SCIENTIFIC COMMITTEE
SIXTEENTH REGULAR SESSION
Online
11–20 August 2020

Updating the WCPO Skipjack Operating Models for the 2019 Stock Assessment

WCPFC-SC16-2020/MI-IP-08

R. Scott[1], F. Scott, G. M. Pilling, P. Hamer and J. Hampton

[1]Oceanic Fisheries Programme, The Pacific Community

Contents

1 Introduction **5**

2 Accounting for uncertainty **7**

2.1 Process Uncertainty 7

2.2 Observation Uncertainty 8

2.3 Model Uncertainty and Parameter Uncertainty 9

3 The skipjack OM grid **9**

3.1 Reference and robustness sets 10

3.2 Comparison with 2019 assessment model grid 10

3.3 Comparison with 2016 OM model grid 11

4 Model validation **11**

4.1 Model diagnostics 11

4.1.1 Overall model fit diagnostics 12

4.1.2 Model fit to data diagnostics 12

4.1.3 Model output diagnostics 13

4.1.4 Model consistency diagnostics 13

5 Discussion **14**

5.1 Procedures for updating the OM grid under the Harvest Strategy approach 14

6 Conclusions **14**

A Appendix **18**

A.1 Model validation tables 18

A.2 Model validation figures 20

Executive Summary

An initial grid of operating models for WCPO skipjack has previously been presented to, and considered by SC14. The initial grid was developed from the 2016 skipjack stock assessment and formed the basis of the preliminary HCR evaluations presented to SC15. A new assessment for WCPO skipjack was agreed at SC15. In this paper we compare the skipjack OM grid, as developed from the 2016 stock assessment, with the 2019 stock assessment. We consider the steps and procedures for updating the OM grid for the revised assessment and we further consider the range of diagnostic outputs that should be presented when selecting the suite of models for the OM grid.

The revised skipjack OM grid based on the updated 2019 assessment (below) is very similar to the previous OM grid based on the 2016 stock assessment. The axes of uncertainty considered in the robustness set and their respective settings have changed very little.

Axis	**Levels**		**Options**		
	Reference	Robustness	**0**	**1**	**2**
Process Error					
Recruitment Variability	**2**		**1982-2018**	**2005-2018**	
Observation Error					
Catch and effort	**1**	1	**20%**	30%	
Size composition (ESS)	1		**estimated**		
Tag recaptures	**1**	2	**status quo**	low	none
Model Error					
Steepness ‡	**3**		**0.8**	**0.65**	**0.95**
Mixing period (qtr) ‡	**2**		**1**	**2**	
Growth ‡	**2**			**low**	**high**
Movement	**1**	1	**estimated**	El Nino/La Nina	
DD catchability (k) ‡	**2**	1	**0**	-0.5	-0.9
Implementation Error					
Effort creep	**2**	1	**0%**	2%	3%

Table: Skipjack OM uncertainty grid. Scenarios shown in bold are proposed for the reference set. ‡ denotes those scenarios for which a dedicated fit of MULTIFAN-CL is required.

Overall, the diagnostics of model fit to data are consistent with those of the 2019 assessment and show broadly consistent model fits across the OM grid. None of the models provide implausible outputs that would indicate inadequate fits.

The overall trends in depletion estimated for the revised OM grid are broadly consistent with those of the previous 2016 OM grid and the estimates correspond very well in the terminal years. It should not always be necessary to update the suite of operating models each time a new stock assessment is conducted. If the OM grid remains broadly consistent with the full assessment conducted by the monitoring strategy there should be no need to change it.

The grid of models outlined in this report form the basis of the updated MP evaluations provided in WCPFC-SC16/MI-IP-03.

We invite WCPFC-SC to consider the following questions:

- Have all important sources of uncertainty been considered?
- Do the ranges of parameter values adequately reflect uncertainty in the dynamics of the resource?

in addition we invite WCPFC-SC to note

- It should not always be necessary to update the suite of operating models each time a new stock assessment is conducted. If the OM grid remains broadly consistent with the full assessment conducted by the monitoring strategy there should be no need to change it.
- The grid of models outlined in this report form the basis of the updated MP evaluations provided in WCPFC-SC16/MI-IP-03.

1 Introduction

Within the MSE framework (Figure 1) the Operating Model (OM) simulates the real world by attempting to capture all existing knowledge and data processes for the exploited populations and associated fisheries. Our knowledge of the dynamics of populations and fisheries is often incomplete. The OM should therefore allow for the evaluation of the consequences for management by testing different hypotheses about those dynamics. In this respect a suite of different OMs may be identified, each one representing an alternative hypothesis. Conditioning an operating model involves fitting the model to data (much the same as a stock assessment) in order to identify a suite of models that adequately characterises the range of uncertainty so that we can find the Management Procedure (MP) that performs best and is robust to that uncertainty.

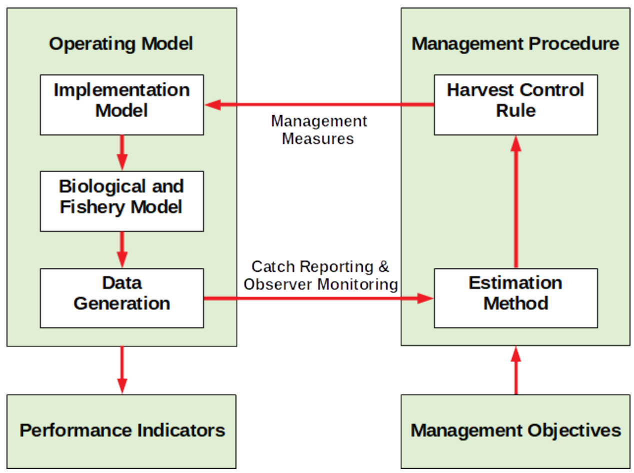


Figure 1: The MSE framework for testing Management Procedures.

An initial grid of operating models for WCPO skipjack has previously been presented to, and considered by SC14 (Scott et al., 2018b). The initial grid was developed from the 2016 skipjack stock assessment (McKechnie et al., 2016) and formed the basis of the preliminary HCR evaluations presented to SC15 (Scott et al., 2019). A new assessment for WCPO skipjack was agreed at SC15 (Vincent et al., 2019). The new assessment has modified parameter settings, a revised, 8 region, spatial structure and provides updated estimates of stock status for the most recent years. The 2019 assessment of WCPO skipjack tuna, as agreed by SC15, employed a grid of 54 models across 4 axes of uncertainty. The grid differed from the previous 2016 assessment uncertainty grid in that tag over dispersion is now estimated, rather than fixed at a particular value, and has therefore been dropped from the grid. Uncertainty in growth is included in the 2019 assessment uncertainty grid.

In this paper we compare the skipjack OM grid, as developed from the 2016 stock assessment, with the 2019 stock assessment. We consider the steps and procedures for updating the OM grid for the revised assessment and we further consider the range of diagnostic outputs that should be presented when selecting the suite of models for the OM grid.

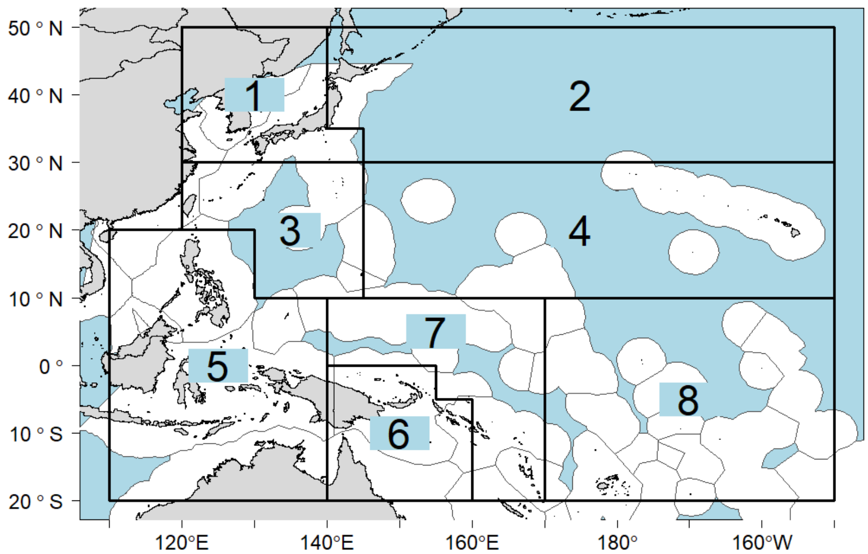


Figure 2: 2019 stock assessment regional structure.

In the case of the WCPO skipjack evaluation framework, changes made to the OM grid may have implications for the design and structure of the management procedure, particularly when the spatial structure of the models change. In WCPFC-SC16/MI-IP-10 we consider the implications of a revised OM grid for the generation of pseudo data and in WCPFC-SC16/MI-IP-09 we re-evaluate the performance of the estimation model to ensure that it continues to provide an appropriate and reliable indication of stock status.

The steps and procedures for updating the OM grid include a re-evaluation of the most important sources of uncertainty (Section 2), definition of the OM grid and consideration of the reference and robustness sets (Section 3) and the interrogation of a range of model diagnostics and outputs to determine the validity of the suite of models that comprise the OM grid (Section 4).

Axis	**Levels**	**Options**		
		0	**1**	**2**
Region structure	2	**8 regions**	5 regions	
Steepness	3	**0.8**	0.65	0.95
Length comp. wtg	3	50	**100**	200
Mixing period (qtr)	2	**1**	2	
Growth	3	**Default**	Low growth	High growth

Table 1: Skipjack 2019 stock assessment uncertainty grid (Vincent et al., 2019).

2 Accounting for uncertainty

Stock assessments conducted by the Pacific Community (SPC) have typically presented a range of model configurations, termed the uncertainty grid. The grid explores the sensitivity of the assessment results to alternative assumptions about model settings for which the data are often uninformative such as growth rates or the steepness of the stock and recruitment relationship. Management advice is then based on the range of model outcomes rather than on a single model.

The stock assessment uncertainty grid is a useful starting point for considering the range of uncertainty that should be included in the suite of OMs for the MSE analyses. However, the assessment uncertainty grid is concerned primarily with those factors that impact the historical trajectory of the stock, as estimated by the stock assessment. When projecting assessment results forwards in time, as performed by the MSE simulations, it may be necessary to consider a different set of sensitivities in order to adequately capture the most important sources of uncertainty.

A number of previous studies have categorised the types of uncertainty based on their different sources (Rosenberg and Restrepo, 1994; Francis and Shotton, 1997; Kell et al., 2007). Punt et al. (2014) recommend that, at minimum, an MSE should consider (i) process uncertainty; (ii) parameter uncertainty and (iii) observation uncertainty. They note, however, that the choice of uncertainty to include in any MSE will be case specific.

A detailed consideration of the elements to be considered when conditioning operating models for WCPO skipjack has previously been submitted to SC14 (Scott et al., 2018b). We summarise the key elements included in the OM grid for each of the three main sources of error and reconsider them in the context of the 2019 stock assessment.

2.1 Process Uncertainty

Process uncertainty arises through natural variability in the biotic and abiotic processes that impact on population dynamics. Perhaps the most significant effect of this natural variability is in the number of fish that survive the larval stages and recruit to the fishery. The scale of, and potential variability in, future recruitment is a particularly important consideration when running projections. The assumption most often adopted is that future variability in recruitment will be similar to that observed in the past.

Estimates of past recruitment for the WCPO skipjack stock show a long-term pattern of steadily increasing recruitment with a more recent period of relatively stable but high recruitment. As for the previous MSE uncertainty grid, two time periods are specified from which to re-sample recruitment; a long-term time series that includes the lower recruitments observed in the 1980's and a short-term time series that assumes that recent, higher recruitment levels will continue into the future. However, the difference in mean recruitment levels between these two periods is less marked than it has been for previous assessments (Figure 19).

2.2 Observation Uncertainty

Observation uncertainty relates to the accuracy with which something is measured. It includes natural errors that occur in any data collection procedure as well as systematic errors (affecting all measurements) that can arise from, for example, the miscalibration of instruments. Observation uncertainty is a key source of uncertainty and a particularly important consideration with respect to the input data to the MP. The key input data include fishery specific estimates of catch, effort, size composition and tag recaptures, all of which will be subject to observation error to a greater or lesser extent.

Consistent with the previous OM uncertainty grid, a C.V. of 20% is applied to future catch and effort whilst the size frequency data are generated by sampling from a multinomial distribution with a fishery specific effective sample size as determined from a MULTIFAN-CL fit using the self-scaling multinomial minimisation option (Kleiber et al., 2018). Settings for the generation of simulated future catch, effort and size frequency data subject to different levels of observation uncertainty are further considered in WCPFC-SC16/MI-IP-10 along with further details of the methods employed.

When generating tag recapture data for the future the number and spatial distribution of tag releases must be specified by the user. Observation uncertainty is introduced into the tag recapture data based upon the OM estimation of the multinomial probability of recapture given the release samples. In this sense the probability of recapture of tagged fish is determined from the internal calculations of MULTIFAN-CL and cannot be specified by the user. The user must, however, specify the quantity of tags to be released, the regions from which those releases will be made and the fishery selectivity from which the length distribution of the releases will be generated.

Assessment	**Tag Release Programmes**							
Region	**SSAP**		**RTTP**		**PTTP**		**JPTP**	
	1977-1980		1989-1992		2006-2018		1998-2018	
1	0		0		0		19	(154)
2	0		0		0		47	(384)
3	1	(82)	0		0		55	(578)
4	1	(162)	0		0		30	(287)
5	2	(2662)	6	(2179)	3	(7332)	2	(205)
6	3	(3875)	10	(2414)	16	(5652)	2	(56)
7	7	(1084)	8	(934)	12	(1406)	20	(590)
8	9	(3972)	5	(2021)	2	(3424)	9	(506)

Table 2: Tag release summary: Number of tag release events by region and tagging program. The average number of fish tagged and released is shown in brackets.

A summary of historical tag releases is shown in Table 2. As for the previous OM uncertainty grid it is assumed that existing tagging programmes (PTTP and JPTP) will continue into the future.

Future tag releases in each region are set to the sum of the average regional releases for these two programmes.

2.3 Model Uncertainty and Parameter Uncertainty

Model uncertainty arises from the possibility that the model is mis-specified, whereas parameter uncertainty is the possibility that the parameters used to define the model are incorrect, given that the model form is correct. Parameter uncertainty occurs because there is only a limited amount of data from which to estimate the parameters and because the parameters themselves may evolve through time. For the purpose of this analysis we consider model uncertainty and parameter uncertainty together as a single category.

Many of the main sources of model uncertainty have already been considered during the development of the stock assessment (Vincent et al., 2019). The OM grid is developed from this starting point, but makes a small number of minor changes.

One axis of uncertainty has been dropped from the grid. Alternative weighting of the size frequency data in the assessment had the least impact on model results and, as for the previous conditioning of skipjack OMs, has not been carried forward to the OM grid.

Of the three growth models considered for the stock assessment the default and high growth were almost identical (Figure 5). Consequently the default growth model has been dropped and only the low growth and high growth models are considered further.

In addition to the stock assessment axes of uncertainty, autocorrelation in recruitment has previously been included in the skipjack OM grid although it had very limited influence on the model results. Autocorrelation in recruitment was again investigated, however, the extent of autocorrelation in recruitment was very low for all models in the grid (Figure 4c) and the impact on model estimates (specifically depletion) was negligible (Figure 6e). Consequently, autocorrelation in recruitment was not considered further and is not represented in the OM grid.

As for previous skipjack OM grids, a further axis of uncertainty was added to the grid for hyperstability in CPUE (also termed density dependent catchability, (Scott et al., 2015)). The settings for hyperstability in CPUE are the same as those assumed for the previous skipjack OM grid with two values assumed in the reference set of OMs (0 and -0.5) and a more severe hyperstability value of -0.9 considered in the robustness set.

3 The skipjack OM grid

The revised skipjack OM grid based on the updated 2019 assessment (Table 3) is very similar to the previous OM grid based on the 2016 stock assessment. The axes of uncertainty considered in the reference set and their respective settings have changed very little.

Axis	**Levels**		**Options**		
	Reference	Robustness	**0**	**1**	**2**
Process Error					
Recruitment Variability	**2**		**1982-2018**	**2005-2018**	
Observation Error					
Catch and effort	**1**	1	**20%**	30%	
Size composition (ESS)	1		**estimated**		
Tag recaptures	**1**	2	**status quo**	low	none
Model Error					
Steepness ‡	**3**		**0.8**	**0.65**	**0.95**
Mixing period (qtr) ‡	**2**		**1**	**2**	
Growth ‡	**2**			**low**	**high**
Movement	**1**	1	**estimated**	El Nino/La Nina	
DD catchability (k) ‡	**2**	1	**0**	**-0.5**	-0.9
Implementation Error					
Effort creep	**2**	1	**0%**	**2%**	3%

Table 3: Table: Skipjack OM uncertainty grid. Scenarios shown in bold are proposed for the reference set. ‡ denotes those scenarios for which a dedicated fit of MULTIFAN-CL is required.

3.1 Reference and robustness sets

It is considered best practice to divide the suite of OMs into a reference set and a robustness set (Rademeyer et al., 2007). The reference set is considered to reflect the most plausible hypotheses and forms the primary basis for identifying the 'best' management strategy. The robustness set comprises hypotheses that are considered less likely but still plausible.

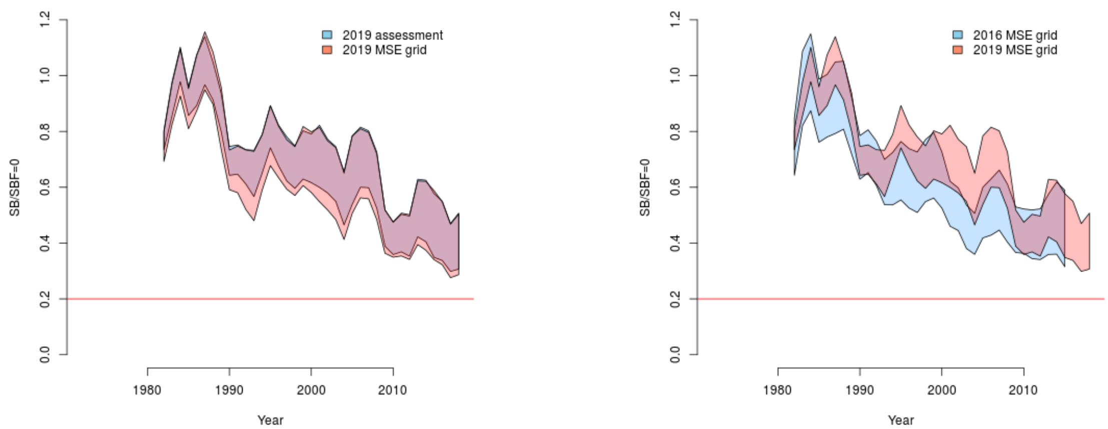
The performance indicators will be calculated from the reference set whilst the robustness set will be used to give a secondary indication of the performance of the management procedure. Work continues to further investigate and define the range of scenarios that should be considered in the robustness set.

3.2 Comparison with 2019 assessment model grid

The revised skipjack OM grid differs from the 2019 assessment uncertainty grid in that alternative weighting of the size frequency data has been omitted; the default growth option is omitted; and hyperstability in CPUE has been included. In spite of these changes the range of depletion levels estimated for the OM grid is very consistent with that estimated by the stock assessment uncertainty grid (Figure 3a). The 95th percentiles of the ranges of estimated depletion overlap throughout the time series with the OM grid estimating slightly lower depletion in some years for the hyperstable CPUE model runs.

3.3 Comparison with 2016 OM model grid

Key differences between the previous 2016 OM grid and the revised OM grid include the change to an eight region spatial structure with 31 fisheries; the re-definition of purse seine effort as number of sets rather than days fished and revised data inputs including re-worked tag release and recapture information. The difference in the range of depletion estimates between the two grids is more marked (Figure 3b), however the overall trend in depletion remains consistent between the two and the estimates correspond very well in the terminal years.



(a) 2019 stock assessment vs 2019 OM grid

(b) 2016 OM grid vs 2019 OM grid

Figure 3: Comparison of depletion (95 %ile range) for the 2019 stock assessment grid (54 models) and the 2019 (24 models) and 2016 (72 models) OM grids.

4 Model validation

The primary purpose of 'conditioning' the OMs is to ensure that all important sources of uncertainty have been appropriately accounted for. It is, therefore, particularly important to consider how accurately each of the models in the OM grid represents the real world and whether the type and scale of uncertainty generated from it represents a plausible future scenario against which a management procedure should be tested. There is no simple test to establish the validity of a model and instead we rely on a collection of indicators, based on diagnostics of the fit of the model to data, and consideration of whether the quantities estimated from it are reasonable.

4.1 Model diagnostics

A range of diagnostics are presented to describe the overall performance of the model; the fit of the model to specific data sets; as well as the plausibility and consistency of model estimates.

Many of these diagnostics are routinely presented for WCPFC stock assessments. Appropriate diagnostics for model fits are further considered with respect to both stock assessments and OM conditioning in WCPFC-SC16/MI-IP-07. In this paper we provide a number of model validation diagnostics with specific reference to the conditioning of OMs for WCPO skipjack. Depending on the information to be conveyed, diagnostics may be presented either for individual model runs or as summary diagnostics across the full grid of OMs.

It would be impractical to present a comprehensive set of all model diagnostics in this report. Instead we present only a subset of the available diagnostic plots and tables. A full set of diagnostics is provided in an accompanying shiny app which can be accessed at https://ofp-sam.shinyapps.io/hierophant/

4.1.1 Overall model fit diagnostics

Diagnostics of the overall model fit provide an indication of the extent to which the model has converged to a stable solution. They include the maximum gradient of the estimated parameters (Table 5, Figure 4b) as well as likelihood profiles of key parameters and model outputs (Figure 22). For the skipjack OM grid the maximum gradient is consistently less than 0.001 across all model runs (the same convergence criteria as for the stock assessment). At a finer scale, likelihood profiles of key parameters and model outputs are often used to determine how well a particular parameter is estimated by the model and to determine how consistently that parameter is estimated by the different sources of information available to the model (CPUE, size composition, tag data, etc.). A key model output is the estimate of stock status (in this case depletion, $SB/SB_{F=0}$) that is used as an important performance indicator (PIs 1,8 Scott et al. (2018a)). The profiles (Figure 22) indicate that depletion is well estimated by the OM with the point estimate of $SB/SB_{F=0}$ corresponding with the lowest value of the total likelihood profile.

4.1.2 Model fit to data diagnostics

Perhaps the most informative set of diagnostics are those that describe the consistency between model estimates and observed data. For the WCPO skipjack assessment these include plots of the observed vs predicted catch and CPUE (Figures 9 and 10), the scale of the effort deviations (Figures 7 and 8), the fit of the predicted size composition to observed length frequency data (Figures 11 and 12) and the model predicted vs observed tag recaptures (Figure 21). We present these diagnostics either for individual model runs or, where practical, across the full OM grid of 24 models. We also focus on the purse seine and pole and line fisheries. The full set of diagnostic plots can be viewed in the MFCL diagnostics shiny app which is described further in WCPFC-SC16/MI-IP-07.

The OM grid corresponds closely to the 2019 stock assessment grid. Consequently model diagnostics are very similar to those presented in the 2019 stock assessment report (Vincent et al., 2019). Model diagnostics across the grid of models are generally very consistent. Effort deviates for the purse seine and pole and line fisheries (Figures 9 and 10) generally show very little variation across the grid of models. Similarly the relationship between model predicted and observed tag recaptures (Figure 21) is broadly consistent across the model grid depending on the mixing period assumption.

Overall, the model fit to data diagnostics are consistent with those of the 2019 assessment and show broadly consistent model fits across the OM grid.

4.1.3 Model output diagnostics

Model output quantities are not, strictly speaking, diagnostics but warrant inspection nonetheless to ensure that implausible values are not being estimated. For the WCPO skipjack OM grid such quantities include the fishery specific selection patterns (Figure 13) and model estimates of biological processes including natural mortality (Figure 14), recruitment (Figures 15 to 20), movement rates as well as adult biomass and stock depletion (Figure 6). Once again, we provide a subset of diagnostics here, the full set being available from the MFCL diagnostics shiny app.

Model outputs are broadly consistent across the grid but vary depending on model settings for the key sources of uncertainty included in the OM grid. For example, fishery specific estimates of selection pattern (Figure 13) differ depending on the assumed growth model (Figure 5) and estimates of natural mortality at age (Figure 14) differ depending on the assumed tag mixing period. The deviation of estimated recruitment from the assumed SRR (Figure 17) and the temporal and spatial distribution of recruits (Figure 18) are consistently estimated across the OM grid of models.

Across the OM grid, none of the models provide implausible outputs that would indicate failure to fit.

4.1.4 Model consistency diagnostics

The final set of diagnostics relate to the internal consistency of the models and their stability in terms of providing consistent model estimates. Retrospective analyses are an important diagnostic tool for determining the robustness of model estimates to varying quantities of data and in recent years have been routinely presented for each new stock assessment. Retrospective analyses were conducted across the OM grid, whereby the final assessment model was refitted to a progressively truncated time series of data over a 5 year period (ie. the terminal year of the assessment is iteratively moved backward from 2018 to 2013). Results are shown for 3 models from the OM grid (Figure 23). The full set of retrospectives can be viewed in the MFCL diagnostics shiny app.

The results show very consistent estimates of depletion are achieved throughout the retrospective period with no indication of persistent retrospective bias and little indication of variation between each successive estimate of depletion.

5 Discussion

5.1 Procedures for updating the OM grid under the Harvest Strategy approach

Under the harvest strategy approach the stock assessment procedure, as it is currently performed, will take on a different role. It will no longer be used as the source of stock status to manage the fishery. This role is now conducted by the selected management procedure (i.e. the combination of an agreed data collection program, estimation model and harvest control rule). However, a full stock assessment will continue to be routinely conducted as part of the monitoring strategy (see WCPFC-SC16/MI-IP-02). The settings for this full stock assessment may vary (which is in contrast to the fixed estimation model within the management procedure that should not change through time). The assessment will be based on all available data and the most recent modeling approaches and will represent the best available science. The full assessment will be used to determine if the OM grid continues to adequately represent stock status and to check that all important sources of uncertainty are adequately represented in the simulation framework. Therefore, it should not always be necessary to update the suite of operating models each time a new stock assessment is conducted. If the OM grid remains broadly consistent with the full assessment conducted by the monitoring strategy there should be no need to change it.

It is recommended that the full stock assessment, conducted as part of the monitoring strategy, does not take place in the same year that the management procedure is run. This serves both to reduce the workload for a stock in a given year and to clearly delineate the two separate tasks of running the management procedure and checking the validity of the OM grid.

6 Conclusions

Model diagnostics for the revised OM grid are consistent with those of the 2019 assessment and show broadly consistent model fits across the range of uncertainty axes. None of the models provide implausible outputs that would indicate failure to fit.

The overall trends in depletion estimated for the revised OM grid are broadly consistent with those of the previous 2016 OM grid. The estimates correspond particularly well in the terminal years. We note that it should not always be necessary to update the suite of operating models each time a new stock assessment is conducted. Provided that the OM grid remains broadly consistent with the full assessment conducted by the monitoring strategy there should be no need to change it.

The grid of models outlined in this report form the basis of the updated MP evaluations provided in WCPFC-SC16/MI-IP-03.

Acknowledgments

We gratefully acknowledge funding for this work from the New Zealand Ministry of Foreign Affairs and Trade (MFAT) funded project "Pacific Tuna Management Strategy Evaluation". In addition we thank the Center for High Throughput Computing (CHTC UW-Madison) for generously providing access to their computing resources.

References

Francis, R. I. C. C. and Shotton, R. (1997). "Risk" in fisheries management: a review. *Canadian Journal of Fisheries and Aquatic Science*, 54:1699–1715.

Hoshino, E., Hillary, R., Davies, C., Satria, F., Sadiyah, L., Ernawati, T., and Proctor, C. (2020). Development of pilot Empirical harvest control rules for tropical tuna in Indonesian archipelagic waters: Case studies of skipjack and yellowfin tuna. *Fisheries Research*, 227(doi.org/10.1016/j.fishres.2020.105539).

Kell, L., Mosqueira, I., Grosjean, P., Fromentin, J., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M., Poos, J., Scott, F., and Scott, R. D. (2007). FLR: an open source framework for the evaluation and development of management strategies. *ICES Journal of Marine Science*, 64:640–646.

Kleiber, P., Hampton, J., Davies, N., Hoyle, S. D., and Fournier, D. (2018). *MULTIFAN-CL User's Manual*. http://www.multifan-cl.org/.

McKechnie, S., Hampton, J., Pilling, G. M., and Davies, N. (2016). Stock assessment of skipjack tuna in the western and central Pacific Ocean. WCPFC-SC12-2016/SA-WP-04, Bali, Indonesia, 3–11 August 2016.

Punt, A. E., Butterworth, D., de Moor, C., De Oliveira, J., and Haddon, M. (2014). Management strategy evaluation: best practices. *Fish and Fisheries*, (DOI:10.111/faf12104).

Rademeyer, R., Plaganyi, E., and Butterworth, D. (2007). Tips and tricks in designing management procedures. *ICES Journal of Marine Science*, 64:618–625.

Rosenberg, A. and Restrepo, V. (1994). Uncertainty and risk evaluation in stock assessment advice for U.S. marine fisheries. *Canadian Journal of Fisheries and Aquatic Sciences*, 51:2715–2720.

Scott, F., Scott, R. D., Pilling, G., and Hampton, J. (2018a). Performance indicators for comparing management procedures using the MSE modelling framework. WCPFC-SC14-2018/MI-WP-04, Busan, South Korea, 5–13 August 2018.

Scott, R. D., Scott, F., Davies, N., Pilling, G., and Hampton, J. (2018b). Selecting and conditioning the operating models for WCPO skipjack. WCPFC-SC14-2018/MI-WP-03, Busan, South Korea, 5–13 August 2018.

Scott, R. D., Scott, F., Yao, N., Pilling, G., and J., H. (2019). Results of inital evaluations of management procedures for skipjack. WCPFC-SC15/MI-WP-05, , Pohnpei, Federated States of Micronesia, 12–20 August 2019.

Scott, R. D., Tidd, A., Davies, N., Pilling, G., and Harley, S. (2015). Implementation of alternative CPUE/abundance dynamics for purse seine fisheries within MULTIFAN-CL with application to effort-based projections for skipjack tuna. WCPFC-SC11-2015/SA-IP-02, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

Vincent, M., Pilling, G. M., and J., H. (2019). Stock assessment of skipjack tuna in the western and central pacific ocean. WCPFC-SC15-2019/SA-WP-05 (rev 2), Pohnpei, Federated States of Micronesia. 12-20 August, 2019.

A Appendix

A.1 Model validation tables

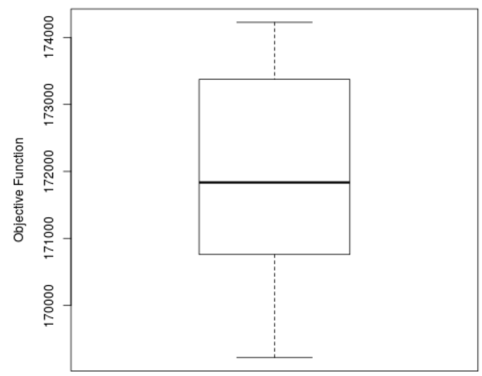
Axis	**Letter**	**Levels**	**Options**		
			0	**1**	**2**
Steepness	A	3	**0.8**	0.65	0.95
Mixing period (qtr)	B	2	**1**	2	
Growth	C	2	**-**	Low growth	High growth
Hyperstability CPUE	D	2	**0**	-0.5	

Table 4: WCPO Skipjack model uncertainty grid. Values in bold show settings for the diagnostic case. Each model can be identified by its unique letter-option combination (e.g A0B0C2D0).

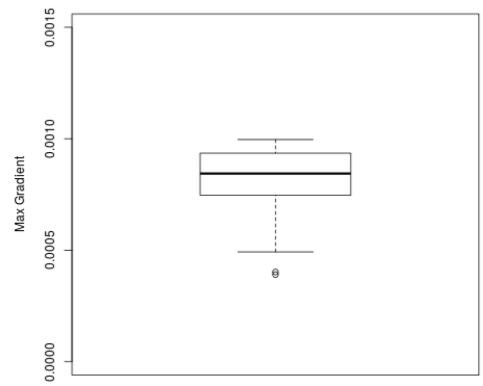
Model	Likelihood					Max	Model Estimates		
	Total	CPUE	Catchability	Size Comp	Tag	gradient	SB/SBF0	MSY	BMSY
A0B0C1D0E0	170378.2679	2542.9191	116.9353	-200505.0879	27466.811	8e-04	0.403	554600	952000
A1B0C1D0E0	170378.2766	2542.9384	116.9256	-200505.1081	27466.8132	0.0005203	0.372	576300	1305000
A2B0C1D0E0	170378.2742	2542.9209	116.9515	-200505.0872	27466.7841	0.0009532	0.423	554200	671600
A0B1C1D0E0	173187.7	2541.3117	120.3201	-200382.8212	24533.369	0.000781	0.5	688300	1185000
A1B1C1D0E0	173187.7033	2541.3145	120.317	-200382.7981	24533.3428	0.0009831	0.473	703800	1568000
A2B1C1D0E0	173187.721	2541.3117	120.3165	-200382.7999	24533.3292	0.0008149	0.517	692300	870800
A0B0C2D0E0	172122.5828	2492.5907	112.5802	-200230.5391	25502.6217	0.0007366	0.363	498300	980500
A1B0C2D0E0	172142.3866	2466.1601	113.1386	-200229.2582	25507.4098	0.0009354	0.33	526100	1330000
A2B0C2D0E0	172142.3876	2466.1897	113.1226	-200229.2305	25507.3642	0.0009422	0.384	492000	717900
A0B1C2D0E0	174784.3933	2510.7894	117.1219	-200110.5754	22698.1421	0.0008333	0.439	579900	1175000
A1B1C2D0E0	174822.7637	2464.0794	117.4581	-200109.6374	22705.2073	0.0007246	0.41	601000	1542000
A2B1C2D0E0	174784.3935	2510.7795	117.1047	-200110.5807	22698.1726	0.0009899	0.457	576600	887300
A0B0C1D1E0	169940.8432	2820.0185	120.2885	-200317.6142	27436.3146	0.0009539	0.398	564300	874600
A1B0C1D1E0	169940.836	2820.0567	120.2831	-200317.6343	27436.3083	0.000887	0.367	584300	1221000
A2B0C1D1E0	169940.8472	2820.0217	120.2839	-200317.6385	27436.3355	0.0009335	0.418	567200	583000
A0B1C1D1E0	172844.6822	2751.7318	123.3578	-200214.2523	24494.3638	0.0008482	0.474	682400	1081000
A1B1C1D1E0	172844.6759	2751.7149	123.3807	-200214.2414	24494.3527	0.0009904	0.448	696700	1455000
A2B1C1D1E0	172845.7747	2747.9919	123.4248	-200214.1057	24496.7966	0.0007109	0.491	689200	759900
A0B0C2D1E0	171361.5166	3090.2898	114.9377	-200095.6456	25528.7374	0.0006534	0.32	477500	966900
A1B0C2D1E0	171361.5611	3090.3508	114.8968	-200095.645	25528.6702	0.0009557	0.288	507500	1325000
A2B0C2D1E0	171361.5507	3090.3423	114.9175	-200095.6638	25528.6884	0.0008776	0.341	469600	697500
A0B1C2D1E0	174082.163	3066.5752	120.0571	-199964.5716	22695.648	0.0008876	0.387	550700	1135000
A1B1C2D1E0	174082.1585	3066.6103	120.0309	-199964.5452	22695.6163	0.0008888	0.359	573900	1504000
A2B1C2D1E0	174082.1758	3066.5703	120.0558	-199964.5875	22695.6563	0.0007189	0.405	546500	848500

Table 5: WCPO Skipjack: likelihood components and model estimates summary table for the 24 model grid. (Depletion is calculated as SB/SB$_{F=0}$latest).

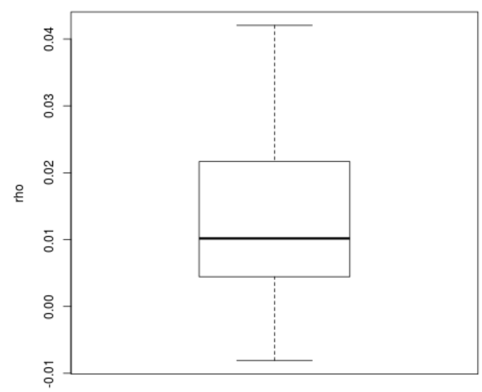
A.2 Model validation figures



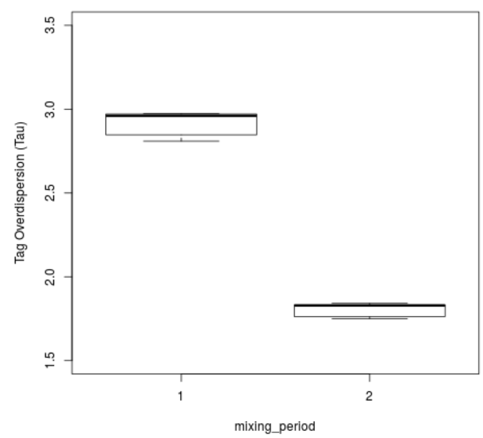
(a) Objective function



(b) Maximum gradient



(c) Autocorrelation estimate



(d) Tag overdispersion estimate

Figure 4: OM diagnostics across the 2019 MSE grid (72 models).

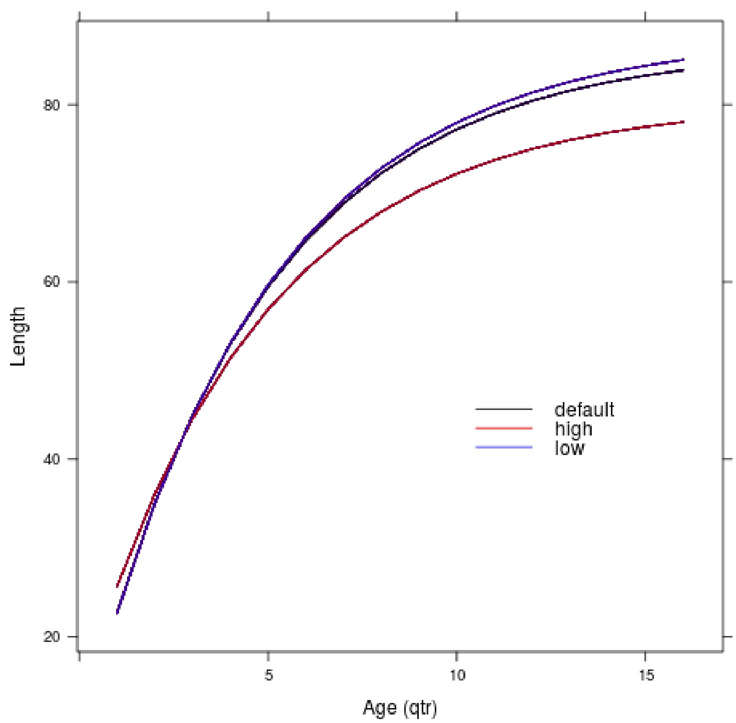
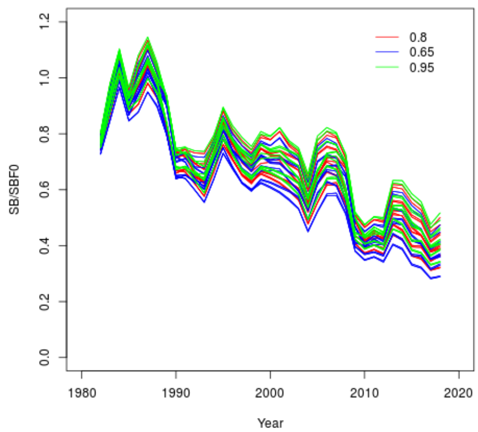
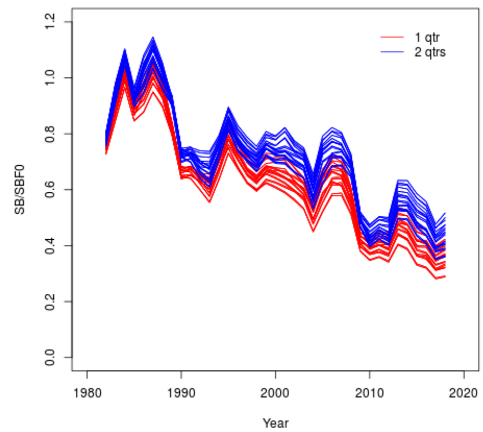


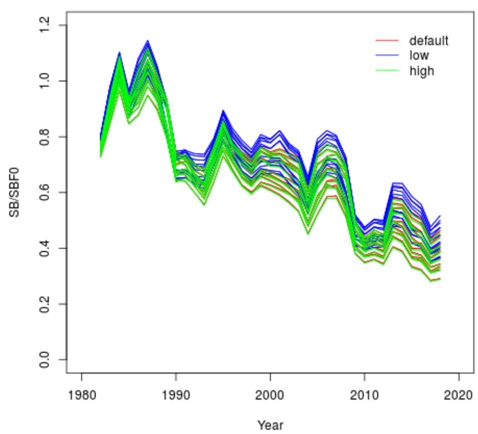
Figure 5: Growth models.



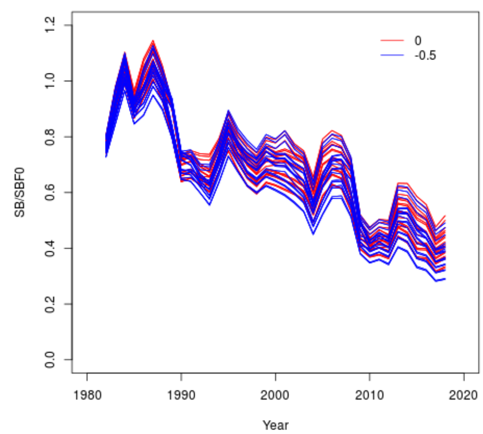
(a) Steepness of the SRR



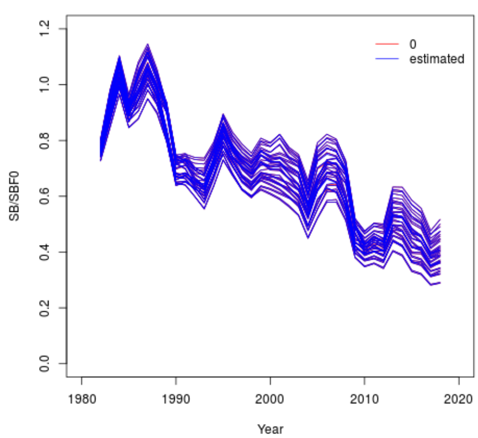
(b) Tag mixing period



(c) Growth



(d) Density dependent CPUE



(e) Recruitment autocorrelation

Figure 6: OM diagnostics across the 2019 MSE grid (72 models).

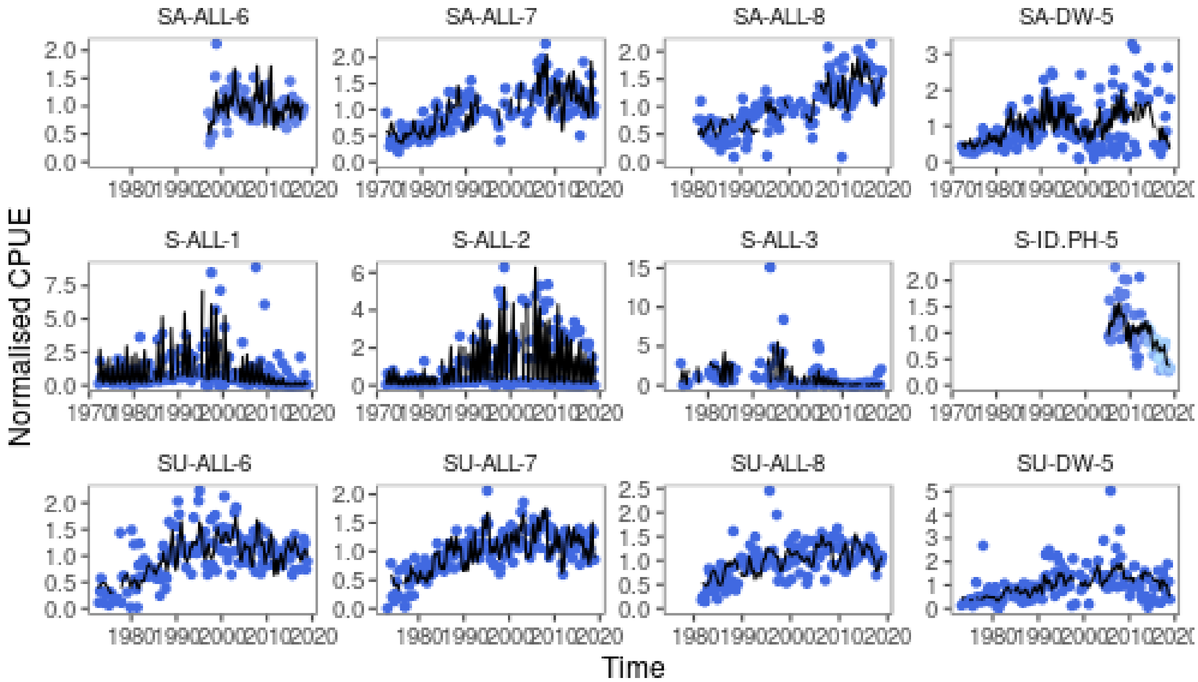


Figure 7: Observed vs prediced CPUE for purse seine fisheries (single model estimates A0B0C1D0E0).

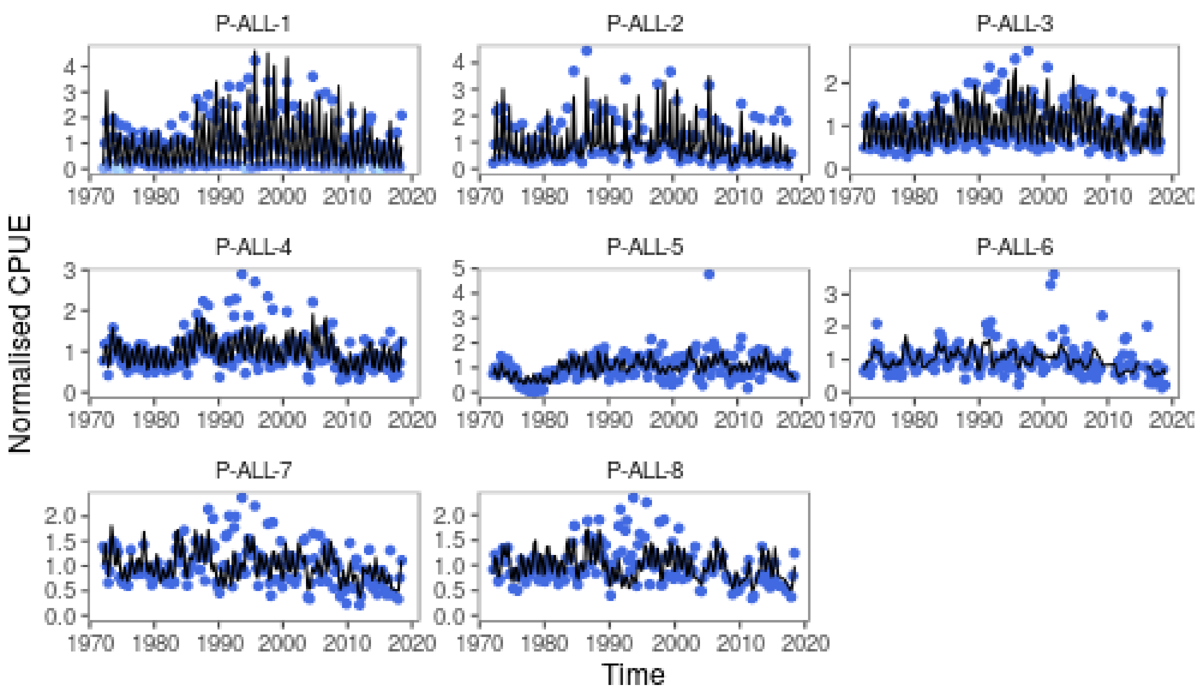


Figure 8: Observed vs prediced CPUE for pole and line fisheries (single model estimates A0B0C1D0E0).

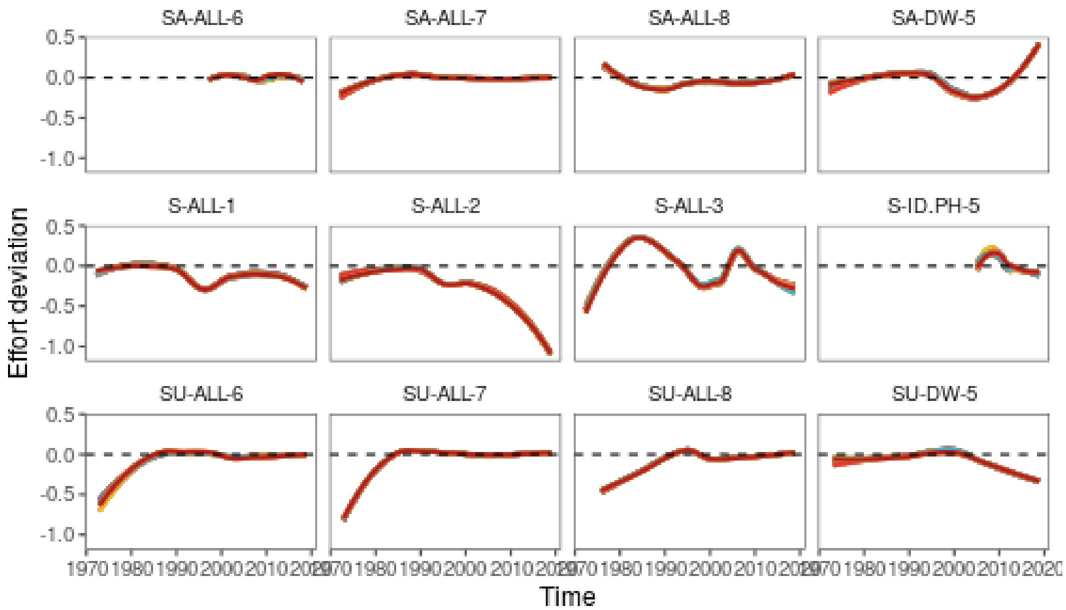


Figure 9: Effort deviates for purse seine fisheries across the OM grid (24 models).

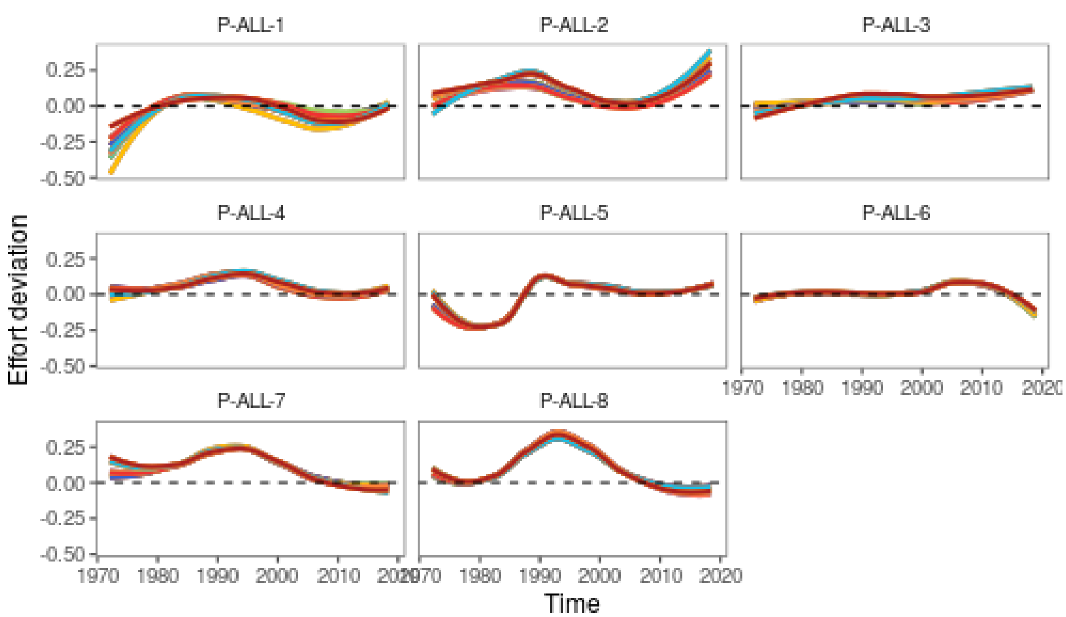


Figure 10: Effort deviates for pole and line fisheries across the OM grid (24 models).

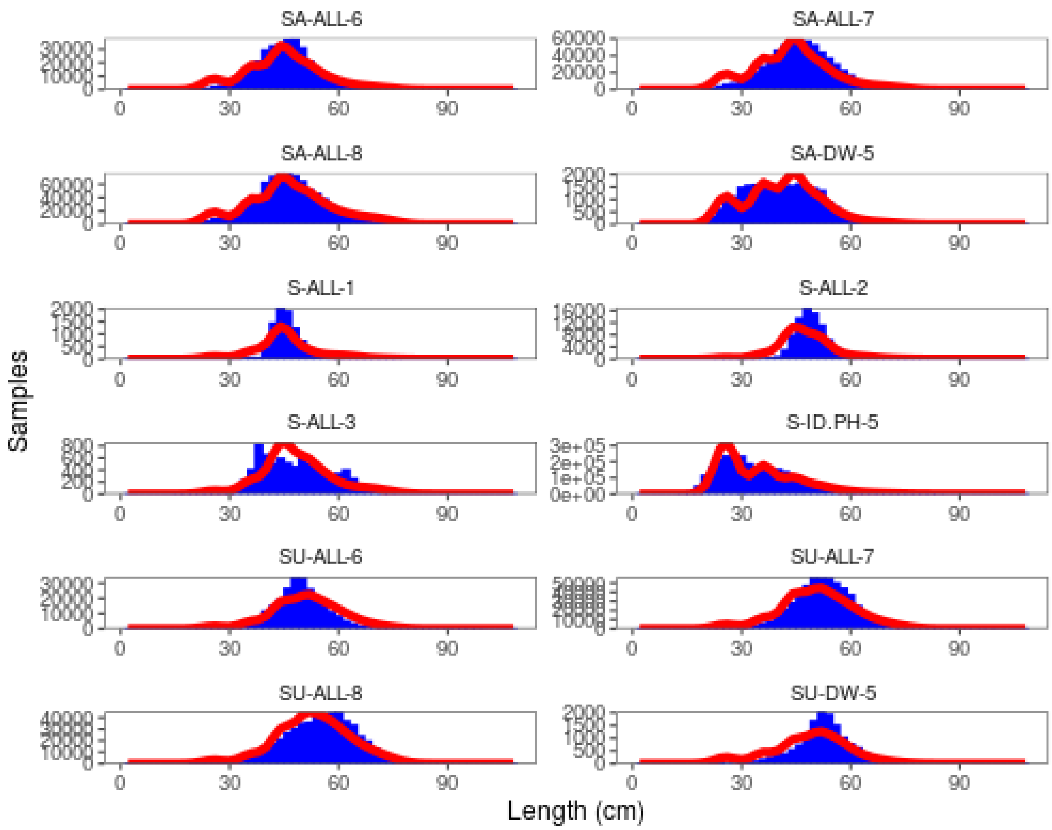


Figure 11: Fits to size composition data (all time periods combined) for purse seine fisheries (single model estimates A0B0C1D0E0).

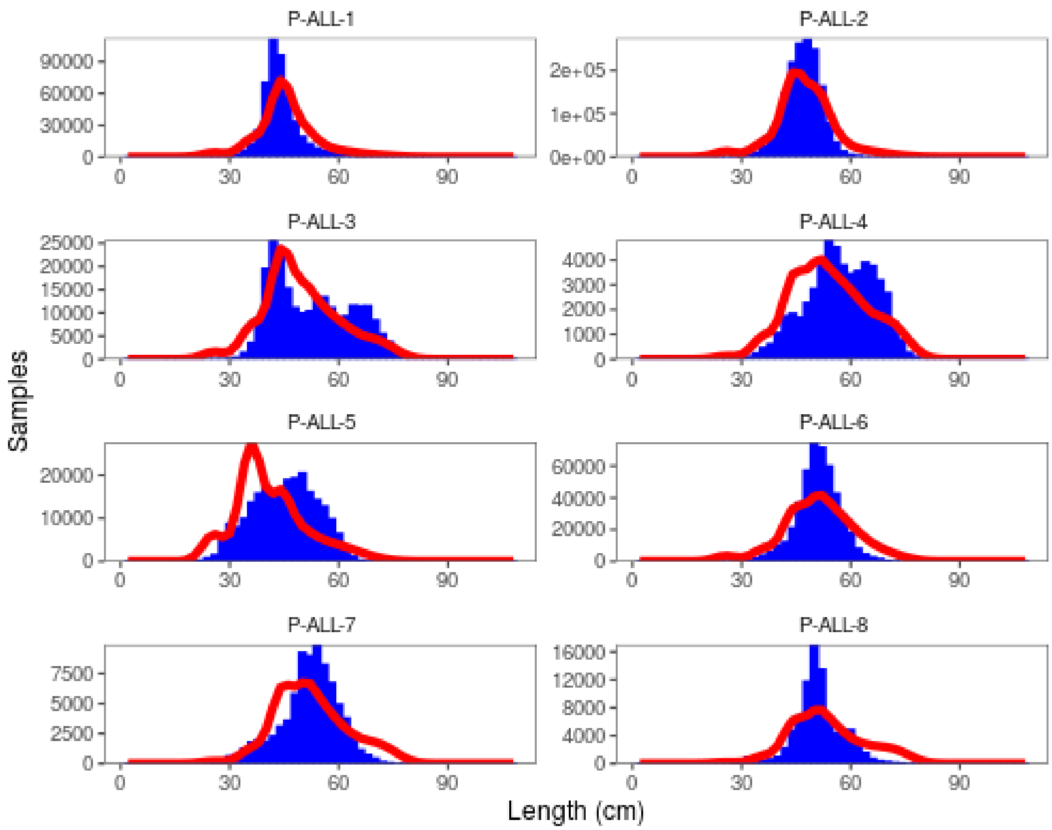


Figure 12: Fits to size composition data (all time periods combined) for pole and line fisheries (single model estimates A0B0C1D0E0).

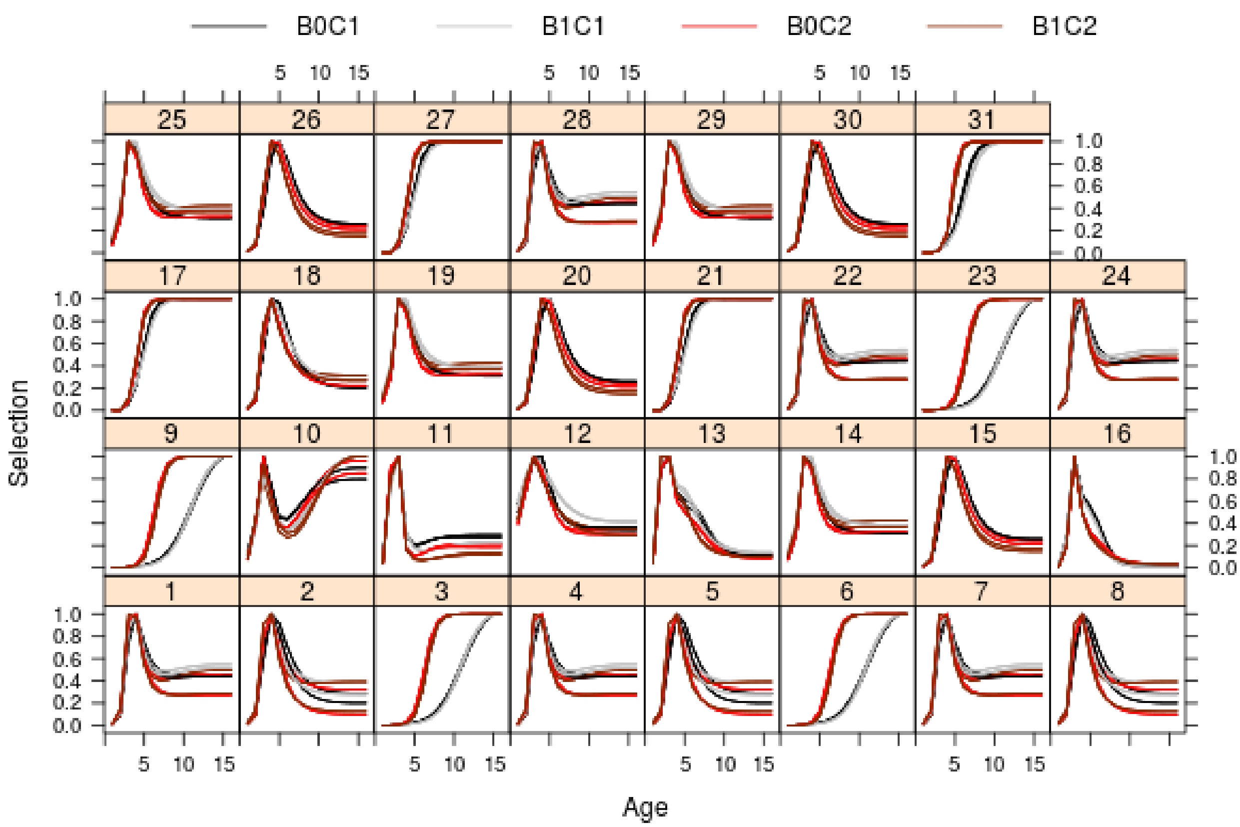


Figure 13: Fishery specific selection patterns across the OM grid (24 models). Low growth and 1qtr tag mixing (B0C1, black), low growth and 2qtr tag mixing (B1C1, grey), high growth and 1qtr tag mixing (B0C2, red), high growth and 2qtr tag mixing (B1C2, brown).

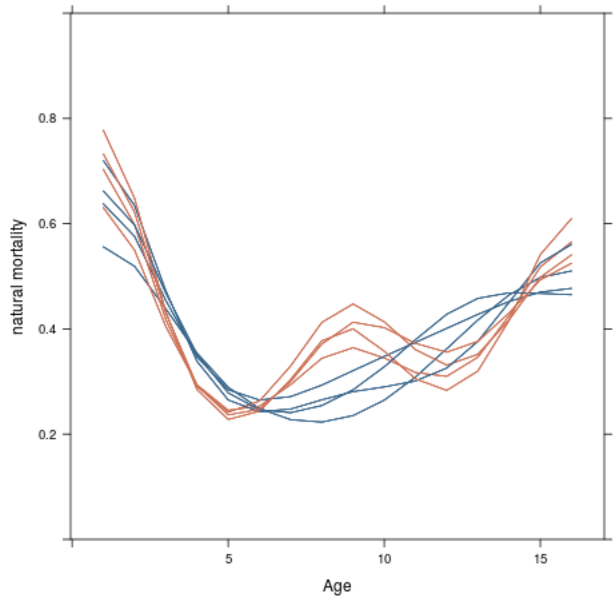


Figure 14: Natural mortality at age across the OM grid (24 models). 1qtr tag mixing (blue), 2qtr tag mixing (red).

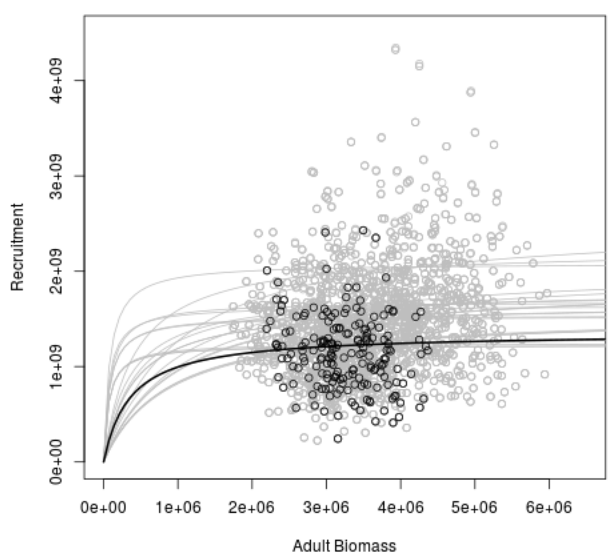


Figure 15: Stock and recruitment pairs and the fitted Beverton Holt stock and recruitment relationship across the OM grid (24 models, diagnostic case shown in black).

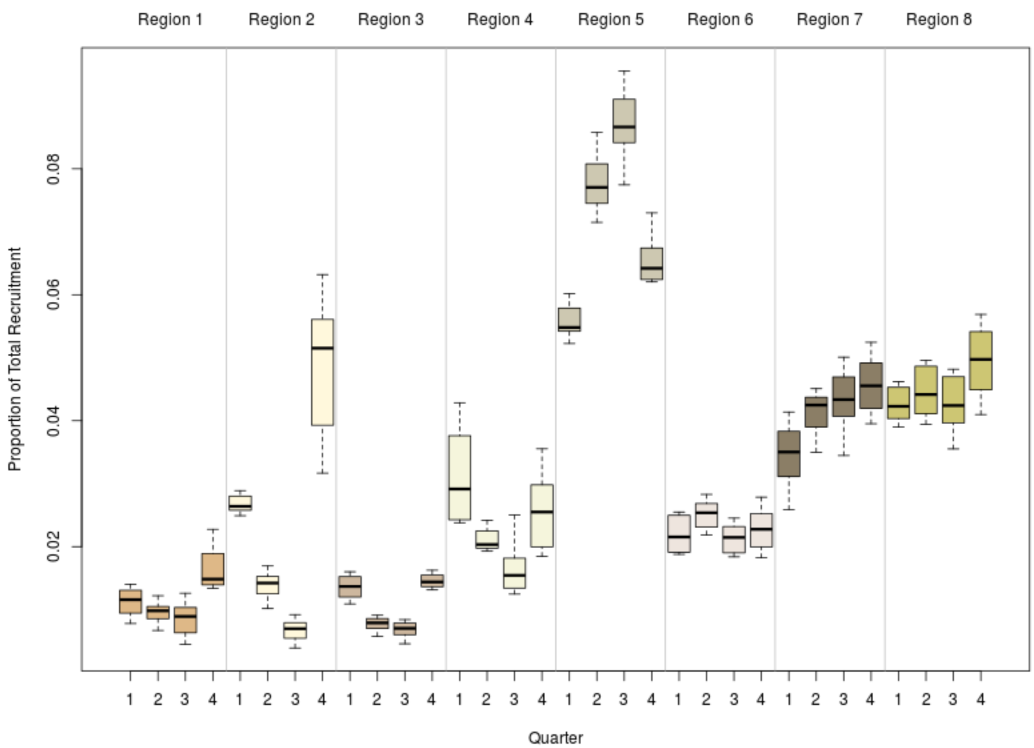


Figure 16: Quarterly regional recruitment distribution across the OM grid (24 models) averaged over the year range 1982:2018.

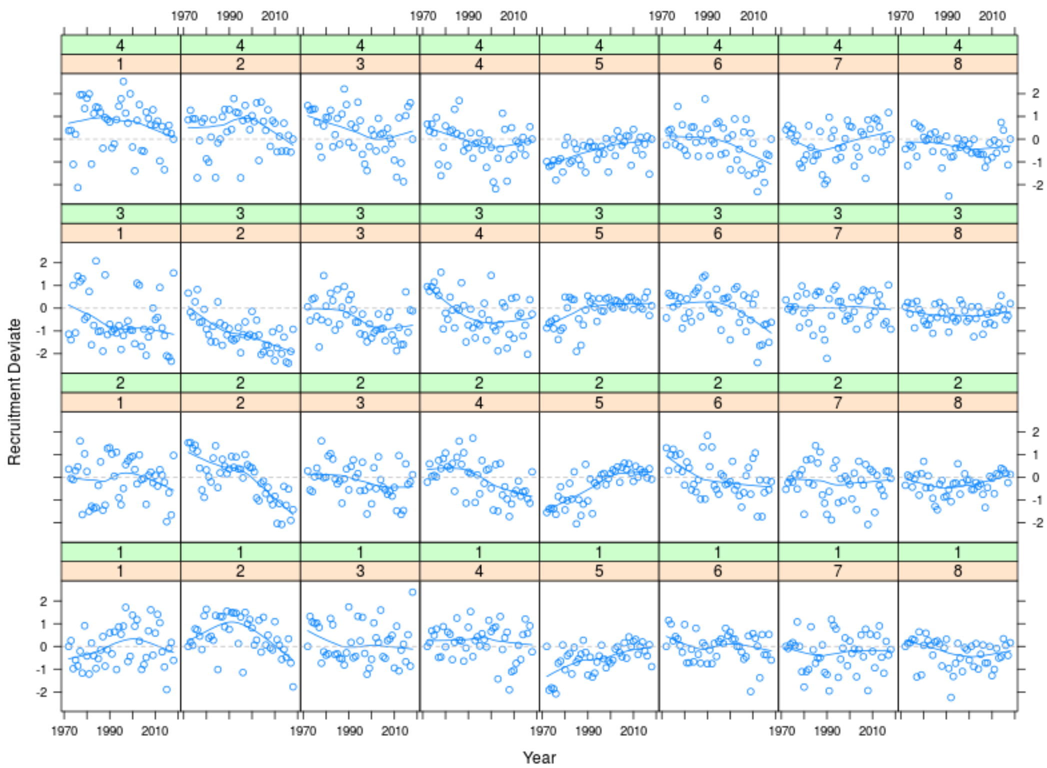
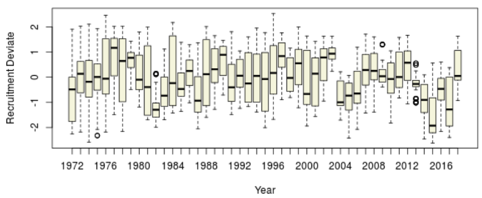
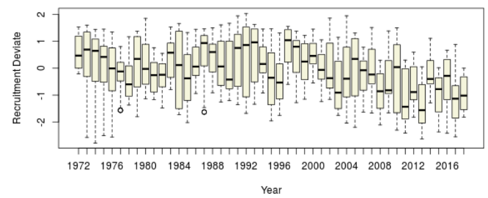


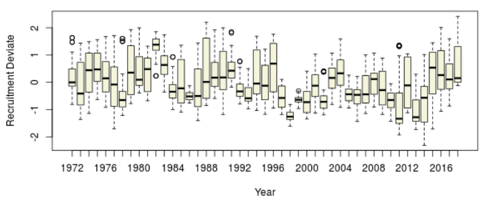
Figure 17: Recruitment deviates by region (columns) and season (rows) for the diagnostic case OM.



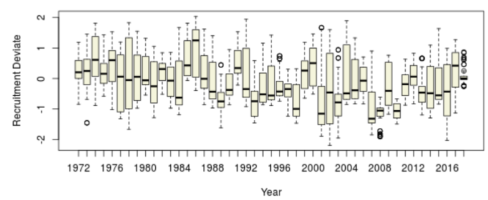
(a) Region 1



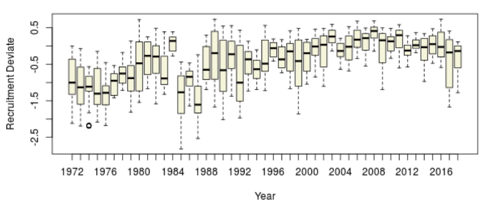
(b) Region 2



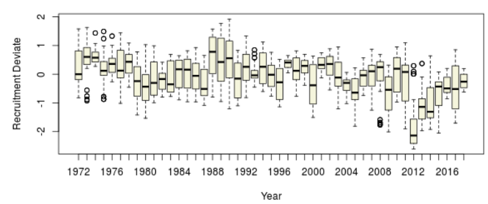
(c) Region 3



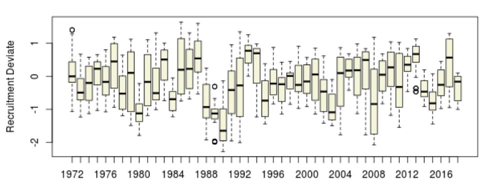
(d) Region 4



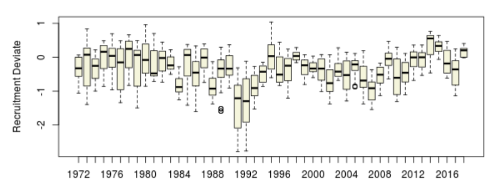
(e) Region 5



(f) Region 6



(g) Region 7



(h) Region 8

Figure 18: Recruitment deviates by year (seasons combined) and region across the OM grid (24 models).

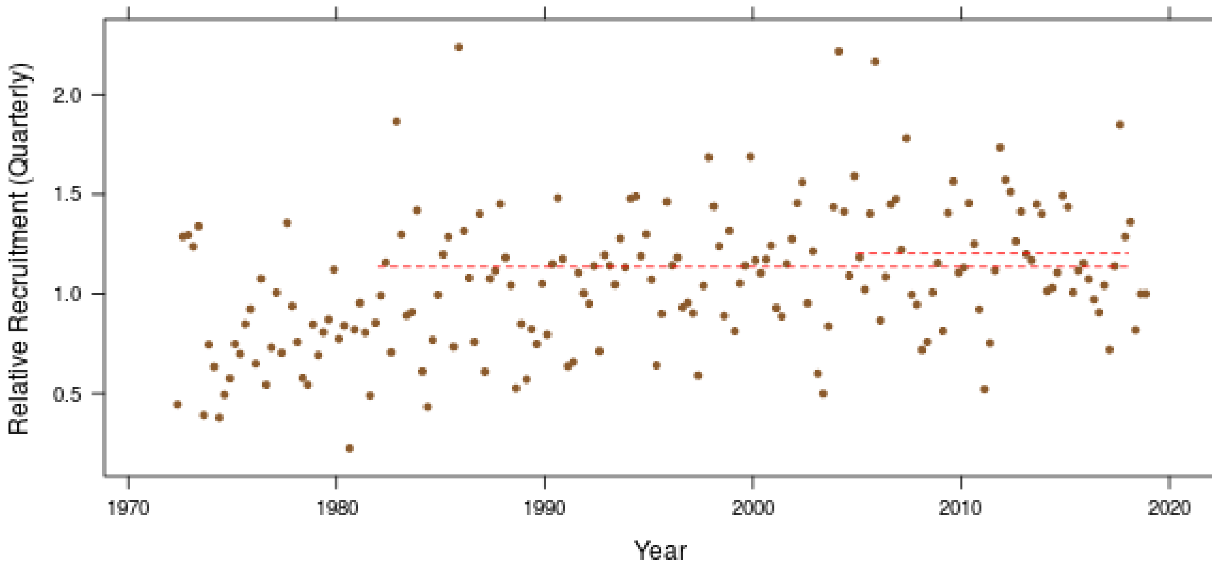


Figure 19: Quarterly relative recruitment estimates for the diagnostic case. Dashed lines show the long-term (1982:2018) and short-term (2005:2018) mean recruitment levels.

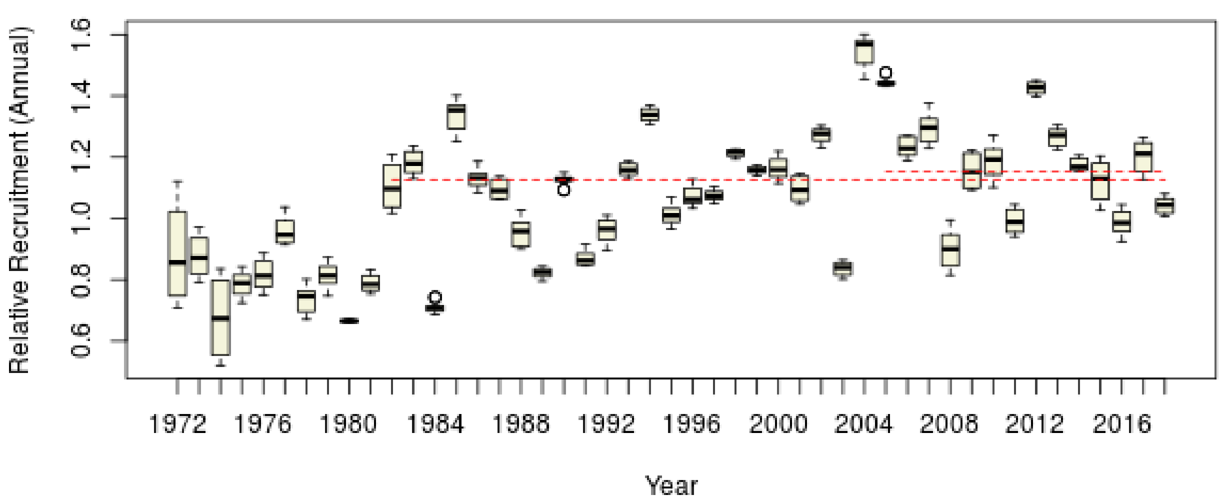


Figure 20: Annual relative recruitment estimates across the OM grid of 24 models. Dashed lines show the long-term (1982:2018) and short-term (2005:2018) mean recruitment levels, calculated across the grid of 24 models.

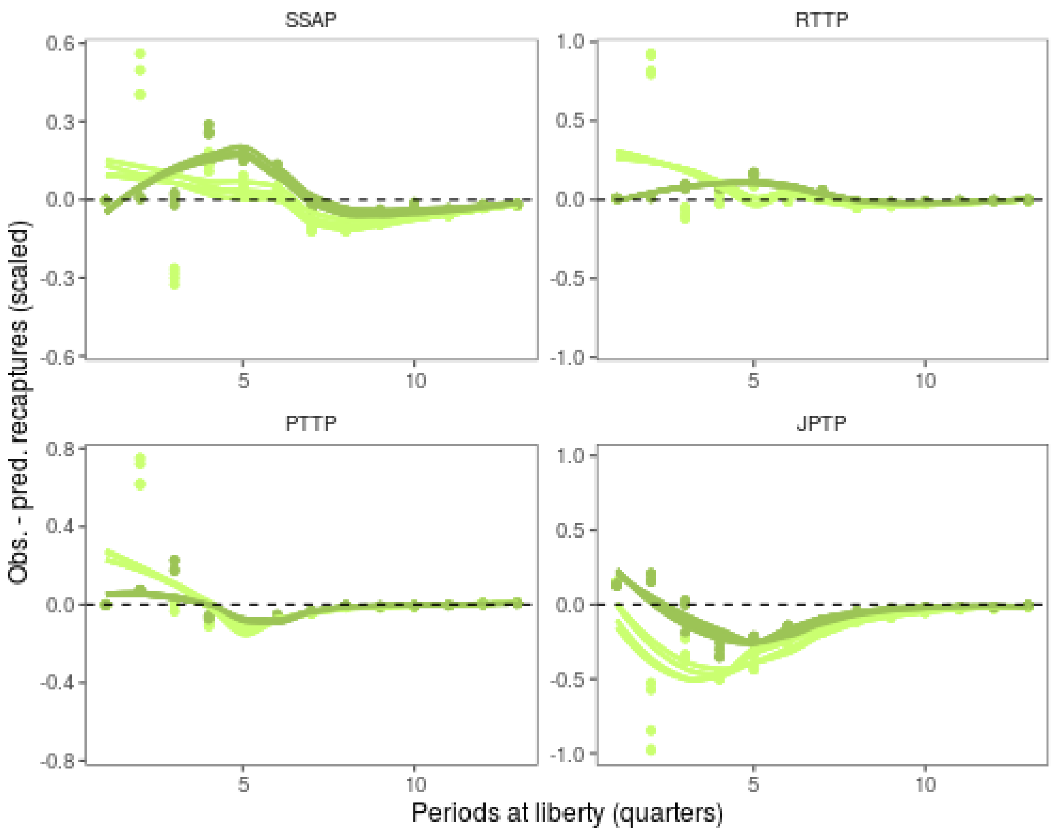


Figure 21: Observed - predicted tag recaptures for all time periods by tag release program across the OM grid (24 models). Model runs are shown for 1qtr mixing (light green) and 2 qtr mixing (dark green).

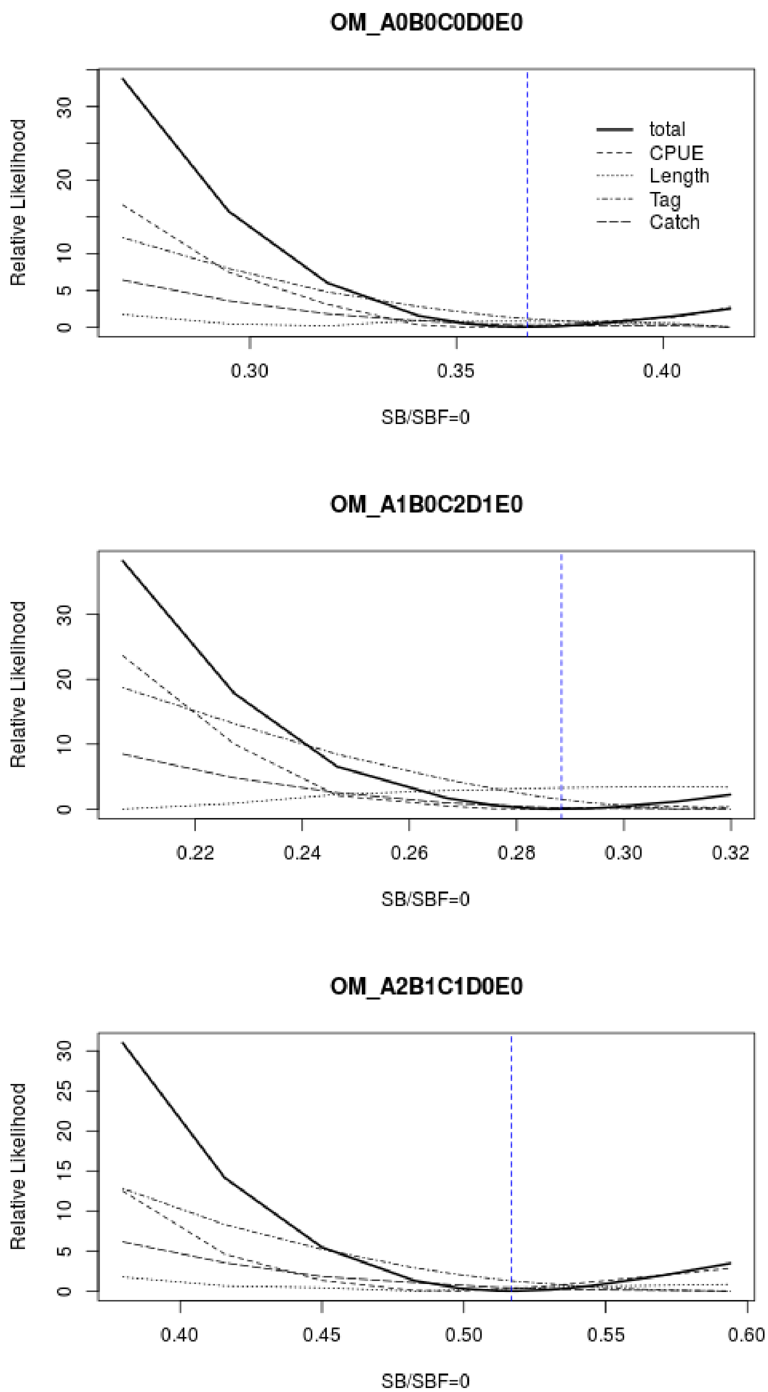


Figure 22: OM likelihood profiles for the diagnostic case model and the two models giving the lowest (A1B0C2D1E0) and highest (A2B1C1D0E0) estimates of SB/SBF0 (calculated as SB/SBF0latest). Vertical blue line shows the model estimate of SB/SBF0

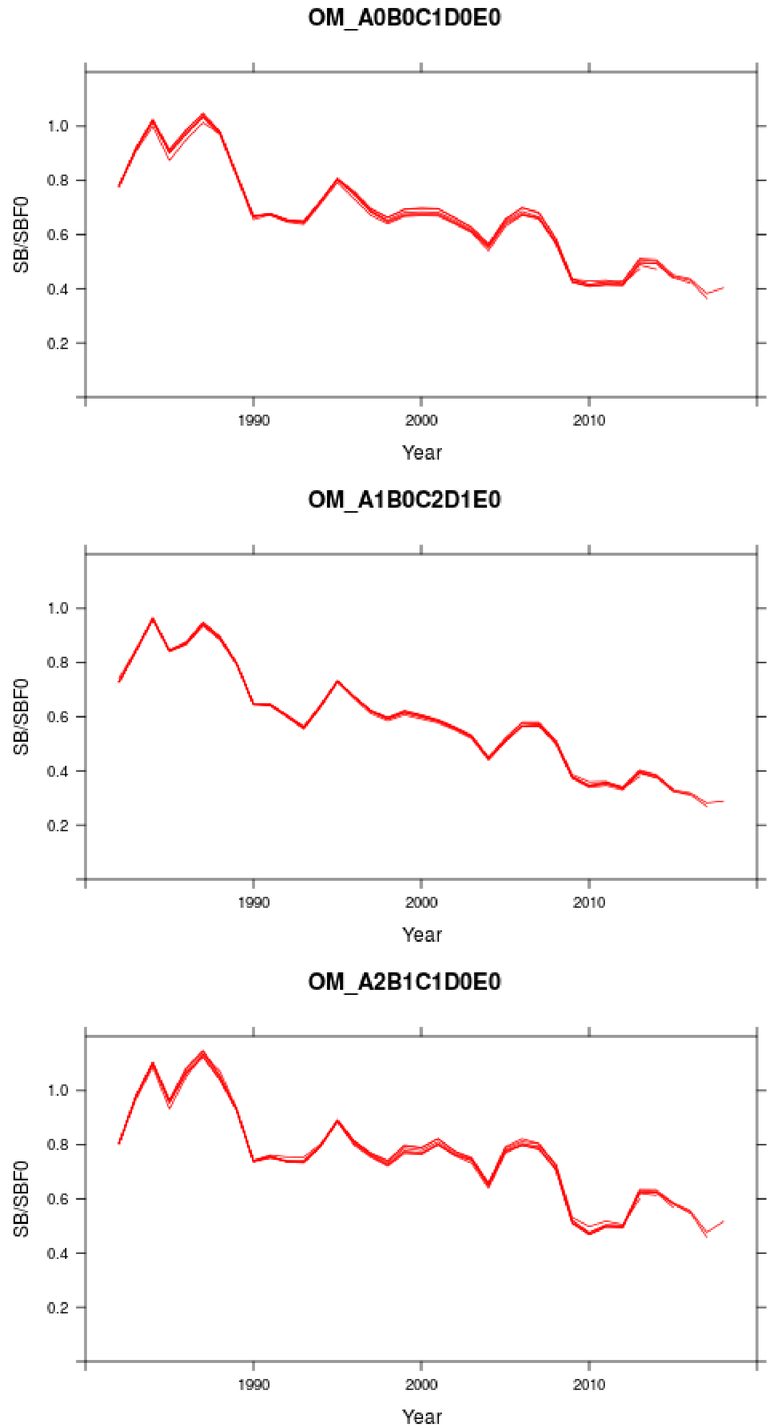


Figure 23: Retrospective analyses (2018:2013) for the diagnostic case model and the two models giving the lowest (A1B0C2D1E0) and highest (A2B1C1D0E0) estimates of SB/SBF0 (calculated as SB/SBF0latest)