SCIENTIFIC COMMITTEE
FIFTEENTH REGULAR SESSION
Pohnpei, Federated States of Micronesia
12–20 August 2019

The WCPO Skipjack MSE Modelling Framework

WCPFC-SC15-2019/MI-IP-02

R. Scott[1], F. Scott, N. Yao, G. M. Pilling, J. Hampton and N. Davies[2]

[1]Oceanic Fisheries Programme, The Pacific Community
[2]Te Takina Ltd.

Contents

Executive Summary

The modelling framework for testing candidate harvest control rules for Western and Central Pacific Ocean (WCPO) stocks and fisheries has undergone considerable development since 2016. Work has focussed primarily on developing the two main components of the management strategy evaluation (MSE) modelling framework, the operating model and the management procedure. Here we describe the current status of the modelling framework for skipjack and highlight recent developments and changes to the grid of models that comprise the MSE uncertainty grid, the specifications for the assessment model settings within the management procedure and a series of model validation diagnostics that have been run for both the operating model and the management procedure.

The skipjack MSE uncertainty grid is similar to that presented to SC14. The model scenarios comprising the grid are separated into a reference set (most plausible hypotheses forming the primary basis for selecting the 'best' management procedure) and a robustness set (less likely but still plausible). At the request of members, alternative scenarios for effort creep and hyperstability have been moved from the robustness set to the reference set, in addition the tag overdispersion parameter value of 8 has been replaced with a value of 6. Work continues to finalise the outstanding elements of the uncertainty grid.

More recent work has focussed on the management procedure component of the skipjack modelling framework, specifically with respect to the estimation model which provides the necessary information on the status of the resource (in this case depletion) that is used by the harvest control rule (HCR) to determine the necessary management action.

We outline the basis for the generation of pseudo data within the framework and detail the model specifications for the estimation model which provides estimates of stock status to the harvest control rule (HCR). We present model validation diagnostics including the maximum gradient of the estimated parameters, likelihood profiles of key parameters and model outputs, and retrospective analyses to establish the validity of the respective components of the framework.

This document describes the current status of the MSE modelling framework for skipjack in the WCPO. The framework continues to be developed and modified with work currently focussed on refining the models that comprise the robustness set (Scott et al., 2019b). However, the main components of the framework are now relatively well established and it is anticipated that the general framework described here will form the basis of future evaluations to test candidate harvest control rules and management procedures for skipjack.

1 Introduction

The modelling framework for testing candidate harvest control rules (HCRs) for Western and Central Pacific Ocean (WCPO) stocks and fisheries has undergone considerable development since 2016. A number of papers have been submitted to recent meetings of the Scientific Committee (WCPFC-SC) that document these developments and associated analyses with specific respect to the code development and testing of MULTIFAN-CL (Davies et al., 2017, 2018), the overall modelling framework (Scott et al., 2017b, 2018a), the conditioning of the operating models (Scott et al., 2018c), the generation of pseudo data (Scott et al., 2017c, 2018d), and the calculation of performance indicators (Scott et al., 2018b).

More recently, work has focussed on the management procedure component of the skipjack modelling framework, specifically with respect to the estimation model which provides the necessary information on the status of the resource (in this case depletion) that is used by the harvest control rule (HCR) to determine the necessary management action.

This document provides an updated overview of the management strategy evaluation (MSE) modelling framework. It contains a summary of the previous work that has already been presented to WCPFC-SC as well as information on the more recent work that will be presented in this and other papers to WCPFC-SC15.

1.1 The Harvest Strategy Approach

CMM 2014-06 (WCPFC, 2014) outlines the necessary elements of a harvest strategy as having:

1. defined operational objectives, including timeframes, for the fishery or stock;
2. target and limit reference points for each stock;
3. defined acceptable levels of risk of breaching limit reference points;
4. a monitoring strategy using best available information to assess performance against reference points;
5. decision rules (HCRs) that aim to achieve the target reference point on average and avoid limit reference points with high probability; and
6. an evaluation of the performance of the proposed HCRs against management objectives including risk assessment.

The modelling framework outlined in this paper describes the approach and procedures to evaluate the performance of proposed HCRs against management objectives as specified in item 6 above.

1.2 Testing and Developing Management Procedures

Because we do not have perfect knowledge of the dynamics of the resource, uncertainties will always be present in any assessment of the status or productivity of the fishery. By including all important sources of uncertainty in the evaluation framework we aim to find the management procedure that performs best and is robust to that uncertainty.

2 MSE Framework Overview

The MSE evaluation framework outlined in Figure 1 illustrates the process used to test the relative performance of candidate HCRs. As outlined in previous reports, the framework simulates the biological dynamics of the fish population; the stock assessment process and the consequent management action as specified by the candidate HCR. The evaluations are run for a representative time frame, in this case for a period of 30 years. Throughout the simulation period, performance indicators are calculated that enable the comparison of the performance of one HCR relative to another against agreed fishery objectives. From this information the best performing HCR can be identified.

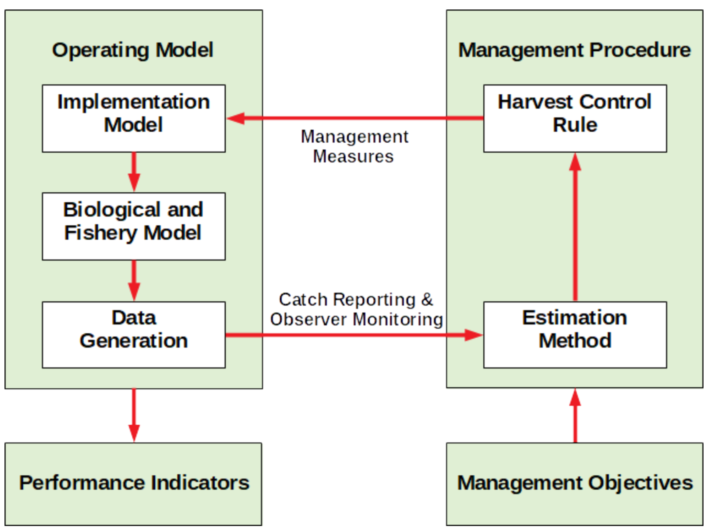

Figure 1: Conceptual diagram of the MSE framework (after Punt et al. (2014)).

The biological dynamics of the fish population as well as the fishing process are simulated by the operating model (OM). The estimation model (EM) and HCR are simulated by the management procedure (MP) component of the simulation framework. See Scott et al. (2018a) for a more detailed description of the MSE framework.

3 Operating Model

The operating model is a critical component of the MSE framework. It simulates the biological dynamics of the fish population as well as the dynamics and behaviour of the fishing fleets that exploit it. Typically a suite of operating models will be developed that encompasses the range of uncertainty associated with the assessment and management of the fishery. This suite of models is primarily developed through a process termed "conditioning" whereby a range of alternative model formulations are fit to the available data.

3.1 OM Conditioning

A first consideration of the conditioning of the skipjack OM (Scott et al., 2018c) was presented to SC14 along with a proposed grid of models to be used for the evaluations. It was noted at the time that certain elements of the MSE uncertainty grid for skipjack (the grid of models that covers the most important sources of uncertainty, Table 1) would require further investigation and that its contents and structure may evolve depending on the results of future analyses. For example, uncertainty in stock movement dynamics due to short-term environmental effects; the impacts of alternative future tag release programs and the potential value of, and trend in, effort creep were only vaguely specified in the initial grid.

Another area highlighted for further investigation was the tag over-dispersion setting which has the effect of increasing or decreasing the dependency of the assessment on tag data relative to other data sources. Three values of overdispersion were initially selected (2.5, 4, 8). Whilst the middle value of 4 appears to be a reasonable assumption and consistent with the level of overdispersion estimated for the tag recapture data, the bounds for sensitivity analyses (2.5 and 8) are less well determined and may be too wide. Model runs for which overdispersion was set to 8 consistently failed when running the evaluations. The overdispersion parameter value of 8 has been replaced with a value of 6 which results in greater model stability, although further investigation of the most appropriate values to assume for this parameter is recommended.

Work continues to finalise the outstanding elements of the uncertainty grid. More detailed information on the most recent analyses to investigate settings for the MSE uncertainty grid are available in Scott et al. (2019b).

3.1.1 MSE Uncertainty Grid

The skipjack MSE uncertainty grid (Table 1) is similar to that presented to SC15 except that, at the request of members, alternative scenarios for effort creep and hyperstability have been moved from the robustness set to the reference set and the option for tag overdispersion at 8 has been replaced with a value of 6.

Axis	Levels		Options		
	Reference	Robustness	**0**	**1**	**2**
Process Error					
Recruitment variability	**2**		**1982-2014**	**2005-2014**	
Recruitment autocorrellation	**2**		**0**	**estimated**	
Observation Error					
Catch and effort	**1**	1	**20%**	30%	
Size composition	1		all models (see Scott et al. (2018c))		
Tag recaptures	**1**	2	**status quo**	low	none
Model Error					
Steepness ‡	**3**		**0.8**	**0.65**	**0.95**
Mixing period (qtr) ‡	**2**		**1**	**2**	
Tag overdispersion ‡	**3**		**4**	**2.5**	**6**
Movement	**1**	1	**estimated**	El Nino/La Nina	
DD catchability (k) ‡	**2**		**0**	**-0.5**	-0.9
Implementation Error					
Effort creep	**2**	1	**0%**	**2% cont.**	3%

Table 1: 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.

3.1.2 Model Validation

Throughout the process of conditioning the OMs, consideration must be taken of how the suite of models correspond to the real world system, or at least our perception of it. At the simplest level, the models should appear to be realistic representations of the stock and fisheries. 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. Such model diagnostics include the maximum gradient of the estimated parameters, likelihood profiles of key parameters and model outputs, and retrospective analyses similar to those conducted for stock assessments.

The maximum gradient is a measure of how well the model has converged to a solution. The lower the maximum gradient the better the model fit is considered to be. Maximum gradients for the initial fits of the OM model grid (Figure 4) are consistently below 0.01, indicating that all of the models have achieved a satisfactory level of convergence.

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. (2018b)).

Conducting likelihood profiles is computationally intensive and it is not practical to run them for each of the models across the entire OM grid. Instead likelihood profiles have been conducted for a selection of models (Figure 6), chosen to represent a range of model configurations. The profiles 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 overall profile.

Retrospective analyses indicate how sensitive the model is to the addition of new data and are routinely conducted for each stock assessment. Analyses conducted for the 2016 assessment of WCPO skipjack (Figure 5) indicate good consistency in model estimates of depletion with reducing time series of data and little evidence of retrospective bias (i.e. where the model consistently under or over estimates stock status in the terminal years).

3.2 Pseudo Data Generation

A critical function of the operating model is the generation of so called "pseudo data". The simulated pseudo data represent the information that is routinely collected from the fishery to monitor and assess the status of the stock. Within the evaluation framework these data are used by the management procedure to determine the overall level of depletion that forms the primary input to the HCR. It is, therefore, particularly important that these simulated pseudo data are accurate representations of the type, and quality, of data that are collected in reality.

The assumptions underlying the generation of pseudo data are an important consideration since the results of any analysis will be predicated on the assumptions and conditions under which the analysis was conducted. The performance of a given management procedure will depend on the type, quantity and quality of data that are available to it. Consequently, the identification and selection of the best performing management procedure must be made with the implicit understanding that the conditions under which it was tested will also apply when the procedure is put into practice.

Approaches for generating pseudo data in MULTIFAN-CL have been outlined in previous reports (Scott et al., 2017a, 2018c) and the methods employed in the current framework broadly follow the same procedure. The key assumptions and conditions for generating pseudo data for the evaluations is summarized in Table 2 and described in further detail below.

The evaluations can be considered, broadly, as having a historical period, prior to the implementation of the HCR (1972-2015), and a future period, during which time a HCR is implemented (2016-2045). Functionality exists within MULTIFAN-CL to generate pseudo data for the historical time period (Davies et al., 2018), however, for the purposes of these current evaluations the historical data remain unchanged and pseudo data are generatated only for the future time period. As such variability in the historical period arises solely from the different model error settings outlined in Table 1.

Data Type	Distribution	Conditions
Recruitment	Log-normal	SRR deviates resampled from historical observations (2005:2014)
Catch	Log-normal	Time invariant, CV applied to all fisheries
Effort	Log-normal	Time invariant, CV applied to all fisheries
Length Compositions	Multinomial	Time invariant, fishery specific, multinomial effective sample sizes determined from MULTIFAN-CL model with self-scaling multinomial.
Tag Releases	User defined	Biennial, second qtr, all regions (release length comp determined from PL fisheries selection), time invariant, region specific, tag releases per release event.
Tag Recaptures	Multinomial	Determined from scenario specific fishing mortality rates and assuming a 60% reporting rate. (Note: overdispersion not currently enabled for tag recapture simulations).

Table 2: Pseudo data generation summary for skipjack.

3.2.1 Generating Catch, Effort and Size Composition Data

The generation of catch, effort and size composition data follows the approach previously outlined in Scott et al. (2018d). For each management period in the evaluations, fishery specific, quarterly catch and effort are determined from MULTIFAN-CL projections with log-normal observation error based on a user defined coefficient of variation (c.v.). Separate coefficients can be applied for either catch or effort but cannot currently be applied to individual fisheries. In other words the catch c.v. applies to all catch projected fisheries and the effort c.v. to all effort projected fisheries.

Size composition data are generated from a multinomial sample of predicted catches for each fishery, derived from the fishery specific selection pattern applied to the projected population in the region in which that fishery operates. The multinomial sample is based on a fishery specific effective sample size which is held constant through time and is based on the estimated effective sample sizes determined from a model fit using the self scaling multinomial option in MULTIFAN-CL (Table 2).

3.2.2 Generating Tag Release and Recapture Data

The generation of pseudo tag recapture data requires a number of assumptions to be made about the design, scale and frequency of future tagging programmes. The numbers and spatial distribution of fish that are tagged and released will clearly determine the quantity of recaptures that might be expected and which fisheries are likely to recapture them.

Two tagging programmes are currently in operation for WCPO skipjack tuna, the Pacific Tuna Tagging Program (PTTP) that tags and releases skipjack predominantly in the western equatorial region and the Japanese tagging programme (JPTP) that tags and releases skipjack throughout the region but to a greater extent in the more northerly subtropical area of assessment region 1 (for the 5 region assessment model). The PTTP is characterised by large numbers of tag releases from relatively short dedicated tagging cruises and alternates annually between targetting skipjack tuna in one year and targetting bigeye and yellowfin tuna the next. In contrast the JPTP releases fewer tagged fish per event but operates on a more continual basis.

For the purpose of the evaluations future tag releases are assumed to occur from multiple release events that occur on a biennial basis and take place in each region of the assessment during the second quarter of the year (Table 3). The number of releases from each event is set to comparable levels to those observed in recent tagging programmes (Figure 2) and the length distribution of the released fish is determined from the selection pattern of pole and line fisheries in each region of the assessment.

Pseudo tag recaptures are generated based on the multinomial probability of recapture given the estimated age-specific fishing mortality and the reporting rate probability for each tagging event. A value of 0.6 has been assumed for the reporting rate which applies to all fisheries for all simulated tag release events.

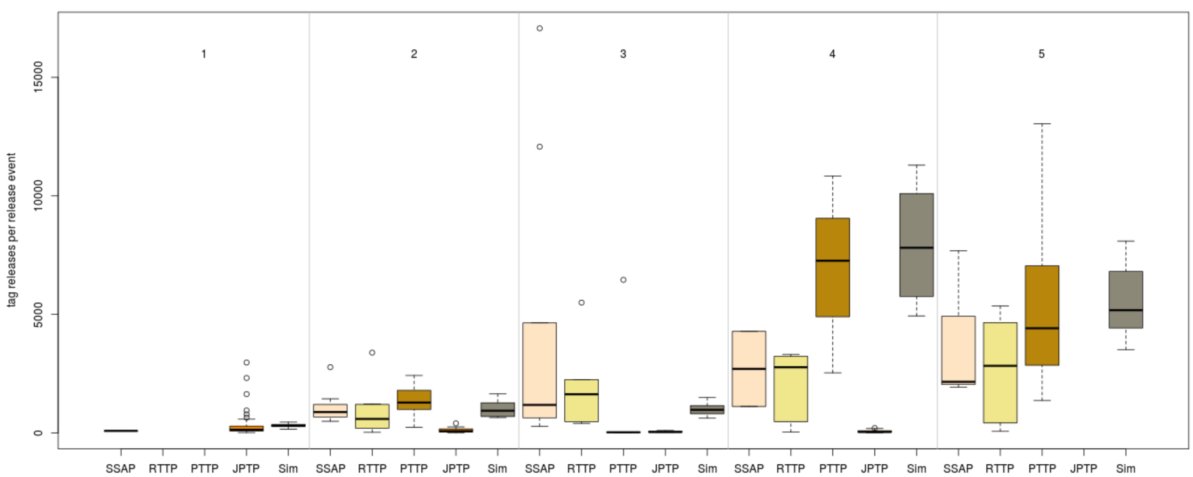

Figure 2: WCPO skipjack tag releases: Number of tag releases per release event by region for historical tag release programmes and simulated future data. Tag programs shown are the JPTP, the PTTP and its predecessors (SSAP and RTTP) and the values assumed for the simulations (Sim).

3.3 Implementation Model

Implementation error occurs when the management actions specified by the HCR are not followed precisely by the fishery. Some level of implementation error is almost inevitable, particularly in

Region	Year	Month	fishery	N releases
1	2017	5	P-JPN-1	300
2	2017	5	P-ALL-2	1335
5	2017	5	P-ALL-5	5641
3	2017	5	P-ALL-3	825
4	2017	5	P-ALL-4	6879
1	2019	5	P-JPN-1	300
2	2019	5	P-ALL-2	1335
5	2019	5	P-ALL-5	5641
3	2019	5	P-ALL-3	825
4	2019	5	P-ALL-4	6879
	...			
	...			

Table 3: Number of releases of tagged fish by region, year and month and the fishery from which the selection pattern was used to determine length compositions of tagged fish at time of release.

cases where a single species HCR is applied to fisheries that can switch to targetting other species. Of greater concern is a situation where the HCR consistently over-estimates the catch or effort necessary to manage the fishery towards a target. In severe cases this can ultimately lead to stock collapse. Such situations can occur in, for example, the case of widescale mis-reporting of catch or effort but may also occur due to persistent technological advances in fishing technology that progressively increases the efficiency of the fleet, a phenomenon known as effort creep.

Previous discussions within WCPFC on the issue of implementation error have identified three broad areas in which it may occur, including *intrinsic error* arising from factors around the operation of the fishery; *compliance error* arising from non-compliance with agreed measures (both of which are discussed above), and *management error* whereby the Commission agrees on management measures that do not fully meet the outputs of the HCR. We consider the potential for management error to be a genuine concern However, we have not included this in our evaluations. We have evaluated the HCRs under the assumption that they will be implemented, albeit with some allowance for the potential of effort creep in selected fisheries. We note the recommendations of the recent MSE technical review Scott et al. (2019a) that the implementation model should be kept as simple as possible and that although implementation error may be important, if the output of the MP is not respected and implemented correctly then the MP is effectively different to that which was originally tested.

3.3.1 Effort Creep

Several recent studies have tried to quantify the level of effort creep in WCPO purse seine fisheries and to identify potential indicators of effort creep (Tidd et al., 2015; Pilling et al., 2016; Muller et al., 2018). Initial estimates suggested that effort creep has been around 3% p/a in some fisheries although subsequent analyses suggest it may be at lower levels in recent years. Estimates of catchability from recent stock assessments for WCPO skipjack provide an indication of potential effort creep for the fishery aggregations defined in the stock assessment (Muller et al., 2018) and indicate that catchability has generally stabilised or declined in recent years.

Whilst there is considerable potential for effort creep, particularly in the associated set purse seine fishery component, there is no clear understanding of how it manifests or what scale it might be. For the evaluations we assume a simple linear trend in effort creep whereby effective effort increases progressively throughout the evaluation period and can update these values as necessary on the basis of ongoing work.

4 Management Procedure

A management procedure is a simulation-tested set of rules used to determine managment actions, in which the data, the methods for analysing the data (including any method of stock assessment) as well as the HCR are pre-agreed and pre-specified (Butterworth et al., 1997). Two types of MP can be distinguished:

- **Empricial MPs** where resource monitoring data (such as survey estimates of abundance or CPUE) are input directly to the HCR with minimal processing.
- **Model-based MPs** where an assessment model is used to generate an estimate of stock status.

Empirical methods have the advantage of being conceptually simple and typically require relatively little computer power for testing. In contrast model-based approaches allow for formal estimation of stock status using an assessment model that can vary in complexity depending on the type and quality of data available to the MP. Experience with the evaluation of management procedures has led to some guidelines for their properties and behaviours. Model-based MPs tend to result in less variation in catch or effort from one management period to the next but can also be slower to respond to major changes in resource abundance (De Oliveira et al., 2008).

The two main components of the MP are the estimation model (EM) and the harvest control rule (HCR). The WCPO skipjack MSE evaluation framework employs a model-based management procedure in which estimates of stock status ($SB/SB_{F=0}$), determined from a stock assessment model, are input to the HCR.

4.1 Estimation Model

The estimation model is a particularly important component of the framework as it determines the estimate of resource status that forms the primary input to the HCR. The resulting estimate of stock status must be sufficiently reliable to ensure the resulting management action is appropriate, but must also be determined in a sufficiently short time frame to make the simulations viable. SC14-MI-WP-02 noted that the estimation model was, at the time, one of least developed components of the MSE framework for skipjack and that further work would be required in this area.

Model-based management procedures require a stock assessment model to be run as part of the management procedure to determine the status of the resource at each management time step. The 2016 skipjack stock assessment using MULTIFAN-CL took around 16 hours to converge making it unsuitable for use within a simulation framwork. Previous studies have investigated options for alternative stock assessment methods (Scott et al., 2017b) including so-called 'reduced' MULTIFAN-CL assessments that consider a smaller spatial area and fewer fisheries, but none of the approaches produced sufficiently reliable estimates of stock status to use within the evaluations.

More recent work has focussed on approaches based more closely on the 2016 stock assessment and on procedures for refitting MULTIFAN-CL to pseudo data (Scott et al., 2017a, 2018d). The refitting of models has been explored both for the historical (estimation) period and for extended, projection time periods. The best results were achieved using initial parameter values that were closer to the OM fitted solution, requiring fewer function evaluations to reach convergence and running much faster.

Here, we outline the procedure for fitting the estimation model within the simulation framework. For the purpose of the evaluations, pseudo data are generated only for the projection period. Input data (catch, effort, size composition and tag recapture data) for the historical period, prior to the first projection year (2016), remain unchanged.

Axis	**Setting**
Steepness	**0.8**
Mixing period (qtr)	**1**
Tag overdispersion	**4**
Autocorrelated recruitment	**off**
Movement	**estimated**
DD catchability (k)	**0**

Table 4: Skipjack Estimation Model settings.

4.1.1 Refitting MULTIFAN-CL

The model is initialised with parameter estimates from the operating model fit for the estimation period. Model settings for steepness, tag mixing, etc. (Table 4) are adjusted accordingly and the model run for a sufficient period to allow it to refit to the new data and the new model settings. No parameters are fixed during the refitting process, all parameters continue to be freely estimated. The model is run for 3 phases with the penalty on catch deviations progressively increased in each phase. Phase 1; penalty of 100 and 100 function evaluations. Phase 2; penalty of 10,000 and 100 function evaluations. Phase 3; penalty of 100,000 and 1,000 function evaluations. The resulting estimate of stock depletion ($SB/SB_{F=0}$) at the end of each successive management period is used as the input to the HCR.

Importantly, the estimation model employs a different model configuration to that used for the operating model. A fixed model structure is used for all evaluations (Table 4) based on the reference case of the most recent (2016) assessment. As such, the various assumptions of alternative stock dynamics that are captured in the grid of operating models are not mirrored exactly within the management procedure. This approach ensures that any potential model mis-specification that may occur in the stock assessment process is captured in the evaluations.

4.1.2 Model Validation

As for the operating model, the validity of the estimation model is an important consideration. The estimation model must provide a sufficiently reliable estimate of stock status to ensure that the output of the HCR and the resulting management action are appropriate. However, as for the operating model, there is no single test to determine the validity of the estimation model and instead we rely on a collection of model diagnostics.

The consistency of the estimation model can be examined by comparing model estimates from successive management periods in a similar way that a retrospective analysis compares model estimates for different time periods of data. Successive estimates of depletion for each management period for a selection of model scenarios (Figure 7) show good consistency with only a slight tendency to under-estimate stock status in some years.

Likelihood profiles, similar to those run for the operating model, were also run for a selection of estimation models (Figure 8). The profiles show that depletion is typically less well estimated by the estimation model, in comparison wtih the operating model, with $SB/SB_{F=0}$ estimates often lying a little above or below the maximum likelihood estimate but typically within 5% to 10% of that value.

Running likelihood profiles for MULTIFAN-CL can be a time consuming and computationally intensive process. Running profiles for all model combinations would be impractical, consequently profiles have been run for only a selection of models. Further information on the procedures for running likelihood profiles along with results of the analyses described above is provided in Appendix B.

4.2 Harvest Control Rules

In theory any style or formulation of HCR can be implemented in the framework. The majority of evaluations considered so far, however, have utilised a very simple 'broken stick' style of HCR (Figure 3) that takes an estimate of stock status ($SB/SB_{F=0}$) as input and outputs a scaler that can be applied to either catch or effort to determine the level of fishing for the next management period.

For the current skipjack evaluations, the scaler returned by the HCR is used to scale future fishing effort for purse seine fisheries (see Table 5) and future catch for all other fisheries specified in the operating model.

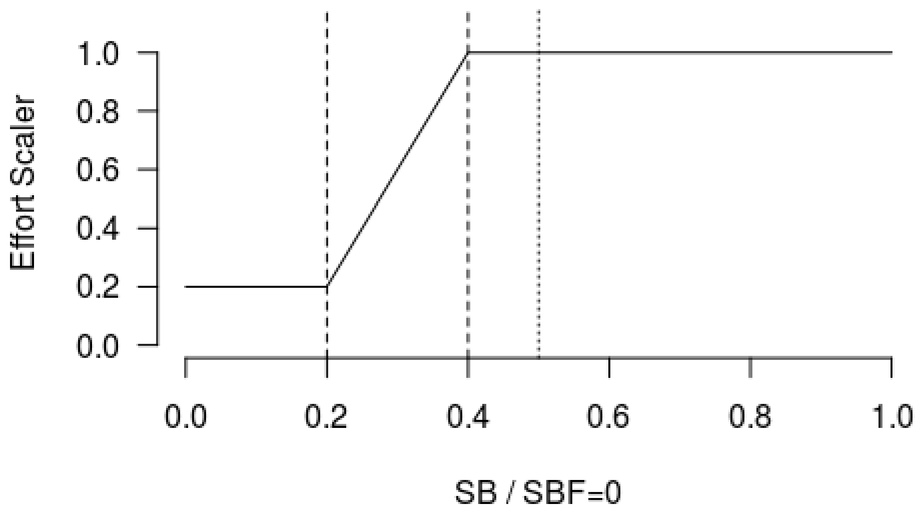

Figure 3: Example HCR

5 Model Outputs

The primary outputs from the modelling framework are the MULTIFAN-CL report files and associated information from which the performance indicators are calculated. MULTIFAN-CL outputs a number of files some of which can be very large. Only a selection of these files are necessary for calculating the performance indicators. The minimum set of essential files include the plot.rep and catch.rep files from the last iteration of the OM, the parameter estimates for the growth model as specified in the .par file and the proj.frq that specifies the catch and effort values as input to the OM projection model. Where effort creep is implemented in the evaluations a second proj.frq must also be returned to keep track of difference between 'effective' effort and 'perceived' effort. In addition a small results object is also returned to keep a record of the stock status and resulting HCR output for each management cycle for.

5.1 Performance Indicators

Performance indicators are calculated from the outputs of the OM using the files listed above. Currently 6 performance indicators are calculated for the skipjack evaluations. Further information on the calculation of performance indicators for WCPO skipjack are provided in Scott et al. (2018b).

1. **Indicator 1**. Maintain SKJ, YFT, BET biomass at or above levels that provide fishery sustainability throughout their range.
2. **Indicator 3**. Maximise economic yield from the fishery (average expected catch)
3. **Indicator 4**. Maintain acceptable CPUE.
4. **Indicator 6**. Catch stability.
5. **Indicator 7**. Stability and continuity of market supply (effort variation relative to a reference period).
6. **Indicator 8**. Stability and continuity of market supply (probability of and deviation from $SB/SB_{F=0} > 0.5$).

and a further 4 indicators remain under consideration pending further discussion on how they might best be calculated or approximated.

1. **Indicator 5**. Taking Article 30 of the WCPFC convention into account: Maximise SIDS revenues from resource rents / Take into account the special requirements of developing states and territories.
2. **Indicator 9**. Food security in developing states (import replacement).
3. **Indicator 10**. Avoid adverse impacts on small scale fisheries.
4. **Indicator 11**. Minimise bycatch.

We note the comments of CCMs concerning the definition of these performance indicators (WCPFC14-2017-11, Attachment C) and the ongoing discussions around their calculation, in particular for those outstanding indicators for which values cannot yet be calculated. We consider the calculation of these outstanding indicators to be a priority concern that will need to be addressed as soon as possible. We stress that the lack of a calculated value for a performance indicator, at this stage, does not imply it has reduced priority in the framework.

We continue to seek feedback from members on the definition and calculation of performance indicators.

6 Summary

This document describes the current status of the MSE modelling framework for skipjack in the WCPO. The framework continues to be developed and modified with work currently focussed on refining the models that comprise the robustness set (Scott et al., 2019b). However, the main components of the framework are now relatively well established and it is anticipated that the general framework described here will form the basis of future evaluations to test candidate harvest control rules and management procedures for skipjack.

References

Butterworth, D., Cochrane, K. L., and De Oliveira, J. (1997). *Global Trends: Fisheries Management*, chapter Management procedures: a better way to manage fisheries? The South African experience. American Fisheries Society Symposium.

Davies, N., Fournier, D. A., Takeuchi, Y., Bouye, F., and Hampton, J. (2017). Developments in the MULTIFAN-CL software 2016-2017. WCPFC-SC13-2017/SA-IP-05, Rarotonga, Cook Islands, 9–17 August 2017.

Davies, N., Fournier, D. A., Takeuchi, Y., Bouye, F., and Hampton, J. (2018). Developments in the MULTIFAN-CL software 2017-2018. WCPFC-SC14-2018/SA-IP-02, Busan, South Korea, 9–17 August 2018.

De Oliveira, J., Kell, L., Punt, A. E., Roel, B., and Butterworth, D. (2008). *Advances in Fisheries Science: 50 years on from Beverton and Holt*, chapter Managing without best predictions: the management strategy evaluation framework, pages 104–135. Blackwell Publishing Ltd.

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. (2014). *MULTIFAN-CL User's Guide*. 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.

Muller, B., Pilling, G., and PNA Office (2018). Updating indicators of effort creep in the WCPO purse seine fishery. WCPFC-SC14-2018/MI-IP-05, Busan, Korea, 3–11 August 2016.

Pilling, G., PNA Office, Norris, W., and Hampton, J. (2016). Examining indicators of effort creep in the WCPO purse seine fishery. WCPFC-SC12-2016/MI-WP-08, 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).

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Scott, F., Scott, R., Yao, N., Hillary, R., Kitakado, T., Davies, N., Pilling, G., and Hampton, J. (2019a). Recommendedport of the second expert consultation workshop on management strategy evaluation. WCPFC-SC15-2019/MI-IP-03, Pohnpei, Federated States of Micronesia, 12–20 August 2019.

Scott, F., Scott, R. D., Davies, N., Pilling, G., and Hampton, J. (2018a). Technical developments in the MSE modelling framework. WCPFC-SC14-2018/MI-IP-02, Busan, South Korea, 5–13 August 2018.

Scott, F., Scott, R. D., Pilling, G., and Hampton, J. (2018b). 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. (2019). *FLR4MFCL: R functions for MULTIFAN-CL*. R package version 1.1.4.

Scott, R., Davies, N., Pilling, G., and Hampton, J. (2017a). Generating pseudo data in MULTIFAN-CL. WCPFC-SC13-2017/MI-IP-02, Rarotonga, Cook Islands, 9-17 August 2017.

Scott, R., Scott, F., Yao, N., Pilling, G., and Hampton, J. (2019b). Modelling key uncertainties in skipjack management strategy evaluation. WCPFC-SC15-2019/MI-WP-06, Pohnpei, Federated States of Micronesia, 12–20 August 2019.

Scott, R. D., Davies, N., Pilling, G., and Hampton, J. (2017b). Developments in the MSE modelling framework. WCPFC-SC13-2017/MI-WP-04, Rarotonga, Cook Islands, 9–17 August 2017.

Scott, R. D., Davies, N., Pilling, G., and Hampton, J. (2017c). Generating pseudo-data in MULTIFAN-CL. WCPFC-SC13-2017/MI-IP-02, Rarotonga, Cook Islands, 9–17 August 2017.

Scott, R. D., Scott, F., Davies, N., Pilling, G., and Hampton, J. (2018c). 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., Davies, N., Pilling, G., and Hampton, S. (2018d). Generating pseudo data in MULTIFAN-CL. WCPFC-SC14-2018/MI-IP-03, Busan, South Korea, 5–13 August 2018.

Tidd, A., Pilling, G., and Harley, S. J. (2015). Examining productivity changes within the tropical wcpo purse seine fishery. WCPFC-SC11-2015/MI-WP-06, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

WCPFC (2014). Conservation and Management Measure on establishing a harvest strategy for key fisheries and stocks in the Western and Central Pacific Ocean. Conservation and Management Measure 2014–06, Eleventh Regular Session of the WCPF Commission, Apia, Samoa, 1–5 December 2014.

A Model Tools and Software

The evaluation framework is constructed predominantly from software packages and programs written either in C++ or R (R Core Team, 2018). The framework relies heavily on the stock assessment software MULTIFAN-CL which, in the case of WCPO skipjack, is used as the basis for both the OM and the MP. File manipulation, data transfer and other associated tasks are conducted using a dedicated R package (FLR4MFCL) that has been developed specifically to support analyses using MULTIFAN-CL. The evaluations are run on a remote server using the University of Wisconsin HTCondor system (CHTC) which provides access to a bank of PCs with operating systems based on either Scientific Linux 6 or Centos 7. Currently the evaluations are run using R version 3.5.2 and FLR4MFCL version 1.1.3.

A.0.1 MULTIFAN-CL

MULTIFAN-CL (Kleiber et al., 2014) is the primary stock assessment method used for WCPFC stocks. It implements a statistical, size based, age-structured and spatial-structured population model that has been developed specifically for the assessment of fish stocks for which extensive age sampling of catches is not feasible but for which length frequency sampling data are available.

Initial developments of MULTIFAN-CL added many important features, most notably spatial structure, fish movement and capacity for tagging data which are particularly important for the assessment of WCPO skipjack. More recent developments of the assessment software have provided the functionality to generate simulated input data sets (or pseudo data) both for historical (estimation) and future (projection) time periods (Davies et al., 2018; Scott et al., 2018d). The combination of fast, efficient minimisation of highly complex models and the ability to generate simulated data sets for a range of statistical assumptions and distributions makes MULTIFAN-CL a powerful tool for testing management strategies and specifically for the development of operating models within a management strategy evaluation framework.

A.0.2 FLR4MFCL

A drawback of MULTIFAN-CL is that it relies on a collection of flat text input files to provide the necessary input data for each phase of the analysis. These input files have become increasingly large and unwieldy as more information on the stock is collected and as the progressive development of the software provides greater flexibility and functionality for the assessment process.

Running the simulation framework requires extensive editing of these input files. FLR4MFCL (Scott, 2019) has been developed as a subsidiary package of FLR (Kell et al., 2007) which provides a flexible platform for quantitative fisheries science based on the R statistical language. Source code for the FLR4MFCL package is hosted on a public github site and is available from https://github.com/PacificCommunity/ofp-sam-flr4mfcl.

A.0.3 CHTC

HT Condor is an open-source high-throughput computing software framework for coarse-grained distributed parallelisation of computationally intensive tasks. It allows large simulation jobs to be broken up into many smaller jobs that can be run individually on remote PCs or dedicated servers. OFP-SAM operates a small, in-house HT-Condor system at SPC HQ, Noumea, but has been granted access to the HT-Condor network at the Center for High Throughput Computing (CHTC) at the University of Wisconsin which substantially increases the available processing power for running large simulations.

Under this approach, the simulations are broken down to smaller, more manageable units that can be run in parallel across a large number of PCs and the outputs subsequently combined to form the overall simulation results. The smallest unit that the simulations can presently be broken down to is a single run (for 30 years) for one OM, one HCR and one set of random deviates (recruitment variability, observation error, etc.). Under the current assumption of a 3 year management period, each unit will run 9 cycles of the management procedure with an approximate run-time of around 10 to 11 hours. As such the evaluation of a single HCR across all OMs and for a representative number of evaluations can be achieved in around 24 to 36 hours, depending on the number of iterations required. The necessary tasks of retrieving the results from the remote PCs and post processing the output does, however, incur additional time that can vary with network speed.

B Model Validation Diagnostics

Sample diagnostics are presented for both the operating model and the estimation model. Where possible diagnostics have been calculated across the full grid of model combinations, however, due to the high computational demand of running some of the diagnostics (likelihood profiles in particular) they are shown here for only a small subset of the grid.

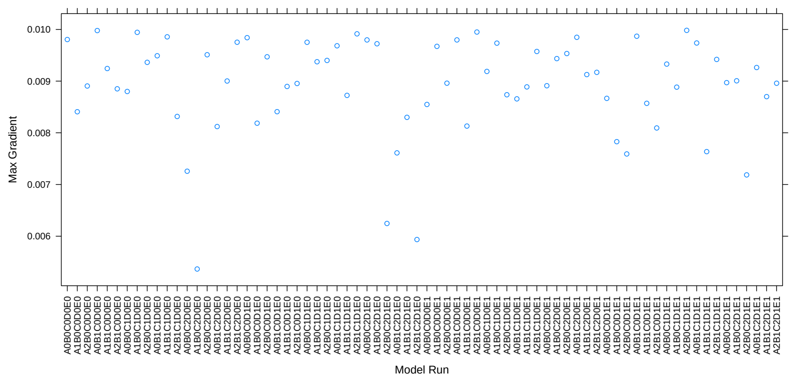

Figure 4: Maximum gradient across the grid of models. (All models were run for an additional 18000 function evaluations after the final fitting phase)

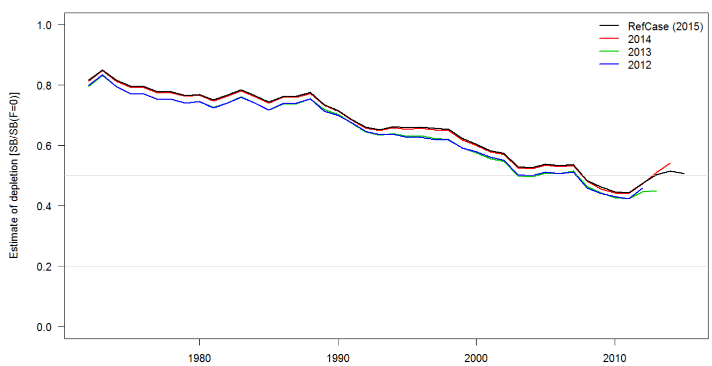

Figure 5: Retrospective analysis conducted for the 2016 reference case assessment of WCPO skipjack (McKechnie et al., 2016) showing estimates of depletion from 4 assessment models with reducing time series of data.

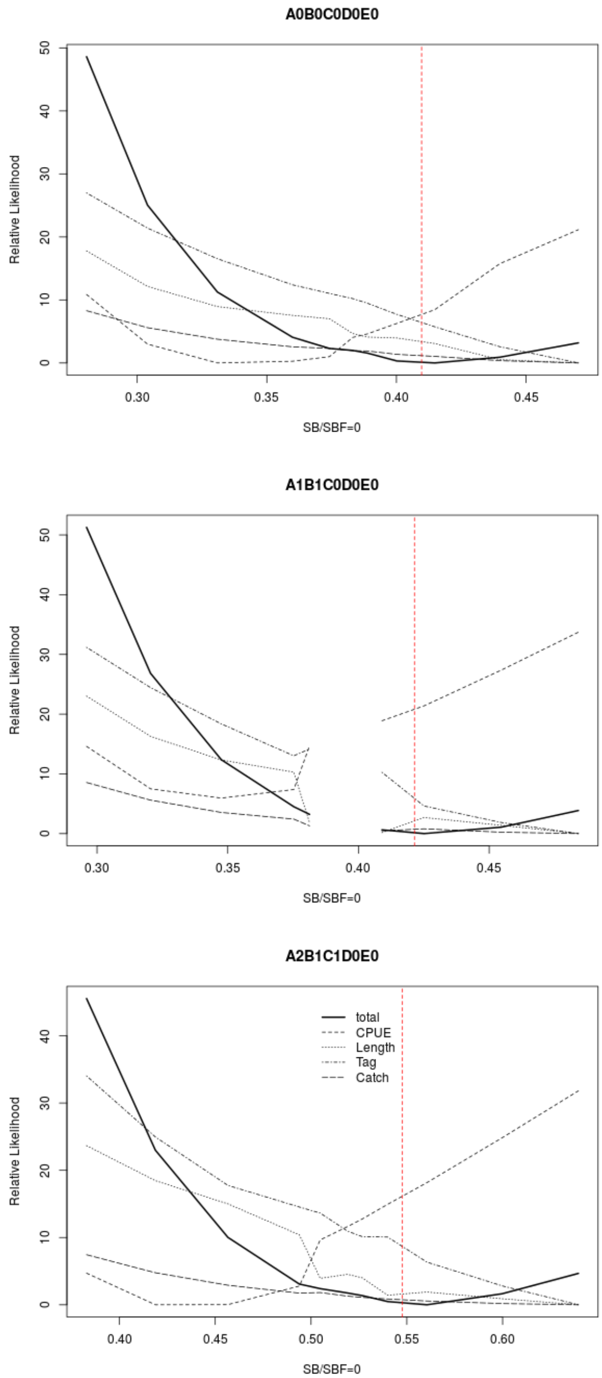

Figure 6: Likelihood profiles for depletion ($SB/SB_{F=0}$ in 2015) for select cases of operating models. Vertical red lines show the point estimate of depletion for each model. Profile runs for some depletion values were poorly converged giving spurious model estimates that have been omitted from the plots.

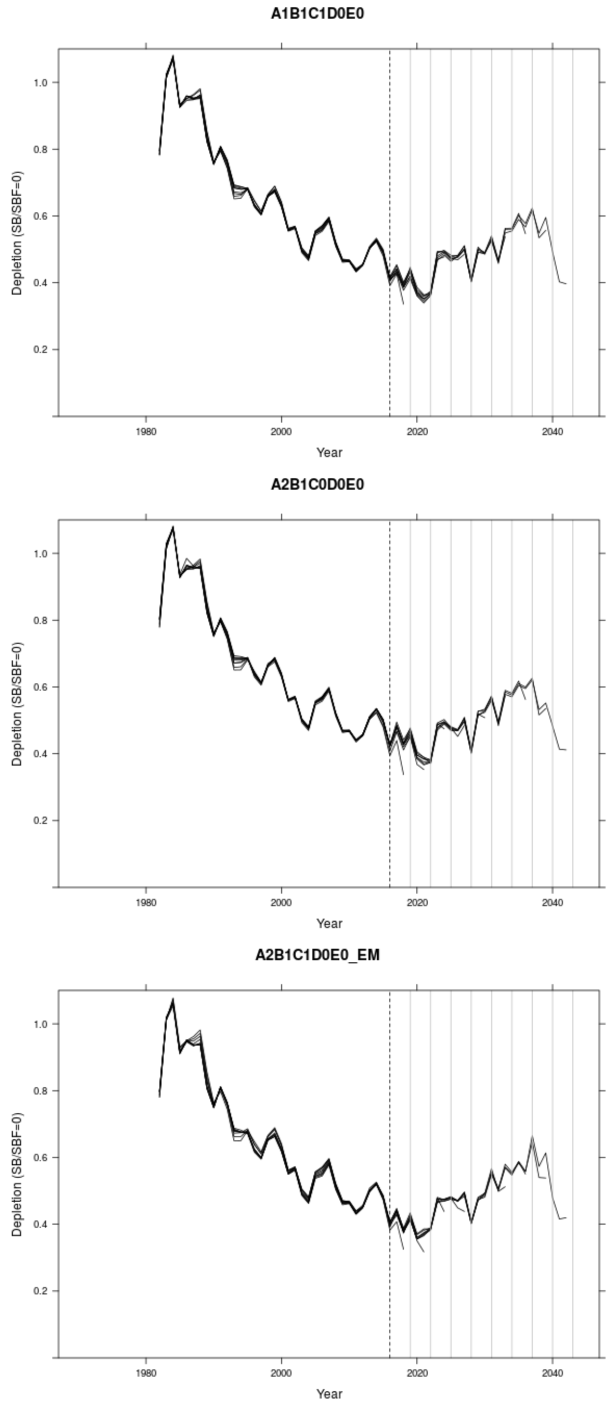

Figure 7: Retrospective analyses for depletion ($SB/SB_{F=0}$ in 2015) for select cases of estimation models. Vertical lines show the first projection year and subsequent management periods. Each scenario has been run with the same random number seeds and recruitment deviates.

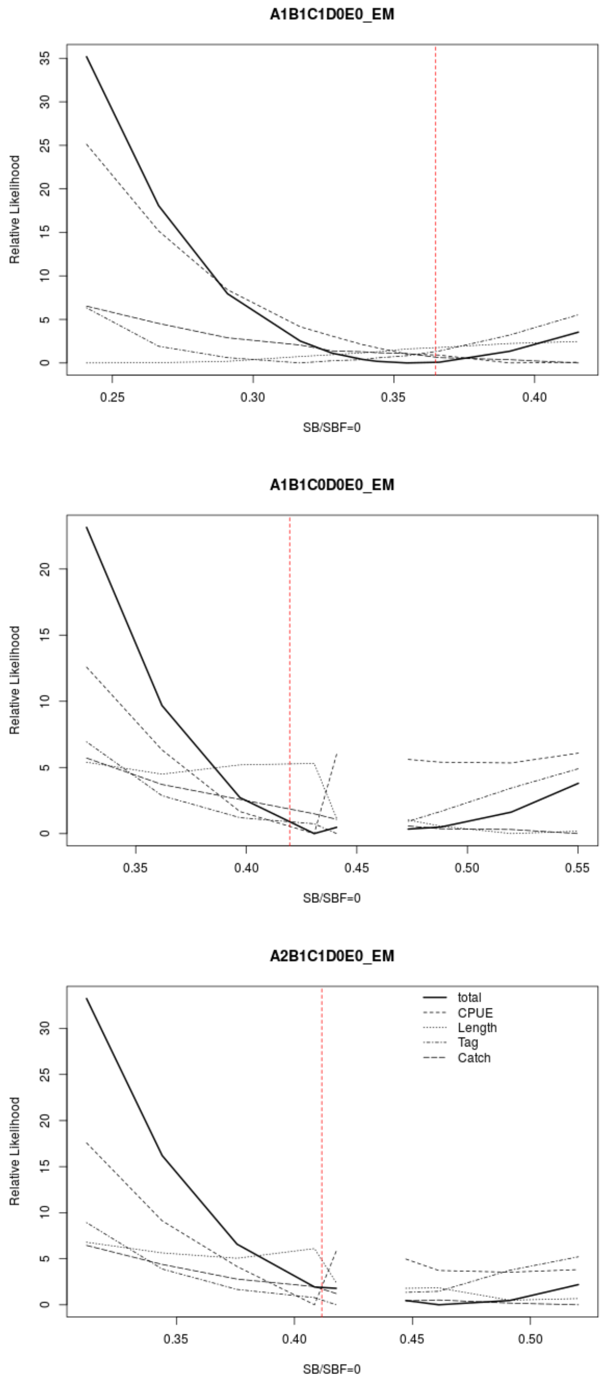

Figure 8: Likelihood profiles for depletion ($SB/SB_{F=0}$ in 2015) for select cases of estimation models. Vertical red lines show the point estimate of depletion for each model. Profile runs for some depletion values were poorly converged giving spurious model estimates that have been omitted from the plots.

C HCR Design

C.1 Effort Correction for Archipelagic Waters

The archipelagic waters (AW) correction is based on the total fishing effort in 2012 and on the effort of fisheries 9 and 10 as specified in the 2014 skj.frq file (fishery 9 effort prior to standardisation). The effort correction is a simple scaler based on the proportion of 2012 fishing effort inside and outside AWs (Eqn.1).

$$S_5 = \frac{S_{HCR} * E_{EEZ-AW} + E_{AW}}{E_{EEZ-AW} + E_{AW}} \tag{1}$$

where :

S_5	adjusted effort scaler to be applied in Region 5
S_{HCR}	effort scaler determined from the harvest control rule
E_{EEZ-AW}	fishing effort outside of AWs in 2012
E_{AW}	fishing effort inside AWs in 2012

C.2 Catch and effort projected fisheries

Fishery	**Region**	**Type**	**Fishery**	**Region**	**Type**
PL	1	catch	PL	3	catch
PS	1	effort	PS-ASS	3	effort
LL	1	catch	PS-UNA	3	effort
PL	2	catch	LL	3	catch
PS-ASS	2	effort	Dom-PH	4	catch
PS-UNA	2	effort	Dom-ID	4	catch
LL	2	catch	PS	4	effort
PL	5	catch	PL	4	catch
PS-ASS	5	effort	PS-ASS	4	effort
PS-UNA	5	effort	PS-UNA	4	effort
LL	5	catch	Dom-VN	4	catch
			LL	4	catch

Table 5: Fisheries for which the HCR controls either catch or effort.