www.csiro.au

Indian Ocean Yellowfin Tuna Management Procedure Evaluation Update April 2020

Prepared for the IOTC MSE Task Force originally scheduled for March 2020

Dale Kolody (email: dale.kolody@csiro.au), Jemery Day, Paavo Jumppanen

CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart TAS 7000, Australia
April 2020.

Citation

Kolody, D, Day J, Jumppanen, P. 2020. Indian Ocean Yellowfin Tuna Management Procedure Evaluation Update April 2020. Report prepared for the Indian Ocean Tuna Commission Informal Management Strategy Evaluation workshop 2020.

This document will be assigned a number in the IOTC archive at a future date.

Contents

Acknowledgments

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

Summary

This working paper describes developments on the Indian Ocean Tuna Commission (IOTC) yellowfin (YFT) reference set and robustness test operating models (OMs), since the 2019 Working Party on Tropical Tunas (WPTT) and Working Party on Methods (WPM). In the following (for historical reasons), we mostly use the term MP and Management Strategy (MS) interchangeably, though we subscribe to the specific definition of MP as a subset of MS (as defined in the CCSBT and IWC, in which the MP aims for full specification and simulation testing of data collection and analytical methods). Management Strategy Evaluation (MSE) is the simulation testing process, using complex operating models, for evaluating performance of alternative MSs (or MPs). The intent was to obtain feedback on presentation requirements for the 2020 Technical Committee on Management Procedures (TCMP) meeting, and recommendations on further analyses and revisions for the OMs in preparation for the WPM and WPTT 2020 (but priorities changed due to the Covid-19 pandemic and remain uncertain).

There were requests for a number of YFT OM developments from the 2019 WPM and WPTT. These requests were made during a particularly busy WPTT meeting schedule, under the assumption that the OM changes would parallel the improvements realized through the YFT stock assessment update that was happening in parallel (at least that was our interpretation). However, the YFT assessment was aborted near the end of the formal WPTT meeting, and there was not time to discuss the implications of the various assessment innovations and their appropriateness for either future assessments or OM conditioning. The intent has always been that OM re-conditioning should not be required every year – only when there is evidence that the OM fails to adequately represent the plausible system uncertainty. In the absence of a revised stock assessment and a systematic discussion of proposed assessment/OM changes, we have interpreted the OM development requests as guidance for further investigation. However, it is not clear from the explorations undertaken to date that that we have improved the YFT OM relative to the previous iteration.

We did resolve some minor OM technical issues that have a large effect on a small number of atypical OM specifications, but do not expect that this would have much influence on overall MP performance given the emphasis on central tendency performance (e.g. 50 % probability of rebuilding spawning biomass by specific dates):

- The discontinuity between historical and projected CPUE, as presented to WPM (2020) has been greatly improved by:
  - A small number of OM configurations had a large recruitment spike shortly before the end of the time series. This was not a CPUE problem as originally thought, but it is dubious and not supported by much data. It was resolved by simply extending the recruitment constraint period from 4 to 12 quarters in the Stock Synthesis (SS) configuration (noting that random noise is introduced to the initial numbers-at-age during the MP evaluation projections to compensate for these constraints).
  - Adopting model-specific CPUE CV, autocorrelation, and initial error, which interprets systematic lack of fit between the conditioning CPUE series and MP CPUE

series as a correlated observation error on a case by case basis. This does not appear to make a noticeable difference in the current OM, because the CPUE CV tends to be less than the minimum constraint of 0.2.

We investigated the recommended OM changes from the 2019 WPTT, and found that they merit further consideration from the broader MSE group:

- Spatial Structure:
  - We propose not retaining the recommended 2 area structure for various reasons, including i) some important implications for the relative scaling of CPUE interpretation were not discussed by the WPTT and WPM, and ii) this implicitly requires the removal of tags from the majority of the OM grid.
- M options:
  - The WPTT recommended adding the (high) Western and Central Pacific M vector to the OM. Similar levels have been investigated in the IOTC (and ICCAT) in the past and found not to be plausible.
  - The proposed low M option was from the Atlantic (derived from a maximum observed age estimate) and is very similar to the tag-informed estimate arising from the Indian Ocean as used in previous iterations. We propose to retain the original low M vector.
  - We propose to retain the original intermediate M value from the previous OM (because it smoothed out the bimodality in OM ensemble characteristics in previous versions of the OM).
- Growth (length-at-age)
  - We propose retaining the Fonteneau growth curve at this time (along with Dortel model 2). The Dortel model 3 log-normal requested cannot be represented adequately in SS in the traditional manner. We are exploring other options (i.e. platoons and growth morphs), which have more flexibility. These options are computationally slow and cannot be represented in the OM code at present. But they are potentially interesting from the perspective of representing size-based fishery selectivity more realistically.
- Standardized CPUE series – the alternative standardized CPUE series recommended for the OM were not updated in 2019. The CPUE consultant agreed that the CPUE Working Group Terms of Reference for 2020 should include the production of specific CPUE-related recommendations and series to support the MSE work.

We also attempted to address some of the longstanding concerns about the OM and assessment: i) the retrospective pattern (in which new data persistently suggest that the previous assessment should have been more optimistic), and ii) the possibly-related issue of implausibly high exploitation rates (that often result in numerical problems in which the observed catch cannot be fully extracted). Explorations included:

- CPUE hyperdepletion – Our simple analysis suggests that this mechanism might help to reduce the retrospective problem slightly, although might come with the cost of a worse fit to the data.

- We revisited the use of environmental variables as a method of moving fish seasonally (and hence potentially reducing the very high exploitation rate for any individual strata), and again concluded that it does not seem useful:
    - The MSE code was modified to accept seasonal movement variability
    - SS results from seasonally averaged environmental variables produced almost identical results to the 2018 assessment that included seasonal and interannual variability.
    - A 144 model grid with (seasonally averaged) environmental variables was almost identical to the 144 model grid that did not include any environmental links (as used in previous MSE iterations), both in terms of estimated stock status, and general diagnostics (including the bimodal distribution of catch likelihood terms).
- We explored some of the iterative reweighting methods used to improve the internal consistency of variance-related parameters in other assessments. While yielding slightly different results, this seems to represent a small impact relative to the uncertainties introduced by the OM grid structure.

MP evaluations have not been repeated at this time, because (aside from the shortage of time): i) the minor OM changes that we have adopted are not likely to change the MP performance substantially relative to the previous iteration, and ii) there are many unresolved issues in the OM, relating to the last set of recommendations from the WPTT, that merit critical review by a broader section of the IOTC scientific community (ideally including the proponents of the specific recommendations).

A proposed list of development priorities is attached for discussion by the IOTC MSE Task Force. We continue to encourage other members of the IOTC scientific community to engage with the MSE process. The current phase of YFT/BET MSE support has funding until June 2021, however, staff allocations were expected to be reduced over the Jun-Oct 2020 interval, as this is normally the slow period for MSE development.

1. Introduction

This working paper describes developments on the IOTC yellowfin (YFT) reference set and robustness test operating models since WPTT (2019) and WPM (2019). The original intent was to obtain feedback about the results, presentation requirements for the 2020 TCMP meeting, and recommendations on further analyses and revisions for the OM, in preparation for the WPM and WPTT 2020. Given the evolving Covid-19 situation, and unclear implications for upcoming IOTC meetings, we will be seeking general feedback from the IOTC MSE Task Force including revised project milestones to work toward. This document is primarily an update on Kolody and Jumppanen (2019), and many details parallel the companion paper produced for bigeye (Kolody et al 2020). The intended audience is already familiar with the scope of the work, background and technical jargon. Other interested parties may need to consult the history of project reports found in https://github.com/pjumppanen/niMSE-IO-BET-YFT/.

There were requests for a number of YFT MSE developments from the WPM (2020) and particularly WPTT (2020). These requests were made during a particularly busy WPTT meeting schedule, under the implicit assumption that OM changes would parallel an updated YFT assessment. A number of investigations to the stock assessment were undertaken before (Urtizberea et al 2019a) and during WPTT (2019). However, there was not time to thoroughly discuss the assessment results, and it was not accepted as a new assessment. It is our interpretation that these new explorations were useful, but the overall perception of stock status was largely unchanged, and long-standing concerns about the stock assessment did not appear to be resolved. The MSE work has always been undertaken with the expectation that OM re-conditioning should not be required, unless there is substantive new evidence that the OM is fundamentally flawed (and/or several years out of date), in a way that will likely bias MP performance evaluation results. Given that the YFT explorations presented to WPTT 2019 were ultimately not adopted, we interpret the specific requests for YFT MSE development as guidelines for further investigation, which may be helpful for future OM and stock assessment iterations.

We have attempted to move the YFT MSE forward on three fronts: i) improving some technical implementation details related to consistency between historical and projected CPUE series, ii) investigating the appropriateness and feasibility of the 2019 WPTT OM change requests, and iii) investigating diagnostics and potential structural options for reducing the fundamental problems with the YFT assessment and OM. With respect to point (iii) – we note that recent versions of the stock assessment and OM are generally pessimistic, and suggest that disruptive management actions will be required for IOTC to meet management objectives. Given the catch history and declining longline CPUE trends, this inference is not surprising. However, there are some troubling characteristics with most or all of the models in the ensembles, including i) there appears to be a retrospective pattern, in which the newest data seem to consistently suggest that the previous assessment should have been somewhat more optimistic, and ii) in most or all models, there are implausibly high exploitation rates estimated for some age-classes, often with numerical problems in which the observed catch cannot be fully extracted.

The MP evaluations were not repeated at this time, because i) it is not clear that what we are proposing as the new reference set has made tangible improvements to the previous OM (and the stock status inferences remain largely unchanged), and ii) there are many unresolved issues in the OM that should be reviewed and endorsed by the broader IOTC scientific community, iii) a full OM update would require additional CPUE series that were not provided by the CPUE working group in 2019 (this has been flagged for the Terms of Reference for future CPUE working group activities), and iv) this does not seem like a priority if the June 2020 TCMP meeting is cancelled.

The various OM assumption option abbreviations are defined qualitatively in Table 1; individual models and OM ensembles in Table 2. The OM grid change requests from WPTT (2020) are reproduced in Table 3, and discussed in section 3.

Table 1. Model assumption option abbreviations (as used in the text and figures). Bold indicates the assessment reference case assumption. Some abbreviations may relate to explorations that have not yet been examined, relate to bigeye, and/or are not reported in the current document.

Abbreviation	Definition
	Stock-recruit function (*h* = steepness)
h70	Beverton-Holt, *h* = 0.7
h80	**Beverton-Holt, *h* = 0.8**
h90	Beverton-Holt, *h* = 0.9
Rh70	Ricker, *h* = 0.7
Rh80	Ricker, *h* = 0.8
Rh90	Ricker, *h* = 0.9
	Recruitment deviation penalty
sr4	$\sigma_R = 0.4$
sr6	**$\sigma_R = 0.6$**
sr8	$\sigma_R = 0.8$
	Natural mortality multiplier relative to SA-base
M12	1.2
M10	**1.0**
M08	0.8
M06	0.6
	Tag recapture data weighting (tag composition and negative binomial)
t00	$\lambda = 0$
t0001	$\lambda = 0.001$
t001	$\lambda = 0.01$
t01	$\lambda = 0.1$
t10	**$\lambda = 1.0$**
t15	$\lambda = 1.5$
	Assumed longline CPUE catchability trend (compounded)
q0	**0% per annum**
q1	1% per annum
q3	3% per annum
q5	5% per annum
	Tropical CPUE standardization method (error assumption for all series)
iH	**Hooks Between Floats ($\sigma_{CPUE} = 0.3$)**
i10H	Hooks Between Floats ($\sigma_{CPUE} = 0.1$)
iC	Cluster analysis ($\sigma_{CPUE} = 0.3$)
i10C	Cluster analysis ($\sigma_{CPUE} = 0.1$)
	Tag mixing period
x3	3 quarters
x4	**4 quarters**
x8	8 quarters
	Longline selectivity
SL	**Stationary, logistic, shared among areas**
SD	Double-normal (potentially dome-shaped), shared among areas (except NW region which retains logistic?)
	Size composition input Effective Sample Sizes (ESS)
ESS2	ESS = 2, all fisheries
ESS5	**ESS = 5, all fisheries**
CLRW	ESS = One iteration of reweighting; the output ESS from the reference case assessment analogue, capped at 100.
CL75	ESS = One iteration of reweighting; the output ESS from the reference case assessment analogue raised to the power of 0.75 and capped at 100.

Table 2. Yellowfin model and ensemble definitions referred to in the text.

Model Name	Definition (assumption abbreviations are defined in Table 1)
ref2018	Reference case Stock Synthesis assessment extracted from the Fu et al (2018) stock assessment, used as a reference point for examining the impact of alternative assumptions and exploratory analyses (includes environmental-linked movement).
meanSeasEnv	Single model as ref2018, except environmental time series are replaced by the seasonal averages (i.e. no inter-annual variability).
NoEnv	Single model analogue of ref2018 from OMgridY20.1, with no environmental links.
OMgridY20.1	The *tentatively proposed* YFT reference set ensemble (133 converged models from the 144 model fractional grid), includes models which might fail other diagnostic criteria (e.g. catch likelihood constraint).
OMgridY20.2	Identical to OMgridY20.1, except with the seasonally averaged environment data linked to movement (summary plots exclude ~10 % of models (due to convergence failure)).
OMrefY20.1	The default YFT reference set ensemble (58 models - OMgridY20.1 converged models and models with catch likelihood < 10^{-5} retained).
OMrefY19.4	The YFT reference set OM ensemble presented to the 2019 WPTT.

2. Revisiting CPUE projection assumptions in the IOTC MP evaluations

In the 2019 WPM, we reported that there often appears to be a discontinuity between the historical CPUE observations and the first projected CPUE (i.e. a sudden jump in CPUE in the first year of projections). The issue was thought to arise from two sources of inconsistency: i) the lack of fit between predicted and observed CPUE, and more importantly ii) the MP must always use the same historical CPUE series, regardless of which CPUE series was used in the OM conditioning. In addition to data weighting (variance-related) assumptions, there are a number of assumption options that determine the actual CPUE data used in the OM conditioning grid, e.g. for yellowfin, there are:

- 2 × Catchability trend assumptions
- 2 × Methods to account for targeting in tropical regions
    - trop_cl0_hb1_hk1_TW2005_discard2 + temp_cl1_hb0_hk1_TW2005_discard2
    - trop_cl0_hb1_hk1_TW2005_discard2 + temp_cl0_hb1_hk1_TW2005_discard2
- 2 × Regional-scaling factor assumptions

These assumptions interact to produce 8 alternative interpretations of historical abundance, all of which the WPM and WPTT consider to be sufficiently plausible that they should be represented in the reference set OM with equal weighting. Some of these series are probably not very different, but only one is equivalent to the CPUE series that the MP will use.

Our approach for resolving this issue is discussed in more detail in the bigeye companion paper (Kolody et al 2020), where it now appears to be more important. Further investigation revealed that the most problematic cases of CPUE discontinuity for YFT were associated with dubious recent recruitment spikes. These were not identified in the previous iteration, because our general model diagnostics were based on the standard MP evaluation outputs, that trim off the 10th percentile tails. The large recruitment spikes mostly occurred 5-8 quarters before the last model timestep. The timing and magnitude of these spikes is not consistent among models and they are not strongly informed by data. These spikes have been largely eliminated by simply increasing the recent recruitment constraint from 4 to 12 quarters. Additional error is introduced to the initial numbers-at-age to compensate for the lack of recruitment variability in the conditioning.

Figure 1 shows some random example cases of YFT predicted CPUE with the MP CPUE observations, including three approaches for the projections:

i) Catchability calculated over all historical CPUE observations, fixed MP CPUE CV = 20%, auto-correlation = 0.5.
ii) Catchability calculated over the final 3 years of historical CPUE observations (as proposed by WPM 2020), fixed MP CPUE CV = 20%, auto-correlation = 0.5.
iii) Catchability calculated over the whole time period, with model-specific CV and auto-correlation based on the deviation between the MP CPUE and model-specific vulnerable numbers (with a minimum CV of 20%).

In all cases, the first projected CPUE observation error is correlated with the last historical error. For most or all models, the difference between case (i) and (iii) is negligible. The quality of fit between the MP CPUE and the predicted vulnerable numbers is better than we would have reason to expect for commercial CPUE (and hence the 20% minimum CV for projections is mostly or always active). Figure 1 (D-F) shows the very large systematic lack of fit to the historical data that arises if case (ii) is adopted, and the implications for the elevated projection CV and auto-correlation, if we want to be internally consistent with that approach.

Figure 2 shows the corresponding distributions of error characteristics for the whole OMgridY20.1 ensemble. The median auto-correlation is higher than 0.5 assumed previously, but the effect of the auto-correlation is closely linked to the CV (i.e. auto-correlation becomes less important as the CV decreases)

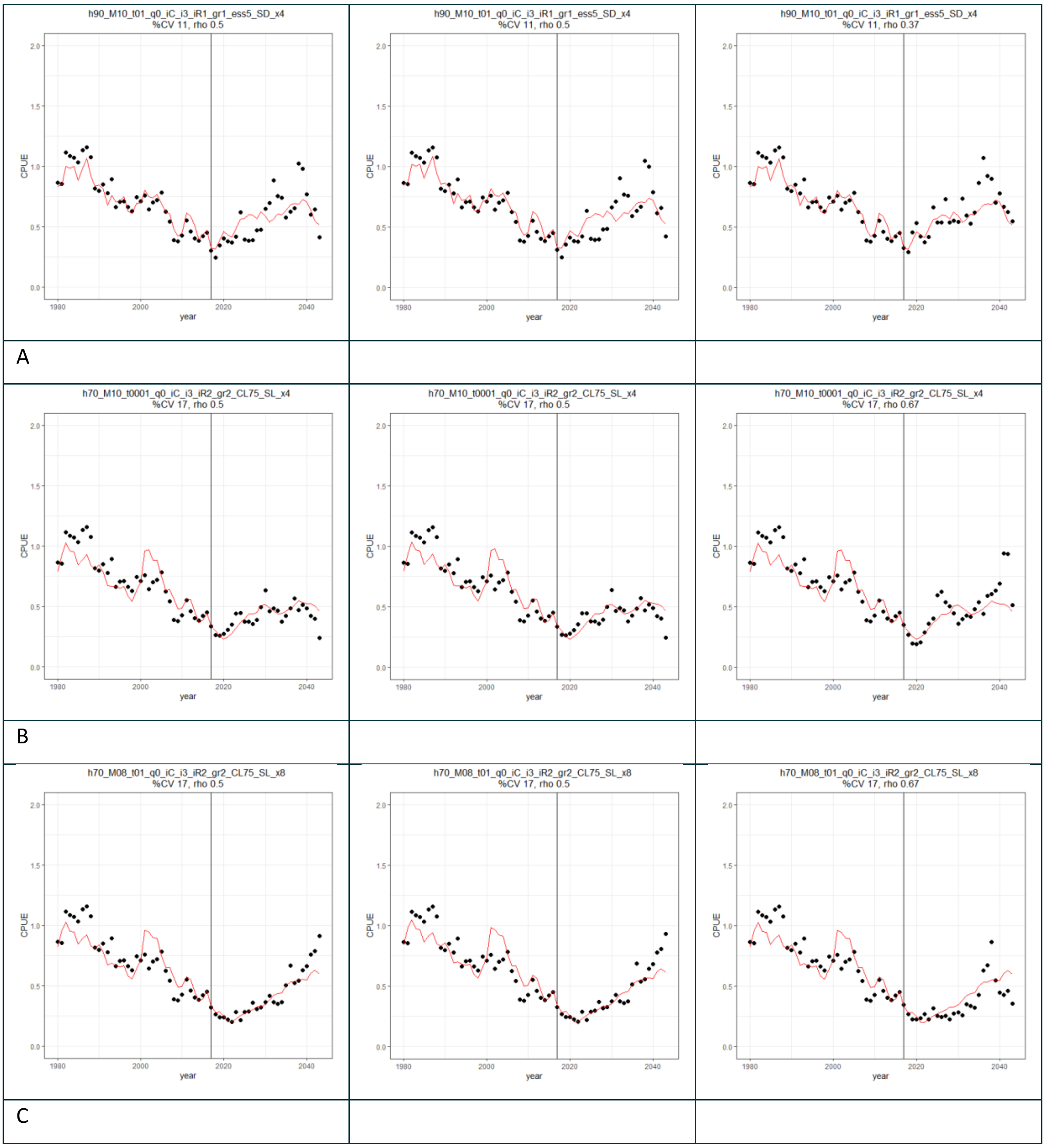
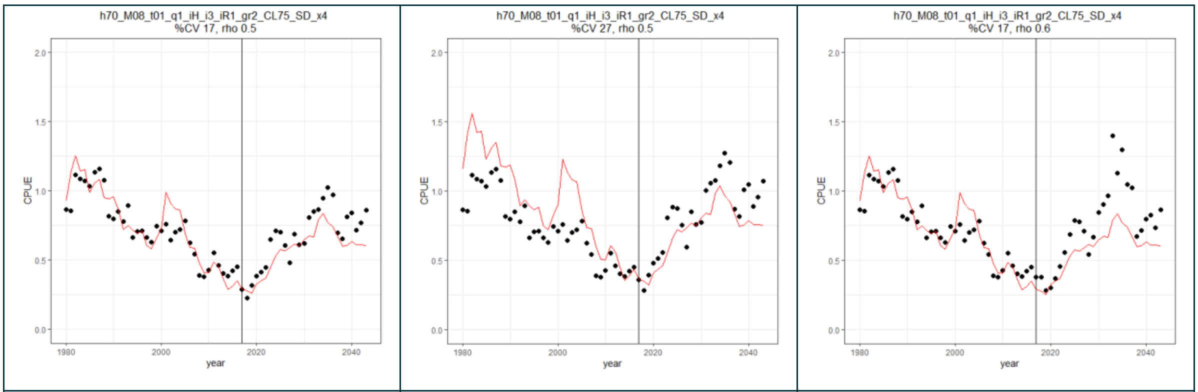
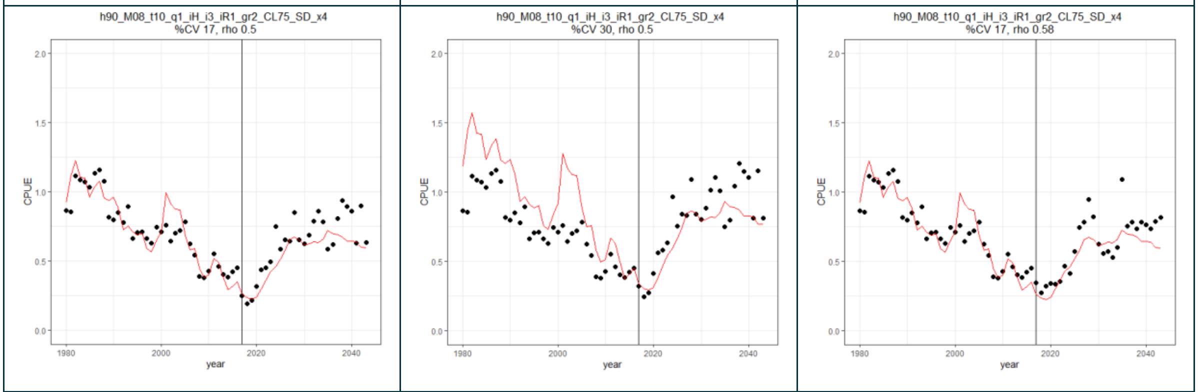


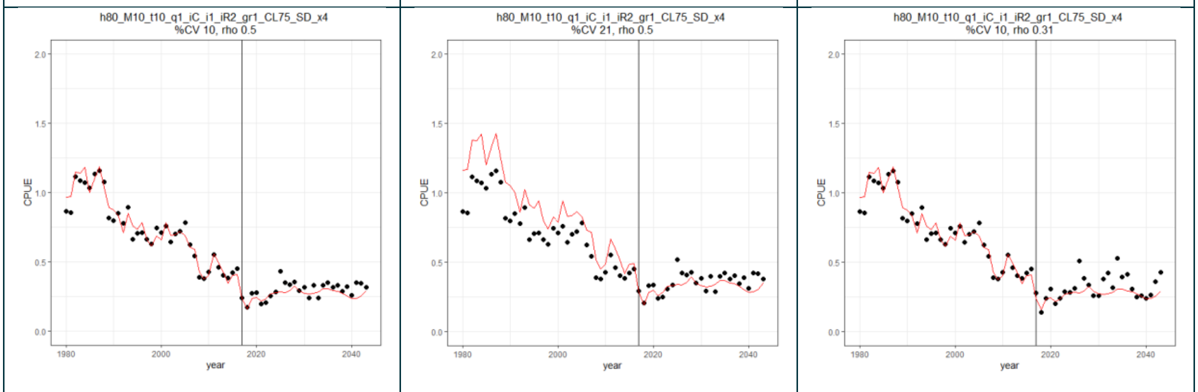
Figure 1. Some examples from the YFT OM ensemble contrasting the predicted (lines) and observed (circles) CPUE series for the MPs, derived from the three CPUE projection options outlined in the text. Vertical line indicates the first projection year. Left column – q calculated over whole period, projection CV = 20% and auto-correlation = 0.5 (for all models as used in previous MSE iterations); middle column – q calculated over final 3 years only, projection CV = 20% and auto-correlation = 0.5; right column – q calculated over whole period, CV and auto-correlation correspond to the individual model (or 20%, whichever is higher). Vertical line is the last year of real observations. (Figure 1 continued on next page)



D

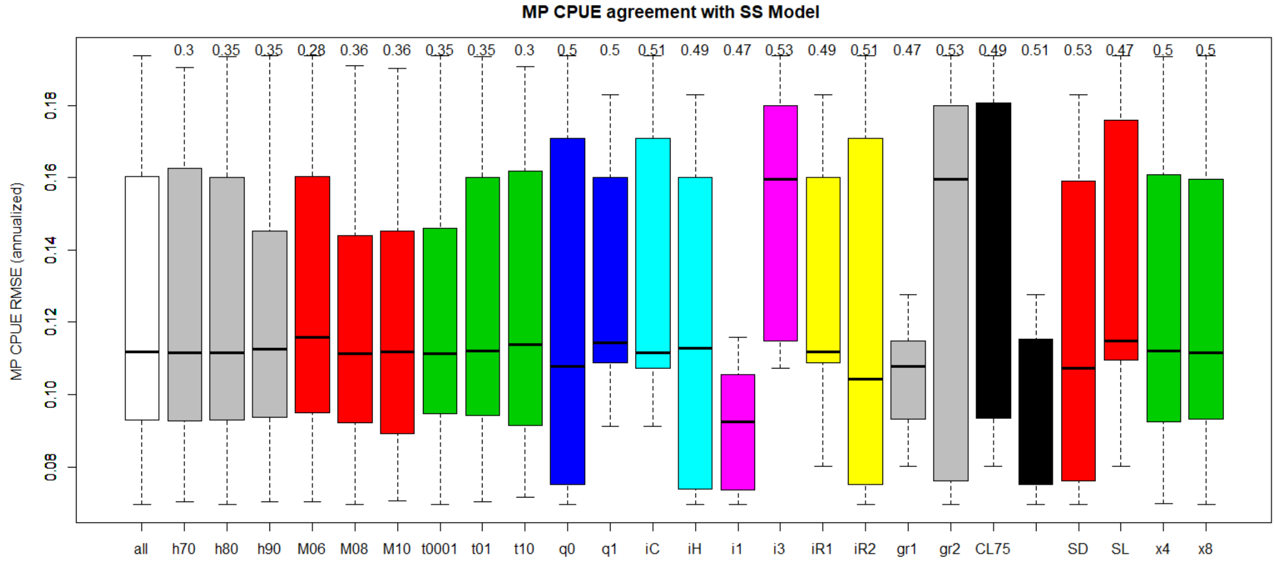


E



F

(Figure 1 cont.)



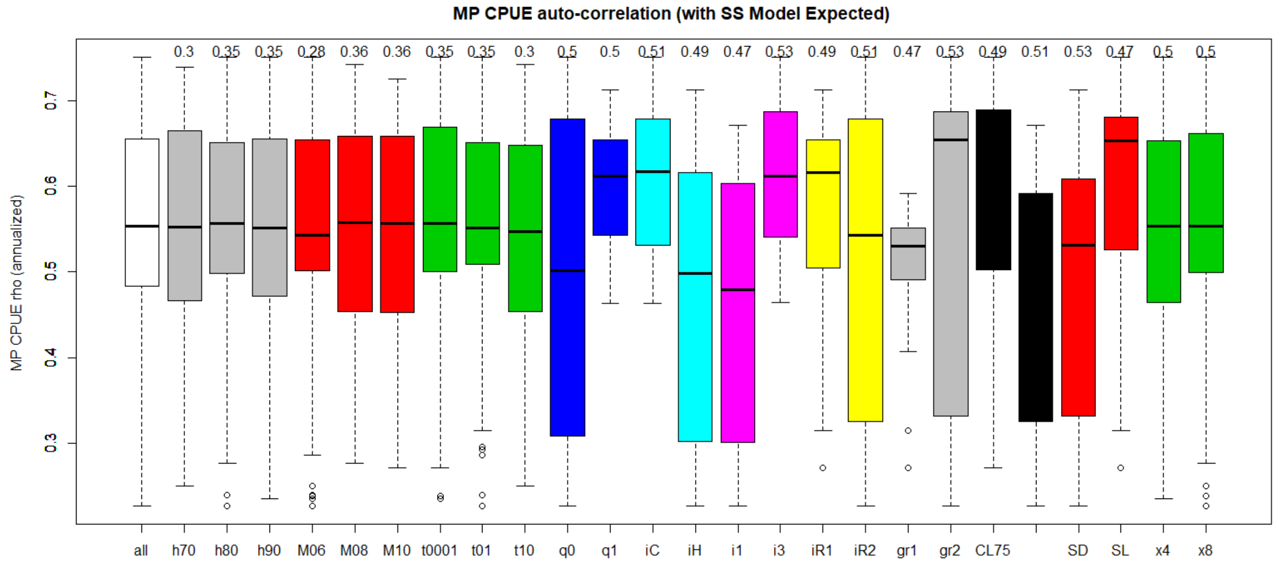


Figure 2. OMrefY20.1 agreement between the predicted and MP CPUE (spatially-aggregated, annualized RMSE, top, and lag (1 year) autocorrelation, bottom), marginalized over individual grid assumptions.

3. Revisiting the YFT OM development requests from 2020.

Table 3 lists the YFT OM structure proposed by WPTT 2020, while Table 4 describes our tentatively proposed OM following the considerations detailed in this document. Key points of difference include:

1) We propose to proceed with only the 4-area structure at this time, dropping the 2-area structure for the following reasons:
   a. Merging the 2 and 4 area models in a grid unbalances the intended influence of the tagging data. i.e. The WPTT agreed that tags were not appropriate for the 2 area structure because of the low mixing rate of tagged and untagged fish. If the 2 and 4 area models are balanced, the tags will not apply to 67% of models and only 17% of models will receive the full tag weight.
   b. The 2 area model has adopted a different interpretation of the regional scaling factors from the historical approach. Urtizberea et al (2019a) notes that: "*…the model can be very sensitive to scaling factor and in the case of the reference case, the model could not converge assuming the same catchability in both regions. Thus, the scaling factor is considered when the longline CPUEs are aggregated within the regions defined in the reference case but without assuming the same catchability between the reference case regions.*" This is a fundamental change in how the CPUE data are interpreted, and was seemingly not discussed by the WPM or WPTT.
   c. The OM projection software is not structured to simultaneously handle models with differing regional structures within a single OM object.
   d. The 2 area models did not appear to substantially change the stock status inferences (Urtizberea et al (2019b)).
   e. The proposed assumption of no movement in the 2 area model (motivated by the lack of tagging data to directly estimate movement), implies a curious situation in which there are 2 effectively isolated populations. If this is believed to be realistic, then the obvious approach would be to assess and manage the 2 areas independently.

2) Different YFT natural mortality (M) vectors are compared in Figure 3. At this time, we propose to retain the M06, M08 and M10 vectors as used in previous MSE iterations. M10 is the base case (common to Table 3 and Table 4), M06 is a low M option that was estimated in a previous Indian Ocean assessment (strongly supported by the tags), and M08 is intermediate (added to reduce bimodality in earlier OM ensemble results). The M06 estimate is very similar to the vector requested by WPTT (2019), that was derived from the maximum observed age estimate in the Atlantic (provided by Agurtzane Urtizberea, AZTI). We are reluctant to retain the WCPO M estimate because i) it was not properly presented and discussed at the WPTT, and ii) it closely resembles the high M assumption that was explored previously in the OM, and the high M from the Atlantic, which was rejected by ICCAT (and possibly the 2019 IOTC assessment team) as implausible (i.e. not enough large fish available to fit the CL distribution). We hope to examine the level of uncertainty expected from the maximum observed age estimate (i.e. recognizing that

the observed maximum age is very much a function of the exploitation history and sampling approach).

3) At this time, we propose to retain the Fonteneau growth curve (and Dortel (2014) model 2). The Dortel model 3 lognormal curve (D3ln) requested by WPTT (2020) cannot be easily represented in SS, because there is limited control over the variance in length-at-age (this model estimates a large increase in variance with age, for fish that are near the length asymptote). We will continue to explore the possibility of using D3ln using SS platoons (or growth morphs), which offer more flexibility. The approach partitions each age-class into multiple semi-independent units (Figure 4), and offers a new perspective for investigating more realistic size-based selectivity. However, this comes with a cost (5 platoons = 5X computational overhead), and this cannot be represented within the MSE projection code at this time.

Other points of OM evolution or divergence from the 2018 or 2019 assessment include:

- We are proposing to retain the OM structure inherited from the 2018 assessment, including using Stock Synthesis (SS) (Methot and Wetzel 2013) SS3.24Z software (rather than SS3.3x as explored in the aborted 2019 assessment) because: i) It is not clear that the 2019 analyses offered any new insights over the 2018 assessment, ii) we were not sure if there would be unexpected compatibility issues between the MSE software and SS3.3x that could be resolved in time for the MSE task force, and iii) The 2019 CPUE Working Group did not update the alternate CPUE series that were requested for the OM. In consultation with the CPUE Working Group consultant (Simon Hoyle, pers. comm.) there is strong agreement that the representation of CPUE uncertainty is critical to the MSE work. It was proposed that future CPUE working group Terms of Reference (starting in 2020) should include recommendations about the CPUE uncertainty to include in the MSE, including the provision of specific CPUE series. We expect that the IOTC assessments will eventually migrate to SS3.3x, with the OMs to follow.
- The discontinuity between historical and projected CPUE is improved using model-specific error characteristics, and recruitment variability for the most recent 12 cohorts is constrained (as discussed in the previous section).
- The SS maximum fishing mortality setting was raised to 6.0 and 7 hybrid F iterations (up from 2.9 and 4, respectively, in the stock assessment). This is discussed in more detail in the sections below.

Table 3. YFT OM structure requested by WPTT (2020).

Spatial Structure – Equal weighting on both that are otherwise unbalanced
(Second option to be reviewed intersessionally by MSE Task Force)

- 4 regions
- 2 regions; merge 1+2, 3+4

Stock-recruit function (*h* = steepness)

- Beverton-Holt, *h* = 0.7
- Beverton-Holt, *h* = 0.8
- Beverton-Holt, *h* = 0.9

Natural mortality multiplier relative to reference case M vector

- 2019 Base case
- 2019 Atlantic
- Tremblay-Boyer et al. 2017 (WCPO)

Tag recapture data weighting (tag composition and negative binomial)
(Applies to 2 area structure only)

- $\lambda = 0.001$
- $\lambda = 0.1$
- $\lambda = 1.0$

Growth curve

- Dortel et al. (2015) – model 2
- Dortel et al. (2015) – model 3 with lognormal error

Assumed longline CPUE catchability trend (compounded)

- 0% per annum
- 1% per annum

Tropical longline CPUE standardization method

- Hooks Between Floats
- Cluster analysis

Longline CPUE error assumption (quarterly observations)

- $\sigma_{CPUE} = 0.3$
- $\sigma_{CPUE} = 0.1$

longline CPUE Regional-scaling factors

- reference case
- alternate

Tag mixing period

- 4 quarters
- 8 quarters

Table 4. OMgridY20.1 - proposed YFT reference set OM uncertainty dimensions (abbreviations defined in Table 1).

Abbreviation	Definition
	Spatial Structure – 4 regions only (**removed 2 region option**)
	Stock-recruit function (*h* = steepness)
h70	Beverton-Holt, *h* = 0.7
h80	Beverton-Holt, *h* = 0.8
h90	Beverton-Holt, *h* = 0.9
	Natural mortality (multiplier relative to reference case M vector M10)
	~1.2+ (**removed WCPO vector that seems implausible for IOTC and Atlantic**)
M10	1.0 - Base case (1.0)
M08	0.8 – (**Intermediate vector to smooth bimodal OM results**)
M06	0.6 - (**Very similar to and adopted instead of "Atlantic" M-low request**)
	Tag recapture data weighting (tag composition and negative binomial)
t0001	λ = 0.001
t01	λ = 0.1 (**removed in previous iterations, re-introduced on WPTT request**)
t10	λ = 1.0
	Growth curve
gr1	Dortel et al. (2014) model 2
gr2	Fonteneau (c. 2012) (**Dortel model 3 lognormal under exploration**)
	Assumed longline CPUE catchability trend (compounded)
q0	0% per annum
q1	1% per annum
	Tropical longline CPUE standardization method
iH	Hooks Between Floats
iC	Cluster analysis
	(neither series was updated in 2019 because iC was not produced)
	Longline CPUE error assumption (quarterly observations)
i3	σ_{CPUE} = 0.3
i1	σ_{CPUE} = 0.1
	longline CPUE Regional-scaling factors
iR1	reference case
iR2	alternate
	Tag mixing period
x4	4 quarters
x8	8 quarters
	Longline fishery selectivity
SL	Stationary, logistic, shared among areas
SD	Stationary, double-normal (potentially dome-shaped), shared among regions
	Size composition input Effective Sample Sizes (ESS)
ESS5	ESS = 5, all fisheries
CL75	ESS = One iteration of reweighting from reference case model (fishery-specific), raised to the power of 0.75, capped at 100.

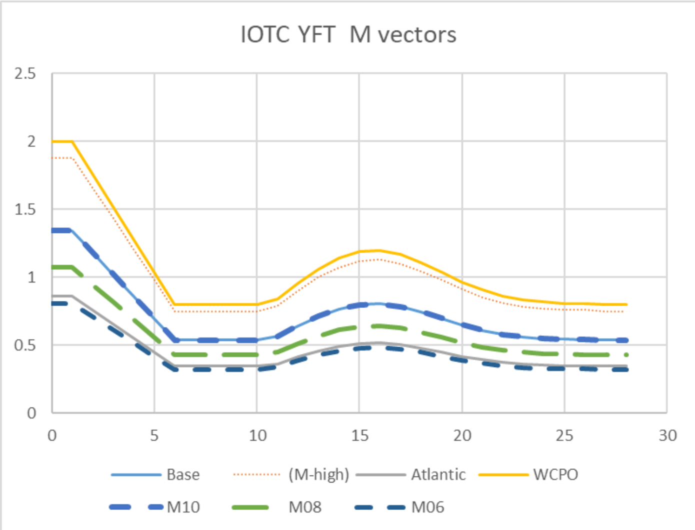


Figure 3. Alternative YFT M vectors (M06, M08, M10, as used in the previous version of the OM; Base, Atlantic and WCPO as proposed by 2020 WPTT, plus M-high from Atlantic for comparison).

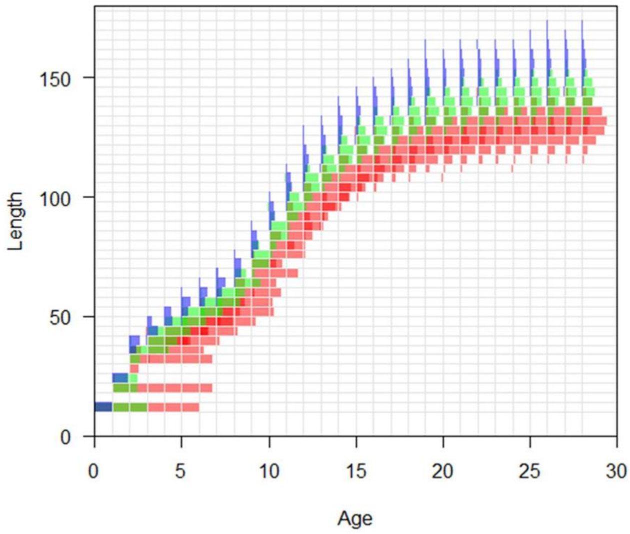


Figure 4. Example from the initial attempt to represent Dortel model 3-lognormal growth curve using 3 platoons. i.e. It seems possible to represent the required skewness in the older age distribution using SS platoons, but not the distributional differences among ages (though this may be possible with growth morphs).

4. Exploratory analyses

4.1 Can hyperdepletion resolve the retrospective pattern and high exploitation rates?

Matsumoto et al. (2018) identified a troubling retrospective pattern in the 2017 YFT assessment, that remains relevant. Removing x years of data results in a more depleted stock status estimate for year T-x relative to that observed in T-x when all data are included. When subsequent catches were taken, the population did not decline as much as would have been expected if the T-x assessment been correct, so the stock status seemingly should have been more optimistic than previously estimated. Since this pattern is consistently repeated, it seems reasonable to expect that it will probably continue into the future, and the most recent assessment will probably be deemed too pessimistic when examined in the future.

Given that CPUE are the most informative data with respect to relative abundance, one mechanism for introducing the retrospective pattern might be a non-linear relationship between CPUE and abundance. Hyperdepletion is the situation in which CPUE exaggerates the level of depletion, and is commonly recognized in the early development of many tuna longline fisheries. We imposed several different values for the SS non-linear abundance-CPUE relationship parameter H (equally for all longline fleets), where $I = QN^{1+H}$ (see Figure 5). If the retrospective pattern is a simple result of this non-linearity, we would predict that the retrospective pattern would be more exaggerated with negative values of H (hyperstability), and diminish with increasing H (eventually reversing direction, such that historical assessments would be shown to be too optimistic).

H values from -0.5 to 0.5 had an impact on the stock status inferences and retrospectives, but not in the simple way that we might have expected if it is the only relevant factor (Figure 6 - Figure 7). It might be argued that H = 0.1 has the best retrospective pattern, but (qualitatively) it does not seem that different from H=0. The (total) negative log-likelihood favours H = 0, but H = 0.1 is not far off (Table 5). The catch likelihood favours H = 0.1, but H = 0 is similar (and both are marginal cases with respect to the catch likelihood filtering criterion used in previous MSE iterations as discussed in the following section). We did not anticipate that the catch likelihood would increase as H increased > 0.1. Our expectation was that higher H would consistently support a more optimistic assessment (less depletion), resulting in less difficulty extracting the catch and a catch likelihood < 10^{-5}. This general mechanism may be worth further investigation, but we would argue against adopting H <> 0 for the OM at this time. If the CPUE-abundance non-linearity is operating, it is probably far more complicated, e.g. differing by region, varying as a function of time and confounded with other CPUE factors such as technological change and the actual amount and distribution of effort within regions. The IOTC CPUE Working Group might be better positioned to speculate on this issue.

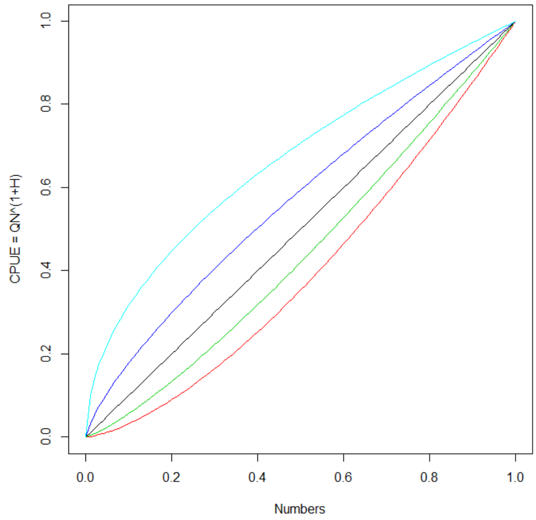


Figure 5. SS non-linear CPUE – abundance relationship for values of H from +0.5 (red, hyperdepletion) to -0.5 (pale blue, hyperstability).

Table 5. Yellowfin SS objective function values for a range of fixed values for the hyperdepletion parameter H.

H (non-linearity parameter)	Objective function value (relative)		Catch Likelihood
-0.5	11350.0	(1989.29)	9.90
-0.2	9793.26	(432.55)	6.98×10^{-1}
-0.1	9402.45	(41.74)	6.73×10^{-2}
0 (assessment value)	**9360.71**	**(0)**	8.41×10^{-4}
0.1	9367.11	(6.4)	**1.60** $\times10^{-4}$
0.2	9430.90	(70.19)	4.38×10^{-1}
0.5	10087.0	(726.29)	16.0

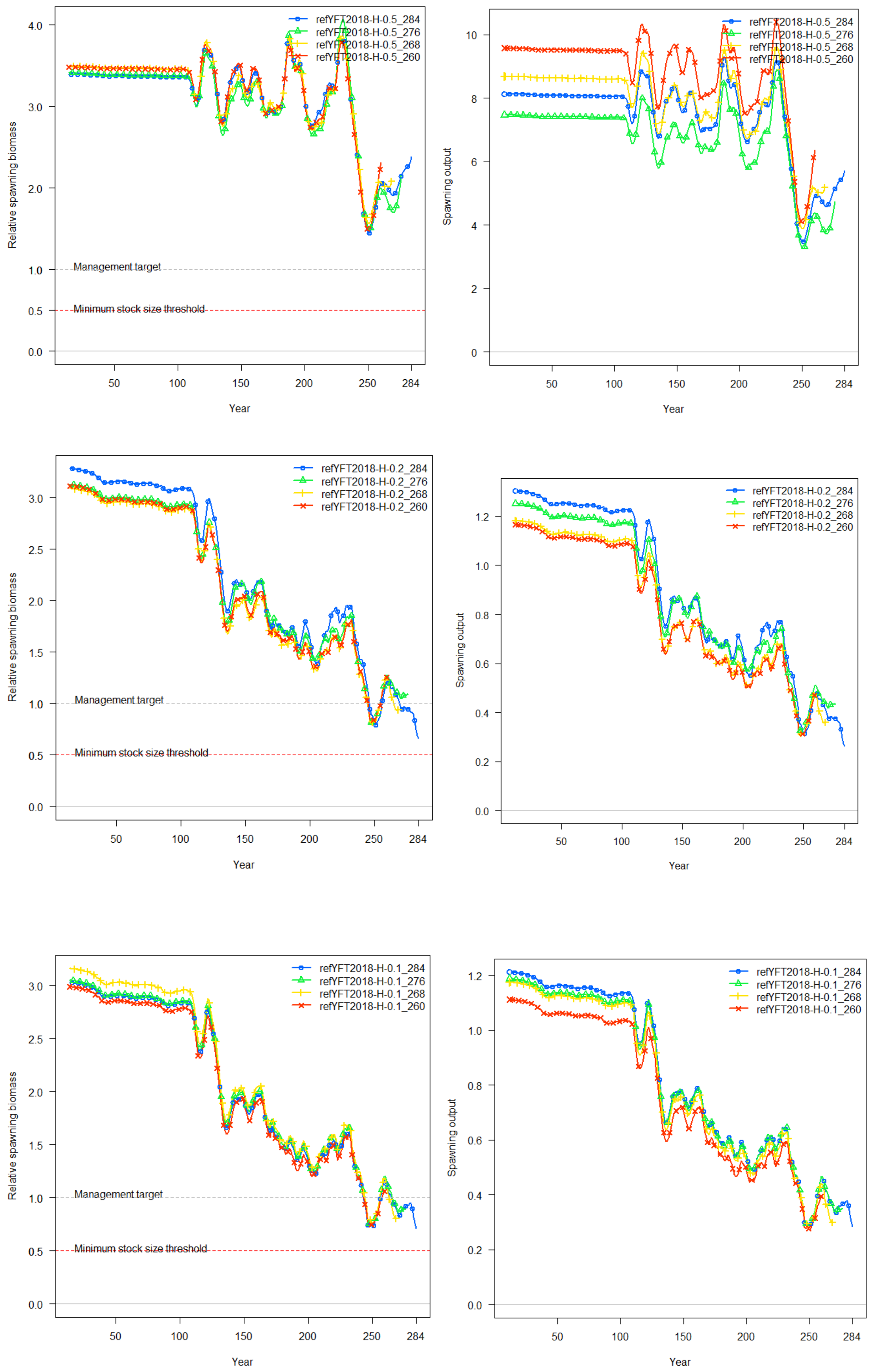


Figure 6. Retrospective yellowfin spawning biomass time series from reference case models with non-linear catchability parameter fixed at (-0.5, -0.2, -0.1, top to bottom). i.e. These all assume some degree of hyperstability.

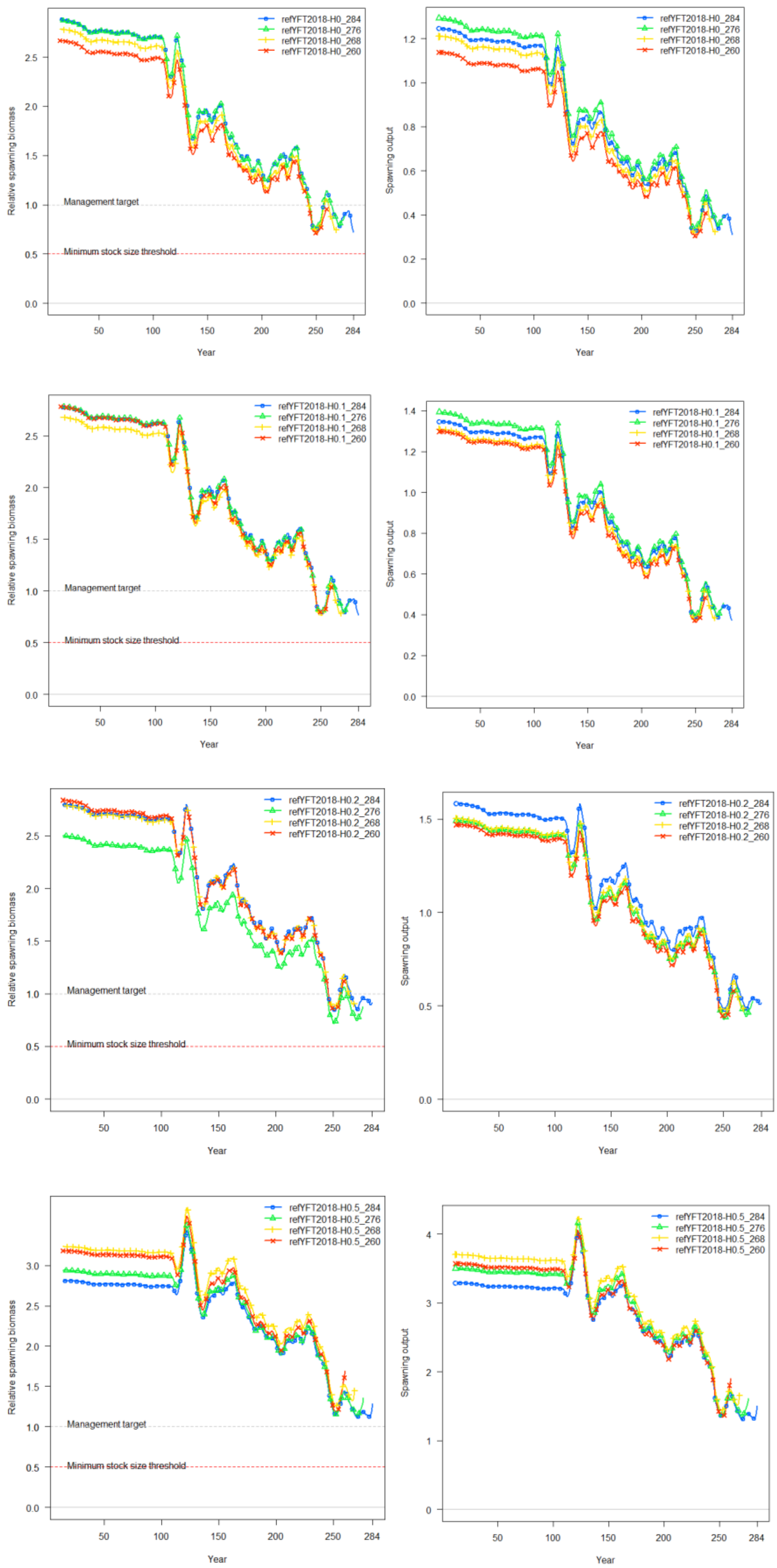


Figure 7. Retrospective yellowfin spawning biomass time series from reference case models with non-linear catchability parameter fixed at (0, 0.1, 0.2, 0.5, top to bottom). i.e. Top plot is the reference case assessment, lower three assume increasing degree of hyperdepletion.

4.2 Revisiting the Catch likelihood as a plausibility diagnostic

Many of the IOTC Stock Synthesis (hybrid F option) model configurations (for YFT and BET at least) hit the maximum fishing mortality constraint (set to 6.0 in the OM, corresponding to an exploitation rate, U of ~99.5% for the most highly selected age class). If the model cannot remove the observed catch (or not enough iterations are employed to solve the Baranov equations), this results in a “non-trivial” catch likelihood. Some SS analysts are content to implicitly interpret this as catch over-reporting observation error. However, we consider the catch likelihood to be a flag that there is probably some fundamental problem with the model, in a manner that is likely to be too pessimistic, at least in the specific time period when it is active. i.e. In assessments, we tend to believe the catch data are the most reliable data – if we believe the catch data, but the model predicts that there are not enough fish to be removed, something must be wrong. In many cases this might not much matter to the assessment (e.g. unusual environmental conditions might have caused an atypical and transient fish distribution one time 20 years ago). But we might also expect that hitting the Fmax bound could have considerable unexpected consequences on model behaviour, similar to hitting a parameter bound.

The bimodality of the catch likelihood distribution within the OM ensemble (e.g. Figure 15) is in part related to the maximum F setting, and was used as a plausibility constraint in previous OM iterations. For BET this iteration (Kolody et al 2020), the vast majority of models would fall into the righthand mode if the SS default Fmax of 2.9 was applied (and would be rejected if the 10^{-5} catch likelihood threshold was applied). Raising the BET Fmax to 6.0 moves almost all of the models to the left-hand mode. However, for YFT, the Fmax = 6.0 (max U = 99.5%) does not remove the problematic catch likelihood issue. We suggest that these models should definitely be considered implausible. A much lower threshold would probably be more appropriate, but any threshold will be arbitrary and binary (e.g. why should Fmax = 2.9 (U = 95%) be the threshold?). Alternatively, it may be possible to convert the catch likelihood into some sort of continuous weighting factor. Whatever approach is agreed in the interim, we recommend that further effort should be spent investigating fundamental issues of model structure and/or data interpretation.

Some possibilities are explored in the following, with limited success. Note that we have not considered the uncertainty in the historical catch data - this remains an ongoing topic of WPTT discussion.

4.3 Longline fishing mortality trend as a plausibility diagnostic

Related to the issue of the catch likelihood and retrospective patterns (discussed below), we thought that implausible models might be identifiable if there is a systematic deviation between the actual (standardized) effort observed in the longline fishery, and the fishing mortality estimated by the model. If the retrospectives suggest that the most recent assessment is likely to be too pessimistic, this might also correspond to recent F estimates that are trending upward faster than the standardized effort.

To test this, we examined the NW region, assuming that we could simply calculate the effective longline effort as the catch/(standardized CPUE). Figure 3 corresponds somewhat to what we predicted. i.e. There is reasonable agreement between observed effort and estimated F, for most of the time series, but there is a substantial upward trend in the ratio in recent years (i.e. black line, 48 quarter smoother). Unfortunately, this analysis is oversimplified. The LL fishery catch series differs from the catch series corresponding to the logbook CPUE standardization data set, and the CPUE consultant confirms that there are likely to be non-trivial temporal trends in the divergence (Simon Hoyle, NIWA, pers. comm.).

It was never clear whether this analysis would yield a result that was fundamentally different from the systematic lack of fit to the CPUE series, and the primary reason for presenting it now relates to the curious ratio of the two series. There appears to be an ~4 year cycle in the ratio (e.g. Figure 3, loess smoothers of 4-19 quarters). This does not appear to be an artefact of the loess smoother. We are not aware of any oceanographic process with this periodicity, nor could the CPUE consultant think of any obvious mechanism or analytical artefact that would introduce this cycle to the standardized CPUE series (Simon Hoyle, NIWA, pers. comm.). It might be random noise, but we thought it might stimulate some interesting discussion.

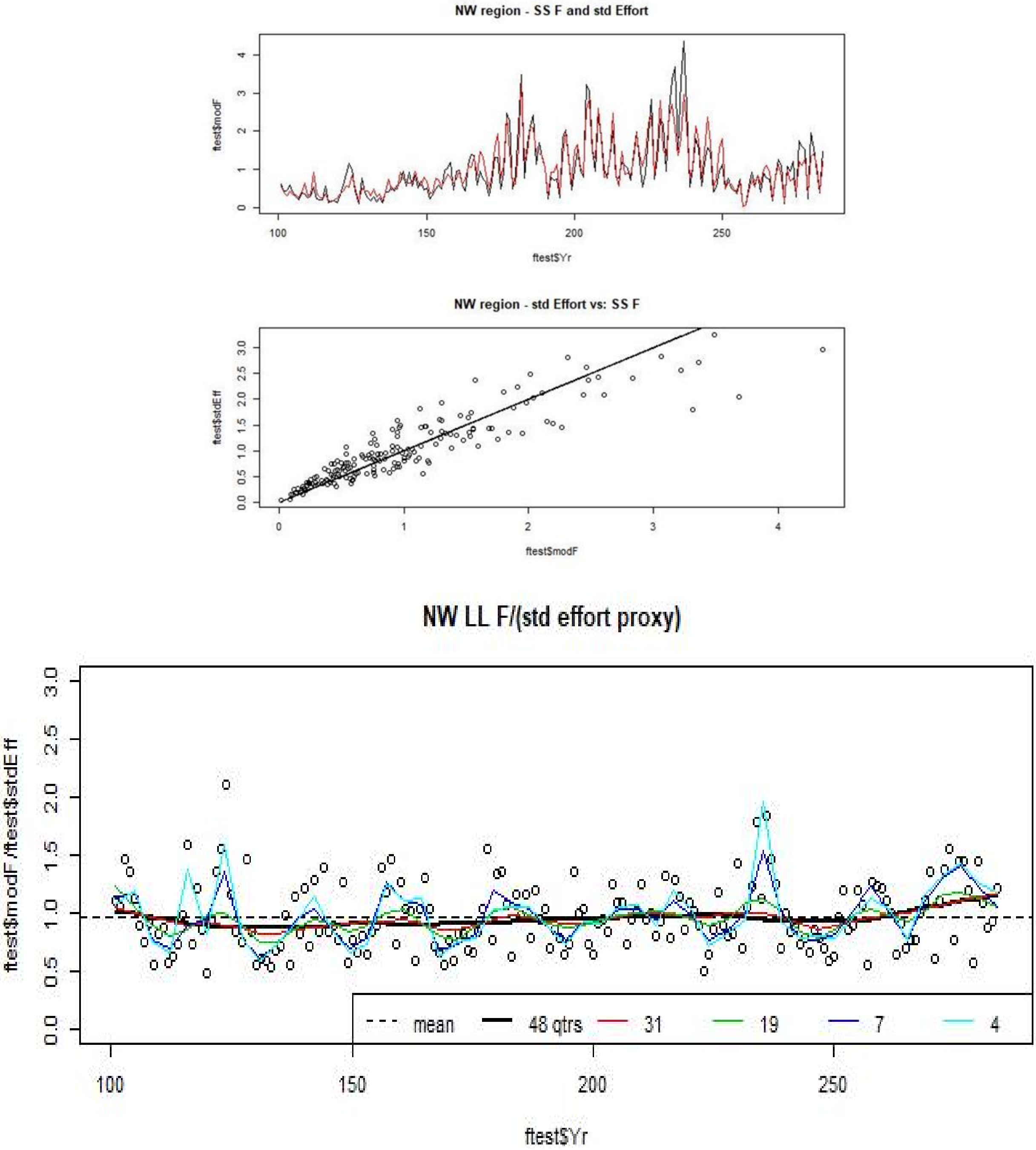


Figure 8. Comparison of ref2018 Northwest LL estimated fishing mortality and the standardized longline effort (as defined in the text). Time series (top), scatterplot (middle), and ratio of time series (bottom). In the bottom panel, numbered series represent loess smoothers with the indicated "span" argument, X-axis represents quarterly observations.

4.4 Revisiting the Environment-Movement Link from the Stock Assessment

In previous versions of the YFT OM, we did not use the environment-movement link in the OM (unlike the assessment), because i) it did not seem to make very much difference, and ii) it is not obvious what to do with projections (and hence could be a large effort for questionable gain). However, there seem to be strong seasonal patterns in the system (i.e. recruitment, movement, and/or catchability) and this could be one mechanism causing the catch likelihood problem (i.e. the fish might not be distributed properly). If the failure to remove the observed catches is an artefact of failing to represent seasonal movement, this may constrain the model in various unexpected ways. While the original environment-movement mechanisms proposed in the assessment sounded plausible, we do not consider that the relevant physical processes have been firmly established. But we recognize that they do provide a mechanism for testing seasonality in a non-seasonal SS model structure.

Figure 9 - Figure 11 show that the reference case OM (*NoEnv*, without environmental links) is qualitatively similar, but somewhat more pessimistic than, the reference case assessment (*ref2018*). However, if the assessment environmental time series are replaced with seasonally averaged time series (*MeanSeasEnv*), the dynamics appear extremely similar. There would be at least two advantages to using seasonal averages over the original data: i) they could be used in projections, and ii) there would be no missing value problems. Given that we can largely reproduce the effects of the environment-movement link using only the repeating seasonally-averaged data, we opted to run a full (144 model) grid (OMgridY20.2). However, comparing OMgridY20.2 (seasonal-averages) and OMgridY20.1 (non-environmental) OMs, we see very little difference in the distribution of stock status characteristics (Figure 12 - Figure 14), and the catch likelihood distributions are similarly bimodal (Figure 15).

The SS models (with seasonal or interannual plus seasonal environmental links) do not result in large seasonal changes in abundance by region that approach the apparent seasonal variability in the CPUE series (e.g. compare Figure 11 and Figure 16). It remains possible that the mechanism is important, but the data resolution may not be adequate, e.g. model quarters might not align consistently with the fish and/or fishery movements, or the environmental data might not be the best choice. As tested, there does not appear to be any net benefit to retaining the seasonal movement.

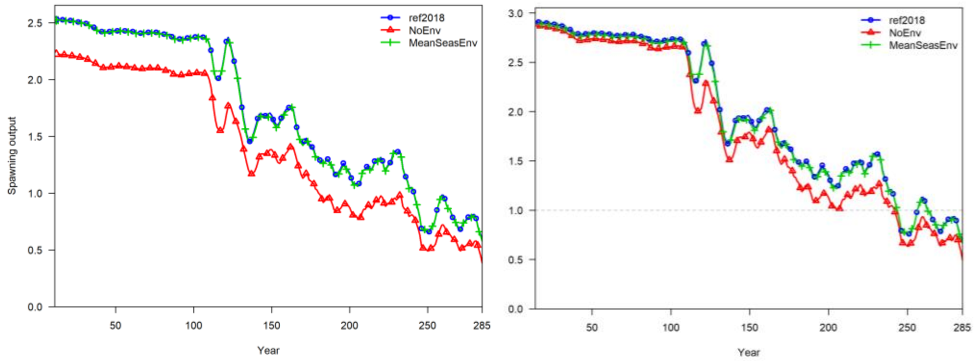


Figure 9. Comparison of the absolute (left) spawning output and relative to SBMSY (spawning biomass) (right) from the reference case (ref2018, blue), using no environmental correlates (NoEnv, red) and using the seasonal mean of environmental correlates (MeanSeasEnv, green).

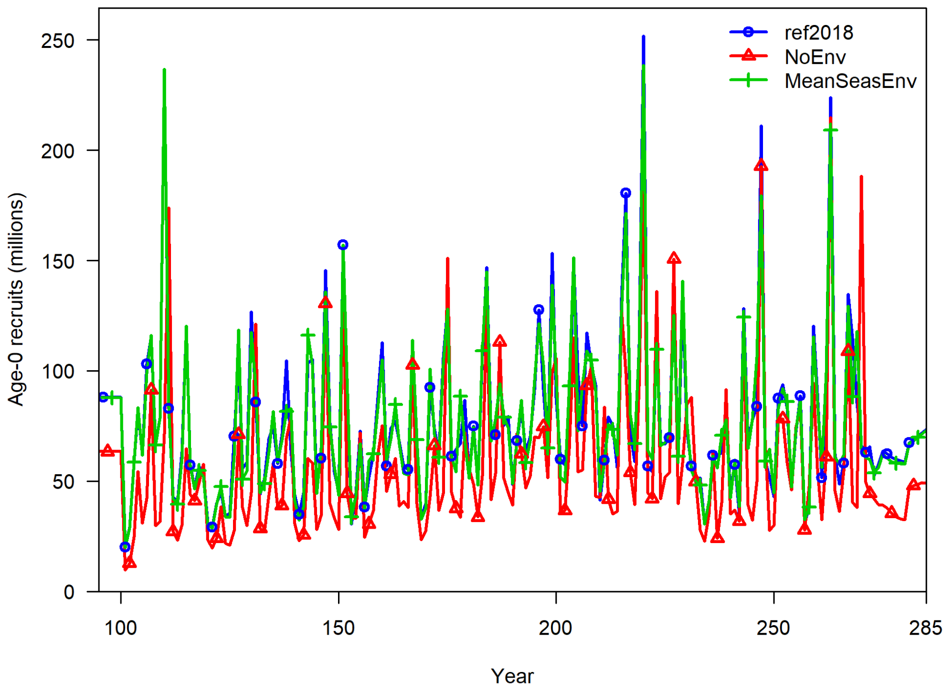


Figure 10. Comparison of the recruitment deviations from the reference case (ref2018, blue), using no environmental correlates (NoEnv, red) and using the seasonal mean of environmental correlates (MeanSeasEnv, green).

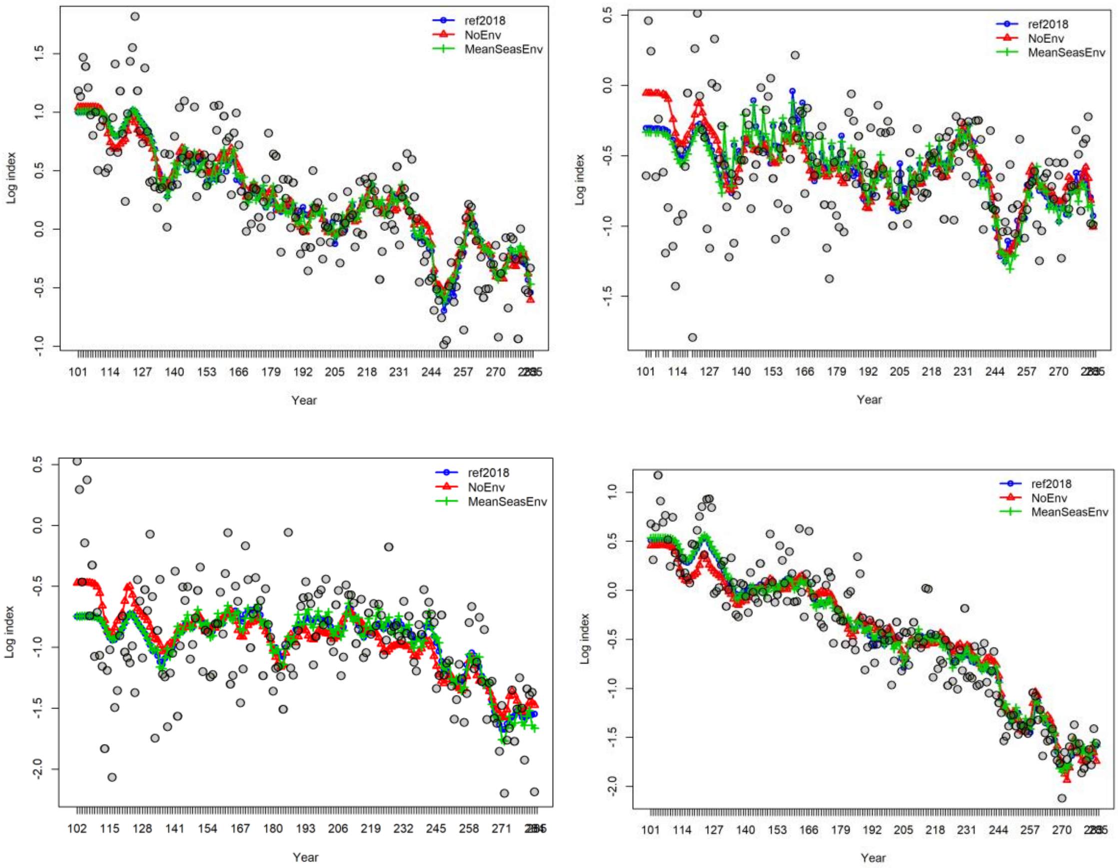
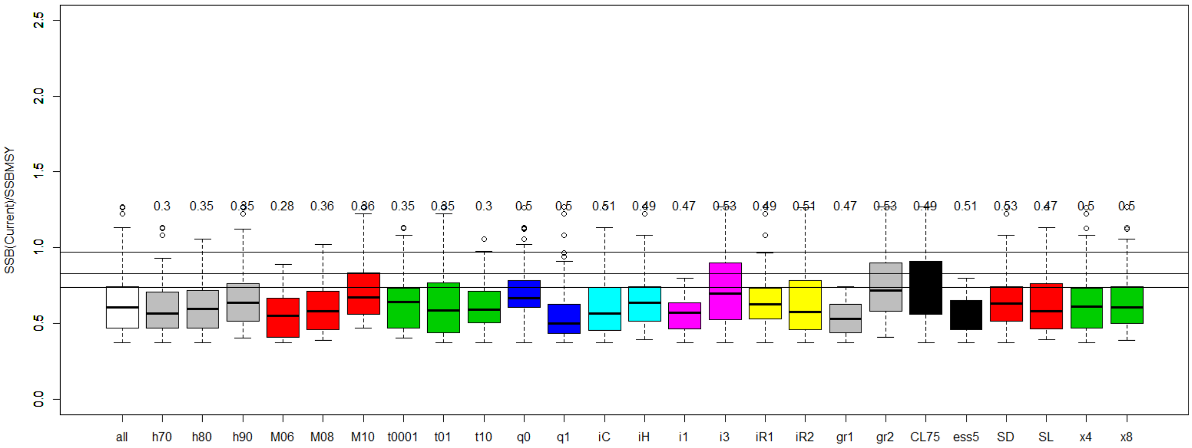


Figure 11. Fits to the log CPUE for fleet 26 (top left), 27 (top right), 28 (bottom left) and 29 (bottom right) for the reference case (ref2018, blue), using no environmental correlates (NoEnv, red) and using the seasonal mean of environmental correlates (MeanSeasEnv, green).



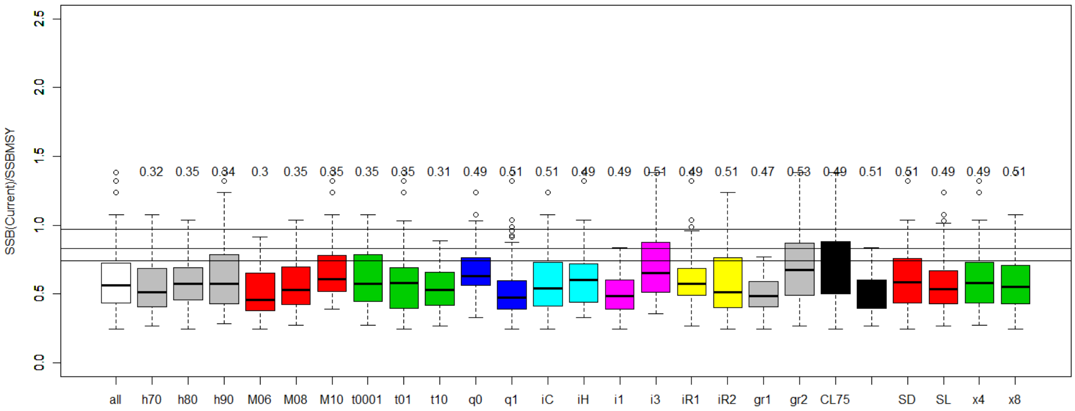
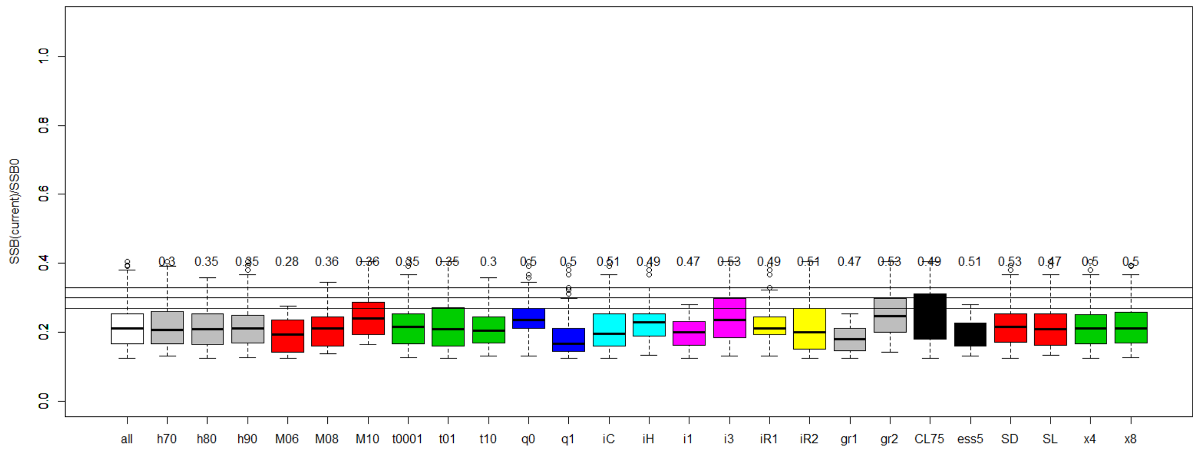


Figure 12. Comparison of depletion relative to SSB(MSY) from (converged models in) OMgridY20.1 (top, without any environmental links) and OMgridY20.2 (bottom, with seasonal environmental-linked movement). Boxplots partitioned by grid options.



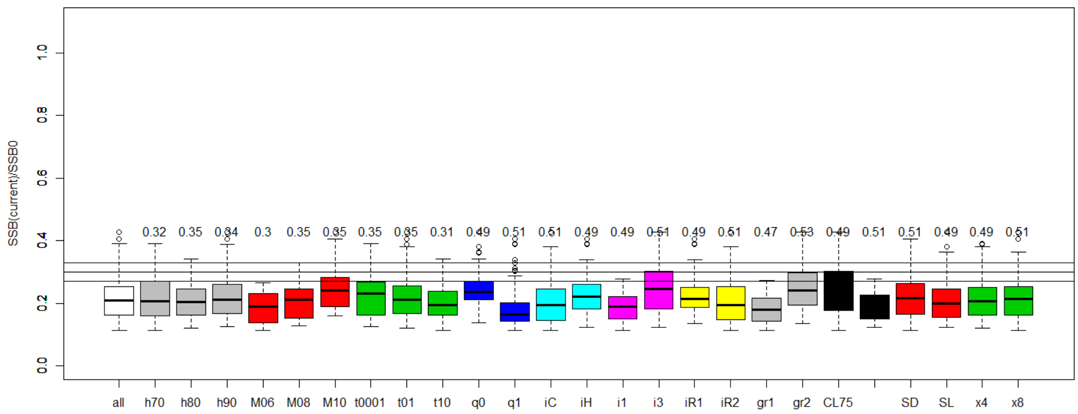
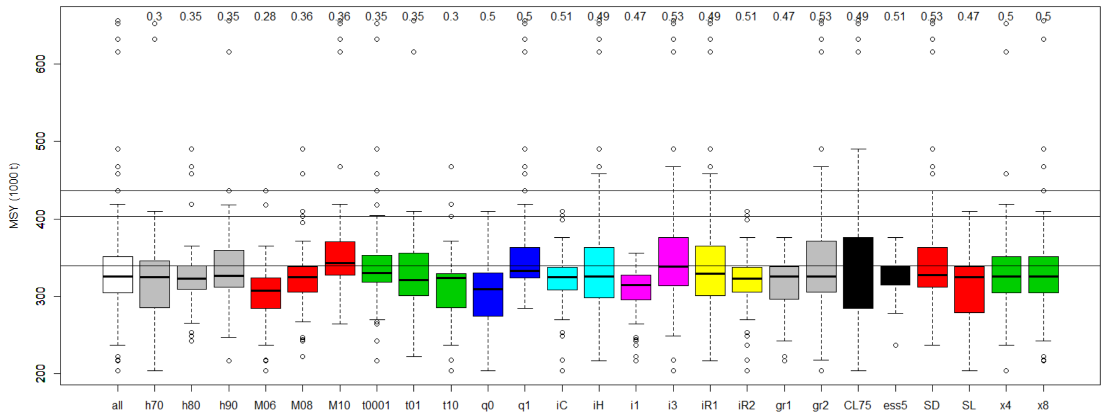


Figure 13. Comparison of depletion relative to B0 estimates from (converged models in) OMgridY20.1 (top, without any environmental links) and OMgridY20.2 (bottom, with seasonal environmental-linked movement). Boxplots partitioned by grid options.



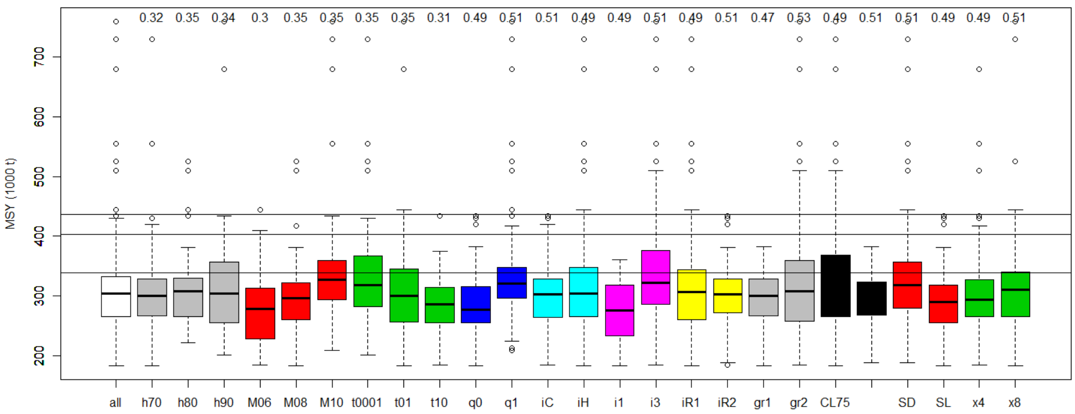
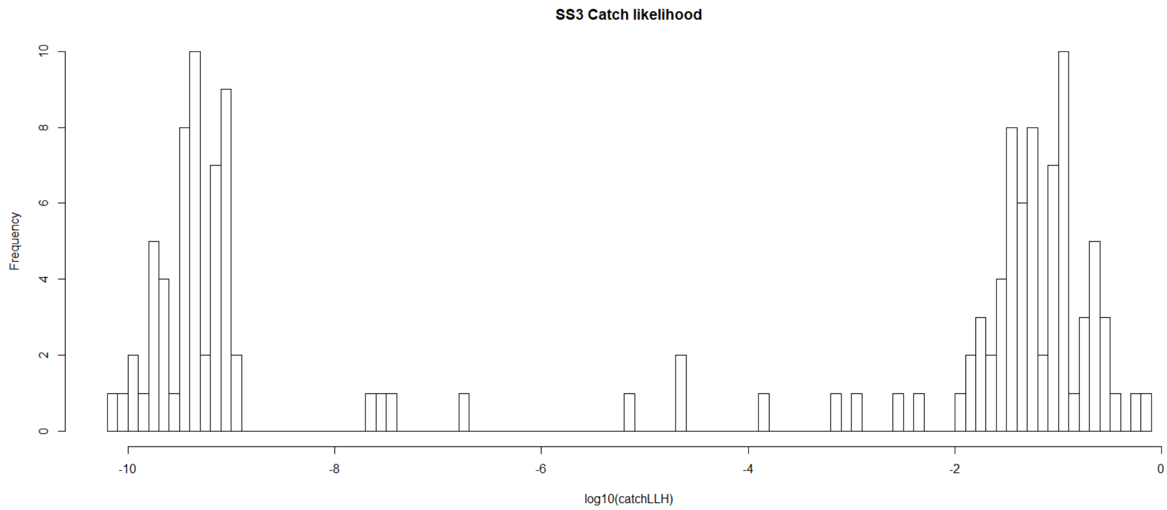


Figure 14. Comparison of MSY estimates from (converged models in) OMgridY20.1 (top, without any environmental links) and OMgridY20.2 (bottom, with seasonal environmental-linked movement). Boxplots partitioned by grid options.



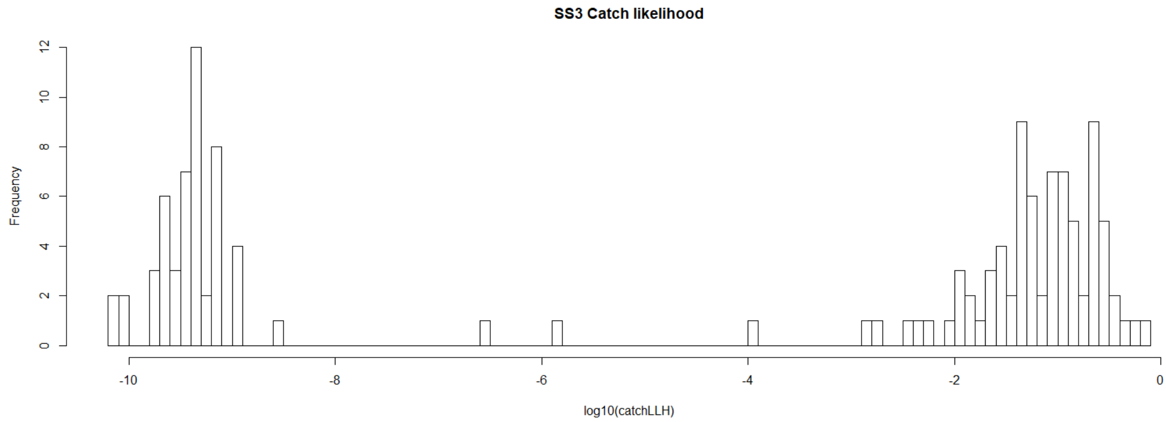
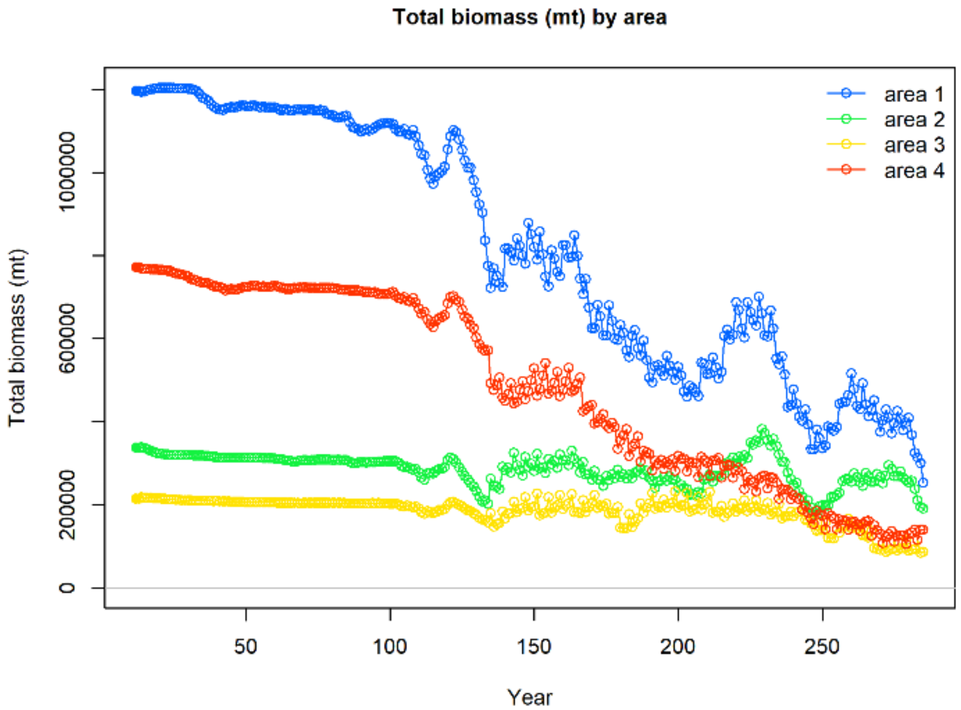


Figure 15. Comparison of catch likelihood distributions from (converged models in) OMgridY20.1 (top, without any environmental links) and OMgridY20.2 (bottom, with seasonal environmental-linked movement).



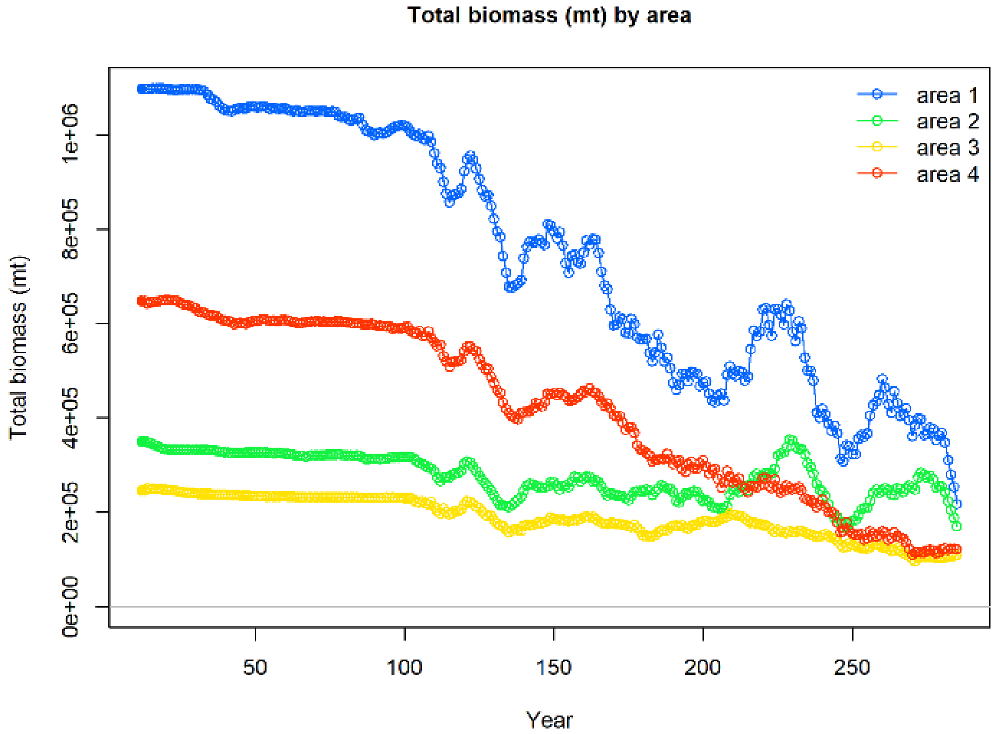


Figure 16. Total biomass by region for model h80_M10_t10_q0_iH_i1_iR1_gr1_CL75_SL_x4 with (top) and without (bottom) seasonal migration.

4.5 Revisiting Iterative Reweighting of data

Given the large amounts, varying quality, and disparate types of data in the yellowfin assessment, combined with the structural simplifications that are required to produce tractable estimators, it is inevitable that different data sources will influence model inferences in different ways (i.e. resulting in conflicts). Iterative reweighting (sometimes called re-scaling or tuning, the latter of which is not to be confused with "MP tuning") is a method of improving the consistency between variance-related assumptions and the model quality of fit to the data. Most of the indices (CPUE, surveys and composition data) used in fisheries underestimate their true variance by only reporting measurement or estimation error and not including process error. The variance-related assumptions include the recruitment deviation penalty, CPUE observation error, size composition sample sizes and tag-related assumptions, etc. If the internal consistency of the model is improved, it is expected that the statistical properties of the model will be improved. However, the approach considered here seeks a first order improvement, and does not address issues of systematic lack of fit (i.e. due to model specification errors and/or data biases).

The intent of this exploration was i) to see if assessment results were sensitive to different versions of iterative reweighting (recognizing that there is no uniquely best method), and ii) to see whether there is any merit in considering an automated approach to iterative reweighting in the context of the IOTC OMs.

4.5.1 Iterative reweighting method

We explored a repeatable iterative reweighting procedure recommended as current best practice in the Pacific Fishery Management Council (2018), and which has been adopted in other jurisdictions. The specific process involved:

1. Start with the standard error for the log of relative abundance indices (CPUE) as specified in the input file. Stock Synthesis then allows an estimate to be made for an additional adjustment to the relative abundance variances appropriately. This is a free parameter that can be estimated any time and does not require multiple iterations of reweighting.

An iterative fitting procedure was used for the remaining adjustments. For the recruitment bias adjustment ramps:

2. Adjust the maximum bias adjustment and the start and finish bias adjustment ramps as predicted by r4ss after running Stock Synthesis at each step (Methot and Taylor 2011).

For the length composition data:

3. After model fitting, SS (r4ss) exports a CL sample size (multiplier) that can be used to calculate the CL sample size that would be consistent with the model predictions. The 'Francis method' (Francis, 2011), calculates a separate multiplier for each fleet (i.e. applies to all years equally). If the multipliers are very different from unity, this suggests that the model fits the data better or worse than expected given the input sample sizes. The model fitting was predicated on the previous input sample size, so adopting the new input sample size and rerunning the model will result in a new set of multipliers, and the process may need repeating until it converges adequately.
4. Repeat steps 2 - 3, until all are converged and stable (with proposed changes < say 2%).

We note that this process should include the tagging data as well, but this option is not incorporated into any of the automated diagnostics currently available in Stock Synthesis or r4ss and would require further development work. However, the expectation is that this would probably not be very influential in this case, because the tag weighting did go through an analogous process in an earlier iteration of the assessment, with the assumptions retained as the default (and other elements of the assessment have not substantively changed in the meantime).

4.5.2 Iterative reweighting the 2018 yellowfin tuna reference case

The reference 2018 yellowfin tuna assessment (ref2018, blue) was only partly iteratively reweighted using the procedure described above due to the long run times involved, with the expectation that subsequent steps would only make minor changes to the results.

The impact of partial iterative reweighting is shown in two plots of relative biomass (compare the red time series in Figure 17 with the blue time series), with the plot on the left showing absolute spawning output and the plot on the right rescaled relative to SBMSY (spawning biomass).

Differences in the estimates of recruitment events are shown in Figure 18 with differences in the fits to CPUE (in log scale) shown in Figure 19.

From these results we tentatively conclude that our approach does not offer any new insight over the approaches that were explicitly or implicitly used in the assessments in the past. The impact on stock status (point estimate) inferences was small, particularly relative to uncertainty introduced from other elements of the OM grid. And we cannot conclude that this method is preferable, particularly since it does not include the tags (though documented reproducibility is probably desirable in whatever approach is used). The approach did confirm that the CPUE is probably modestly over-weighted in the assessment and OM. We consider this to be intentional and desirable, in keeping with the general principle that relative abundance data are usually the most informative and need to be well fit, i.e. if you cannot fit the CPUE, or you do not believe the CPUE, it is questionable what your assessment can provide. Toward this same goal, the CL data were generally not over-fit.

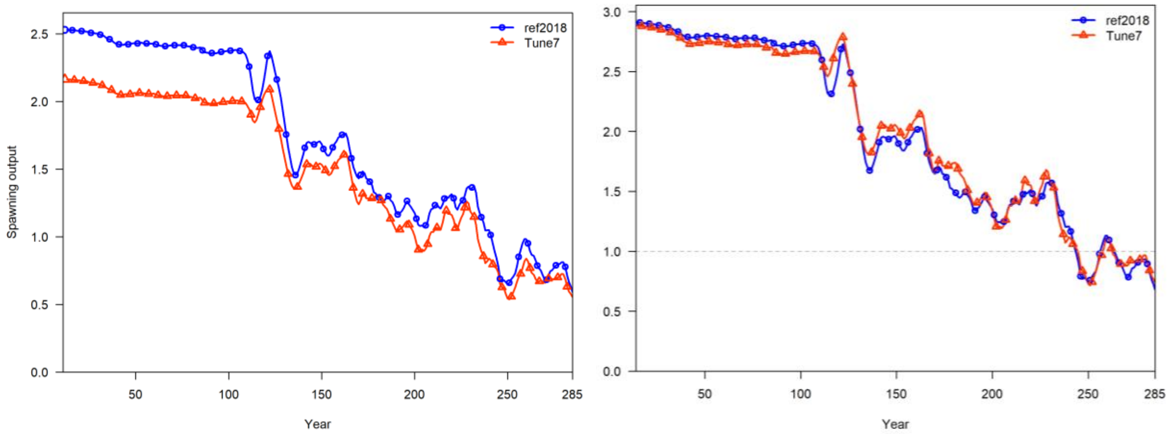


Figure 17. Comparison of the absolute (left) spawning output and relative to SBMSY (spawning biomass) (right) from the reference case (ref2018, blue) and from the partially tuned model (Tune 7, red).

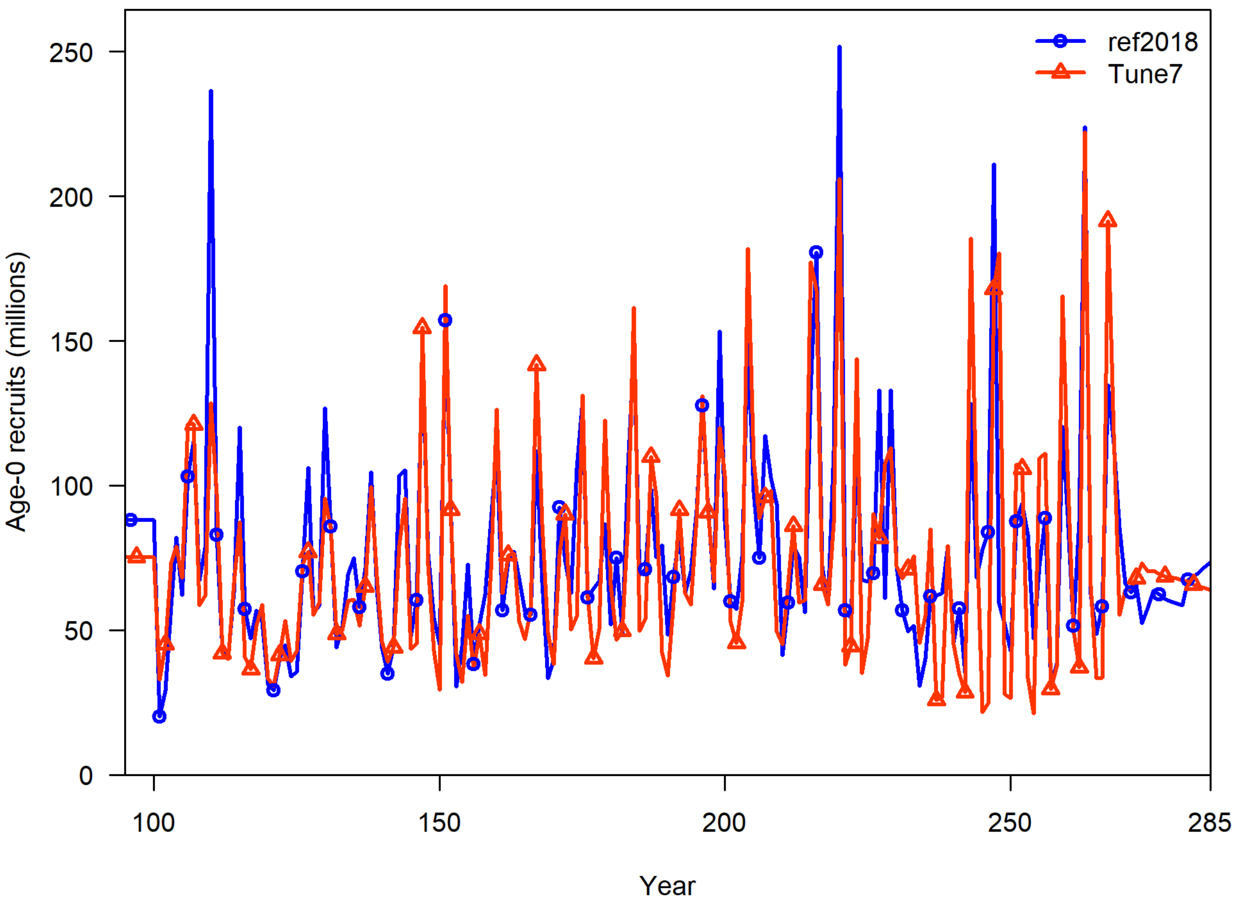


Figure 18. Comparison of the recruitment deviations from the reference case (ref2018, blue) from seven steps in iteratively reweighting the model (Tune 7, red).

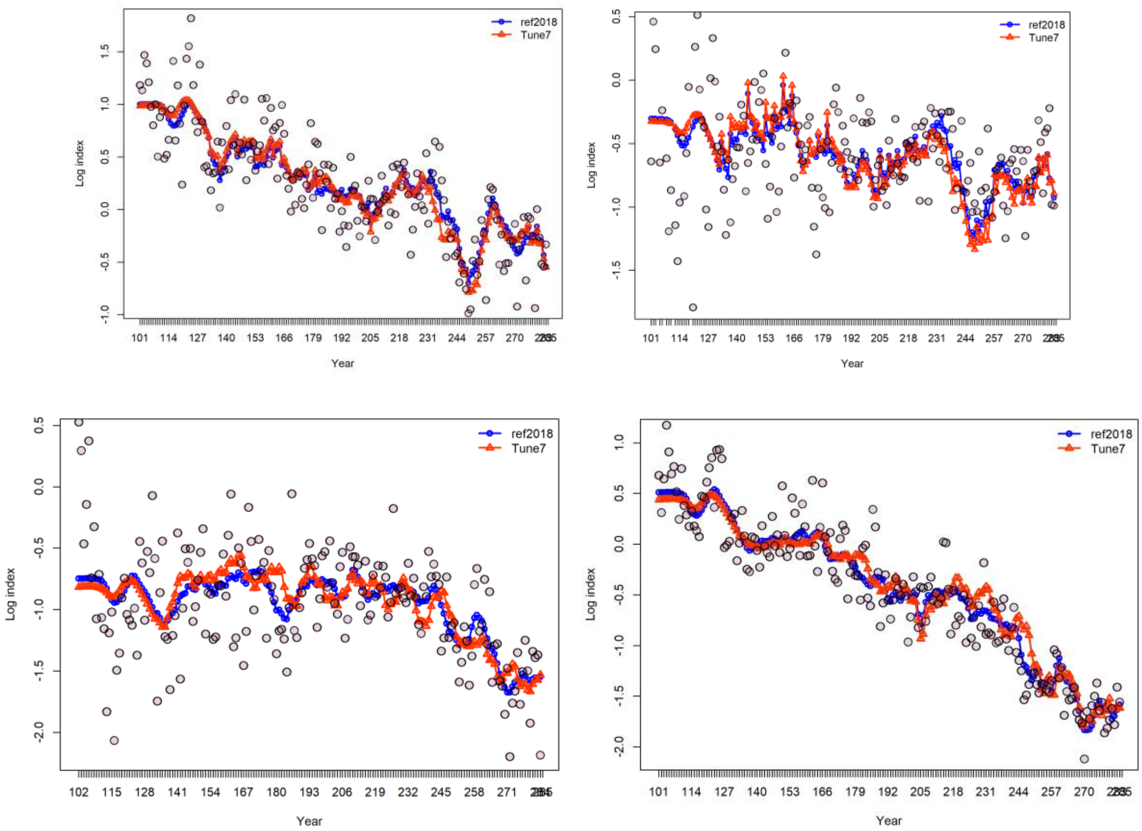


Figure 19. Fits to the log CPUE for fleet 26 (top left), 27 (top right), 28 (bottom left), and 29 (bottom right) for the reference case (ref2018, blue) and the partially reweighted model (Tune7, red).

5. Yellowfin reference set Operating Model OMrefY20.1

5.1 Characteristics of the conditioned YFT OM

At this time, what we are tentatively proposing as the YFT reference set OM (OMrefY20.1) is not much different to the previous iteration – the main differences are

- re-introduction of the intermediate tag weighting option (t01)
- increasing the period of recent recruit constraints from 4 to 12 quarters
- revised approach for handling the CPUE observation error (which probably has a trivial effect in the current OM).

Several graphical summary diagnostics are provided below, for the subset of 133 converged models from the 144 model OMgrid20.1. To be consistent with previous iterations, we would have undertaken further steps, including:

- Additional iteration of bounds checking and relaxation/re-fitting as required
- Detailed inspection of individual models at the “corners” (i.e. highest and lowest B/B(MSY), highest and lowest MSY) for strange behaviour.

The stock status summaries associated with OMrefY20.1 are presented in Figure 12-Figure 14, from which we note a couple key points:

- Results are qualitatively similar to previous iterations
- Stock status is more uncertain and generally more pessimistic than the assessment. As would be expected, the CPUE catchability trend appears to be the most influential assessment option affecting stock status (and regardless of the catchability trend, higher relative weighting of the CPUE data appears to support more pessimistic outcomes).

CPUE fit summaries indicate:

- The fit to the observed (annual) CPUE (Figure 21) is usually better as good as we could hope (e.g. regional annualized median RMSE < 20%, max ~32%), while the fit to the (regionally-aggregated, annual) MP CPUE is very good (RMSE < 20%) (Figure 2).

Figure 20 shows that the historical recruitment deviation time series does not have consistent long term trends (this is indexed for individual models in Figure 23). The size composition data fits are similar to previous iterations (Figure 22). Interactions among model fit indices, assumption levels and stock status indicators are shown in Figure 23.

Figure 24 shows stock status characteristics for OMrefY20.1 (the 58 model OM ensemble that remains if the 10^{-5} catch likelihood is applied to OMgridY20.1). Figure 25 shows the equivalent plots from OMrefY19.4 (the last iteration of the YFT OM presented to the WPTT 2019). The pooled distributions are similar, but the effects of some of the individual factors differ more than might be expected. We speculate that this is most likely due to the differing fractional factorial designs coupled with the differential rejection of models with certain factors due to the differing interactions. Notably the high CPUE weight (CV=0.1) is the most rejected assumption option – in the current framework it seems difficult to reconcile the catch and CPUE.

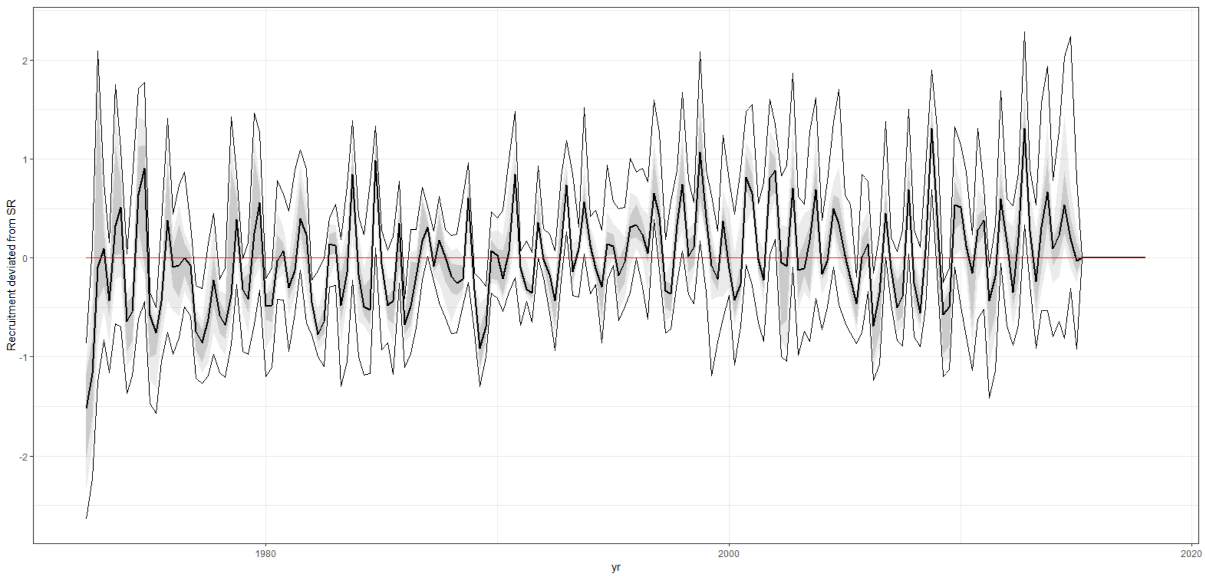


Figure 20. OMrefY20.1 recruitment deviations showing 0, 10, 50, 75, 90 and 100th percentiles

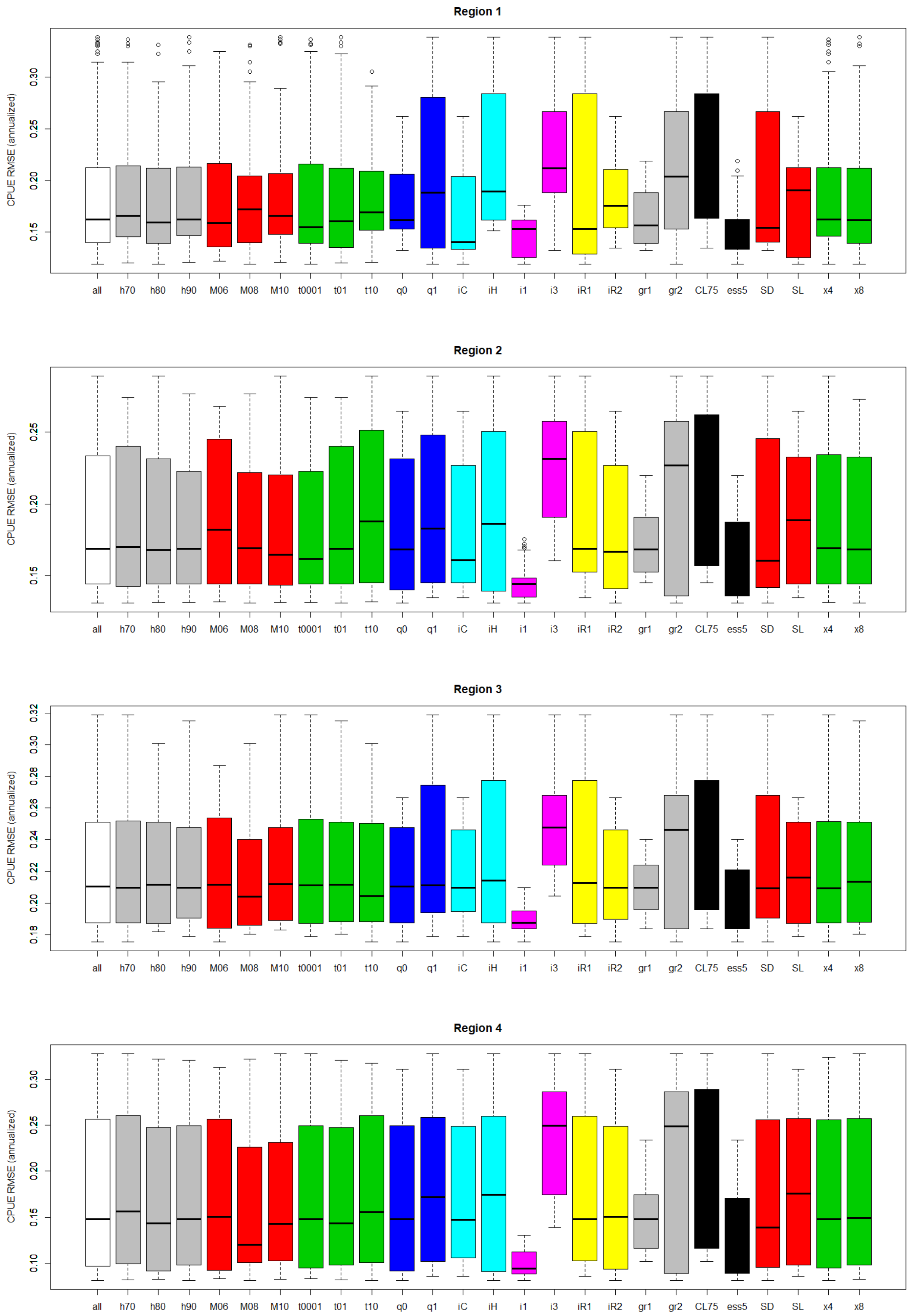


Figure 21. OMrefY20.1 agreement between the predicted and observed CPUE by region (annualized RMSE), marginalized over individual grid assumptions.

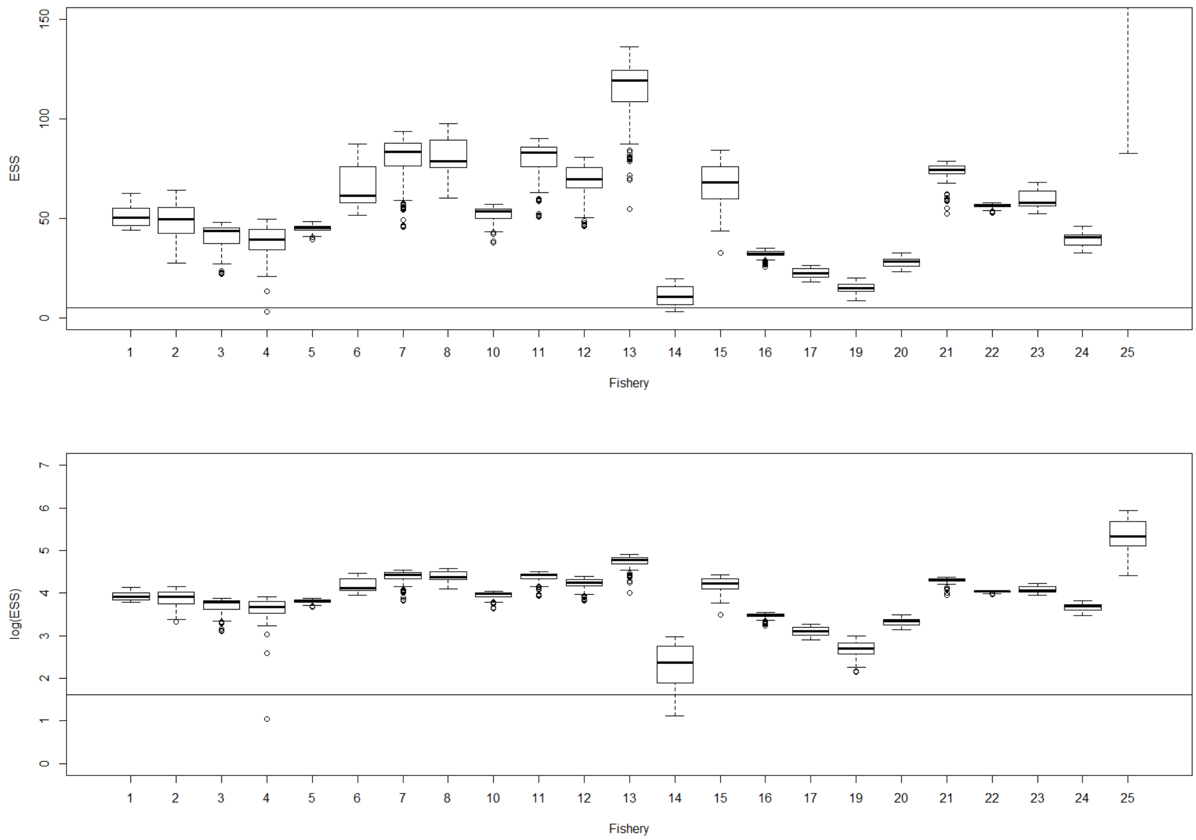
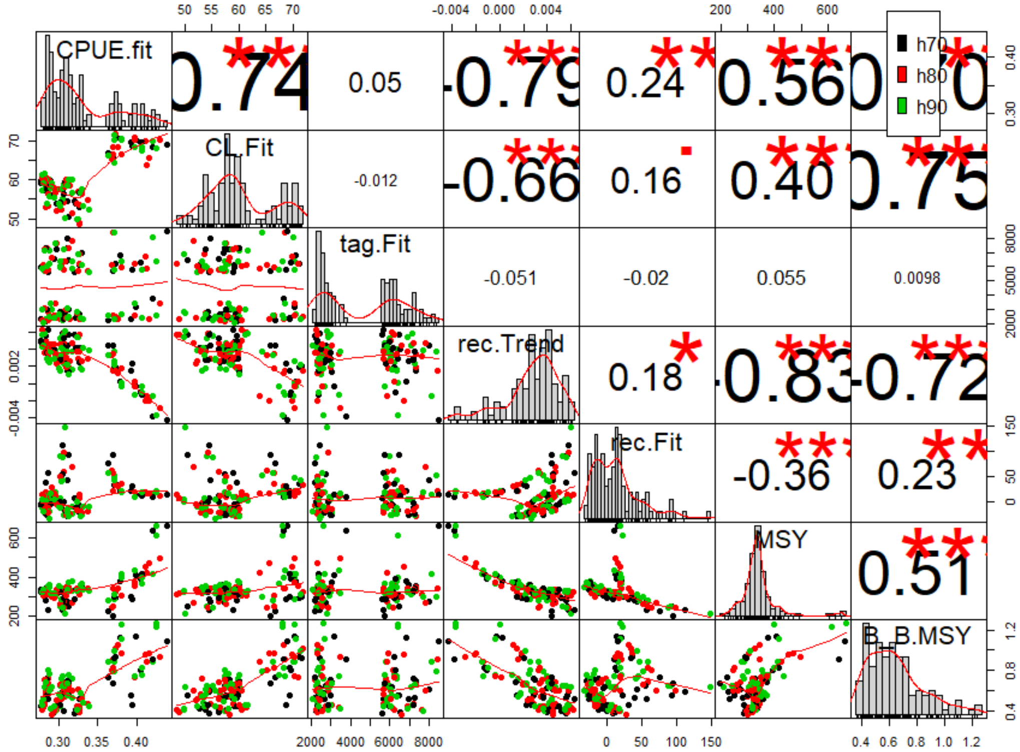


Figure 22. OMrefY20.1 fit to the size composition data by fishery. Each box represents the distribution across all models in the grid.



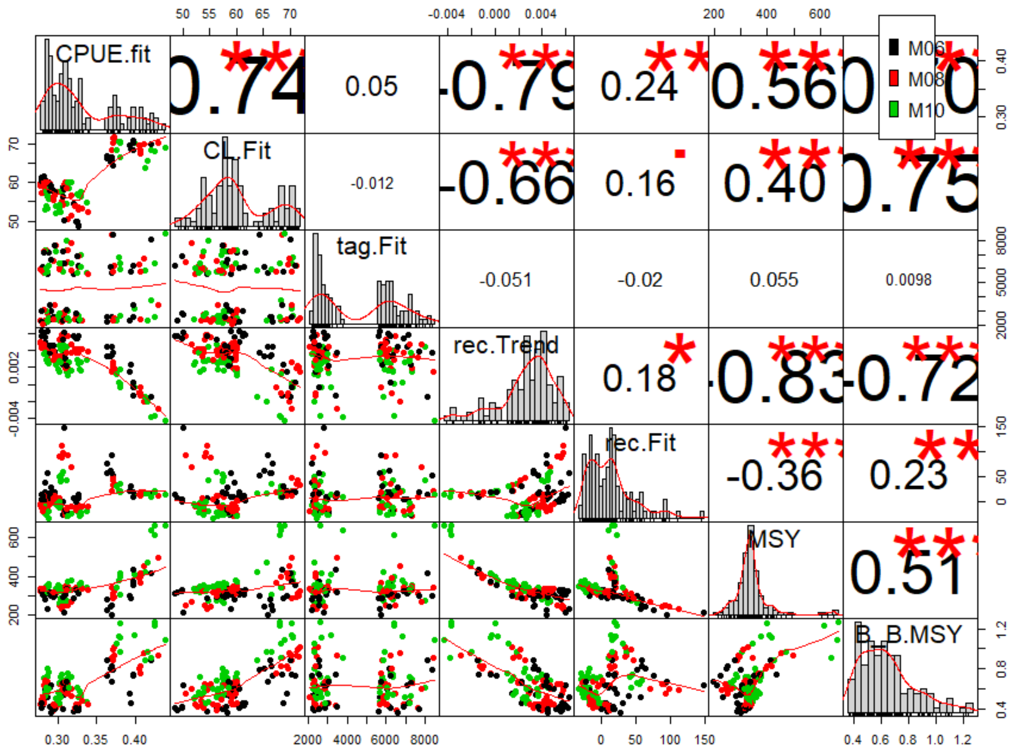
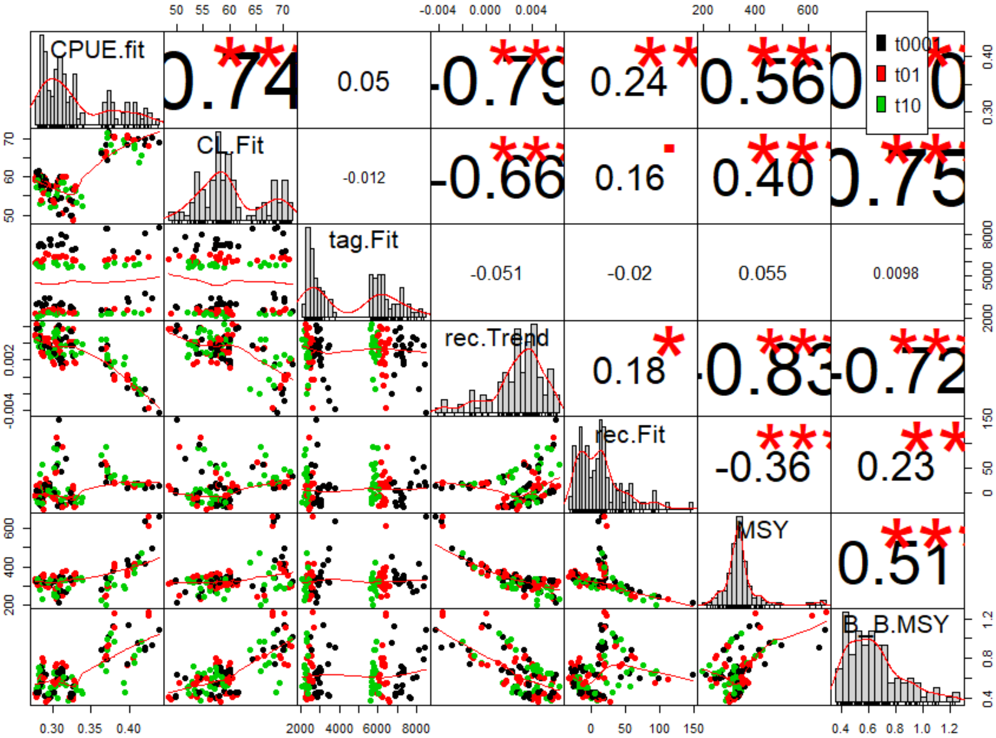


Figure 23. Multiway comparison of OMrefY20.1 model characteristics, partitioned by assumption by colour.



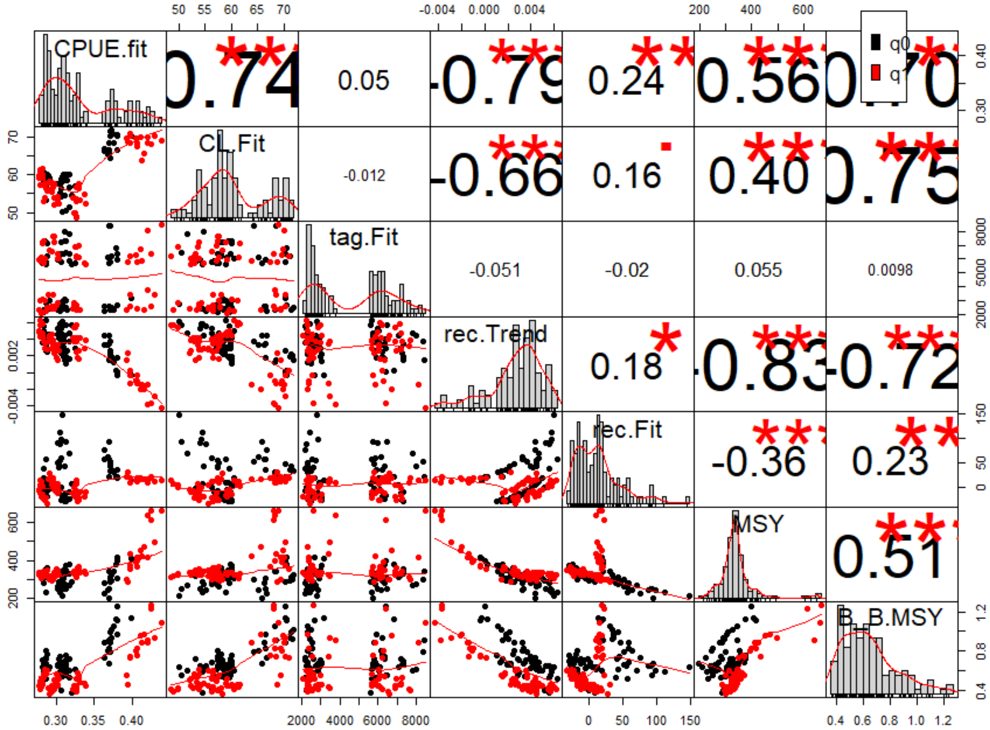
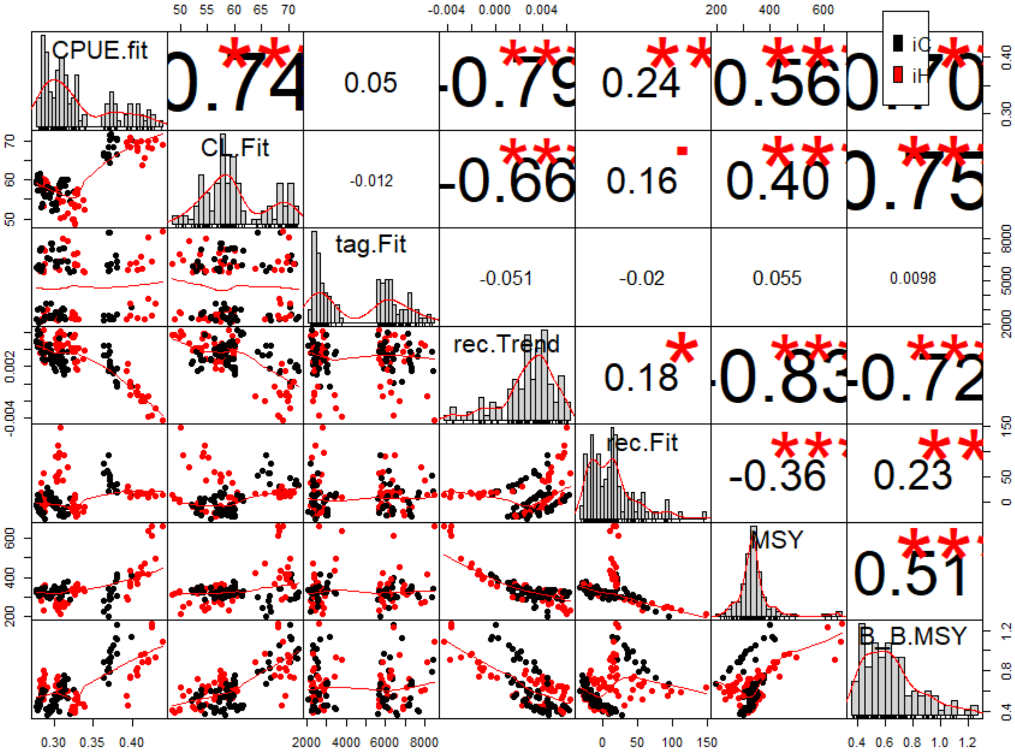


Figure 23 (cont.)



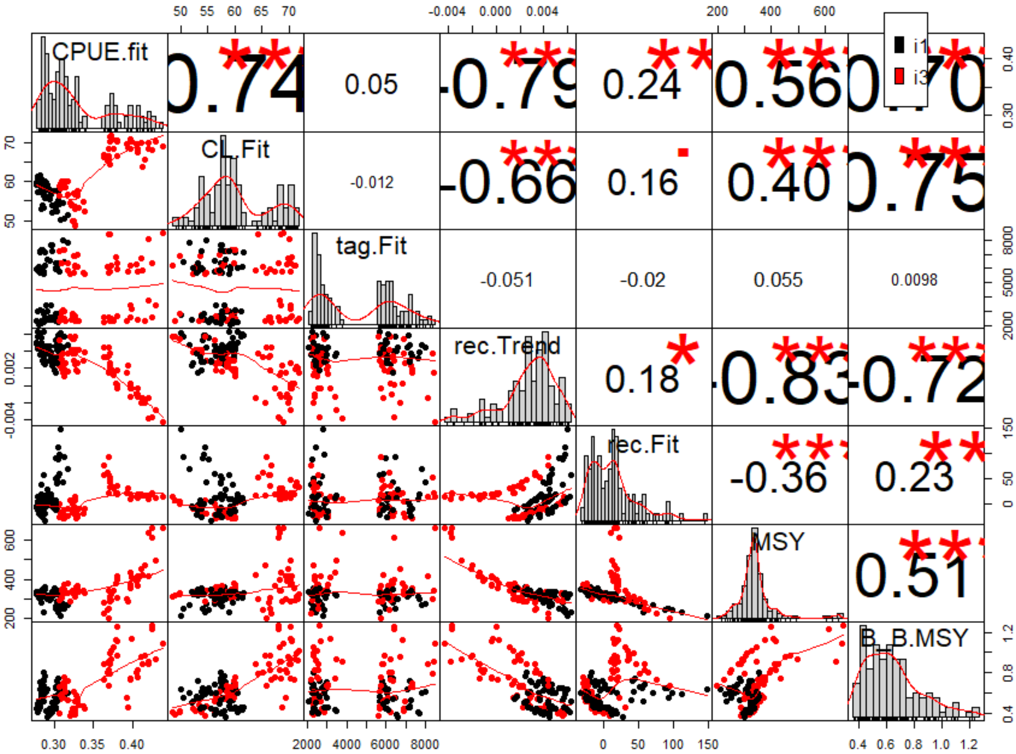
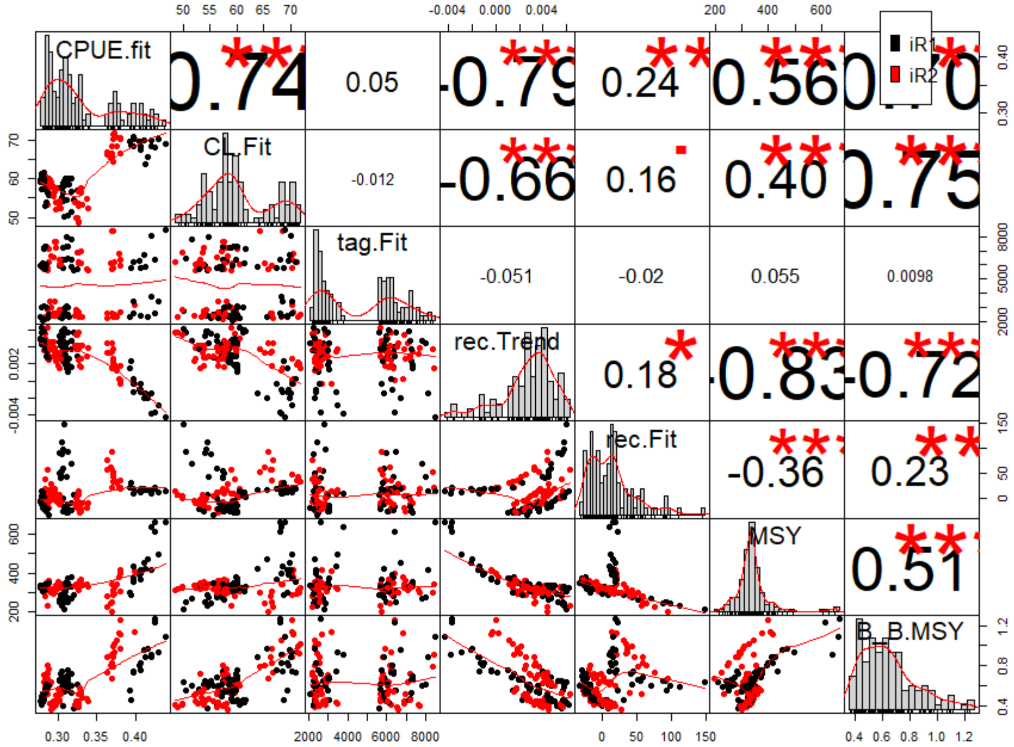


Figure 23 (cont.)



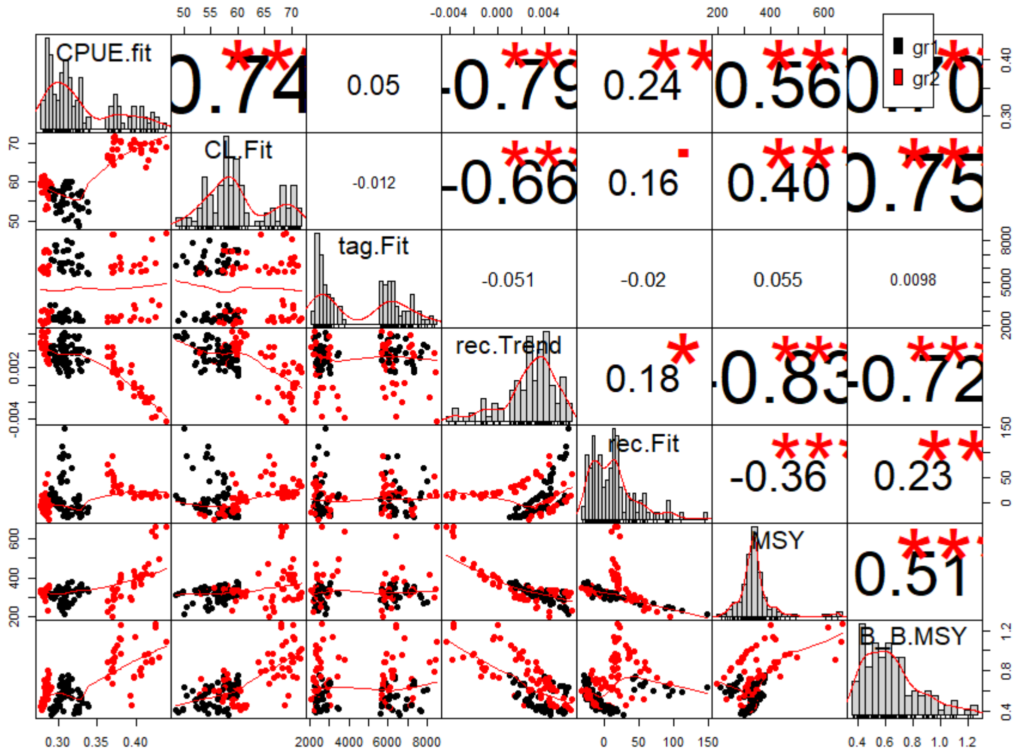
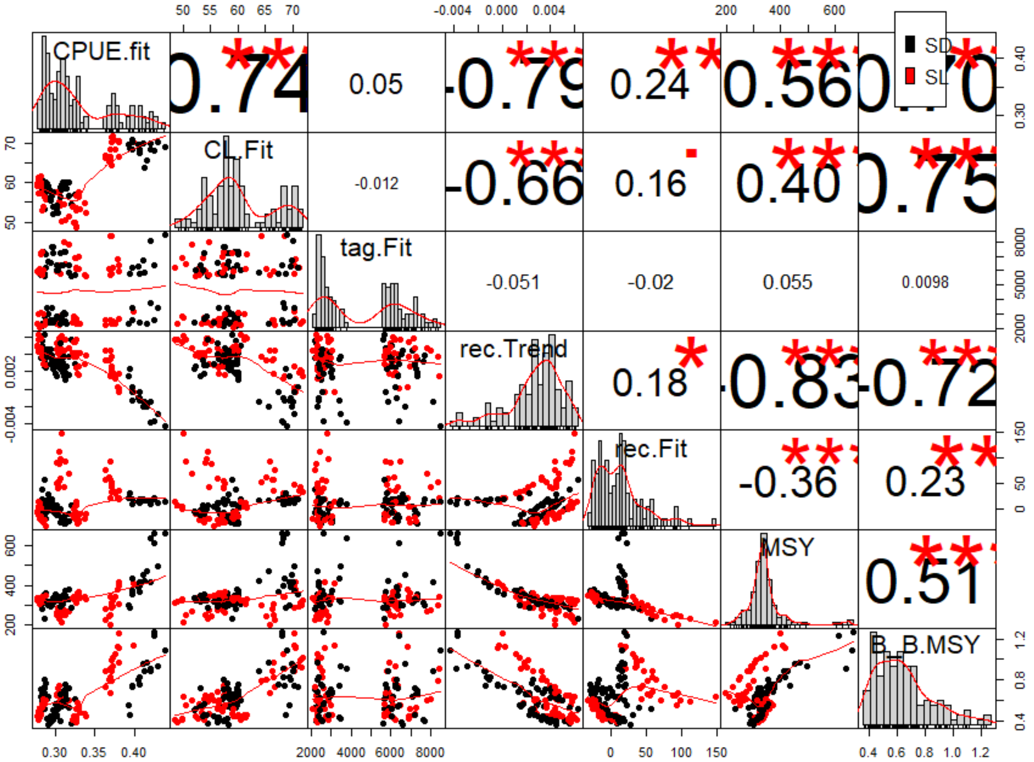


Figure 23 (cont.)



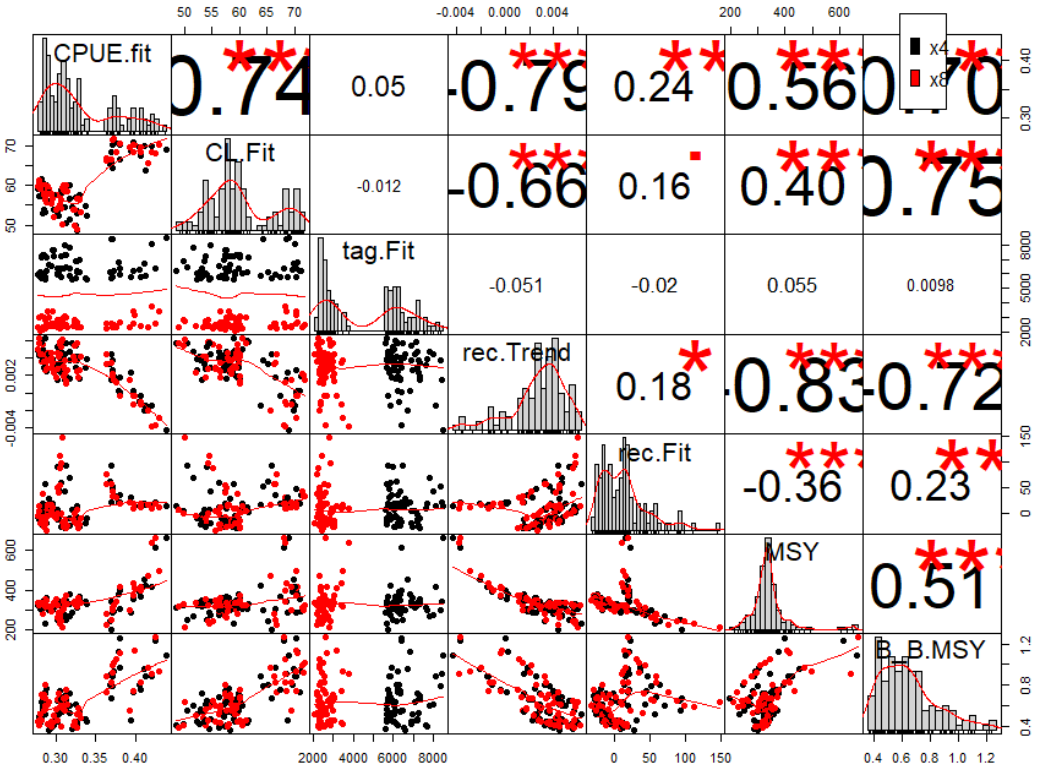


Figure 23 (cont.)

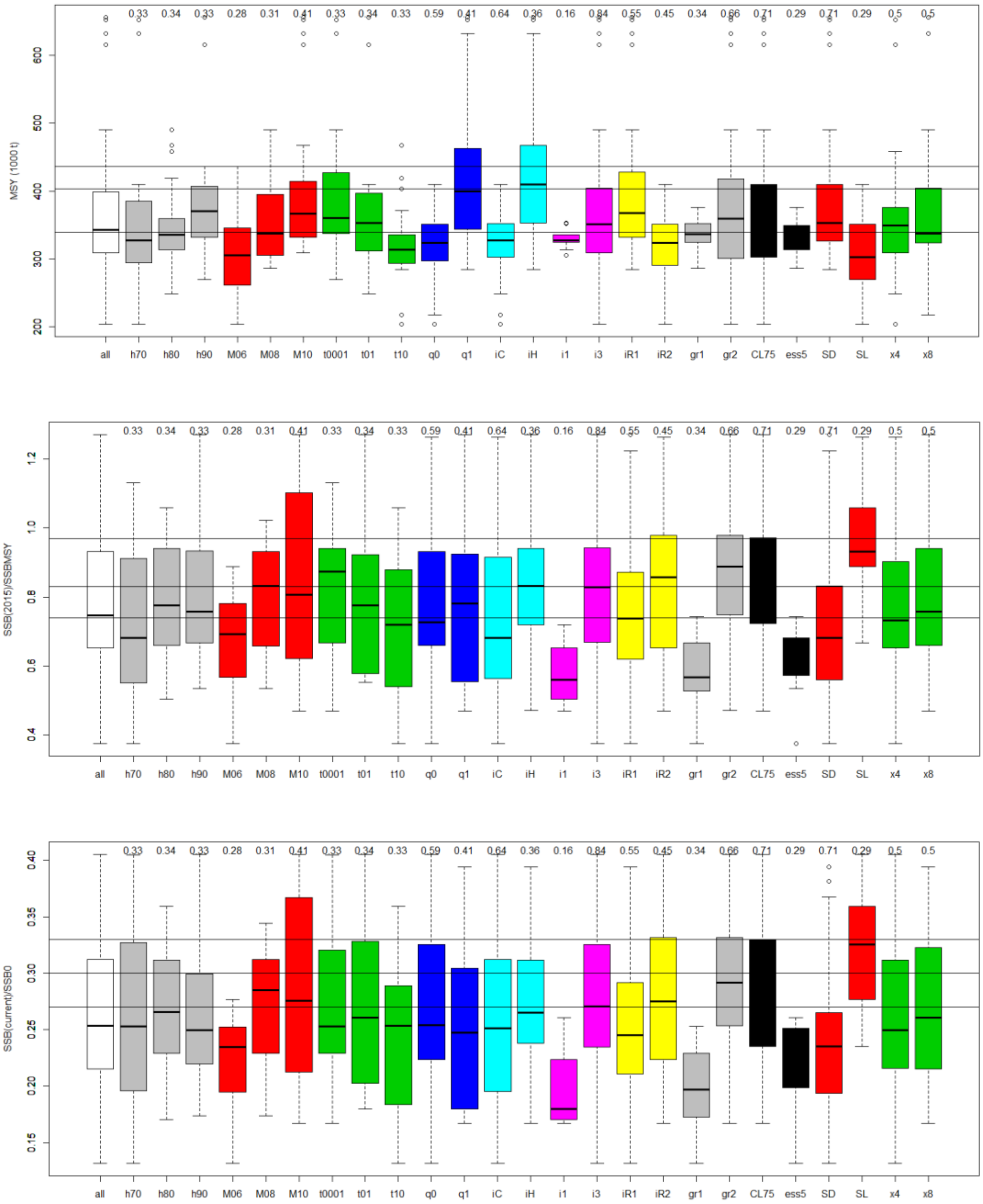


Figure 24. Stock status summaries from OMrefY20.1, marginalized over grid assumptions.

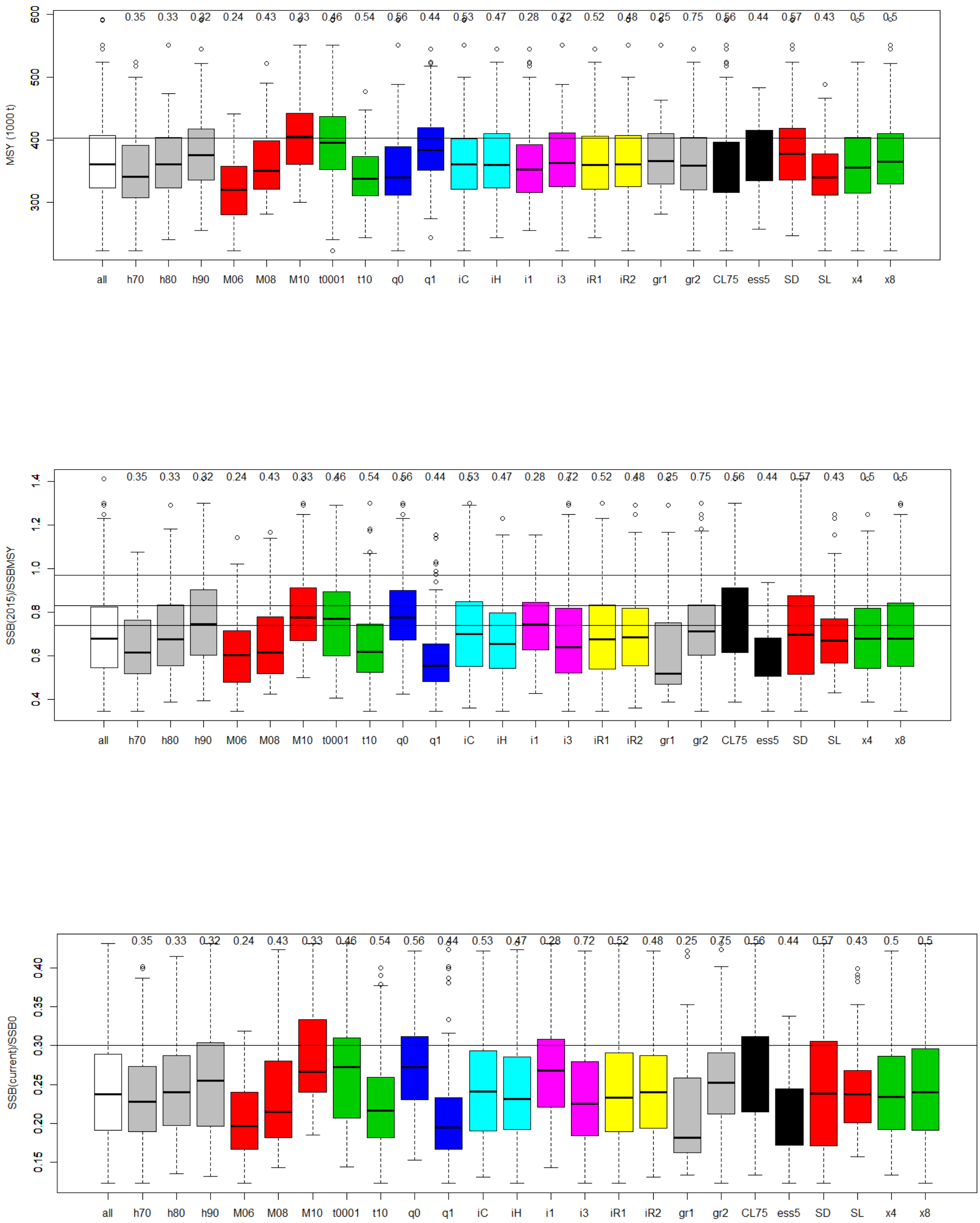


Figure 25. Stock status summaries from OMrefY19.4 (previous iteration), marginalized over grid assumptions.

6. Yellowfin - Key Points for IOTC MSE Task Force 2020 Consideration:

We welcome feedback (or endorsement) on all elements of the MSE work, and suggest the following priority points for consideration:

1. At the time of writing this document, it remains unclear how the MP development timeline will be altered as a result of COVID-19 disruptions to the IOTC community. The original plan and budget for the current phase of the project assumed that the bulk of the phase 3 MSE development work would occur Jan2020-Jun2020, and Oct2020-Jun2021, with minimal time scheduled between the 2020 TCMP and WPTT (Jul-Oct 2020).

2. We recommend retaining the basic approach to generating the OM ensemble that has evolved over previous iterations, including:

   - Each model in the ensemble consists of the Maximum Posterior Density estimates from a grid of Stock Synthesis models with multiple interacting assumption options.
   - Fractional factorial design is used to produce a manageable number of conditioned models (an ensemble of around 50 – 150), which includes all interactions between the 3 level assumptions and allows all main effects of all 2 level options to be estimable. Inclusion of all 2-way interactions would be desirable, but adds a large computational overhead and did not seem to affect MP evaluation results perceptibly in previous tests.
   - Repeated convergence tests are used to minimize the probability of outlier behaviour caused by inevitable numerical sensitivity (flat or polymodal likelihood surfaces).
   - The OM is a random sample of 500 models from the weighted grid of plausible models. To date, “weighting” has consisted of either retention or rejection of individual models, with all retained models weighted equally. Retention criteria include “adequate” numerical convergence, and other more subjective criteria.

3. The YFT OM investigations described in this document support two changes to the OM structure (which strongly influence a handful of conditioned models, but likely have a minor effect on MP tuning and behaviour overall):
   - Expanding recent recruitment constraints from 4 to 12 quarters
   - Using the new model-specific approach for linking historical and projected CPUE.

4. Additional model features are under investigation, which we would propose to continue and report against at WPTT/WPM 2020:

- The use of SS platoons or morphs to potentially represent the Dortel et al. (2014) model 3 lognormal growth curve, and explore the implications of size-based fishery selectivity.
- Representation of M uncertainty:
    - Are high M values plausible for IOTC yellowfin?
    - How sensitive are M estimates derived from methods based on maximum observed age?

5. We remain uncertain how to improve use of diagnostics to evaluate model plausibility for retention/rejection or weighting in an OM context:
    - We do employ coarse diagnostics to identify undesirable model behaviour (e.g. recruitment deviation trends, systematic lack of fit to CPUE series), etc, but these tend to manifest on a continuum, and to date have only proved useful for identifying and rejecting a small number of outlier models.
    - Iterative reweighting of model data is useful for improving the internal consistency of model statistical properties, but this is not practical in the context of a large grid and (as long as the model is in a reasonable space to begin with) this does not seem likely to have much effect on the OM dynamics. Furthermore, the common approaches do not seem to be very helpful for addressing structural and data problems that are likely to introduce important biases.
    - Retrospectives and the catch likelihood term both seem to suggest that the YFT OM might be biased towards pessimism, however our attempts to explore the issue (by introducing seasonal migration, and exploring CPUE hyper-depletion), did not suggest any obvious improvements. We suggest that the catch likelihood remains relevant for flagging questionable models, but it is not clear that this is being used in the most appropriate manner.

6. We suggest that the following priorities should be addressed before the YFT OM is formally updated:
    - Await the recommendations from the 2020 CPUE Working Group (i.e. including the most appropriate and up-to-date CPUE series to include, regional-scaling factors and potentially new advice on catchability trends).
    - Await the results of (and potentially contribute to the activities of) any collaborative efforts to improve the YFT stock assessment that are proposed for 2020. Update the stock synthesis software if this is the basis for recognized improvements.

7. If time permits, implement an MP based on fitting a joint process and observation error model (instead of the observation error production models tested to date). However, we would also continue to encourage other interested parties in the IOTC scientific community to develop and test their own MP ideas, such as this one, if they have this expertise. MP development in other RFMOs has suggested that competing MPs could lead to new approaches, shared experiences, and better performance.

References

Dortel, E, Sardenne, F, Bousquet, N, Rivot, E, Million, J, Le Croizier, G, Chassot, E. 2014. An integrated Bayesian modeling approach for the growth of Indian Ocean yellowfin tuna. Fisheries Research. 163. 10.1016/j.fishres.2014.07.006.

Francis RICC. 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. **68**: 1124‑1138.

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

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

Kolody, D, Jumppanen, P. 2019. Update on IOTC Yellowfin Tuna MSE Operating Model Development October 2019. IOTC-2019-WPM10-09.

Matsumoto, T, Yokoi, H, Satoh, K, Kitakado, T. 2018. Diagnoses for stock synthesis model on yellowfin tuna in the Indian Ocean. IOTC–2018–WPTT20–42_Rev1.

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

Methot RD, Taylor IG. 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci., **68**: 1744-1760.Pacific Fishery Management Council. 2018. Terms of Reference for the Groundfish and Coastal Pelagic Species Stock Assessment Review Process for 2017-2018 http://www.pcouncil.org/wp-content/uploads/2017/01/Stock_Assessment_ToR_2017-18.pdf.

Urtizberea, A, Fu, D, Merino, G, Methot, R, Cardinale, M, Winker, H, Walter, J, Murua, H. 2019a. Preliminary assessment of Indian Ocean yellowfin tuna 1950-2018 (Stock Synthesis, v3.30). IOTC-2019-WPTT21-50.

Urtizberea, A, Fu, D, Merino, G, Methot, R, Cardinale, M, Winker, H, Walter, J, Murua, H. 2019b. Preliminary assessment of Indian Ocean yellowfin tuna 1950-2018 (Stock Synthesis, v3.30). IOTC-2019-WPTT21-50 (powerpoint presentation).

WPM 2019. Report of the 10th Session of the IOTC Working Party on Methods. Pasaia, Spain, 17-19 October 2019. IOTC–2019–WPM10–R[E].

WPTT 2019. Report of the 21st Session of the IOTC Working Party on Tropical Tunas. Donostia-San Sebastian, Spain, 21 - 26 October 2019. IOTC–2019–WPTT21–R[E].

CONTACT US

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

FOR FURTHER INFORMATION

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