Conditioning an operating model for Indian Ocean albacore

Iago MOSQUEIRA (WMR)*

8th Session IOTC Working Party on Temperate Tuna - Assessment. 25-29 July 2022

Introduction

This document presents the process of conditioning and the current state of the Operating Model (OM) for Indian Ocean albacore. The OM is being used for the initial evaluation of Management Procedures for the stock following the guidelines providing by recent meetings of the TCMP (IOTC 2021b). The OM is based on a grid of alternative runs based on the stock assessment for albacore (Langley 2019) carried out and accepted by WPTmT in 2019.

Three system characteristics of the Operating Model and the Observation Error Model are likely to have the greatest influence in the performance of an MP: scale, noise and trend. The strategy for development of an MP described here tries to ensure that a realistic range of options for those three quantities are present in the OM set.

Model runs have been selected based on four criteria related to their fit to the data, prediction skill, and ability to explain recent catches. A large proportion of model runs did not pass these tests. Finally, the remaining runs were resampled using sampling weights based on their prediction skill for the two CPUE indices to be used in future projections.

The WPTmT 2019 SS3 albacore stock assessment

The last session of the IOTC Working Party on Temperate Tunas, WPTmT (IOTC 2019), reviewed and approved a new stock assessment (Langley 2019) for the albacore stock. The model has been constructed using the Stock Synthesis platform (Methot and Wetzel 2013), version 3.30. This is a seasonal, two-sex model, where catch data and indices of abundance are split across four areas (Figure 1), mostly to account for differences in the sizes of fish caught in the Northern and Southern areas.

Data is available from the beginning of the industrial fisheries (Figure 2), but their quality, and the amount of information contained, has varied over time through changes in the activities of some of the fleets, and lack of sampling in others.

Longline CPUE indices by area have been incorporated that are the result of a collaborative study across all longline fleets (Hoyle et al. 2019). The standardised indices were derived from operational-level longline data from the three fleets (Japan, Taiwan and Korea), and using cluster analyses to consider the effects of target change, vessel effects and spatial effects. Indices have been included for the period 1979–2017 (Figure 3, as trends in years earlier than 1979 cannot be explained with the catches taken in those years.

*Wageningen Marine Research. Haringkade 1, Postbus 68, 1976CP, IJmuiden, The Netherlands, iago.mosqueira@wur.nl

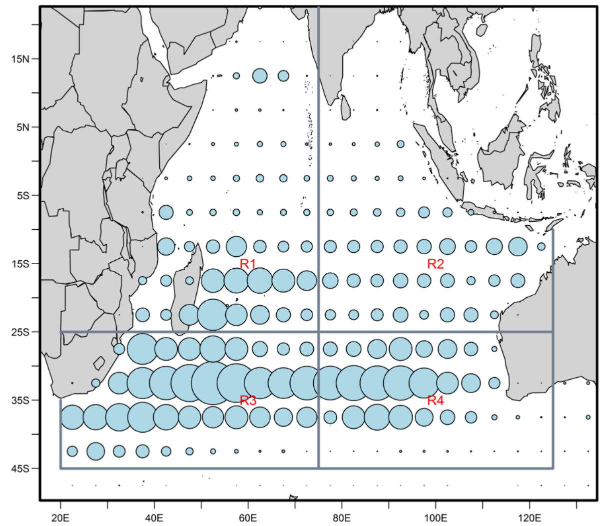


Figure 1: Spatial structure of the Indian Ocean albacore SS3 stock assessment model.

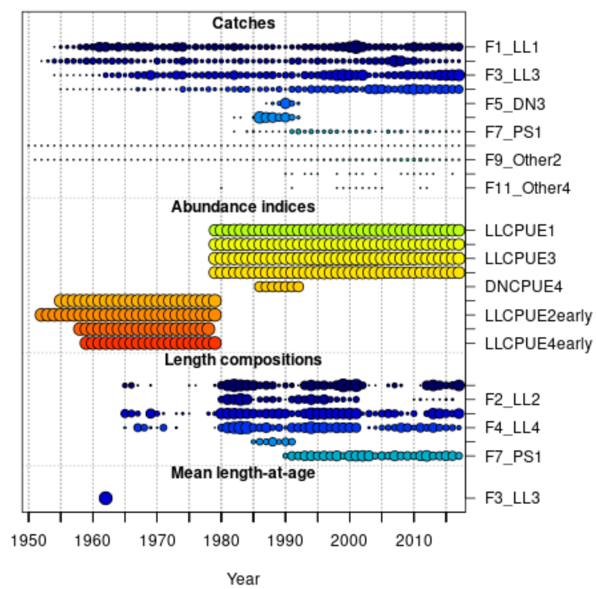


Figure 2: Temporal coverage and sources of catch, relative abundance and length composition data employed in the stock assessment model.

The 2019 stock assessment resulted in estimates of biomass that were approximately one third lower than those obtained by the 2016 model. Four main changes in the new assessment configuration ((Langley 2019)) might be behind those differences:

- Refinements to the spatial distribution of catches from the longline fishery. Catches for the LL1 and LL2 fisheries have increased, while those for LL3 and LL4 have decreased.

Figure 3: Indices of abundance employed in the SS3 stock assessment model in the later period.

Figure 4: Time series (recruitment, SSB, catch and fishing mortality) for the three stock assessment runs used for advice by WPTmT (IOTC, 2019).

- The main CPUE index in the assessment (LL3) has been revised and extended, and now shows a greater decline in stock abundance.
- A new set of growth parameters, obtained from work carried out on sampled from the Indian Ocean, are now used in the model.

- Changes in the configuration of the longline length composition data, which altered the estimates of both biomass and depletion level.

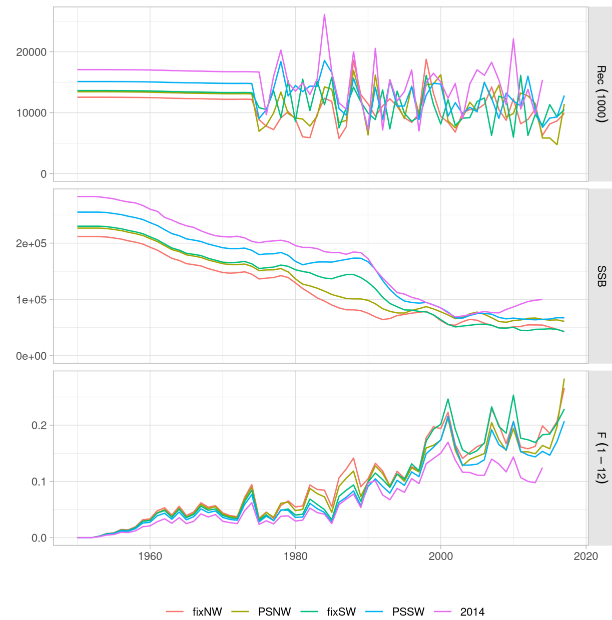


Figure 5: Stock trajectories for the three 2019 base case runs, compared with the WPTmT 2016 (SA2014) stock assessment.

Model diagnostics

An initial set of model diagnostics, following (2020), are being explored to (i) assess the quality of fit of each run in the grid, and (ii) use to provide weights to use in the resample procedure. Convergence is assessed by the value of the gradient at the solution, which is checked to be smaller than 1e-4 (Carvalho et al. 2020).

Retrospective analysis

Retrospective analysis is a form of hindcasting commonly used to assess the stability of a model formulation to updates in the data. An statistic, for example Mohn's *rho* (Mohn 1999), is then used to quantify the strength of the retrospective pattern. A subjective rule might be established where model runs with values larger than a certain limit are deemed invalid. For example, (2014) proposed that values outside the -0.15 to 0.20 range, should indicate an undesirable retrospective pattern for longer lived species.

The retrospective pattern for the estimated SSB from the base case model is shown in Figure 6. This plot also includes a one-step ahead projection of SSB based on the known total catches. The usefulness of a retrospective statistic is less clear in an operating model context, but has value at signalling models for which future bias could be a cause of instability in the application of an MP.

Runs tests

Runs tests on the CPUE and length-frequency data sources are common diagnostics of goodness of fit (Carvalho et al. 2017). The Wald-Wolfowitz runs test, a non-parametric statistical test that checks the randomness hypothesis for a data sequence, can be used to identify residuals patterns that should not be considered random.

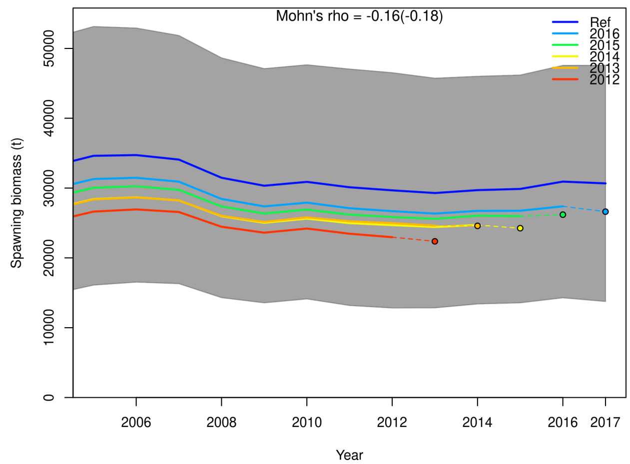


Figure 6: Five year restrospective runs with a one step ahead forecast of SSB according to total catch.

The runs test for the CPUEs and for the length-frequency data sources in the base case stock assessment (Figures 7 and 8) indicate that the model fit to most data series (those in red) present significant patterns. Model fits where the fit to the chosen CPUE does not pass the Wald-Wolfowitz runs test could be identified and not incorporated to the base case OM.

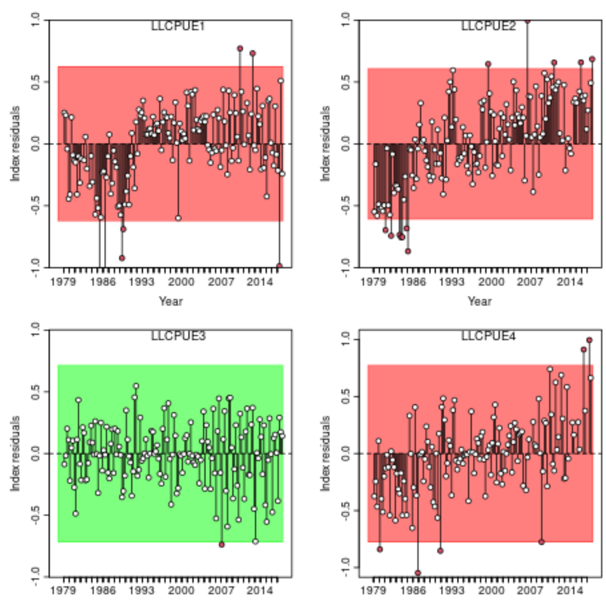


Figure 7: Runs tests of the four ongoing CPUE series.

Hindcasting cross validation

A proposed model-free hindcasting technique (HCXval) uses cross-validation to compare observations to their predicted future values (Kell, Kimoto, and Kitakado 2016). The prediction skill of a model is then computed from the prediction residuals. A robust statistic for evaluation the prediction skill can be constructed using

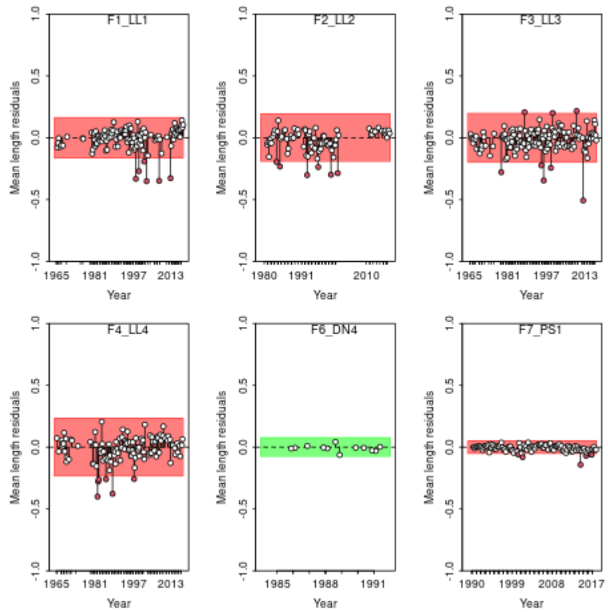


Figure 8: Runs tests for the various sources of length frequency data.

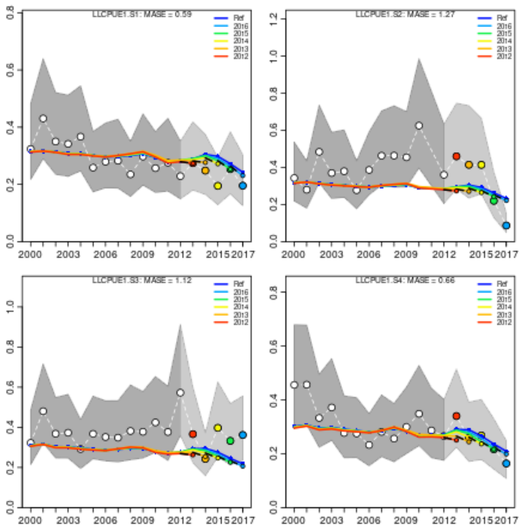
the mean absolute scaled error (MASE) of (2006). The R package *ss3diags* contains functions that simplify the calculation of prediction skill and the computation of their MASE for both CPUE and length frequencies.

The prediction skills of both the Northwestern (LLCPUE1) and Southwestern (LLCPUE3) indices are presented in Figure 9. A MASE score larger than 1 indicates the model does only as well as a random walk at predicting the quantity, while a value of 0.5 indicates the model is twice as good as a random walk. For both indices, the model prediction skill appears to be better for seasons 1 and 4, the later being the spawning season for this stock. These results could indicate that a single season CPUE, or a combination of seasons 1 and 4, could provide a better indication of stock status and trends to inform the management procedure.

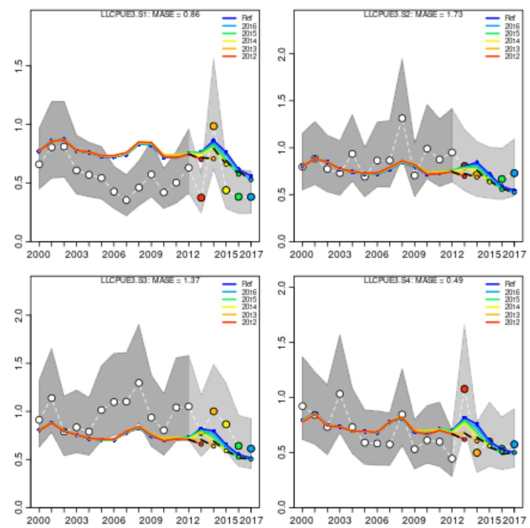
Parameter uncertainty

The previous albacore operating model did not incorporate parameter uncertainty, only considering the structural uncertainty as characterized by the variables included in the grid. A comparison of both sources of uncertainty should help informing on whether this should still be the case. Figure 10 presents the uncertainty in the estimates of current status, as generated from a Multivariate Log-Normal distribution (Winker et al. 2019).

A easy measure of parameter uncertainty, to compare with the structural uncertainty contained in the model grid, can be obtained for the estimate of biomass at the very start of the series (1952) obtained through the MVLN method (Figure 11). It appears that the scale of variability is lower, as expected, when only parameter uncertainty is considered.



(a) a



(b) b

Figure 9: Hindcasting cross-validation results by season for the LLCPUE1 and LLCPUE3 indices.

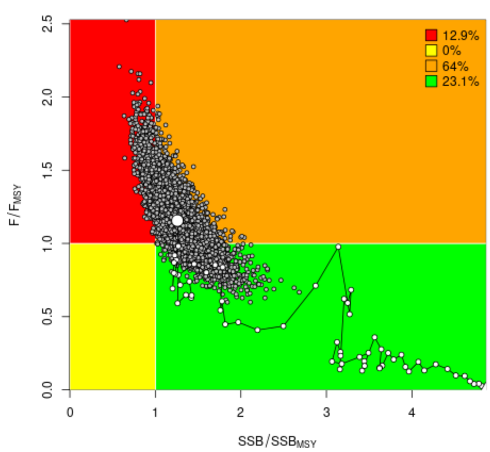


Figure 10: Uncertainty in the estimates of status (F/FMSY and B/BMSY) of the albacore base case model.

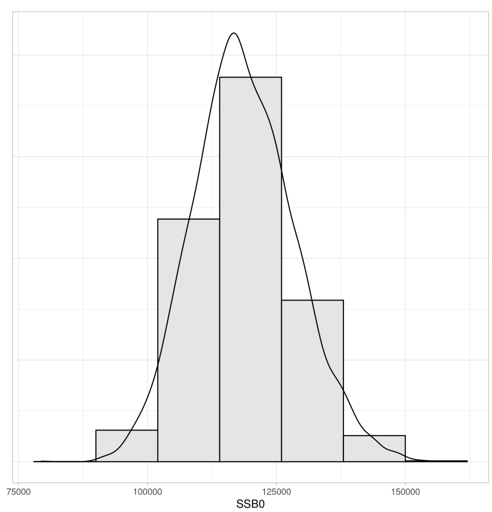


Figure 11: Uncertainty in the estimates of SSB in 1952 from the albacore base case model.

Operating Model grid

A simplified model grid was discussed and accepted by WPM (IOTC 2021a) that reduced the number of factors and levels from those used in the previous OM, based on the 2016 assessment model. The configuration of the model grid was informed by an analysis of the effect of each level and factor on the uncertainty of the previous one, which consisted of 1,440 model runs.

The model grid currently in used includes the following factors and levels:

- Natural mortality (M): 0.20, 0.25, 0.30 or 0.35, for all ages.
- Standard deviation in recruitment deviates (sigmaR): 0.4, 0.6, or 0.8.
- Steepness of the stock-recruitment relationship: 0.7, 0.8 or 0.9.
- LL CPUE series (cpues): Northwest (12) or Southwest (14).
- Length-frequency data likelihood weighting (lfreq): 0.01, 0.01, 0.1 or 1.
- Catchability increase of the LL CPUE series (llq): 0% or 1% per year.

Therefore the *full* grid consists of 432 model runs, together with five restrospective runs for each of them. Running these models requires around 1,200 h of computation time.

Effect of individual grid factors and levels

An exploration has been carried out of the individual effect on model output of adopting an alternative value for each variable, one at a time. These *main effects* provide an useful indication of how much variability in dynamics and status is controlled by a single variable, and how much is the result on 2nd or higher order interactions. The estimates of the scale indicator (virgin spawning biomass) are shown in Figure 12, while those for trend (SSB in 2017 over SSB at MSY) are presented in Figure 13.

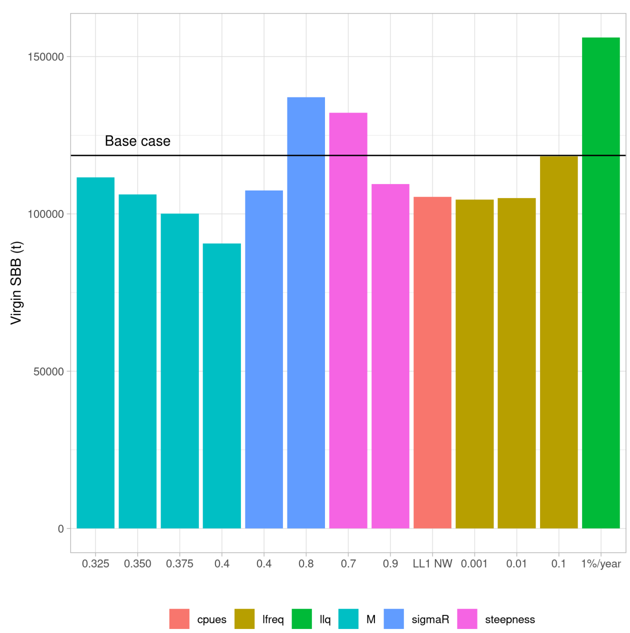


Figure 12: Changes in estimates of virgin spawning biomass (SSB0) under each factor and value. Horizontal line shows the reference estimate for the base case model run.

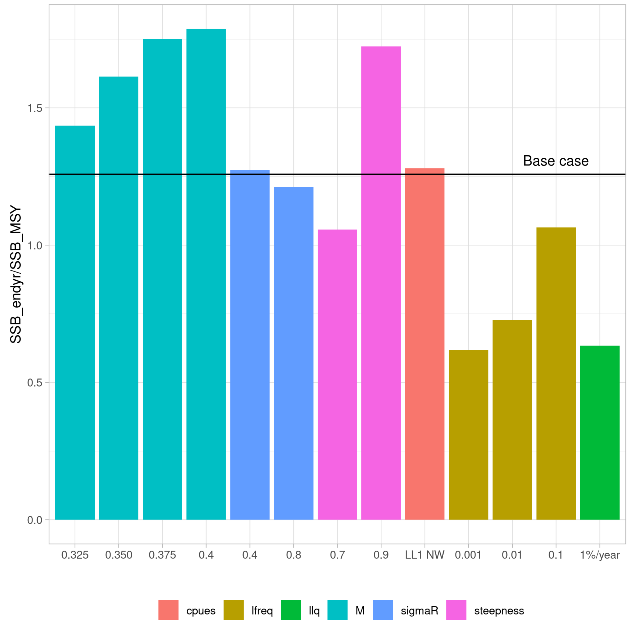


Figure 13: Changes in estimates of spawning stock biomass (SSB) in the last year (2017) over the SSB at MSY, under each factor and value. Horizontal line shows the reference estimate for the base case model run.

Partial factorial operating model

A previous meeting of the MSE taskforce of WPM (IOTC, 2021), recommended that all operating models constructed around a grid of SS3 model runs should apply a partial factorial design to the full grid. This should allow running and inspecting a reduced number of models without a loss in its capacity to map stock uncertainty. Federov's exchange algorithm, as implemented in R package 'AlgDesign'[1], was used for this. An evaluation of the sampling design indicated that a value of 84 runs would be sufficient to obtain a robust representation of the grid uncertainty (close to 95%), as quantified by the minimax normalized variance (Figure 14).

Once the selection criteria outlined below were applied to this subset, 38% of the chosen runs were rejected, leaving an OM grid composed of 52 model runs. Given that the size of the full grid, much reduced in this latest version, was considered not too large, and expecting a similar rejection rate as obtained in the partial factorial grid, the decision was taken instead to run the full grid of 432 models.

Selection process

Model runs from the full grid were then evaluated applying four diagnostic criteria. The objective was to retain only those model runs where convergence was assured, led to reasonable estimates of stock status, and had sufficient prediction skill:

1. Identify models leading to estimates of virgin biomass outside of a reasonable range (SB0 > 1e7), or overly optimistic estimates of recent stock status (SB 2017 / SBMSY > 3).
2. Signal models with no clear convergence of Stock Synthesis run, if final gradient > 1e-4

[1] https://github.com/jvbraun/AlgDesign

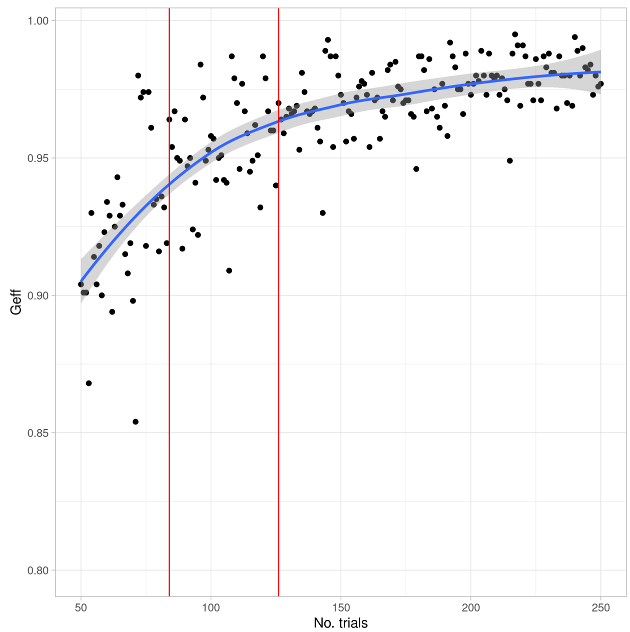


Figure 14: Minimax normalized variance expressed as an efficiency with respect to the optimal approximate theory design (Geff) for an increasing number of trials.

3. Identify runs where their MASE of selected CPUE indices (LLCPUE1 NW and LLCPUE3, SW) in seasons 1 and 4 were greater than 1.
4. Determine model runs that could not take the 2018, 2019 and 2022 nominal catches reported to IOTC (41 615, 39 246 and 41 051 t, respectively) without a yearly increase in fishing mortality larger than 25% per year.

These four selection criteria were applied to the full OM grid of 432 Stock Synthesis model runs. The number of runs failing each of the individual tests were as follows:

- Unrealistic values: 10.
- Convergence: 26.
- MASE CPUE indices 1 and 3: 121.
- Catch 2018-2020 could not be taken: 376.

Model runs were kept in the OM set only if they passed all four of those selection filters. This was the case for only 39 model runs out of 432 model configurations in the original grid, or around 9%. This model selection process leads to an uneven distribution of runs across four out of the six grid factors (Figure 15): natural mortality (M), steepness, weight of the length frequency data (lfreq) and yearly increase in LL CPUE catchability (llq).

Time series

The time series plot for this reference OM shows a large uncertainty in initial biomass and in current status, most notable on the level of fishing mortality (Figure 16). The comparison with the stock trajectories estimated by the base case stock assessment model (Figure 17) differ specially in the recent levels of fishing mortality and stock abundance. After projecting the stock for the catches reported between 2018 and 2020, the biomass

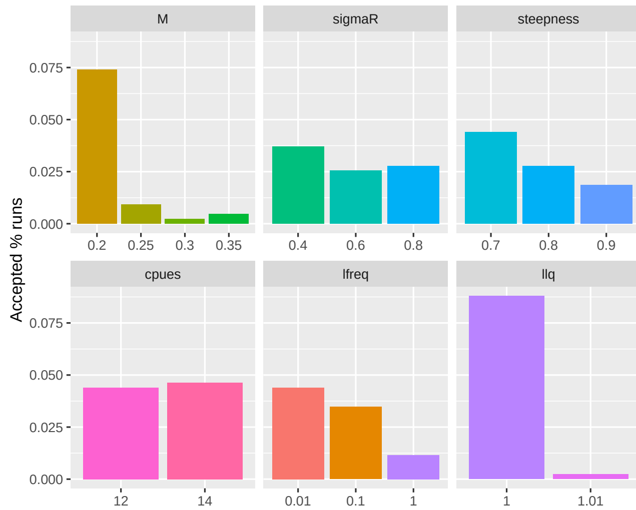


Figure 15: Percentage of accepted OM runs across all factors and levels.

levels appear to have turned to even lower values.

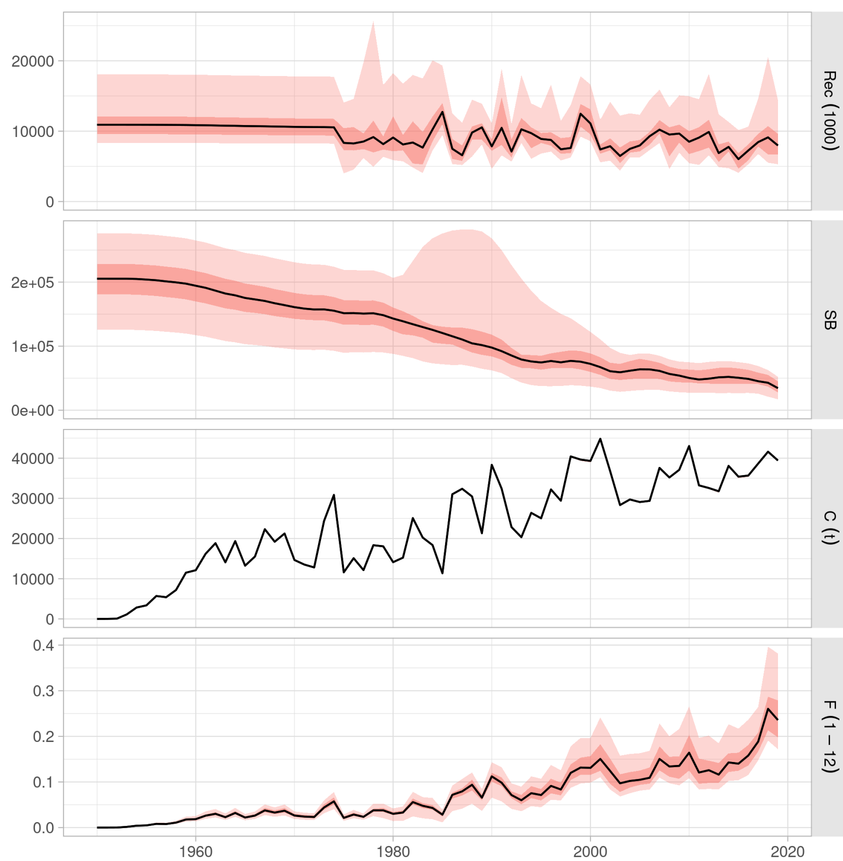


Figure 16: Time series (recruitment, SSB, catch and fishing mortality) for the resampled operating model, projected ahead to 2019 based on total reported catches.

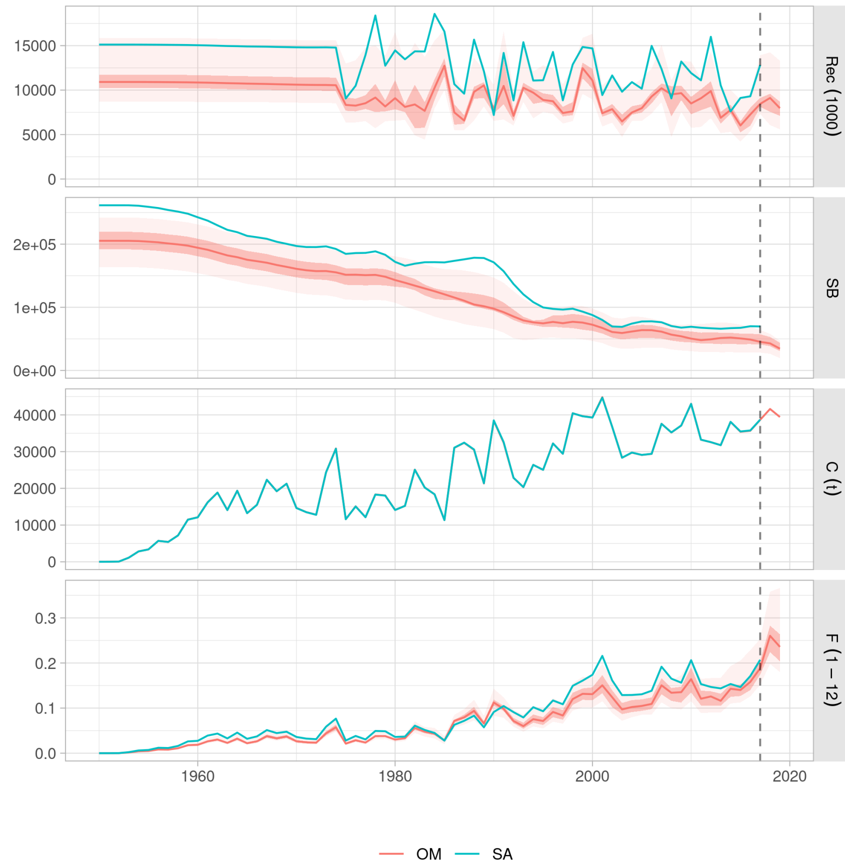


Figure 17: Time series (recruitment, SSB, catch and fishing mortality) for the resampled oeprating model, projected ahead to 2019 based on total reported catches (OM) and the base case stock assessment model run (SA).

Projections for MP evaluations using the reference OM will start from a stock status that is in a majority of cases (75%) worse than what the base stock assessment estimated (Figure 18). This will likely call for a recovery phase. Management objectives might need to be realigned to ensure stock recovery in a reasonable time frame. Current tuning objectives, P(Kobe = green) of 50%, 60%, and 70% in 2030-2034, might only be achievable through a sharp decrease in allowable catches in the short term.

Displaying the time series of the main metrics for the OM, but split across levels of each grid factor (Figures 19 to 21) shows how and in which way the OM runs diverge. Natural mortality values scale the biomass of the stock up or down, as expected (Figure 19), while both the yearly increase in LL catchability (Figure 21) and the choice of CPUE (Figure 19), appear responsible for the two alternative trends in biomass over the 1980 to 2000 period.

The uncertainty in stock dynamics is also reflected in the range of estimates for MSY reference points (Figure 22).

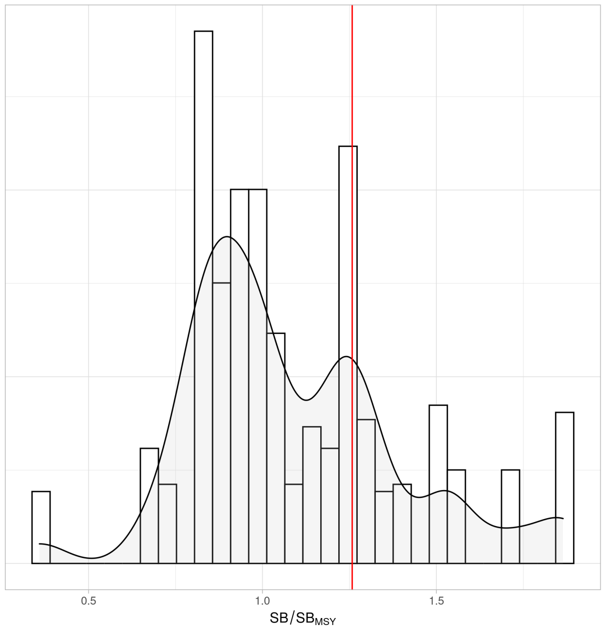


Figure 18: Distribution of estimated values for stock status (SB_{2017}/SB_{MSY}) compared to that obtained from the base case stock assessment model run (red line).

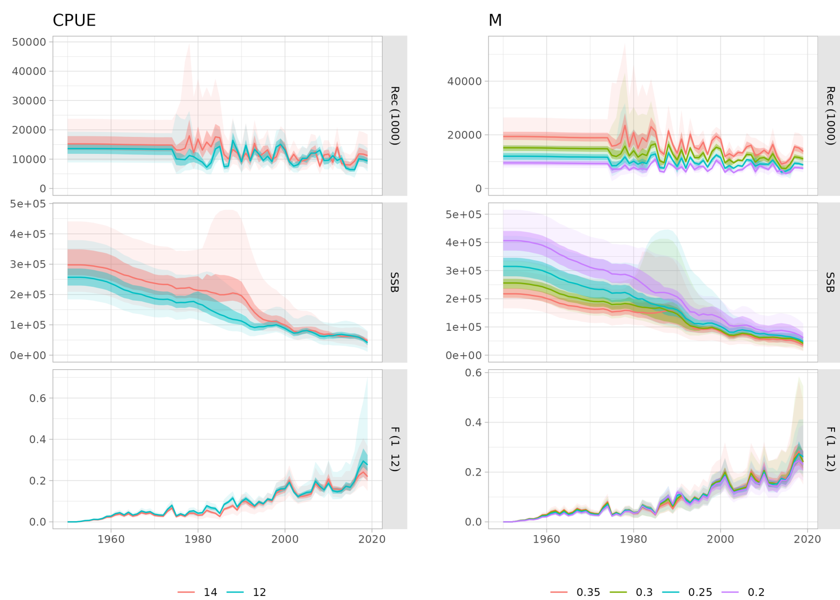


Figure 19: Time series (recruitment, SSB, catch and fishing mortality) for the resampled operating model, split across CPUE and natural mortality grid levels.

Importance of grid factors and levels

The effect of the various factors and levels in the model grid on some quantities of interest was explored through a series of regression trees (Breiman et al. 1984), computed using the R package *rpart*. Three variables

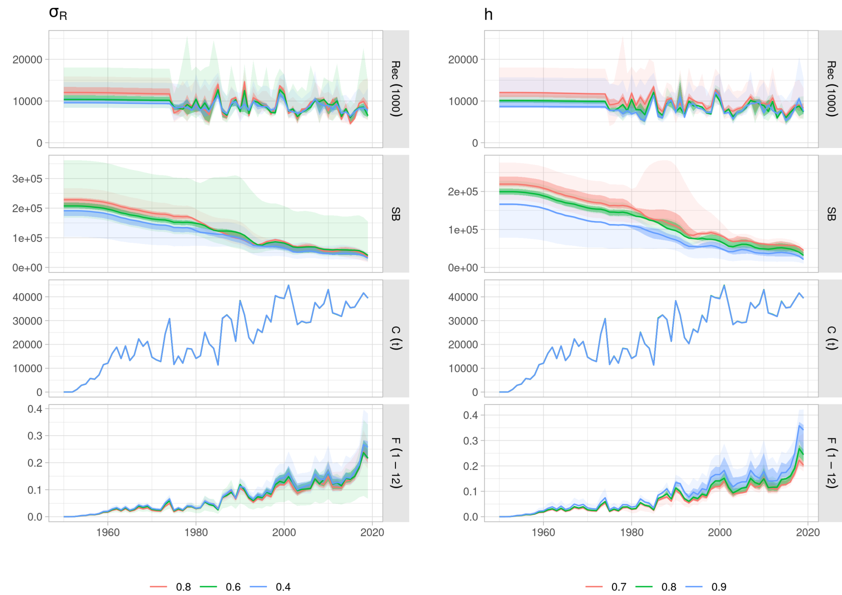


Figure 20: Time series (recruitment, SSB, catch and fishing mortality) for the resampled operating model, split across recruitment variance and steepness grid levels.

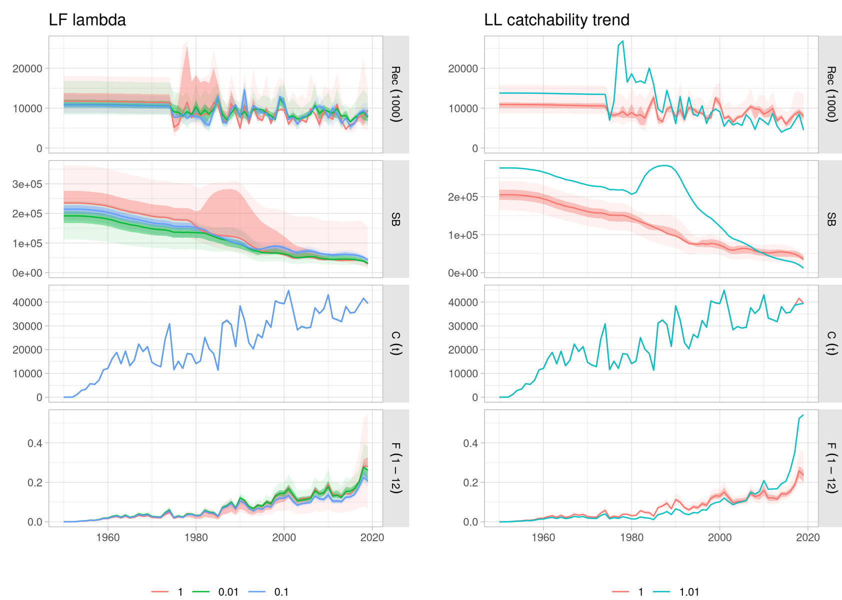


Figure 21: Time series (recruitment, SSB, catch and fishing mortality) for the resampled oeprating model, split across LL length frequency likelihood weight and yearly increase in LL CPUE catchability grid levels.

were selected to represent the three main OM system characteristics: scale (K), noise (recruitment variance) and trend (depletion).

Figures 23 to 25 present the results of this analysis. The estimates of initial stock size (K) depend on two grid

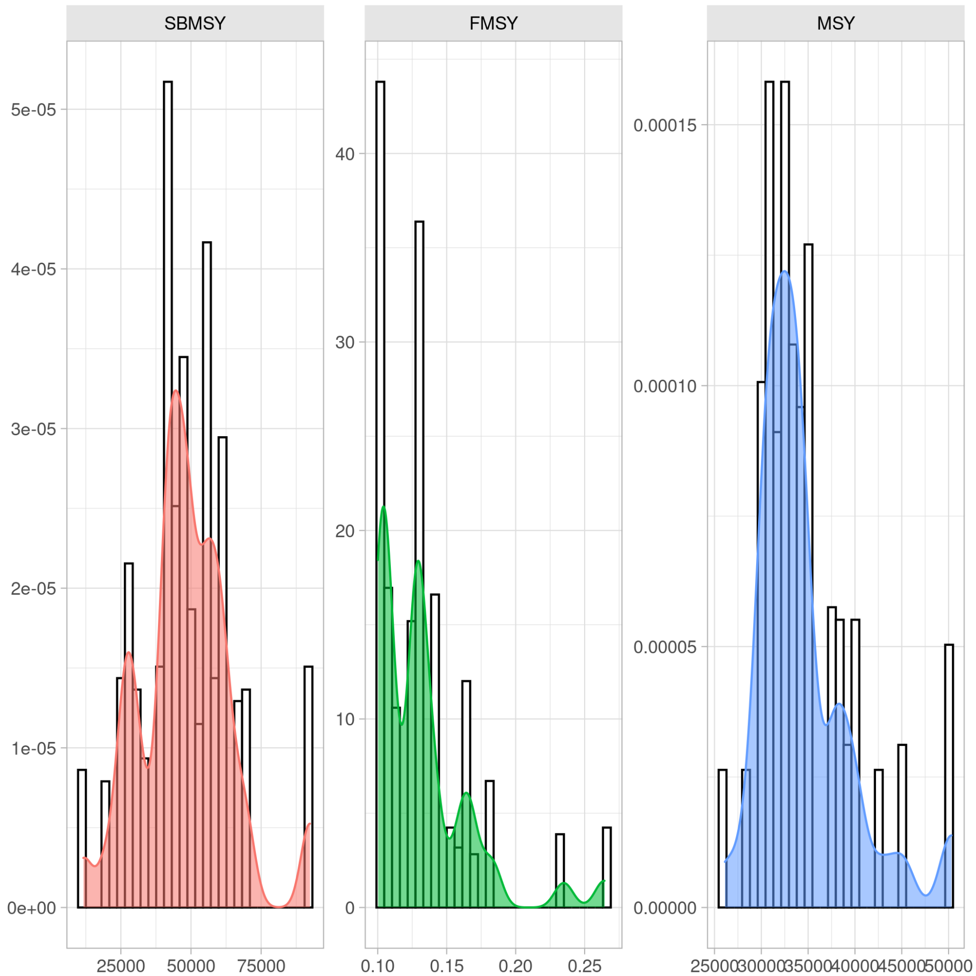


Figure 22: Distribution of estimated MSY reference points across the full OM grid.

factors, natural mortality and steepness, and also on the variability in recruitment deviances (Figure 23). The choice of recruitment variability in the model fitting (sigmaR), has relatively limited importance on the actual process error, as represented by the variance in recruitment deviances 25. Instead, the choice of CPUE and the assumed catchability trend in the CPUE fleet has a three-fold effect on this model characteristic. Depletion in the last year, computed as $SB_{2017} SB_0$, appears to be greatly determined by a combination of factors 24, lead by the catchability trend and steepness. Stock status at around the BMSY level, 27% depletion, is only found on model runs with high steepness values and no changes in longline COUe catchability.

Short-term projections

The future dynamics and robustness of the OM have been first evaluated through a series of projections. They attempt to establish whether the population is able to sustain recent catch or fishing mortality levels over a number of years, and the expected impact of a short recovery to MSY conditions. Following the OM update based on the nominal catches for the 2018-2020 period, abundances in 2021 have been computed by assuming a level of mean fishing mortality ($\bar{F}$) over ages 1 to 12 equal to that estimated in 2020.

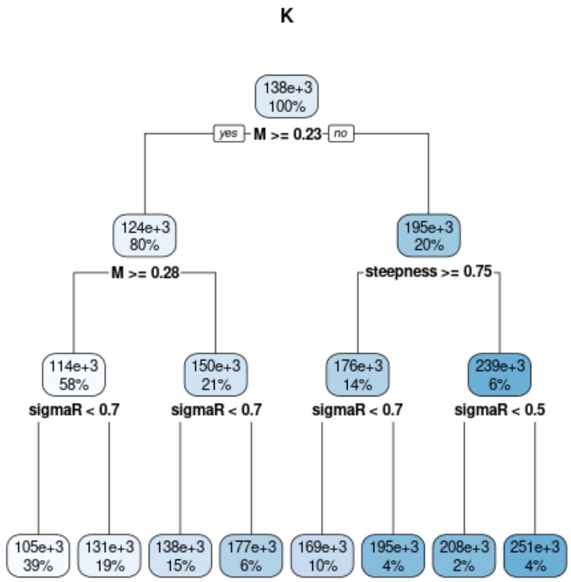


Figure 23: Regression tree for all factors and levels in the filtered OM grid over carrying capacity (K).

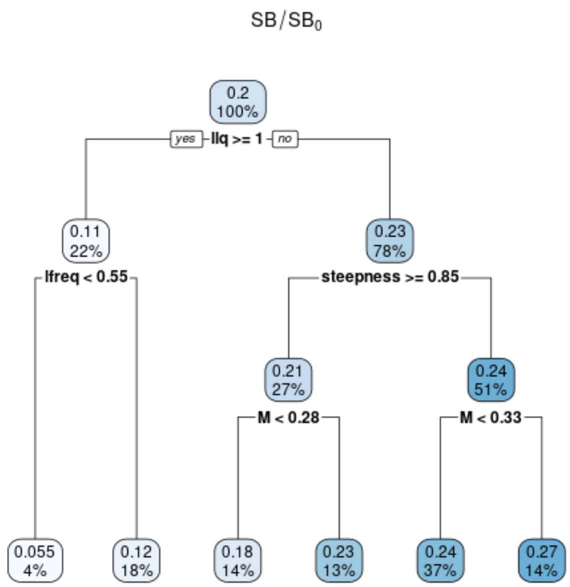


Figure 24: Regression tree for all factors and levels in the filtered OM grid over stock status ($SB/SB0$), defined as depletion level from virgin SSB in 2017.

Applying a catch level of 35,000 t, approximately equal to the catches of 2016, and lower than catches reported between then and 2019, leads to 76% of model iterations not being able to explain those catches if yearly increases in fishing mortality are limited to be no greater than 25% 26.

When the OM is projected for the respective MSY catch levels across each run (Figure 27), only 8% of the

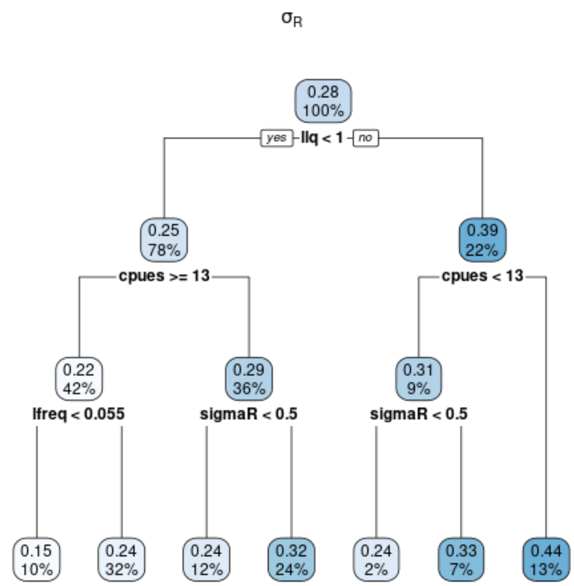


Figure 25: Regression tree for all factors and levels in the filtered OM grid over the variance in stock-recruitment residuals (σ_R).

iterations are unable to sustain that catch level over the same period. The large variability in the estimates of reference points 22 is reflected here in the disparity of catch targets able to maintain the stock at healthy levels, and the probability of catches like those observed in the recent past being extracted.

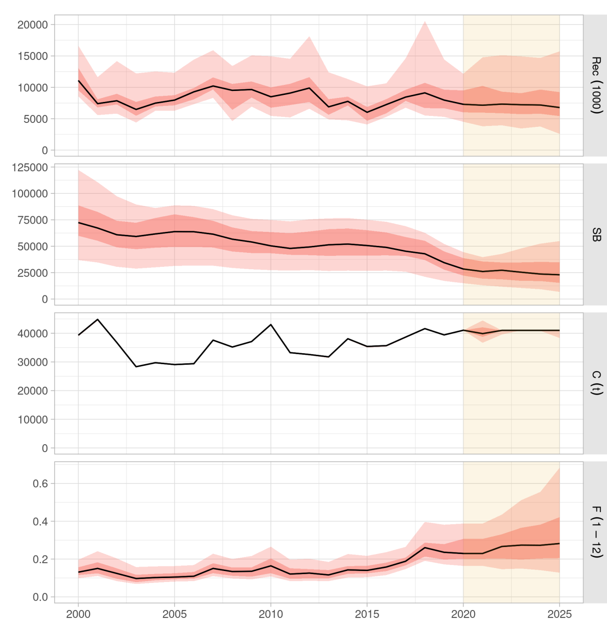


Figure 26: Projection of the full OM for a catch level of 35,000 t over the 2020-2023 period.

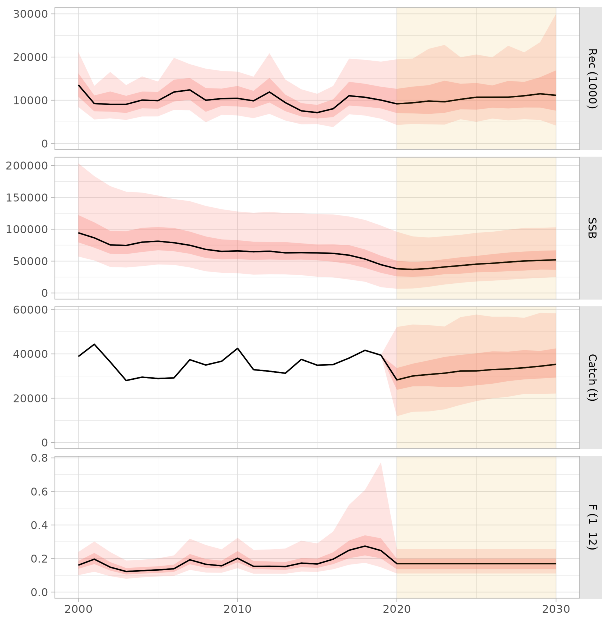


Figure 27: Projection of the full OM for a catch level corresponding to the FMSY estimated for each model run over the 2020-2023 period.

Discussion

The OM presented here is the third iteration of a model grid-based conditioning for Indian Ocean albacore. The grid has been substantially reduced in size from the previous version, but still a large number of model formulations lead to results that can not be accepted. The selection steps outlined above led to almost half of those models being rejected, given the problems in convergence, extreme results in estimated carrying capacity or recent biomass, or their inability to predict the main source of information on stock abundance, the LL CPUE series. Management procedures for this stock will always depend on the ability of the standardized indices of abundance to track faithfully changes in stock abundance. Models that cannot make good use of this information, measured here by their prediction skill, have very limited value for simulation testing of any MP.

The average stock status determined by the OM is in clear contrast with that obtained from the latest stock assessment in which it is based. The base case model run indicates a healthier stock, with respect to the management MSY reference points. The nature of the catch and CPUE series, effectively a one way trip, appear to make the model runs heavily dependent on the values chosen for some of the model parameters. The model grid was constructed in an attempt at covering the uncertainty around set assumptions, by generally going for both higher and lower values. But the two and three-way interactions led to uncertainty in dynamics and status being more complex than that. A large proportion of model runs that cannot explain recent nominal catch, or show very little prediction power for the indices of abundances. This severely limits their usefulness as platforms for exploring the possible effect of alternative management procedures.

The new stock assessment model for Indian Ocean albacore discussed during WPTmT08 might present a view of the stock dynamics, and specially its status, different enough to justify a call for reconditioning the OM. This process has already taken place twice, partly due to the step by step process for adoption of MPs at IOTC, but also because of the difficulties encountered with these OMs. Even after the model selection process, some runs in the grid present a stock with dynamics that diverge from those of most of the grid, and lead to short term risks that might not be realistic.

Decoupling the OM conditioning from the stock assessment is an option that should be considered for this stock. Methods for conditioning such as those recently presented to the WPM MSE taskforce (Hillary and Mosqueira 2022), present a possible methodological alternative. Approximate Bayesian Computation (ABC) requires priors to be defined for multiple population and fishery parameters, and this could be constructed using stock assessment outputs. So rather than a complete break from the stock assessment model, this methodology proposes adopting an approach that recognizes that the objectives of both types of models, assessment and operating, are not the same. The stock assessment step should still be part of the MP review process, and carried out to provide a check on the behaviour and effect of the MP, and identify any warning sign that would require a review of the MSE analysis.

Some lessons from the conditioning exercise might be identified by WPTmT as useful for the stock assessment work. The diagnostics employed to select and weight model runs are now all available in standard form, and should be computed as they provide important insights on the model quality and ability to provide management advice. The lag between data and (possible) management action for this stock in IOTC is larger than for other stocks, and information on recruitment strength is only available after a number of years, given the selectivity patterns of most fleets. Both factors make forecasting future stock status for the stock under different management options particularly challenging, and makes more necessary any evaluation of the ability of the model to do so.

Acknowledgements

This work has benefited from the continuing feedback from members of IOTC WPM, WPTmT and SC.

Funding

©FAO 2019-2022. This work is being funded by the Indian Ocean Tuna Commission - Food and Agriculture Organization of the United Nations (IOTC-FAO). The views expressed in this document are those of the author(s) and do not necessarily reflect the views or policies of FAO or IOTC.

References

Breiman, Leo, Jerome Friedman, Charles J Stone, and Richard A Olshen. 1984. *Classification and Regression Trees*. CRC press.

Carvalho, Felipe, André E. Punt, Yi-Jay Chang, Mark N. Maunder, and Kevin R. Piner. 2017. "Can Diagnostic Tests Help Identify Model Misspecification in Integrated Stock Assessments?" *Fisheries Research* 192 (August): 28–40. https://doi.org/10.1016/j.fishres.2016.09.018.

Carvalho, Felipe, Henning Winker, Dean Courtney, Maia Kapur, Laurence Kell, Massimiliano Cardinale, Michael Schirripa, et al. 2020. "A Cookbook for Using Model Diagnostics in Integrated Stock Assessments." *Fisheries Research*, Accepted manuscript under revision.

Hillary, R., and I. Mosqueira. 2022. "Exploring a Wider Approach to Om Conditioning in Iotc Mse Work." IOTC–2018–WPM13(MSE)–04. Eden Island, Seychelles: Indian Ocean Tuna Commission.

Hoyle, S. D., D. Fu, D. N. Kim, S. I. Lee, T. Matsumoto, K. Satoh, S. -P. Wang, and T. Kitakado. 2019. "Collaborative Study of Albacore Tuna Cpue from Multiple Indian Ocean Longline Fleets in 2019." Mahe, Seychelles: IOTC.

Hurtado-Ferro, Felipe, Cody S. Szuwalski, Juan L. Valero, Sean C. Anderson, Curry J. Cunningham, Kelli F. Johnson, Roberto Licandeo, et al. 2014. "Looking in the Rear-View Mirror: Bias and Retrospective Patterns in Integrated, Age-Structured Stock Assessment Models." *ICES Journal of Marine Science* 72 (1): 99–110. https://doi.org/10.1093/icesjms/fsu198.

Hyndman, Rob J., and Anne B. Koehler. 2006. "Another Look at Measures of Forecast Accuracy." *International Journal of Forecasting* 22 (4): 679–88. https://doi.org/10.1016/j.ijforecast.2006.03.001.

IOTC. 2019. "Report of the 7th Session of the Iotc Working Party on Temperate Tunas: Assessment Meeting." IOTC-2019-WPTmT07(AS)-R. Shizuoka, Japan: Indian Ocean Tuna Commission.

———. 2021a. "Report of the 10th Session of the Iotc Working Party on Methods." IOTC-2021-WPM12-R[E]. Indian Ocean Tuna Commission.

———. 2021b. "Report of the 4th Session of the Iotc Technical Committee on Management Procedures." IOTC-2019-TCMP04-R[E]. Indian Ocean Tuna Commission.

Kell, Laurence T., Ai Kimoto, and Toshihide Kitakado. 2016. "Evaluation of the Prediction Skill of Stock Assessment Using Hindcasting." *Fisheries Research* 183 (November): 119–27. https://doi.org/10.1016/j.fishres.2016.05.017.

Langley, Adam. 2019. "Stock Assessment of Albacore Tuna in the Indian Ocean Using Stock Synthesis for 2019." In *Seventh Session of the Iotc Working Party on Temperate Tunas: Assessment Meeting*. IOTC–2019–WPTmT07(AS)–11. Mahe, Seychelles: IOTC.

Methot, Richard D., and Chantell R. Wetzel. 2013. "Stock Synthesis: A Biological and Statistical Framework for Fish Stock Assessment and Fishery Management." *Fisheries Research* 142 (May): 86–99. https://doi.org/10.1016/j.fishres.2012.10.012.

Mohn, R. 1999. "The Retrospective Problem in Sequential Population Analysis: An Investigation Using Cod Fishery and Simulated Data." *ICES Journal of Marine Science* 56 (4): 473–88. https://doi.org/10.1006/jmsc.1999.0481.

Winker, Henning, John Walter, Massimiliano Cardinale, and Dan Fu. 2019. "A Multivariate Lognormal Monte-Carlo Approach for Estimating Structural Uncertainty About the Stock Status and Future Projections for Indian Ocean Yellowfin Tuna." In *22nd Session of the Iotc Working Party on Temperate Tunas*. IOTC-2019-WPTT21-51. Mahe, Seychelles: IOTC.