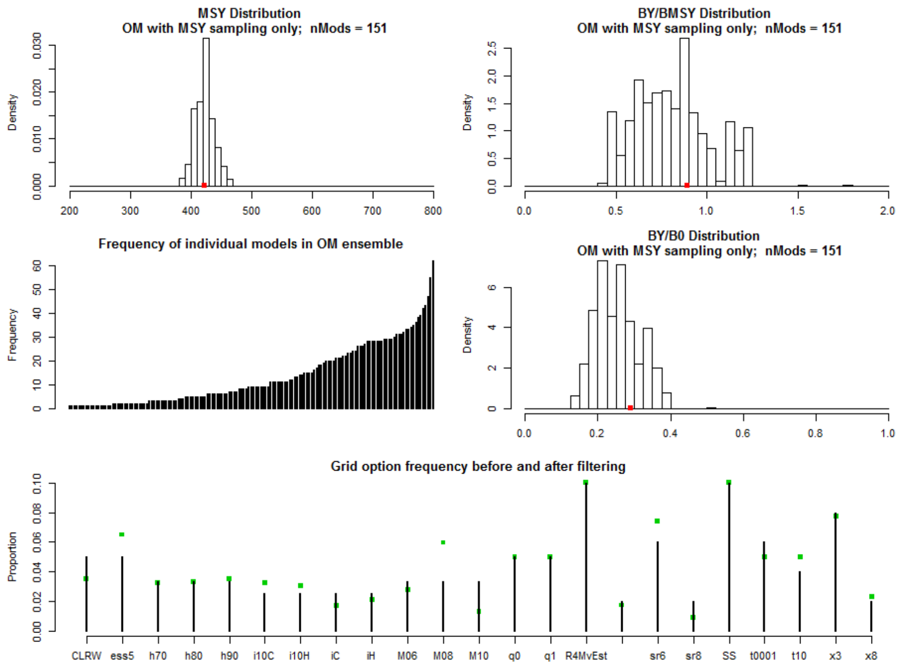


Figure 16. A potential OM based on only sampling OM-ref17.2grid with respect to MSY.

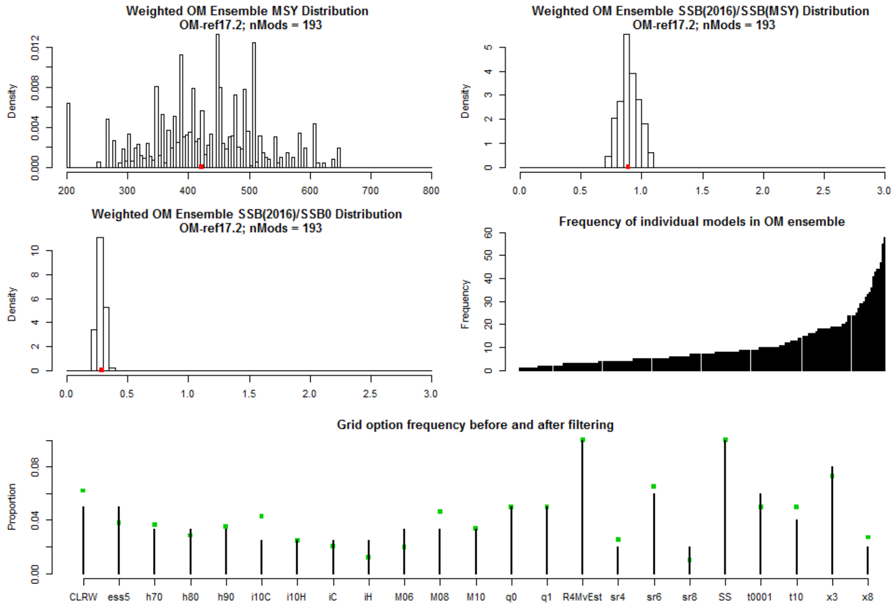


Figure 17. A potential OM based on only sampling OM-ref17.2grid with respect to recruitment deviation trends and SSB/SSBMSY.

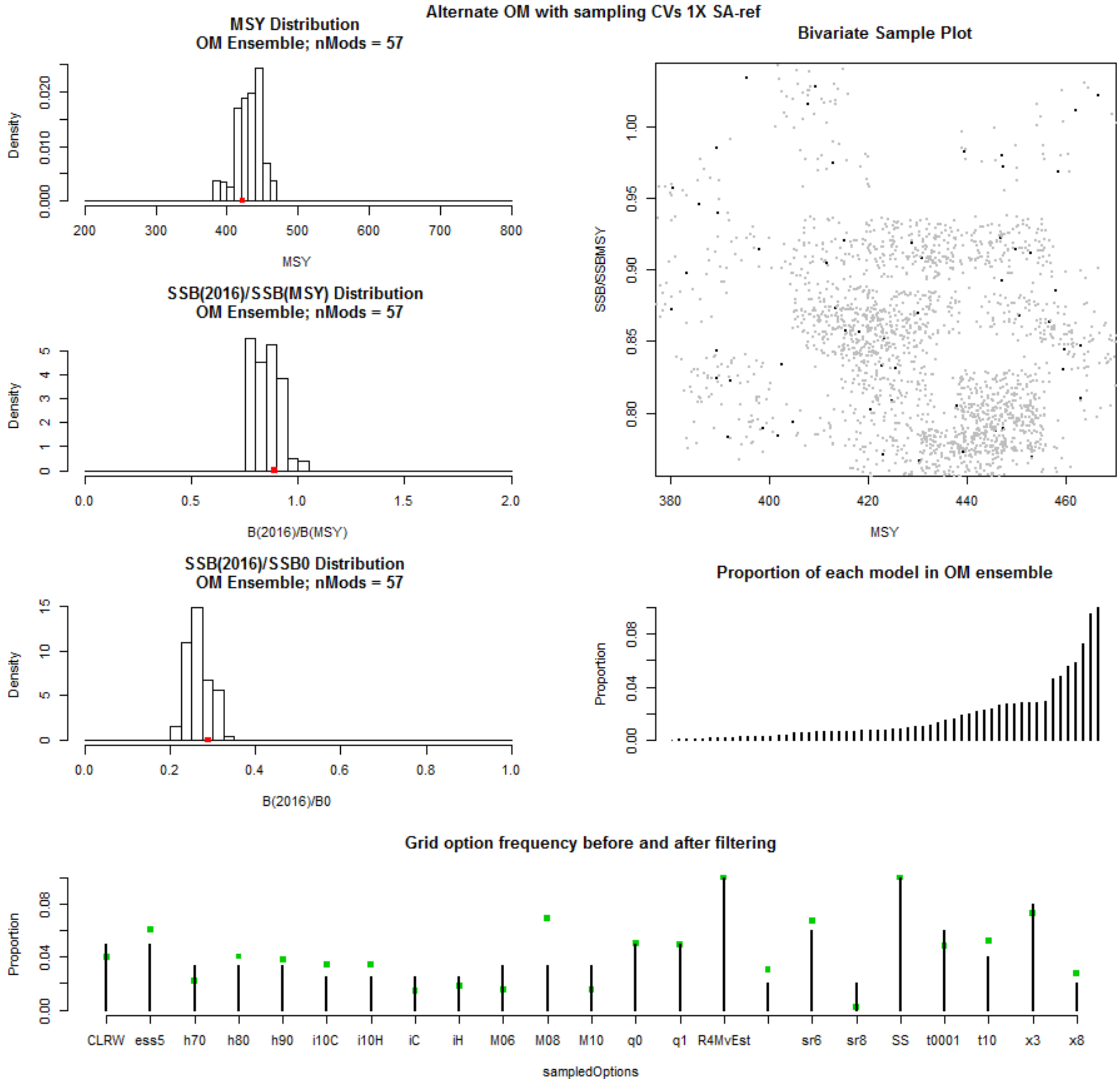


Figure 18. A potential alternative to OM-ref17.2, in which the sampling of the MSY and SSBY/SSBMSY distributions with a CV equal to the reference case assessment.

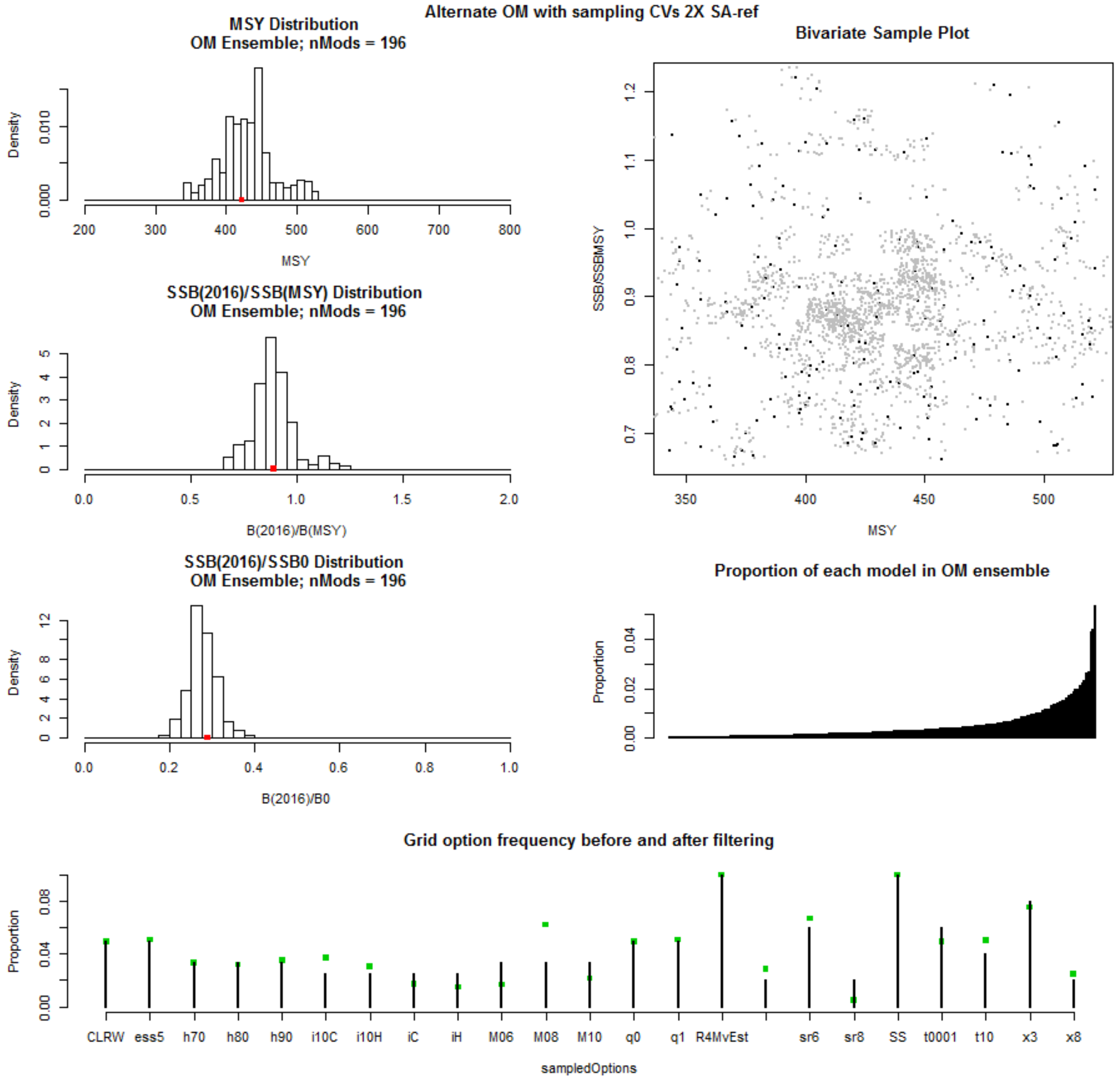


Figure 19. A potential alternative to OM-ref17.2, in which the sampling of the MSY and SSBY/SSBMSY distributions with a CV double to the reference case assessment.

6 MP performance with OMref17.2 and the TCMP tuning objectives

Results from candidate MPs defined in Table 4 are presented. See Kolody and Jumppanen (2016) for the full specification of these MPs, noting that the PT4010F option was added as described in Kolody and Jumppanen (2017). This project is aiming for the *sensu stricto* definition of Management Procedures, in which the MP consists of:

i) pre-defined data collection

ii) pre-defined analytical methods (including assessment model specification or data processing)

iii) Harvest Control Rule to specify the management action

All three elements of the MP are simulation-tested together.

These MPs were tuned according to the criteria defined in Table 5. For expedience, all tuning for this report was conducted with 200 realizations from OMref17.2. Previous testing indicated tuning with only 216-realizations resulted in performance within 5% of the tuning objective when subsequently applied to a full suite of 2160 realizations. This level of tuning precision is considered adequate for the purposes of this report, but the full set of 2000 will be used for the TCMP.

Table 4. Qualitative definitions of the MPs used in this report (see Kolody and Jumpannen 2016 for full details). In all cases, TAC settings were made every 3 years starting in 2019, with a 15% TAC change constraint.

Label	Definition
PT4010	A catch-based "40:10-type" HCR coupled with a surplus production model.
PT4010F	An F-based "40:10-type" HCR coupled with a surplus production model.
IT	A CPUE-based HCR that "aims" for a desirable CPUE target by increasing or decreasing the TAC, depending whether CPUE is above or below the target, and whether COUE is trending up or down.
CCt	Constant catch set to achieve the desired tuning level
CC001	Fishing moratorium (for testing purposes)
CC407	Constant catch 407Kt (2015 catch) (for testing purposes)

Table 5. MP Tuning objectives defined for yellowfin and bigeye.

Label	Source	Definition
Y1/B1	TCMP YFT/BET objective 1	Pr(mean(SB(2019:2039))/SB(MSY) = 1.0) = 0.5
Y2	TCMP YFT objective 2	Pr(mean(SB(2024))/SB(MSY) = 1.0) = 0.5
B2	TCMP BET objective 2	Pr(Green Kobe 2019:2039) = 0.75

Four MPs are evaluated for tuning objectives Y1 and Y2 (Figure 20 and Figure 21). The summary statistic comparison plots for the 4 MPs are included for tuning objective Y1 (Figure 22, the new TCMP figures could not be prepared in time for this document). From these plots it is evident that:

- The IT MP rapidly stabilizes the median SB near the target level, at which time ~25% of simulations consistently remain below SBlim. There is a general decline in catch over the projection period.
- MP PT4010 is similar in behaviour to IT, but with larger fluctuations exceeding Btarg and Ftarg by a greater extent. The performance characteristics between the catch-based and F-based PT4010 rules was negligible.
- Tuning target Y2 is slightly more conservative, but results in generally similar MP dynamics (Figure 21).
- The constant catch rule does not appear to perform obviously worse than the feedback-based rules, and closer inspection reveals this to be related to a potentially serious problem with OMref17.2 (and the broader OM structure). In the current simulation tests, the constant catch scenarios clearly indicate that more than half of the scenarios frequently result in an unattainable TAC (the number is similar for the other MPs, but not obvious from the figures). If MP TAC recommendations are unattainable, the MP feedback control mechanism does not operate as intended. Furthermore, the MP parameters derived through tuning may be sensitive to assumptions about how dynamics operate in situations that are subject to numerical catch limits. e.g. In this case, the tuning parameter for MP PT4010 is the fraction of estimated MSY that should be taken if SSB > 0.4 SSB0. The tuned value is ~3, which is clearly not desirable under normal circumstances (i.e. we would expect a value near 1 unless the production model is badly biased).

An exploratory attempt to tune MPs for the more conservative B2 objective (defined for BET) suggested that it was not achievable with OMref17.2 and the default MPs (perhaps due to the TAC change constraints).

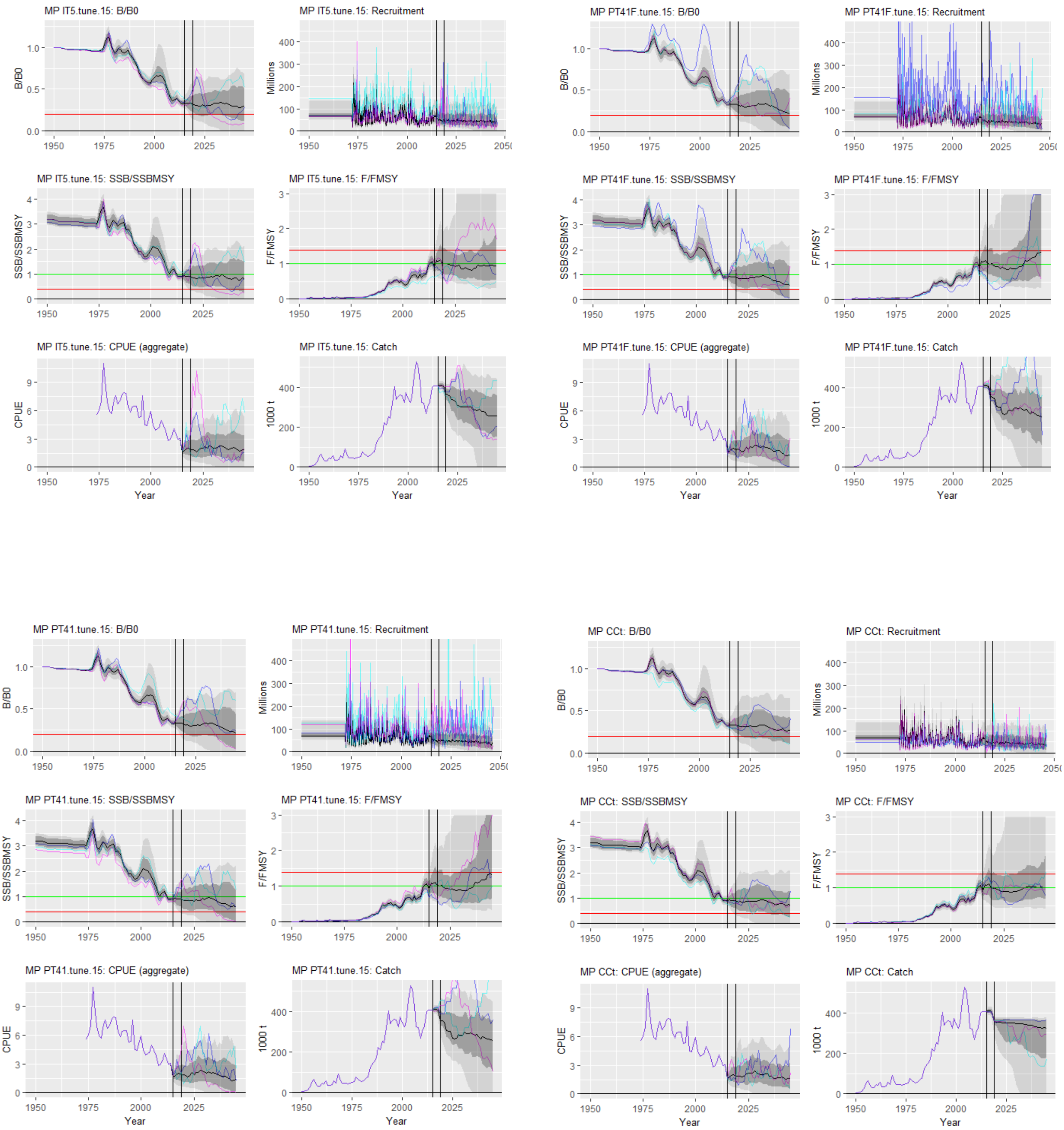


Figure 20. MSE results for 4 MPs, tuned to target Y1 with OMref17.2. Vertical lines indicate the first projection year, and the first active MP setting (constant current catches are assumed in the intervening period). Percentiles (10, 25, 75, 90) are represented by shading and a black line for the median. Three 3 individual realizations are included to illustrate that there will generally be much more variability than the summary distribution percentiles suggest.

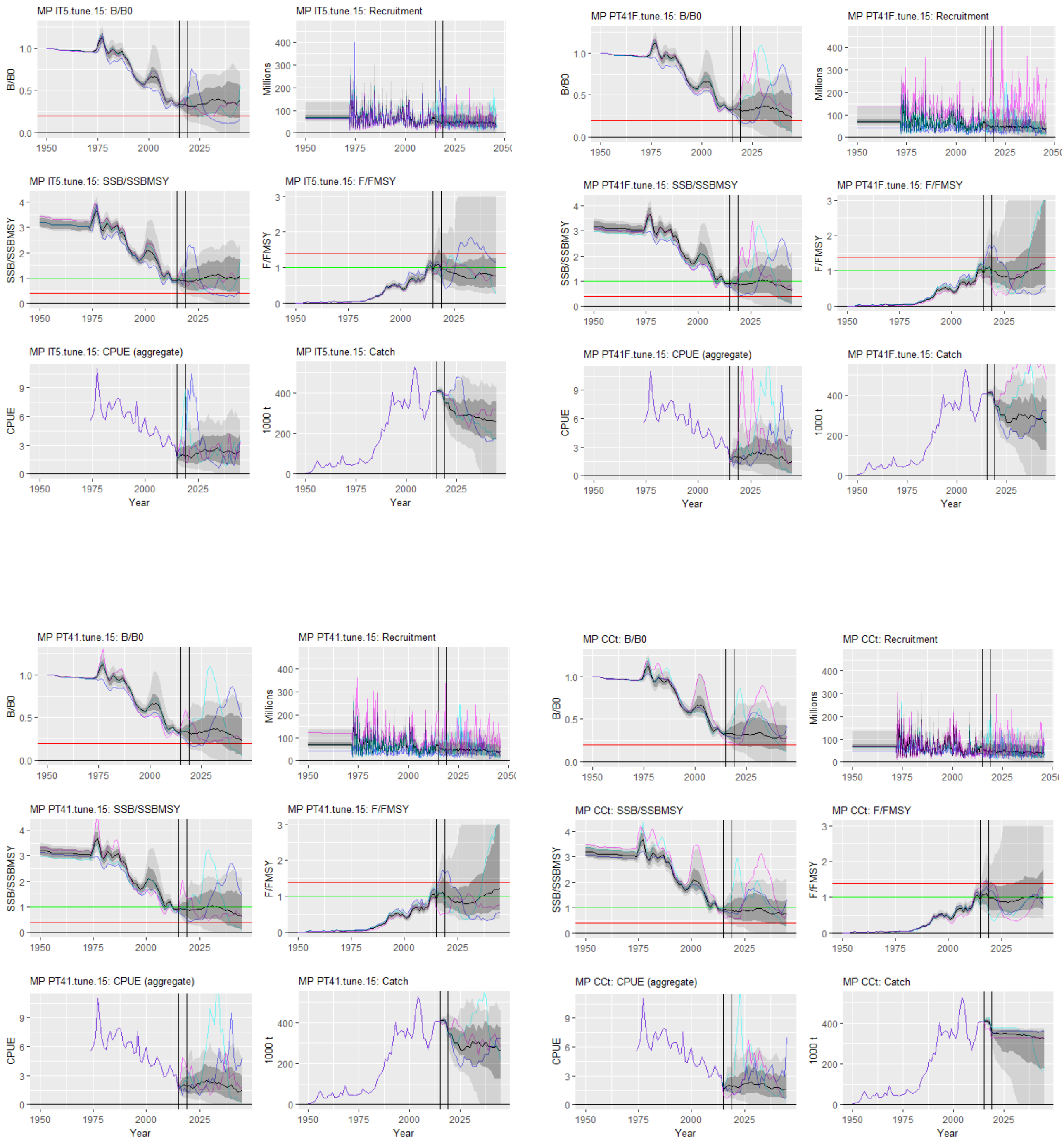


Figure 21. MSE results for 4 MPs, tuned to target Y2 with OMref17.2.

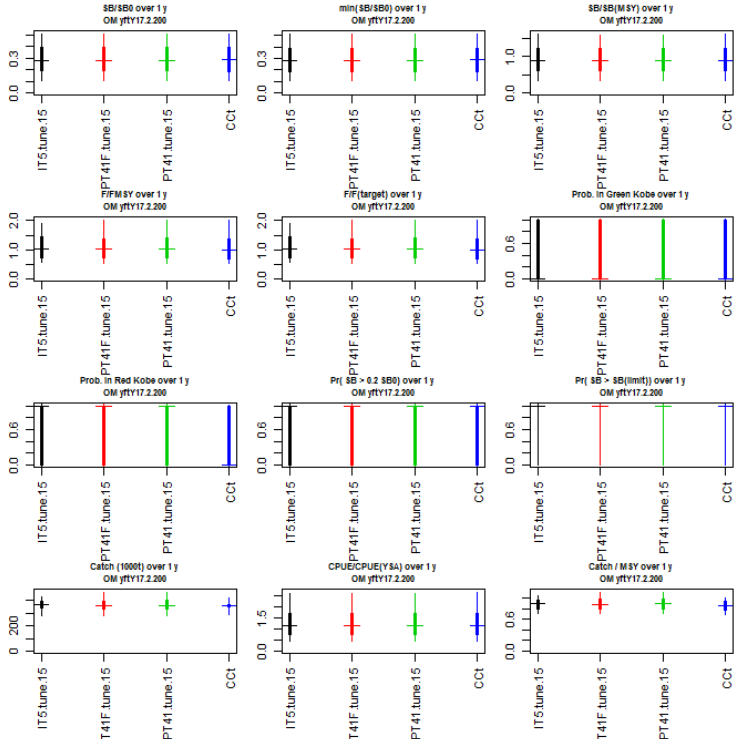
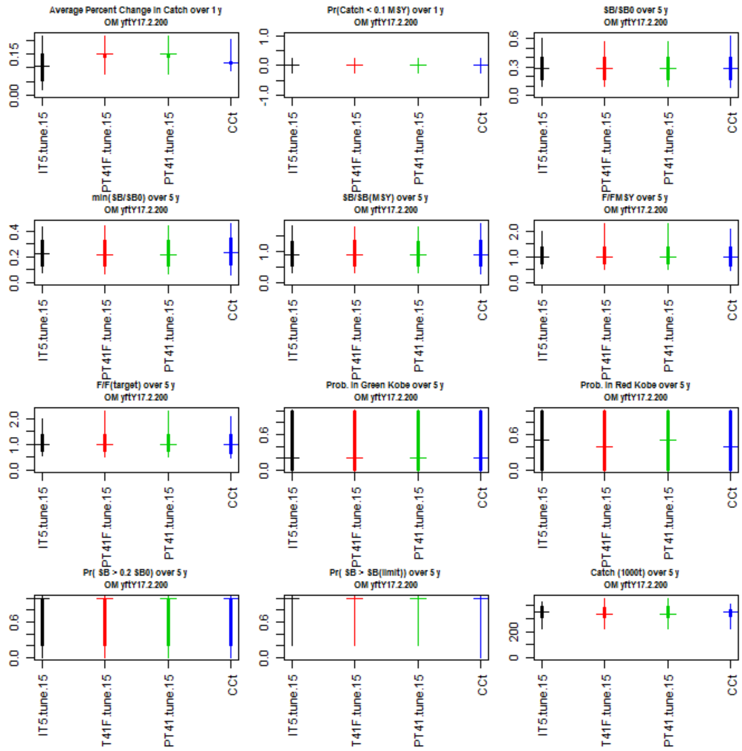
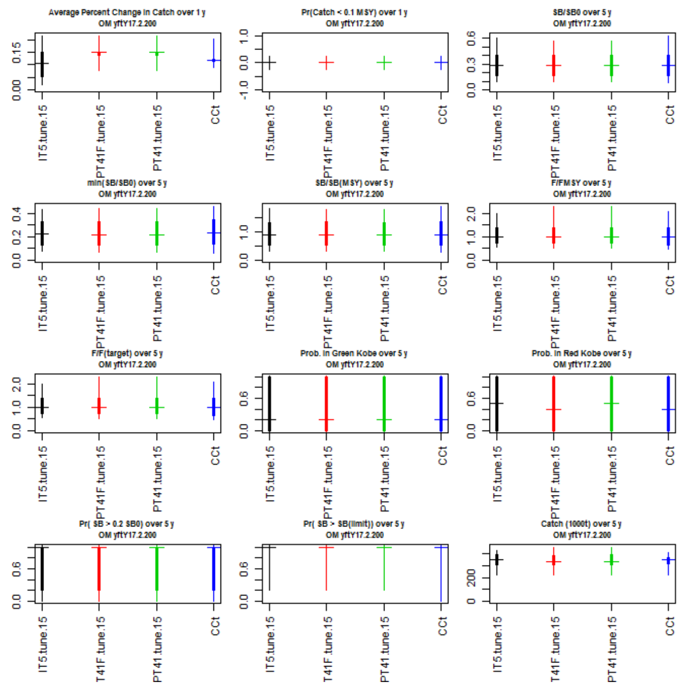


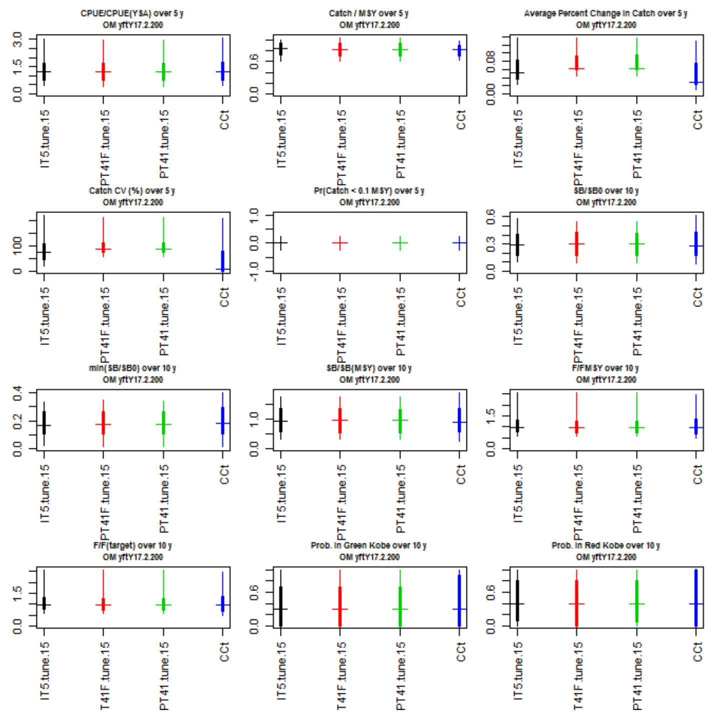
Figure 22. "Udon-Soba" plots comparing the 4 MPs, evaluated with OMref17.2, for tuning objective Y1.



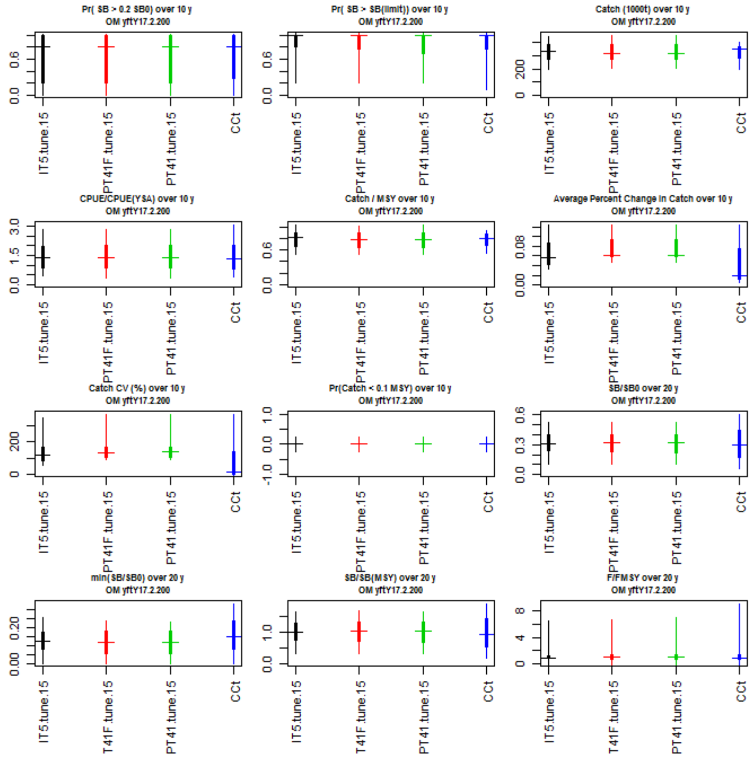
(Figure 22 continued)



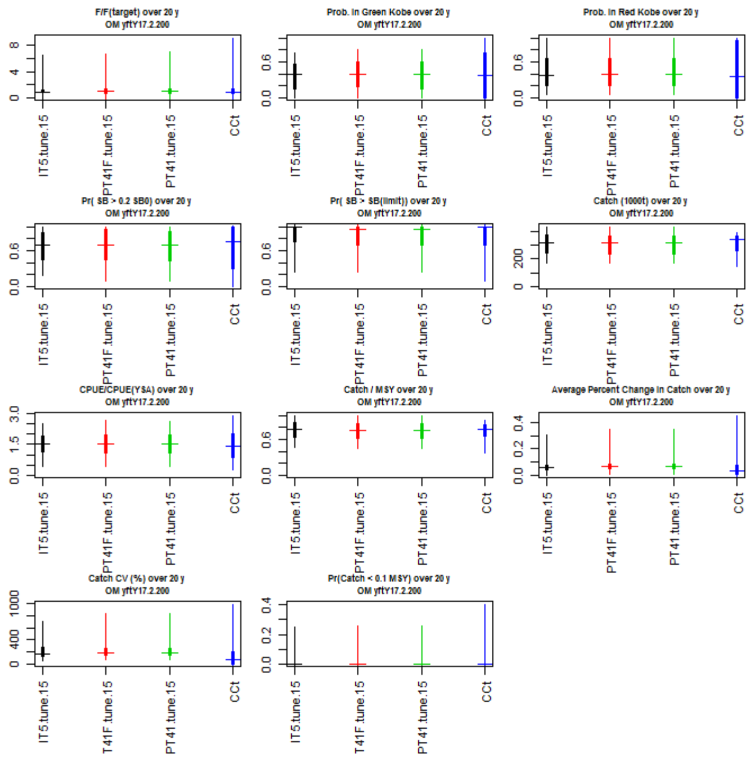
(Figure 22 continued)



(Figure 22 continued)



(Figure 22 continued)



(Figure 22 continued)

7 Yellowfin Robustness OMs

There is a substantial list of potential robustness OM scenarios identified in section 3. Progress has been made in adapting the simulator to describe some of these scenarios, but no results are reported at this time. We hope to get guidance from the informal MSE group on prioritization.

8 Discussion

The informal WPM MSE group represents the last chance for the broader IOTC community to provide feedback before results are presented to the TCMP 2018. As part of the IOTC informal WPM MSE meeting in March 2018, we will be presenting options and seeking feedback on the workplan priorities between Mar2018 and the TCMP (May 2018) and WPTT/WPM (Oct 2018).

Foremost among the concerns is finding a workable solution for the numerical issues arising from the high fishing mortality rates associated with OMref17.2. The problem was listed as the number one concern in the authors' self-critique of phase 1, but it was not considered an inevitable problem at the time (Kolody and Jumpannen 2016):

> *"...we would question whether either approach is very realistic when fishing mortality rates are so high that quotas cannot be met, catch rates are uneconomical, and incentives exist for fleets to stop fishing, move among areas and/or switch species targeting."*

Possible solutions may include:

- A review of OMref17.2 may conclude that it is too pessimistic (though this conclusion seems unlikely unless there is also a justification for rejecting the 2016 stock assessment).
- Some kind of bio-economic modelling may be required to explicitly model catch and effort responses to high fishing mortality.
- Alternative tuning objectives could reduce the frequency of high F events.
- Increasing the movement rates among areas and quarters may retain the aggregate productivity characteristics of the individual model in a way that more realistically reduces the high F impact.

We continue to welcome feedback on any aspect of the OM formulation, software or workplan. It should be recognized that a number of subjective decisions need to be made is an MSE process, subject to the judgement of the developers and time constraints. Ideally, MSE in an RFMO context should be undertaken with the active engagement of many parties, including at the technical level, to represent the broad scientific experience within the working parties. We continue to encourage other member scientists to download the source code, and scrutinize OM assumptions, performance characteristics and MP formulations, and present alternative views where appropriate (please contact the authors ahead of time, to ensure that the latest version of the code is available from github).

References

Kolody, D, Jumppanen, P. 2016. IOTC Yellowfin and Bigeye Tuna Management Strategy Evaluation: Phase 1 Technical Support Project Final Report. Indian Ocean Tuna Commission Working Paper IOTC-2016-WPTT18-32.

Kolody, D., Hoyle, S. 2015. Evaluation of tag mixing assumptions in western Pacific Ocean skipjack tuna stock assessment models. Fisheries Research 163 (2015) 127–140.

Langley, A. 2015. Stock assessment of yellowfin tuna in the Indian Ocean using stock synthesis. IOTC-2015-WPTT17-30.

Langley, A. 2016. An update of the 2015 Indian Ocean Yellowfin Tuna stock assessment for 2016. IOTC-2016-WPTT18-27.

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

Martell, SJD, Pine, WE, Walters, CJ, 2008. Parameterizing age-structured models from a fisheries management perspective. Canadian Journal of Fisheries and Aquatic Sciences, 2008, 65(8): 1586-1600, https://doi.org/10.1139/F08-055

MSE 2017. Report of the 6th Workshop on MSE of IOTC WPM Scientists, Bangkok, Thailand, 1-4 April 2017.

TCMP 2017. Chair report of the 1st IOTC Technical Committee on Management Procedures. Yogyakarta, Indonesia, 20 May 2017. IOTC–2017–TCMP01–R[E]

WPM 2016. Report of the 7th Session of the IOTC Working Party on Methods. Victoria, Seychelles, 11–13 November 2016. IOTC–2016–WPM07–R[E]

WPM 2017. Report of the 8th Session of the IOTC Working Party on Methods. Beau Vallon, Seychelles, 13-15 October 2017. IOTC–2017–WPM08–R[E]

WPTT 2016. Report of the 18th Session of the IOTC Working Party on Tropical Tunas. Seychelles, 5-10 November 2016. IOTC–2016–WPTT18–R[E]

WPTT 2017. Report of the 19th Session of the IOTC Working Party on Tropical Tunas. Seychelles, 17-22 October 2017. IOTC–2017–WPTT19–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