SCIENTIFIC COMMITTEE
THIRTEENTH REGULAR SESSION
Rarotonga, Cook Islands
9–17 August 2017

Developments in the MSE modelling frameowrk.

WCPFC-SC13-2017/MI-WP-04

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

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

Contents

1 Introduction **4**

1.1 MSE Framework 4

1.1.1 The Operating Model 4

1.1.2 The Management Procedure 5

1.2 Tools for MSE 6

2 Operating Model: Data Generation **6**

2.1 Catch and effort and CPUE data 7

2.2 Pseudo length frequency data 7

2.3 Refitting MULTIFAN-CL 9

3 Management Procedure: Estimation Method **10**

3.1 a4a model 10

3.1.1 Creation of an age-specific tuning index 11

3.1.2 Creation of a tag based tuning index 12

3.1.3 Results 14

3.2 Reduced MULTIFAN-CL model 15

3.2.1 Modification of MULTIFAN-CL input files 16

3.2.2 Modification of the doitall 16

3.2.3 Results 17

4 Tools for MSE **18**

4.1 MULTIFAN-CL 18

4.2 FLR 20

4.3 High-throughput Computing: HT-Condor 20

5 MSE Modelling Framework **21**

A List of Abbreviations **25**

Executive Summary

This paper presents an account of recent developments in the construction of the MSE framework for WCPFC stocks and fisheries and focusses primarily on some of the technical issues related to the development of the modelling framework. In many cases, the results presented in this paper represent the current state of ongoing analyses and are not considered to be final. We specifically consider two issues, the generation of pseudo data from the operating model, and the development of an estimation model for skipjack that can be used within the management procedure to estimate stock status. In addition, we briefly identify a number of tools and software approaches that may be used for building the evaluation framework.

Although many of the issues addressed here are generic to the development of the MSE framework for WCPFC stocks and fisheries, our analyses and examples have been based on the WCPO skipjack stock. The results presented here are provided for information and to update the WCPFC-SC on the approaches being taken to construct the simulation framework and the progress being made in developing the necessary tools and software.

Recent developments to MULTIFAN-CL to facilitate the generation of simulated data are examined and tested. Although these results are preliminary in nature, they indicate that further work is required to generate simulated data that better reflect the characteristics of observed data both for catch and effort observations and for simulated length frequency data. Whilst the current implementation introduces observation error into the pseudo data, there are important sources of process error that remain unaccounted for.

Options investigated for developing an estimation model that can generate suitable estimates of stock status within a feasible time period included i) a4a (an age based statistical assessment, implemented in R/FLR/ADMB), and ii) options for a simplified MULTIFAN-CL assessment model. The a4a approach failed to produce reasonable estimates of stock status, possibly due to the approach taken to convert length-based input data to catch at age information, but also due to the difficulty associated with creating an index of abundance. The simplified MULTIFAN-CL assessment models yielded estimates of stock status that were more in line with those of the most recent assessment and currently represent the most feasible option for estimating stock status within the management procedure.

We invite WCPFC-SC to consider the approach being taken to develop an MSE framework for WCPFC stocks and fisheries, noting that the results presented here are preliminary and that development of the framework is ongoing. Specifically we invite SC13 to

- consider the approach outlined here for the development of the MSE simulation framework, noting that our initial focus has been on WCPO skipjack and that other stocks and fisheries will need to be considered as the work proceeds.
- consider the proposed software solutions for developing and running the simulation framework.

1 Introduction

This paper presents an account of recent developments in the construction of the management strategy evaluation (MSE) framework to be used to test the performance of candidate HCRs for WCPFC stocks and fisheries. It focusses primarily on some of the technical issues related to the development of the modelling framework, but also touches on some procedural and data management issues that may need to be taken into consideration as the harvest strategy work progresses.

In many cases, the results presented in this paper are not considered to be final. Instead they represent the current state of ongoing analyses. With respect to the modelling framework, recent work has focussed on the development of methods for generating pseudo data from the operating model (Section 2) and the creation of supporting software (Section 4). In particular, we highlight an unresolved problem associated with the development of an efficient and reliable estimation model for skipjack that can be used within the management procedure to estimate stock status (Section 3). Although many of the issues addressed here are generic to the development of the MSE framework for WCPFC stocks and fisheries, our analyses and examples have been based on the WCPO skipjack stock.

1.1 MSE Framework

Previous analyses to evaluate the performance of candidate harvest control rules for WCPFC fisheries have been based on a relatively simple modelling framework (Kirchner et al., 2014; Scott et al., 2016a) that did not take explicit consideration of a number of important sources of uncertainty. These approaches included variability in future recruitment, to take partial account of process error, whilst model error was approximated through the application of noise to the terminal biomass estimates. No other potential sources of uncertainty were included. In order to more rigorously test candidate HCRs it is necessary that a more advanced modelling framework be developed that will allow all of the important sources of uncertainty to be more explicitly taken into consideration.

1.1.1 The Operating Model

The evaluation framework (Figure 1) is constructed from two main components, an operating model (OM) and a management procedure (MP). The OM is a mathematical representation of the "true" system. It simulates the real world by attempting to capture all existing knowledge and data processes for the exploited populations and associated fisheries. It should also allow the evaluation of the consequences of contrasting hypotheses about the dynamics of those populations and fisheries. Very often the OM will include a greater level of complexity than that used for the stock assessment so that all important sources of uncertainty might be appropriately included in the evaluation process. The expert consultation workshop on MSE (Scott et al., 2016b) noted the current features

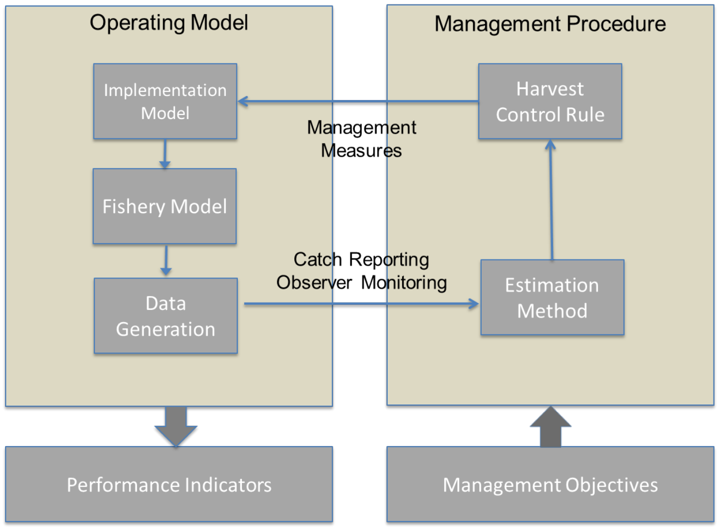

Figure 1: Conceptual diagram of the MSE framework (after Punt et al. (2014)).

available in MULTIFAN-CL, and the developments planned for the future (e.g. the generation of pseudo-data), and considered it to be an appropriate tool for developing the OM. It further considered that the approach used by WCPFC to capture uncertainty in the assessment results through the uncertainty grid to provide a starting point for capturing key uncertainties in the OM.

1.1.2 The Management Procedure

The management procedure (MP) represents the system for managing a fishery, from data collection through stock assessment to formulation of the management action. The MP may be based on current or alternative assessment methods and management approaches. In its simplest form it includes an estimation method to derive estimates of stock status (or some other relevant quantity) and a pre-defined set of management actions according to some specified rules (e.g. a HCR) that take into account the outcome of the estimation method. The estimation method and the HCR are intrinsically linked since the output of one feeds directly into the other. Whilst managers are very often primarily interested with the design of the HCR (which has a direct impact on the management measures) it is important to note that the two components should be considered together and that the MSE evaluates the performance of the MP as a whole and not just that of the HCR being considered.

The choice of model to use for the estimation method is important. It needs to provide sufficiently reliable estimates of stock status so that the HCR will determine appropriate management measures. But, within the MSE context, it must also achieve those estimates in a reasonable amount of time so that multiple evaluations can be performed across a grid of alternative HCRs and operating model scenarios. The 2016 assessment of skipjack tuna in the WCPO using MULTIFAN-CL takes around 16 hours to reach a converged solution. This is clearly too long. An evaluation for 10 management cycles with 200 iterations across the current skipjack uncertainty grid (54 models) would require 108,000 assessments to be run for a single HCR. Later in this paper we describe some software solutions that can help reduce the time it takes to run such evaluations but even so, the use of a full MULTIFAN-CL assessment model within the MP is not currently a viable option and an alternative approach is required. Hilborn (2003) predicted a shift in fisheries management towards the use of management procedures that are based on simple rules and rely on less complex models, whilst more complex models are used to evaluate the robustness of those management procedures. We investigate a number of options for developing a simpler estimation model for skipjack in the WCPO (Section 3).

1.2 Tools for MSE

A recommendation for the development of MSE frameworks is to base both the OM and MP on existing software that has been developed either for broad application or has been specifically developed to evaluate management strategies (Punt et al., 2014). The wide use and extensive testing of such software will minimise the chances of errors due to software coding. The development of additional bespoke software, however, is almost inevitable but can be minimised by the use of existing solutions. We describe the sources of software currently available, the approaches taken to use existing software where possible, and the methods and design approaches taken for the development of new software (Section 4).

2 Operating Model: Data Generation

Because we are trying to evaluate the potential future performance of candidate HCRs and MPs the OM must be able to generate data, conditional on the projected future state of the stock, that resemble the type of data routinely collected for managing the fishery in terms of their statistical properties and characteristics. This is the data generation process identified in Figure 1. The stock assessment software MULTIFAN-CL has recently been modified to enable the generation of such pseudo data. A more detailed description of the generation of pseudo data by MULTIFAN-CL is provided in WCPFC-SC13-2017/SA-IP-05 in terms of the technical details of using MULTIFAN-CL as a data generator and in WCPFC-SC13-2017/MI-IP-01 with specific regard to generating pseudo data for skipjack. The main points are summarised here.

Initial developments of MULTIFAN-CL to produce pseudo data have focussed on the catch, effort and length (or weight) frequency data that make up the input *.frq* file. Additional functionality also exists to generate pseudo tag data although this feature has not yet been fully tested and is not investigated further here. The generation of pseudo data is currently only possible for the future, projection period. Pseudo data cannot yet be generated for the historical time period, although this feature is also currently under development. The ability to generate pseudo data for the historical time period would be particularly useful when testing the "conditioning" of the OM.

Conditioning the operating model is the process of fitting the operating models to the available data to generate a set of plausible representations of the stock and fishery. Alternative states of nature are represented by these different models which may have different assumptions for those parameters that are less well estimated or highly influential for stock dynamics (e.g. alternative rates of growth or natural mortality, stock and recruitment dynamics, spatial stock structure and mixing, etc.). In this way, the suite of OMs is derived along similar lines to the uncertainty grid of assessments that is routinely presented for WCPFC stock assessments.

Simulated data are generated from these OMs to apply the candidate management strategies in the simulations. Conditioning is a critical phase of the development of the operating model. It ensures that the characteristics and statistical properties of the simulated data are consistent with those of the observed data. The expert consultation workshop (Scott et al., 2016a) noted the considerable work required in conditioning operating models, especially when these are spatially explicit, and in particular it stressed the importance of and difficulty associated with simulating realistic data.

2.1 Catch and effort and CPUE data

In this exercise we have used MULTIFAN-CL to generate pseudo catch, effort and length frequency data from the 2016 WCPO skipjack reference case assessment model and compare the historical observed data to the generated pseudo data. These comparisons are largely based on visual inspection.

There are clear shifts in CPUE for some fleets between the historical, observed data and future generated data (see WCPFC-SC-2017/MI-IP-02 for more information). This is particularly evident for some of the pole and line fisheries, but also the case for some of the seine fisheries too.

We have assumed here a constant standard deviation of 0.1 for catch which was held constant through time. These initial settings have allowed for preliminary testing of the software but may not represent appropriate values to be carried forward when testing harvest control rules. Further work will be necessary to identify more appropriate fleet specific variance estimates for both catch and effort such that future generated data more appropriately represent historical observations.

2.2 Pseudo length frequency data

Pseudo length frequency data are generated assuming a multinomial distribution for the observed length frequency data and requires that a user defined effective sample size (ESS) be supplied. For a small ESS the simulated data are highly variable, whilst for a large ESS the simulated data conform more rigidly to the multinomial distribution. The term *"effective"* sample size is used because it is

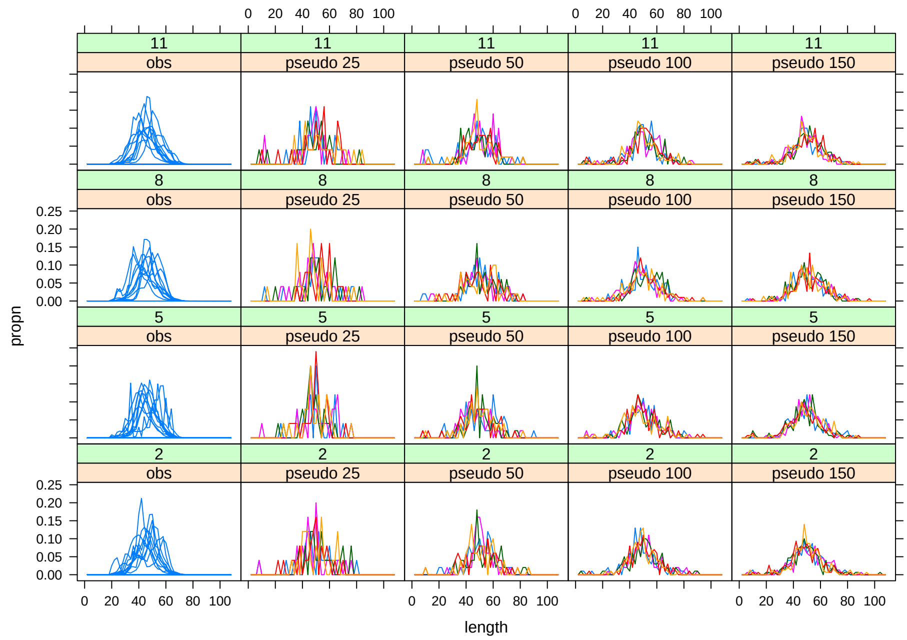

Figure 2: Quarterly length frequency data (months 2,5,8,11, by rows) for the skipjack associated purse-seine fishery in assessment region 2 (S-ASS-ALL-2). Historical observations for the period 2005 to 2015 are shown in the left column, and five iterations of pseudo data generated from MFCL, each for increasing effective sample sizes (25, 50, 100, 150) in subsequent columns.

usually much smaller than the actual size of the sample that generated the observations. This is because length frequency data are often highly correlated and estimates of uncertainty based on the actual sample size is an underestimation of the true error in the observations (Hulson et al., 2011). We investigated a range of ESSs for generating length frequency data for the associated purse seine fishery in region 2 of the WCPO skipjack assessment.

The characteristics of the pseudo length frequency data differ from those of the observed length frequencies (Figure 2). At low effective sample sizes the pseudo generated length frequencies show greater variability in the tails of the frequency distribution. As the effective sample size increases the variability in the tails decreases but continues to be greater than that of the observed data. In addition, throughout the range of effective sample sizes the modal distribution of the pseudo length frequency data remains invariant in comparison to the observed data for which greater variation in modal distribution is apparent.

The pseudo length frequency data are generated from multinomial re-sampling of the observed length frequency data only. This approach takes account of observation error in the composition data but does not allow for process error. Process error can arise if, for example, the fishery catches larger (or older) fish than usual, perhaps as a result of fish movement. Under this scenario, the entire length frequency distribution would be shifted to the right (towards larger fish). The opposite would occur if smaller fish were caught. This would have the appearance of random variation in

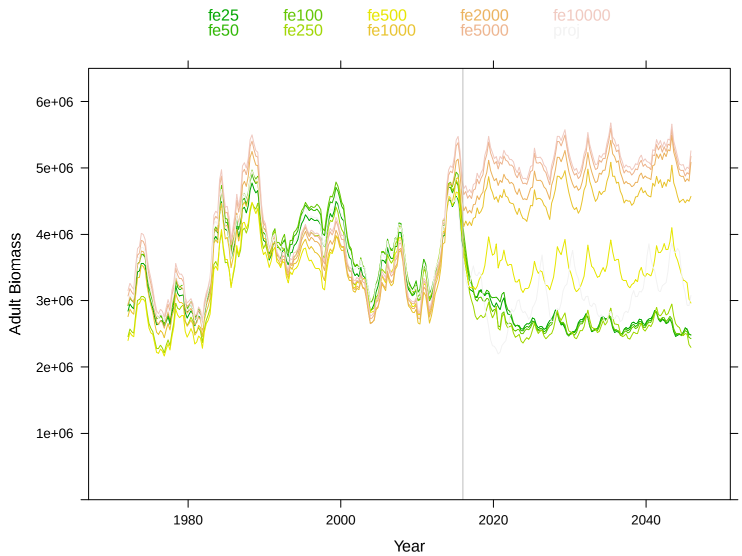

Figure 3: Estimates of adult biomass derived from the refitted MULTIFAN-CL model with increasing (green to brown) function evaluations. Vertical grey line denotes the transition between historical observed data and future pseudo data.

selectivity. A more sophisticated approach to generating length frequency data that also takes account of process error (through, for example, selectivity deviations) may be a more appropriate way to generate simulated length frequencies that better resemble the observed data.

2.3 Refitting MULTIFAN-CL

As a further analysis we ran a series of assessment runs over a range of function evaluations to see how well and how quickly MULTIFAN-CL was able to fit the pseudo generated data. We used the converged model solution from the 2016 skipjack reference case assessment as the starting point for these runs under the assumption that if the pseudo data were an accurate representation of the historic data series the model would find a converged solution to the extended data set in a relatively small number of function evaluations. We note that a more direct comparison would be to generate pseudo data for the historical time series and to re-fit the 2016 reference case model to the new data using exactly the same model settings, time periods etc. but for the reasons described above this is not possible at present.

We generated catch, effort and length composition data for a 30 year period and, taking the 2016 reference case assessment as a starting point, refitted MULTIFAN-CL to the extended data set to see how well and how quickly the model converged to a solution.

Running MULTIFAN-CL over an increasing range of function evaluations revealed relatively poor levels of convergence, for a model data set that has been increased by 30 years, for model runs of less than 5,000 function evaluations. Estimates of adult biomass from each of the nine runs with increasing function evaluations are shown in Figure 3. Adult biomass is initially estimated to be relatively low, but the estimates increase progressively as the number of function evaluations increases and eventually appears to stabilise at higher biomass levels. These terminal values differ from those of the projections from which the pseudo data were derived and indicate that more work is required in order to get reasonable estimates of stock status from the generated pseudo data.

3 Management Procedure: Estimation Method

Construction of the MP is typically more straightforward than the task of designing and developing the OM, but not without its complications. In the case of WCPO skipjack a particular problem, that was highlighted during the MSE expert consultation workshop, is the choice of model to use within the MP to estimate stock status. This component is referred to in Figure 1 as the estimation method because it can be based on either empirical or model based methods. The expert workshop considered that empirical estimation methods (e.g. trends in CPUE) would be unlikely to work well in instances where purse seine fisheries account for a large proportion of the catch because purse seine CPUE is notoriously difficult to interpret and unlikely to provide a sufficiently reliable signal to the HCR. For skipjack, therefore, investigations have concentrated on a model-based approach.

In many MSE applications the assessment method that is routinely used to assess the stock has been used within the MP to provide information on stock status to the HCR. Tuna assessments conducted for the WCPFC, however, are based on seasonally stratified, spatially disaggregated models that employ large quantities of tagging data in addition to fleet specific catch, effort and length composition data. The models are large, complex and take a long time to run which makes them undesirable for use within a simulation framework that will need to run for a large number of scenarios and for many iterations.

We investigated two alternative approaches to estimating stock status for WCPO skipjack. The first was based on the stock assessment software a4a that has been developed as part of the FLR project (`http://flr-project.org`, see section 4.2). The second approach attempted to trim down the current MULTIFAN-CL assessment model to produce a simplified and faster running model.

3.1 a4a model

The a4a initiative (Jardim et al., 2015) aims to develop a more intuitive stock assessment environment based on a risk type analysis that more explicitly accounts for major sources of uncertainty. The approach uses a statistical catch at age assessment model, implemented in R (R Core Team, 2016), making use of the FLR framework (Kell et al., 2007) and using automatic differentiation implemented in ADMB (Fournier et al., 2012) as the optimisation engine.

a4a is an age based assessment method. To examine its ability to assess WCPO skipjack, it was therefore necessary to convert the length based MULTIFAN-CL input data for the 2016 stock assessment to be age based. This was done using a relatively simple length slicing approach that used the fitted von Bertalanffy growth parameters, as estimated from the 2016 reference case MULTIFAN-CL assessment, to convert the data into age increments with one year intervals. Although MULTIFAN-CL takes length based information as inputs, its internal calculations are conducted by age thus allowing the same age based values for natural mortality, maturity and weight to be used as those used in the 2016 reference case assessment.

The a4a method is based on catch at age information and, as for many stock assessment models, requires one or more indices of abundance with which to "tune" the model. These indices may be either age specific or aggregated across ages as a total biomass index. Where possible they may be derived from fishery independent data, such as research surveys, but are often constructed from subsets of the total fishery data for which the underlying relationship between catch rates and abundance is considered to be well represented. We investigated options for creating age specific tuning indices from the available catch and length composition data (Section 3.1.1), and total biomass indices using the tag release and recapture data (Section 3.1.2).

3.1.1 Creation of an age-specific tuning index

The relationships between CPUE and abundance for fleets fishing in region 2 of the stock assessment area (Figure 4) show a consistent trend for increasing CPUE with increasing abundance. In all cases, the r-squared statistic for each of the fits was small (< 0.35) indicating that, although positive, no strong relationship existed for any of the fishing fleets for which catch and effort data are available. Calculations for other regions of the assessment (not shown) gave similar, or worse fitting results. Although none of the fisheries examined showed a strong relationship between CPUE and abundance, the pole and line fishery in region 2 (P-ALL-2) showed a positive relationship that was significant in three of the four quarters and was therefore selected as a potential candidate for use as a tuning index.

The more internally consistent an age based tuning index is the more informative it will be when estimating rates of cohort decline and exploitation levels. Internal consistency here refers to the extent to which catches at one age are correlated to catches at the next age lagged by one year; at two ages higher lagged by 2 years, and so on. In other words, if catches are high for one year old fish in a given year, it might be expected that catches of two year old fish would be high the following year. P-ALL-2 shows relatively good internal consistency for 0 group and 1 group fish (Figure 5) but the relationship rapidly declines for older ages and at lags greater than 1 suggesting that this tuning series would not be particularly informative to the stock assessment.

Preliminary a4a assessment runs using P-ALL-2 as either a total biomass or age structured tuning index gave generally poor results. Model diagnostics suggested poor fits either to the overall catch data or to the tuning indices, and model outputs (adult biomass and fishing mortality) appeared

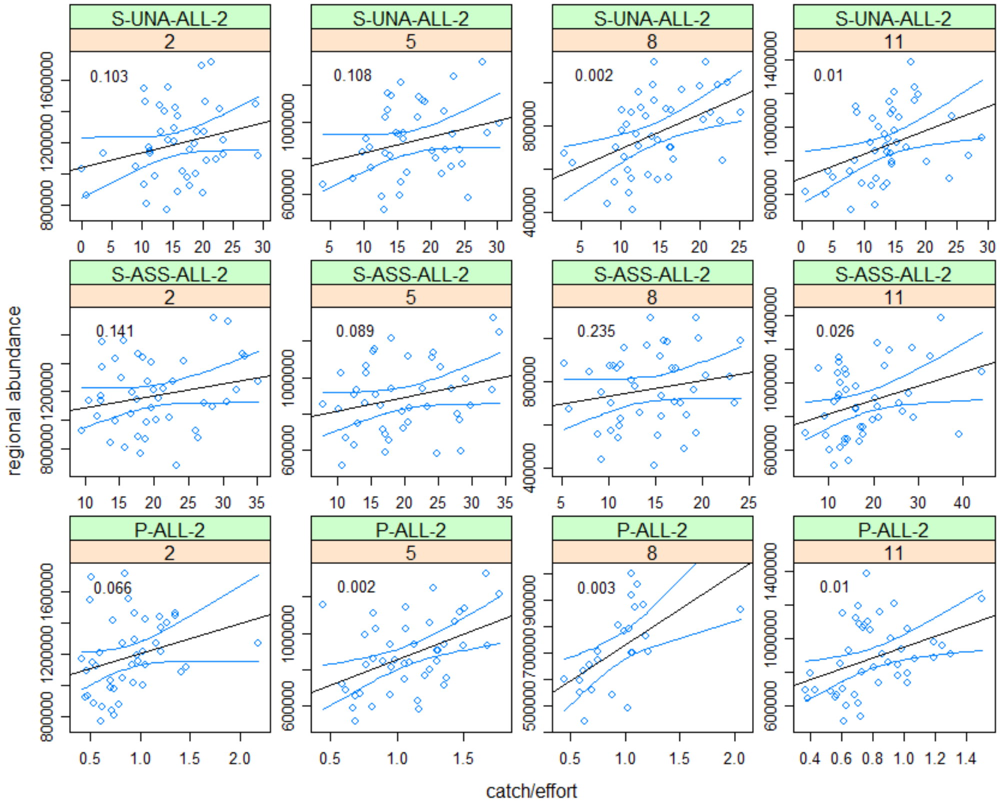

Figure 4: Skipjack total biomass and CPUE, by quarter (months 2,5,8,11), for fleets fishing in region 2 of the assessment area. Upper and lower 95% confidence intervals are shown for a simple linear regression. p-values for each fitted model are shown in the upper left corner of the panels

unrealistic.

3.1.2 Creation of a tag based tuning index

Tag release and recapture data are used within MULTIFAN-CL to inform on stock movement rates and estimates of natural mortality. However, methods are available to use tag recapture information to determine absolute estimates of total abundance. We investigated a simple Petersen method to estimate total stock abundance from recaptures in time period j of fish tagged and released in time period i as follows

$$N_{i,j} = \frac{n_{i,j}.c_j}{m_{i,j}} \quad (1)$$

where $n_{i,j}$ is the total number of fish tagged and released at time period i that are estimated to be available for recapture in time period j; c_j is the total weight of fish caught in time period j, and $m_{i,j}$ is the total number of fish tagged in time period i and recaptured in time period j.

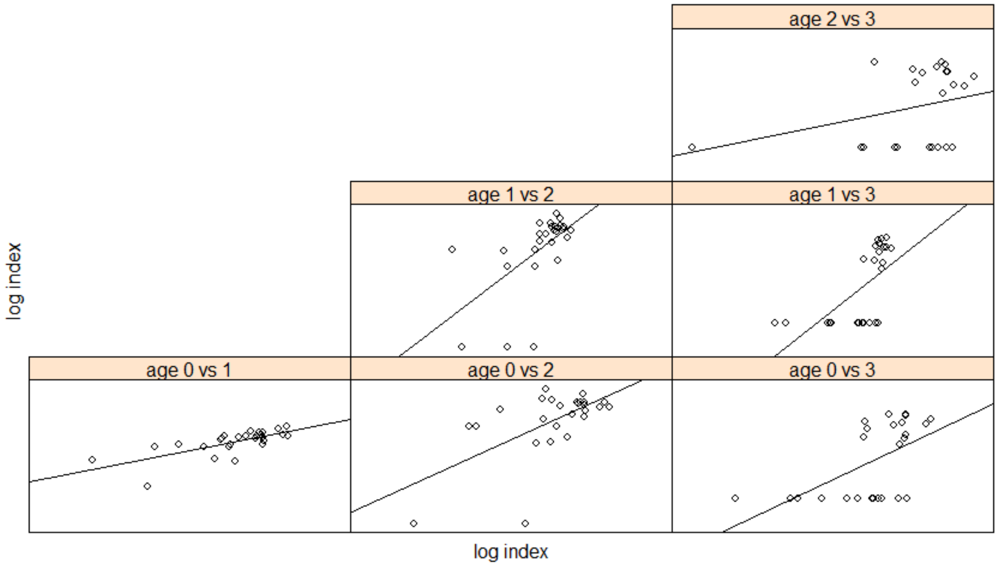

Figure 5: P-ALL-2 internal consistency plot showing pairwise comparisons (on the log scale) of numbers at age relative to numbers at age lagged by one, two or three years.

The numbers of fish estimated to be available for recapture following release was determined from the number of actual tags released in time period i corrected for tag induced instantaneous mortality t, tag failure f and natural mortality M.

$$n_{i,j} = n_i.(1-t).e^{-(M+f)(i-j)} - \sum_{k=i}^{j} m_{i,k}.e^{-(M+f)(j-k)} \tag{2}$$

Prior to calculating the index the tag recaptures were raised by their respective tag reporting rates (Peatman et al., 2016). For these initial investigations tag failure and tag induced mortality were both assumed to be zero. Recaptures of tagged fish that were at liberty for less than one quarter were removed from the analysis to allow for mixing of the tag releases into the population. Similarly release events for which less than three tags were recaptured were also removed from the analysis to prevent excessively high estimates of biomass that can occur with very low recapture rates.

Tag based estimates of total biomass for the SSAP, RTTP and PTTP tagging programs are shown in Figure 6 along with the 2016 reference case estimate of total skipjack biomass as estimated by MULTIFAN-CL. For all three tagging programs estimates calculated for each release and recapture event (open circles) show considerable variation. Averaging the estimates for each year-quarter (solid circles) reduces the variation slightly. Tag based estimates of total stock abundance from the SSAP and PTTP are approximately centred around the MULTIFAN-CL estimates for the corresponding period. Estimates for the RTTP, however, are consistently lower than those of MULTIFAN-CL.

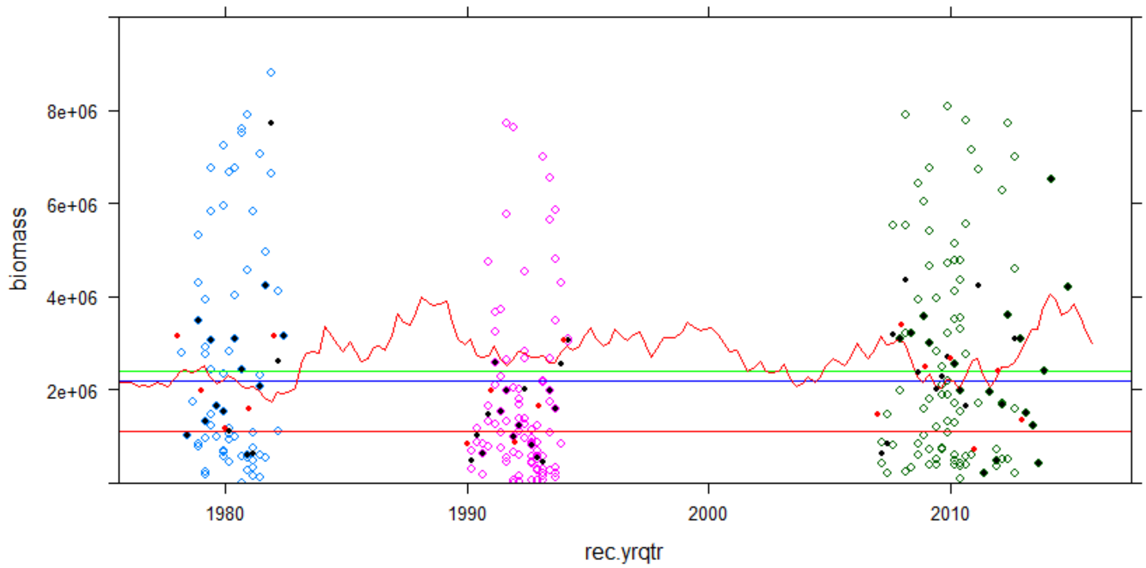

Figure 6: Tag-based Petersen biomass indices for skipjack. Biomass estimates for individual tag release events (open dots) and year-quarter means (solid dots) are shown for the SSAP (blue), RTTP (red) and PTTP (green) tagging programs against adult skipjack biomass (wiggly red line) as determined from the 2016 reference case assessment. Solid horizontal lines denote the median biomass estimate from each of the tagging programs.

3.1.3 Results

Preliminary results from the a4a assessments showed generally poor fits to the data and yielded model estimates of stock abundance and fishing mortality that were not considered to be realistic. Given the poor internal consistency of the age structured indices and the high variability and lack of temporal correlation in the tag based biomass estimates this result is perhaps not surprising. We do not consider the assessment outputs here in any further detail.

We note that this was a relatively quick study that was carried out to investigate the potential use of a relatively simple assessment model within the management procedure and that the problems encountered in this instance might be common for other potential approaches. With more time a number of issues could be further investigated to resolve some of the problems we encountered. Perhaps one of the most important areas for further investigation would be the approach taken to slice the length based catch data into age based information. We have based our model on annual age increments. A quarterly age increment would potentially preserve more information in the cohort structure and give better internal consistency to both the catch at age matrix and the tuning indices. The a4a software appears to have been developed primarily for input data with annual increments and applications using quarterly time steps would require further investigation. We note the approach taken by Scott et al. (2015) to take account of uncertainty in biological processes (including the length slicing process) by integrating over multiple scenarios using a model averaging approach (Millar et al., 2015).

The a4a assessment software benefits from being a comparatively simple assessment approach and from being developed as part of a modelling environment that was designed specifically for developing simulation frameworks. Its simple structure, and use of ADMB for minimisation, make it extremely fast and its design allows it to interact seamlessly with other components of the simulation framework without requiring extensive manipulation of input files. However, until it can be used to derive reasonable estimates of stock status for WCPFC stocks it cannot be considered suitable for use within the MP of the MSE framework.

3.2 Reduced MULTIFAN-CL model

As we note above, the full MULTIFAN-CL assessment for many WCPO stocks can take a long time to run to completion. The actual time taken can vary depending on the speed of the computer and the number of function evaluations required to reach satisfactory convergence in each phase of the assessment. In some instances it may be possible to reduce the number of function evaluations. For example, a large number of function evaluations are often required in the final phase to ensure the hessian matrix is well estimated. In instances where the hessian matrix is not required the final phase of the assessment can be shortened considerably with little effect on the final model estimates. Even so, the time taken to fit a five region model with 23 fishing fleets, such as the current WCPO skipjack assessment, remains prohibitively slow for use within a simulation framework.

The time taken for MULTIFAN-CL to reach a converged solution can also be shortened by reducing the complexity of the model. We investigated the use of alternative model structures that had fewer regions and fewer fleet components to see if run times could be reduced to acceptable levels whilst maintaining sufficiently reliable estimates of stock status and exploitation rates. The reduced assessments were based on a subset of the reference case model comprising regions 2,3 and 5 only of the current WCPO skipjack assessment area (Fig 7). In the first instance we retained the spatial structure for the three regions and estimated movement rates between them. To create an even simpler model we merged regions 2,3 and 5 into a single region that required no estimation of movement across regional boundaries.

It is important to note that whilst the estimation method within the MP might be based on a reduced area assessment, the OM will continue to cover the entire region of the stock. Performance indicators, and other diagnostics, can therefore be calculated for the whole stock region (or sub-regions if desired) to determine the status of the whole stock, even though the HCR is based on the status of only part of the stock.

3.2.1 Modification of MULTIFAN-CL input files

MULTIFAN-CL assessments run from three initial starting files, a *.frq* file that specifies the fleet specific catch, effort and length composition data; a *.tag* file that holds the tag release and recapture information and a *.ini* file that contains the necessary information to initialise the model. A number of changes were necessary to the MULTIFAN-CL input files in order to run both the single region and three region skipjack assessment models.

In both cases the fishing fleets were restricted to the 12 fleets that operate in regions 2,3 and 5, corresponding to fleets 4 to 15 from the WCPO skipjack assessment, (McKechnie et al., 2016). The fleet specific area identifiers in the *.frq* file were then adjusted as appropriate for each fishery depending on whether a single region or 3 region model was being developed. Similar adjustments were also made to the relative region size, data flags and movement matrix components of the *.frq* file.

In both cases the *.tag* files required substantial modification so that only data relevant to the reduced assessment were carried forward. The first step was to remove all records of tagged and recaptured fish for those fish that had been released from regions outside of the reduced area. Release numbers were then adjusted downwards to take account of those fish that had been released inside the reduced area but recaptured outside of it. Finally the retained release events were renumbered so as to have a continuous, ascending order.

Once modification of the tag file was complete the *.frq* and *.ini* files were further updated to have the corresponding number of tag release events and appropriate entries for estimates of tag reporting rates, tag groupings and penalties, diffusion coefficients and regional recruitment distribution.

3.2.2 Modification of the doitall

MULTIFAN-CL typically runs in a series of phases, the outputs of one phase being used as the inputs to the next along with any adjustments to the model settings. This allows parameters to be held constant initially and progressively freed as the model converges towards a solution. The file that automates this phased fitting approach and controls the various adjustments to the model settings has become known as the *"doitall"*.

Changes were made to the doitall only for those settings that were related either to the change in number of regions or the change in number of fleets in the reduced models. For the three region model the doitall comprised 10 phases (the same as the 2016 reference case model). However, a number of phases could be removed completely for the single region model for which only 5 model fitting phases were necessary.

A reduction in the number of phases can substantially speed up the model fitting process. The 3 region model showed little improvement in run time over the reference case model, however, run

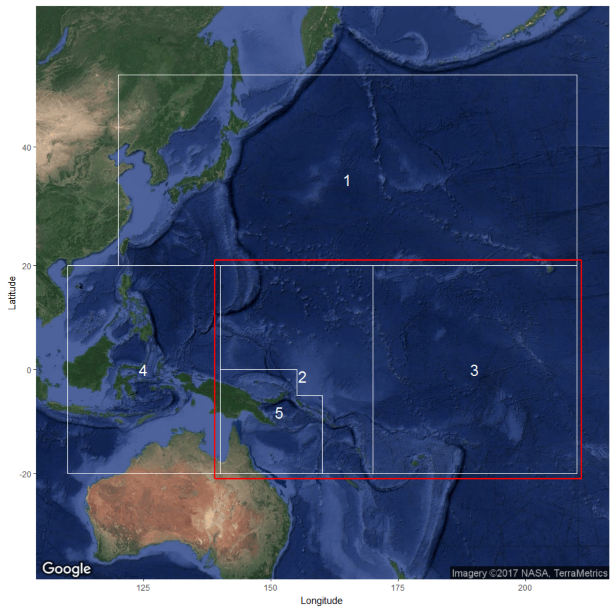

Figure 7: The geographical area covered by the WCPO skipjack stock assessment. The reference case assessment is based on a 5 region model structure (white lines). The reduced assessment models were conducted for the areas 2, 3 and 5 only (red lines).

time for the single region model was significantly reduced to just 20 minutes.

3.2.3 Results

The model diagnostics of the reduced area assessments have not yet been examined in detail but summary plots of the resulting estimates of adult biomass (Figure 8) show the three region model achieves similar trends and absolute levels to those of the same regions in the reference case assessment whilst the single region model estimates biomass to be at a much lower level. When the adult biomass estimates are plotted on a relative scale, however, (Figure 9) the overall trends appear similar (although with some differences) for all three models.

Estimates of stock status ($SB/SB_{F=0}$, Figure 10), calculated for regions 2, 3 and 5 only, are similar for the reference case and 3 region model, but the single region model estimates lower adult biomass relative to un-fished biomass throughout the model time series.

Although the single region model provides lower estimates of stock status than the other model formulations, this does not mean that it cannot be used as an estimation method within the MP. The design of the HCR can be tailored to accommodate the type of information supplied to it by the estimation method. In this respect it is important to separate the overall objectives for the fishery (to maintain the stock as a whole around the TRP and away from the LRP) from the trigger points of the HCR which may be set to maintain a component of the stock at some other level. However, it is too early to enter into this discussion in any detail. Much more testing of the reduced assessment

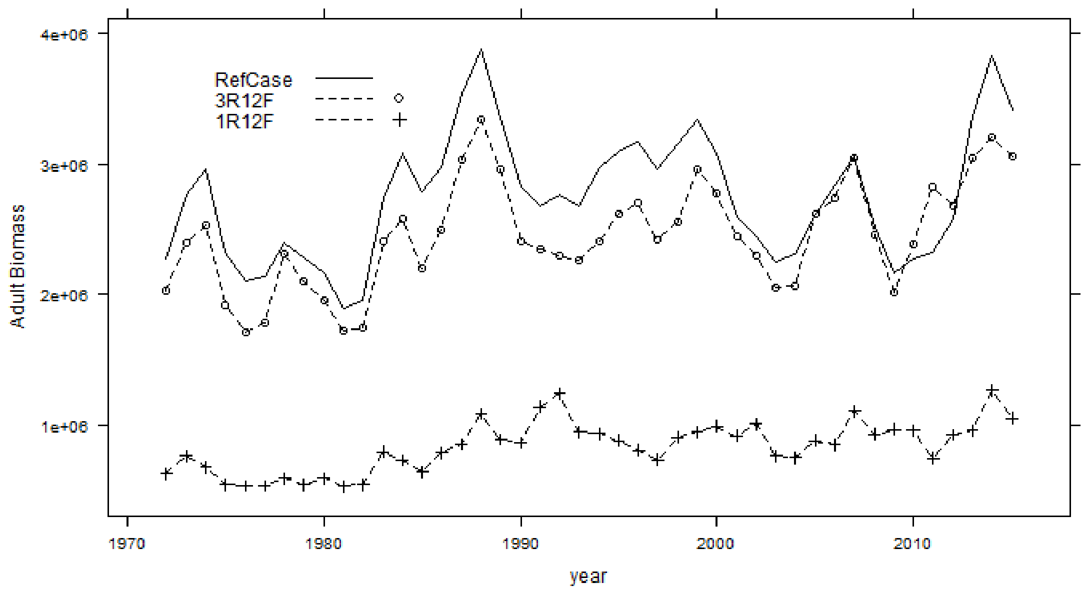

Figure 8: Estimates of adult biomass from the 2016 reference case assessment (summed over regions 2,3 and 5 only) and reduced area assessments based on either three region, twelve fleet (3R12F) or a single region (1R12F) configurations.

approach is required before substantive discussions on the design of the HCR can take place.

4 Tools for MSE

The design and construction of an MSE framework for testing HCRs and MPs for WCPFC stocks and fisheries is far from agreed and many more trials and exploratory analyses still need to be carried out before a finalised framework can be presented. A number of key tools and software approaches have been initially identified as potential building blocks with which to construct the framework. We describe some of them briefly here.

4.1 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 and has been identified as a potentially useful tool for developing the OM. In addition to being used for the OM, MULTIFAN-CL may have a role as the estimation method within the MP for specific stocks.

The use of the same model to implement the OM as is used in the MP to estimate stock status is generally not advised. If the modelling approaches in both elements of the simulation framework

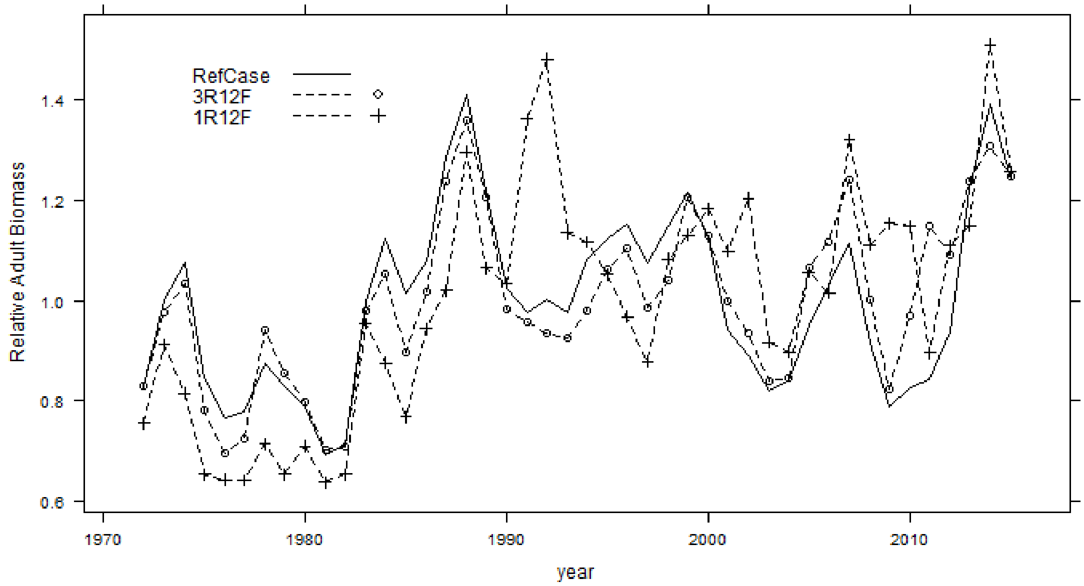

Figure 9: Estimates of relative adult biomass from the 2016 reference case assessment (summed over regions 2,3 and 5 only) and reduced area assessments based on either three region, twelve fleet (3R12F) or a single region (1R12F) configurations.

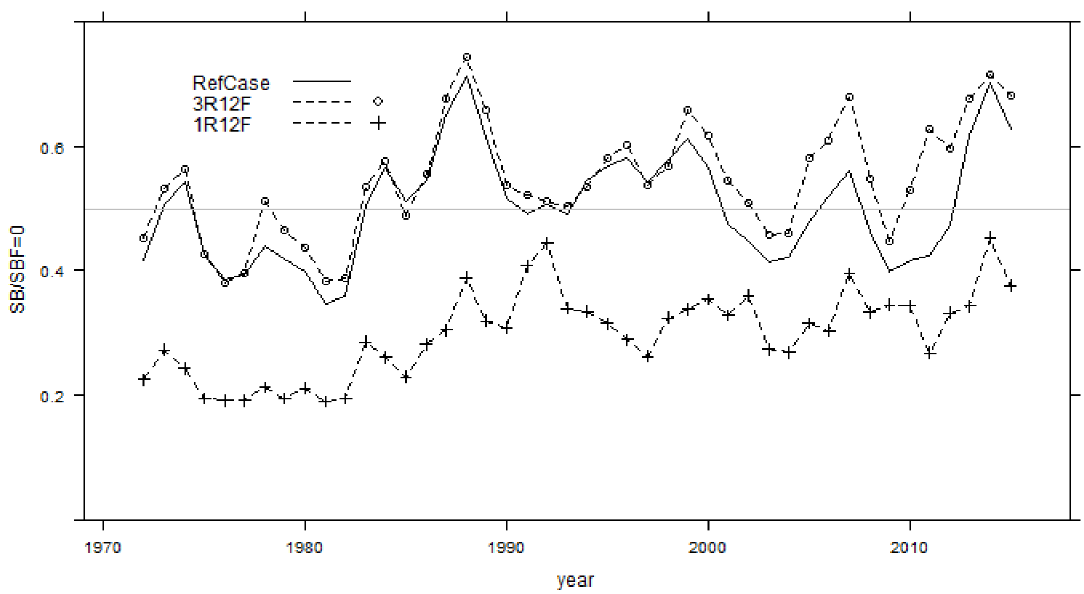

Figure 10: Estimates of SB/SBF=0 from the 2016 reference case assessment (summed over regions 2,3 and 5 only) and reduced area assessments based on either three region, twelve fleet (3R12F) or a single region (1R12F) configurations..

are too similar it is possible for analysts to over-estimate the amount of uncertainty that they are really including in their evaluations. MULTIFAN-CL is, however, a very flexible model allowing alternative model structures to be developed and explored. In this way, the models used for the OM and the MP can be quite different even though the same software is used to implement them. We have investigated here the potential to use an alternative MULTIFAN-CL model within the MP that may prove to be a viable approach. We will, however, need to be careful as we proceed, to ensure that efforts to include uncertainty in the OM are not mirrored exactly, and therefore negated, in the MP.

4.2 FLR

The FLR initiative (Kell et al., 2007; Hillary, 2009) is a development effort directed towards the evaluation of fisheries management strategies. Its overall goal is to develop a common framework to facilitate collaboration within and across disciplines (e.g. biological, ecological, statistical, mathematical, economic, and social) and in particular, to ensure that new modelling methods and software are more easily validated and evaluated, as well as becoming widely available once developed. Specifically, the framework details how to implement and link a variety of fishery, biological, and economic software packages so that alternative management strategies and procedures can be evaluated for their robustness to uncertainty before implementation. The FLR software is freely available and hosted on github (`http://www.flr-project.org/`). It has been used extensively to develop MSE frameworks and to test MPs in a wide range of fisheries fora (Miller et al., 2008; Bastardie et al., 2010; Mosqueira, 2011; Tserpes et al., 2016).

FLR takes advantage of S4 classes in R enabling a hierarchical and extensible package design. In this way subsidiary packages can be developed for more specific purposes that inherit the basic functionality of the parent package. FLR4MFCL (`https://github.com/PacificCommunity/ofp-sam-flr4mfcl`) has been developed as one such subsidiary package to FLR providing specific functionality for MULTIFAN-CL.

4.3 High-throughput Computing: HT-Condor

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 recently been granted access to the HT-Condor network at the Center for High Throughput Computing (CHTC) at the University of Wisconsin (`http://chtc.cs.wisc.edu/`) which substantially increases the available processing power for running large simulations. There are however, a number of restrictions on individual run times and the size of files transferred across the network which may have implications for the design of the simulation framework.

5 MSE Modelling Framework

This paper presents an account of recent developments in the construction of the MSE framework for WCPFC stocks and fisheries with particular focus on some of the technical issues related to the development of the modelling framework. In many cases, the results presented here are preliminary and represent the current state of ongoing analyses. For this reason we have not drawn many firm conclusions from the work so far. We have explored options to solve some of the technical challenges associated with developing an MSE modelling framework, but other options remain to be investigated and other challenges are still to be addressed.

Based on the work presented here, some key approaches can be identified as likely solutions and therefore prioritised for further development:

- *The use of MULTIFAN-CL for the OM*
  As noted above, MULTIFAN-CL has been identified as an appropriate tool to use as the OM within the simulation framework. Initial work has been undertaken to develop MULTIFAN-CL for data generation although, as we have shown here, further development work in this area is still required.
- *The use of a reduced MULTIFAN-CL model for the Estimation Method*
  Although not ideal (and with specific regard to skipjack) a simplified MULTIFAN-CL model currently represents the most successful approach for estimating stock status within the MP. Run times for fitting the reduced model can be substantially reduced but are still longer than desired.
- *The use of FLR4MFCL for developing the framework*
  FLR4MFCL is based on existing software that has been used extensively to develop MSE frameworks and to test MPs. The FLR4MFCL package provides specific functionality for MULTIFAN-CL but also takes advantage of the methods and tools that have been developed for more generic application in MSE modelling approaches.
- *The use of HT-Condor and CHTC for running the simulations*
  Access to the CHTC HT-Condor flock significantly increases the processing power available to OFP-SAM for running the simulations. There are however, a small number of restrictions on the sizes of files that can be passed across the network connection and the run times for individual simulations that may affect the way that the simulations are run.

In addition to the above, a number of issues remain to be resolved that we have not yet, or only just, begun to consider. These include, amongst others, the generation of pseudo tag data; alternative options for the estimation method (for stocks other than skipjack); the design of the implementation model (that controls the extent to which the management measures resulting from the HCR are actually implemented in the OM); the scenarios to use for the suite of OMs and the methods for selecting a "best" MP from the performance indicators.

Progress has been made in a number of areas with regard to the development of the MSE framework but a considerable amount of work remains. Recent funding to SPC-OFP from the New Zealand Ministry of Foreign Affairs (MFAT) will provide support for three scientific positions to work exclusively on the development and implementation of harvest strategies for the WCPFC. These positions have recently been advertised and we anticipate will be filled in the very near future.

Acknowledgements

We gratefully acknowledge funding for this work from the European Union through their funding support for the WCPFC "Simulation testing of reference points" and Project 75 (Technical Support for the Development of Harvest Strategies" and also from the New Zealand Ministry of Foreign Affairs and Trade (MFAT) funded project "Pacific Tuna Management Strategy Evaluation".

References

Bastardie, F., Vinther, M., Nielsen, J., Ulrich, C., and Storr Paulsen, M. (2010). Stock based vs. fleet-based evaluation of the multi-annual management plan for the cod stocks in the Baltic Sea. *Fisheries Research*, 101:188–202.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielson, A., and Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2):233–249.

Hilborn, R. (2003). The state of the art in stock assessment: where we are and where we are going. *Scientia Marina*, 67(Suppl. 1):15–20.

Hillary, R. (2009). An introduction to FLR fisheries simulation tools. *Aquatic Living Resources*, 22:225–232.

Hulson, P-J, F., Hanselman, D., and Quinn, T. (2011). Effects of process and observation errors on effective sample size of fishery and survey age and length composition using variance ratio and likelihood models. *ICES Journal of Marine Science*, 68(7):1548–1557.

Jardim, E., Millar, C. P., Mosqueira, I., Scott, F., Chato Osio, G., Ferretti, M., Alzorriz, N., and A., O. (2015). What if stock assessment is as simple as a liner model? The a4a initiative. *ICES Journal of Marine Science*, 72(1):232–236.

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.

Kirchner, C. H., Berger, Pilling, G., and Harley, S. (2014). Management strategies (objectives, indicators, reference points and harvest control rules): the equatorial skipjack purse seine fishery as an example. WCPFC-SC10-2014/MI-WP-02.

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.

Millar, C., Jardim, E., Scott, F., Osio, G., Mosqueira, I., and Alzorriz, N. (2015). Model averaging to streamline the stock assessment process. *ICES Journal of Marine Science*, 72(1):93–98.

Miller, D., Shelton, P., Healey, B., Brodie, W., Morgan, J., Butterworth, D., Alpoim, R., Gonzalez, D., Gonzalez, F., Fernandez, C., Ianelli, J., J-C., M., Mosqueira, I., Scott, R., and Vazquez, A. (2008). Management strategy evaluation for Greenland halibut (Reinhardtius hippoglossoides) in NAFO Subarea 2 and Divisions 3LKMNO. Scientific council meeting, NAFO SCR Doc. 08/25.

Mosqueira, I. (2011). Development of a management strategy evaluation process for IOTC. IOTC-2011-SC14-36.

Peatman, T., Caillot, S., Leroy, B., McKechnie, S., Roupsard, F., Sanchez, C., Nicol, S., and Smith, N. (2016). Analysis of tag seeding data and reporting rates. WCPFC-SC12-2016/SA-IP-13, 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 (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Scott, F., Jardim, E., Millar, C., and Cervino, S. (2015). An applied framework for incorporating multiple sources of uncertainty in fisheries stock assessments. *PLoS ONE*, 11(5).

Scott, R., Pilling, G. M., Brouwer, S., and Hampton, J. (2016a). Evaluation of candidate harvest control rules for the PNA skipjack purse seine fishery. WCPFC-SC12-2016/MI-WP-06, Bali, Indonesia, 3–11 August 2016.

Scott, R., Pilling, G. M., Hampton, J., Reid, C., and Davies, N. (2016b). Report of the expert consultation workshop on management strategy evaluation. WCPFC-SC12-2016/MI-WP-05, Bali, Indonesia, 3–11 August 2016.

Tserpes, G., Nikolioudakis, N., Maravelias, C., Carvalho, N., and Merino, G. (2016). Viability and management targets of Mediterranean demersal fisheries: The case of the Aegean Sea. *PLoS ONE*, 11(12):15pp.

A List of Abbreviations

Abbreviation	Meaning	Description
a4a	Assessment for all	A non-linear catch-at-age model implemented in R/FLR/ADMB with low parametrization requirements.
ADMB	AD Model Builder	a C++ application which implements automatic differentiation using specialized classes and operator overloading. http://admb-foundation.org
FLR	Fisheries Library in R	A collection of tools for quantitative fisheries science, developed in R, that facilitates the construction of bio-economic simulation models of fisheries systems. http://www.flr-project.org
FLR4MFCL		A subsidiary package to FLR providing specific functionality for MULTIFAN-CL
Git	...	An open source version control system for tracking changes in computer files and for coordinating collaborative development of software.
HCR	Harvest Control Rule	
HTC	High-throughput computing	
MFCL	MULTIFAN-CL	Integrated stock assessment model http://www.multifan-cl.org/
MP	Management Procedure	
MSE	Management Strategy Evaluation	
OM	Operating Model	
SS3	Stock Synthesis 3	Integrated stock assessment model http://nft.nefsc.noaa.gov/Stock_Synthesis_3.htm