



**SCIENTIFIC COMMITTEE  
TENTH REGULAR SESSION**

Majuro, Republic of the Marshall Islands  
6-14 August 2014

---

**DEVELOPMENTS IN THE MULTIFAN-CL SOFTWARE 2013-14**

---

**WCPFC-SC10-2014/SA-IP-08**

**Nick Davies<sup>3</sup>, Dave Fournier<sup>2</sup>, John Hampton<sup>1</sup>, Shelton Harley<sup>1</sup>,  
and Fabrice Bouyé<sup>1</sup>**

<sup>1</sup> Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia

<sup>2</sup> Otter Research Ltd, North Saanich, British Columbia, Canada

<sup>3</sup> Te Takina Ltd, Whangarei, New Zealand

## Table of Contents

Executive summary .....	4
1 Introduction .....	5
2 Development overview .....	5
2.1 Team .....	5
2.2 Calendar .....	5
2.3 Collaboration and versioning.....	5
2.4 Developer’s workshops.....	6
2.5 Postings to website .....	6
2.5.1 7 November 2013 – version 1.1.5.5.....	6
2.5.2 14 February 2014 – version 1.1.5.6.....	6
2.6 Independent Peer Review of the 2012 bigeye tuna stock assessment .....	6
2.7 Tool development.....	7
2.7.1 R4MFCL.....	7
2.7.2 Testing framework.....	7
2.7.3 Viewer .....	8
2.7.4 Condor parallel processing facility .....	8
2.8 User’s guide .....	8
3 New features .....	8
3.1 Sex-disaggregation .....	8
3.1.1 Rationale .....	8
3.1.2 Methods and Testing .....	9
3.1.2.1 Tagging data shared over multiple species/sexes .....	9
3.1.2.2 Catch data shared over multiple species/sexes .....	10
3.1.2.3 Length- and weight-frequency data shared over multiple species/sexes .....	11
3.1.2.4 Single equilibrium yield over multiple sexes .....	11
3.1.2.5 –makepar operation for multiple sexes .....	12
3.2 Time-variant selectivities .....	12
3.2.1 Rationale .....	12
3.2.2 Methods and testing.....	12
3.3 Fit to conditional age-length data .....	13
3.3.1 Rationale .....	13
3.3.2 Methods and Testing .....	13
3.4 Multinomial distribution for size composition data.....	13
3.4.1 Rationale .....	13
3.4.2 Methods and Testing .....	13
3.5 Likelihood component table.....	13
3.5.1 Rationale .....	13

3.5.2	Methods and Testing .....	13
3.6	Fit the BH-SRR to annual recruitments .....	14
3.6.1	Rationale .....	14
3.6.2	Methods and Testing .....	14
3.7	Initial biomass approximates the equilibrium unexploited biomass.....	14
3.7.1	Rationale .....	14
3.7.2	Methods and Testing .....	14
3.8	Region-specific yield estimation.....	15
3.8.1	Rationale .....	15
3.8.2	Methods and Testing .....	15
3.9	Fixed terminal recruitment deviates .....	16
3.9.1	Rationale .....	16
3.9.2	Methods and Testing .....	16
3.10	Logistic-normal size composition likelihood.....	16
3.10.1	Rationale.....	16
3.10.2	Methods and Testing.....	16
3.11	Exclude reporting rates from tag predictions during mixing period .....	16
3.11.1	Rationale.....	16
3.11.2	Methods and Testing.....	16
4	Other enhancements and bug fixes .....	17
4.1	Correction to fishing mortality calculation.....	17
4.2	Tag diagnostics report.....	17
4.3	Correction to using annual biomass in the BH-SRR fit.....	17
4.4	Non-integer tag releases.....	17
4.5	Increased independent variables .....	17
4.6	Windows compilations.....	18
5	Application of new features.....	18
6	Future work.....	18
7	Discussion .....	18
8	References.....	19
9	Annex.....	31
9.1	Benchmark testing .....	31
9.2	Beverton-Holt spawner stock-recruitment relationship: solution in respect of steepness	32

## EXECUTIVE SUMMARY

This paper presents an account of developments made to the MULTIFAN-CL software project as carried out by the team at SPC-OFI from August 2013 to July 2014. Ianelli et al.(2012) reported thirteen recommendations from an independent peer review panel specifically relating to MULTIFAN-CL. The rationale for these recommendations is to address the key areas of uncertainty for the tuna assessments reported to the Scientific Committee of the Western and Central Pacific Fisheries Commission (SC), through improving the biological description of population processes, better estimation of observation error, and the improved modelling of fishing mortality. These recommendations translate into developments to the MULTIFAN-CL software and those identified for 2013-14 are listed below with respect to their current status of completion.

<b>Peer review recommendation</b>	<b>Status of completion</b>
f. Sex-disaggregation	Development 100%; Testing 100%
a. Time-variant selectivities	Development 100%; Testing 50%
e. Fit to conditional age-length data	Development 100%; Testing 20%
j. Multinomial distribution for size composition data	Development 100%; Testing 0%
g. Likelihood component table	Development 100%; Testing 0%
m. Fit the BH-SRR to annual recruitments	Development 80%; Testing 0%
<b>Other new feature</b>	<b>Status of completion</b>
Initial biomass approximates the equilibrium unexploited biomass	Development 90%; Testing 20%
Region-specific yield estimation	Development 100%; Testing 20%
Fixed terminal recruitment deviates	Development 100%; Testing 100%
Logistic-normal size composition likelihood	Development 90%; Testing 20%
Exclude reporting rates from tag predictions during mixing period	Development 100%; Testing 20%

This represents substantial progress towards the set of recommendations (developments for six out of 13 completed), with the testing of these new features scheduled for Aug. – Dec. 2014. In addition to these, five other new features were added during the 2014 tuna stock assessments, which were either incorporated into the assessments, or were of value during the exploratory phases of model development.

It is suggested that the high priority tasks for the forthcoming 2014-15 year include most of the remaining recommendations of the independent peer review panel, but also a task that has long been important to the MULTIFAN-CL project. Other developments may be assigned to subsequent years. These are outlined in the following list in respect of the suggested period for their implementation.

<b>Peer review recommendation</b>	<b>Implementation</b>
l. Tag likelihood relative weighting	2014-15
d. Tags inform movement	2014-15
m. Fit the BH-SRR to annual recruitments	2014-15
b. Non-uniform size bins	2014-15
i. Tail compression	2014-15
h. Selectivity ogives	2014-15
k. Maturity at length	2015+
c. Long-term tag loss	2015+
<b>Other new feature</b>	<b>Implementation</b>
Seasonal selectivity	2014-15
Tags inform growth	2015+

# 1 INTRODUCTION

MULTIFAN-CL is a statistical, age-structured, length-based model routinely used for stock assessments of tuna and other pelagic species. The model was originally developed by Dave Fournier of Otter Research Ltd for application to south Pacific albacore tuna (Fournier et al. 1998).

MFCL is typically fitted to total catch, catch rate, size-frequency and tagging data stratified by fishery, region and time period. Recent tropical tuna assessments (e.g. Davies et al. 2014; Harley et al. 2014, and Rice et al. 2014) encompass a time period of 1952 or 1972 to 2012 in quarterly time steps, and model multiple separate fisheries occurring in 5 to 9 spatial regions. The main parameters estimated by the model include initial numbers-at-age in each region (usually constrained by an equilibrium age-structure assumption), the number in age class 1 for each quarter in each region (the recruitment), growth parameters, natural mortality-at-age (if estimated), selectivity-at-age by fishery (constrained by smoothing penalties or splines), catch (unless using the catch-conditioned catch equation), effort deviations (random variations in the effort-fishing mortality relationship) for each fishery, initial catchability and catchability deviations (cumulative changes in catchability with time) for each fishery (if estimated). Parameters are estimated by fitting to a composite likelihood comprised of the fits to the data and penalized likelihood distributions for various parameters.

Each year the MFCL development team works to improve the model to accommodate changes in understanding of the fishery, to fix software errors, and to improve model features and usability. This document records changes made since August 2013 to the model and other components of the MFCL project both for the current release version (1.1.5.6), and the current unreleased development version, and updates the report for the previous period, 2012-13, (Davies et al. 2013).

## 2 DEVELOPMENT OVERVIEW

### 2.1 Team

The senior developer of MFCL is Dave Fournier, of Otter Software in Canada. Assisting with programming is Nick Davies, with occasional programming carried out by John Hampton (SPC, New Caledonia), and Pierre Kleiber (NMFS Hawaii). Other tasks include testing and debugging (ND, SDH, PK, JH, and Fabrice Bouye (SPC)); documentation (PK, ND); and planning and coordination (DF, ND, JH, and Shelton J Harley). Related project software are developed or managed by FB (MFCL Viewer, Condor, Gforge), PK (R scripts), ND and SH (R4MFCL, Condor).

### 2.2 Calendar

August – November: Testing, planning and ongoing code development, Developer’s workshop

December – February: Testing and ongoing code development

March: Developer’s workshop

April-July: Testing, ongoing code development and support for stock assessments

### 2.3 Collaboration and versioning

The project management website based on the open source GForge software was established in 2008-09 has been maintained and provides the nucleus for source code management and versioning. The repository for MFCL source code development is held on the website and uses the open source software SVN (<http://tortoisesvn.net/>). Code developments are consecutively committed to the repository while tracing the different versions chronologically. The repository and overall development are coordinated via the GForge website <http://gforge2.spc.int/>, which is administered by Fabrice Bouye [fabriceb@spc.int](mailto:fabriceb@spc.int).

Problems with MFCL operation or compilation have been reported to the project management website so as to maintain a list of desired enhancements, and to allocate tasks among the project team. Some of the tasks identified during the previous reporting period (2012-13) have been addressed in the current period in the way of model developments. A main trunk exists for the MFCL source code, and a development branch has been created to hold these recent developments to the source currently being developed and tested. A formal testing procedure has been designed before source code is committed

from the development branch to the trunk, and a manual for the testing of new compilations, standardizing the source code compilation procedure, and posting of executables has been drafted.

A version of the source code for ADMB (<http://admb-project.org/>) has been added to the project management website in a separate repository. Minor modifications were required to the ADMB source (currently held in a development branch) to facilitate the recent MFCL developments.

## 2.4 Developer's workshops

Developer's workshops were held at North Saanich, British Columbia, Canada, 29 August – 6 September 2013, and at Whangarei, New Zealand, 3 to 12 March 2014. The participants were the primary developer Dr Dave Fournier and Mr Nick Davies.

The main areas of development during the first workshop were:

- Multi-species/sex model development
- Time-varying selectivities
- A likelihood function for conditional age-at-length data

The main areas of development during the second workshop were:

- Logistic-normal likelihood for size-composition data
- Assumed initial conditions such that  $B_{init}$  approximates  $B_0$
- MinGW compilation for producing executables for Windows OS
- Region-specific yields

The specific details of these developments are provided in Section 3.0.

## 2.5 Postings to website

There have been two postings of MFCL updates to the website since August 2013.

### 2.5.1 7 November 2013 – version 1.1.5.5

Fixes were made to the zero fishing mortality scenario (when `fish_flags[55]` is set to 1) for the calls to routines that generate recruitment predictions from the Beverton-Holt stock recruitment relationship and for the calculation of average recruitment given the period defined.

### 2.5.2 14 February 2014 – version 1.1.5.6

A minor modification was made to this re-build version that enables non-integer tag release length frequency data to be input in the `.tag` file.

## 2.6 Independent Peer Review of the 2012 bigeye tuna stock assessment

An outcome of an independent peer review of the 2011 bigeye tuna stock assessment (Ianelli et al. 2012) was a set of recommendations for improvements and developments to the MFCL software. These aim not only to improve the software's application in the context of the bigeye assessment specifically, but also its stock assessment application more generally.

From this list, a subset of the recommendations were identified for the developments to be undertaken for 2013:

- The implementation of a 'multi-sex' version of MULTIFAN-CL, with application to example multi-sex stock assessments. (recommendation "f")
- Completion of the development and implementation for time-varying selectivities (recommendation "a")
- Completion of the development and implementation for conditional age-at-length data to be included in the likelihood function. (recommendation "e")

- Completion of the development and implementation of an output table which lists all of the likelihood components by fleet and automates the process of computing effective sample sizes (and other summary statistics related to the model fit). (recommendation "g")
- Completion of the development and implementation for assuming a multinomial likelihood for the compositional data in the first phases and only transition to the robust normal likelihood in the later phases. (recommendation "j")

Completing these developments has comprised the majority of work done between August 2013 and July 2014. Substantial progress has been made towards these tasks and is reported in Section 3.0 of this report. Work towards the recommendations proposed for 2014-15 is outlined in Section 6.

## 2.7 Tool development

### 2.7.1 R4MFCL

The R scripts for working with MFCL, developed and released on the internet (<http://code.google.com/p/r4mfcl/>), have been updated to adapt to the recent MFCL release version file formats. These scripts are used to manipulate the input files, so that submitting model runs can be automated from R. Other scripts can be used to read in the output files, analyze the results, and generate plots and tables. Further refinements of these tools were undertaken as part of the 2014 stock assessments that consolidated new features to the utilities package. The repository was updated in 2013 for the current and development versions of the package.

### 2.7.2 Testing framework

The testing framework for MFCL compilations first developed in 2011-12, was applied extensively during 2013-14. This framework ensures the repeatability and traceability of testing by streamlining the process for new source code developments through a system of model testing procedures and folder locations. The testing criterion is based upon pair-wise comparisons of model run results obtained using an existing MFCL compilation (usually the current release version) versus those from a development compilation. Tests are undertaken over multiple processor platforms and architectures, with application to multiple input testing data sets, and with various options for the MFCL operation, viz. single or multiple model evaluations, or full do it all model fits to convergence. This ensures a thorough integrity-check of model quantities and components of the objective function prior to the distribution of new versions.

Since March 2013 MULTIFAN-CL source code has undergone substantial developments for multi-threading of the tagged population calculations, combined tagging observations for multiple species/sexes, multi-sex equilibrium yield calculation, time-variant selectivities for individual fisheries, and adding a likelihood term for age-length composition data. These developments are described in Section 3.0.

In adding these new features to the development version, regular testing of this versus the release version was undertaken to ensure the integrity of existing operations. Known as "benchmark tests" these are summarized in Annex 9.1. The development version was tested in March 2013 versus the release version, which defines the development version as the **benchmark** source code. Subsequent development versions were then tested relative to the benchmark to establish their integrity, after which they are defined as the new benchmark development version. The recent developments were validated using two levels of tests.

#### 1. Establish the accepted development version

The first level of testing ensures the integrity of existing model features by undertaking tests using single-species data for ALB2012, BET2011, YFT2011, SKJ2011, STM2012, and SWO2013 to conclude that single model evaluations and the fitted solutions were sufficiently close to regard the development version estimates essentially similar to the benchmark version. This indicates integrity of the development version for undertaking single-species model evaluations. The benchmark tests undertaken in 2013-14 is summarized in Annex 9.1. with the most recent being done in May 2014.

Tests using multi-species data disaggregated among species were done which entails comparing the fitted solutions of the development version code versus those solutions obtained using the corresponding data for each species fitted individually (Figure 1). These tests concluded that the operations applying to each population in the disaggregated model have integrity and effectively emulate the solutions obtained when each population is modelled individually.

These tests concluded the development version has integrity for single and multi-species operations where disaggregated data are used, and can therefore be used as the benchmark development version of the MULTIFAN-CL source code.

## 2. Establishing integrity of new features

This second level of testing entails a detailed examination of new features. The inputs and model configuration are customized for the new features and the operation of the new algorithms are evaluated in respect of the original formulations. During 2013-14 extensive testing was done for including catch, length- and weight-frequency data aggregated among species/sexes used in model calculations to ensure that these data are used correctly in the likelihood calculations and produced the expected results. Similarly, tests of tagging release and recapture data aggregated among species/sexes used in model calculations were done. Further details of these tests are provided in Section 3.0 in respect of each new feature.

### 2.7.3 Viewer

A development version of the MFCL viewer that can display the results of a multi-species or multi-sex application was updated as new output was added to the report files. Development of this version is ongoing since the final output format of the multi-species/sex application is not yet complete.

The release version of the viewer was updated to:

- Better display graphs in respect of the real year-quarter within the model calculation period
- Display plot.rep results for solutions lacking yield estimates (useful for examining initial model runs)

### 2.7.4 Condor parallel processing facility

The Condor ([www.condor.wisc.edu](http://www.condor.wisc.edu)) facility has been used routinely for managing multiple MFCL model runs on a grid currently numbering over 100 processors; being windows or Linux platforms, and either 32- or 64-bit architecture. This grid enables intensive model runs for: testing MFCL development versions prior to release; undertaking stock assessments that entail multiple model runs (e.g. sensitivity analyses), structural uncertainty analyses, and management strategy evaluation. The Condor version used by SPC-OFP was recently updated in 2014.

## 2.8 User's guide

A revision has been completed that documents the developments in version 1.1.5.6 and this has been posted on the <http://www.multifan-cl.org/> website.

## 3 NEW FEATURES

All new features that have been implemented in the MULTIFAN-CL source code have been added firstly to the development version. Once these features have been tested for their integrity, with no impacts on existing features, then the development version is merged into the release version of the code. The current release version is 1.1.5.6 and it is proposed to merge many of the new features described below into this version during 2014-15. The new features added to the development version are listed in respect of the status of completion in Table 6 and are described in the following sub-sections.

### 3.1 Sex-disaggregation

#### 3.1.1 Rationale

Ianelli et al. (2012) recommendation f. was to "Extend MFCL to allow gender to be explicitly represented. This will allow the impacts of differences in growth and natural mortality between the sexes to be represented. The current approach to modeling, for example, length-specific natural mortality (e.g.

WCPFC-SC4-2008/ ME-WP-1) seems unnecessarily complicated given the lack of gender-structure in the model”.

This is a substantial development to the model since it changes the fundamental structure by expanding the dimensions at the regional level and among all parameters. It has important benefits for modeling Pacific tuna and billfish populations which often have sex-specific growth rates, or regional differences in growth (Nicol et al. 2011) by reducing process error caused by assuming single parameters for both sexes. Parameters specific to sex, species, or stocks may be accommodated and, if required, shared among selected dimensions. This may reduce the number of parameters required in fitting models for species that share fisheries having the same characteristics.

### 3.1.2 Methods and Testing

For this development, test data sets were created for bigeye and yellowfin for 6 regions, using a reduced model time period (1990-2009) so as to facilitate rapid computation while yet retaining the full extent of parameters required for the large tuna assessment models. The input “*frq*” file format was changed to integrate multiple sex/spp/stock catch, effort and size composition data within one file. Source code has been adapted for the increased dimensions and model fits to multi-species data have been completed.

#### 3.1.2.1 *Tagging data shared over multiple species/sexes*

This was one of the most substantial remaining developments for the multi-sex model. It takes account of how lengths of tagged fish are converted to ages and may include a new feature to condition the probabilities of age-at-length by the model predictions of catch-at-age for the fishing method used to catch, tag and release fish. Currently the model assigns tagged fish to the two age classes adjacent to that of the corresponding mean length-at-age. It was considered that this current approach be retained for the development version, and the improved approach was noted for a later implementation.

Tagging data (releases and recaptures) for a given release event may be disaggregated or aggregated among the sexes. Whereas in the case of the observations being disaggregated and the predictions for each sex are fitted separately, for the case where observations are aggregated, the predictions for the relevant stratum must first be aggregated before calculating the likelihood.

A new format for the \*.tag input file was drafted to include species flags in the header record for each tag release group. Numerous new pointers were created to identify tag release groups for which the sex was not known and the tag\_flags were modified to enable duplication of the release groups that are shared among sexes. The algorithm for analyzing combined tagging data was as follows:

- Observations were duplicated over both sexes
- The initial release tagged population was apportioned over both sexes according to the predicted region-specific sex ratio.
- Tagged population dynamics were applied as normal
- Tag release groups that are combined over both sexes are identified and the predicted recaptures for each sex are aggregated before fitting to the observed tag recaptures that are combined over both sexes.

A proof of concept source code was completed and targeted features were checked. A single release group within the 68 in the multi-sex data example being used was assigned to be combined over both sexes. Testing the operation of this feature entails two components: deterministic comparison, and fitted comparison.

The solution .par from the sex-disaggregated model with all observations disaggregated among sexes is used to produce model quantities from a first deterministic single-evaluation. Sex-specific predictions for tag recaptures were output for two sex-specific release events. A second deterministic single-evaluation of the model using the identical .par is done, while aggregating the observations for selected tag release events among both sexes. The specific fishing incidents and tag release events are identified and the sum of the observations are recorded. The deterministic comparison of the predictions among the two evaluations illustrated the integrity of the code operation.

For the comparison of fitted models, firstly the quantities from the sex-disaggregated model (denoted **mod1**), with all observations disaggregated among sexes was used to produce a set of predictions for the specific fishing incidents and tag release events as identified and output from the deterministic comparison.

The selected observations for the tag release events aggregated among both sexes for were then used in re-fitting the model, where some are disaggregated, and some aggregated. The solution of this revised model fit is denoted, **modagg**. Due to the change in the likelihoods for each data type, it is likely that the solution .par will differ somewhat with respect to mod1, however the differences can be attributed to the aggregated observations being simply the summation of those input to mod1.

The fitted model's predictions for the aggregated release group was relatively similar to the sum of the predictions for the model fitted to the disaggregated data; 723.5 versus 755.5, respectively (Table 1). Despite the differences in the predictions, the recaptures are predicted to occur over similar time periods for both models (Figure 2). A higher overall biomass is predicted by the model fitted to the aggregated data, and this might be attributed to the lower number of total recaptures predicted. The main differences among the models that may explain the difference in predicted total biomass were higher absolute recruitments (while relative recruitment patterns remain very similar), and a substantial shift in the relative proportions of recruitments among regions such that a higher proportion was assigned to region 3, that being the region in which tag groups 18 and 56 were released. A marginally worse total likelihood (by 0.07%) was obtained for the model fitted to aggregated data.

The tests of the models using either disaggregated or aggregated tagging data illustrated the integrity of the operation of the new feature to accommodate tagging data aggregated among sexes.

### 3.1.2.2 *Catch data shared over multiple species/sexes*

The format of the input.frq file (version 7) accommodates more than one species or sex by separating the input data for each on separate rows. This format was adapted further in a new version (8) that includes two additional columns to denote which of the species/sexes are included in a row of fisheries data. This facilitates input of combined data as follows:

- A row must be entered for each species/sex even if the data are combined for both, in which case the first row will be used for entering the data. The second row is used for creating the "region" for the second species/sex and the extra fishery for the second, or more, species/sexes.
- A "1" in the 5th and 6th columns denote which species/sex the data relates to.
- A "1" in the 7th and 8th columns denote whether the data are combined for both species/sexes, i.e. if both columns are "1" this indicates catches are combined.
- Note that the effort in the duplicated rows is identical since this is the same fishing incident, but the catch may be different if it is dis-aggregated with respect to species/sex.

The method for testing the input and use of catch data combined over both sexes was the same as applied for combined tagging data (Section 3.1.2.1) with the comparisons being made between the observed and predicted catches for both species/sexes (i.e. the corresponding fisheries in the 1st and 2nd groups, for sex 1 and 2 respectively) for the model fitted with all data disaggregated over both species/sexes (**mod1**) and for the model fitted with the data aggregated (**modagg**). The data for making this comparison were taken from the plot.rep output report sections:

```
# Observed catch by fishery (down) and time (across)
```

```
# Predicted catch by fishery (down) and time (across)
```

Given the positive result obtained for the deterministic test, only the catches for the fishing event having aggregated data were compared for the fitted comparison.

For **modagg** the observed catches for fisheries 1 and 14 were the same for both species, being the aggregate value for each, whereas for **mod1** the observed catches remain disaggregated for the associated fishing incidents (Figure 3). For **modagg** the predicted catches for these events remain very similar to those of **mod1** which confirms that during the fitting procedure the predictions are aggregated among the species when the likelihood is calculated relative to the aggregated catch. It is noted there was

a slight change in the **modagg** predicted value for Fishery 1 (catch in numbers) compared to that of **mod1** which indicates the effect of reducing the information available for the model fit.

### 3.1.2.3 Length- and weight-frequency data shared over multiple species/sexes

The method for input of length- and weight-frequency data combined over species or sexes is the same as that for catch data (Section 3.1.2.3) and the testing of the model fit to these combined data was the same applied for tagging and catch data.

Comparisons were made of the observed and predicted length-frequencies for both species/sexes (i.e. fisheries in the 1st and 2nd groups) for the model fitted with all data disaggregated over both species/sexes (**mod1**) and for the model fitted with the data aggregated (**modagg**). The data for making this comparison were taken from the length.fit output report. Given the positive result for the deterministic test, only the length-frequencies for the fishing event having aggregated data were compared. Despite having observations aggregated among species/sexes, the predictions of the **modagg** model (denoted CLFagg) were very similar to that of **mod1** (Figure 4).

The same comparisons were made of the observed and predicted weight-frequencies for both species/sexes (i.e. fisheries in the 1st and 2nd groups) for the model fitted with all data disaggregated and for the model fitted with the data aggregated. The data for making this comparison were taken from the weight.fit output report. Despite having observations aggregated among species/sexes, the predictions of the **modagg** model (denoted CWFagg) were very similar to that of **mod1** (Figure 5).

### 3.1.2.4 Single equilibrium yield over multiple sexes

The algorithm for calculating equilibrium yield quantities using the BH-SRR was adapted for doing a multi-sex yield calculation. Firstly, the biomass was summated over the regions corresponding to each sex. Secondly, the Newton-Raphson procedure that estimates  $F_{MSY}$  and returns the equilibrium yields for a given range of total mortalities was modified for the circumstance of multiple sexes. Thirdly, the BH-SRR for only the female sex was applied in predicting the equilibrium recruitments, with an assumed recruitment sex ratio determining recruitments to the male population. The routine get\_yield\_at\_multiplier2() was modified so as to calculate the equilibrium catch for both sexes given the value of lambda, and these catches were summed. This calculation was tested using a deterministic comparison with a multi-species model having identical species for which a converged solution has previously been obtained (**mod1**), where equilibrium yields are calculated separately for each species. This model was then configured for multiple sexes, such that each sex is identical (**mod2**), but the equilibrium yield is calculated among the sexes, and is expected to be exactly double that of each species in the multi-species model. This was achieved by duplicating all parameters in the .par for one of the species, replicating the data for one of the species in the .frq, and the specifications in the .ini.

A fitted model was obtained using the structural assumptions for the multi-sex model, **mod2**, having identical data for each sex, and disaggregated in respect of sex. The fitted solution was denoted **mod2\_fit**. This was compared relative to **mod2**, and any differences might be attributable to the effects of the penalty term on the BH-SRR for the multi-sex model being somewhat different to that of the multi-species model.

For the deterministic comparison the multi-species model **mod1** produced identical equilibrium yields for each species as expected given the construction of the input .par such that identical parameters were used for each species. A deterministic single evaluation of the multi-sex model (**mod2**) generated an equilibrium yield function that was exactly double that of each species of the multi-species model and the MSY was consequently exactly double (Table 2). All other parameters including the recruitments and BH-SRR were identical to that of the multi-species model. This illustrates the correct operation of the equilibrium yield calculation for a multi-sex model where the BH-SRR is used for the female sex only in deriving recruitments for both sexes, and the yield is the sum of the equilibrium catches among the sexes. It was noted that  $F_{MSY}$  for the multi-sex model was double that of the multi-species model. This is attributable to the multi-sex yield calculation where the equilibrium biomass relates only to the females; thus, the equilibrium fishing mortality is expressed only in respect of this sex.

The fitted solution for **mod2\_fit** is very similar to that of the deterministically derived model **mod2** (Table 3, Figure 6). This can be expected given that the data for each sex and the structure and

assumptions were identical. This test demonstrates that the BH-SRR is being calculated correctly during the model optimization, and the solution obtained with the penalty term calculated for the female sex only, is very similar to that of a multi-species model where BH-SRR penalties are derived for both species. This is not surprising given the low relative weight of this penalty term.

#### 3.1.2.5 *-makepar operation for multiple sexes*

The multi-sex implementation entails the input of the new flags: `region_flags[3]` and `species_flags[]` which are reported in the `.par` file. A `-makepar` run was completed and the `00.par` was modified to include values for these new flags. A first phase of a `doitall` was run and operates correctly which confirms the model will perform a full fitting procedure. It will be necessary to create a means of input for these new flags using a modified version of the `.ini` file.

## 3.2 Time-variant selectivities

### 3.2.1 Rationale

Ianelli et al. (2012) recommendation a. was to “Test the options for time-varying selectivity – allowing for time-varying selectivity may address some of the issues related to the sometimes poor fits to the length- and weight-frequency data”. The intention of this is for fishery-specific selectivity to vary through time in a manner which may be termed as “time-blocks”, having shifts in the selectivity patterns due to changes in fishing practice, perhaps due to altered target species. The current kludge used for this is to split the time period within a fishery to create two, or more, new fisheries. For each of the “time-split” fisheries, individual selectivities must be estimated. This is unduly complex, and the improvement as recommended is to estimate time-variant selectivity for a given fishery.

### 3.2.2 Methods and testing

An algorithm was developed for sharing selectivity parameters over specified “blocked” periods of a fishery’s realizations. The `sel_dev_coffs` were used to store the deviations estimated for each time-block that were applied to the selectivity parameters for the first time-block of that fishery. Numerous pointers were developed to indicate the number of breaks, the blocked fisheries, and a new input file holding the first year of each time-block. `fish_flags[71]` was assigned to supply the number of breaks in the time-blocked fishery, and which activates the time-block option. This feature needed to take account of the fishery grouping in respect of the selectivities and was consequently complex. The algorithm was applied to spline and logistic selectivity functions, with only the preliminary stages of its application to double-normal and age-specific (no functional form) selectivities being developed.

The algorithm entailed estimating selectivity parameters for the first time period defined (the initial time-block), and for each subsequent time-block, a set of deviates were estimated and applied to the initial time-block’s parameters to generate a separate selectivity pattern. Although the assumed functional form for the fishery’s selectivity is constant over the model calculation period, the parameters alter among the time-blocks by means of the estimated deviates that change the shape of the form in each time-block subsequent to the initial time-block. The estimated selectivity parameters for the initial time-block are held (as normal) in `selcoff`, and the deviates from the initial time-block’s selectivity parameters are held in `sel_dev_coffs`.

A new input file was constructed “`selblocks.dat`” having the format where the years for the first year of each time-block, subsequent to the initial time-block, are listed as a vector on each row for the relevant fisheries. A new output file was created called “`blocked_fisheries_report`” by routine `ests_write1` (`plot.cpp`) conditional upon `sum_ff71_flag`, having a format in sections for each fishery with time-variant selectivity being labelled, and within each section a row for the selectivity-at-age within each time-block, including the initial time-block.

The proof of concept code has been checked and the selectivities were found to be equal among grouped fisheries and shared correctly among time-blocks. The design for tests has been prepared using the albacore tuna model which has 148 fisheries of which 72 are fisheries that have been split into separate periods to create time blocks. These will be reduced to 32 fisheries when time-variant parameters are estimated.

### 3.3 Fit to conditional age-length data

#### 3.3.1 Rationale

Ianelli et al. (2012) recommendation e. was to “Allow conditional age-at-length data to be included in the likelihood function. This will allow the ageing data from current sampling (e.g. WCPFC-SC6-2010/GN IP-04) to be formally included in the assessment”. Observations of fish age-at-length provides direct information to the model for estimating the growth function parameters describing mean length-at-age. Growth estimates, e.g. length at maximum age are uncertain in some tuna assessments, e.g. bigeye (Harley et al. 2014), and this has significance for the model’s productivity estimates.

#### 3.3.2 Methods and Testing

To allow the ageing data from current sampling to be formally included in the assessment a negative log-likelihood formulation for these observations was drafted, being the sum over all the length intervals, of the multinomial function in respect of the predicted proportion of fish  $\rho_{ij}$  having a particular age  $j$  given it has a length  $i$ ,

$$\rho_{ij} = \frac{p_j q_{ij}}{\sum_k p_k q_{ik}}$$

A new class was added to the project to contain the new members needed for this development and was made conditional upon `parest_flags[240]` to activate this feature, and the new `age_length` likelihood term has been added. A test data set of observations of fish ages at length was prepared based upon the growth function for species 1, and was input to the proof of concept code. While the code appears to run well with the new input data, data structures, and plausible likelihood values are calculated, further checks and comprehensive testing are required. A test data set comprised of bigeye age-length observations has been compiled.

### 3.4 Multinomial distribution for size composition data

#### 3.4.1 Rationale

Ianelli et al. (2012) recommendation j. was to “Add an option which allows the analyst to assume a multinomial likelihood for the compositional data in the first phases and only transition to the robust normal likelihood in the later phases”. For the release version the generally assumed error distribution available for fitting length and weight composition data is the robust normal likelihood, with this option being specified using `parest_flags[141]`.

#### 3.4.2 Methods and Testing

This feature is in the preliminary stages of development with an option added for the size composition likelihood specified by `parest_flags(161) = 6` that applies the multinomial distribution. The multinomial likelihood has been formulated and added to the development version code. This feature is yet to be fully documented and tested.

### 3.5 Likelihood component table

#### 3.5.1 Rationale

Ianelli et al. (2012) recommendation g. was to “Create an output table which lists all of the likelihood components by fleet and automates the process of computing effective samples sizes (and other summary statistics related to model fit)”. This is a powerful diagnostic tool for model development that facilitates the examination of the effects of alternative relative weightings for the input data types and to display conflicts among these data.

#### 3.5.2 Methods and Testing

The release version currently reports values for likelihoods and penalties in screen output and report files, and this feature aims to report them in a single tabulated report. A routine `print_likelihood_components()` has been added that generates a report file “`test_plot_output`” with the following content:

- total length component of likelihood for each fishery: by fishery and realisation for each

- total weight component of likelihood for each fishery: by fishery and realisation for each
- total catch components of likelihood: by fishery and realisation for each
- tag likelihood by tag release by fishery groups: by release group and fishery grouping for each
- BH-SRR penalty
- Effort deviates penalty
- Catchability deviates penalties by fishery and by group

The content of the table is yet to be tested and a function must be added to R4MFCL in order to summarise the detailed report produced into various levels as may be required for the model diagnostics.

### 3.6 Fit the BH-SRR to annual recruitments

#### 3.6.1 Rationale

Ianelli et al. (2012) recommendation **m.** was for “Annual recruits for BH-SRR. Consider fitting the stock-recruitment relationship to the annual rather than seasonal recruitments”. Many tuna models assume a quarterly temporal stratification which frequently results in high variability in the estimated recruitments, most likely due to the seasonality in the underlying pattern of recruitments. This high variability may unduly affect the fit of the BH-SRR and it may be preferable to fit the relationship to the annual recruitments.

#### 3.6.2 Methods and Testing

Within the algorithm for constraining the initial biomass ( $B_{init}$ ) to approximate the unexploited equilibrium biomass ( $B_0$ ), see Section 3.7), the routine for fitting the BH-SRR has been adapted for the option of assigning average annual biomass to the x-variable and total annual recruitments to the y-variable. These assignments are made conditional upon `age_flags(94)==3` and consequently the BH-SRR parameters are estimated in respect of the average annual values. This facility may be readily adapted to any other assumption for the initial population conditions and to be generic for assuming an annual BH-SRR in calculating equilibrium quantities. Completing the development of this feature for generic application is proposed for 2014-15.

### 3.7 Initial biomass approximates the equilibrium unexploited biomass

#### 3.7.1 Rationale

In the classic biomass depletion model, the dynamic population model biomass trajectory commences from an abundance that is equivalent to that estimated using the estimated spawner stock-recruitment relationship (BH-SRR) in an unexploited equilibrium state ( $B_0$ ). For some tuna assessments (e.g. yellowfin – Davies et al. 2014, and bigeye – Harley et al. 2014) the initial biomass is substantially higher than  $B_0$  to a level that may be considered counter-intuitive or results from initial recruitments that are implausibly high. A feature that constrains the initial population conditions to approximate  $B_0$  (denoted as  $B_{init} \sim B_0$ ) will satisfy the assumptions of the biomass depletion model.

#### 3.7.2 Methods and Testing

The fundamental assumption that the initial population is equivalent to that estimated using the estimated BH-SRR in an unexploited equilibrium state ( $B_0$ ) facilitates solving for the relationship's parameters in respect of the steepness parameter (Annex 9.2). This greatly simplifies the implementation of this feature in that it does not add to the overall likelihood, and the BH-SRR parameters can be solved for analytically.

Before adapting the BH-SRR relationship to the new formulation it was necessary to calculate  $\phi$ , which entails deriving the initial equilibrium population over the first  $n$  recruitments (number of recruitments per year). Several new routines were added to achieve this conditional upon the `age_flags(94)==3`, i.e. the `age_flag` that specifies the initial population condition assumptions. Given that it was unreasonable to specify the initial population on the basis of a single season's or quarter's recruitment, it was necessary to derive the BH-SRR using annualized recruitments. Using the analytical solution for  $\alpha$  and  $\beta$  (Annex 9.2) these parameters were excluded from the vector being estimated.

Preliminary testing has been undertaken using the 2011 yellowfin tuna assessment model which is characterized by  $B_{init}$  being approximately  $1.5B_0$  (spawning biomass). A comparison of the BH-SRR estimates from the model without and with the  $B_{init} \sim B_0$  assumption illustrates the effect of using annualized recruitments and spawning biomass (Figure 7), such that the predicted recruitments are about four times larger than the model using quarterly recruitments. Consequently the equilibrium yield is also four times larger with  $MSY=607,100$  mt versus 134,700 mt. The quarterly recruitments were similar over most of the model calculation period, but in the first year those for the model with the  $B_{init} \sim B_0$  assumption were around the average of the BH-SRR predictions (Figure 8) which can be expected due to the constraint applied. Making the assumption appears to affect estimates of absolute abundance, which were on average lower, and while initial biomass was close to the  $SB_0$  level (0.98), the high early recruitments immediately following the first year result in high biomass levels similar to that of the model where  $B_{init}$  is unconstrained (Figure 9).

The preliminary tests indicate this new feature is operating as formulated, however, formal testing is required to ensure its integrity and that existing features are unaffected. Also the reporting of the estimates derived under the assumption must be appropriately formatted and labelled.

### 3.8 Region-specific yield estimation

#### 3.8.1 Rationale

Some tuna models predict a substantial proportion of total biomass to occur in specific regions, e.g. the 2014 yellowfin model predicts around 70% of biomass is in the western equatorial region. These regions therefore account for a substantial component of the annual productivity and an understanding of the equilibrium productivity derived from such regions is of utility to fisheries managers.

#### 3.8.2 Methods and Testing

The routine for calculating yields specific to each region already present in the current release version of the code, was reviewed and tested in the development version. It is activated by the setting of `age_flags(140)=1`. Rigorous testing was done of the routine `get_equilibrium_age_structure()` that calculates equilibrium numbers at age in each region. The option for the correction for log-normal bias in the BH-SRR predictions was added so as to be consistent with the conventional approach that calculates the yield over the aggregated regions. The output to the file `plot.rep` was formatted to report the region-specific equilibrium yields, biomass, and fishing mortalities.

Preliminary testing of the feature was done using the 2011 yellowfin tuna assessment model. For the model where region-specific yields were estimated (spatially dis-aggregated calculation), total  $MSY$  over all regions was around 20% lower than the conventional method, (spatially-aggregated). Total equilibrium biomass at  $MSY$  was around 10-16% higher, and the  $MSY$  fishing mortality level was about 27% lower (Table 4). While appearing counter-intuitive at first, this result reflects the differences in the equilibrium yield calculations among the approaches. The spatially-aggregated method finds the optimum yield for a single population, with a given fishing mortality-at-age. Consequently, the productivity of the entire population age-structure is optimised in respect of the total average fishing mortality-at-age that produces the  $MSY$ . In contrast, the spatially-disaggregated method applies a common fishing mortality scalar among all regions upon the **region-specific** fishing mortality-at-age, and so the maximum yield is found over all regions (sum of region-specific yields). Consequently, the aggregated maximum yield is unlikely to be optimised for each region. For example, say there are two regions and the fish mortality is already at  $MSY$  for region 1, but is only at 10% of  $MSY$  for region 2. Then it might be that one can raise both by say 5% and increase the combined sustainable yield, with the disadvantage of overfishing region 1. Hence, the spatially-disaggregated approach produces a lower estimate, given some regions are at a suboptimal level. This is illustrated when comparing the region-specific yield curves generated by the spatially dis-aggregated method on the example data set (Figure 10).

Further testing of this approach is required to ensure the integrity of the formulations as implemented, and possibly the spatially dis-aggregated method may be refined, e.g. to apply the fishing mortality multiplier specifically to each region when optimizing the yield for each.

## 3.9 Fixed terminal recruitment deviates

### 3.9.1 Rationale

It is common for recruitment estimates in the terminal model time periods to be poorly estimated due to the low number of observations for recent cohorts. This has been investigated using retrospective analysis where data from the terminal time periods (the last three years) were successively removed and the model fitted to each case (Davies et al. 2014). The terminal recruitments and biomass estimates were compared among the retrospective models for their robustness to the loss of data, and were found to be uncertain. Excluding the estimation of the terminal temporal recruitment deviates in the most recent model time periods is therefore advantageous in reducing process error in current or latest model quantities of interest to managers, e.g. *B<sub>latest</sub>*.

### 3.9.2 Methods and Testing

This feature is controlled using `parest_flags(400)` which is set to the number of model time periods from the last period that are excluded from the estimation of temporal recruitment deviates. In respect of the recruitments from the specified time periods, the operation of this flag:

- Excludes them from the calculation of mean recruitment;
- Excludes them from the vector of estimated parameters;
- Sets the temporal deviates to zero.

This feature was tested using the yellowfin assessment for 2014 and the effect is clear in a comparison of recruitments between models with and without the feature (Figure 11).

## 3.10 Logistic-normal size composition likelihood

### 3.10.1 Rationale

Francis (2014) suggests that the multinomial likelihood performs worse than the logistic normal when applied to size-composition data. To further explore this suggestion for its potential improvement of tuna stock assessments, an option was added for the logistic-normal likelihood to the development version of MULTIFAN-CL.

### 3.10.2 Methods and Testing

As parameterized by Francis (2014) the logistic-normal likelihood entails 5 parameters and these are activated by `parest_flags(290 to 298)` for length-frequency data and `parest_flags(280 to 288)` for weight-frequency data (Table 5). An enhancement to the published formulation was also made that adds robustness, and this was achieved by assuming a multivariate student-t probability density function.

Only preliminary testing of this new feature has been done using a bigeye tuna data set; including estimating the relative weighting, and tests over a range of robustifying constants. Further testing of this new feature is proposed in 2014-15.

## 3.11 Exclude reporting rates from tag predictions during mixing period

### 3.11.1 Rationale

Instances are possible where large numbers of recaptures are made immediately following a tag release event such that the reporting of these recaptures occurs with minimal error due to physical loss or careless handling of the recovered tags. In such cases, it may be preferable not to account for the reporting rate when solving the Newton-Raphson catch algorithm for tags recaptured during the mixing period.

### 3.11.2 Methods and Testing

This feature was implemented using `tag_flags(2)==1` that is specific to particular tag-release groups and it excludes the reporting rate when calculating the predicted recaptures in the tag mixing period for input to the Newton-Raphson catch estimation routine. The flexibility was added in respect of the type of reporting rate being estimated, i.e. being either fishery-specific, or both fishery- and tag release group-specific.

Only preliminary tests of this feature have been done to date using the tuna data sets available for the 2014 assessments; with further testing is proposed in 2014-15.

## **4 OTHER ENHANCEMENTS AND BUG FIXES**

### **4.1 Correction to fishing mortality calculation**

The routine `fishing_mortality_calc()` was corrected for a potential error which could occur in certain instances relating to the fishery subscript for a temporary storage variable. The routine was corrected by taking the address of `catchability(ir,ip)` instead because this returns a reference. This correction produced no effects upon existing features or model estimates obtained using the full suite of test data in the testing framework.

### **4.2 Tag diagnostics report**

For the current release version, two output files are produced (`fish-rep-rate.dat`; `fish-rep-rate2.dat`) which report the fit to the tagging observations obtained when the estimated reporting rates are specific to both tag release groups and individual fisheries. Modifications were made to the development version for this to be reported in the standard report file `plot.rep`; it is flexible in respect of the type of reporting rates being estimated; and, accommodates the option for the reporting rates being excluded from the catch calculation during the mixing period (see Section 3.1.1).

### **4.3 Correction to using annual biomass in the BH-SRR fit**

An error was identified in the development version relating to the BH-SRR fit estimated using annual biomass and recruitments. This occurred for instances where a single annual recruitment is estimated (e.g. striped marlin and albacore models). The arithmetic exception occurred (divide by zero) because a conditional `if`-block was satisfied within a `for`-loop when close to the end of the model period and the spawner-recruit lag period was assumed  $> 0$ . The logic was made robust in the case of models having annual recruitments, and where biomass is already annual.

### **4.4 Non-integer tag releases**

In order to accommodate pre-processing of length frequency data of tagged fish released, such as when correcting for initial release mortality (Berger et al. 2014), this variable was altered from an integer to a real number. Using the development version, the following objects were declared as real double numbers:

```
dmatrix  initial_tag_release_by_length
dvector& itr1=initial_tag_release_by_length(itr)
```

The 2011 skipjack assessment model was used for testing this change since it includes 196 tag release groups and may readily illustrate the effects of small floating point changes due to transforming the release frequencies from integers to real numbers. Tests included comparing the results (the tag data likelihood values and model quantities) of the release and development versions for:

- A single evaluation using the solution parameter file
- A `doitall` fit solution

In both cases, the test results showed the tag data likelihood values and model quantities were identical between the release version (integers) and the development version (real numbers) illustrating that there was no effect due to changing to a real number for the declaration of the tag release frequencies. On this basis, the code changes were incorporated into the release version 1.1.5.6 subsequently posted (Section 2.5.2).

### **4.5 Increased independent variables**

The tropical tuna models developed during the 2014 assessments have greater spatial disaggregation and more complex fishery definitions, requiring the estimation of around 8700

parameters in the case of yellowfin. Consequently, the maximum dimension of the parameter vector was increased to 10,000 in both the release and development versions.

#### **4.6 Windows compilations**

Producing executables for Windows operating systems may be simplified by compiling directly from the Linux-based project using the MinGW compiler (<http://www.mingw.org/>). The compiler was installed and tested for producing both 32- and 64-bit windows compilations. Further testing of this approach for compiling release executables is proposed for 2014-15 before adopting this as the routine method for creating Windows compatible MULTIFAN-CL executables for release.

### **5 APPLICATION OF NEW FEATURES**

The following new features were employed for the 2014 tropical tuna assessments (Davies et al. 2014, Harley et al. 2014, Rice et al. 2014):

- Increased independent variables
- Non-integer tag releases
- Tag diagnostics report
- Fixed terminal recruitment deviates

Certain other new features were explored during the developmental stages of these assessments but were not incorporated because they had not yet undergone rigorous testing. However, they were of value in exploring particular potential assumptions for the reference case models.

### **6 FUTURE WORK**

A future work plan for the development of new features in MULTIFAN-CL is suggested in Table 7 with those having high priority being undertaken in 2014-15, while others may be addressed in subsequent years. Tasks of immediate priority and relating to the new features recently added (Section 3.0) are discussed in the following section.

### **7 DISCUSSION**

Only two postings of the MULTIFAN-CL release version were made in 2013-14 and these related to relatively minor revisions. This was because the majority of progress made in this time was with the development version to which substantial changes have been made, and does not reflect the overall development of the project.

The key achievement in 2013-14 was the completion of the new feature for dis-aggregating the population by species or sexes. This development, initiated in 2011, has comprised the main focus of the project. The final stages presented in this report were to: allow for instances where catch, size composition and tagging data may be input as dis-aggregated, combined, or both in respect of species or sexes; for the equilibrium yield calculations of the multi-sex model to use a BH-SRR estimated solely for the females with the distribution of recruitments among both sexes; and, the total equilibrium yield to be calculated from the catches of both sexes. Following the successful testing completed to date, this feature may now be applied in an assessment context.

The release and development versions are now substantially different due the large scale of the multi-species/sex development that entailed rigorous testing. It is recommended that high priority be assigned to merging the development version to the repository trunk, and posting this as the next release version. Considering the scale of the proposed future developments (Table 7), such divergence between the release and development versions is unlikely to re-occur in the foreseeable future, and advancing the code with regular release versions will be more streamlined.

Eight of the ten new features completed in 2013-14 (described in Section 3) require testing before they can be released. Many of these have high utility in forthcoming assessments, particularly: time-variant selectivity; fit to conditional age-length data; likelihood component table; and, region-specific yield estimation. This testing is considered to be a high priority before undertaking further new developments so as to ensure these new features are included in the release version available for the 2015 stock assessments.

Of the 13 recommendations made by the independent peer review panel (Ianelli et al. 2012) for developments in MULTIFAN-CL, six have been completed with testing either complete, or to be completed in Aug.-Dec. 2014. This represents substantial progress towards these recommendations, especially given that recommendation **f.** entailed fundamental changes to almost every aspect of the software.

It is suggested that the remaining seven recommendations be addressed largely during 2014-15, with perhaps maturity at length and long-term tag loss (**k.** and **c.**, respectively) having a lower priority and being addressed in the subsequent year. A high priority task, that is outside of the peer review recommendations, is to implement a scheme to estimate seasonal variability in selectivity coefficients (Table 7). It is anticipated that this would substantially improve the south Pacific albacore tuna assessment for which catch compositions are highly seasonal. The current approach for splitting fisheries in respect of season results in complex fisheries definitions (148 fisheries), and seasonal coefficients would substantially reduce the total number of selectivity parameters to be estimated.

Many of the new features result in format changes to the input and output files, or in entirely new reports being produced, e.g. likelihood component table. Unfortunately a lag has occurred in the support tool development required to accommodate these format changes or additional reports. Although some advances have been made with the Viewer's capability for multi-sex model reports, substantial further development is needed for this to be fully operational in applying this model structure in an assessment context. Other examples are: the region-specific yield report which must be added to the Viewer; and, a function to be added to the R4MFCL package that summarizes the likelihood component table. It is suggested that priority be assigned to these advances in the support tools for MULTIFAN-CL to ensure the new features in the soon-to-be posted release version can be fully and effectively utilized by analysts.

## 8 REFERENCES

Berger, A., McKechnie, S., Abascal, F., Kumasi, B., Usu, T., and Nicol, S. 2014. Analysis of tagging data for the 2014 tropical tuna assessments: data quality rules, tagger effects, and reporting rates. WCPFC-SC10-SA-IP-06, Majuro, Republic of the Marshall Islands, 6 – 14 August 2014.

Davies, N., Fournier, D., Hampton, J., Hoyle, S., Bouyé, F., and Harley, S. 2013. Recent developments in the MULTIFAN-CL stock assessment software. WCPFC-SC9-2013/SA-IP-07

Davies, N., Harley, S., Hampton, J. and McKechnie, S. 2014. Stock assessment of yellowfin tuna in the western and central Pacific Ocean. WCPFC-SC10-2014/SA-WP-04

Fournier, D.A., Hampton, J., and Sibert, J.R. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. *Can. J. Fish. Aquat. Sci.* **55**:2105-2116

Francis, R.I.C.C. 2014. Replacing the multinomial in stock assessment models: a first step. *Fisheries Research* **151**: 70-84

Harley, S., Davies, N., Hampton, J., and McKechnie, S. 2014. Stock assessment of bigeye tuna in the western and central Pacific Ocean. WCPFC-SC10-2014/SA-WP-01

Ianelli, J., Maunder, M., and Punt, A. 2012. Independent review of 2011 WCPO bigeye tuna assessment. WCPFC-SC8-SA-WP-01

Nicol, S., Hoyle, S., Farley, J., Muller, B., Retalmai, S., Sisor, K., and Williams, A. 2011. Bigeye tuna age, growth and reproductive biology (project 35). WCPFC-SC7-2011/SA-WP-01

Rice, J., Harley, S., Davies, N., and Hampton, J. 2014. Stock assessment of skipjack tuna in the western and central Pacific Ocean. WCPFC-SC10-2014/SA-WP-05

**Table 1.** Observed and fitted recaptures for a model with tagging disaggregated among species, "Disagg", and a model where release groups 18 and 56 were aggregated, "Tagagg" where release group 50 is the duplicated release group for species 2.

	Disagg		Tagagg	
	Rel_grp 18	Rel_grp 56	Rel_grp 18	Rel_grp 50
Observed	537.9997	254.0003	792.0007	792.0003
Predicted	524.2118	231.2873	723.5487	351.5859

**Table 2.** Equilibrium yield estimates for a multi-species model having identical parameters and structure for each species (mod1) and for a multi-sex model (mod2) having the identical parameters and structure but treating one of the species as the female sex.

	mod1		mod2
	Sex1	Sex2	Sexes 1 and 2
MSY	2.250e+04	2.250e+04	4.500e+04
F <sub>MSY</sub>	4.895e-02	4.895e-02	9.790e-02

**Table 3.** Likelihood, parameters, and equilibrium yield estimates for multi-sex models, one derived deterministically to have identical parameters and structure for each sex (mod2), and the other (mod2\_fit) being the fitted solution for mod2.

	mod2_fit_sex1	mod2_sex1	%dif_sex1_mod2_fit	mod2_fit_sex2	mod2_sex2	%dif_sex2_mod2_fit
av %_mat_age	0.4839	0.4839	0.000	0.4839	0.4839	0.000
totpop	15.0652	15.0675	0.015	15.0652	15.0675	0.015
K	0.0676	0.0669	-1.071	0.0676	0.0669	-1.070
L1	23.3445	23.6042	1.112	23.3447	23.6042	1.112
La	162.6863	162.9192	0.143	162.6866	162.9192	0.143
obj	928748.2369	928748.1961	NA	928748.2369	928748.1961	NA
gradient	0.000802669	0.002995242	NA	0.000802669	0.002995242	NA
alpha	6578000	6569000	-0.137	6578000	6569000	-0.137
beta	23770	23810	0.168	23770	23810	0.168
MSY				45090.00	45000.00	-0.200
F <sub>MSY</sub>				0.09801	0.0979	-0.112

**Table 4.** Comparison of the equilibrium yield quantities for a model calculation over all regions aggregated (spatially-aggregated) with a region-specific model calculation (spatially dis-aggregated).

	Spatially-aggregated	Spatially dis-aggregated	% Diff
MSY	150700	121600	19.31%
F multiplier at MSY	1.236	1.300	-5.18%
F at MSY	0.0934	0.0682	26.94%
Total biomass at MSY	1614000	1782000	-10.41%
Adult biomass at MSY	725900	838700	-15.54%

**Table 5.** Parameters of the logistic-normal likelihood applied to size composition data and the associated MULTIFAN-CL parest\_flags for assigning initial values and activating their estimation.

Parameter	Length-frequency flag	Weight-frequency flag
Log-length variance	290	280
Length-rho	291	281
Log-length degrees of freedom	292	282
Switch for type of probability density function (normal or multivariate student-t)	293	283
Epsilon robustness constant	294	284
Sample size relative weighting	295	285
Activate psi estimation (autoregressive moving average parameter)	296	286
Activate the likelihood calculation for psi	297	287
Activate the Francis' power parameter	298	288

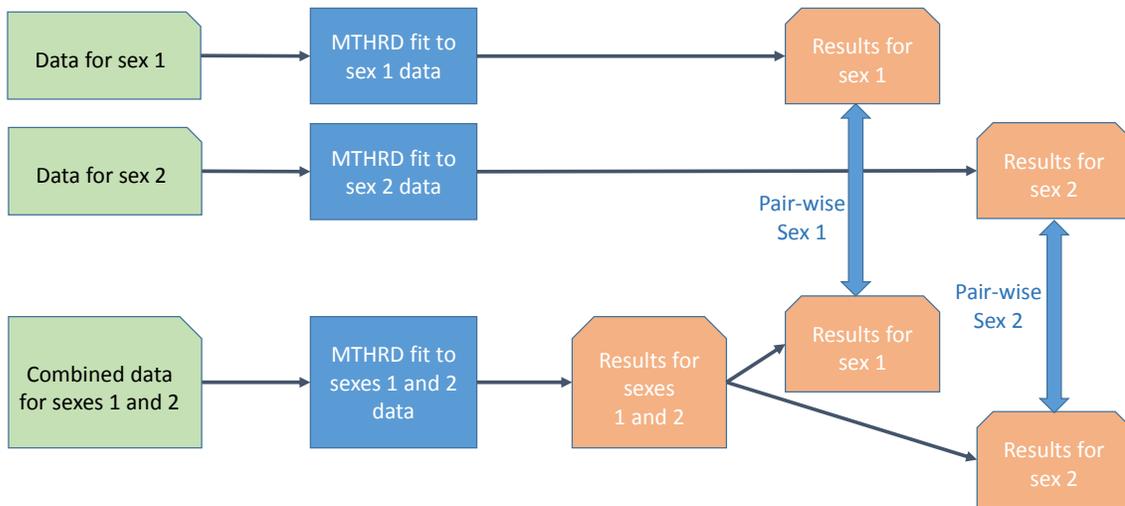
**Table 6.** New features added to MULTIFAN-CL with respect to their state of completion as of August 2014.

<b>Peer review recommendations</b>		
Task	Description	Status of completion
f. Sex-disaggregation	Extend MFCL to allow gender to be explicitly represented.	Development 100%; Testing 100%
a. Time-variant selectivities	Test the options for time-varying selectivity – allowing for time-varying selectivity may address some of the issues related to the sometimes poor fits to the length- and weight-frequency data.	Development 100%; Testing 50%
e. Fit to conditional age-length data	Allow conditional age-at-length data to be included in the likelihood function.	Development 100%; Testing 20%
j. Multinomial distribution for size composition data	Add an option which allows the analyst to assume a multinomial likelihood for the compositional data in the first phases	Development 100%; Testing 0%
g. Likelihood component table	Create an output table which lists all of the likelihood components by fleet and automates the process of computing effective samples sizes	Development 100%; Testing 0%
m. Fit the BH-SRR to annual recruitments	Annual recruits for BH-SRR. Consider fitting the stock-recruitment relationship to the annual rather than seasonal recruitments.	Development 80%; Testing 0%
<b>Other developments</b>		
Task	Description	Status of completion
Initial biomass approximates the equilibrium unexploited biomass	Abundance in the first model period is equivalent to that estimated using the estimated BH-SRR in an unexploited equilibrium state	Development 90%; Testing 20%
Region-specific yield estimation	Calculate the equilibrium productivity and biomass from each region in a spatially disaggregated model	Development 100%; Testing 20%
Fixed terminal recruitment deviates	Exclude the estimation of the terminal temporal recruitment deviates in the most recent model time periods	Development 100%; Testing 100%
Logistic-normal size composition likelihood	Add the logistic-normal likelihood as parameterized by Francis (2014)	Development 90%; Testing 20%
Exclude reporting rates from tag predictions during mixing period	Do not to account for the estimated tag reporting rate when solving the Newton-Raphson catch algorithm for tags recaptured during the mixing period	Development 100%; Testing 20%

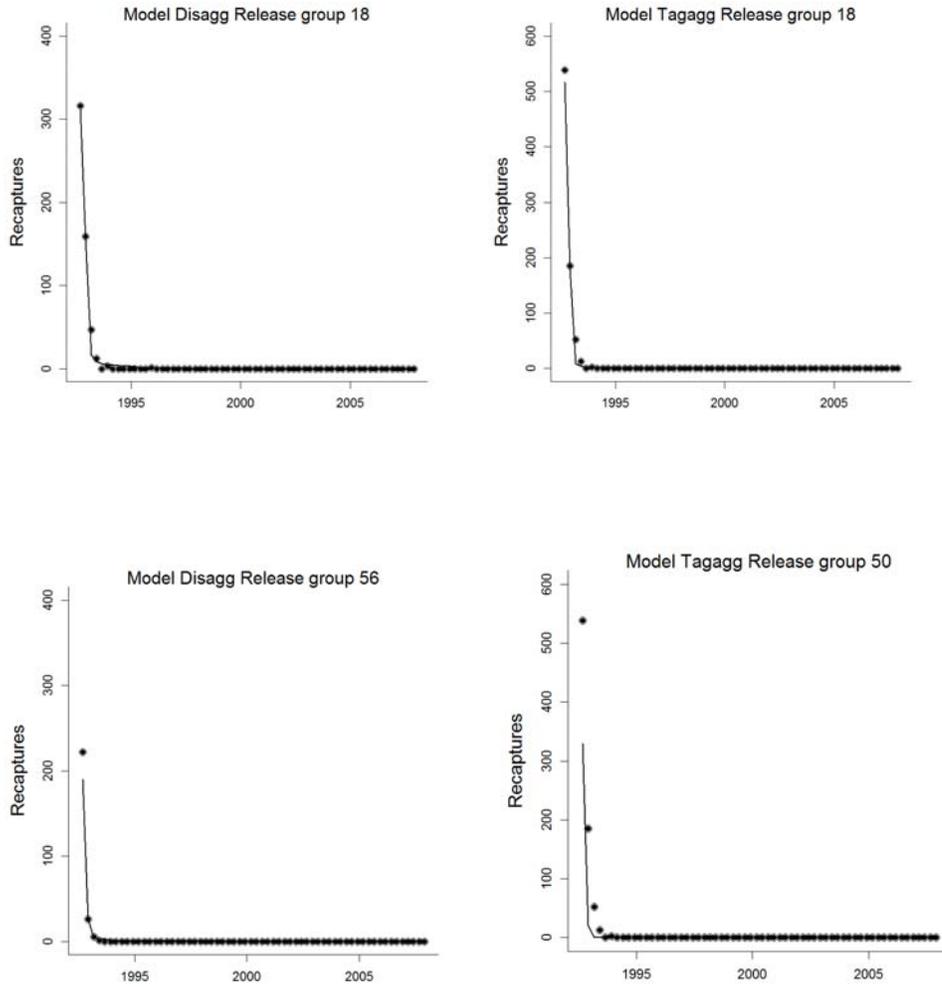
**Table 7.** New features to be added to MULTIFAN-CL in 2014-15 and subsequent years.

<b>Peer review recommendations</b>		
Task	Description	Implementation
l. Tag likelihood relative weighting	An option to add a likelihood weight to the tagging data component should be added	2014-15
d. Tags inform movement	Include an option which allows the tagging data to inform movement only rather than movement and mortality	2014-15
m. Fit the BH-SRR to annual recruitments	Annual recruits for BH-SRR. Consider fitting the stock-recruitment relationship to the annual rather than seasonal recruitments.	2014-15
b. Non-uniform size bins	Allow the length bins to be of different widths. One might, for example, want many narrow length bins for the smaller lengths, but fewer but wider length bins for the larger lengths.	2014-15
i. Tail compression	Include a “tail compression” option, which would pool all length- and weight-data for large and small sizes based on a specified percentage.	2014-15
h. Selectivity ogives	Allow for more general selectivity options, including selectivity patterns where the first age for which selectivity is non-zero is pre-specified. This should help to avoid selectivity being non-zero owing to the functional form for selectivity rather than data.	2014-15
k. Maturity at length	When maturity data are based on length, converting to ages should be done within the model.	2015+
c. Long-term tag loss	Allow for long-term and initial tag-loss. Currently initial tag-loss is implemented by reducing the number of animals tagged when inputting data to the model and no account can be taken of long-term tag-loss.	2015+
<b>Other developments</b>		
Task	Description	Implementation
Seasonal selectivity	Implement a scheme to estimate seasonal variability in selectivity coefficients	2014-15
Tags inform growth	Add length increments of tag recaptures to inform growth estimation.	2015+

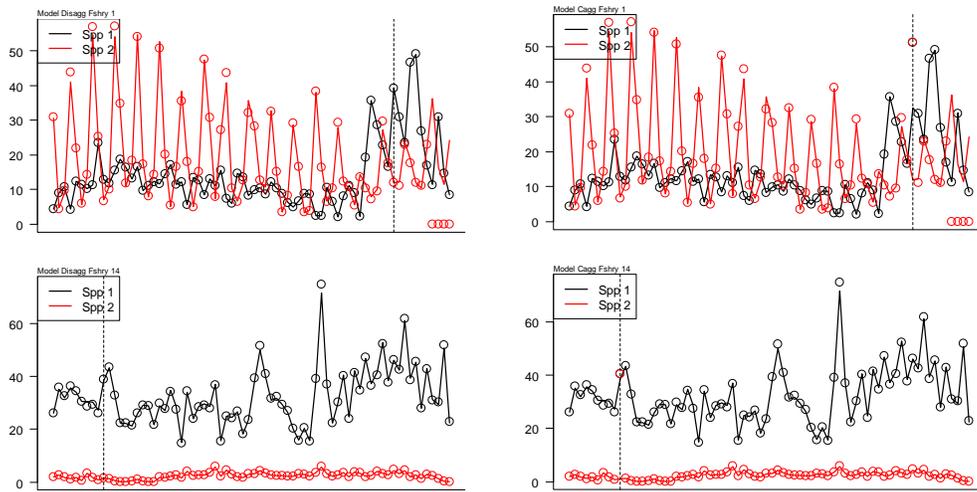
### Flow diagram – multi-sex model testing



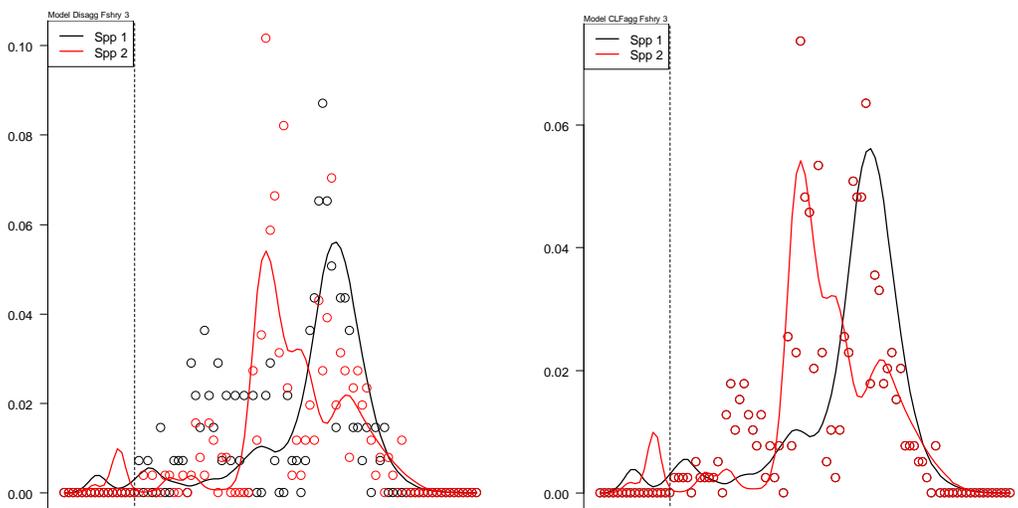
**Figure 1.** Flow diagram for testing the integrity of sex-disaggregated and sex-aggregated models.



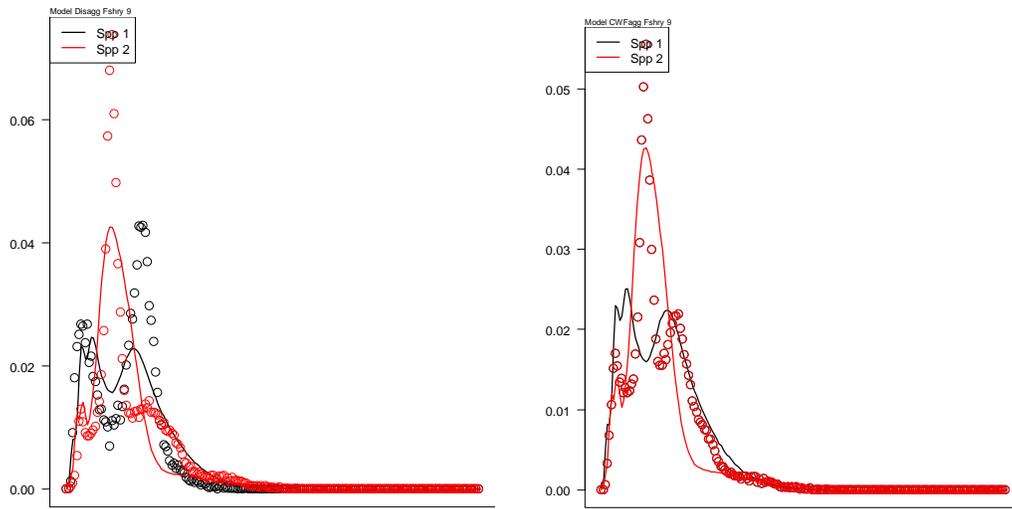
**Figure 2.** Observed and fitted recaptures with respect to recapture period for a model with tagging disaggregated among species, "Disagg", and a model where release groups 18 and 56 were aggregated, "Tagagg" where release group 50 is the duplicated release group for species 2.



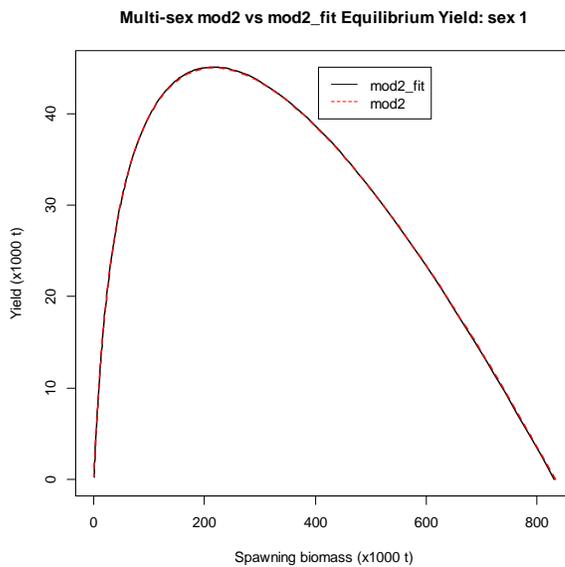
**Figure 3.** Observed (circles) and predicted (lines) catches for each species by fishery and for the model with observed catches disaggregated (Disagg) and aggregated (Cagg) in respect of catches. Vertical dashed lines indicate the fishing incident for which observed catches were aggregated.



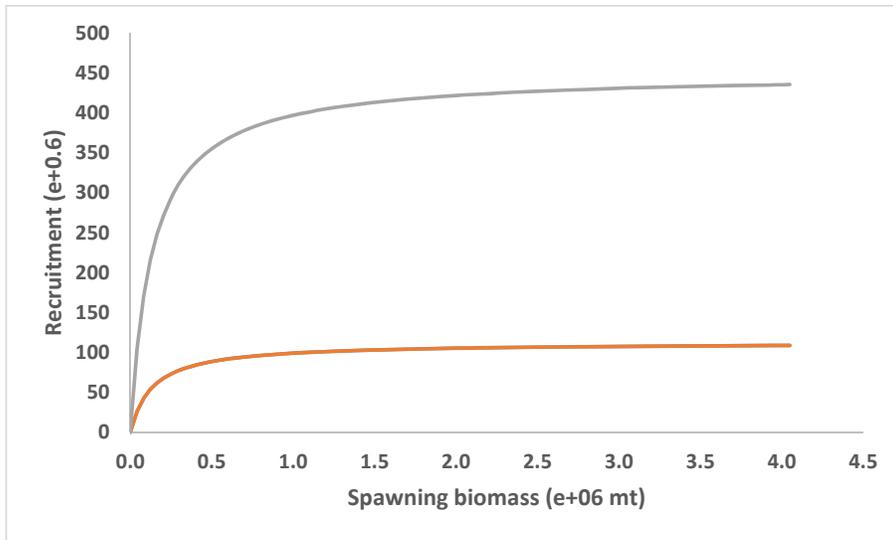
**Figure 4.** Observed (circles) and predicted (lines) length frequencies for each species by fishery and for models with observed catches and length frequencies disaggregated (Disagg) and aggregated (CLFagg).



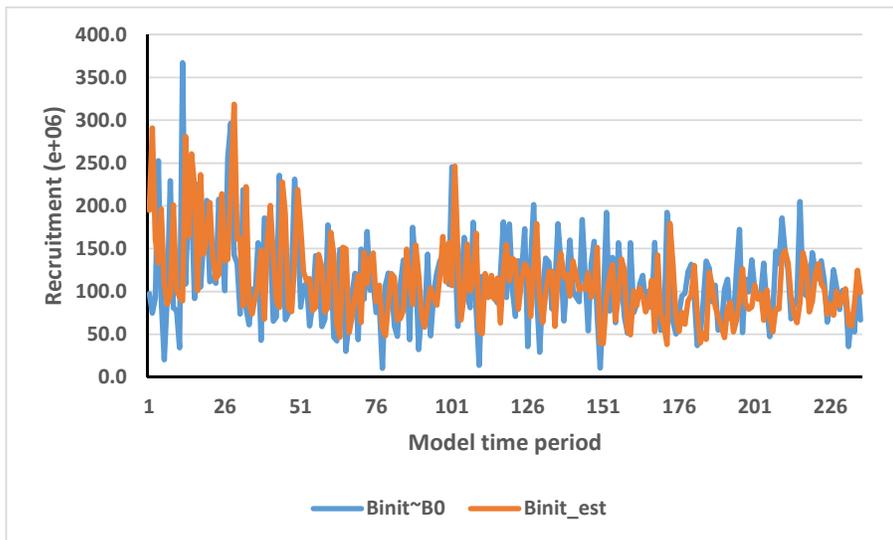
**Figure 5.** Observed (circles) and predicted (lines) weight frequencies for each species by fishery and for models with observed catches and weight frequencies disaggregated (Disagg) and aggregated (CWFagg).



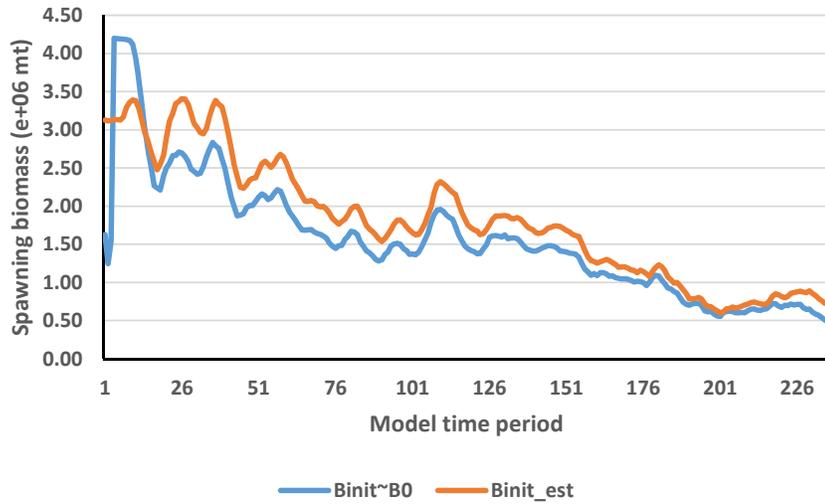
**Figure 6.** Equilibrium yield for two multi-sex models, the first derived deterministically using parameters from a multi-species model with identical species (mod2) and the second being the fitted solution for this model (mod2\_fit) that converged to a very similar solution to that of mod2.



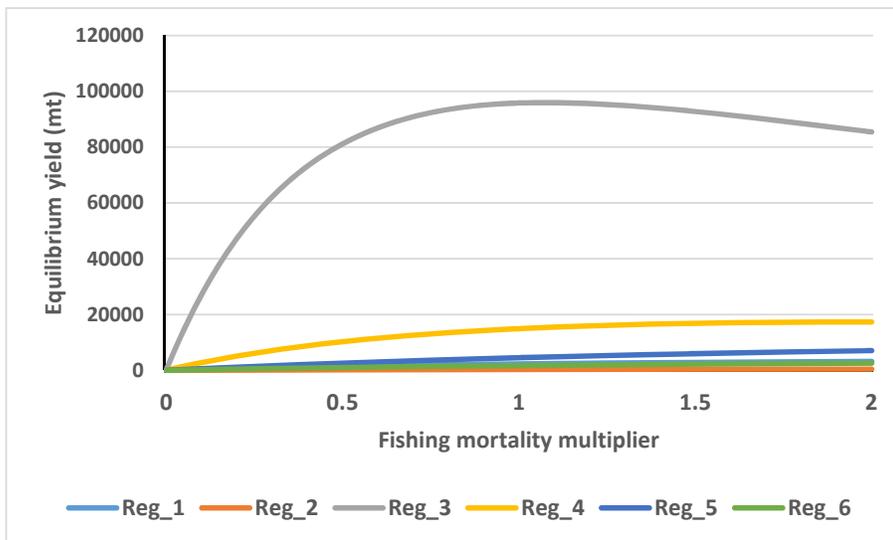
**Figure 7.** BH-SRR predictions for models with (grey line) and without (orange line) the  $B_{init} \sim B_0$  assumption showing the effect of assigning annual recruitments and biomass for the model with the assumption.



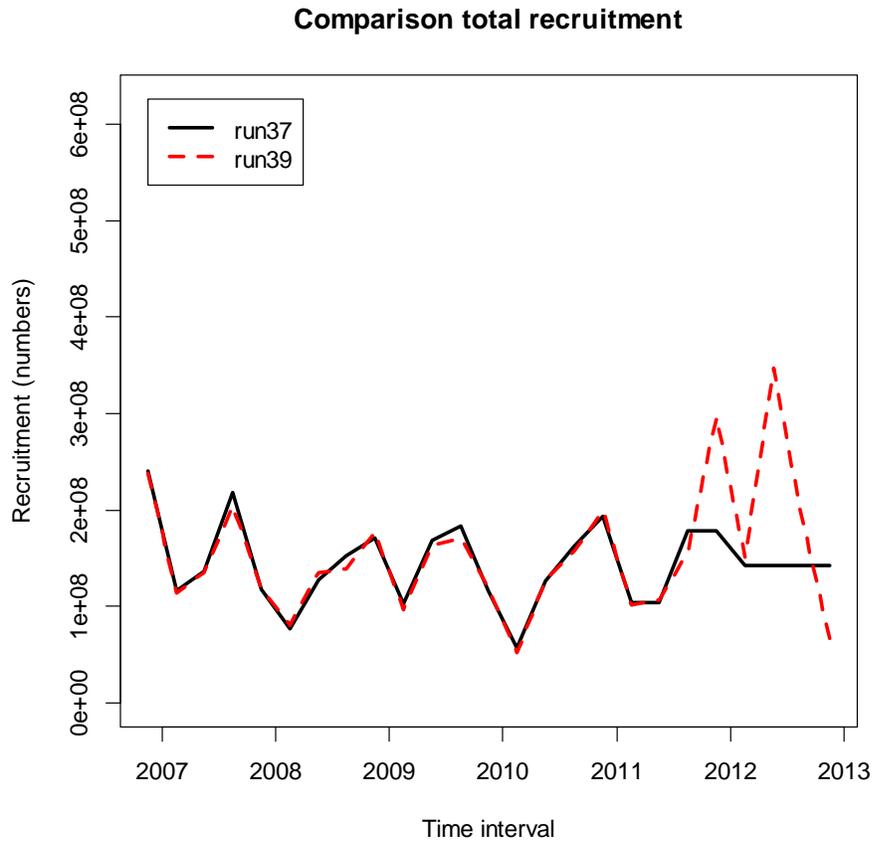
**Figure 8.** Estimated quarterly recruitments for models with ( $B_{init} \sim B_0$ ) and without (Binit\_est) the  $B_{init} \sim B_0$  assumption.



**Figure 9.** Estimated quarterly spawning biomass for models with ( $B_{init} \sim B_0$ ) and without (Binit\_est) the  $B_{init} \sim B_0$  assumption.



**Figure 10.** Equilibrium yields specific to regions calculated using the spatially dis-aggregated method relative to the fishing mortality multiplier applied over all regions.



**Figure 11.** Comparison of quarterly recruitment estimates for models without (run37) and with (run39) the estimation of terminal recruitments as controlled by `parest_flags(400)`.

## 9 ANNEX

### 9.1 Benchmark testing

Table 9.1.1. Summary of benchmark tests of the development version using single species data undertaken August 2013 to July 2014.

Compilation name	Compilation folder & Exec.name	Compilation date	Test folder	Test date	Test configurations	Results folder	Key results
MMERG	multi_spp_merge_2013	Mar-13	2013-03-15_fit_SSM_MSM/MSM_2013_02_15_skj_2011_fit	15/03/2013	af145=1	2013-03-15_fit_SSM_MSM_skj	Obj. = 89688.802
MMERG	multi_spp_2013_04_05	Aug-13	2013-11-20_fit_skj	20/11/2013	af145 = 1; penwt=af145/1000	2013-11-20_fit_skj	Essentially identical: MMERG Obj. 90547.279
MTHRD	multi_spp_2013_11_01	Nov. 2013			penwt=0.001		MTHRD Obj. 89689.512; MSY=3.606e+05; B0=5930000
MTHRD_bchmrk	multi_spp_2013_11_11	May. 2014	2014_05_05_repNov13	5/05/2014	af145 = 1; penwt=af145/1000	2014-05-04_fit_skj	penwt=0.001; Obj.=89689.511554852230; MSY=3.606e+05; B0=5.930e+06
MMERG	multi_spp_2013_05_06	Aug.2013	2014_05_05_repNov13_MMERG	5/05/2014	af145 = 1; penwt=af145/1000	2014-05-04_fit_skj	penwt=0.001; Obj.=90550.486338439616; MSY=3.515e+05; B0=5.843e+06
<b>Benchmark testing 11 May 2014</b>							
MTHRD_bchmrk	multi_spp_2013_11_11; exec_txf/mfclo64_stepbase	May. 2014	2014_05_11_fit_skj	11/05/2014	af145 = 1; penwt=af145/1000	2014-05-11_fit_skj	penwt=0.001; Obj.=89689.511554852230; MSY=3.606e+05; B0=5.930e+06
MTHRD_2013_11_01	multi_spp_2013_11_01; exec_txf/mfclo64_2013_11_01	May.2014	2014_05_11_fit_skj	11/05/2014	af145 = -3; penwt=10.0**af145	2014-05-11_fit_skj	penwt=0.001; Obj.=89689.511554852230; MSY=3.606e+05; B0=5.930e+06
<b>Benchmark testing 17 May 2014</b>							
MTHRD_2013_11_01	multi_spp_2013_11_01; exec_txf/mfclo64_2013_11_01	May. 2014	2014_05_17_fit_skj	17/05/2014	af145 = -3; penwt=10.0**af145	2014-05-17_fit_skj	penwt=0.001; Obj.=89689.511554852230; MSY=3.606e+05; B0=5.930e+06
MTHRD_2014_05_13	multi_spp_2014_05_13; exec_txf/mfclo64_2014_05_13	May.2014	2014_05_17_fit_skj	17/05/2014	af145 = -3; penwt=10.0**af145	2014-05-17_fit_skj	penwt=0.001; Obj.=90547.251906630234; MSY=3.631e+05; B0=5.974e+06

## 9.2 Beverton-Holt spawner stock-recruitment relationship: solution in respect of steepness

The Beverton-Holt spawner stock-recruitment relationship (BH-SRR) may be solved in respect of the steepness parameter when the assumption is true that the initial population conditions are in unexploited equilibrium.

Assume that the BHSRR is parameterized as

$$R = \frac{\alpha B}{\beta + B} \tag{x1}$$

where  $R$  is the recruitment and  $B$  is the (reproductive) biomass.

### Theorem

Assume that the population distribution etc. is in equilibrium. Then for each value of the steepness  $S$  (defined below) and the recruitment  $R$ , there exist uniquely determined values  $\hat{\alpha}$  and  $\hat{\beta}$  such that (x1) holds i.e.

$$R = \frac{\hat{\alpha} \hat{B}}{\hat{\beta} + \hat{B}} \tag{x2}$$

where  $\hat{B}$  is the biomass at equilibrium.

### Proof

The steepness  $S$  is defined to be the ration of the predicted recruitment,  $R$  at  $0.2\hat{B}$  to the predicted recruitment at  $\hat{B}$ , i.e.

$$S = \frac{\frac{0.2\hat{\alpha}\hat{B}}{\hat{\beta} + 0.2\hat{B}}}{\frac{\hat{\alpha}\hat{B}}{\hat{\beta} + \hat{B}}} \tag{x3}$$

We see that  $\hat{\alpha}$  cancels out and we can solve (x3) for  $\hat{\beta}$ .

$$\hat{\beta} = \frac{\hat{B}(1 - S)}{(5S - 1)} \tag{x4}$$

To solve for  $\hat{\alpha}$  we need the notion of the reproductive potential,  $\phi$ , of one fish. This is defined to be the equilibrium reproductive biomass for the constant recruitment of one fish.  $\phi$  has the nice multiplicative property that for constant recruitment  $R$ ,

$$\hat{B} = R\phi \tag{x5}$$

(Note that this is only true if the mortality rates are density independent.) Substituting this in the BH-SRR for  $\hat{B}$  yields

$$R = \frac{\hat{\alpha}R\phi}{\frac{R\phi(1 - S)}{5S - 1} + R\phi}$$

Or

$$R = \frac{\hat{\alpha}}{\frac{(1-S)}{5S-1} + 1}$$

Which simplifies to

$$\hat{\alpha} = \frac{4SR}{(5S-1)}$$

$$\hat{\beta} = \frac{R\phi(1-S)}{(5S-1)}$$

Which are the desired equations.