CSIRO OCEANS AND ATMOSPHERE
www.csiro.au

Options for Tier 5 approaches in the SESSF and identification of when data support for harvest strategies are inappropriate.

Malcolm Haddon, Neil Klaer, Sally Wayte, and Geoff Tuck

CSIRO Oceans and Atmosphere, GPO Box 1538, Hobart, TAS 7001, Australia.

June 2015 FRDC Project: 2013/202

National Library of Australia Cataloguing-in-Publication entry

ISBN: 978-1-4863-0493-6 (print.)
Notes: Includes bibliographical references.
Subjects: Fish stock management.
Commercial Fisheries Harvest Strategies.

Author: Haddon, Malcolm.
Other Authors: Wayte, Sally, Klaer, Neil, Tuck, Geoffrey
Title: Options for Tier 5 approaches in the SESSF and identification of when data support for harvest strategies are inappropriate FRDC Project: 2013/202
Year of Publication: 2015

This publication (and any information sourced from it) should be attributed to:
Haddon, M., Klaer, N., Wayte, S., and G. Tuck (2015) *Options for Tier 5 approaches in the SESSF and identification of when data support for harvest strategies are inappropriate.* CSIRO. FRDC Final Report 2013/200. Hobart. 115p.

Researcher Contact Details

Name: Malcolm Haddon
Address: Oceans and Atmosphere, CSIRO, Castray Esplanade, Hobart, GPO Box 1538, Tasmania, 7001, Australia
Phone: 03 6232 5097
Fax:
Email: Malcolm.Haddon@csiro.au

FRDC Contact Details

Address: 25 Geils Court
Deakin ACT 2600
Phone: 02 6285 0400
Fax: 02 6285 0499
Email: frdc@frdc.com.au
Web: www.frdc.com.au

Contents

1 Acknowledgments

The authors would like to thanks E.J. Dick for sending through R code for running the Depletion-Based Stock Reduction Analysis (as in Dick and MacCall, 2011). The code for the Depletion Corrected Average Catch came from the NOAA Toolbox. Jemery Day is thanked for modifying the R code to facilitate using it with the CSIRO MSE simulation framework. We also thank the rest of the CSIRO SESSF team and SESSF RAG members for discussions concerning data-poor needs, methods, and limitations.

2 Executive Summary

2.1.1 Background

The quota species within the SESSF are managed using stock assessments to estimate recommended biological catches (RBCs) for each species, which after considering State catches, discards, social, economic, and indigenous requirements, lead to a Total Allowable Catch (TAC). This occurs in the context of the Commonwealth Harvest Strategy Policy (HSP). The principle conservation requirement in the HSP is that a fished stock should stay above the Limit Reference Point (LRP) at least 90% of the time. To ensure this happens, a tiered set of harvest strategies have been developed each with their own data requirements, stock assessment method and decision control rule. The HSP was introduced in 2007 and since then it has become apparent that even the relatively data-limited Tier 3 and 4 harvest strategies are not appropriate for some quota species and may be providing misleading management advice. The HSP is currently under review and this seems likely to expand the need for stock assessments, even for data-poor minor or by-product species. These issues meant there was an urgent need for a higher Tier than Tiers 3 and 4, which would be able to handle the spectrum of data-poor fisheries from those with only limited catch data to those with biological information as well as detailed catches.

In this present work, Management Strategy Evaluation (MSE) was used to test some of the factors that can lead to the assessment methods in the existing harvest strategies being inappropriate for assessing the status of some species. This aspect of the project focused on the effects of sample size (precision) and of bias on the outcomes of Tier 3 and 4 assessments applied to lightly and highly depleted stocks.

MSE was also used to test and compare seven different data-poor methods ranging from simple median, average, and 3rd highest catch estimates (for truly catch only data), and model assisted catch-only methods that included the Depletion-Corrected Average Catch, the Depletion-Adjusted Catch Scalar, and the Depletion-Based Stock Reduction Analysis (which are aimed at species for which some biological information is also available). The MSE tested data-poor scenarios where the initial depletion level was in fact either heavily depleted, on target, or only lightly depleted. The data-poor methods were applied to each of these scenarios while assuming the simulated stocks were also in each of these states (thus each assumption was tested against each simulated reality to determine how sensitive each method was to making incorrect assumptions). The main candidate Tier 5 methods were also applied to three species for which there were well developed Tier 1 assessments so as to illustrate typical outputs and the strengths and weaknesses of the methods.

2.1.2 Objectives

The objectives of the project were to:

1. *Establish guidelines, using SESSF case studies, for when the particular tier harvest strategy for a given stock becomes inappropriate and make explicit recommendations as to what response would then be appropriate,*
2. *Determine options for alternative harvest strategies when none of the present tiers is appropriate (i.e. potential Tier 5 approaches), and*
3. *Produce presentations and explanatory documents for distribution across RAGs and MACs, describing the criteria and new Tier 5 harvest strategies.*

Each of these was addressed during the project, with the third objective, an extension of the findings to RAGs, MACs, and other stakeholders, expected to extend beyond the life of the project.

2.1.3 Recommendations

- The measures of central tendency (median catches, 3rd highest catches, maximum constant yield) were useful for truly data-poor species.
- The model assisted assessment methods, depletion corrected average catch (DCAC) and depletion based stock reduction analysis (DB-SRA) provide more details and estimates of risk associated with their sustainable catch estimates.
- Assuming stocks to be only lightly depleted risks over-fishing in data-poor species.
- Data-poor assessment methods and their reporting should be automated, as much as possible, to produce recommended biological catch levels for chosen species.
- The outputs of the DCAC and DB-SRA are greatly influenced by the assumed initial depletion and final depletion levels. The RAGs should approve the levels selected for each species which use these methods.
- The final depletion level selected for the DB-SRA influences the outcome so a range should be explored for each assessment to make the RAG aware of the risks associated with each assumed level.
- The SESSF RAGs, perhaps with MAC agreement, will need to decide which species currently assessed using either Tier 3 or Tier 4 should be nominated to be assessed using one of the Tier 5 methods. Given the amount of information available for currently assessed SESSF species it should be possible to use the DB-SRA.
- Depending on the requirements within the revised Harvest Strategy Policy (HSP), the RAGs will also have to decide what approaches should be used with any new species included in the HSP.
- The Tier 5 candidate approaches examined do not constitute an exhaustive list. The development of data-poor assessment methods and related harvest strategies is not a static field and notice should be taken of future data-poor assessment methods as they become available (in particular the Catch-MSY method and its potential derivatives).

Keywords
Flathead, School Whiting, Jackass Morwong, Management Strategy Evaluation, MSE, SESSF, Tier 3, Tier 4, Tier 5, data-poor

3 Introduction

3.1 Background

The SESSF has had a tiered set of harvest strategies in place since 2007, and the assessment methods and harvest control rules specific to each tier have since been formally management strategy evaluation (MSE) tested to ensure that they meet the Commonwealth harvest strategy policy objectives (Little *et al*., 2011; Wayte, 2009; AFMA Project 2006/815, FRRF Project RUSS). This testing highlighted some problems with existing strategies and provided solutions which were implemented (Wayte and Klaer, 2010). There are two major issues remaining with the current tiered system: (i) to answer when it is most appropriate to move species from one tier to another (when is a given tier inappropriate), and (ii) how to assess particularly data-poor species that have CPUE indices that do not appear to reflect abundance or may only have a relatively short time-series of representative catch data. Generally, the Tiered harvest strategy approach implemented in the SESSF appears to be performing well (Smith *et al*, 2014). However, as with all systems, continued improvement and accounting for exceptions as they arise is required.

At present, the most data-poor tier level in the SESSF is the Tier 4 harvest strategy that uses current and target CPUE and catch levels to determine an RBC. One of the assumptions required for the Tier 4 approach to be valid is that CPUE provides a reliable index of relative abundance for the species (Haddon, 2014). It is becoming increasingly clear that CPUE is not a reliable index of abundance for a number of current Tier 4 species, so there is a need for an alternative harvest strategy and tier for such species. One species, royal red prawn (*Haliporoides sibogae*), has been recognised by the resource assessment group as not appropriately fitting within any of the existing tiers and yet, because there is no current alternative, a Tier 4 analysis continues to be used. Similarly with the Tier 3 approach, the management advice for some species has been highly variable from year to year (e.g. Mirror Dory) and its reliability with some species has been questioned (another failure of the underlying assumptions; Klaer, 2014) so alternatives are required.

Various procedures for assessing the status of data poor species that do not have a reliable index of abundance or snapshots of age information have been examined for Australian Commonwealth fisheries (FRRF project RUSS and FRDC project 2010/044), providing a list of candidate data-poor Tier 5 methods that could be recommended for use in the SESSF. In comparison to tiered assessment approaches implemented by other nations, Australia is unusual in that the SESSF does not have a procedure, for example, that uses catch history alone to arrive at TAC recommendations (e.g. New Zealand uses a Constant Annual Yield and the USA now often uses the Depletion-Based Stock Reduction Analysis approaches; Dick and MacCall, 2011). Globally, there are on-going efforts to develop workable stock assessment methods and related harvest strategies for such data-poor stocks; with, for example, a Wakefield Symposium on Data-Poor Approaches being held in May 2015. There is good reason to conclude that there are many options that could be used to bridge the gap between the currently available tiers in the SESSF and the Ecological Risk Assessment (ERA), which, of course, does not provide the RBC required for by-product and minor species.

The current internationally recognised approach for testing new harvest strategies, including assessment methods and harvest control rules, is Management Strategy Evaluation (MSE). This project used MSE to examine the performance of identified data-poor procedures to compare their outputs with existing results for Tier 1 species in the SESSF. In addition, the effect of precision and accuracy of data collected from the fishery (how representative the data is of the stock in question) on the performance of existing Tiers 3 and 4 was tested to determine the potential effect of each on the performance of those harvest strategies.

As MSE is now a global standard for the testing of alternative harvest strategies and associated data requirements, the SESSF is now well placed to focus this powerful procedure on a wide range of actual fishery problems. While the method is standard, the range of potential problems that could be examined is limitless, so care is required to direct effort to well identified major issues. The objectives of this project were formulated based on high priority problems in the SESSF identified by AFMA, SESSF RAGs and MACs, but have outcomes that will be useful for other fisheries.

More details concerning the background of the particular problems within the SESSF and the context in which they occur are provided in Appendix 10 (page 27).

3.2 Current Research

The idea of using a fishery's catch history as a means of assessing stock status has led to some intense debate in the literature. Pauly (2013) argues that if all that is available is catch-data then efforts must be made to use that data, and more fisheries should at least have catch data collected. Hilborn and Branch (2013), however, argue that catch-data alone will be misleading so often that it is dangerous to use such methods to provide management advice. The strategy of using MSE testing of such harvest strategies is more effective than merely arguing about their potential value and potential biases and can provide increased clarity to the debate about when catch-only methods are useful and when they are not; the need for such formal testing applies to all catch-only data methods.

This project examines an array of data-poor assessment methods, but those considered do not constitute an exhaustive list of possible methods. For example, the fishery status classification of Anderson et al. (2012) extends the methods originally proposed by Froese and Kesner-Reyes (2002) to assess stock status using only the time series of catch data (as a proportion of the peak maximum catch). By examining stocks with full quantitative stock assessments, they were able to categorize the sequence of development of a fishery; from its early developmental stage, to full exploitation, over-exploitation and potential collapse. This method would be difficult to implement without a suitable harvest control rule, but if one could be developed it would become worthy of testing.

Those methods that use a form of stock reduction analysis are essentially attempting to identify plausible combinations of population dynamics that would at least be consistent with the observed catches. Martell and Froese (2014) have taken that idea and produced a method that explores the region of plausible dynamics by including a simple model of those dynamics. By using a simple Schaefer surplus production model and setting bounds on the parameters of that model, biomass trajectories that are inconsistent with the observed fishery can be eliminated (i.e. the observed catches might lead some combinations to go extinct or to expand well above the hypothetical carrying capacity). By conducting a Monte Carlo analysis of the possibilities and using the outcomes of the plausible model parameter set, Martell and Froese (2014) are able to generate estimates of MSY along with uncertainty estimates about the management statistics.

Martell and Froese's (2014) harvest strategy has similarities with the approach described by Bentley and Langley (2012) who describe a method that employs "Feasible stock trajectories", although their underlying model is more complex than that used by Martell and Froese (2014).

Carruthers et al. (2014) provide a detailed review of data-poor assessment methods and a review of the literature on data-poor harvest strategies is provided by Dowling *et al*., 2015a), while guidelines for the development of data-poor harvest strategies are given in Dowling *et al*. (2015b).

4 Objectives

There were three objectives with the first two aimed at improving current and potential future practice and the third aimed at communicating the outcomes from this study to the people who will need to implement any recommended changes.

- Establish guidelines, using SESSF case studies, for when the particular tier harvest strategy for a given stock becomes inappropriate and make explicit recommendations as to what response would then be appropriate.

- Determine options for alternative harvest strategies when none of the present tiers are appropriate (i.e. potential Tier 5 approaches)

- Produce presentations and explanatory documents for distribution across RAGs and MACs, describing the criteria and new Tier 5 harvest strategies.

5 Methods

5.1 Appropriateness of Selected Tiers

The first objective here is to "*Establish guidelines, using SESSF case studies, for when the particular Tier harvest strategy for a given stock becomes inappropriate and make explicit recommendations as to what response would then be appropriate.*"

The selection of which SESSF harvest strategy tier to use is currently based primarily on the available information that can be used to make an assessment of stock status. To be explicit about why this can be a problem, the selection does not currently take into account the capability of a selected tier to estimate stock status from that available data. The first objective therefore requires an examination of the consequences of the application of the current Tier 3 or Tier 4 harvest strategies when the data may not be of sufficient quality to support that application. Errors in observations can be categorised as related to precision and bias (or accuracy; **Figure 11** page 45), which can be examined explicitly and separately in the context of SESSF tiers and species.

To date, MSE work in the SESSF (Wayte 2009, Klaer and Wayte 2011, Little et al., 2011) has assumed that sampled data from the simulated fish population are at levels of precision that reflect average apparent observed levels across sampled species. It also assumes that sampling is random and unbiased and therefore accurately represents the stock. Where stocks are spatially heterogeneous or have an extensive geographical distribution and sampling occurs unevenly across different areas then biases may enter the data simply through uneven sampling of natural variation. In the simulated sampling within the MSE conducted here the precision and bias within sample collections was varied across all major sampled data sources (CPUE, length, age) to determine how, for each harvest strategy or tier, this modifies the risk to the stock, as defined by the harvest strategy policy (staying above the LRP 90% of the time), and to assess the ability of the HS to achieve and maintain the target depletion level.

The effect of data precision is simplest to implement, and requires testing of a range of assumed variance values for sample collection from the simulated population. Data accuracy (bias) is more difficult to implement but can be addressed using plausible scenarios. Examples are a bias towards sampling of more longer/older fish from the population, or a linear trend in catchability in a CPUE index. The latter specifically allows testing of the effect of gear/vessel improvements over time (often termed 'effort creep') that may not have been accounted for in CPUE standardisations, and how that may affect the outcome of the application of the Tier 4 harvest strategy. A total of 36 different combinations of bias and precision were tested with 18 on Tier 3 and 18 on Tier 4 (**Table 6** page 52).

Specific details and equations of how bias and precision were implemented in the MSE and illustrations of how they influence the simulated fisheries data are provided in Section 12.3 (page 45).

5.2 Management Strategy Evaluation

5.2.1 Introduction to MSE

A SESSF management strategy evaluation (MSE) simulation framework has already been developed (Wayte 2009, Fay *et al*., 2009; Klaer and Wayte 2011; Little et al., 2011), and provides a flexible platform for testing harvest strategies as they apply to SESSF species in particular, but also more generally. New projects requiring MSE testing thus no longer require the development time for the detailed operating model that incorporates uncertainty with the dynamics of a fish stock, sampling of data required for stock assessment, and the implementation of established SESSF harvest strategies.

The stock assessment methods and harvest strategies proposed here are new, however, and hence require adjustments to the simulated sampling of data collected from the fishery to examine effects of different levels of precision and accuracy. It also requires the implementation of new data-poor assessment methods and harvest control rules. However, these modifications to the existing system are relatively minor, thus allowing this proposal to be built as a one-year project, with greater focus on planning and running of appropriate simulations and interpretation of the results.

It is standard practice to base MSE testing on species for which good information is available, so that the results across a full range of harvest strategies can be compared (in this case Tiers 1 to 5). The SESSF data rich species used in the following analyses were Tiger Flathead (*Neoplatycephalus richardsoni*) and School Whiting (*Sillago flindersi*), and testing is carried out under a range of stock depletion levels for each species, including being above, at, and below the target depletion level (initial depletion levels used were $0.18B_0$, $0.48B_0$, and $0.78B_0$). The intention of this was to determine whether each assessment method tested was capable of recovering a depleted stock, maintaining it close to the target, and of fishing a stock down in a controlled fashion.

The Tier 1 harvest strategy in the SESSF involves a fully quantitative stock assessment, which has a variety of standard data sources including length and age composition and also an abundance index (generally in the SESSF this is CPUE). In the case where one of those sources contains an unknown bias, if the bias is sufficiently or consistently large the assessment will show a conflict among data sources, thereby allowing recognition and investigation of the source of that bias, and dealing with it in some way, at least in alternative model structures via sensitivity analysis. However, Tiers 3 and 4 both rely on a single source of input data (age composition and CPUE trends, respectively), which implies that such biases would not be detected.

The MSE and its details are described in the Appendix in Section 11 (page 36).

5.2.2 Summary Statistics used to Compare Relative Performance

To compare the relative performance of each Tier's methods under different conditions or the relative effectiveness of the different candidate Tier 5 methods, four performance statistics were estimated and plotted for each of the scenarios run. These statistics were: 1) the average annual catch, 2) the spawning biomass depletion relative to the unfished spawning biomass, 3) the average annual variability in the catch across the projection period, and 4) the probability of the spawning biomass going below the limit reference point ($20\%B_0$). Formal descriptions and related equations are provided in section 12.3.5 (on page 53).

5.2.3 The Effects of Bias and Precision on Tiers 3 and 4

In the MSE testing of the effects of bias and precision on the outcomes from the Tier 3 and Tier 4 harvest strategies the MSE simulation framework for the specified stock was first put into an initial state of depletion by applying a known catch history and varying the recruitment dynamics. Only two levels of depletion were aimed for, those being a stock that was only lightly depleted and one that was heavily depleted. These were relative levels of depletion and were not precisely set each time because the random variation that is an important part of capturing the uncertainty when running an MSE meant that the initial depletion level was always slightly different. Once the initiation was complete the selected combination of bias and precision was applied and the population dynamics projected forward for 30 years. In each scenario the selected bias or precision was introduced in a linearly increasing fashion for the first 20 years and then continued for a further 10 years (e.g. see **Figure 16** and **Figure 17** on page 50). For each of the two Tiers there were 18 different scenarios composed of different combinations of initial depletion, precision of the CPUE and sample size of the age samples, and then finally different bias levels in both the CPUE and age samples. The full list of alternatives are listed in **Table 6** (page 52).

The outcomes of each scenario are plotted as a series of boxplots by species and depletion level (e.g. **Figure 23** on page 62) to provide a visual depiction of relative performance.

5.2.4 The Candidate Tier 5 Assessment Methods

The Tier 5 methods considered can either be fixed, where a single catch level is set at the start of the projections, or dynamic, where there is feedback from any response of the stock and the analyses are updated regularly using new data from the fishery. As with the Tier 3 and 4, bias or imprecision would not be detectable in either of these Tier 5 approaches.

The assessment methods considered here do not include all possible methods and new approaches continue to be developed (e.g. Martell and Froese, 2014). The Tier 5 harvest strategy being explored here is unlike the other SESSF Tiers in that it will contain an array of possible assessment methods each of which may be able to generate an estimate of sustainable catch (see **Figure 60**). However, the notion of a species being data-poor covers a wide range with some species literally only having catch data while others may have catch and an array of biological information relating to growth, mortality, productivity, and in some cases a range of possible initial and final depletion levels. To reflect this range the proposed Tier 5 can be any one of a range of assessment methods with the final selection being a reflection of exactly what information is available and should be decided or at least confirmed by the RAG involved.

Seven different methods were considered in the MSE testing of potential Tier 5 methods (**Table 1**). Four were purely catch-only methods that attempt to determine some form of central tendency of the sustainable catch for each species. The idea is that if recent catches have been relatively stable and the RAG involved consider that the stock status does not appear to have altered significantly, then the observed catches represent a sustainable catch. There are numerous uncertainties with this approach, not least being that it is possible to sustainably over-fish a stock albeit at a lower yield than would be possible if the stock were in a less depleted state. To make partial allowance for this it is possible to use a fractional multiplier (**Table 1**, method 5). Alternatively, if a fishery is in the process of developing, with catches exhibiting an increasing trend, then emphasizing early smaller catches may under-estimate the potential yield, hence the option of using the third highest catch from a specified period. In each case the RAG involved in the assessment would need to provide guidance or at least agreement for the period chosen over which to summarize the catches and whether or not to update the sustainable catch estimate and if so at what interval.

These purely catch-only methods, if they are fixed and are not updated, inherently aim to maintain the status quo, which may be all that is required for minor by-product species.

Table 1. Some alternative catch-only methods for setting an RBC. T5 lists the numeric code used in the diagrams and Code the two-letter code. $C_{0..-x}$ implies the catch from the current year to –x years before hand; 0..-9 is the previous ten years.

T5	Code	Brief Description	RBC
4	C3	Third highest landings over the last 10 years	third highest($C_{0..-9}$)
3	MC	Median catch from the last 10 years	median($C_{0..-9}$)
3	MC	Median catch from the last 3 years	median($C_{0..-2}$)
5	CY	Scaled average catch from a reference period - MCY	$c\overline{Y}$
6	DB	DB-SRA – depletion based – stock reduction analysis	median(DB-SRA)
7	DC	DCAC – depletion corrected average catch	median(DCAC)
8	DA	DACS – depletion adjusted catch scalar	median(DACS)

For more valuable but still data-poor species, some further more adaptable methods might be required. Three were tested here and more are becoming available. The three tested were the Depletion-Based Stock Reduction Analysis (DB-SRA), the Depletion Corrected Average Catch (DCAC), and the Depletion Adjusted Catch Scalar (DACS) (**Table 1**). Each of these produces not only an estimate of what should be a sustainable catch but also characterizes the uncertainty about the estimate. This would allow for refinements of the Harvest Control Rule or the specific catch level selected. For example, if the estimate was extremely uncertain then using some value lower than the median estimate of sustainable catch would be an option. Here the tests were conducted using the median catch. Discounts of either 0% or 25% of the predicted catch were applied to each of the candidate Tier 5 methods.

A total of 200 different scenarios were run and compared (a full list is given in **Table 9**, page 56) but they included different initial conditions of stock depletion and assumed different initial degrees of depletion for each of the alternative candidate methods.

Again the outcomes of each scenario were plotted as boxplots (e.g. **Figure 27** on page 69) to provide a visual depiction of relative performance. In addition, the median trajectories across the 100 replicates run for each scenario were plotted up for each scenario to enable the longer term behaviour of each scenario to become apparent (e.g. **Figure 28** on page 70). This facilitates the interpretation of the boxplots which often represent information summed across the projection period or selected from the final year. The median values tend to be relatively smooth and gradual, which obscures known variability. To further communicate the form of any single trajectory five randomly selected trajectories from contrasting scenarios were plotted to illustrate typical variability expected in a particular fishery (e.g. **Figure 29** on page 71).

In an attempt to illustrate the outcomes across all scenarios succinctly, phase plots of catch levels versus depletion levels were made for all scenarios across species and discounts (see **Figure 52** on page 94) and also a table summarizing the scenarios and their outcomes (see **Table 11** on page 95).

5.3 Application of Tier 5 Methods to Current Fisheries

To demonstrate the application of some of the Candidate Tier 5 methods and to provide illustrations for the third objective of communicating the new approaches to stakeholders, the Depletion Corrected Average Catch and the Depletion-Based Stock Reduction Analysis were applied to two different species: 1) Flathead (*Neoplatycephalus richardsoni*) and 2) School Whiting (*Sillago flindersi*). Flathead has a time series of catches stretching back to 1915 (98 years) so the DB-SRA method was applied to successively shorter sub-sets of that data to determine the effect.

For the fixed methods using median or third largest catch the last 10 years (or three) years of historic catches were used to set the RBCs.

In all cases the assumed final depletion was 48% (on target) although with the Flathead data an alternative final state of 35% was also run so as to illustrate the effect of altering the assumed end point.

5.4 Communication of New Methods

PowerPoint presentations are under development that will include the application of the candidate methods described here to an array of species in the SESSF to illustrate and formally describe the methodology and allow the RAG, MAC, and other interested stakeholders to become aware of and even apply the Tier 5 methods once they are agreed upon.

Worked examples have proven to be the most effective way to demonstrate and communicate novel analytical techniques. This will be required to gain acceptance by the RAGs of these new approaches.

6 Results/Key Findings

6.1 When is a particular Tier Appropriate?

Determining the appropriateness of a particular assessment (and in the SESSF the associated Harvest Strategy) has not been attempted formally. Ideally one would use simulation testing to determine how well, given data typical of a fishery, it was possible to estimate the stock status performance measures (whether that is the B_{Curr}/B_0 of a Tier 1, the F_{curr} of the Tier 3, or the scaling factor SF_{curr} of the tier 4). If that were known then a decision could be made as to how precise an estimate was necessary before a particular tier was deemed inappropriate. But even if this process were to be conducted regularly it would not capture all the possible issues concerning the appropriateness of different assessments. The precision of any estimate is certainly related to how variable the data being used tends to be, but can also be greatly affected by whether or not the data used in a stock is truly representative of the stock as a whole. This assumption is again difficult to test although high levels of variation in data between years would be indicative that something about the sampling is not managing to encompass the full variation within the stock as a whole (**Figure 1**).

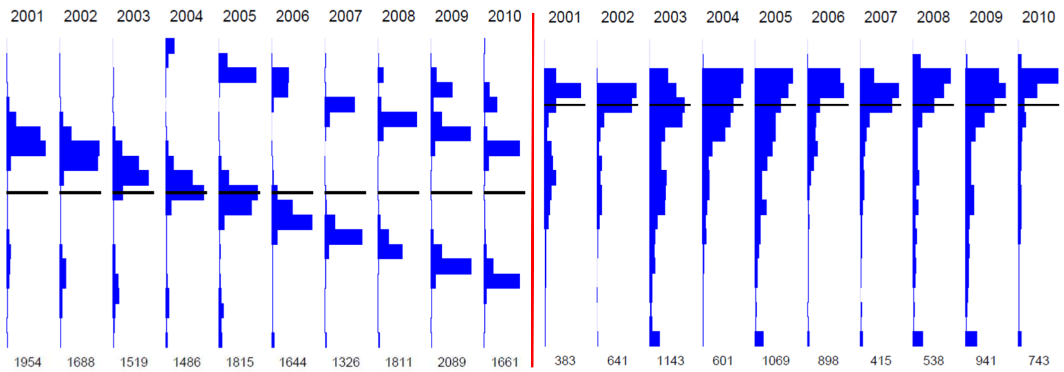

Figure 1. The age samples for Blue Grenadier (*Macroronus novaezelandiae*; left-hand graph) and Blue-Eye Trevalla (*Hyperoglyphe antarctica*; right-hand graph) for the years 2001 – 2010 (Klaer *et al*, 2014), illustrating the variation between years. Blue Grenadier shows almost ideal data with clear year classes progressing each year and consistency through time (although with some ageing error apparent in the spread around the particular strong year classes). Blue-Eye Trevalla, on the other hand, shows inconsistencies every year with annual progressions of year classes being vague and ephemeral at best. For example, 2006 and 2007 have similarities but differ markedly from 2008 and 2009, apparently indicating completely different age structures.

6.1.1 Tier Assumptions

A minimum requirement for a particular tier to be appropriate would be that the species' biology and the available data adhere to the assumptions inherent in the methods associated with the Tiers. Klaer (2014) and Haddon (2014) list the assumptions for Tiers 3 and 4 respectively. But meeting the assumptions is not always able to be cleanly determined. Thus, for example, for the Tier 4 harvest strategy to be valid requires the CPUE to actually provide an index of relative abundance. But it is not clear how far the relationship between CPUE and stock biomass can deviate from a simple linear relationship before the Tier 4 HS would become unworkable. Because it has previously been updated each year this may correct small deviations from the assumption of linearity. This could be tested using an MSE framework but there are many ways in which CPUE could deviate from a linear relationship with biomass and to test them all would involve very many simulations and a lot of time. In many cases where Tier 4 assessments have been known to have been invalid, the absence of an alternative that could provide catch level advice meant that the assessments were not rejected. The simplest example of this is where all the deep water Oreo species are currently assessed using the Tier 4 approach. The fundamental assumption behind the Tier 4 analyses (which use catch and standardized catch rate time series of data; Little et al., 2011) is that catch rates reflect relative abundance of the stock and are representative of the whole stock. Neither of these assumptions are met, especially since the advent of the 700m closure for deepwater species in Australia. Oreo catch rates vary from extremely low to extremely high, depending upon whether the aggregations of fish are targeted or not (**Figure 2**).

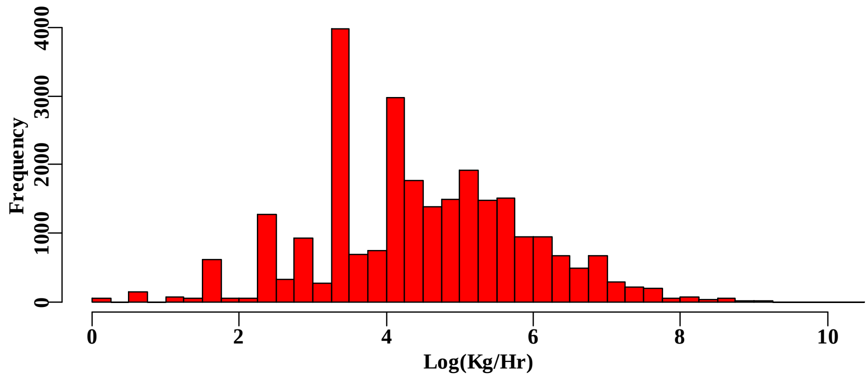

Figure 2. The log-transformed CPUE for the mixed Oreo category, to the end of 2013, which includes *Allocyttus verrucosus* (Warty Oreo), *Neocyttus rhomboidalis* (Spikey Oreodory), *Neocyttus psilorhynchus* (Rough Oreodory), *Allocyttus niger* (Black Oreodory) and a further mixed category (Oreodory). Note the spikiness of the lower levels of CPUE containing large numbers of records. The first five spikes relate to 5, 10, 15-20, 30, and 60 kg/hr.

It remains with the RAG and the full assessment process to ascertain whether or not the application of a particular Tier is appropriate. This requires the assessment scientists to present the analyses along with a listing of where the assumptions may deviate from those that are required. Whether or not to apply a data-poor method instead of the current Tier should be determined before applying the alternative so as to circumvent the possible accusation that the method has been selected because it generates the catch levels preferred by different stakeholder groups.

There are currently no standard, routine methods, or formal criteria that can be applied to determine whether a fisheries stock assessment is appropriate which can be applied independently of the assessment and management process in which it is embedded. In the SESSF it is the Resource Assessment Groups (RAGs) that determine whether or not to accept a stock assessment, and this tends to be done on a weight-of-evidence approach that attempts to account for consistency through time, the relative quality of fit of the model to the data, and whether the model structure correctly represents the stock dynamics as far as they are known. Most often a stock assessment might be rejected on the basis of qualitative reviews of the match between the model structure used and what is known about the fished stock. In a Tier 1 assessment if two data streams are in conflict, with one implying things are improving and another implying things are declining, it would be more usual for one of the data streams to be rejected rather than the assessment; alternatively the model can be left to determine the optimum fit across all data streams and biological information available. At very least the sensitivity of the assessment outcomes to including or excluding (or down-weighting) each data-stream would be examined.

While there are no formal criteria presently available, beyond the classical statistical fit criteria, it would undoubtedly be helpful for keeping processes open and understood, if such more formal criteria were developed. This is not suggesting that the current less formal review of the applicability of an assessment be discontinued, but rather that at least some more formal aspects be recognized and made part of the RAG's routine so as to make communication and understanding simpler.

6.2 The Effects of Sample Size, CV, and Bias on Tiers 3 and 4

6.2.1 Tier 3

The Tier 3 assessment method and harvest strategy appears capable of achieving the Target Reference Point of 48%B_0 for Tiger Flathead (and similar species) for all levels of age sample size even when starting from low or high levels of initial depletion, but this is the case only if there is no or only slightly positive sampling bias (Wayte, 2009). If the sampling has a significant positive bias the outcomes ended either at the target or just below the limit if the stock started well above the target but can lead to missing the target in about 75% of occasions when the stock started below the target. This latter outcome was simply a reflection that the catches were badly over-estimated (see section 13.2 and 13.2.1 on page 61 for full details).

With School Whiting (and similar species) the Tier 3 was only able to achieve the target or remain above it if there was negative bias (see **Figure 23** on page 62). Positive bias led to depleted states and with the maximum positive bias the medians were effectively on the limit reference point. In all cases of different sample sizes and separately, with no bias the median depletion ended at or just below about 40% instead of the target of 48%. Once again, positive bias in the aging samples generated misleading outputs and undesirable management outcomes.

6.2.2 Tier 4

The effects on the Tier 4 applied to Flathead-like species, of the coefficient of variation (CV) of the CPUE and of bias tended to be highly exaggerated for the very high CVs of 0.6 and 0.9. These two levels led to time series of CPUE which were unlike any seen in the real fishery and so these levels should be ignored for Flathead (see section 13.2.3 on page 64 for full details).

With the remaining levels of CV, not surprisingly, as the CV increases so does the spread around the median levels of catch, the final depletion, and the probability of avoiding the LRP. The catch variation appeared to increase exponentially for CVs of 0.05, 0.15, and 0.45. The effect of positive levels of bias is to over-estimate the sustainable catches, which in turn leads to greater levels of depletion. If the stock starts already below the target then positive bias can force it down to or just below the LRP (see **Figure 25** on page 65). This is of concern as 'effort creep' would lead to positive bias and has undoubtedly occurred with the advent of GPS and colour depth sounders, etc., for which there is no information that can be included in any of the CPUE standardizations.

With the School Whiting like species, the effects of increasing CV on the Tier 4 outcomes was to increase variation in catches, but this species is already highly variable from year to year so the higher values of CV have less of an impact on the outcomes (see **Figure 26** on page 67). When the species starts in a depleted state then the final median depletion level would always be below the target reference point (TRP); this is, however, a function of the targets selected in the Tier 4 HCR. The Tier 4 HCR selects a reference period of years to identify the target catch and CPUE which drive the HCR. It is a mistake to believe these Tier 4 targets are at 48%B_0 rather than simply being a state of the fishery identified to be a good place to be in terms of sustainability and profitability; these targets are merely proxies and may in fact be above or below the equilibrium level represented by 48%B_0 (Haddon, 2014). With such an empirical harvest strategy that is based on data from the fishery and not the implied stock dynamics a proxy is selected. This proxy has been interpreted as a proxy for the biomass TRP of 48%B_0 but in practice it is a proxy for meeting the stock status of meeting the required target. If it happened to achieve 48% this would be simply by chance. The Tier 4 can achieve its selected target but there is no guarantee that this will in fact be at 48%B_0. Thus the outputs demonstrate that the method can generate consistent outcomes across a wide range of precision for both initially heavily and lightly depleted states (although with outcomes below the formal TRP and above the TRP respectively).

The effect of positive bias is very similar to that seen in the Flathead-like species. Strongly positive bias can lead to serious depletion and failure to avoid the LRP in the initially heavily depleted scenarios, although in the lightly depleted scenarios they all remain above the formal TRP.

The impact of positive bias is especially important as positive bias in CPUE could be brought about by improvements in technology and fishing practices. As it appears highly likely that such ‘effort creep’ will have occurred it would be valuable to further explore the possible impact of such positive biases in a more specific manner relating to the advent of events leading to increasing bias as a series of events across just a few years (e.g. the advent of GPS from 1990 – 1992). Having no way of taking such changes and resulting bias into account may be leading to overly optimistic views of each of the fisheries.

In Tier 1 assessments the effects of such bias would be detectable through the time series of CPUE or the ageing or length frequency data being inconsistent with each other. It is primarily the Tier 3 and Tier 4 methods that require further exploratory analyses.

6.3 Candidate Tier 5 Stock Assessment Methods

6.3.1 Introduction

The objective of data-poor and data-limited methods is to estimate a practical level of yield that is likely to be sustainable (MacCall, 2009). By ‘practical’, MacCall means commercial yields rather than overly conservative yields. Strictly all stocks can be considered data-limited so Bentley *et al*. (2014) suggest the preferred term should be data-poor.

The MSE testing highlights that a stock should never be assumed to be in only a lightly depleted state. A reasonable option if the assessment process (i.e. including the RAG’s involvement) determines that a particular stock has only been lightly fished, is to assume 40%B_0 so as to avoid the risk of over-fishing. This is, however, a policy decision and all that can be done here is to point out that, obviously, 40% would be more conservative and there would be a lower risk of over-fishing than selecting 48%B_0. However, there would also be a higher risk of failing to take as much catch as would be sustainable. Given the MSE results, 40% should certainly avoid the stock declining below the LRP (even if the stock were really depleted to the LRP and it was the DCAC method; (see **Figure 33** on page 75) where DCAC only just meets the < 10% probability of being below the LRP).

6.3.2 Central Tendency of Sustainable Catch Methods

Possible candidates for use in a new Tier 5 set of assessments include the methods that involve a measure of the central tendency of catches such as the average or median catch (possibly the 3rd highest catch). Ideally, these average catches would be estimated from periods of stability within each fishery, but in reality, in Australia, such periods are not common. Such central tendency methods involve empirical harvest strategies where the estimated central tendency catch constitutes the sustainable catch estimate (the 'assessment' is the decision rule; **Figure 60**). The recommended sustainable catches would need to be presented in the context of a weight-of-evidence appraisal of whatever stock was being considered. Dowling et al (2015a, b) discuss the use of such catches in the context of a set of catch triggers where a set of catch levels are set that, if met by the fishery, trigger management actions that can vary from a simple review of events to the application of some simple assessment or update of the average catch applied. In the Commonwealth HSP within the SESSF this would entail setting a multi-year TAC that might be reviewed for a breakout or major change each year and reviewed as to its level every few years.

Using a central tendency of catch estimate to set upper limits to catch before further management action, requires the assumption that the stock is currently in an acceptable state or that the catches already observed have not led to serious or undesirable levels of depletion. If the weight-of-evidence appraisal supports this assumption then a recommended biological catch can be made. Reasons for not using this approach include that the time series of catch data is not representative of the fishery (see **Figure 63** on page 109), or that the catch data is too sporadic to obtain a representation of the fishable stock. Specific trigger catch levels could then be set (Dowling *e al*., 2015b). Whether a discount would be required would depend on the final decision rule adopted. In the MSE testing the particular central tendency of catch was used (mean or median) but some other quantile could be used. The 3rd highest catch usually proved to be as capable as the other central tendency methods at avoiding the LRP, so an average or median should be sufficiently conservative as long as the state of depletion is considered to be acceptably far from the LRP at the start (see section 13.4 on page 93, and following pages).

The methods that used fixed estimates avoided the potential for a ratcheting down of catches that can occur in the strategies that include regular updating of the central tendency estimate. It is not the case that an allocated TAC will always be fully taken, especially with a by-product species that is not specifically targeted. If sustainable catch estimates are updated by using the mean or median of a time series that has an upper limit (a TAC) which is often not met, then the upper limit will automatically decline. Such catch estimates should be reviewed at five or ten year intervals in a weight-of-evidence context, especially if more information beyond catches has been collected, but otherwise the fixed methods have advantages over the dynamic or updated methods.

6.3.3 Model Supported Catch-Based Methods

The model supported catch-based methods include the Depletion-Corrected Average Catch (DCAC), the Depletion Adjusted Catch Scalar (DACS), and the Depletion-Based Stock Reduction Analysis (DB-SRA). Among these methods the DACS and DCAC are somewhat simpler to implement than DB-SRA (see section 13.4 on page 93, and beyond). Each of these also has assumptions and input requirements beyond having estimates of natural mortality. Fortunately, these input requirements are not especially strict or onerous and even when relatively strong assumptions are made (such as restricting the initial and final depletion levels to values that would be conservative) these methods can still generate solutions. The advantage of these model supported methods is that whatever estimate of sustainable catch is derived, it comes with an estimate of the uncertainty about the estimate (see **Figure 3** and **Figure 4**), so there is freedom in the harvest strategy to add further precaution if it is deemed necessary. This might depend on whether the RAG considered the catch time series used to be reliable. For example, earlier in the recent history of catches of Blue and Silver Warehou (*Seriolella brama* and *S. punctate*) the two species were not distinguished. For example, "… in 1992 both species were lumped under a global TAC of 4000t, 2000t of which was allocated to the trawl sector. Separate TACs were established in 1993 to avoid issue of high-grading spotteds [Silver Warehou] in favour of Blues." (Smith *et al*., 1994). Such potential flaws in the available catch data could be solved by eliminating the early data, although in the context of Blue Warehou, the early catches are verbally reported to have been large.

Assuming that no stock would be assumed to be initially well above the target (TRP), then no major consistent differences were observed between these three approaches. The DB-SRA provides more information than the other two methods and so if it can be implemented this would be the method of choice. But at least the DCAC should also be run to ensure that the estimates are not significantly different. A comparison of at least two of the methods should assist in discovering any unusual aspects of the available data as some reason would need to be found for any differences. The DB-SRA allocated a predicted depletion level although it does not necessarily hit the allocated value each replicate (see **Figure 4**). The trials run provide a spread of trajectories and the proportion that fall below the 20% depletion line in the final year would provide an indication of the relative risk of the predicted median MSY value of failing to meet the criteria of avoiding the LRP 90% of the time (see **Figure 59** on page 102).

6.4 Application of Tier 5 Methods to Current Fisheries

6.4.1 DCAC

For each species the distribution of sustainable catches was skewed to the left, which reflected the uncertainty that derives from the various assumptions made about the biology and the production model representing the stock dynamics. Nevertheless the estimates of sustainable catches were 2153t, 934t, and 1439t for Flathead, Jackass Morwong, and School Whiting respectively (**Figure 3**).

The depletion-corrected average catch estimation of a sustainable catch invariably led to a reduction relative to the simple average catch over the same period used in the estimation (see section 13.5.2 on page 96); the correction over the average catch for depletion was relatively minor being 137 t for Flathead, 110 t for Jackass Morwong, and 145 t for School Whiting from 1980.

School Whiting was a special case as catches were only minor from 1947, when records begin, to 1980, but have been noisy but higher since. Once again, the assessment process already in place means that the RAGs would need to agree on what time series of catches to use when estimating the sustainable catch. The spread of the sustainable catch estimates in each case were not intended to be an estimate of MSY or other recognized biological reference point. Rather the intent was to estimate a practical level of yield that is likely to be sustainable (MacCall, 2009). By 'practical', MacCall means commercial yields rather than overly conservative yields.

While the adjustment appears relatively minor being about 6, 10.5, and 9% of the average catch these are certainly conservative adjustments. With longer lived animals (e.g. Ocean Perch or Redfish), which have lower natural mortality rates, the adjustments are likely to be greater. This method does not provide an estimate of stock depletion, however, so its capacity to avoid the Commonwealth LRP can only be tested using MSE.

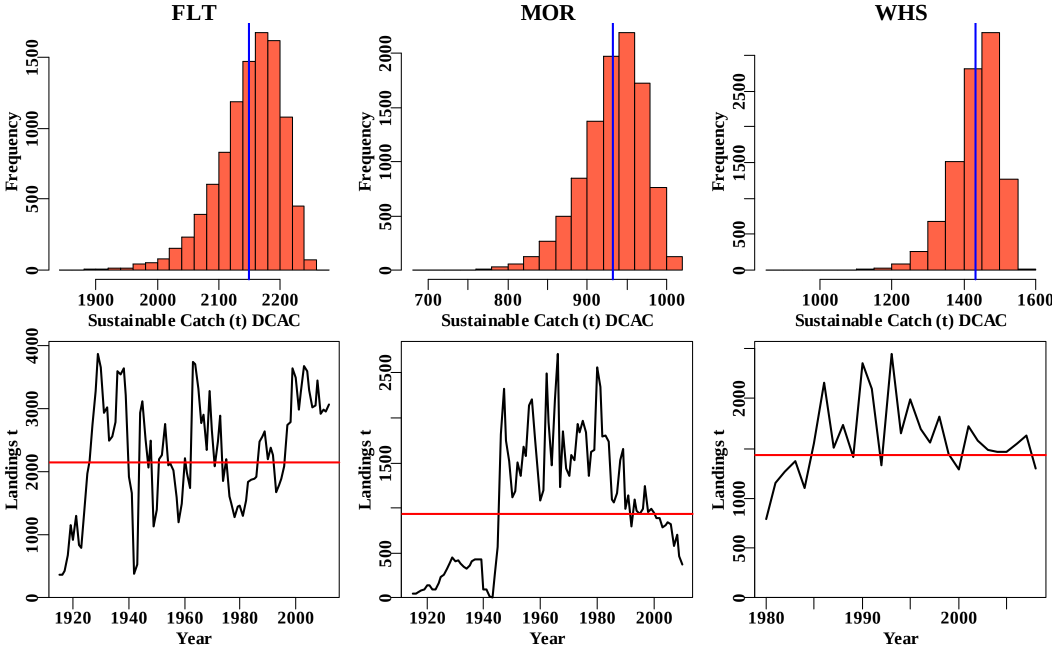

Figure 3. The distribution of sustainable catch levels (top) and mean time-series (bottom) from 10,000 replicate estimate using the DCAC method on Flathead (FLT), Jackass Morwong (MOR), and School Whiting (WHS). The median estimate of the sustainable catch is depicted by the blue line in the top diagrams and the mean estimate for each species are reflected on the catch histories as red lines (2153t, 934t, and 1439t). The assumption was made that the stock was at the target biomass in all these assessments.

6.4.2 DB-SRA

The Depletion-Based Stock Reduction Analysis is more flexible than the DCAC as it can provide estimates of MSY, B_{MSY}, F_{MSY} and depletion levels. These are dependent upon the production model used but nevertheless this enables the method to generate estimates that can be directly interpreted by the HSP. The application of the DB-SRA method is rather more time consuming so instead of 15 seconds, as for the DCAC, running 10,000 replicates to characterize a single species takes between 15 – 20 minutes (see **Figure 55** and section 13.5.3 on page 99).

None of these catch-only methods should be applied blindly but rather only when taking into account all else known about a fishery; this would be important with Jackass Morwong for example. The use of a weight-of-evidence approach enables the assumptions made to be defensible and possibly testable given sufficient resources.

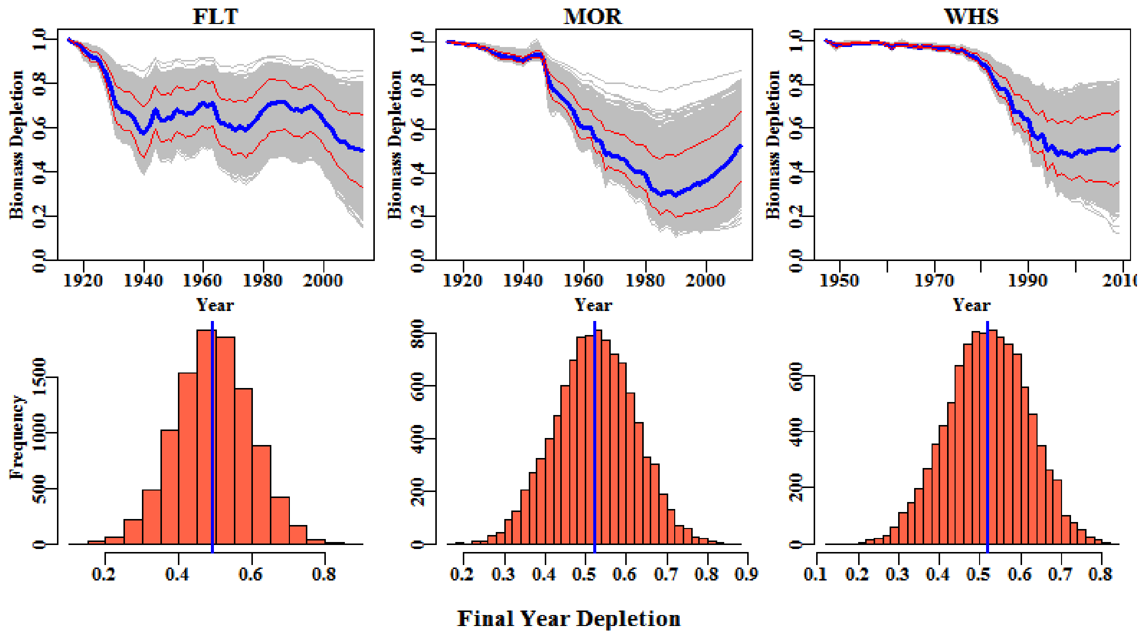

Figure 4. The spawning biomass depletion estimated using the DB-SRA methodology for Flathead, Jackass Morwong, and School Whiting. The blue lines are the median values and red lines are the 95% bounds on the spread of the replicates. The lower graphs represent the frequency distribution in the final year.

Outputs important for management include the MSY, F_{MSY}, and the depletion level, each with estimates of uncertainty included (**Figure 5**). The inclusion of the uncertainty estimates means that any harvest control rule developed for use with the DB-SRA assessment method can attempt to take into account the uncertainty included in the biological parameters and the catch time series used to estimate the management outputs.

The data-poor method MSE testing used only 10 years for the initial estimates of sustainable catch, which constitutes a strong test of the utility of DB-SRA as the method usually requires a long time-series of catches. In reality, for many of the SESSF species longer time series of catch and CPUE are often available. The DB-SRA was run on the Flathead data (1915 – 2012) sequentially removing 10 years at a time from the time series. The effect of decreasing the length of the time series on the estimate of MSY is only minor up until about 1975 but it is in 1985 and onwards where first the spread of possible values increases so that smaller MSY estimates make an appearance, but then in 1995 the mean drops from between 2800 and 2750t down to about 2550t with a shift downwards in the distribution (**Table 2**).

The analyses in the DB-SRA all assumed that the final depletion level in 2012 was 48% and the outcome from the analysis is sensitive to the selected depletion level (see **Fig-**

ure 59 on page 100), which is consistent with conclusions in Wetzel and Punt (2011) and Carruthers *et al*. (2014). For example, when the depletion level in the final year is assumed to be 35%B_0 instead of 48% the MSY output has almost half the range of values (2168 – 3067 rather than 2344 – 4050) of the analysis at 50%B_0, with more difference between the upper bounds and the lower and a 260 t difference between the MSY estimates (**Table 2**). In practice, if this method were used, the selection of the assumed depletion level is a decision that would need to be defended explicitly and agreed upon by the full RAG.

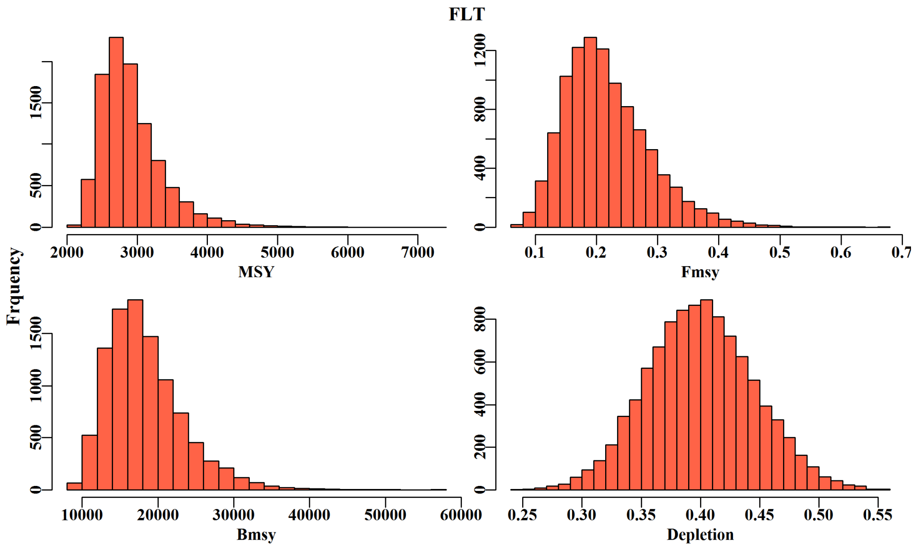

Figure 5. The outputs from the DB-SRA applied to Flathead over the entire catch history using 10,000 replicate runs of the Monte Carlo simulation.

Table 2. The start year, number of years of catch data, and the quantiles of the estimates of the MSY. Note the shift to lower values in the last three rows and the difference relative to the average catches over the same period. In all cases except the last row the assumed final depletion was 48%; in the last row is the outcome from an assumption of 35%.

Start	Years	2.50%	5%	50%	95%	97.50%	Average C
1915	98	2344.537	2382.149	2831.794	3776.656	4050.090	2290.357
1925	88	2342.034	2381.012	2831.601	3776.654	4048.268	2457.239
1935	78	2327.350	2373.559	2827.719	3776.633	4038.496	2404.397
1945	68	2306.799	2364.283	2823.057	3776.583	4031.961	2401.544
1955	58	2267.541	2350.947	2816.476	3775.969	3990.172	2434.862
1965	48	2234.215	2310.502	2802.807	3763.108	3983.085	2486.521
1975	38	2168.726	2255.357	2773.696	3725.866	3943.827	2440.500
1985	28	2067.754	2180.506	2733.887	3668.618	3917.452	2739.179
1995	18	1828.630	1934.611	2569.098	3442.084	3709.652	3051.000
1915	98	2168.842	2200.732	2426.595	2907.005	3067.365	2290.357

Both the DCAC and the DB-SRA require an assumption about the level of depletion in the final year (or a particular year) but with data-poor species this will invariably be a very uncertain value. However, the MSE testing does suggest the outcomes possible if different relatively precautionary depletion levels are chosen (at the risk of under-catching a resource). Despite this the DB-SRA methods holds some advantages over using simple average catches as even with only a few year's data and relatively high natural mortality rates the outcome differs from simply using the average of median catch (**Table 2**).

7 Implications

The costs, in terms of both time for development and in running such assessments, for the seven data-poor methods considered would be relatively minor, with the DB-SRA taking the most time and requiring some code development to simplify or automate its application to standard data extracts. But once parameter files were set up containing the necessary biological and fishery information these would remain stable and only the catch data input files would require updating each year the assessments were conducted. Changes in the fishery would be expected to be relatively slow for many species so these stock assessment methods are good candidates for producing multi-year TACs; especially the central tendency of sustainable catch methods. Catch data are generally well recorded in the SESSF, although there are likely to be variable levels of discards for minor and by-product species so estimates of these would certainly be required. Discard rates would mainly be an issue if they were variable through time or possibly if further regulations were introduced concerning discards.

As long as no stock is assumed to be only lightly depleted the central tendency of sustainable catch methods tend to under-estimate the possible catches and that appears to be the major risk with applying these methods. However, for by-product or minor species for which landings are already small, this may not be a major issue. Nevertheless, the potential for under-catching should be kept in mind when a RAG determines at what interval to review the catch levels.

The primary implication for the SESSF is that there are now tested alternatives to Tiers 3 and 4, which can be applied to relatively minor and by-product species where the application of Tier 3 or 4 methods do not appear to be valid. This will make the assessments for such previously doubtful or uncertain species more defensible.

8 Recommendations

8.1 Appropriateness of Tier Harvest Strategies

When reviewing a stock assessment within each RAG it should become routine that the Tier selected for the assessment be justified or defended in terms of how well the species and the available data meet the assumptions of the stock assessment method applied. If answers are provided for the questions 'Why shouldn't this species be assessed at a lower or a higher tier?' and 'Are the structural assumptions in the assessment method used reasonable for the species concerned?' then the RAG will be in a better position to accept or reject an assessment as being appropriate or not.

As soon as suitable Tier 5 methods are available for selection then decisions need to be made in the Slope, the Shelf, the Shark, and the GAB RAGs as to whether any of the species being taken in their fisheries are in an inappropriate tier and one of the Tier 5 methods selected at least for comparison with current management advice. For example, in the Shark RAG Elephantfish (*Callorhinchus milii*) and Sawsharks (includes *Pristiophorus cirratus, P. nudipinnis,* and *Pristiophoridae*) are undoubtedly bycatch (more strictly by-product) species in the Gummy Shark fishery which are rarely if ever targeted as there is a limited market and they are of relatively low value. Both are currently assessed using a Tier 4 approach, but being bycatch when fishers are targeting a different species questions are always raised as to the validity of using their CPUE. Both of these fisheries would be candidates for a Tier 5 assessment approach. Similar arguments can be made in the other RAGs, though perhaps not the GAB RAG.

8.2 Potential Tier 5 Approaches

When assessing a data-poor fishery using a data-poor assessment method that requires an approximate initial depletion level, no stock should be assumed to be in any state of depletion better than the target of 48%B_0, even if the species has not been targeted in a mixed fishery previously. It can be assumed that in a mixed fishery a by-product species would have been exposed to at least some fishing mortality even if previously the species was not landed to any great extent. Extra analyses can be conducted that explore the importance of the initial depletion level assumed.

For those species where there are limited catch data and little other information available, and yet an estimate of a Recommended Biological Catch is required, it is recommended that methods based on estimating the central tendency of sustainable catches (such as the median catch, average catch, or 3rd highest catch) be used. Those using at least a ten year period of catches, preferably from a relatively stable period within a fishery, and without later update, or only updating at long intervals within a larger context of a weight-of-evidence across that fishery, are to be preferred. This should avoid any artificial ratcheting down of TAC levels that can happen through the fishery dynamics affecting catches rather than the stock dynamics.

For those species for which there is some biological or other data available that would permit the application of model-assisted data-poor stock assessment methods then it is recommended that either the Depletion-Corrected Average Catch (DCAC) or the Depletion-Based Stock Reduction Analysis (DB-SRA) be applied. The DB-SRA has some advantages in its application within the Commonwealth HSP, but both perform acceptably well as long as a stock is not assumed to be in only a lightly depleted state.

These two recommended methods should not be the end-point for the model-assisted methods adopted within the Tier 5 harvest strategy. The Catch-MSY method and any derivatives from it should be explored further as this appears to be more flexible and less demanding in terms of assumptions than the DCAC or DB-SRA. As data-poor methods and harvest strategies are developed both in Australia and elsewhere they should be considered for adoption here if they MSE testing demonstrates they constitute an improvement on the Tier 5 approaches already tested.

8.3 Communication of New Methods

The examples already generated in the 'Application of Tier 5 Methods to Current Fisheries' section, where some of the new data-poor Tier 5 candidate methods were applied to three different species, should be used in the explanatory material to be presented to RAGs and MACs and other interested stakeholders. As the latest catch data becomes available this year further examples using the fullest data sets should be developed to make any presentations directly pertinent to each RAG. The technical details of the methods will be extracted from published material along with the computing algorithms needed to conduct the calculations required (expect where these are trivial – as in 3rd highest catch over the last 10 years) and recorded in more formal documents to be included in different RAG materials.

This communication will be important in progressing the adoption of these methods, which are new to the Commonwealth (and the States) in Australia. For them to be formally adopted into the toolkit of methods acceptable to the harvest strategy policy they need to be accepted, as a minimum, by the SESSF RAG, the SEMAC, and the FAM Board. The various sub-ordinate RAGs also need to accept their utility before their adoption.

8.4 Further Development

Data-poor methods are continuing to be developed here in Australia and elsewhere, especially in the USA and Europe (ICES). These developments should be monitored to ensure that any improved methods developed are considered and adopted here in Australia.

9 Extension and Adoption

The extension of the work and methods presented here is a formal objective of this project and will entail making presentations and explanations to the RAGs and MACs concerned along with other interested stakeholders. Formal explanatory documents describing the methods in detail, based on published literature, will also be presented.

The candidate methods, with some constraints on the assumed initial depletion levels, especially those currently recommended, have been shown to perform in compliance with the HSP requirement of keeping above the 20%B_0 limit reference point better than 90% of the time. There should therefore be no problem with their adoption into the Tier toolkit available in the SESSF. There is no reason apparent why they could not be adopted in other fisheries should those fisheries require such tools.

10 Appendix: Tier 5 Options in the SESSF

10.1 Introduction

10.1.1 Stock Status and Uncertainty

Ideally it would be possible to obtain a measure of the stock status for all fished species so that their management, even when within a mixed fishery, could be balanced across any competing objectives that may have been devised for different stocks or single species fisheries within a mixed fishery (e.g. maximizing catch while remaining sustainable). Such ideal situations are often described to highlight how reality differs from the ideal. In reality, it is the case that even with well documented fisheries stock status cannot be measured directly. We are always limited to making inferences about stock status by taking samples and observations from a fishery itself, and using those samples and observations in some form of stock status assessment. The use of samples implies that it is only ever possible to obtain an uncertain estimate of a stock's status. The sampling can be improved and the development of long time-series of fishery statistics, such as catch rates, catches, age-structure etc., can certainly assist with improving the precision and reducing any bias in estimates as well as increasing our understanding of the dynamics of a given stock. However, the value of such data is dependent upon its quality and representativeness for a fishery; there always remains a degree of uncertainty in any stock assessment and this is especially the case in many data-poor or data-limited fisheries (Haddon *et al*., 2005; Vasconcellus and Cochrane, 2005; Pikitch *et al*., 2012). Nevertheless, fishery managers are required to make decisions in the face of such uncertainty, although it and its implications have not always been well recognized. Recently, around the world, countries and organizations, such as ICES in Europe, the USA, Australia, and New Zealand, which have active and responsive fisheries management, have attempted to account for uncertainty explicitly.

10.1.2 Harvest Strategies

A number of very influential documents for fisheries management were published by the FAO in the mid-1990s, including: the *Code of Conduct for Responsible Fisheries* (FAO, 1995), the *Precautionary Approach to Capture Fisheries* (FAO, 1996), and *Fisheries Management* (FAO, 1997); these latter two documents being parts of the *Technical Guidelines for Responsible Fisheries* series. "Long term management objectives should be translated into management actions, formulated as a fishery management plan or other management framework" (FAO, 1995, p 11). The *Guidelines* appear to be one of the first documents to describe the components of what are now referred to as Harvest Strategies. The need for *targets*, described as the desired outcomes (or desirable state) for a fishery, *limits*, described as undesirable outcomes that are to be avoided, and *harvest control rules* which specify in advance what action(s) should be taken when specified deviations from the operational targets and limits are observed, were all identified explicitly (FAO, 1996; Caddy and Mahon, 1995; Caddy and McGarvey, 1996). Early work on simulation testing of management arrangements (now known as management strategy evaluation; Butterworth and Bergh, 1993; Punt *et al*., 2014) appears to have contributed to this approach to describing harvest or management strategies. Thus, in the *FAO Guidelines* it defines a *management procedure* as a description of the data to collect, how to analyze it, and how the analysis translates into actions. This is the standard way to describe a modern harvest strategy: define the data needed, the analysis of the performance measures (that are used to determine status relative to target and limit reference points), and the control rules used to generate management advice (**Figure 6**).

10.1.3 Tiers of Harvest Strategies

Many of the ideas within fisheries management reflect a focus on the major and most valuable target species, which, because of their value and importance to the fishing industry and markets, tend to have relatively extensive data collections to assist in their stock assessment. Especially in mixed fisheries there tends to be an informal hierarchy by which the main targets or most valuable species, that is the primary drivers in a fishery, gain most assessment attention, then byproduct species, that tend to be landed as opportunity and markets arise, gain some attention, and finally bycatch species that are invariably discarded, rarely receive much attention, if any. With the growth of the use of more formal harvest strategies this hierarchy between grades of species has been put into a categorical scheme of tiers in which the lower the tier (e.g. Tier 1) the more detailed and extensive the available data and hence, usually the more detailed the stock assessment that is possible (Smith *et al.*, 2008). Tiered schemes of harvest strategies were first developed and implemented in the Gulf of Alaska in 1998 (DiCosimo, 2001). The more sophisticated stock assessments provide options for more sophisticated harvest control rules (HCR) that rely on model derived stock performance measures, such as spawning stock depletion (relative to the unfished state).

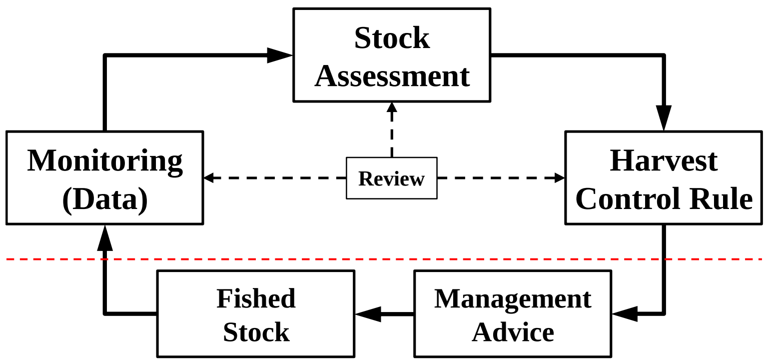

Figure 6. Diagrammatic representation of a standard harvest strategy (everything above the red line) depicting the sequence used when it is implemented. The review stage is there to facilitate adaptive management and modifications should they become necessary.

10.1.4 Guidelines for Selecting a Tier for a Fishery

In Australia, the harvest strategy policy (HSP) and its associated harvest strategies (HSs) were implemented in 2007, however, before that, in 2005, a detailed HS was introduced into the SESSF (Smith *et al.*, 2008). The tiered system of HSs in the SESSF formed a template for the more general Commonwealth HSP. When the original HS for the SESSF was developed the approach was to consider the available data and determine from that what assessments and associated harvest control rules were possible and apply those. The continuation of this approach has led to the status of a number of species being assessed in tiers which are now considered inappropriate for them as a result of the available data either being of insufficient quality or failing to represent the stock as a whole. For example, Blue-Eye Trevalla (*Hyperoglyphe antarctica*) is assessed using a simple Tier 4 HS (Little *et al.*, 2011; Haddon, 2014) but the validity of this assessment is questionable for reasons of data quality (catch per day is used instead of catch per hook, which has been shown to bias the outcome) and for reasons of representativeness. Blue-Eye Trevalla populations exhibit such a high degree of spatial heterogeneity in their biological and fishery characteristics that assessing the whole stock on the basis of a single area is highly likely to bias any outcome.

Given a set of available data more than one assessment method could be applied but rather than simply applying the most detailed possible, a more defensible approach would be to determine how well an assessment could estimate the required performance measures used to determine stock status. If this had been done originally there would have been no need for the first objective in this project:

Establish guidelines, using SESSF case studies, for when the particular Tier harvest strategy for a given stock becomes inappropriate and make explicit recommendations as to what response would then be appropriate.

Demonstrating whether a data set is representative of a fishery is not a simple process as it relates to the details of the sampling that has been done, which, in a mixed fishery extending over a very large geographical area such as the SESSF, is always something of a compromise (**Figure 7**).

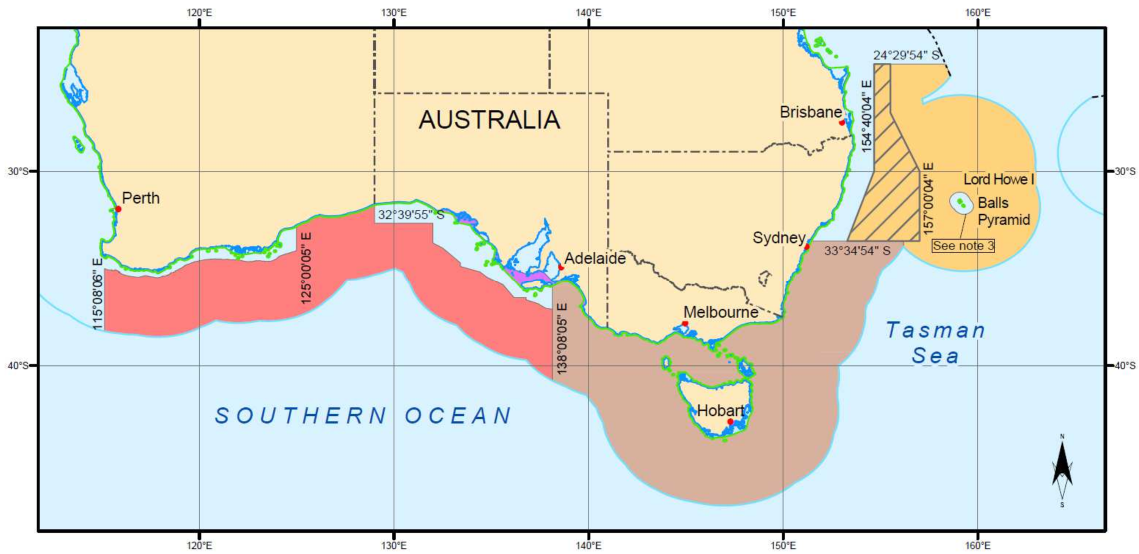

Figure 7. The geographical extent of the SESSF with the trawl fishery subdivisions illustrated red = GAB, brown = Commonwealth trawl, and yellow = East Coast Deepwater trawl. The hatched area in the yellow region is closed to trawlers, although multiple line methods are used there.

Determining whether an assessment can provide valid estimates of the desired performance measures can be done by a consideration of the data available, how and where it was sampled, and the assumptions lying behind the assessment method or HS (**Figure 8**, **Figure 9**, and **Figure 10**. For example, application of the Tier 3 HS (**Figure 9**) will generate outcomes with an unknown degree of bias if applied to data derived from a fishing method that is characterized by a dome shaped selectivity curve rather than a classical logistic selectivity curve (most gill-net fisheries, for example for Blue Warehou, *Seriolella brama*, exhibit dome-shaped selection). The production of the guidelines to meet objective 1 will proceed exactly through a consideration of the data available and the assumptions behind how they were gathered plus a consideration of the assumptions behind the various assessment methods currently used within the different SESSF tiered harvest strategy.

10.1.5 The Australian Harvest Strategy Policy

Australia has numerous different fisheries; although none are large in volume by world standards, some, especially the invertebrate fisheries, are relatively valuable. This limited productivity is a reflection of Australia's geographical location and great age. The generally low productivity of Australian fisheries reflects three things: 1) the low run-off of nutrients from the generally dry and previously eroded ancient continent, 2) the fact that most major coastal current systems flow south from nutrient-poor tropical waters, and finally 3) the small number of permanent areas of upwelling from deeper nutrient rich waters (Haddon, 2007). This diverse range of fisheries constitutes a serious challenge for the specification of a Harvest Strategy Policy that can apply to all Australian fisheries. The Commonwealth introduced a formal harvest strategy in 2007 (DAFF, 2007), while within the States there are attempts to develop formal harvest strategies, with South Australia being the most advanced (Flood et al., 2014).

The Australian Commonwealth harvest strategy policy (HSP) defines its targets and limits in terms of spawning biomass depletion levels or proxies for these measures. Thus, the limit reference point is set at 20%B_0 below which all targeted fishing is supposed to stop; the argument and justification behind this limit being that the risk of recruitment overfishing becomes unacceptable below this level of spawning biomass (Beddington and Cooke, 1983):

> *"… an escapement level of 20% of the expected unexploited spawning stock biomass is used. This is not a conservative figure, but it represents a lower limit where recruitment declines might be expected to be observable."* (Beddington & Cooke, 1983)

The value of 20%B_0 is currently a proxy for $B_{MSY}/2$, but as B_{MSY} is recognised as being extremely difficult to estimate with any precision, recommendations have been made to accept the proxy as the specific limit (Haddon et al., 2013). It is important to be clear that being below 20%B_0 does not automatically imply that recruitment overfishing will occur, the limit reference point is merely the level selected to act as a general indication that a fishery is not performing as well as it should and that below that level the risk of significantly reduced production is assumed to increase.

The target reference point was set to be the biomass that leads to the maximum economic yield (B_{MEY}) which is defined using a proxy as 1.2 x B_{MSY} (where B_{MSY} is the spawning biomass that should give rise to the maximum sustainable yield at equilibrium). This combination implies, given the assumptions in the HSP (which uses a proxy of 40%B_0 for the B_{MSY}), that the target is 48%B_0. A description and justifications for the selection of these values is provided in Haddon *et al*. (2013).

10.1.6 The SESSF Harvest Strategy

Most fisheries in Australia only use a single harvest strategy in their management but some fisheries use one of a range selected from within a tier system. The South East Scalefish and Shark Fishery (SESSF) has a tier system made up, in theory, of four tiers (Smith *et al.*, 2008; DAFF, 2007). The Tier1 harvest strategy (**Figure 8**) applies to stock where there is a robust quantitative assessment that provides an estimate of the current spawning biomass so the biomass related limit and target reference points can be used directly (Tuck, 2014). The Tier 2 relates to a less robust quantitative assessment, and this was originally proposed with a somewhat higher target than the Tier1; in practice there are no Tier 2 species in the SESSF (and the primary target of 40%B_0, listed in Smith *et al.*, 2008, originally only for the SESSF, was increased on the introduction of the Commonwealth HSP). The Tier 3 harvest strategy (**Figure 9**) requires a good estimate of natural mortality and an estimate of current fishing mortality (using a modified catch curve; Wayte and Klaer, 2010; Klaer, 2014). Finally the Tier 4 HS (**Figure 10**) would apply to those stocks with only information on catches and catch rates (cpue; Little *et al*, 2011; Haddon, 2014). Currently there are no other tiers should Tiers 1, 3, and 4 not be suitable.

10.1.7 Are Tiers Hierarchical in relation to Uncertainty?

The use of a tier of harvest strategies is an attempt to recognize that as the data availability for different fisheries differs so will the degree of uncertainty related to any management advice deriving from any assessment of the available data. The plausibility of the different tier outcomes has previously been related to the data and methods available to be used to analyse the stock performance measures. The assumption is that the closer the analysis is to the underlying dynamics of the populations being fished the more likely it is that the outcomes will reflect true events. This implies that the Tier 1 HS, which includes a formal mathematical model of the population dynamics behind the fishery (**Figure 8**), would be expected to produce an assessment with the least uncertainty. The Tier 3 assessment was generally considered next best in terms of likely uncertainty because it used catch curves and as these use ageing data from the fishery this should reflect the reality of the dynamics at least to a limited extent; the assumption is generally made that the ageing samples are representative of the whole fishery each year (**Figure 9**) and that the assumption of equilibrium does not bias the outcome too much. Finally, the Tier 4 assessment, which only uses catches and cpue (**Figure 10**) was originally considered the least certain of the assessments even though cpue was usually the only index of relative abundance and it was also used in the Tier 1 assessments.

The Tiers are thus intended to reflect the growing uncertainty to be expected from the various assessments possible within each Tier level. Thus when sufficient data are available to apply a Tier 1 HS then it would also be possible to apply Tier 3 and Tier 4 HSs. The assumption has been made that because the Tier 1 would generally be less uncertain than the other two, it would be preferred. If only a Tier 3 or Tier 4 is possible the idea of compensating for the assumed increase in uncertainty arose (Dowling et al., 2013). In an attempt to compensate for the increasing uncertainty assumed to occur with higher tiers, a system of discounts are supposed to be applied to the recommended biological catch levels (RBCs) predicted by each Tier; no discount to the Tier 1, a 5% discount on the Tier 3 RBC, and a 15% discount on the Tier 4 RBC. The intent of these discounts is to attempt to reduce the risk of the recommended catches being biased high in accordance with the increased levels of uncertainty assumed to occur in different assessment methods and harvest strategies. While this discounting principle is simple to understand (a balancing of risk against catch), demonstrating that the different assessment methods used have the perceived relative degrees of risk requires detailed simulation testing. Fay *et al*. (2013), in a preliminary study, have demonstrated that the relative risks can be greatly affected by what appear to be small details in the different harvest control rules. For example, without the meta-rule that limits annual changes to the TAC for a stock to no more than 50%, the Tier 3 harvest strategy does not necessarily perform better than the Tier 4 harvest strategy. With the meta-rule then the ordering is as might be expected with the Tier 3 generally out-performing the Tier 4, although, importantly, the particular outcome is also species and stock dependent.

As stated by Fay *et al*. (2013, p 1):

> *As the outcomes were variable across the species, the harvest strategies, and the methods used to implement precaution, it is not possible to provide a simple conclusion that a single optimum method exists for balancing risk against uncertainty for each Tier level of assessment.*

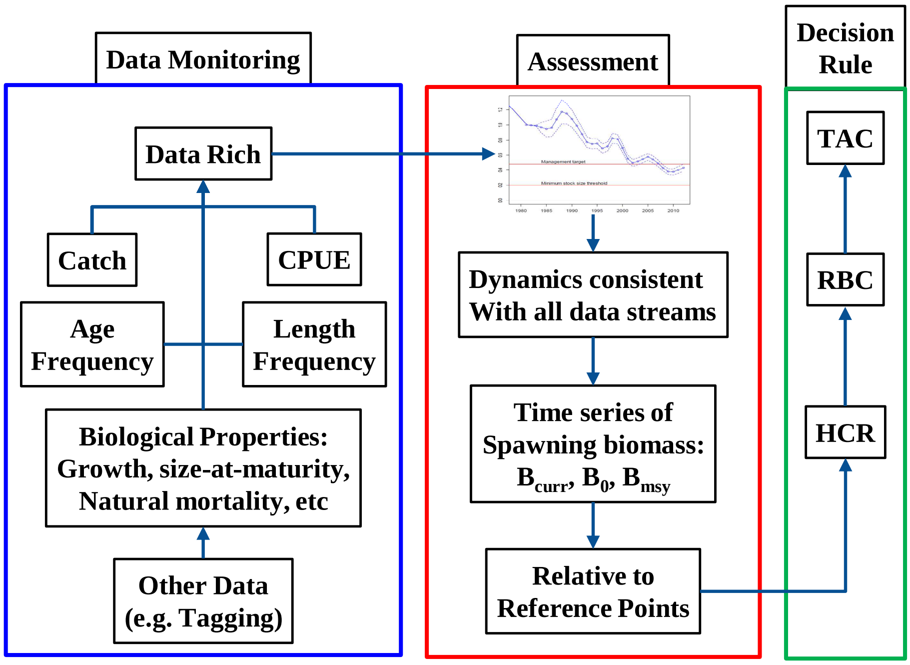

Figure 8. A diagrammatic representation of a Tier 1 analysis as implemented in the SESSF, which involves an integrated age-structured stock assessment model.

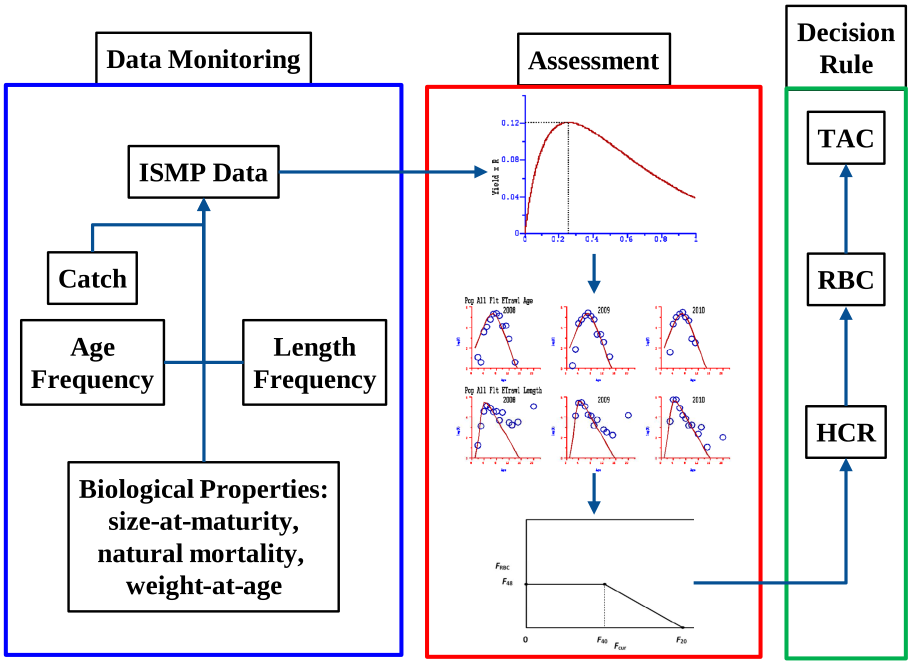

Figure 9. A diagrammatic representation of a Tier 3 analysis (basically a yield-per-recruit plus a modified catch-curve) as implemented in the SESSF.

Fay *et al*'s (2013) work implies that the SESSF Tier system is a convenient name but that the different harvest strategies beyond Tier 1 do not form a tidy hierarchy of certainty. The Tier 1 and 3 approaches attempt to use biological data from the fishery concerned to gain insight into the biological dynamics of the population and/or fishery. The Tier 4 only uses empirical data direct from the fishery statistics. The Tier 4 is thus an empirical harvest strategy that makes no attempt to mimic the stock dynamics in an attempt to understand events within the stock. Instead it assumes that cpue provide a valid measure of the stock status and it uses its relative value through time within a defined harvest control rule to provide management advice (**Figure 10**). There is no automatic reason why an empirical harvest strategy cannot perform perfectly well as long as the performance measure selected (in this case cpue) really does provide a valid index of relative abundance (or at least relative stock status) through time. If it does this successfully, and the target cpue is selected well then a Tier 4 can perform almost as well as a Tier 1. Indeed, if cpue were actually a valid index of relative abundance, and was thus capable of providing useful management advice then a surplus production model, which only uses catches and cpue, could be a useful example of a Tier 2 analysis, (Haddon, 2011).

The key conclusion in Fay *et al*., (2013) is that the degree of risk appears to be idiosyncratic to each particular species.

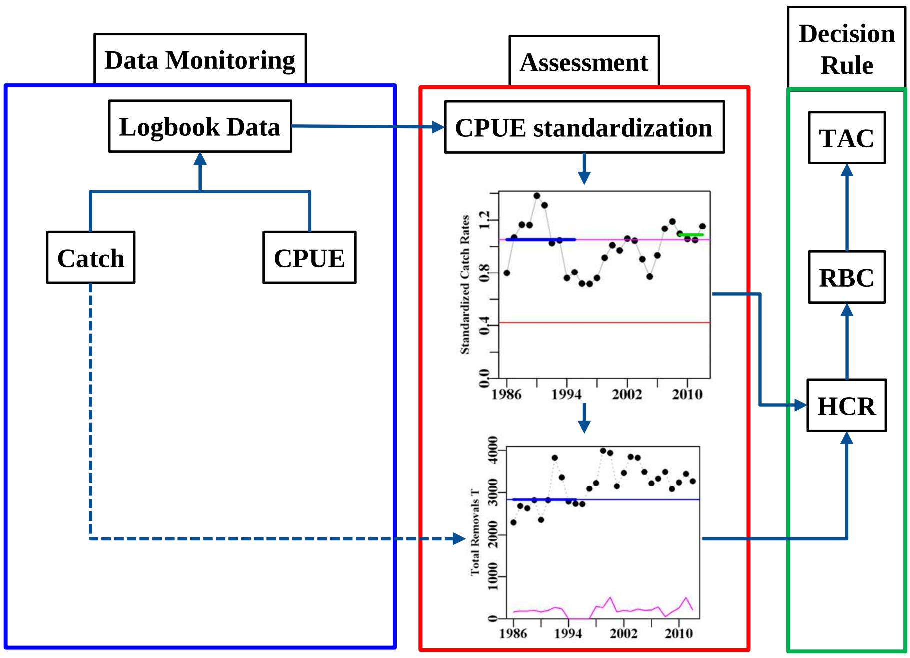

Figure 10. A diagrammatic representation of a Tier 4 analysis as implemented in the SESSF, which involves an empirical consideration of the ratio of current cpue with a specified target.

10.1.8 Ecological Risk Assessment

In addition to the current Tier system of harvest strategies, the Commonwealth has also implemented a system of Ecological Risk Assessments (ERA) in an attempt to document the potential risks to all other species potentially affected by fishing pressure (Smith *et al*., 2007; Hobday *et al*., 2011). These risk assessments aim to approach the needs of ecosystem based fisheries management (EBFM) which requires at least a review of the effects of fishing on non-target species. This has proven effective for many bycatch species which are never or very rarely landed. However, the use of the ERA doesn't provide sufficient management details and advice on sustainable catch levels for those minor species (byproduct species) which are landed as catch but are generally taken in association with primary target species.

10.1.9 Between Tier 3 or 4 and the ERA

In addition to the minor or byproduct species, there are numerous species and fisheries, within the SESSF, for example, to which attempts are made to apply the Tier 3 or Tier 4 harvest strategies and that fail to provide valid advice for an array of different reasons; generally all of these reasons relate to a lack of appropriate information or a failure of some underlying assumptions in the assessment methods used. These examples are from relatively data-poor or data-limited species, and fisheries for data-poor species may be defined as those for which i) a quantitative stock assessment cannot be undertaken because of limitations in the type and/or quality of available data (Haddon *et al.* 2005; Kelly and Codling 2006; Dowling *et al.*, In press). Data-poor does not only mean that data are lacking, as there are many other reasons that a given fishery can be considered data-poor or data-limited; these can include: a) new fisheries with limited observations, b) low value fisheries, c) multi-gear, multi-species fisheries with many small operators or landing sites, d) data quality is low or variable and difficult to verify, and e) spatially structured fisheries where samples collected may not be representative of the whole stock (Haddon *et al.* 2005; Dowling *et al.*, 2015a, b).

Examples where the Tier 4 is applied in the SESSF in situations and to species where its assumptions are broken include species such as all the deep-water Oreo species (*Pseudocyttus Maculatus* – Smooth Oreodory, *Allocyttus verrucosus* – Warty Oreo, *Neocyttus rhomboidalis* – Spikey Oreodory, *Neocyttus psilorhynchus* – Rough Oreodory, and *Allocyttus niger* – Black Oreodory; Haddon, 2014) but also such species as Blue-Eye Trevalla (*Hyperoglyphe antarctica*). In the case of the Oreo species this is often a highly mixed fishery where the species is often simply reported as 'Oreo Dory' rather than particular species. Where the species caught is identified, the catch rates can vary between trivially low to enormously high, depending on the strategy used in trawling. If an aggregation is fished or targeted the cpue can be expected to be very high, but if trawling is merely prolonged, covering an extensive distance, then cpue can range from high to extremely low. This variation means that the analyses become so uncertain as to remain uninformative, and worse can provide misleading management advice. In the case of Blue-Eye Trevalla this is a species with a highly fractured and patchy spatial distribution of the adults, stretching from east coast sea mounts up at -20°S, down to the southern Cascade Plateau at -43.883°S. Even though there is an array of biological and fisheries data that have been collected (Klaer et al., 2014), each area appears to have idiosyncrasies rather than being characteristic of the stock as a whole and no area has been consistently sampled. Despite these impediments, these species, and other species for which it is inappropriate are still considered to fit into the Tier 4 harvest strategy because the catches and cpue are the only data available; currently the unavailability of a higher tier that can provide management advice in terms of catch limits prevents other actions. Applying the ERA would not provide the detailed management advice required for some relatively important quota species.

To achieve objective 1, an array of criteria are required for determining which Tiers it is possible to use for a given species. For those species which should not be considered under the current Tiers there is also a need for a different class of harvest strategies (made up of data required, assessment, harvest control rule) that will fill the gap between the current Tier system and the ERA approach i.e. a set of possible Tier 5 harvest strategies. Hence the two main objectives of this project.

10.1.10 Would a Tier 5 fit into the Harvest Strategy Policy

There is no impediment in the current HSP for the inclusion of new harvest strategies as long as they operate to achieve the intent of the current and future HSPs. This was demonstrated in the Reducing Uncertainty in Stock Status project (Dowling, 2011; Haddon, 2011b; Haddon 2012; Plagányi *et al*., 2013). The fundamental intent of the HSP is to prevent over-fishing and to prevent a stock from being over-fished. As a minimum if a harvest strategy can maintain a stock's biomass above a minimum limit threshold for more than 90% of the time and was successful at limiting fishing mortality when a stock was in a low state, then it could be said to be successfully achieving the intent of the HSP (Haddon, 2012). This would need to be demonstrated, preferably using management strategy evaluation, but other than that the development of alternative higher order Tier harvest strategies should be acceptable. In data-poor stocks the priority becomes one of first avoiding the limit reference point (or, in the case of data-poor species, its proxy) and achieving the proxy target comes second if it is even possible (Haddon, 2011b). The HSP explicitly recognizes that in mixed fisheries it may not be possible to maintain all species at MEY but in all cases all species should be kept above the limit reference point (or its proxy). For data poor species, where catch or landings are the only data readily available, then avoiding the limits may be all that can be successfully or defensibly achieved.

Demonstrating that a new strategy meets the requirements of the HSP strictly requires the use of Management Strategy Evaluation (MSE) to ensure that any new harvest strategy can perform as required. The application of MSE to an array of data-poor assessments and related harvest control rules (harvest strategies) is the primary aim of this current work.

11 Appendix: Management Strategy Evaluation

11.1 Project Objectives

4. *Establish guidelines, using SESSF case studies, for when the particular Tier harvest strategy for a given stock becomes inappropriate and make explicit recommendations as to what response would then be appropriate.*
5. *Determine options for alternative harvest strategies when none of the present Tiers is appropriate (i.e. potential Tier 5 approaches)*
6. *Produce presentations and explanatory documents for distribution across RAGs and MACs, describing the criteria and new Tier 5 harvest strategies.*

The first two objectives require some active examination of current strategies and processes as well as some exploration of the properties and behaviour of potential new Tier 5 approaches. The third objective is primarily about report production and the presentation of methods and any new approaches to potential users.

11.1.1 Guidelines for Selecting a Tier

When the SESSF harvest strategy framework was introduced in 2005 (two years ahead of the Commonwealth Harvest Strategy Policy, which modified the SESSF harvest strategies) there was a need to allocate the main commercial species to the available Tiers. Originally there was a Tier 2, which was considered to be a less robust dynamic stock assessment model (perhaps similar to a Tier 1 but with fewer year's data available). However, eventually it became clear that only Tiers 1, 3, and 4 were used and the notion of a Tier 2 became neglected, at least in the SESSF. It can be argued that all other commercial Commonwealth fisheries only have a single harvest strategy (i.e. standard data collection, a single form of assessment, and a specified harvest control rule); although the mixed species Northern Prawn Fishery has an array of species and associated assessment methods a hierarchical tier system is not used explicitly. For bycatch species and other species that may be impacted by fishing pressure there is also the Ecological Risk Assessment process (Smith *et al*., 2007; Hobday *et al*., 2011).

The allocation of species to specific Tiers in the SESSF was originally based upon the data available for a species rather than whether the harvest strategy (HS) within a given Tier could deliver the required management advice in a consistent and workable manner. This first objective is about developing an explicit set of guidelines for selecting a Tier for a given species. This will include some testing, using management strategy evaluation, of the effects of bias and imprecision being present in the data available for assessment.

11.1.2 Evaluation of Alternative Tier 5 Approaches

There have always been data-poor fisheries and since the 1990s when the importance of explicit management of commercially exploited stocks became more fully developed (FAO, 1995), some attention has been paid to data-poor stock assessment methods (Kruse, *et al*., 2005). However, work on the assessment and management of data-poor and data-limited fisheries gained new impetus when the Magnusson-Stevens Act in the USA (their Fisheries Act) was amended in 2006 to require the determination of annual catch limits without considering whether there was sufficient data to enable such a determination (MacCall, 2009). Since then a number of alternative methods that can be applied to relatively data-poor fisheries have arisen. Some of these methods are purely empirical and driven only by the catch data and others include the catch data augmented by biological information from the species involved. This remains an active research area with many contributors and novel methods still being published (MacCall, 2009; Dick and MacCall, 2011; Bentley and Langley, 2012; Martell and Froese, 2013; Carruthers *et al.*, 2014; Geromont and Butterworth, 2014).

To achieve the second objective, in the Australian context, an array of potential methods selected from the literature will be tested, using management strategy evaluation to determine whether or not they can meet the underlying requirements of the Commonwealth Harvest Strategy Policy (HSP). These requirements are aimed at ensuring sustainability and maximizing profitability. The policy aims to maintain sustainability by preventing over-fishing and preventing the stock from being overfished, and it does this by managing each stock so it stays above some minimum stock size; with the HSP selecting 20%B_0, or *another accepted proxy* as the limit reference point. The policy also aims to maximize profitability by adopting the maximum economic yield (MEY) or *an accepted proxy* as a target reference point (DAFF, 2007). Within data-poor and data-limited fisheries invariably the best that can be achieved is to meet the intention of maintaining sustainability, whereas ensuring maximum profitability would be extremely difficult for most data-limited situations (Haddon, 2012). Nevertheless, it is sometimes possible to devise proxies for what would constitute what might be termed a 'pretty-good' profitability (Hilborn, 2010; Haddon, 2011b) and these can become the target.

11.2 Tier 5 Assessment Evaluation and Catch Determination

The current assessment of fish stocks in the SESSF is conducted under a tiered approach, whereby stocks with reliable and sufficient data, together with a robust assessment, are assessed under a Tier 1 assessment, and stocks with data of less quality and quantity are assessed under Tier 3 (catch curve based) or Tier 4 (catch rate based) assessments. In some circumstances, the data needed or available for even the lower ranked or higher tier assessments are not appropriate for these tiers. This may occur because of insufficient or unrepresentative sampling, market driven catches, insufficient data on catches or biology or biological/fishery characteristics that undermine the assumptions of the current tier assessments (Carruthers et al., 2014). In such cases alternative methods are needed to assess the stock and set annual catches (recommended biological catches, RBCs).

The harvest strategies adopted by the SESSF are composed of an assessment of the current status of the stock using specified data types, with an associated harvest control rule that compares the estimated current status to a target (and a limit) reference point: being either a target biomass (Tier 1), target fishing mortality rate (Tier 3) or target catch rate (Tier 4). The harvest control rule then translates the relative stock status (or depletion level in the case of a Tier 1 assessment) into an RBC (**Figure 8**, **Figure 9**, and **Figure 10**). The general principal of each HCR is that the lower the perceived stock status relative to its target, the lower the catch (the fishing mortality), with the aim of allowing the stock to build back up to meet the target reference point (Smith et al. 2008).

The need to develop assessment methods to estimate the stock status of data-poor or data-limited fisheries is not restricted to Australia. For example, in 2006 the USA's Magnuson-Stephens Fishery Conservations and Management Act (MSA) was amended to require scientifically derived annual catch limits for all federally-managed stocks in the United States (with some exceptions; Newman *et al*., 2015). As a result there has been a great deal of effort and development occurring to produce methods that can generate catch limits in data-limited situations (Carruthers et al., 2014). Such methods can be categorized into three classes:

1. catch-only methods
2. catch methods supplemented by biological parameters (growth, natural mortality)
3. catch methods supplemented by the inclusion of a simple model of dynamics. The data cannot be fitted to the model but implausible parameter combinations can be removed so that constraints are placed on the viable possible catches.

The third class is relatively new (Martell and Froese, 2013) and, except for the DB-SRA method (Dick and MacCall, 2011), will not be considered further except in the discussion of alternatives.

The methods to assess stock status and set annual catches that are described below are categorized into those that only use a time-series of annual catch (catch-only methods) or those that have additional information, on biological parameters for example (catch-supplemented methods).

11.2.1 Catch-Only Methods

Catch-only methods are utilized to estimate an RBC from data-poor methods where insufficient data is available to reliably determine stock status from the currently available methods. Catch–only methods have been used by the South Atlantic Fishery Management Council (SAFMC) and the Mid-Atlantic Fishery Management Council (MAFMC) to manage a number of their stocks (Carruthers et al., 2014). These methods are purely empirical and have no direct relation to the underlying dynamics. Data poor fisheries in New Zealand are assessed under various methods at least partly depending on data availability (Anon, 2012). A simple catch-only method commonly used estimates the maximum constant yield (MCY) as:

1. the average catch over an time period appropriate to the species multiplied by a constant known as the natural variability factor *c*, $MCY = cY_{av}$

The constant *c* attempts to account for natural variation in each stock's productivity; the greater the expected variability, the lower the value of the constant (**Table 3**). If the period over which the average catch method is calculated occurs when the stock was fully exploited, then the method should give an estimate of MCY. However, if it occurs during development of the fishery or during under-exploitation, then the catch will be a conservative estimate of MCY (Ministry of Primary Industry, 2014).

Table 3. The natural variability factor from New Zealand's method four harvest control rule: $MCY = cY_{AV}$. Ministry of Primary Industries (2014, p 29).

Natural Mortality Rate: M	Natural Variability Factor
< 0.05	1
0.05 - 0.15	0.9
0.16 - 0.25	0.8
0.26 - 0.35	0.7
> 0.35	0.6

Catch-only methods can be summarized to illustrate their slightly different approaches (**Table 4**), which vary from using an average, a median, or a maximum catch from a specified period.

The first four methods described below (**Table 4**) do not have a harvest control rule that adjusts the recommended catch (e.g. through a fishing mortality) in attempts to manage the fishery towards a target (Wayte and Klaer, 2010; Little et al., 2011). In the case of the SESSF Tier 4 catch rate based assessment, if the catch rate is lower than the target catch rate, then the catch is scaled down on the assumption that the available biomass is proportional to the current catch rate (Little et al., 2011, Haddon, 2014). In Methods 1 – 4, no other information is used other than catch and it is not generally considered appropriate for catch to be a proxy for abundance. In Method 4, if it can be assumed that the average yield is taken from a period of relatively stable catch and stock biomass, then essentially a target catch has been identified. However, if catches are below the target catch, then this should not necessarily be an indicator that catches can be increased toward the target, as the catches may be low because of low stock size. In fact, a more precautionary approach may suggest that catches should be further reduced under these circumstances, especially if catches have been low for an extended period of time, and adequate justification for the low catch values cannot be provided. If catches can reliably be said to be below the target catch simply because of market or operational decisions, then it may be reasonable to increase catches towards the target reference catch. In effect, the details of the harvest strategy that uses the $c\mathrm{Y}_{\mathrm{av}}$ as a form of control rule has not been fully articulated to provide guidelines for all possibilities when applying the method. In all the methods 1 – 4, the assumption made is that the catches selected by the method represent a stable and acceptable catch level for the fishery. This cannot be considered even an approximate proxy for a target of MEY, however, the assumption is made that whatever catch is selected it is sustainable, meaning that it enables the fishery to avoid the limit reference point.

11.2.2 Model Assisted Catch-Only Methods

If additional information, either estimated or assumed, can be combined with or can supplement a time-series of catches then further methods to assess and set catches can be suggested.

Table 4. Some alternative catch-only methods for setting an RBC.

T5	Code	Brief Description	RBC
4	C3	Third highest landings over the last 10 years	max($C_{0..-9}$)
3	MC	Median catch from the last 10 years	median($C_{0..-9}$)
3	MC	Median catch from the last 3 years (last *x* years)	median($C_{0..-2}$)
5	CY	Scaled average catch from a reference period - MCY	$c\bar{Y}$
6	DB	DB-SRA – depletion based – stock reduction analysis	median(DB-SRA)
7	DC	DCAC – depletion corrected average catch	Median(DCAC)
8	DA	DACS – depletion adjusted catch scalar	median(DACS)

11.2.3 Depletion Adjusted Catch Scalar (DACS)

Carruthers *et al.* (2014) proposed the Depletion Adjusted Catch Scalar (DACS) method, which is a control rule similar to that proposed by Berkson et al. (2011), whereby previous catch levels are adjusted according to periodic estimates of population depletion. The adjustment acts as a control rule, dynamically adjusting the catch. The catch is defined as the mean inter-quartile catch, i.e. the average of all catches greater than the 25th percentile and less than the 75th percentile). The catch is then adjusted by a factor (*b*) according to: half, equal or twice the inter-quartile mean catch when current biomass is considered to be less than 20% of unfished, greater than 20% and less than 60% of unfished, and greater than 60% of unfished levels, respectively; equations (1) and (2).

$$Catch = sbC_{IQ} \tag{1}$$

$$b = \begin{Bmatrix} 0.5 \\ 1 \\ 2 \end{Bmatrix} \quad \text{if} \quad \begin{Bmatrix} B_c < 0.2B_0 \\ 0.2B_0 < B_c < 0.6B_0 \\ B_c \geq 0.6B_0 \end{Bmatrix} \tag{2}$$

Where C_{IQ} is the average inter-quartile mean catch for a pre-defined period, and *s* is a scalar multiple. Carruthers et al. (2014) used an MSE to test the efficacy of the method for *s* scalars of 0.75 and 1.0; equivalent to the discounts used here of 0.25 and 0.0.

11.2.4 Depletion-Corrected Average Catch (DCAC)

Depletion-corrected average catch (DCAC; MacCall, 2009) uses depletion, F_{MSY}/M, M and B_{MSY}/B0 (the same inputs as DB-SRA) with an estimate of average annual catch to provide an estimate of sustainable catch. It also uses Monte Carlo sampling to generate estimates of uncertainty about the average catch estimates. MacCall (2009) used adjustments to well-known simple representation of stock dynamics (e.g. MSY = $0.5MB_0$) to take into account the fact that initial stock depletion includes the windfall catches obtained by removing the biomass during the depletion. While this method does not provide an estimate of MSY it does provide a proxy suitable for a sustainable catch and MacCall recommends this method for the estimations of "… a practical level of yield that is likely to be sustainable."

11.2.5 Depletion-Based Stock Reduction Analysis (DB-SRA)

The depletion based stock reduction analysis (DB-SRA) method can be used if information in addition to a catch time series and estimate of current depletion is available (Dick and MacCall, 2011; Carruthers et al., 2014). The additional information includes: the ratio of F_{MSY} to the natural mortality rate (F_{MSY}/M), the natural mortality rate (M), the most productive stock size depletion relative to unfished (B_{MSY}/B_0) and the age at maturity (A_{mat}). Because it shares so many of the requirements it can thus be considered to be an extension to the DCAC method (MacCall, 2009). Plausible values for the required parameters are drawn from assumed distributions and iterated (through Monte Carlo simulation) using some form of production model to define recruitment (Dick and MacCall, 2011, used a new delay-difference population model but any production model could be used). For each sampled parameter set, the value of B_0 (initial unfished biomass) is found that produces the depletion, given the time-series of catches. Not all combinations of parameters will result in the given level of stock depletion. In some cases, stock biomass will become negative. These implausible combinations of parameter sets are discarded. For a particular set of plausible parameters, the value of F_{msy} and the catch at F_{msy} (*MSY*) can be calculated, and used in future projections. Distributions of these parameters can also be considered.

Feasible Stock Trajectories (FST)

The feasible stock trajectories (FST) method of Bentley and Langley (2012) falls into the same class of models as DB-SRA. Namely, a pool of feasible trajectories (of say 1000) is maintained each year, with those that are deemed infeasible removed from the pool. Similar to DB-SRA each trajectory is defined by a combination of parameters selected from prior probability distributions (e.g. steepness, natural mortality) and variables (e.g. current biomass) which are updated on a yearly basis. Each year the variables are compared against likelihood functions that reflect the range of potential feasible values for the variables. Those trajectories that are infeasible, because one or more of the variables is beyond a feasible range, are removed from the pool and other trajectories are then tested for their feasibility. The suggested control rule for setting catches is the constant catch that achieves a target biomass in a pre-specified number of years. The control rule uses the full pool of potential trajectories to determine a distribution of potential catches that achieve the target biomass. A percentile of this catch distribution is then used as the recommended catch quota, thus integrating across the uncertainty within the pool of feasible trajectories. Here the FST approach was not investigated separately to the DB-SRA approach.

11.2.6 Further Alternatives

The idea of using the catch history as a means of assessing the stock has led to some intense debate in the literature. Pauly (2013) argues that if all that is available is catch-data then efforts must be made to use that and more fisheries should at least have catch data collected. Hilborn and Branch (2013), however, argue that catch-data alone will be misleading so often that it is dangerous to use such methods to provide management advice. The strategy of using MSE testing of such harvest strategies is more effective than merely arguing about their potential value and potential biases; the need for such formal testing applies to all catch-only data methods.

The above is not an exhaustive list of possible methods. For example, the fishery status classification of Anderson et al. (2012) extends the methods originally proposed by Froese and Kesner-Reyes (2002) to assess stock status using only the time series of catch data (as a proportion of the peak maximum catch). By examining stocks with full quantitative stock assessments, they were able to categorize the sequence of development of a fishery; from its early developmental stage, to full exploitation, over-exploitation and potential collapse. This method would be difficult to implement without a suitable harvest control rule, but if one could be developed it would become worthy of testing.

Those methods that use a form of stock reduction analysis are essentially attempting to identify plausible combinations of population dynamics that would at least be consistent with the observed catches. Martell and Froese (2103) have taken that idea and produced a method that explores the region of plausible dynamics by including a simple model of those dynamics., By using a simple Schaefer surplus production model and setting bounds on the parameters of that model, biomass trajectories that are inconsistent with the observed fishery can be eliminated (i.e. the observed catches might lead some combinations to go extinct or to expand well above the hypothetical carrying capacity). By conducting a Monte Carlo analysis of the possibilities and using the outcomes of the plausible model parameter set, Martell and Froese (2014) are able to generate estimates of MSY along with uncertainty estimates about the management statistics.

Martell and Froese's (2014) harvest strategy has similarities with the approach described by Bentley and Langley (2012) who describe a method that employs "Feasible stock trajectories", although their underlying model is more complex than that used by Martell and Froese (2014).

None of the three methods listed here will be included in the MSE testing of HS as they remain very new and as yet unused (although this will change for the Martell and Froese approach, Sabater and Kleiber, 2014). Nevertheless, these methods may become of interest in the near future as more is learnt of how best to implement them in real world situations.

12 Appendix: MSE Methods

12.1 Introduction

A SESSF management strategy evaluation (MSE) framework has been developed over a number of years (Wayte 2009, Fay *et al*., 2009; Klaer and Wayte 2011), and provides a flexible platform for testing data requirements of harvest strategies as they apply to SESSF species in particular, but also more generally. New projects that require MSE testing therefore no longer require the development time for the detailed operating model that incorporates uncertainty with the dynamics of a fish stock, sampling of data required for stock assessment, and the implementation of certain harvest strategies.

The work proposed here is new, however, and hence requires adjustment of the simulation of fisheries sampling to examine effects of different levels of precision and accuracy of data collected from the fishery, and also requires the implementation of new data-poor assessment methods and harvest control rules. However, these modifications to the existing system are relatively minor, thus allowing this proposal to be built as a one-year project, with greater focus on planning and running of appropriate simulations and interpretation of the results.

It is standard practice to base MSE testing on species for which good information is available, so that the results across a full range of harvest strategies can be compared (in this case Tiers 1 to 5). The SESSF data rich species used in the following analyses were Tiger Flathead (*Neoplatycephalus richardsoni*) and School Whiting (*Sillago flindersi*), and testing is carried out under a range of stock depletion levels for each species including being above, at, and below the target depletion level (initial depletion levels used were $0.18B_0$, $0.48B_0$, and $0.78B_0$). The intention of this was to determine whether each assessment method tested was capable of recovering a depleted stock, maintaining it close to the target, and of fishing a stock down in a controlled fashion.

Two other species were considered for inclusion in the testing: Jackass Morwong (*Nemadactylus macropterus*) and Blue Grenadier (*Macroronus novaezelandiae*). However, preliminary testing showed that the characteristic episodic recruitment and delays in stock assessment/TAC application for Blue Grenadier create problems across harvest strategies that will require further investigation. Blue Grenadier was therefore removed from our list of species in favour of stocks with behaviour that is better known and more predictable. The biological characteristics of Jackass Morwong are not greatly different to Tiger Flathead, and the major difference in the history for that species is an apparent environmentally driven regime shift that has affected average recruitment levels (Wayte, 2013). How that regime shift is dealt with in projections creates various future scenarios that are not relevant for the current project. Therefore, Jackass Morwong has also been removed from the list. The remaining species, Tiger Flathead and School Whiting, encompass a relatively wide range of life history characteristics that determine productivity and stock variability, so results using them can be extended to a much larger list of other truly data-poor species.

The Tier 1 harvest strategy in the SESSF, involves a fully quantitative stock assessment, has a variety of standard data sources including length and age composition and also an abundance index information. In the case where one of those sources contains an unknown bias, if the bias is sufficiently or consistently large the assessment will show a conflict among data sources, thereby allowing for recognition and investigation of the source of that bias, and dealing with it in some way, at least in alternative model structures via sensitivity analysis. However, Tiers 3 and 4 both rely on a single source of input data (age composition and CPUE trends, respectively), which implies that such biases would not be detected.

The Tier 5 procedures considered can be either fixed, where a single catch level is set at the start of the projections, or dynamic, where there is feedback from any response of the stock and the analyses are updated regularly using new data from the fishery. As with the Tier 3 and 4, bias or imprecision could not be detected in either of these Tier 5 approaches.

12.2 Initial Conditions

Before each MSE run the operating model needs to be initialized to some predefined state; it needs to be conditioned on a given species and fishery. The conditioning consists of defining the parameters of the population dynamics and of the fishery to relate to a particular species/fishery combination. Thus, values for an array of biological characteristics (growth, recruitment, maturity, and natural mortality) are parameterized within the operating model and these need to be defined before the MSE can be run. In addition, details of the fishery, including, for example, the number of fleets, areas, fishing methods and their respective selectivity patterns, are also parameterized. This conditioning is based on the outcomes of the fully articulated age-structured integrated assessment models that have been developed for both Flathead (Klaer, 2011; Day and Klaer, 2014) and School Whiting (Day, 2010, 2012).

Once the model is conditioned on the selected species biology and the fishery then the dynamics need to be initiated. The model begins in an equilibrium unfished state and the historical catches in each case are used to fish down each simulated fishery. However, instead of only starting the simulations from whatever state the actual fishery is at after its historical catches, the aim in the MSE testing is to determine two things about each harvest strategy, these are whether: 1) it can recover a fishery if it starts in a depleted state, moving it away from any limit reference point and towards the selected target and 2) it can control a fish down from a lightly depleted state down towards the selected target. To set up these conditions of being below the target and being above the target, the historical catches remained the same but the average recruitment levels required to achieve these degrees of depletion is altered. The historical catches, with the altered average recruitment levels, are applied to the unfished equilibrium state until the desired depletion level is achieved in the simulation at the start of year 1, equivalent to 2009; this was the year in which the projections begin using the harvest strategy under test to set the TAC each year. Recruitment variability was introduced for the last 20 years of the historical catches so as to have the projections begin with more realistic levels of variability.

The harvest strategy would include the sampling containing the assumed levels of bias and precision, and then run for a further 30 years (2009 – 2038) with 100 replicate runs made for each scenario considered. In practice, and not surprisingly, the median depletion levels generated by the operating model in 2009, at the start of the projections, varied between Flathead and School Whiting (**Table 5**).

Table 5. Actual depletion levels at the start of 2009, the start of the projection period, as a % of B_0, at the end of the conditioning phase. Individual runs would exhibit differences in their depletion levels in 2009 due to variation in the recruitment time series. All the distributions were skewed to the right (a greater spread above the median than below). 'Highly' means highly depleted and lightly means lightly depleted

Species	Status	Lower 5%	Median	Upper 95%
Flathead	Highly	30.7	32.8	37.3
Flathead	Lightly	56.5	58.8	64.1
School Whiting	Highly	30.9	40.2	55.8
School Whiting	Lightly	55.0	64.8	81.3

12.3 The Appropriateness of Tier harvest strategies

The first objective here is to "*Establish guidelines, using SESSF case studies, for when the particular Tier harvest strategy for a given stock becomes inappropriate and make explicit recommendations as to what response would then be appropriate.*"

The selection of which SESSF harvest strategy Tier to use is currently based primarily on the available information that can be used to make an assessment of stock status. To be explicit about why this can be a problem, the selection does not currently take into account the capacity of any selected Tier at estimating stock status from that available data. The first objective therefore requires an examination of the consequences of the application of the current Tier 3 or Tier 4 when the data may not be of sufficient quality to support that application. Errors in observations can be categorised as related to precision and bias (or accuracy; **Figure 11**), which can be examined explicitly and separately in the context of SESSF Tiers and species.

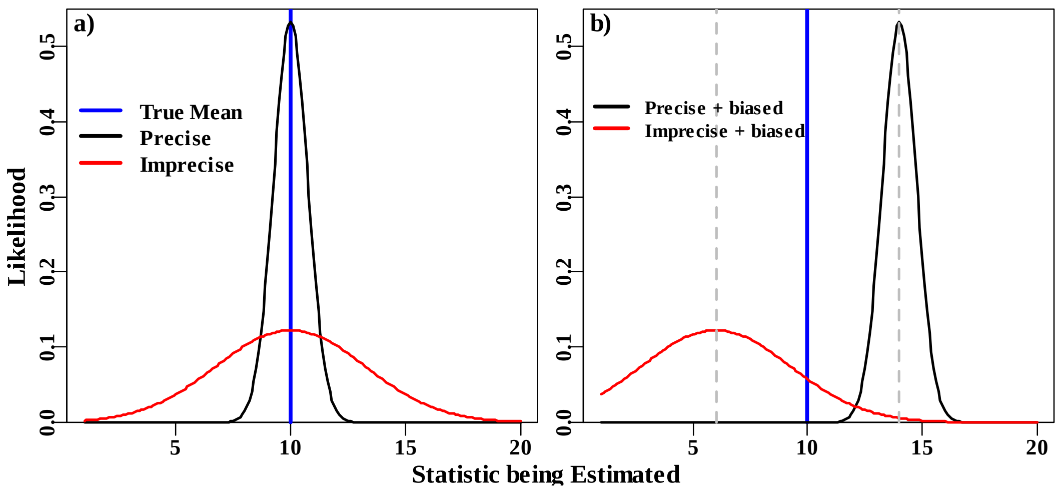

Figure 11. Illustrations of the notions of precision and bias. It is possible to be relatively imprecise but unbiased (a), analogously it is possible to have a precise estimate which is biased (b) such that the estimate completely misses the true mean.

SESSF MSE work to date (Wayte 2009, Klaer and Wayte 2011) has assumed that sampling from the simulated fish population is at levels of precision that reflect average apparent observed levels across many species. It also assumes that sampling is random and unbiased and therefore accurately represents the stock. Where stocks are spatially heterogeneous or have an extensive geographical distribution and sampling occurs unevenly across different areas then biases may enter the data simply through uneven sampling of natural variation. In the simulated sampling within the MSE the precision and bias within sample collections will be varied across all major sampled data sources (CPUE, length, age) to determine how, for each harvest strategy or Tier, this modifies the risk to the stock, as defined by the harvest strategy policy, and the ability of the HS to achieve and maintain the target depletion level.

The effect of data precision is simplest to implement, and requires testing of a range of assumed variance values for sample collection from the simulated population. Data accuracy (bias) is more difficult to implement but can be addressed using plausible scenarios. Examples are a bias towards sampling of more longer/older fish from the population, or a linear trend in catchability in a CPUE index. The latter specifically allows testing of the effect of gear/vessel improvements over time (often termed ‘effort creep’) that may not have been accounted for in CPUE standardisations, and how that may affect the outcome of the application of Tier 4. A total of 36 different combinations of bias and precision were tested with 18 on Tier 3 and 18 on Tier 4 (**Table 6**).

12.3.1 Changed Sampling Precision for Composition Data

The level of precision for composition data is determined by the simulated number of annual samples collected. For length composition the standard value in MSE simulations is 1000, and for age samples is 500. The number of samples collected annually throughout the projected time period will be varied from these defaults to test the effect of composition data precision on stock status outcomes (**Table 6**).

The stated age composition sample size is the total number of samples spread across all fleets in a given year. The sample size taken from each fleet is allocated according to the proportion of catch taken from that fleet in the year of sampling. In the implementation of Tier 3 in the MSE, current *F* is estimated for all fleets that have taken more than 30% of the catch in the last 5 years, and then averaged over fleets using catch weighting. The Tier 3 *F* estimation method is known to be unstable for age samples with a very small sample size (e.g. <= 50). For a fleet that has taken 30% of the catch, the sample size will be 0.3 x total number of samples. To ensure a minimum sample size of 50 requires a total sample size of 166. We have rounded this down to 150, so that the minimum sample size for any fleet used in Tier 3 calculations will be 45.

12.3.2 Changed Sampling Precision for CPUE Data

In the implementation of the Tier 4 harvest strategy in the MSE, the cpue of the fleet with the highest proportion of catch over the last 5 historic years is used in the HCR. For both Flathead the CPUE from the diesel trawl fleet was used, while for School Whiting the CPUE from the Danish seine fleet was used.

When CPUE data are simulated from available biomass in the MSE operating model, a default level of imprecision is assumed for annual CPUE points using assumed CVs from the Tier 1 assessment on which the operating model is based. For Flathead the base cpue fleet CV is 0.05 and for Whiting it is 0.3. The alternative values for CPUE imprecision that are tested are 0.15, 0.45, 0.6 and 0.9; the assumption of linearity between relative abundance and CPUE was not tested at this time.

12.3.3 Applying Bias to Composition Data

In the scenario used, the true (length or age) composition distribution is multiplied by a logistic bias factor that increases each year of the projection for 20 years, and then stays the same for the final 10 years (**Figure 12** and **Figure 13**). Each composition is then rescaled so it sums to 1. The bias factor for each length class in each year is:

$$p_y^l = \left[1 - \frac{1}{1+e^{b(l-l_m)}}\right]\frac{y}{ny}A \qquad (3)$$

where

- b is the logistic slope parameter; a b = -0.6 or 0.6 leads either to a bias towards smaller or to younger fish (**Figure 12**) and b = -3.0 or 3.0 leads either to a bias towards younger or older fish (**Figure 13**);
- l is the mid-point of a given length or age class;
- l_m is the mode of the true length distribution in the first year that the bias is applied, or is the true mean age in the most recent ten historic years;
- y is 1 in the first year, 2 in the second year, .. ny in the ny^{th} to 30^{th} years;
- ny is the number of years for which the incremental bias is applied; and
- A is either 1 or 0.5, to apply either full or half the amount of bias.

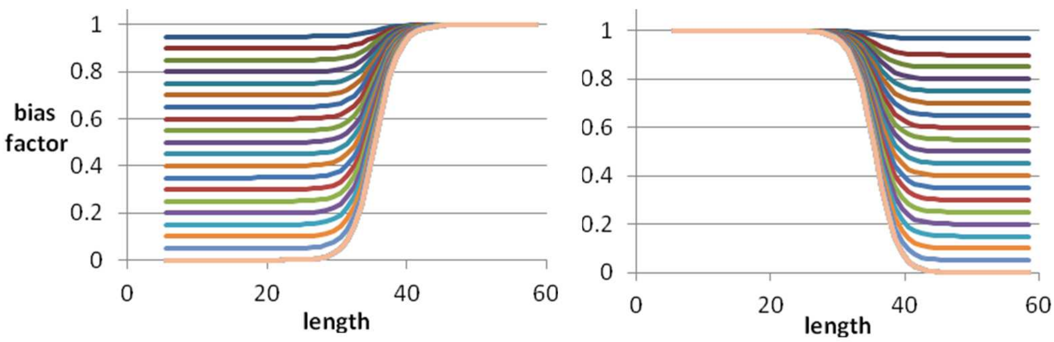

Figure 12. Bias factors applied to length compositions. The top line in each plot is the bias function applied to lengths in the first year of bias application, and the bottommost line is the bias function applied to lengths in the 20th and subsequent years. The plot on the left uses b=0.6, and the plot on the right uses b=-0.6.

The simulations are arranged such that if the compositional bias is applied to the same unbiased distribution in each year in a cumulatively increasing fashion, equ (3), the effect of the bias can be seen, when applied to a stable population structure without variations in the population dynamics (**Figure 14** and **Figure 15** illustrate bias applied to a constant length distribution, while **Figure 16** and **Figure 18** illustrate bias applied to a constant age frequency composition). In the MSE, variations within the population dynamics each year (recruitment catch, etc) imply that the unbiased distribution in each year would invariably be different.

In the MSE, the bias is applied to the true frequency compositions by sex (for each fleet, year, and type [retained/discarded]), error is applied to each of those compositions, then the composition is combined over sexes (if using Tier 3).

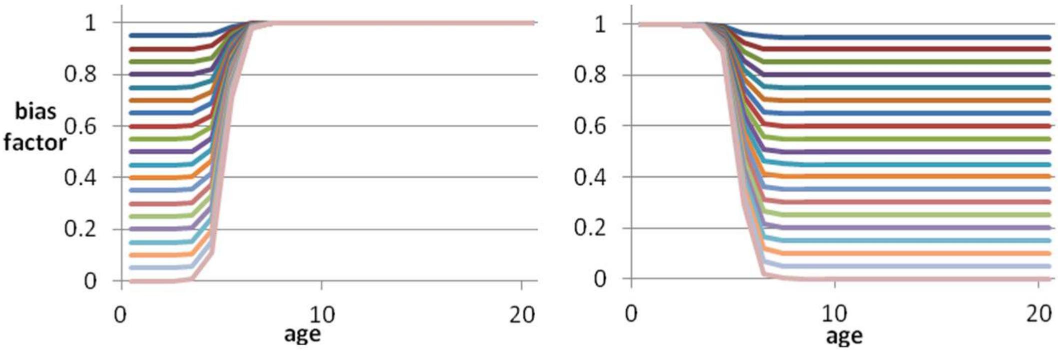

Figure 13. Bias factors applied to age compositions. The top line in each plot is the bias function applied to ages in the first year of bias application, and the bottommost line is the bias function applied to ages in the 20th and subsequent years. The plot on the left uses *b*=3.0, and the plot on the right uses *b*=-3.0.

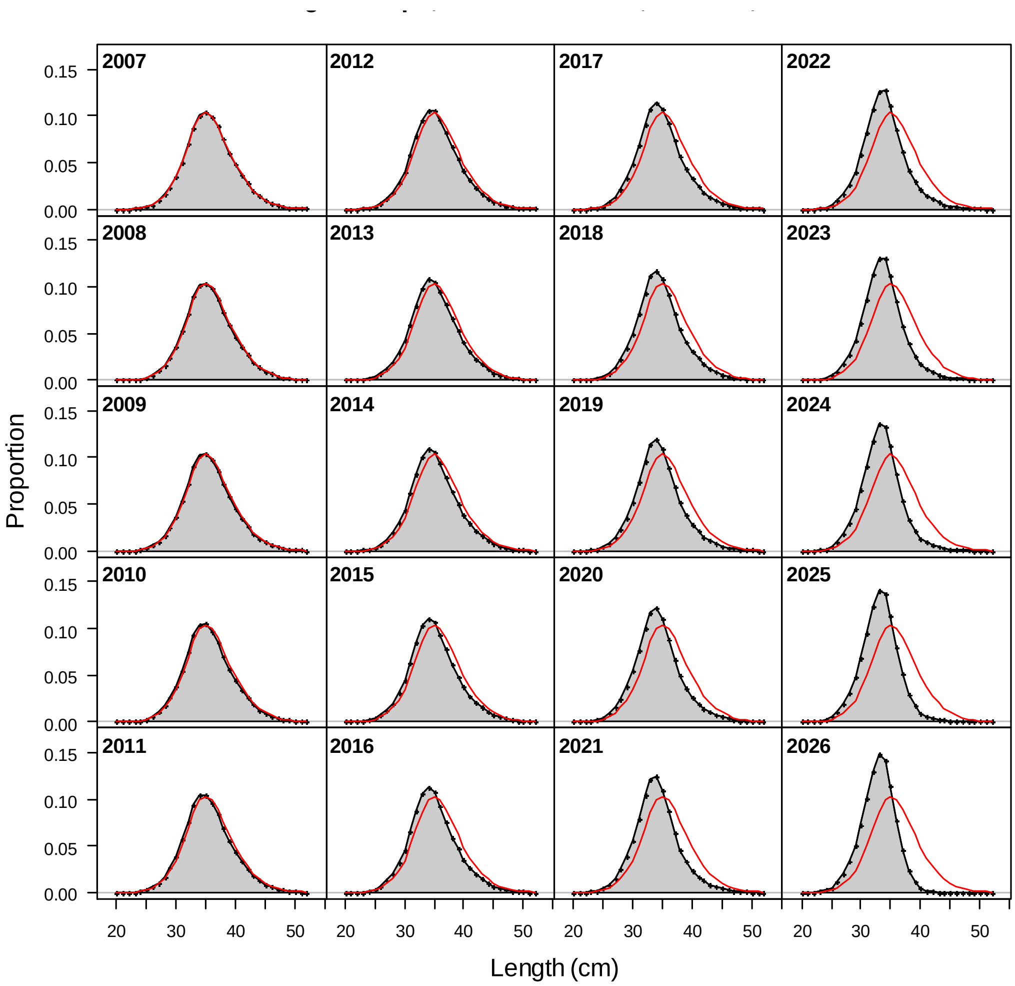

Figure 14. An example of incremental right-skewed bias (biased towards smaller fish) applied to a constant length composition. Bias is applied gradually from 2007. Full logistic bias is applied in 2026, and for the subsequent 10 years (not shown). The red line is the unbiased distribution. The black line with grey fill is the biased distribution that represents the sampling. In reality, and the MSE, such a constant length composition would not occur.

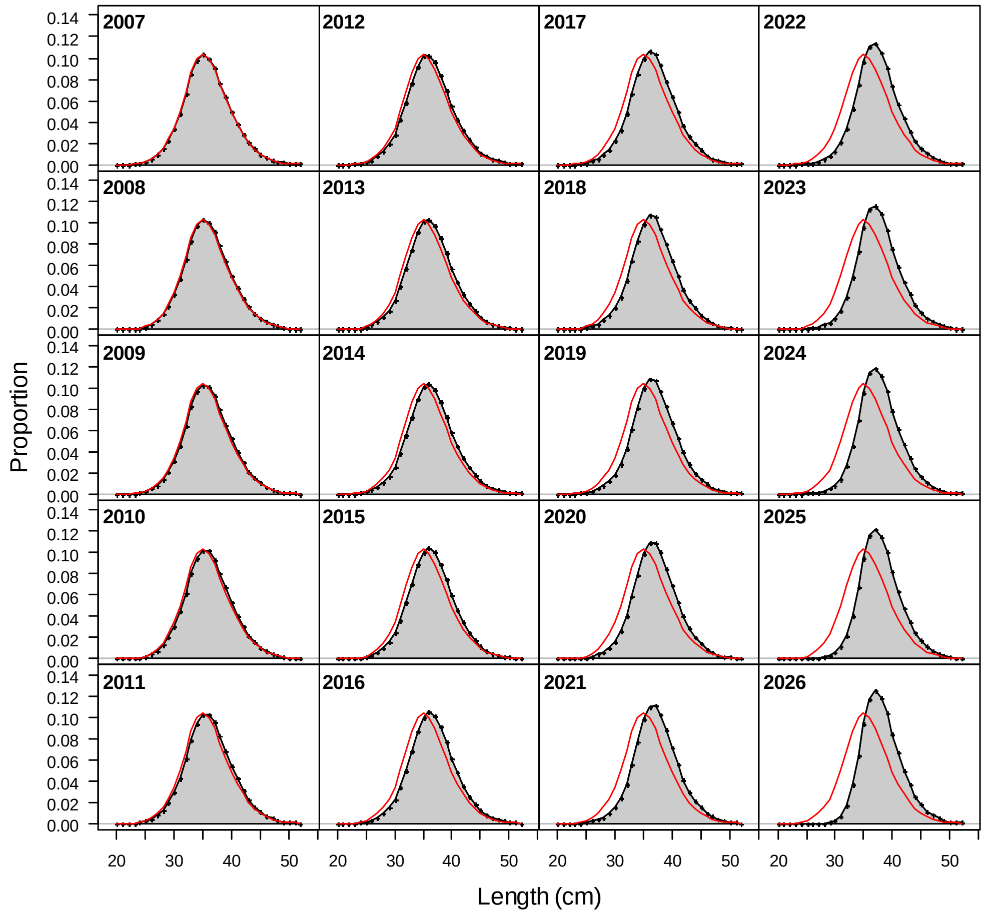

Figure 15. An example of incremental left-skewed bias applied to a length composition (a bias towards larger animals). Bias is applied gradually from 2007. Full logistic bias is applied in 2026, and for the subsequent 10 years (not shown). The red line is the unbiased distribution. The black line with grey fill is the biased distribution that represents the sampling. In reality, and the MSE, such a constant length composition would not occur.

12.3.4 Applying Bias to CPUE data

A linear change will be applied over a range of years starting from the present; this mimics the form of a constant effort creep. Bias multipliers that both increase and decrease the CPUE will be tested. The CPUE generated from the operating model is multiplied by a bias factor which gradually increases for 20 years, and stays at the same level for the next 10 years. The bias factor is:

$$p_y = 1 + \left(CPUEbias - 1\right)\frac{y}{ny} \quad (4)$$

where

y	is 1 in the first year, 2 in the second year, .. *ny* in the *ny*th to 30th years;
ny	is the number of years for which the incremental bias is applied, which was 20 in all runs; and
CPUEbias	takes the values 0.25, 0.5, 1.5 or 2.0.

The bias factor, p_y, by year is shown in **Figure 17**, and an example of the application of a bias multiplier of 2.0 is given in **Figure 19**.

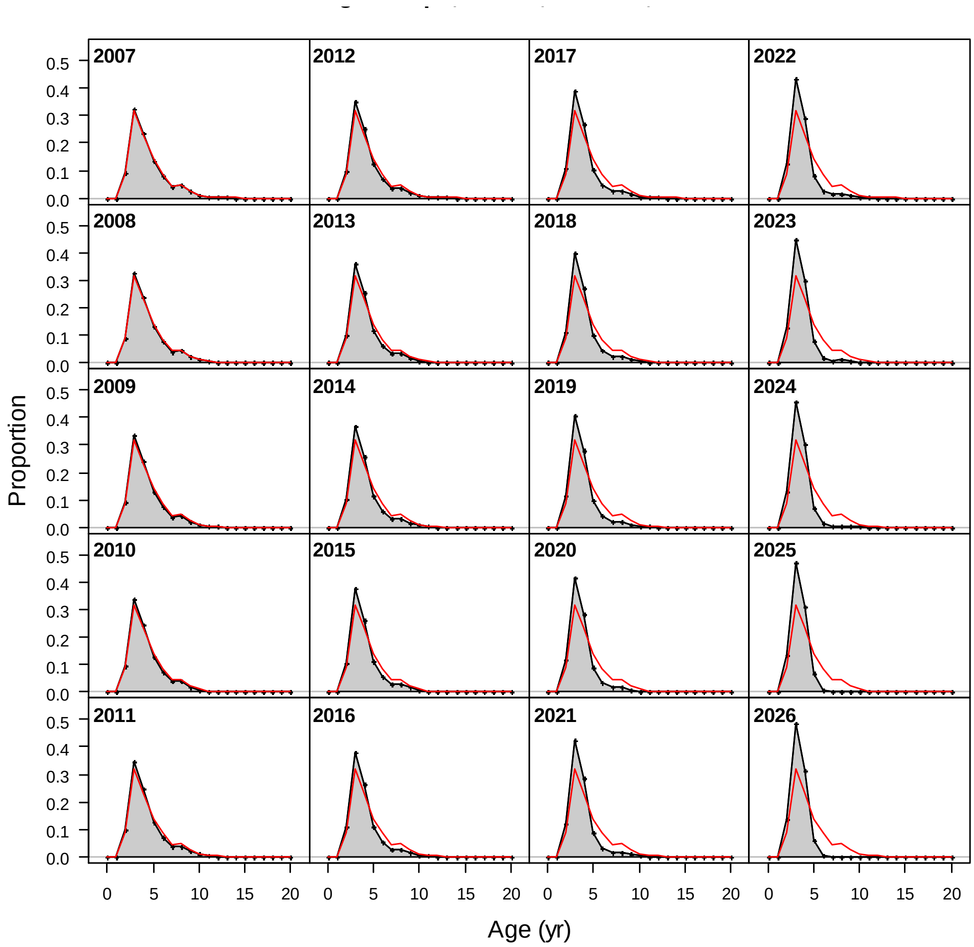

Figure 16. An example of incremental right-skewed bias applied to an age composition (a bias towards younger animals). Bias is applied gradually from 2007. Full logistic bias is applied in 2026, and for the subsequent 10 years (not shown). The red line is the unbiased distribution. The black line with grey fill is the biased distribution that represents the sampling. In the MSE, such a constant age composition would not occur.

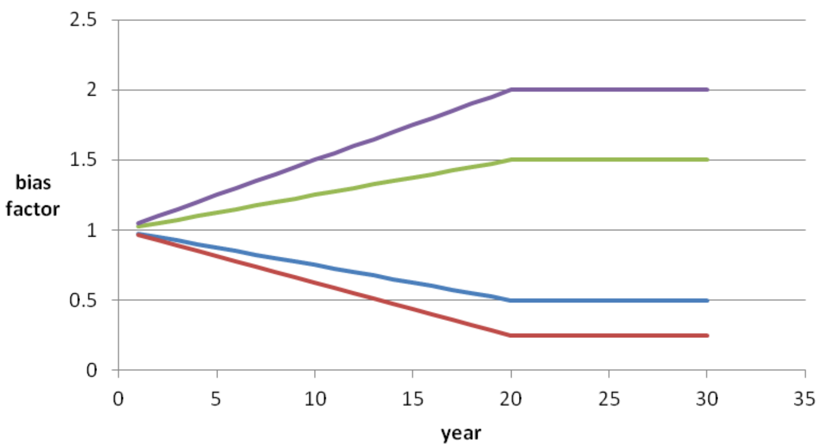

Figure 17. Bias multiplier for CPUE, for the four values of *CPUEbias* (0.25 (red), 0.5 (blue), 1.5 (green) and 2.0 (purple)).

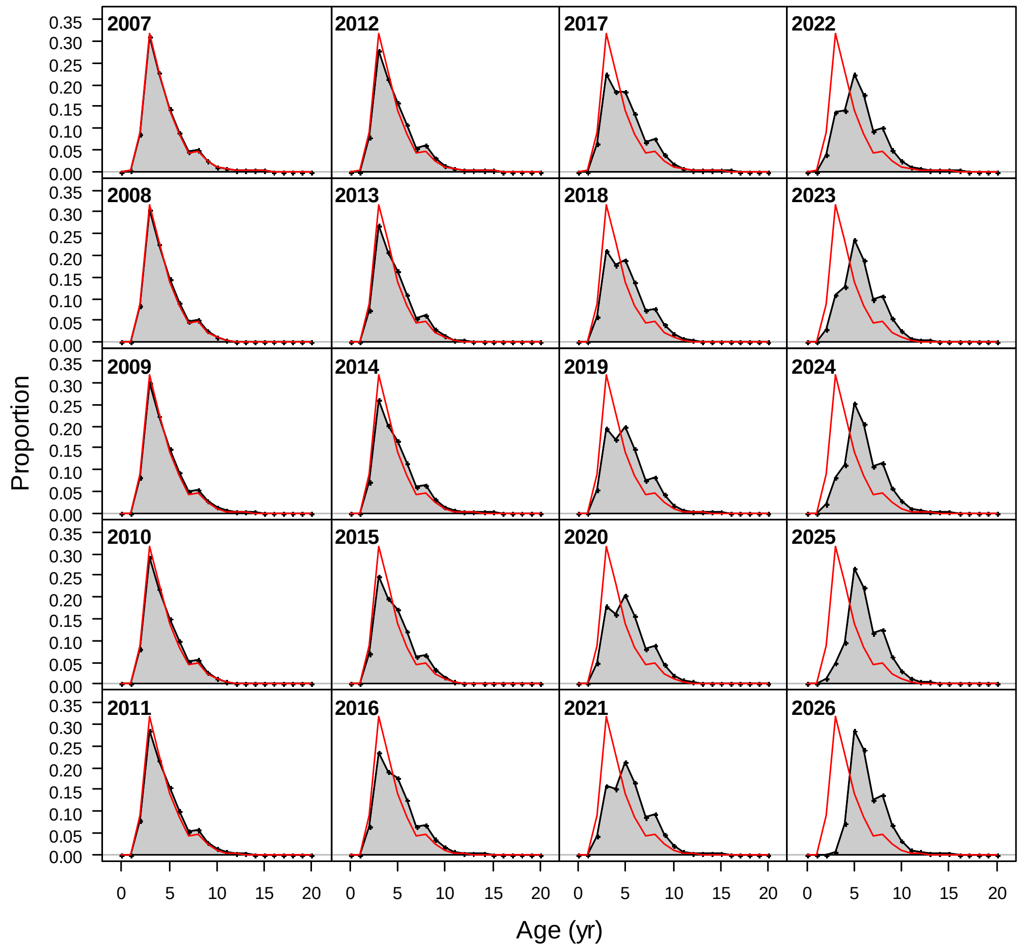

Figure 18. An example of incremental left-skewed bias applied to an age composition (a bias towards older animals). Bias is applied gradually from 2007. Full logistic bias is applied in 2026, and for the subsequent 10 years (not shown). The red line is the unbiased distribution. The black line with grey fill is the biased distribution that represents the sampling. In the MSE, such a constant age composition would not occur.

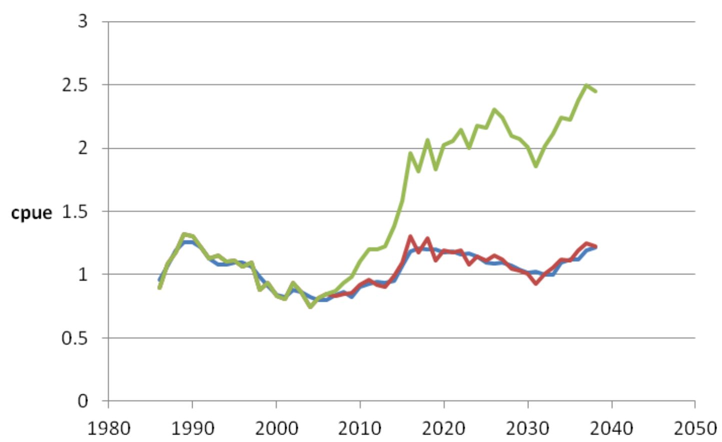

Figure 19. An example of the maximum bias increase applied to a cpue series with random error added. Bias is applied gradually from 2007. The full bias (CPUEbias = 2.0) is applied in 2026, and for the subsequent years. The blue line is true cpue from the operating model, the red has random error added, and the green has bias in addition to the random error.

Table 6. MSE run specifications for testing the effects of sample precision and bias on stock status outcomes for Tiers 3 and 4 per species (major differences among runs highlighted in rose). The base levels of precision (CV) for the CPUE was 0.05 for Flathead and 0.3 for School Whiting. The sample size of 500 was the base line levels of precision for the age structure. The CPUE CV is irrelevant to the Tier 3 and the age sample size is irrelevant to the Tier 4. There are 18 scenarios for Tier 3 and 20 for Tier 4. Heavily depleted = below target, Lightly depleted = above target.

Starting Depletion	Tier	Precision CPUE CV	Age N samples	Bias CPUE	Age
Lightly	3	Base level	500	1	0
Lightly	3	Base level	150	1	0
Lightly	3	Base level	250	1	0
Lightly	3	Base level	1000	1	0
Lightly	3	Base level	500	1	0
Lightly	3	Base level	500	1	-1
Lightly	3	Base level	500	1	-0.5
Lightly	3	Base level	500	1	0.5
Lightly	3	Base level	500	1	1
Heavily	3	Base level	500	1	0
Heavily	3	Base level	150	1	0
Heavily	3	Base level	250	1	0
Heavily	3	Base level	1000	1	0
Heavily	3	Base level	500	1	0
Heavily	3	Base level	500	1	-1
Heavily	3	Base level	500	1	-0.5
Heavily	3	Base level	500	1	0.5
Heavily	3	Base level	500	1	1
Lightly	4	0.05	500	1	0
Lightly	4	0.15	500	1	0
Lightly	4	0.45	500	1	0
Lightly	4	0.8	500	1	0
Lightly	4	0.9	500	1	0
Lightly	4	Base level	500	1	0
Lightly	4	Base level	500	0.25	0
Lightly	4	Base level	500	0.5	0
Lightly	4	Base level	500	1.5	0
Lightly	4	Base level	500	2	0
Heavily	4	0.05	500	1	0
Heavily	4	0.15	500	1	0
Heavily	4	0.45	500	1	0
Heavily	4	0.8	500	1	0
Heavily	4	0.9	500	1	0
Heavily	4	Base level	500	1	0
Heavily	4	Base level	500	0.25	0
Heavily	4	Base level	500	0.5	0
Heavily	4	Base level	500	1.5	0
Heavily	4	Base level	500	2	0

12.3.5 Testing effects of data precision and bias on Tier 3 and Tier 4

For each species (Flathead and School Whiting), changes in precision and bias are introduced from the current year to year 20 in a projection, and then remain at the changed level to the end year of the projection at year 30. This will allow 10 years for the stock to tend towards equilibrium by the end of the projection under the applied change. An important component of the Tier 3 harvest control rule is a maximum RBC change of 50% (Fay *et al*., 2013). This meta-rule was only applied after the first projected year because past catches were not necessarily consistent with the altered R_0 used to obtain the required starting stock status.

Each scenario is defined by its starting depletion (Heavily depleted – below target, or lightly depleted – above target; **Table 5**), the Tier being applied (3 or 4), the precision of the CPUE sampling, and the sample size for ages, and finally the bias applied to the CPUE and the ages (**Table 6**).

12.4 Potential Tier 5 Approaches

When none of the present Tier harvest strategies are appropriate and yet explicit management advice as to sustainable catches is required (when an ERA would be insufficient), then options for alternative harvest strategies and Tiers are needed. These are referred to generically as Tier 5 methods; multiple possible Tier 5 assessment methods might be recommended to allow for the selection of a method most appropriate to particular circumstances but all could be referred to as Tier 5 methods.

The six potential Tier 5 candidate procedures, identified in the introduction, were implemented within the SESSF MSE in preparation for testing against each other (**Table 7**), using four different HS performance measures relating to short-term and long-term stock risk, the total yield, and RBC variability (see section 12.4.1; e.g. Wayte, 2009; Klaer and Wayte 2011). Harvest control rules associated with each procedure were also decided or developed (**Table 4**, **Table 7**).

Runs to be completed for performance testing of the procedures are summarised in **Table 9**. In recognition of current discount factors applying to the RBCs produced by data poor methods in the SESSF, resulting RBC values from the tested harvest strategies were also subjected to discount factors of zero (no discount) and 25%. In addition, the initial stock status in 2009 was assumed to be either 18%B_0 = Low or 78%B_0 = High (see **Table 5** for the actual depletion levels). With the two species, the relatively high or low stock status, and the two different discount levels this equates to 8 times the 25 different run specifications, meaning 200 runs in total (**Table 8**).

For the fixed methods, in which the RBCs are set and held constant throughout the projection, we use the last 10 (or 3 for method 3) years of historic catches to set the RBC.

For the three methods that require a stock status estimate (DB-SRA, DCAC, and DACS) the assumption is made that stock status at the end of the historical catches (the start of the projections; 2009) was 48%B_0, and three levels of assumed change in stock status over the final 10 year period of historical catches were tested: 0, +30% and -30%. This is the same as assuming that the stock status in 2009 was either 48%B_0, 78%B_0, or 18%B_0.

For the feedback or dynamic scenarios of the catch-based methods, the most recent 10 (or 3) years of catches are used to set the RBCs. For the methods that require a stock status estimate we use the last 10 historic years of catches to the current year; thus the number of years can extent from 10 out to 39 in the final year of projection. As with the fixed catch based methods that required a stock status estimate, we assume implicitly that stock status in 1989 was 48%B_0, and test three levels of assumed change in stock status over the period from 10 years prior to start of projections to current year: 0, +30% and -30%.

Table 7. Alternative catch-based Tier 5 approaches to be tested. The code can be either the number or the text label. The 'Δ stock status' relates to the assumed change in stock status over the period prior to the projections (a required input for the dynamic Tier 5 assessments involved. Modified DACS procedure described by Carruthers *et al*. (2014).

Type	Code	Procedure	Catch period	Δ stock status	Source
Fixed	3 MC	Median catch	Last 3 historic		MAFMC
	3 MC	Median catch	Last 10 historic		SAFMC
	4 C3	3rd highest catch	Last 10 historic		SAFMC
	6 DB	DB-SRA	Last 10 historic	0, +30%, -30%	Dick and MacCall, 2011
	7 DC	DCAC	Last 10 historic	0, +30%, -30%	MacCall, 2009
	8 DA	DACS	Last 10 historic	0, +30%, -30%	Berkson et al., 2011, modified
	5 CY	MCY	Tier 4 reference period		NZ
Dynamic	3 MC	Median catch	3 most recent		MAFMC
	3 MC	Median catch	10 most recent		SAFMC
	4 C3	3rd highest catch	10 most recent		SAFMC
	6 DB	DB-SRA	Last 10 historic to most recent	0, +30%, -30%	Dick and MacCall, 2011
	7 DC	DCAC	Last 10 historic to most recent	0, +30%, -30%	MacCall, 2009
	8 DA	DACS	Last 10 historic to most recent		Berkson et al., 2011, modified

Table 8. Combinations of species, initial depletion levels, and potential discount applied to each possible Tier 5 Harvest Strategy. With each of these considered for each of the 25 scenarios listed in **Table 9**, there were a total of 200 separate sets of replicate simulations.

Species	Initial Depletion	Discount
Flathead	H	0
Flathead	H	25
Flathead	L	0
Flathead	L	25
School Whiting	H	0
School Whiting	H	25
School Whiting	L	0
School Whiting	L	25

12.4.1 Harvest Strategy Performance Measures

The performance of each catch-based HS was evaluated by summary plots of the following six performance measures relating to stock level, catch, and variability in catch:

1. average annual catch over the projection period of *f* years:

$$\bar{C} = \left[\sum_{t=y_1}^{y_f} C_t \right] / f \tag{5}$$

where y_1 and y_f are the first and final years of the projection period, respectively, and C_t is the catch in year t, in all cases there were 1 – 30 years of projection;

2. spawning stock biomass (SSB) in the final year relative to unfished SSB (depletion level):

$$D_f = \frac{SSB_f}{B_0} \tag{6}$$

Where D_f is the depletion level in the final year f, SSB_f is the spawning stock biomass in year f, and B_0 is the unfished SSB;

3. catch variability: average absolute percentage inter-annual change in catch (%AAV) over the projection period:

$$\%AAV = 100 \sum_{t=y_2}^{y_f} \left| C_t - C_{t-1} \right| \Big/ \sum_{t=y_1}^{y_f} C_t \tag{7}$$

where y_2 is the second year of the projection period; and

4. probability of the spawning biomass going below the limit reference point (B_{20}) at any time during the projection period.

$$P\left(SSB < LRP\right) = \text{count}_{t=y_1}^{y_f} \left[\left(SSB_t / B_0 \right) <= 0.2 \right] / f \tag{8}$$

where LRP is the limit reference point (= 20%B_0 or its proxy). The Commonwealth Harvest Strategy Policy sets a limit of staying above the LRP at least 90% of the time so if the probability of falling below is greater than 10% this would constitute a failure.

Table 9. MSE run specifications for alternative catch-based Tier 5 Harvest Strategies per species, starting stock status and discount level, each separated by a dashed line. MC – median catch; C3 – 3rd highest catch; CY – MCY; DB – DB-SRA; DC – DCAC; and DA – DACS. Note the scenarios are not in numerical order, but are in the same order as plotted on the diagrams in the data-poor section.

Scenario	Tier5 type	Fixed (x) Dynamic (B)	Catch period	Depletion change 1999 - 2009	Assumed depletion in 2009
1,26,51,76,101,126,151,176	3 MC	x	3		0.48
3,28,53,78,103,128,153,178	3 MC	x	10		0.48
2,27,52,77,102,127,152,177	3 MC	B	3		0.48
4,29,54,79,104,129,154,179	3 MC	B	10		0.48
5,30,55,80,105,130,155,180	4 C3	x	10		0.48
6,31,56,81,106,131,156,181	4 C3	B	10		0.48
7,32,57,82,107,132,157,182	5 CY	x	10		0.48
8,33,58,83,108,133,158,183	6 DB	x	10		0.48
11,36,61,86,111,136,161,186	6 DB	x	10		0.18
14,39,64,89,114,139,164,189	6 DB	x	10		0.78
17,42,67,92,117,142,167,192	6 DB	B	10		0.48
20,45,70,95,120,145,170,195	6 DB	B	10		0.18
23,48,73,98,123,148,173,198	6 DB	B	10		0.78
9,34,59,84,109,134,159,184	7 DC	x	10	0.0	0.48
12,37,62,87,112,137,162,187	7 DC	x	10	+0.3	0.78
15,40,65,90,115,140,165,190	7 DC	x	10	-0.3	0.18
18,43,68,93,118,143,168,193	7 DC	B	10	0.0	0.48
21,46,71,96,121,146,171,196	7 DC	B	10	+0.3	0.78
24,49,74,99,124,149,174,199	7 DC	B	10	-0.3	0.18
10,35,60,85,110,135,160,185	8 DA	x	10		0.48
13,38,63,88,113,138,163,188	8 DA	x	10		0.18
16,41,66,91,116,141,166,191	8 DA	x	10		0.78
19,44,69,94,119,144,169,194	8 DA	B	10		0.48
22,47,72,97,122,147,172,197	8 DA	B	10		0.18
25,50,75,100,125,150,175,200	8 DA	B	10		0.78

12.5 Application of Assessment Methods to Current Fisheries

To provide examples of the application of the suggested methods from this study data for three species (Flathead, School Whiting and Jackass Morwong) were collated and the following methods applied to generate catch estimates. The species selected each had relatively long catch histories so that the effect of the catch history length could also be investigated empirically:

Depletion Corrected Average Catch – generates estimates of sustainable catch along with estimates of uncertainty about those estimates.

Depletion-Based Stock Reduction Analysis – generates estimates of MSY and F_{MSY}, B_{MSY}, and the final depletion of the stock.

Both of these methods use an underlying production model to mimic the stock dynamics, which is what enables them to provide estimates of uncertainty. This is an advantage of the simple average or median catch methods, which very often under-estimate potential catches, although still provide for better advice than no advice or status quo. The MSE testing is for the application of these methods to truly data-poor fisheries, which is why the time line of 10 years was used. In practice, in the SESSF, it is often the case that longer time frames of catch data are available so the MSE is presenting worst case scenarios. To examine this a simple form of inverse retrospective analysis will also be carried out where the full catch data set will be sequentially culled from the beginning forward.

The catch histories were taken from the latest stock assessments for each species so for Flathead this was from 1915 – 2012. For Jackass Morwong it was from 1915- 2010, and for School Whiting it was from 1980 – 2008.

12.5.1 Further Alternative Tier 5 Methods

Martell and Froese (2014) propose the Catch-MSY a method for estimating MSY from catch data, the maximum rate of population increase, *r*, carrying capacity, *k*, and assumptions about relative stock sizes at the first and final year of the catch data time series. The method randomly draws *r-k* pairs parameters from a Schaeffer production model, from a uniform prior distribution, and then determines whether the parameter pairs are feasible, namely the stock does not go extinct or exceed carrying capacity and the final relative biomass estimate falls within the specified range of depletion. From the set of feasible *r-k* pairs, an estimate of MSY is calculated. The lower and upper values for the uniform distribution of the carrying capacity parameter *k* were set at the maximum catch in the time-series and 100 times the maximum catch. Default values (**Table 10**) for the range of values for *r* were based upon resilience estimates from FishBase: High (0.6-1.5), Medium (0.2-1), Low (0.05-0.5) and Very Low (0.015-0.1). Initial and final estimates of depletion were based on the catches relative to the maximum catch:

Table 10. Default values in the Catch-MSY method for the range of depletion values for the start and the end of a time series in relation to how catches have varied through time in each fishery.

	B/k	Catch/Max Catch
First Year	0.5 - 0.9	< 0.5
	0.3 - 0.6	>= 0.5
Final Year	0.3 - 0.7	> 0.5
	0.01 - 0.4	<= 0.5

On comparison of the Catch-MSY Method with 146 stocks with full stock assessments, Martell and Froese (2014) found excellent agreement between estimates of MSY. This method was relatively new and further developments are being undertaken so this was not included in the testing although once completed this should be attempted.

13 Appendix: MSE Results and Discussion

13.1 Current Criteria for Appropriateness of a Tier

Determining the appropriateness of a particular assessment (and in the SESSF the associated Harvest Strategy) is currently something that has not been attempted formally. In fact, there are no standard, routine methods, or formal criteria that can be applied to determine whether a fisheries stock assessment is appropriate or not independently of the assessment and management process in which it is embedded. Ideally, given data typical of a fishery, one would use simulation testing to determine how well it was possible to estimate the stock status performance measures (whether that is the B_{Curr}/B_0 of a tier 1, the F_{curr} of the tier 3, or the scaling factor SF_{curr} of the tier 4); and by 'how well' is meant how precisely and can it be done without bias. If that were known then decisions could be made as to how precise an estimate was necessary before a particular tier was deemed inappropriate (or possibly too expensive to make it appropriate). But even if this process were to be conducted regularly it would not capture all the possible issues concerning the appropriateness of different assessments. The precision of any estimate is certainly related to how inherently variable the data being used tends to be but can also be greatly affected by whether or not the data used in an assessment is truly representative of the stock as a whole. Individual samples are taken within particular geographical and depth bounds from particular vessels. Spatial differences in the biological characteristics of a fished species (e.g. differences in its growth or size at maturity, shape, etc) often occur and if these are large enough that two samples from two areas can appear very different then to obtain representative sampling of a stock can be either very difficult, very expensive, or both. The assumption that the available data represents the stock as a whole is again difficult to test although high levels of variation in data between years would be indicative that something about the sampling is not managing to capture the full variation within the stock as a whole (**Figure 20**).

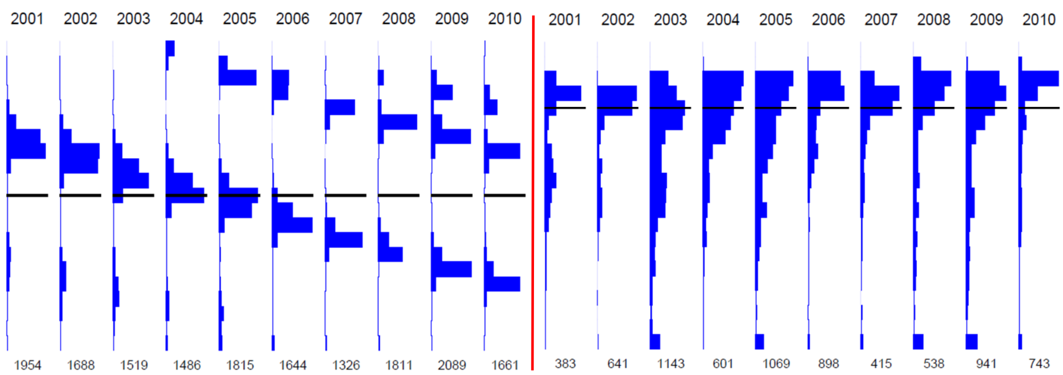

Figure 20. Age distributions sampled from the catches of Blue Grenadier (*Macroronus novaezelandiae*; left-hand graph) and Blue-Eye Trevalla (*Hyperoglyphe antarctica*; right-hand graph) for the years 2001 – 2010 (Klaer *et al*, 2014), illustrating the variation between years by species. Both species have sample sizes that should be sufficient to provide a good representation if the stock were homogeneous in its properties. The Blue Grenadier samples are almost ideal data with clear year classes progressing each year and with consistency through time. Blue-Eye Trevalla, on the other hand, shows inconsistencies every year with annual progressions of year classes being vague and ephemeral at best.

Tier 1 stock assessments based on the Blue Grenadier data use the strong signal in the ageing data combined with other data streams (CPUE, length frequencies, etc.) to provide precise and well defined outcomes. Strictly the Tier 3 modified catch curves cannot be validly applied to the Blue Grenadier data despite the excellent delineation of the cohorts because the selectivity by trawl is dome shaped (**Figure 21**). This implies that the larger, older fish will be under-represented in the catch and any samples from that catch. In turn this would bias any estimates of fishing mortality from the catch curves because the proportion of older fish would be lower than it should be, which would appear as if they had all died. On the other hand, there have been two attempts to produce a Tier 1 assessment using the available Blue-Eye Trevalla data and each time there have been conflicts between any trends apparent in different data streams and plausible solutions cannot be found when trying to fit fully articulated integrated assessment models. If the available data are not representative then a solution is not always possible. As Tukey (1980, p74-75) put it: “The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.”

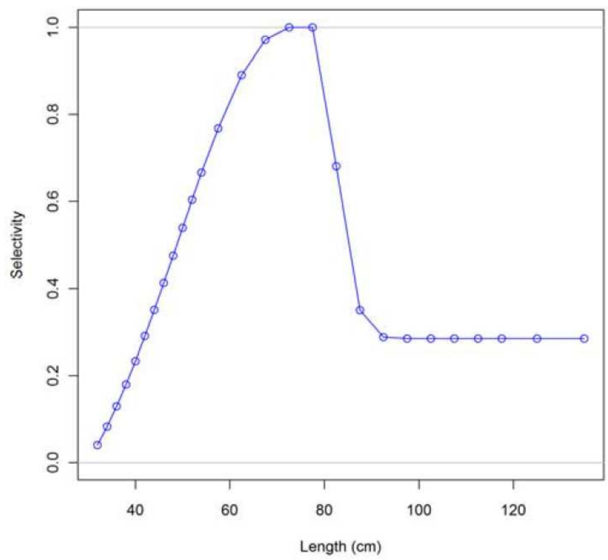

Figure 21. The selectivity curve for Blue Grenadier in the non-spawning fishery in the SESSF. The doming of the selectivity curve implies that there will be fish (and biomass) which are not seen in the samples.

13.1.1 Diagnostic Plots

Currently, when a stock assessment is conducted some diagnostic plots are presented to the different resource assessment groups (RAGs) as a means of displaying graphically how well the model fits the data. Mismatches between the predicted values and those observed often highlight data from particular years as being atypical or at least inconsistent with earlier and later data. The advantage of including a model of the dynamics of a given stock is that this constrains the possible trends within any data stream to at least be consistent through time. The uncertainty and variation inherent in observing any natural system is one reason why the predicted trends in CPUE or age-structures, etc, tend to be smooth and change gradually (although recruitment can certainly differ markedly each year) while the observations from the fishery usually vary far more (**Figure 22**).

While such diagnostic plots are useful for illustrating the relative fit of a model to its data they do not come with specified criteria for quality of fit except where two or models are being directly compared (e,g, likelihood ratio tests, AIC criteria). For example, in **Figure 22**, in 1983 the match between the observed length frequency and the predicted appears to be very good, while that in 1993 does not. This might make the person conducting the assessment look closely at the 1993 data but as the cumulative fit across the whole (and the years not illustrated) is acceptable on average, then the 1993 data is put down to noise and ignored.

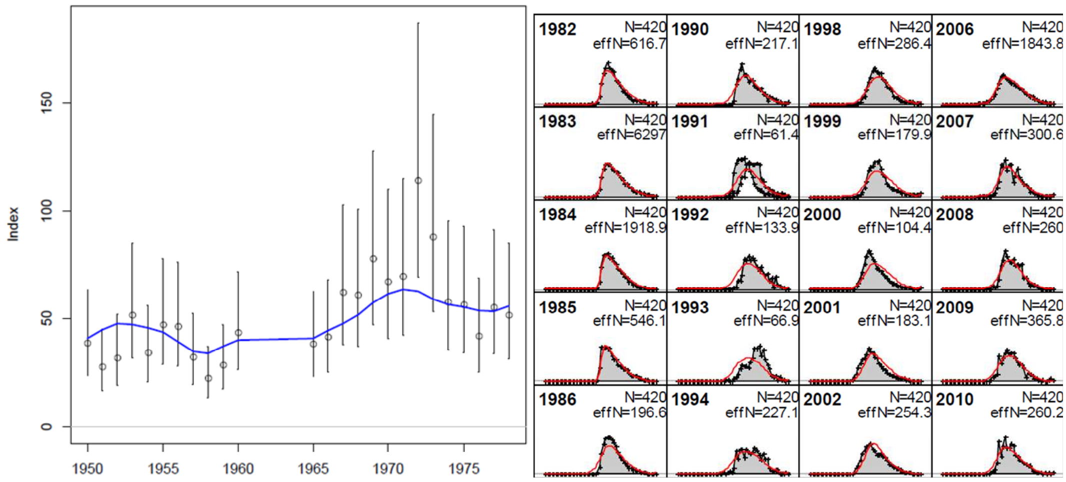

Figure 22. A comparison of the observed with the predicted data from two of the many data streams used in the Tier 1 stock assessment for Tiger Flathead (*Neoplatycephalus richardsoni*). On the left is the CPUE from the early Danish seine fleet with the blue line representing the predicted values and the dots and bars representing the observed CPUE with an estimate of variance around each estimate. On the right is the length composition of the eastern trawl fleet (SESSF zones 10 and 20) for a selection of years. The red predicted lines change smoothly and gradually while the actual observations are less well behaved (this represents an acceptable fit).

There are thus formal methods for comparing the relative quality of the fit of two or more models fitted to the same data but these approaches do not provide guidance when there is only one model and its validity is under question.

13.1.2 Tier Selection

The Resource Assessment Group (RAG) involved with each fishery has the responsibility of providing stock assessment advice to the management agency and part of that is the selection of an appropriate Tier. The selection of which Tier harvest strategy, with its associated assessment method, to apply to a particular fishery, is made or agreed to by the whole RAG rather than just the assessment scientist. So the idea of determining whether a particular tier is appropriate to a given species cannot be determined in a purely statistical manner. Currently there is judgement involved not only with respect to the data used but also about whether the processes involved in the stock dynamics have been modelled in a manner that is realistic or not.

The RAG is free to discuss the quality of fit of data streams to the predicted values, but also whether there are exceptional circumstances occurring in the fishery that might invalidate the application of a particular method. Of particular interest to catch-only status assessment methods is whether the catches are being influenced by factors other than stock availability. The market for a fish can definitely influence whether or not fishers will target or even land a species; currently (2014/2015) for example, the TAC for a number of relatively important species in the SESSF are not being fully landed. The reasons for this have yet to be determined in detail, although from previous Tier 1 stock assessments it does not seem due to a lack of availability of the stocks. This would imply that the current catches are not necessarily indicative of what could be taken. The simple average or median catch methods would be vulnerable to such issues. The DCAC and DB-SRA would be less prone to a problem because they also include an estimate of the final years depletion level, and if catches were artificially low then the use of a higher final depletion might be justified, which should adjust the catches accordingly.

With all stock assessment methods there is room for the development of more formal guidelines for when to reject a stock assessment due to flawed data, and this would be a valuable contribution world-wide. However, if equal importance is whether or not the model or assessment specification provides an adequate representation of the stock dynamics. As well as the usual array of diagnostic plots, it would be a useful addition to all stock assessments to include arguments and a justification for the use of the model structure used (even where that model structure is something very simply such as: CPUE really does provide a linear index of relative abundance through time).

13.1.3 Meeting Assumptions

This would include a discussion of whether or how well the data and fishery meet the assumptions of each stock assessment method and Tier. Thus, if there are arguments that CPUE does not consistently represent the relative abundance of the complete stock through time, then the application of the SESSF Tier 4 would not be appropriate. This is the same as claiming that the Tier 4 in that case would be unable to provide an adequate (precise and unbiased) estimate of the target catch multiplier $SF_{curr,}$ and so could only be expected to provide misleading management advice.

There have been moves to produce a minimum specification of what should be presented in each stock assessment. It is recommended that a section that explicitly discusses and defends the degree to which the assessment selected for a species is having its assumptions met by the data available. Currently such opinions are given verbally to the RAGs during the exposition of the assessment but it should be made explicitly so that it is easier to act upon. This would be especially important in the case of possible Tier 5 catch-only methods because they use a production model to simulate the fishery and this is not fitted to any data except the biological properties of a species (as best they are known).

13.2 Effects of Data Precision and Bias on Tiers 3 and 4

13.2.1 Flathead Tier 3

With Flathead there are clear differences in the relative performance of the 18 different scenarios within the Tier 3 harvest strategy (**Figure 23**; **Table 6**). Changing the sample size of the age sample for either initial depletion arrangement only had minor effects across the range of sample sizes used, with the greatest effect being a slight downward trend in catch variation as sample size increases. Not surprisingly the catches were higher in the lightly depleted scenarios but the effects of sample size were mirrored across both lightly and heavily depleted groups.

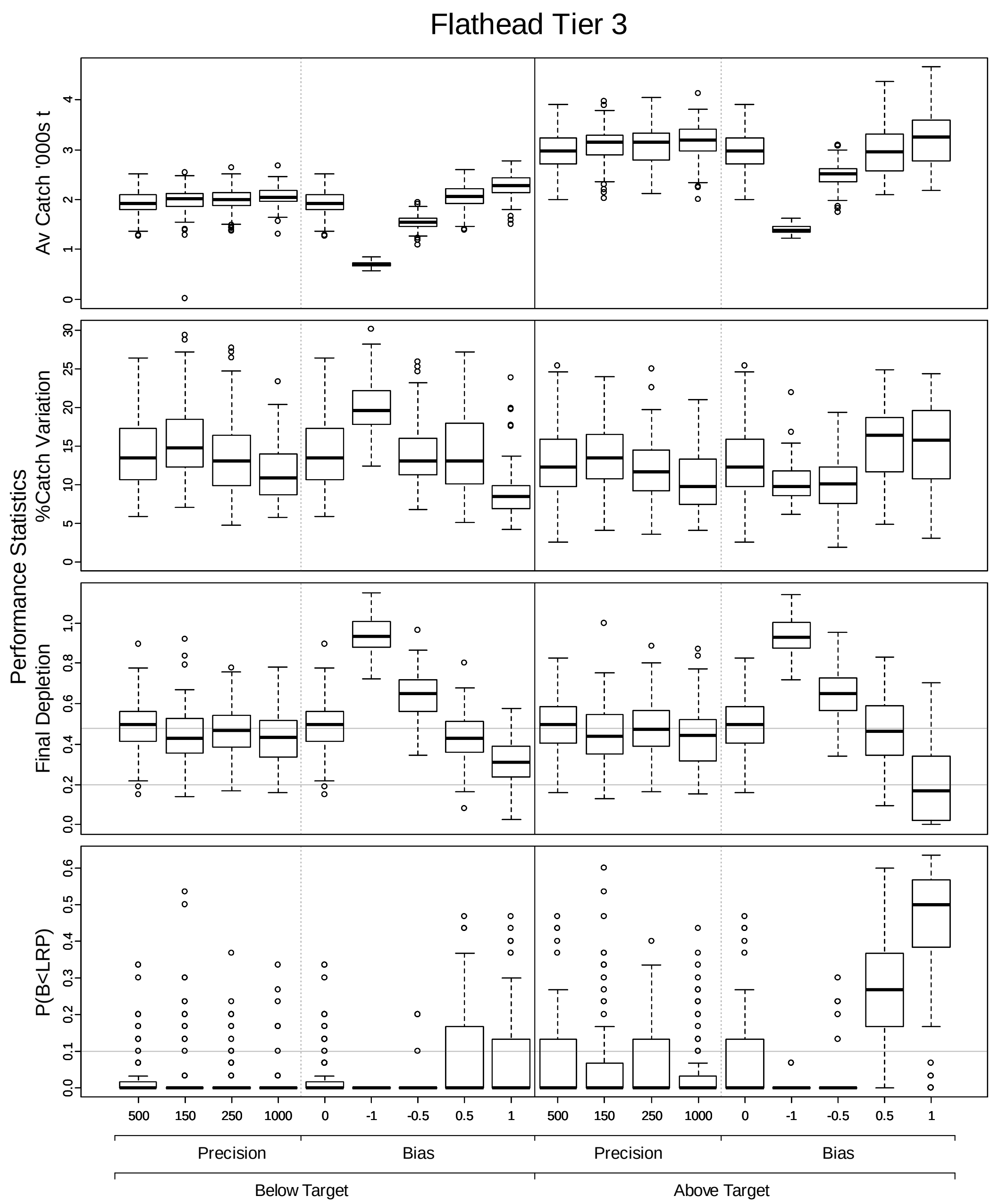

Figure 23. Alternative scenarios tested in the Tier 3 on Flathead; the 150, 250, 500, and 1000 are the sample sizes of ages and the -1 … 1 values are the bias values applied to the ageing samples. The 'Above' and 'Below' relate to the starting depletion levels.
For those stocks that were initially only lightly depleted (78%B_0), while the median probability of not meeting the Limit RP was very low, the spread of values was relatively high with the upper 75th percentile being above the LRP in all cases except the sample size of 1000 (**Figure 23**).

In contrast, the effects of the scale of bias on the age sample applied across the projection period had marked effects (**Figure 23**). Not surprisingly the most biased samples, both positive and negative had the greatest effects although there was also an interaction with the initial depletion level.

The scenarios that were highly depleted initially (18%B_0) exhibited an upward trend in the average catch with increasing bias (from -1 to 1.0), but there were decreasing trends with the average absolute variation (%AAV) and the final depletion level. The median probability of meeting the Limit RP was very low in all cases, although the spread of values increased with increasing bias so there was a greater than 25% chance of exceeding the LRP in the +0.5 and +1.0 bias scenarios (**Figure 23**).

The scenarios that were only lightly depleted initially (78%B_0) also exhibited an upward trend in average catches and a downward trend in final depletion, although in each case the trends were more exaggerated than those in the initially highly depleted stocks. However, the %AAV exhibited rather different behaviour with relatively large increases in the more positively biased scenarios. The median probability of meeting the LRP was low for the negatively biased samples but increased very strongly from zero to positive biases until, for the bias of 1.0 scenario almost all replicate runs failed to avoid the Limit Reference Point (**Figure 23**).

The Tier 3 appears capable of achieving the Target RP of 48%B_0 for Flathead for all levels of age sample size even when starting from low or high levels of initial depletion, but this is the case only if there is only no or only slightly positive sampling bias.

13.2.2 School Whiting Tier 3

The average catches of School Whiting change only slightly with increasing age sample size, however, their absolute level is very sensitive to the state of initial depletion with catches almost doubling for the scenarios starting with a relatively light initial depletion (**Figure 24**). Similarly an increasing trend in average catches is exhibited with increasing bias in the sampling for age for both initial depletion levels, except the catches were more than double in those scenarios beginning in a lightly depleted state (78%B_0) so the changes were more exaggerated between bias levels. Despite the changes in average catches between the two initial depletion states, the average absolute variation in catches appeared very similar with only minor differences between scenarios. In both the increasing age sample size and the increasing bias the %AAV declined slightly but all scenarios exhibited approximately the same degree of variation with the inter-quartile distances being somewhat wider in the initially lightly depleted scenarios (**Figure 24**).

In the scenarios relating to different age sample sizes in terms of the final depletion levels there is little difference between different sample sizes or different initial depletion levels (**Figure 24**).

The patterns in final depletion between the two initial depletion levels were also effectively the same. A negative bias led to much smaller catches and almost complete recovery of the stock up to predicted unfished average levels (although in a naturally varying species such as School Whiting the notion of a stable unfished equilibrium biomass is admittedly artificial). In the most positively biased age samples the final estimates of the median depletion was almost down at the LRP, with the intermediate degrees of bias laying in between the extremes (**Figure 24**).

For all the age sample size scenarios and the negative to zero bias scenarios (across both initial depletion states) the probability of the spawning biomass falling below the LRP again exhibited the same patterns each with very low median values, of probability, and only outlying individual replicate runs failing the LRP. However, for the two positively biased age samples there was more of an impact, especially with the bias of 1.0, in which far more replicates failed to avoid the LRP, with the median probability in the initially lightly depleted scenario falling on the 10% failure line (**Figure 24**).

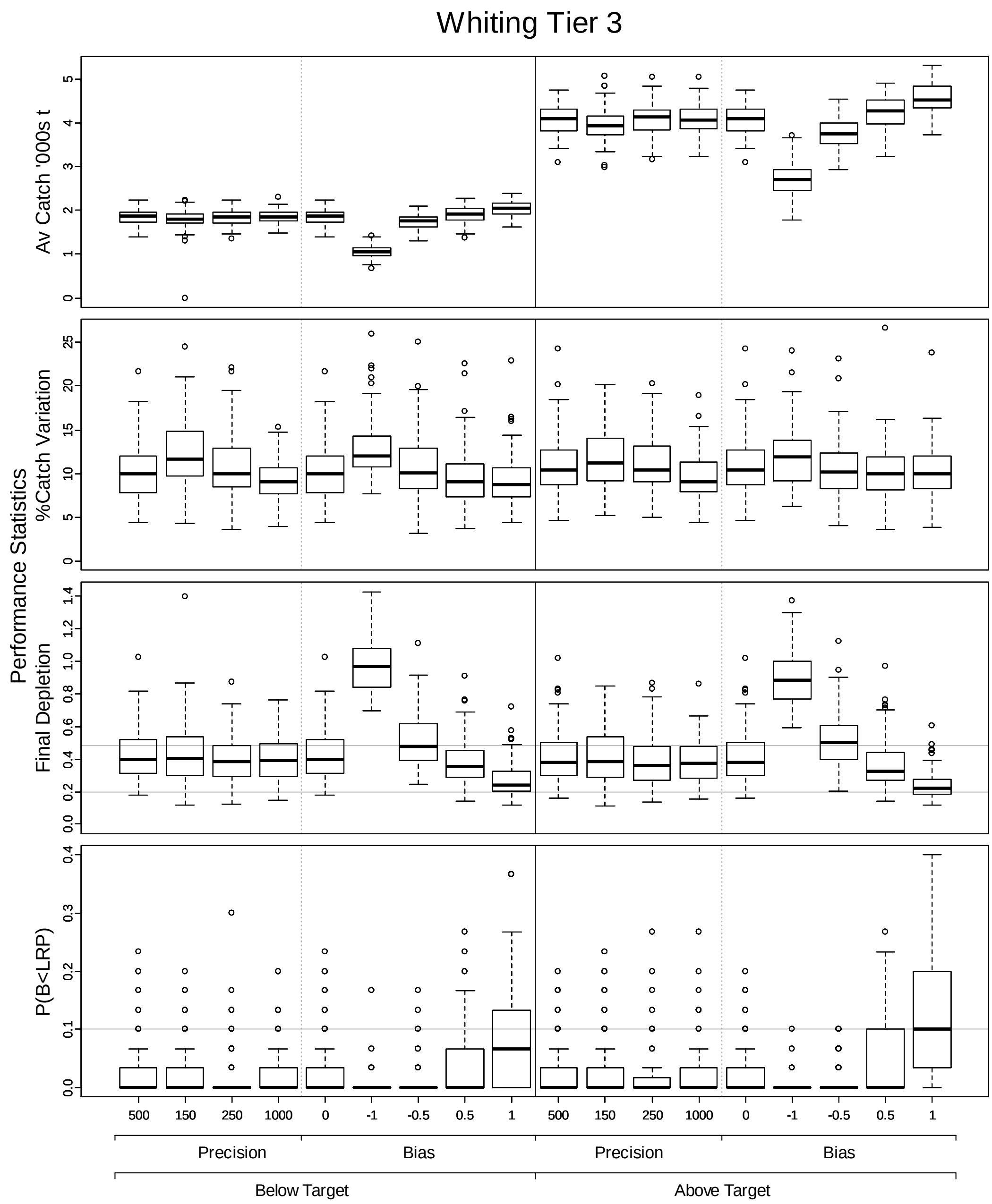

Figure 24. Alternative scenarios tested in Tier 3 on School Whiting; the 150, 250, 500, and 1000 are the sample sizes of ages and the -1 … 1 values are the bias values applied to the ageing samples. The 'Above' and 'Below' relate to the starting depletion levels.

13.2.3 Flathead Tier 4

The effect of differences in the CV of catch rate estimates on the outcome of the Tier 4 analysis can be very great. A CV of even 0.45 can lead to CPUE trends exhibiting enormous variation between years, far more so than is exhibited by real fisheries, so the outcomes from the CVs of 0.6 and 0.9 constitute extreme expected behaviour.

The median value of the average catch slightly declines with increasing CV while the associated spread of values increases markedly. Similarly, above a CV of 0.15 the %AAV more than doubles the values obtained with the CVs of 0.05 and 0.15 (**Figure 25**).

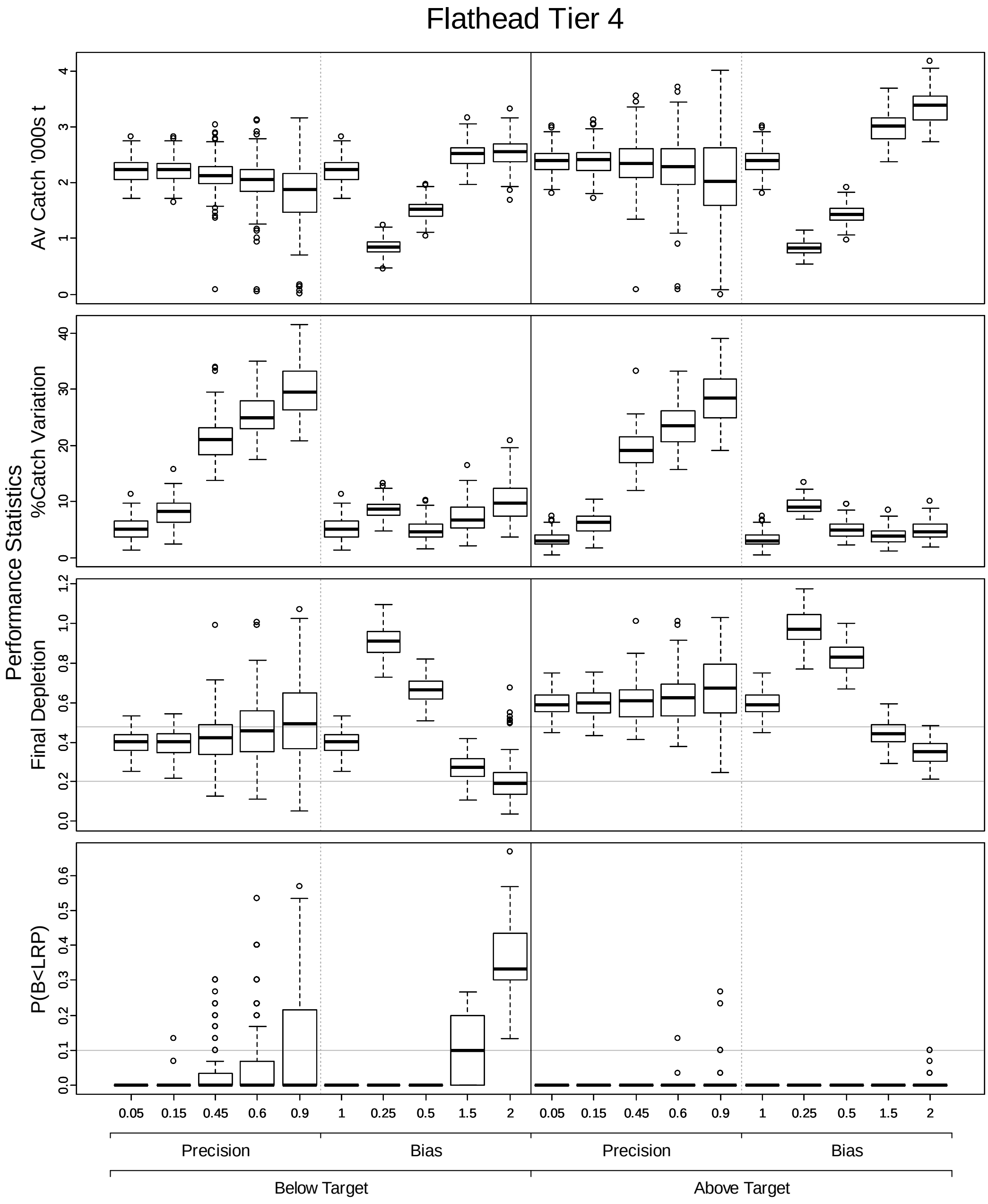

Figure 25. Alternative scenarios tested in the Tier 4 on Flathead; the 0.05 … 0.9 relate to the CV applied to the generate the observed CPUE from the operating model and the 0.25 … 2 values are the bias values applied to the CPUE series. The ‘Above’ and ‘Below’ relate to the starting depletion levels.

The primary effect of increasing the CV value on the final depletion level and on the failure to avoid the LRP was to increase the spread of values, although the median final depletion value also increased slightly up to the Target LP for the greatest CV with all the others falling below the target of 48%B_0. The median probability of failing to avoid the LRP was very low for all scenarios except for the highest CVs and the most positively biased in the initially heavily depleted (18%B_0) stocks, where the median was on the limit for the bias of 0.5 and all replicate runs failed to avoid the LRP in the bias = 1 scenario (**Figure 25**); in the lightly depleted initial state all scenarios succeeded in avoiding the LRP.

13.2.4 School Whiting Tier 4

School Whiting have similar outcomes to the Flathead within the Tier 4 HS for a number of particulars, however, there are significant differences. While the increase in the CV leads to increases in the variability of the average catches, the final depletion, and the probability of failing to avoid the LRP, the effects on the median value for each of those statistics was only relatively minor. This was the same in both the lightly and heavily depleted initial states, although the median values for the average catch and final depletion were both higher in the lightly depleted scenarios and the probability of failing to avoid the LRP was very close to zero for the lightly depleted scenarios but had median values between 0 and 7% in the scenarios from the initially highly depleted scenarios with the upper interquartile bound touching the 10% line for the top two (unrealistic) CV values (**Figure 26**).

The pattern of %AAV was very similar between the two initial depletion states but the range was more extreme and differences between levels exaggerated in the initially lightly depleted scenarios.

Once again the effect of bias on the median statistics was more extreme than the effect of the CV changes with the final depletion levels being inversely related to the average catch levels. But in the initially highly depleted scenarios the average catches of the low biases (0.25 and 0.5) were greater than from the lightly depleted scenarios but the stronger biases were less than their respective scenarios in the lightly depleted scenarios. Despite this the final depletion levels were all lower in the highly depleted initial state scenarios than the respective scenarios in the lightly depleted replicates. Thus the lightly depleted scenario with the maximum positive bias had a median final depletion value very close to the target 48%B_0, while that in the initially heavily depleted scenario had a median value close to the LRP (**Figure 26**).

13.2.5 The Effect of Inconsistently Occurring Biases

In this work only biases that occur consistently through time are examined. This assumes that if biases are inconstant, perhaps through the distribution of sampling changing markedly between years, any effects are assumed to appear as process error through time in any analysis. Thus, the apparent variability of the data, and any related population size signal, would be greatly increased so that the data source may even become uninformative although not necessarily in conflict with other data sources. This assumption would be reasonable as long as any biases remained constant in their intensity. If in some years there was much higher levels of bias than in others that may confuse any intrinsic signal in the data, which, if it were directional, might lead to incorrect outcomes.

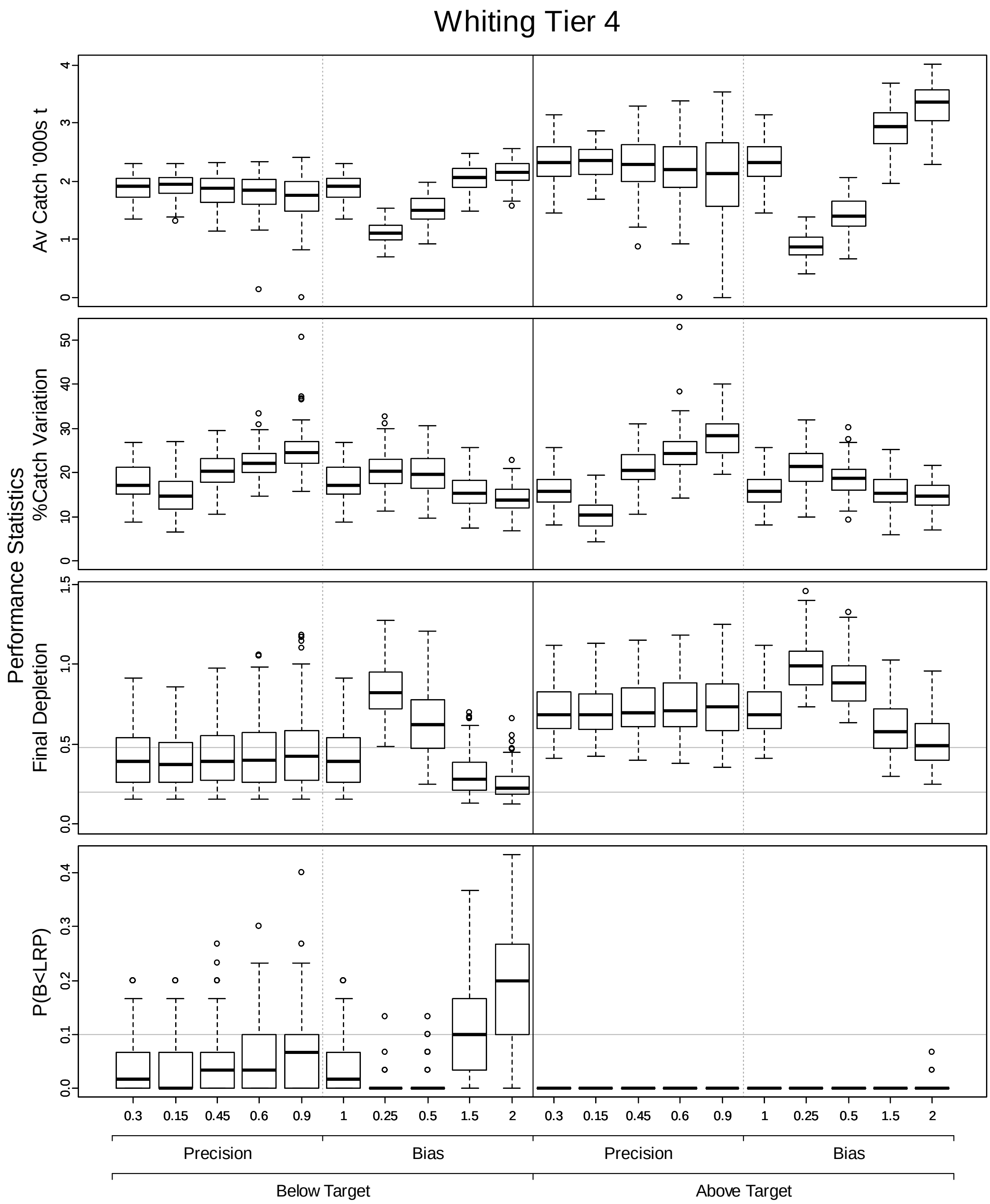

Figure 26. Alternative scenarios tested in the Tier 4 on School Whiting; the 0.05 … 0.9 relate to the CV applied to the generate the observed CPUE from the operating model and the 0.25 … 2 values are the bias values applied to the CPUE series. The ‘Above’ and ‘Below’ relate to the starting depletion levels.

13.3 Data-Poor

The outcomes from the 200 different scenarios are illustrated in **Figure 27 – Figure 50**, with triplets of graphs relating to a particular species. Actual initial stock status ('H' means highly depleted, so starting below the target, and 'L' means lightly depleted, so starting above the target of 48%B_0), and finally whether a discount of 25% was imposed on the catch recommendation or not. There are 24 sets of plots: these include boxplots, median trajectories across all scenarios, and randomly selected individual trajectories from a selection of two scenarios from each of the 8 combinations of species, initial status, and discount. The box plots illustrate the four harvest strategy performance measures (see section 12.4.1), each of which relates to the whole projection period of 30 years except the spawning biomass depletion level, which relates to the final year only. Such box plots can provide a snapshot or summary of the dynamics but in order to visualize the effects of the different harvest strategies upon the stock dynamics, trajectories of expected depletion, cpue, and annual catch are provided for each scenario to illustrate the range of outcomes. Finally, some randomly selected replicate runs (out of the 100 replicates in each case) are used to illustrate the unsmoothed trajectories that may reflect reality more closely than the relatively smooth median trajectories.

13.3.1 Flathead

Those scenarios that assumed the stock was up above the target (assumed to be at 78%B_0) in 2009, irrespective of whether the stock was actually above or below the target in the operating model (**Table 5**), invariably over-estimated safe levels of catches. This led to potentially severe to catastrophic final depletion levels, high catch variability, and significantly high probabilities of failing to avoid the limit reference point (**Figure 27**, **Figure 30**, **Figure 33**, **Figure 36**). Only the DCAC in the scenarios with a lightly depleted stock status and a 25% discount on catches managed to keep the median depletion level above 20%B_0 (**Figure 30**).

The scenarios that assumed the stock was below the target (assumed to be at 18%B_0) in 2009, irrespective of whether the stock was actually above or below the target in the operating model (**Table 5**), generally under-estimated safe levels of catches. This in turn led to final depletion levels often well above the target and generally zero chance of failing to avoid the limit reference point. The DCAC, once again differed from the rest in that it permitted, on average, higher catches than the other harvest strategies, so that the final depletion was generally lower, although still above the target except in the scenarios where the stock was actually heavily depleted and no discount was applied to catches. However, even in that circumstance both scenarios 65 and 68 (**Figure 33**) both median final depletions were above the limit reference point, although a few replicates failed to avoid the LRP.

The scenarios which assumed that the stock was at the target of 48%B_0 in 2009 had different outcomes depending on the real initial status and whether there was a discount or not. If the initial stock status was only lightly depleted then in all cases the median probability of falling below the LRP was very small, although a few replicates in methods 3, 4, 7, and 8 did go below when there was no discount on catches (**Figure 27**). If there was no catch discount then all scenarios except 7 and 8, the MCY and DB-SRA using the initial catch without updating, had a final depletion level below the target, although in all cases above the limit. Whereas with the catch discount all scenarios had a final depletion above the target, with some being well above the target (**Figure 30**).

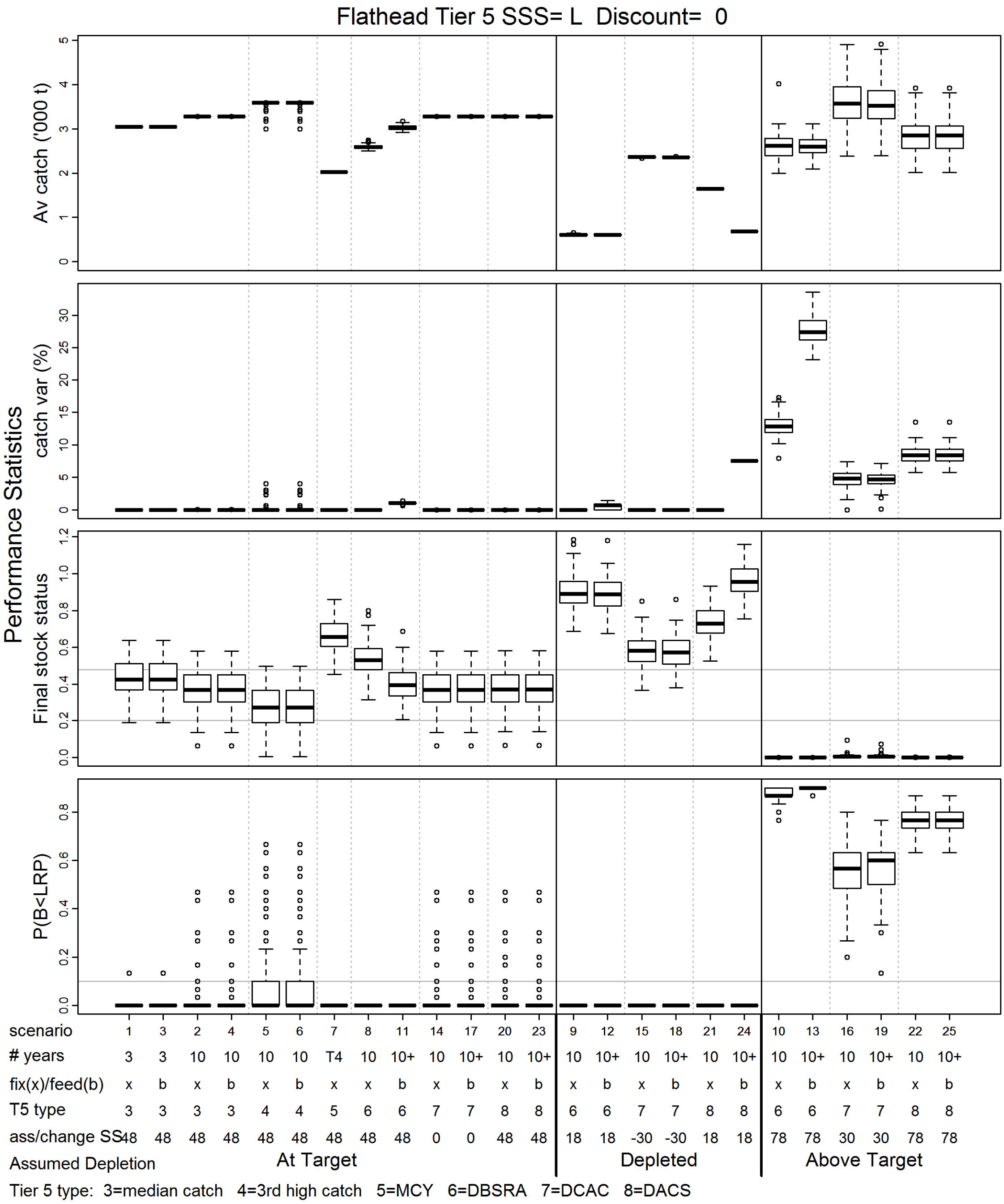

Figure 27. 'ass/change SS' means the current stock status assumed by the method, and for DCAC, 'change SS' means the assumed change in stock status over the catch history. '# years' is the number of years of catches used in the method: T4 means use Tier 4 reference period, 10+ means use last 10 historic years to current. Fix/feed implies fixed TAC setting vs dynamic (or feedback).

When the initial stock status was actually highly depleted so the stock started below the target then the discount on catches was very influential on the outcomes of those scenarios which assumed the stock was at the target when each harvest strategy was first applied. In all cases where there was no discount, the catches were set too high with the consequence that the final depletion was effectively zero except for the MCY method and the DB-SRA with a fixed catch estimate (scenarios 57 and 58; **Figure 33**).

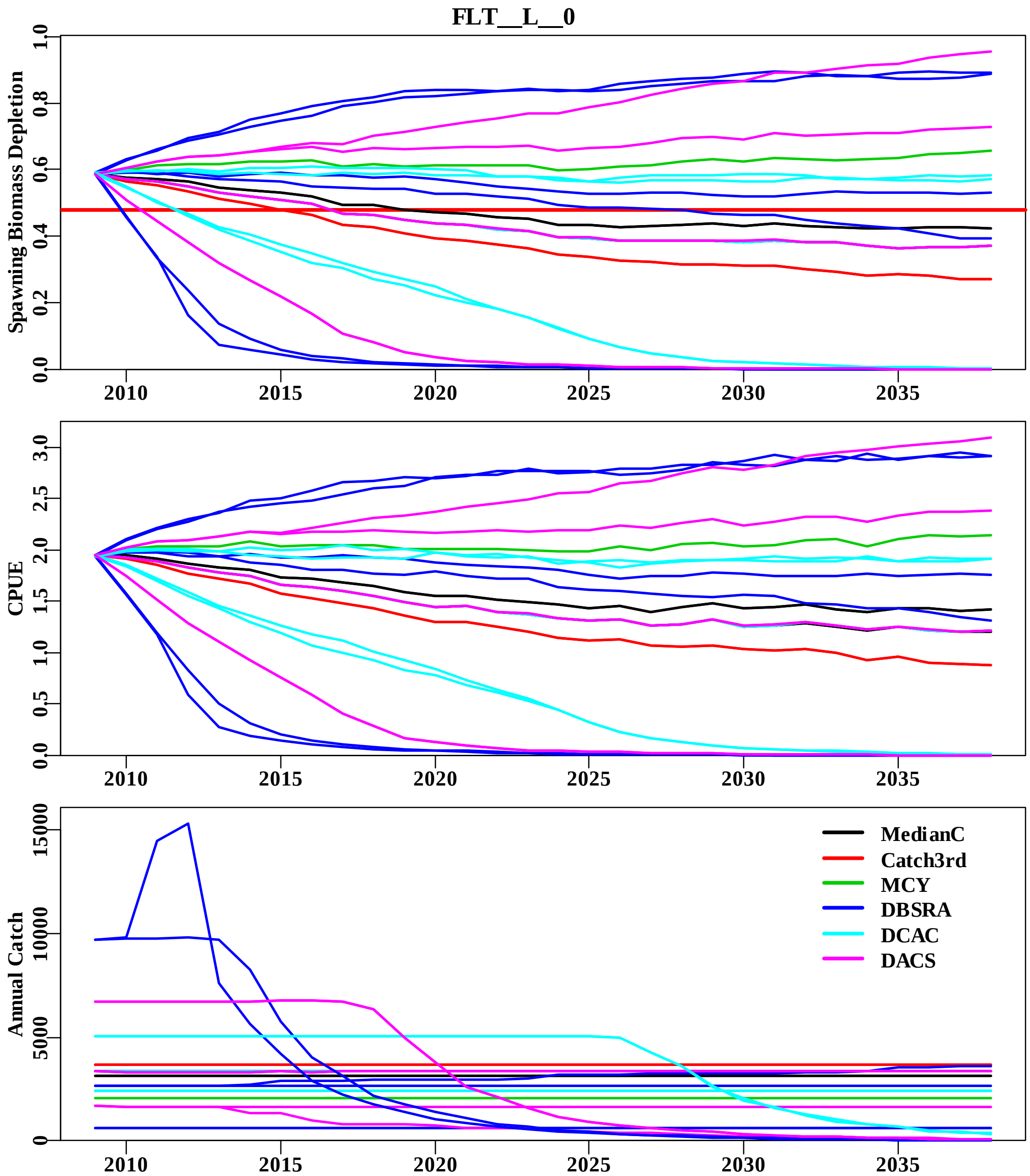

Figure 28. The median trajectories across the 100 replicates for each of the 25 scenarios described in **Figure 27**. The red line in the top figure is the target spawning biomass depletion level.

With a catch discount, on the other hand, some approaches were at or above the target and all were above the limit except the third highest catch with a fixed initial catch strategy (**Figure 36**), which had a median depletion level on the limit; naturally that strategy also had a median probability of falling below the limit which was above the 10% threshold.

In all sets of scenarios the wide bounds placed on the initial conditions meant there were some initial conditions that were implausible, with extremely high annual catches, which in almost all cases eventually led to stock collapse (**Figure 28**, **Figure 31**, **Figure 34**, **Figure 37**). Despite these being implausible in real-life management (given the history of catches, nobody would allocate annual catches of 7-10,000 tonnes of Flathead),

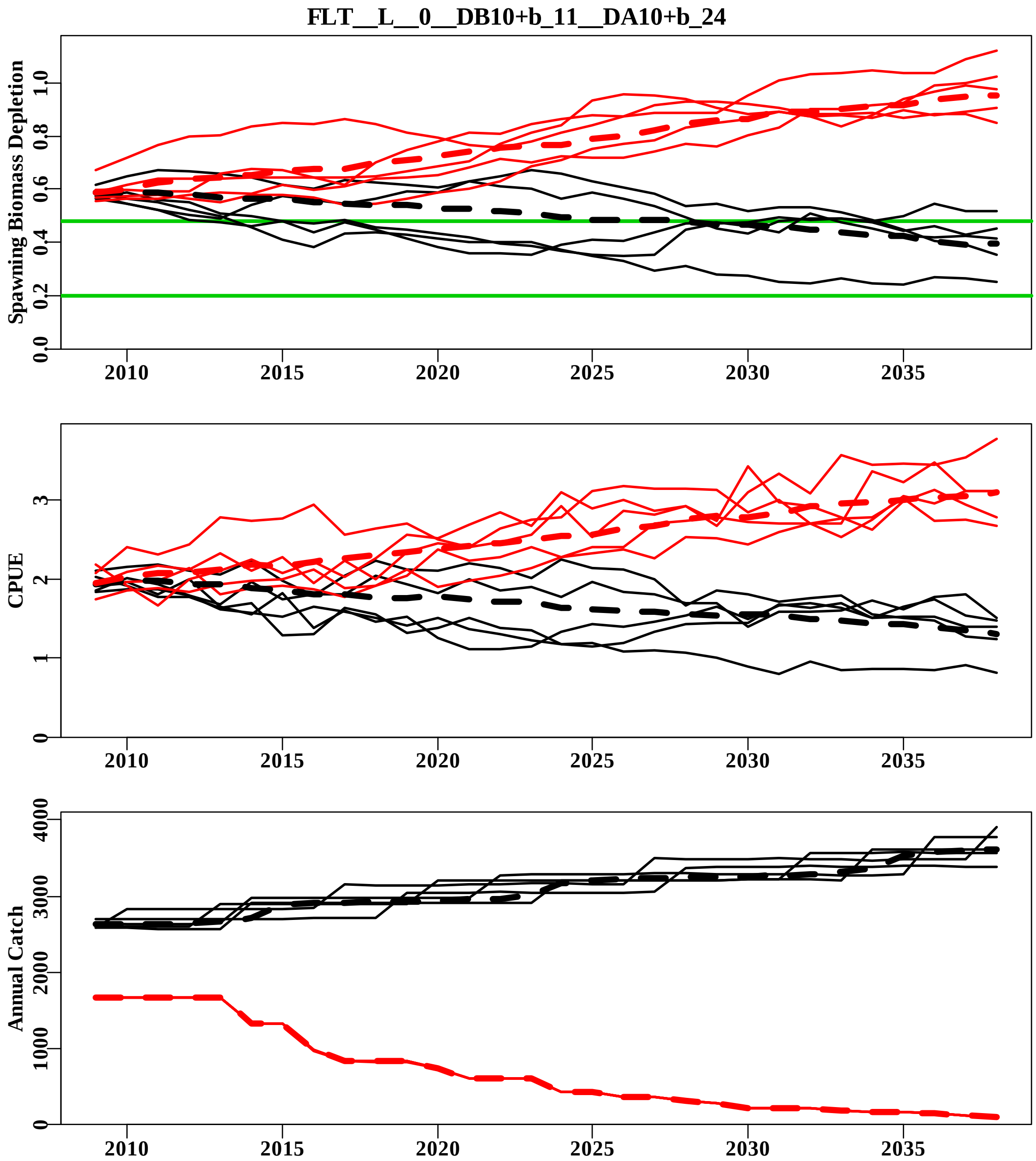

Figure 29. Five randomly selected trajectories from each of two scenarios selected to illustrate contrasting behaviour among the scenarios. The figure heading identifies the species, the actual starting stock status, the discount on catches, the Tier 5 approach (black lines relate to the first, red lines to the second Tier 5), the number of years used, and whether the scenario used a fixed TAC or a dynamic one ('b'), finally the scenario number as depicted on the boxplots and in **Table 9**. The dashed lines are the medians of the 100 replicates for each of the selected scenarios.

these scenarios illustrate that the full range of possibilities have been considered. Thus the median trajectories of spawning depletion range from zero up to almost 100% (1.0).

By considering the trajectories it is clear that in those stocks which begin as only lightly depleted and above the target most harvest strategies avoid increasing well above or below the target with the exceptions being concentrated in those scenarios that assumed the stock was either lightly or heavily depleted rather than at the target. It also becomes clear that the prevalence of stocks finishing above the target increases with the 25%

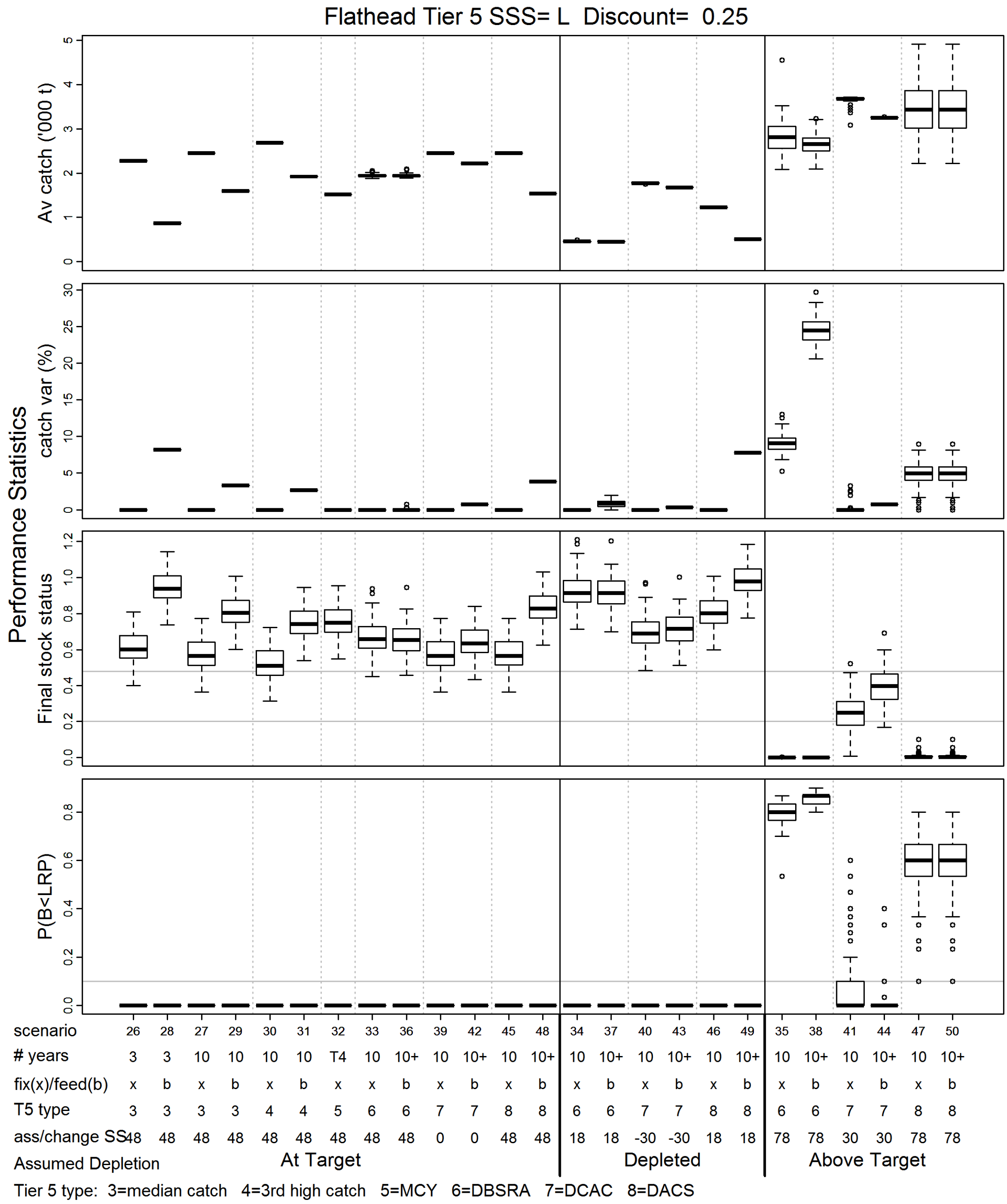

Figure 30. 'ass/change SS' means the current stock status assumed by the method, and for DCAC, 'change SS' means the assumed change in stock status over the catch history. '# years' is the number of years of catches used in the method: T4 means use Tier 4 reference period, 10+ means use last 10 historic years to current. Fix/feed implies fixed TAC setting vs dynamic (or feedback).

catch discount (**Figure 31**). In the scenarios starting with a stock initially depleted below the target then most scenarios ended up being highly depleted. It was primarily those scenarios that assumed the stock was well below the target, which accepted that reduced catches were all that would ever be produced, and ended with depletion levels above the limit and some above the target (**Figure 34**). However, where there was a discount on catches in the initially depleted stock scenarios, then most scenarios finished with final depletions above the limit and many above the target (**Figure 37**).

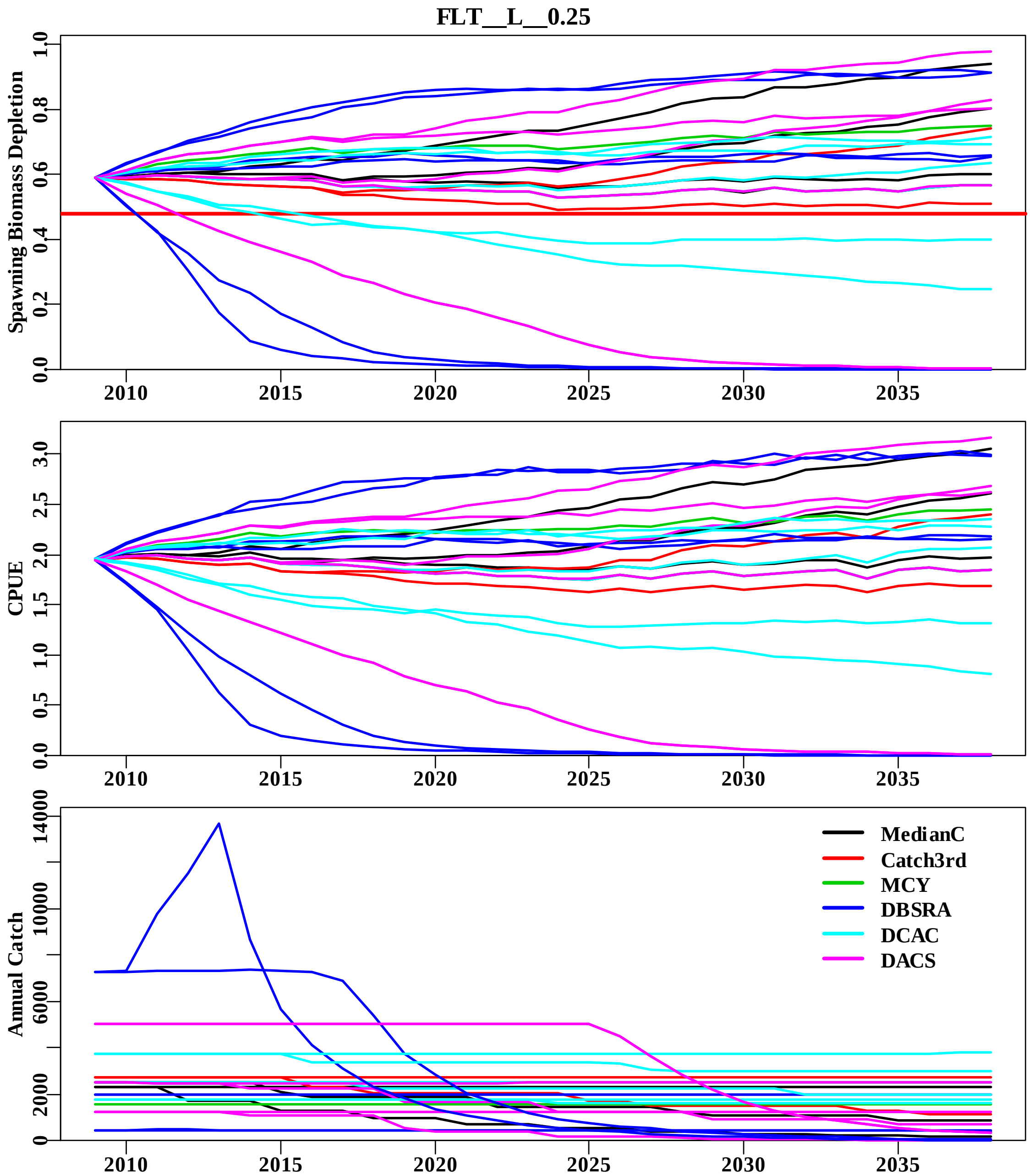

Figure 31. The median trajectories across the 100 replicates for each of the 25 scenarios described in **Figure 30**. The red line in the top figure is the target spawning biomass depletion level.

The random trajectories from the various scenarios selected for each combination of Flathead, initial stock status, and discount, illustrate that the individual replicates tend to be very different from the median trajectory (**Figure 29**, **Figure 32**, **Figure 35**, **Figure 38**). In some cases relatively large changes in CPUE and catch can occur between adjacent years, usually with associated changes in the relative stock depletion. The smoothness of the median trajectories misrepresents what might be expected in a real fishery. By considering individual scenarios in detail it is possible to discern the drivers behind change. For example, in the initially lightly depleted stock when there is no discount, trajectories are illustrated for the DACS method which is updated through time. This updating has the effect of dropping the catches to very low levels even though the

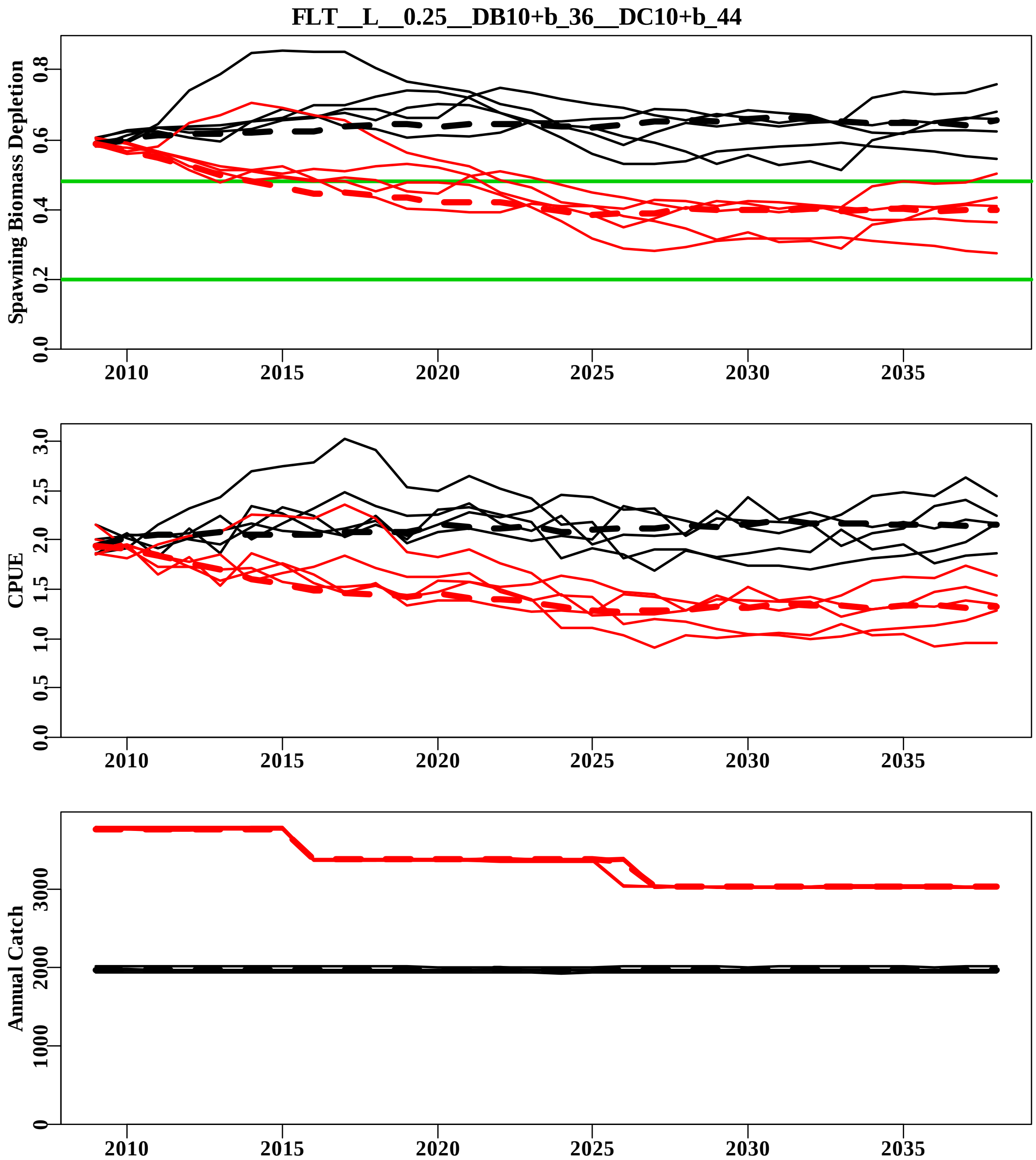

Figure 32. Five randomly selected trajectories from each of two scenarios selected to illustrate contrasting behaviour among the scenarios. The figure heading identifies the species, the actual starting stock status, the discount on catches, the Tier 5 approach (black lines relate to the first, red lines to the second Tier 5), the number of years used, and whether the scenario used a fixed TAC or a dynamic one ('b'), finally the scenario number as depicted on the boxplots and in **Table 9**. The dashed lines are the medians of the 100 replicates for each of the selected scenarios.

spawning biomass depletion and CPUE are both increasing to relatively high levels. Such behaviour provides no beneficial trade-offs; very high catch rates at the cost of almost no catch are not really beneficial. Similarly, in the highly depleted set of scenarios that included a discount when updating DCAC, in which it was assumed the stock started well above the target, led to a very high initial catch level, which led to a rapid decline in the CPUE until finally the high catches could not be maintained and the stock and catches collapses (**Figure 38**). In a real world situation within Australia, it should

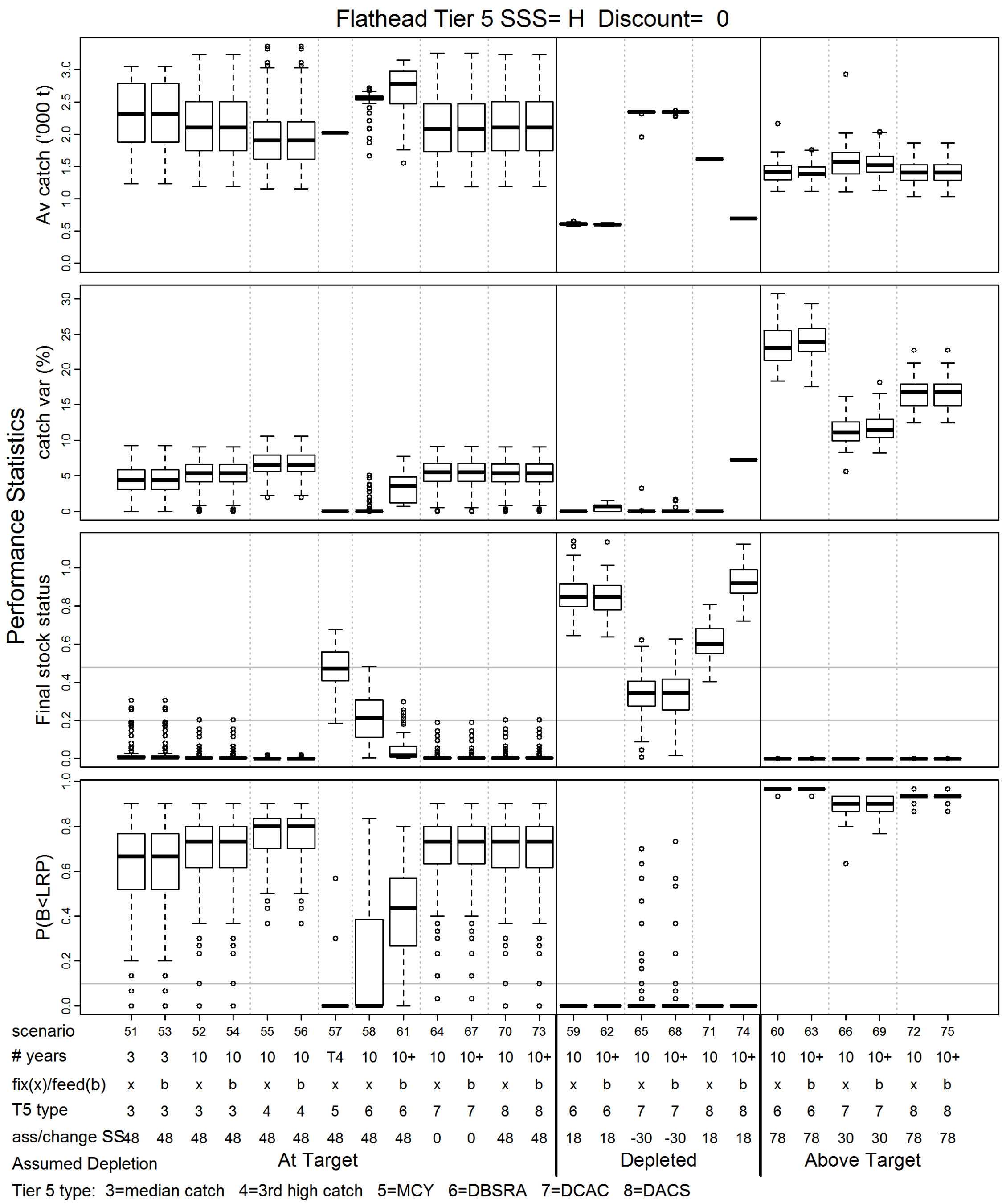

Figure 33. 'ass/change SS' means the current stock status assumed by the method, and for DCAC, 'change SS' means the assumed change in stock status over the catch history. '# years' is the number of years of catches used in the method: T4 means use Tier 4 reference period, 10+ means use last 10 historic years to current. Fix/feed implies fixed TAC setting vs dynamic (or feedback).

be assumed that the initial catch levels were at historical high levels and that catch rates, even if only available sometimes would indicate problems.

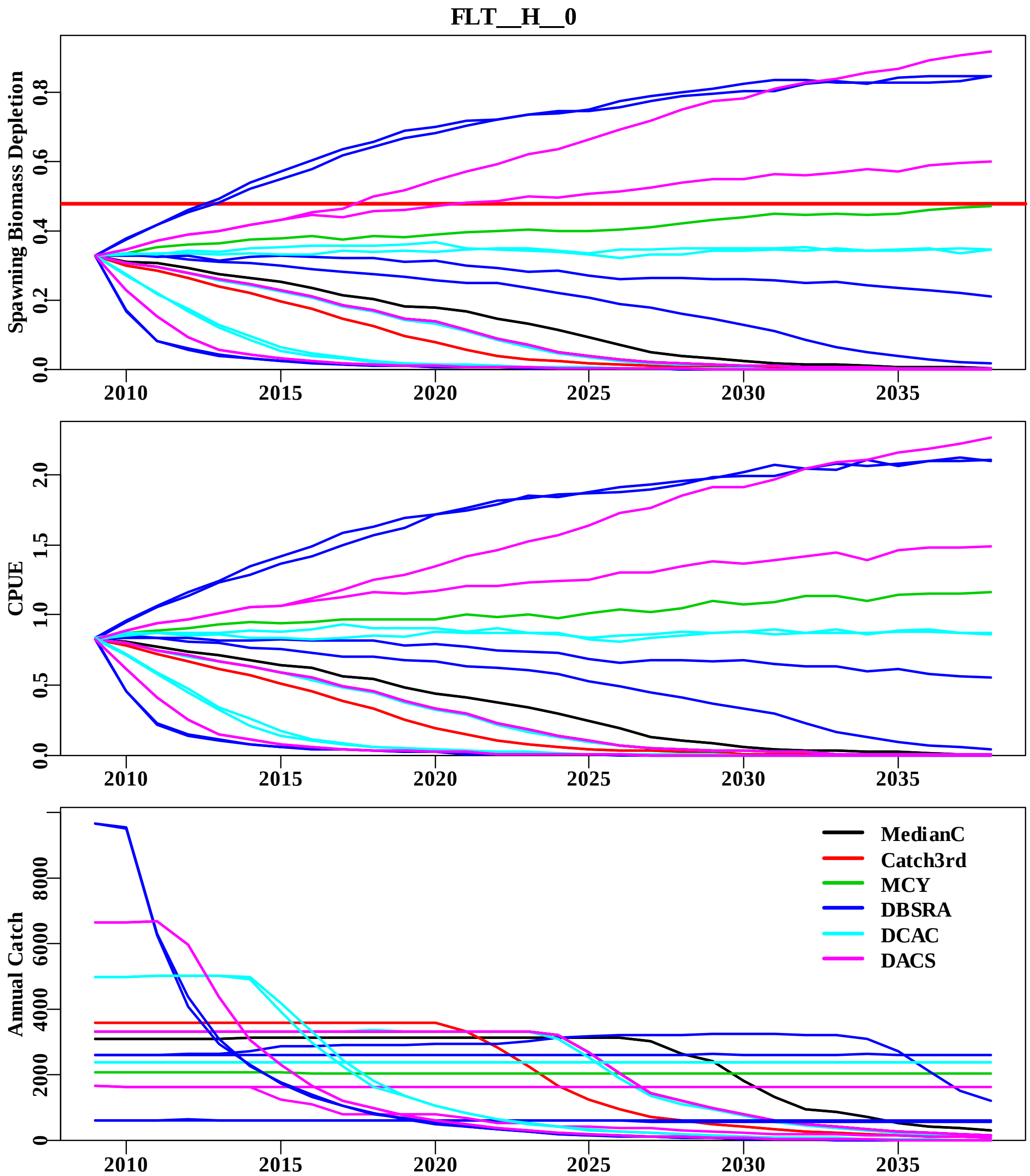

Figure 34. The median trajectories across the 100 replicates for each of the 25 scenarios described in **Figure 33**. The red line in the top figure is the target spawning biomass depletion level.

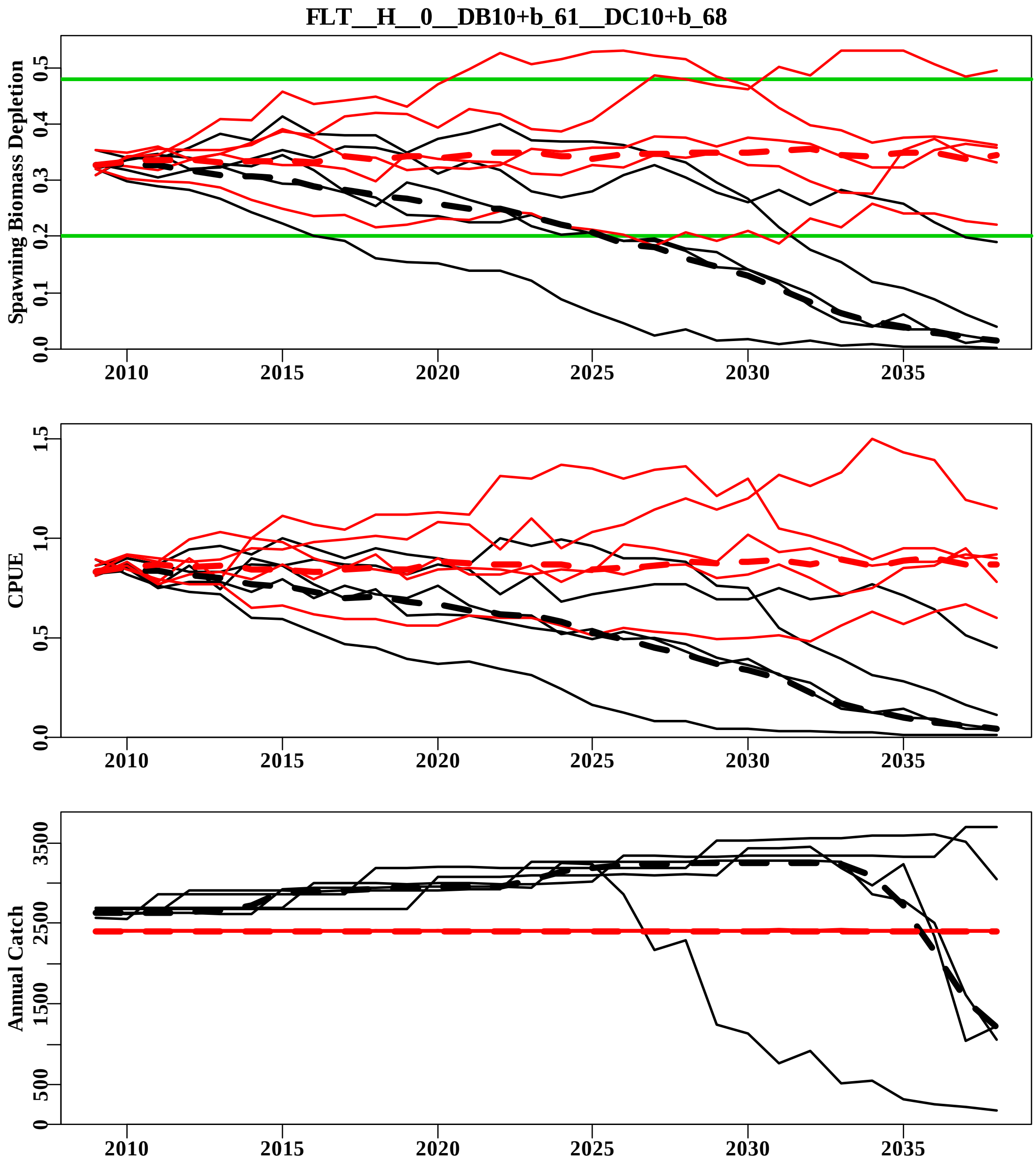

Figure 35. Five randomly selected trajectories from each of two scenarios selected to illustrate contrasting behaviour among the scenarios. The figure heading identifies the species, the actual starting stock status, the discount on catches, the Tier 5 approach (black lines relate to the first, red lines to the second Tier 5), the number of years used, and whether the scenario used a fixed TAC or a dynamic one ('b'), finally the scenario number as depicted on the boxplots and in **Table 9**. The dashed lines are the medians of the 100 replicates for each of the selected scenarios.

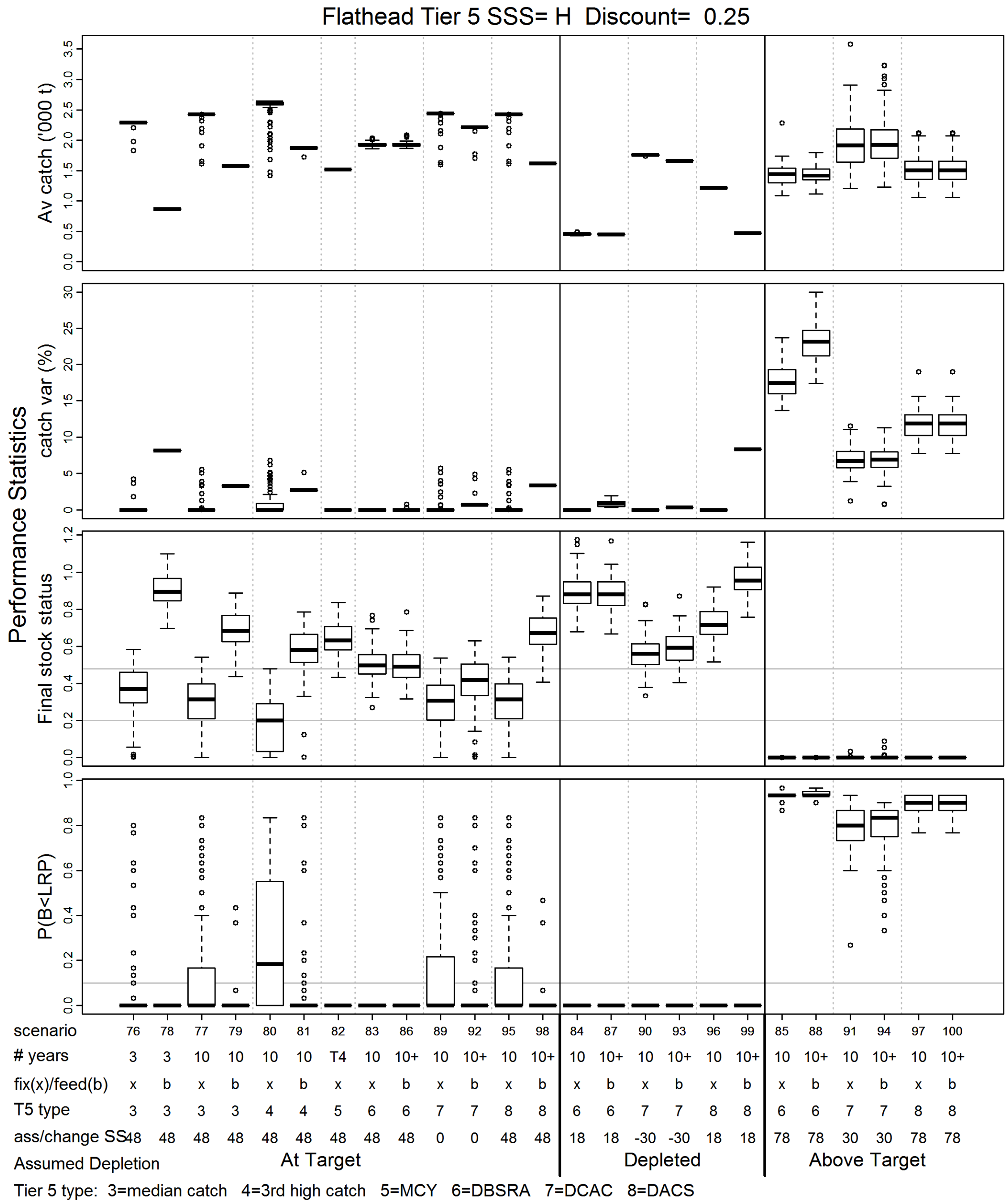

Figure 36. 'ass/change SS' means the current stock status assumed by the method, and for DCAC, 'change SS' means the assumed change in stock status over the catch history. '# years' is the number of years of catches used in the method: T4 means use Tier 4 reference period, 10+ means use last 10 historic years to current. Fix/feed implies fixed TAC setting vs dynamic (or feedback).

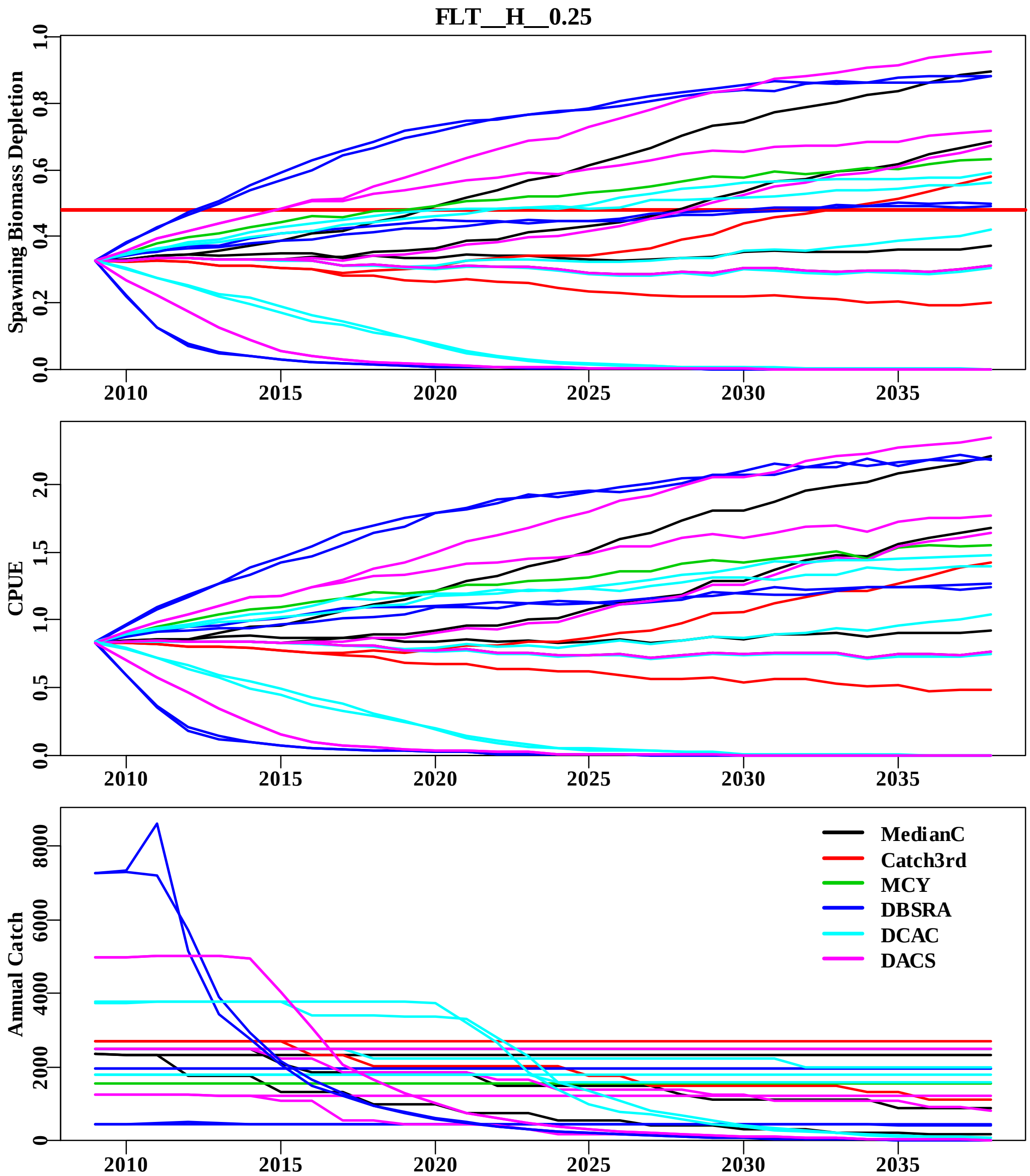

Figure 37. The median trajectories across the 100 replicates for each of the 25 scenarios described in **Figure 36**. The red line in the top figure is the target spawning biomass depletion level.

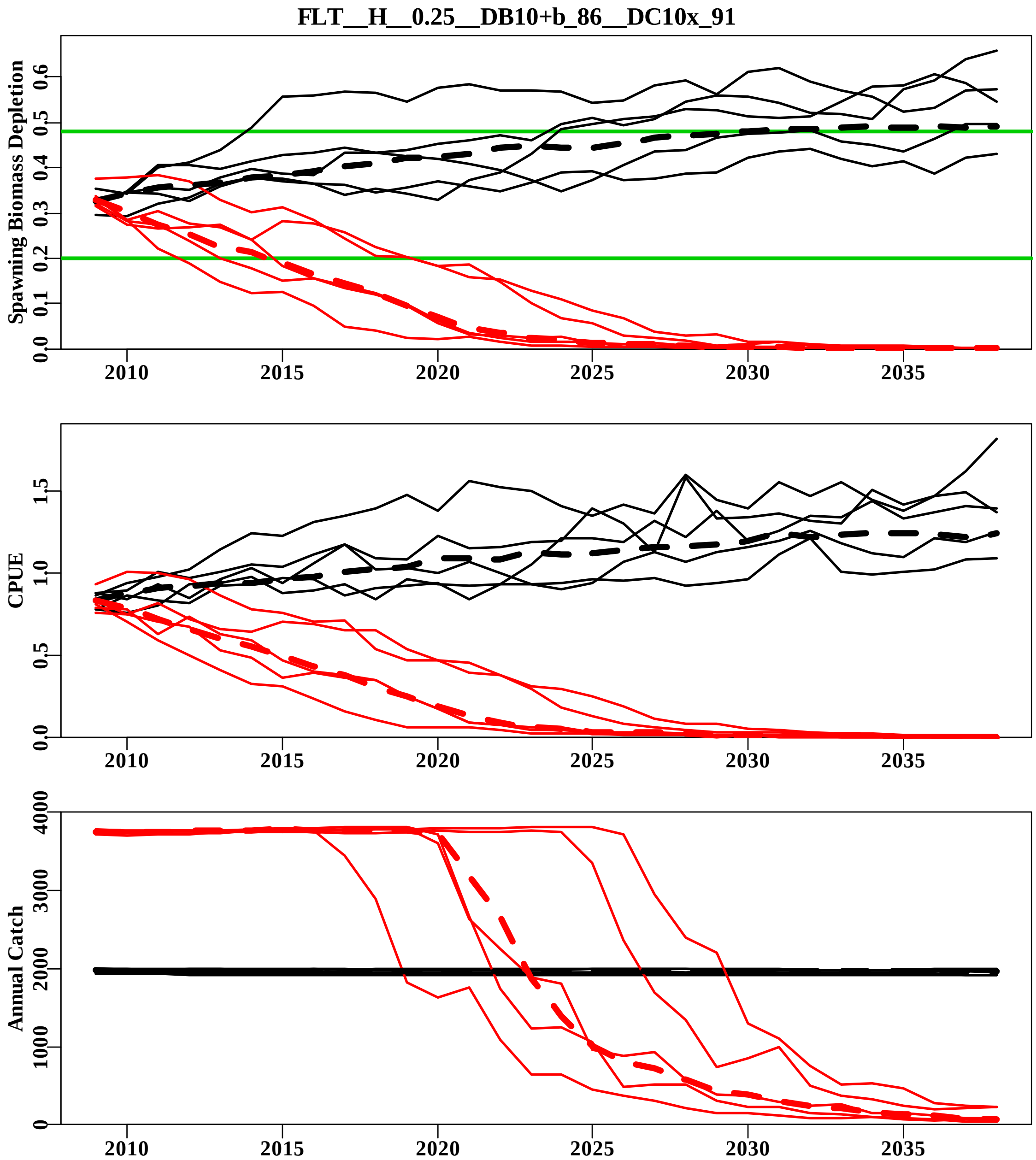

Figure 38. Five randomly selected trajectories from each of two scenarios selected to illustrate contrasting behaviour among the scenarios. The figure heading identifies the species, the actual starting stock status, the discount on catches, the Tier 5 approach (black lines relate to the first, red lines to the second Tier 5), the number of years used, and whether the scenario used a fixed TAC or a dynamic one ('b'), finally the scenario number as depicted on the boxplots and in **Table 9**. The dashed lines are the medians of the 100 replicates for each of the selected scenarios.

An alternative way of illustrating how the dynamics occur is to include a phase plot of catches against the depletion level in which the time series nature of the data is implicit (**Figure 51**, **Figure 52**). The four sets of scenarios can be combined into a figure with four panels. Catches above 5000 t have been truncated to provide more separation between the scenarios with more realistic catch levels (realism being defined as within historical bounds).

13.3.2 School Whiting

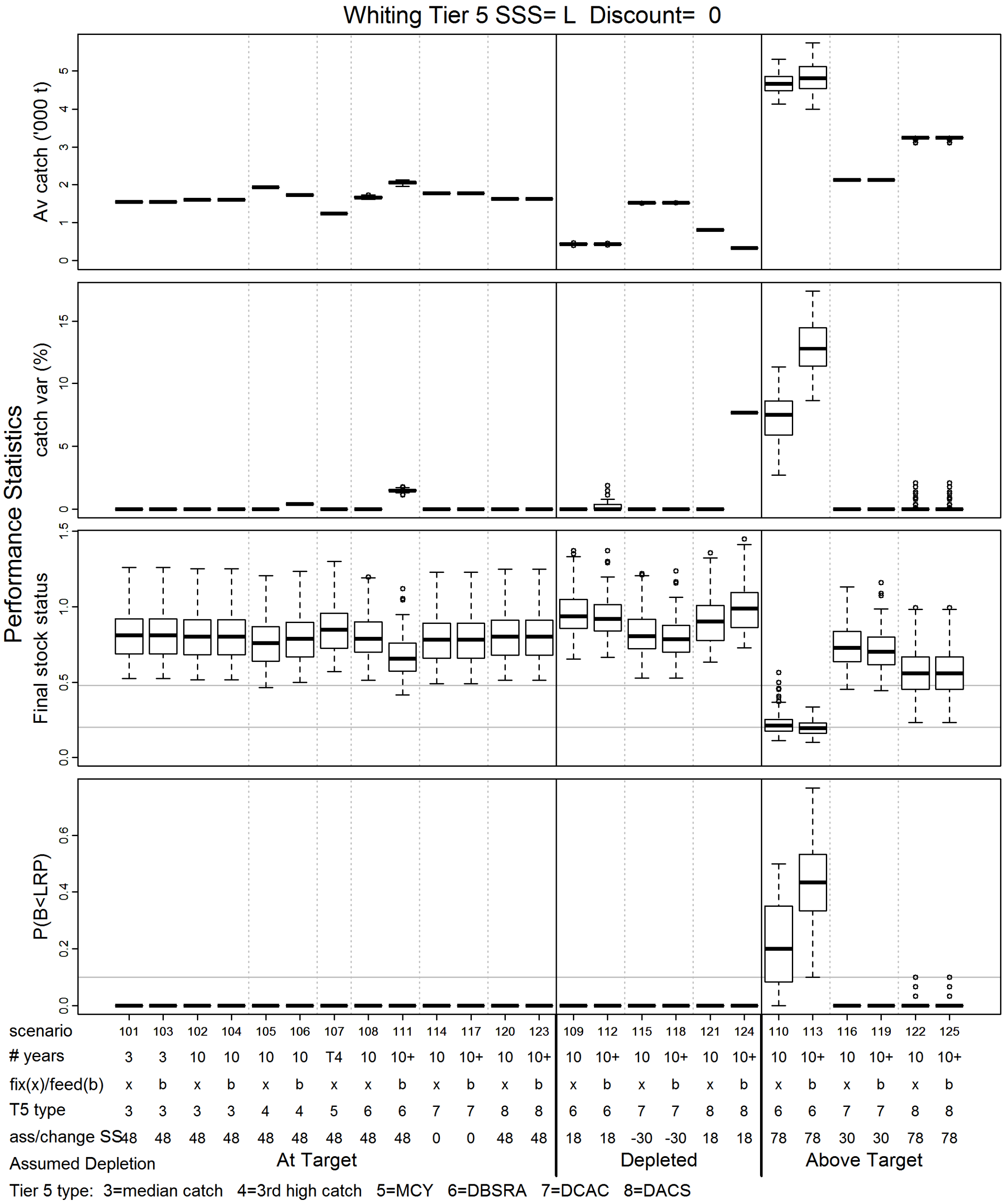

Figure 39. 'ass/change SS' means the current stock status assumed by the method, and for DCAC, 'change SS' means the assumed change in stock status over the catch history. '# years' is the number of years of catches used in the method: T4 means use Tier 4 reference period, 10+ means use last 10 historic years to current. Fix/feed implies fixed TAC setting vs dynamic (or feedback).

With School Whiting those scenarios where the stock was up above the target (assumed to be at 78%B_0) in 2009 (**Table 5**) and with no discount, all led to relatively low level of catch (except for DB-SRA when it assumed the stock was actually up at 78% (**Figure 39**). DB-SRA in the assuming above the target section led to excessive catches, high catch variability, finishing with a median depletion effectively on the limit reference point, and relatively high probability of failing to stay above the LRP. All other scenarios were exactly different from this (**Figure 39 - Figure 41**).

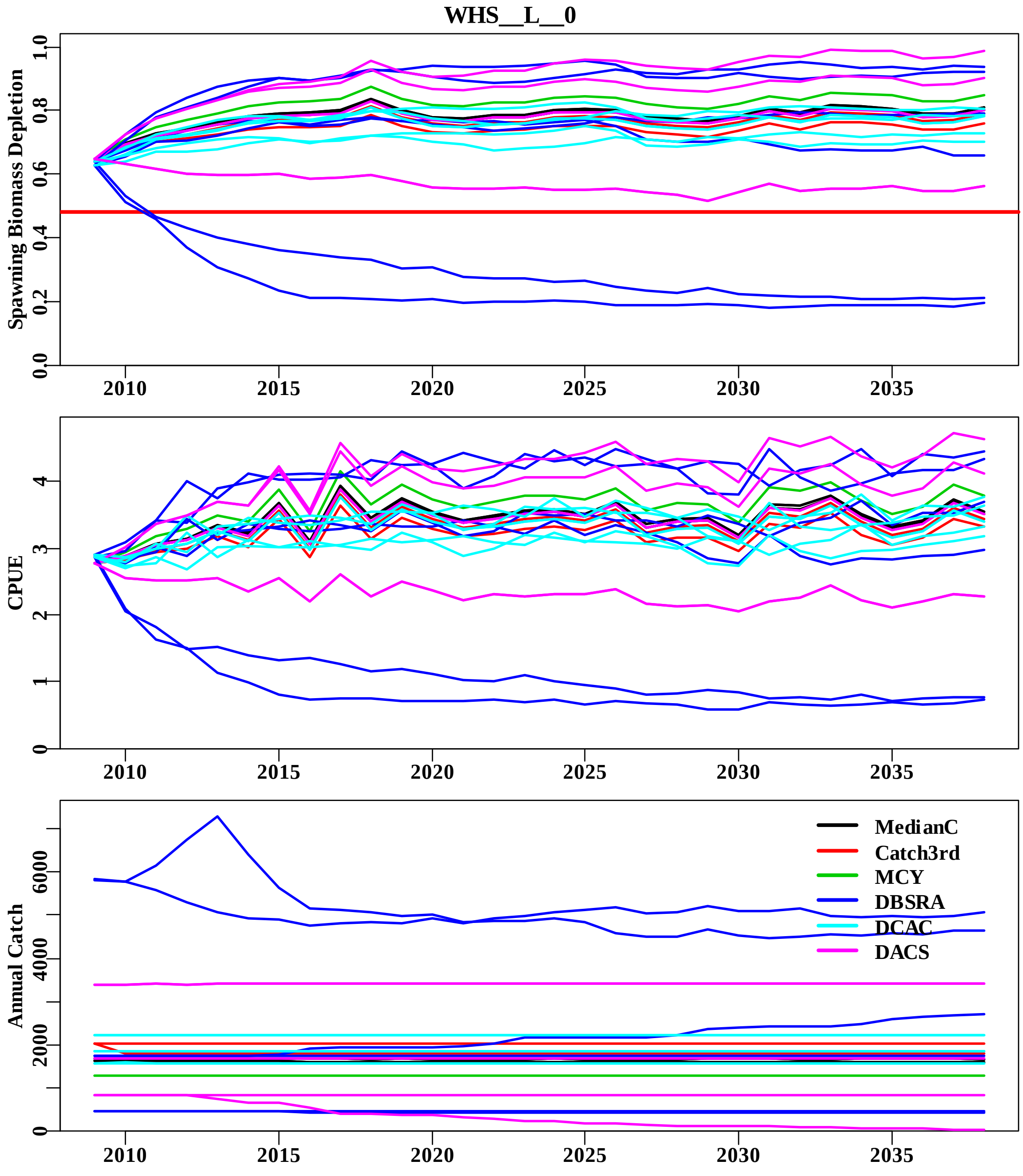

Figure 40. The median trajectories across the 100 replicates for each of the 25 scenarios described in **Figure 39**. The red line in the top figure is the target spawning biomass depletion level.

In the median trajectories from the lightly depleted School Whiting with no discount only the two DB-SRA that were assuming the stock to be above the target very quickly became depleted with the fixed strategy taking longer to deplete than the dynamic (**Figure 40**).

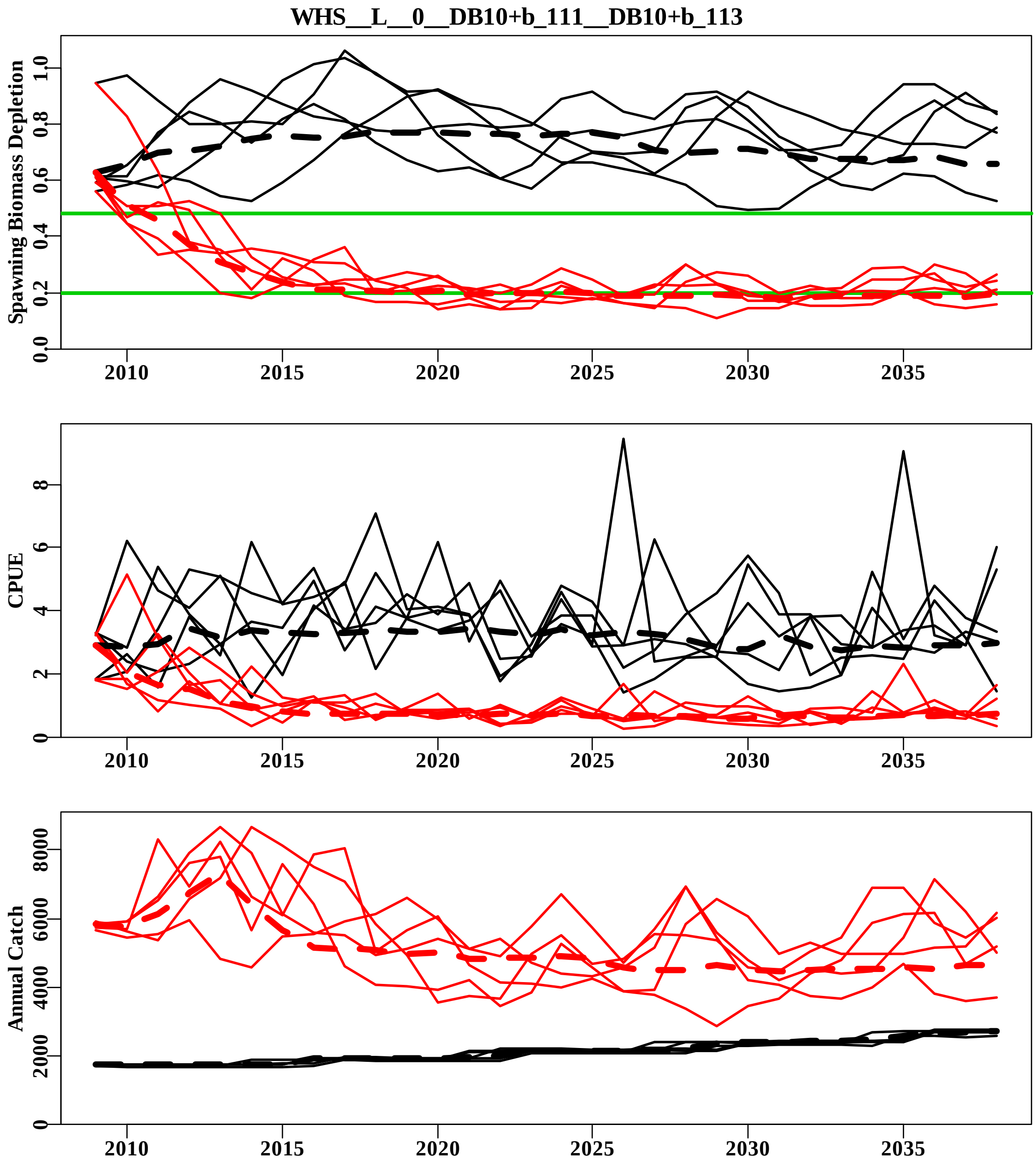

Figure 41. Five randomly selected trajectories from each of two scenarios selected to illustrate contrasting behaviour among the scenarios. The figure heading identifies the species, the actual starting stock status, the discount on catches, the Tier 5 approach (black lines relate to the first, red lines to the second Tier 5), the number of years used, and whether the scenario used a fixed TAC or a dynamic one ('b'), finally the scenario number as depicted on the boxplots and in **Table 9**. The dashed lines are the medians of the 100 replicates for each of the selected scenarios. The black lines represent the assumed stock status was 78%B_0 and the red 48%B_0.

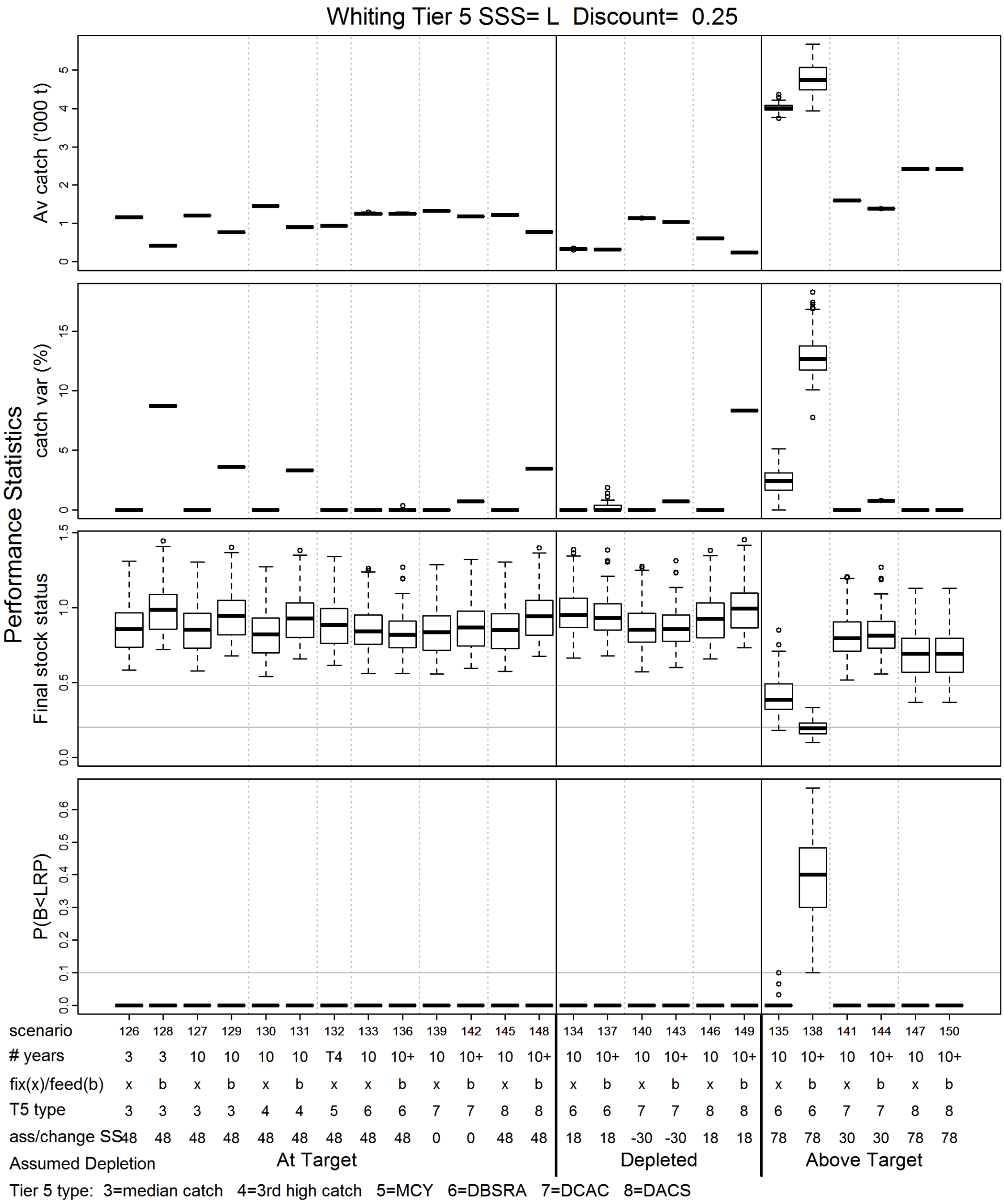

Figure 42. 'ass/change SS' means the current stock status assumed by the method, and for DCAC, 'change SS' means the assumed change in stock status over the catch history. '# years' is the number of years of catches used in the method: T4 means use Tier 4 reference period, 10+ means use last 10 historic years to current. Fix/feed implies fixed TAC setting vs dynamic (or feedback).

Comparing School Whiting results between identical scenarios except for including a 25% discount on catches leads to the catches taken being less which leads to an even lighter level of final stock depletion. More significantly, the DB-SRA with a fixed strategy, which assumed the stock status was really 78%B_0 also had reduced catches which were sufficient to end with a depletion level between the target and limit, and only had one replicates that went above the LRP. The outcome for the dynamic DB-SRA in the same section was not improved (**Figure 42**).

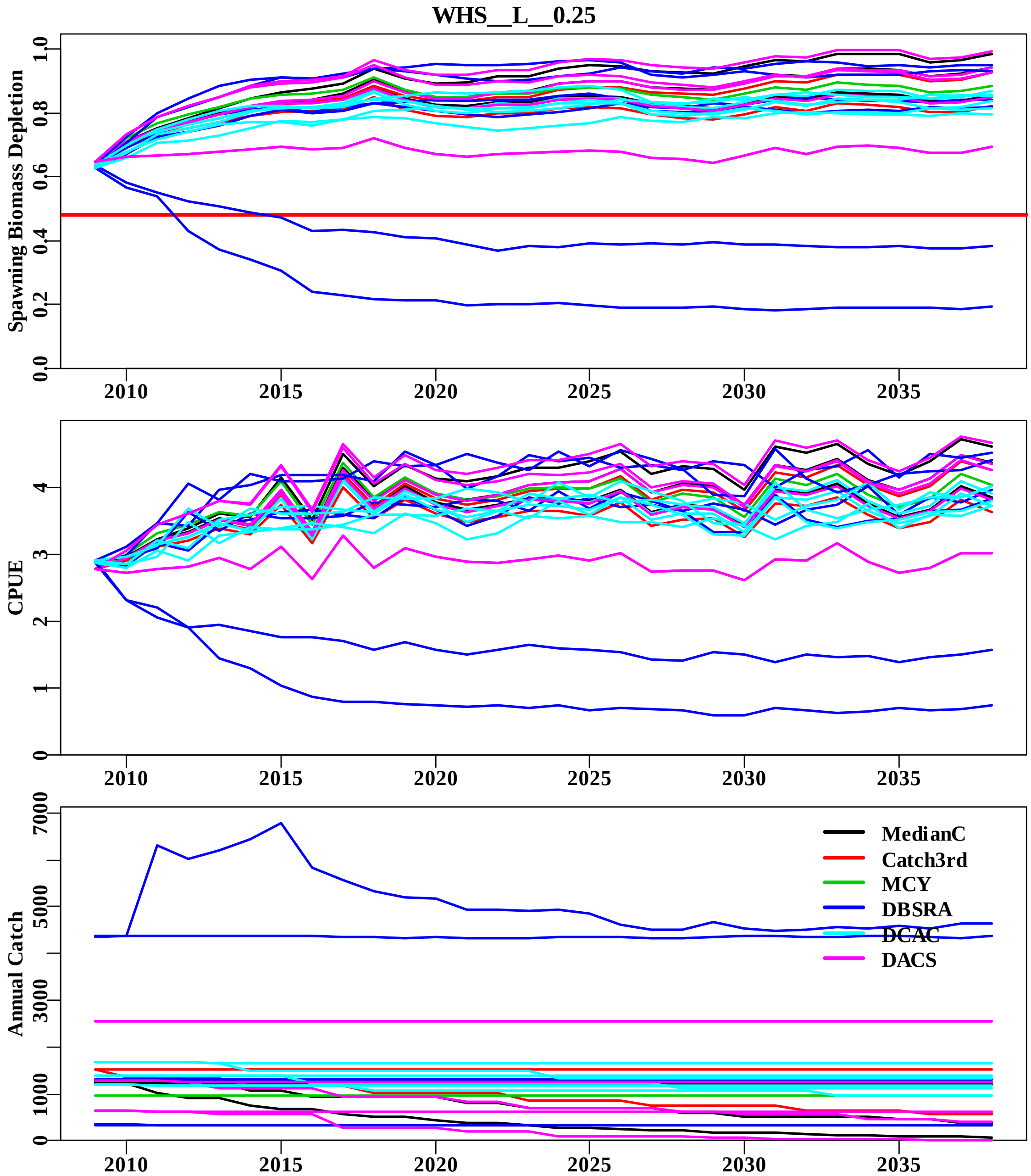

Figure 43. The median trajectories across the 100 replicates for each of the 25 scenarios described in **Figure 27**. The red line in the top figure is the target spawning biomass depletion level.

Once again the median trajectories reflect the boxplots (**Figure 42**) in that only the two DB-SRA scenarios from the assume lightly depleted section deplete to any extent. The constant catch scenario performs better than the updating or dynamic catch DB-SRA (**Figure 43**).

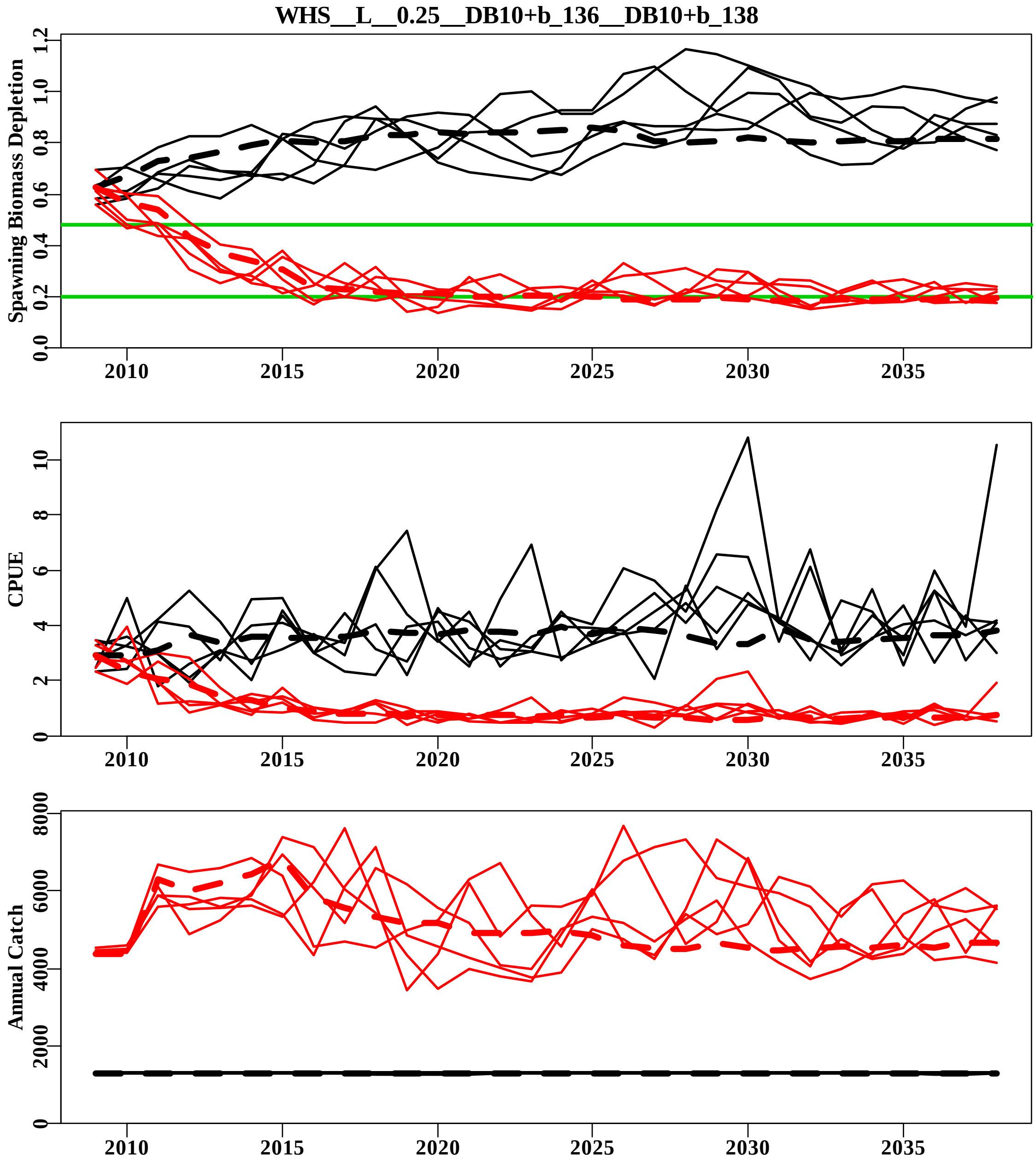

Figure 44. Five randomly selected trajectories from each of two scenarios selected to illustrate contrasting behaviour among the scenarios. The figure heading identifies the species, the actual starting stock status, the discount on catches, the Tier 5 approach (black lines relate to the first, red lines to the second Tier 5), the number of years used, and whether the scenario used a fixed TAC or a dynamic one ('b'), finally the scenario number as depicted on the boxplots and in **Table 9**. The dashed lines are the medians of the 100 replicates for each of the selected scenarios. The red lines represent the assumed stock status was 78%B_0 and the black 48%B_0.

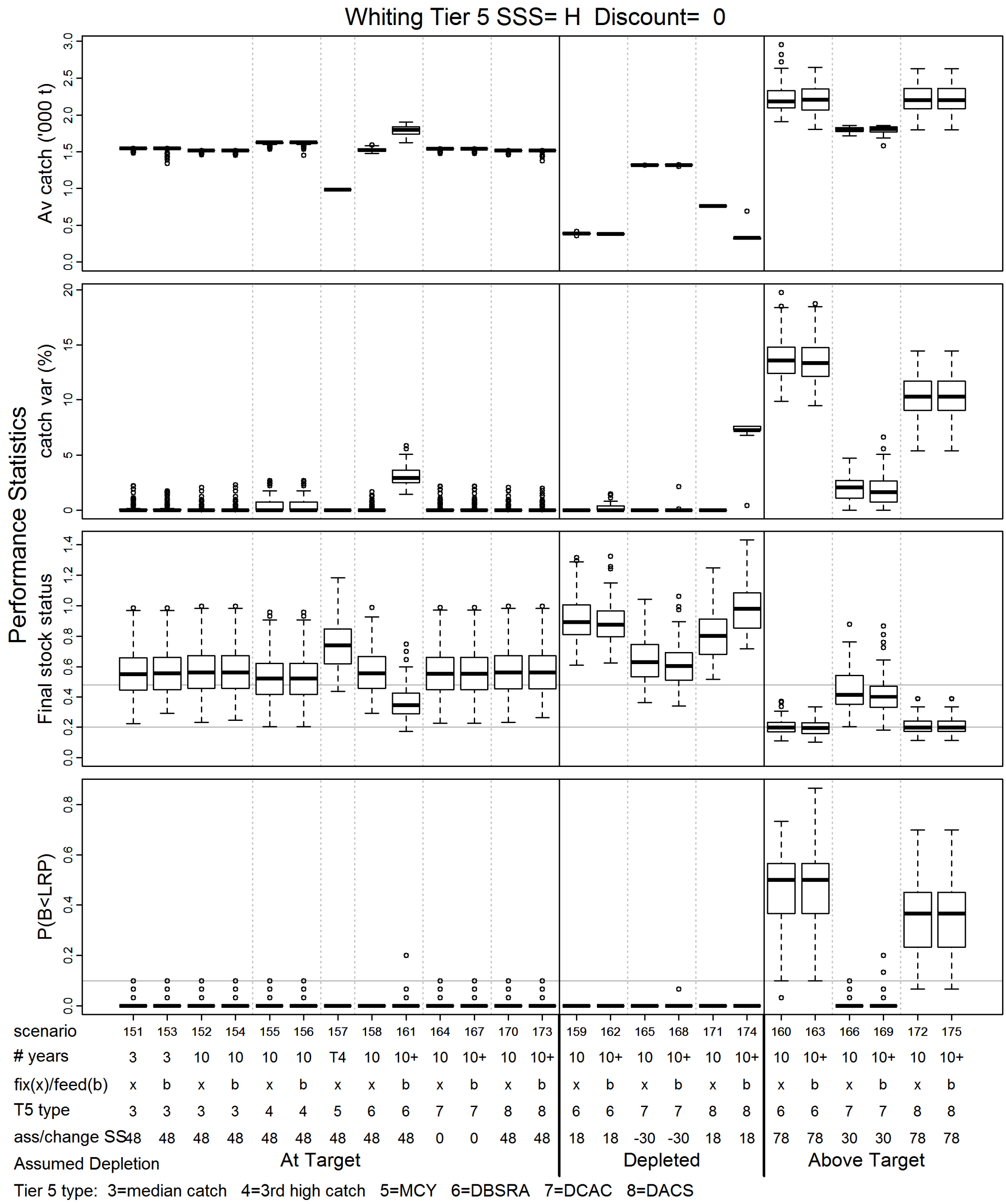

Figure 45. 'ass/change SS' means the current stock status assumed by the method, and for DCAC, 'change SS' means the assumed change in stock status over the catch history. '# years' is the number of years of catches used in the method: T4 means use Tier 4 reference period, 10+ means use last 10 historic years to current. Fix/feed implies fixed TAC setting vs dynamic (or feedback).

Where the School Whiting stock was actually in a highly depleted state (**Table 5**) and zero discount, again the catches for many of the tested scenarios were under-estimated which led to low catch variation and final depletion levels close to or above the target depletion level and only low likelihoods of failing to stay above the LRP (**Figure 45**). For those scenarios that assumed the stock was highly depleted then all methods DB-SRA, DCAC, and DACS badly under-estimated catches and led to relatively high stock levels. In those scenarios which assumed the stock was highly depleted the DB-SRA and DACS both over-estimated sustainable catches which led to very high catch variability, final median depletions sitting on the LRP, and a high probability of failing to stay above the LRP (**Figure 45**); the DCAC, however, performed relatively well, mostly staying above the LRP and finishing just below the target RP. These outcomes are reflected in the plot of the median trajectories from the different scenarios (**Figure 46**).

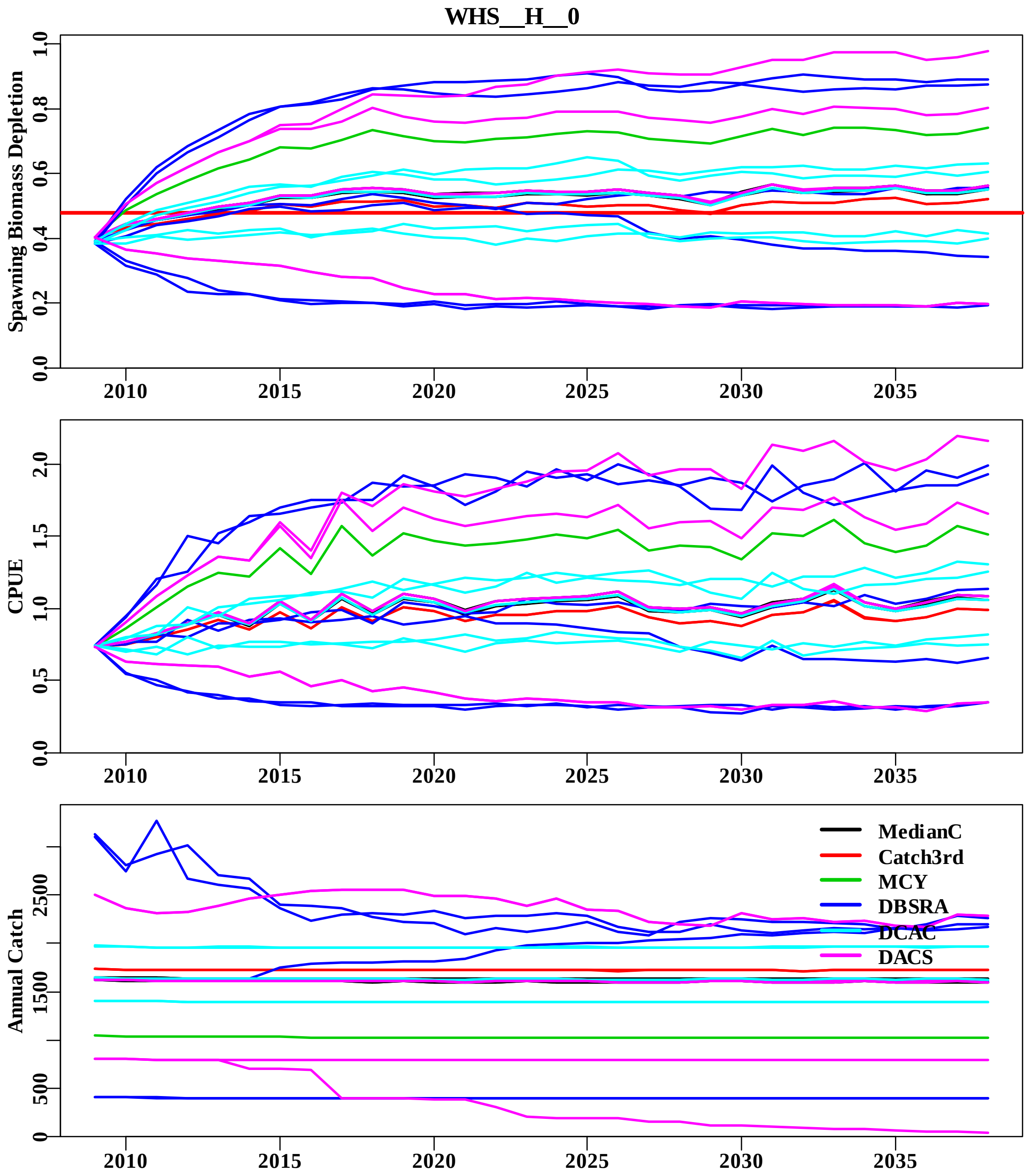

Figure 46. The median trajectories across the 100 replicates for each of the 25 scenarios described in **Figure 45**. The red line in the top figure is the target spawning biomass depletion level.

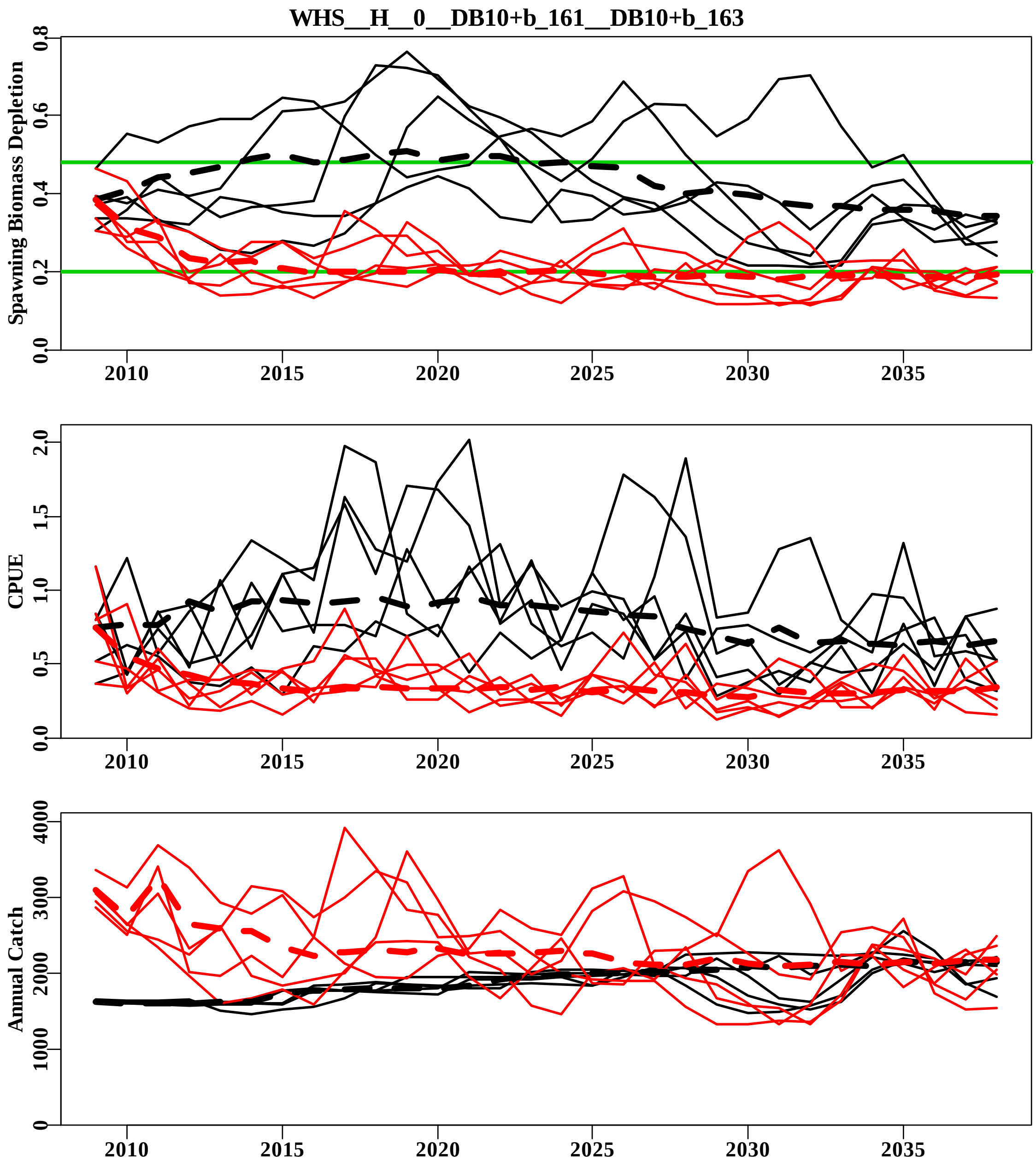

Figure 47. Five randomly selected trajectories from each of two scenarios selected to illustrate contrasting behaviour among the scenarios. The figure heading identifies the species, the actual starting stock status, the discount on catches, the Tier 5 approach (black lines relate to the first, red lines to the second Tier 5), the number of years used, and whether the scenario used a fixed TAC or a dynamic one ('b'), finally the scenario number as depicted on the boxplots and in **Table 9**. The dashed lines are the medians of the 100 replicates for each of the selected scenarios.

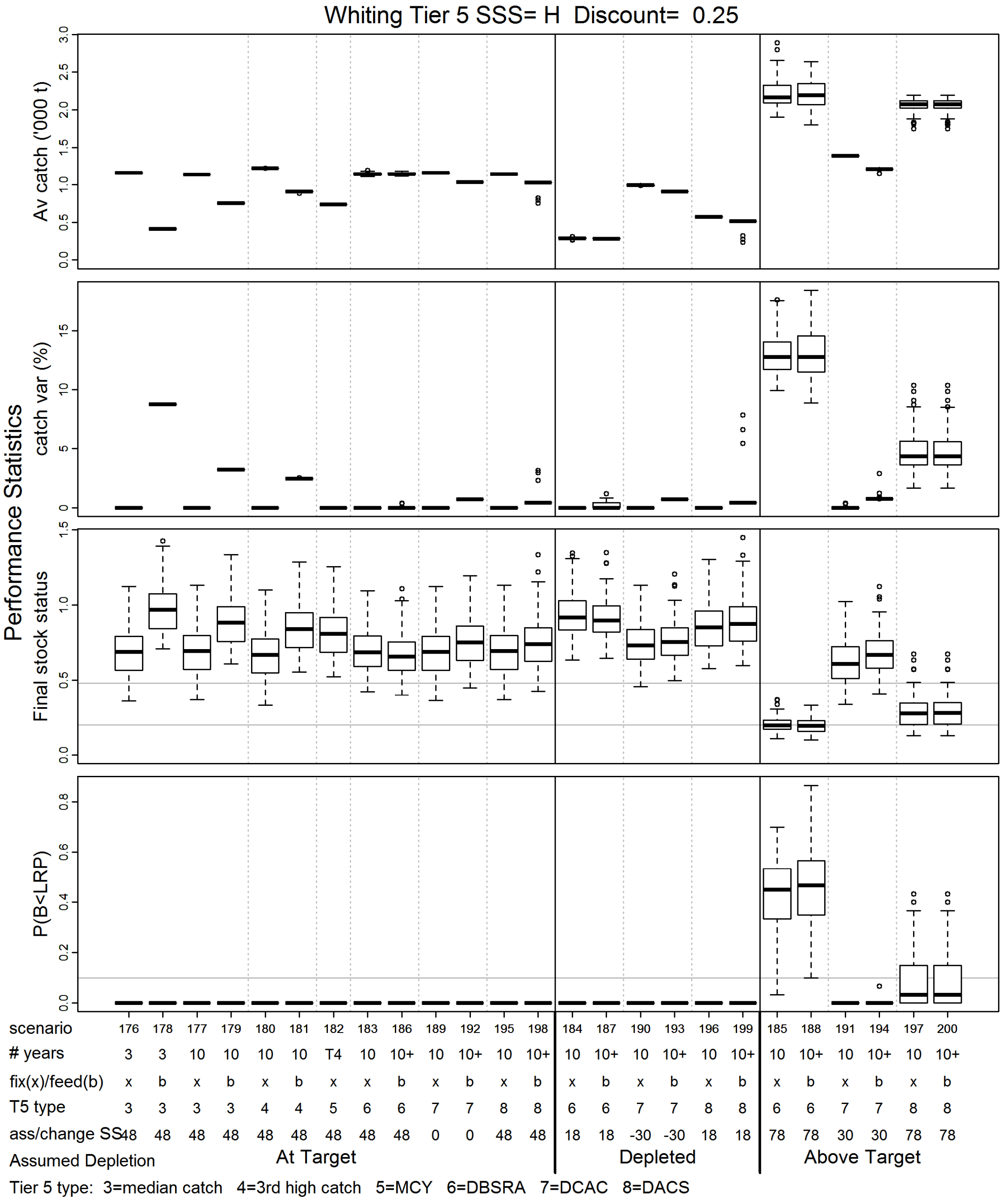

Figure 48. 'ass/change SS' means the current stock status assumed by the method, and for DCAC, 'change SS' means the assumed change in stock status over the catch history. '# years' is the number of years of catches used in the method: T4 means use Tier 4 reference period, 10+ means use last 10 historic years to current. Fix/feed implies fixed TAC setting vs dynamic (or feedback).

Where the School Whiting stock was actually in a highly depleted state (**Table 5**) and a discount of 25%, the catches were naturally lower and this led to relatively high catch variation in three of the scenarios that were assuming the stock to be at 48%B_0; it also led to increased variation in the final depletion levels although all which assumed the stock to be below or at the target ended well above the target. (**Figure 48**). Those scenarios where the assumption was the stock was up at 78%B_0 behaved differently. The

DCAC again performed better than the DB-SRA and DACS although it finished a long way above the TRP. The DACS finished with the lower quartile biomass depletion level on the LRP so there were a relatively high proportion of runs that failed to stay above the LRP. The performance of the DB-SRA was not effectively improved over a zero discount. Again these boxplot findings were reflected in the median trajectories (**Figure 46**) and the individual trajectories (**Figure 49**), although like the zero discount scenarios the random individual trajectories (**Figure 50**) illustrate the greater variability of School Whiting (than Flathead) by the wide range of variation apparent in the dynamics altered scenario.

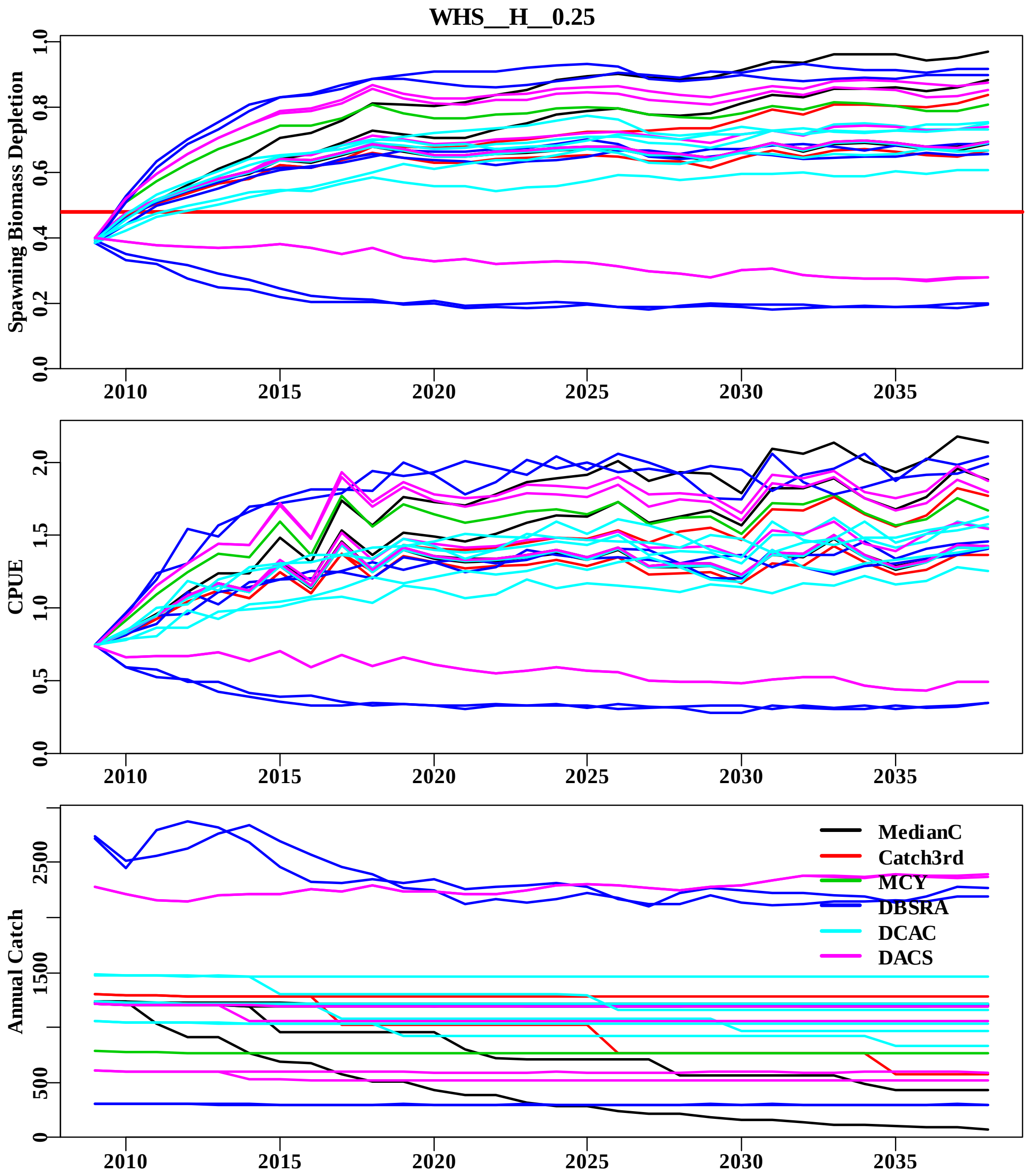

Figure 49. The median trajectories across the 100 replicates for each of the 25 scenarios described in **Figure 27**. The red line in the top figure is the target spawning biomass depletion level.

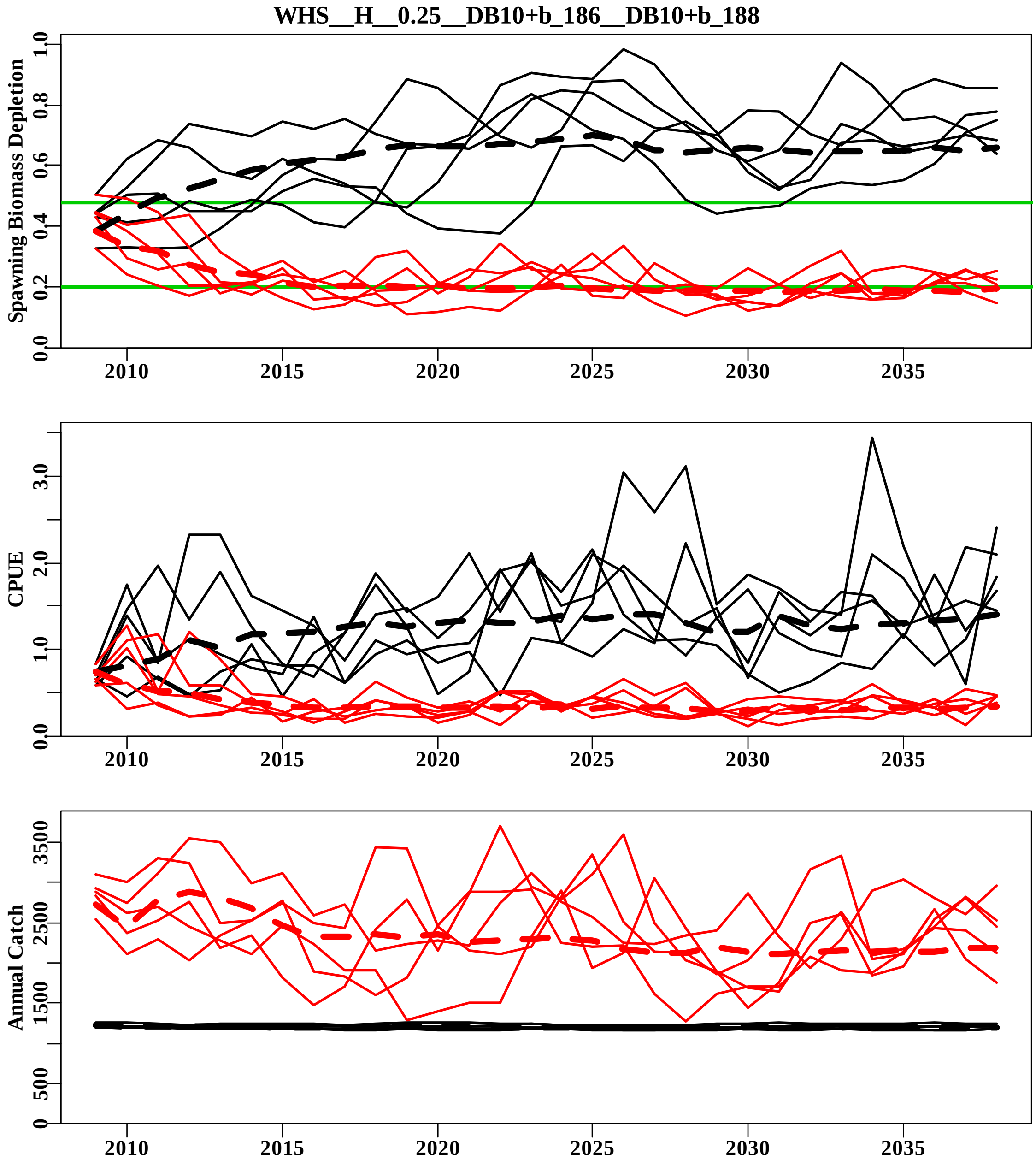

Figure 50. Five randomly selected trajectories from each of two scenarios selected to illustrate contrasting behaviour among the scenarios. The figure heading identifies the species, the actual starting stock status, the discount on catches, the Tier 5 approach (black lines relate to the first, red lines to the second Tier 5), the number of years used, and whether the scenario used a fixed TAC or a dynamic one ('b'), finally the scenario number as depicted on the boxplots and in **Table 9**. The dashed lines are the medians of the 100 replicates for each of the selected scenarios.

13.4 Summarizing Across Species and Scenarios

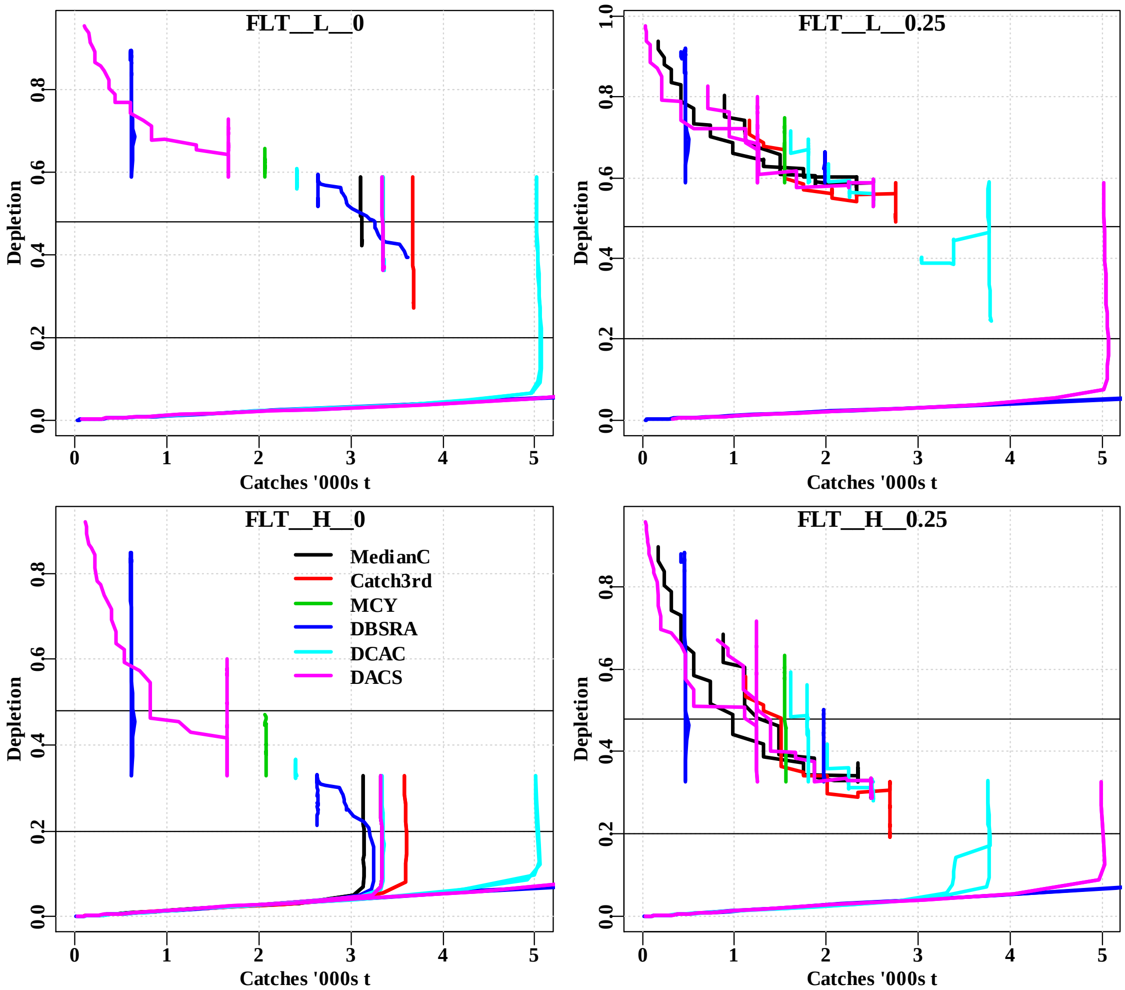

Figure 51. A phase plot of Flathead depletion level against the respective catches for all scenarios within each of the four combinations of initial stock status and discount level. Vertical coloured lines represent constant catches. The legend applies to all panels. The thin black line represent the policies limit and target reference points.

For Flathead in the lightly depleted scenarios catches greater than 4000t lead to depletion below the LRP and in the heavily depleted scenarios catches above 3000t have the same effect (**Figure 51**). Clearly, when initial catches are too high this leads to stock collapse without further intervention. Mixed up within each panel is whether the stock is assumed to be at, below, or above the target depletion level when applying the methods in the first year. The 30 years of projection are indicated by the length of each coloured line. The shorter the line the less catch variation in a strategy. If the lines are vertical this indicates stable catches (**Figure 51**). The differences between the dynamic or fixed versions of a harvest strategy are apparent. The scenarios where the initial stock status was lightly depleted indicate more scenarios that start with relatively high catches which finally decline once the stock biomass declines beyond a critical level. It is not impossible for the median catches to increase beyond the initial catch levels (e.g. **Figure 34**) even when stock levels begin to collapse; this occurs especially with DB-SRA and DACS, but only when the assumption is made that the stock is only lightly depleted up at 78%B_0. This suggests a possible weakness in the proposed methods where it responds inappropriately to the consequent catches. Nevertheless, stable catches appear more successful more often than those data-poor methods that rely on updating through time. All those scenarios where the assumed stock status is lightly depleted predict high catches and most often these fail to keep above the LRP and lead to high to very high depletion levels. Irrespective of the method or discount used the depletion level of a stock should never be assumed to be better than at the target or below.

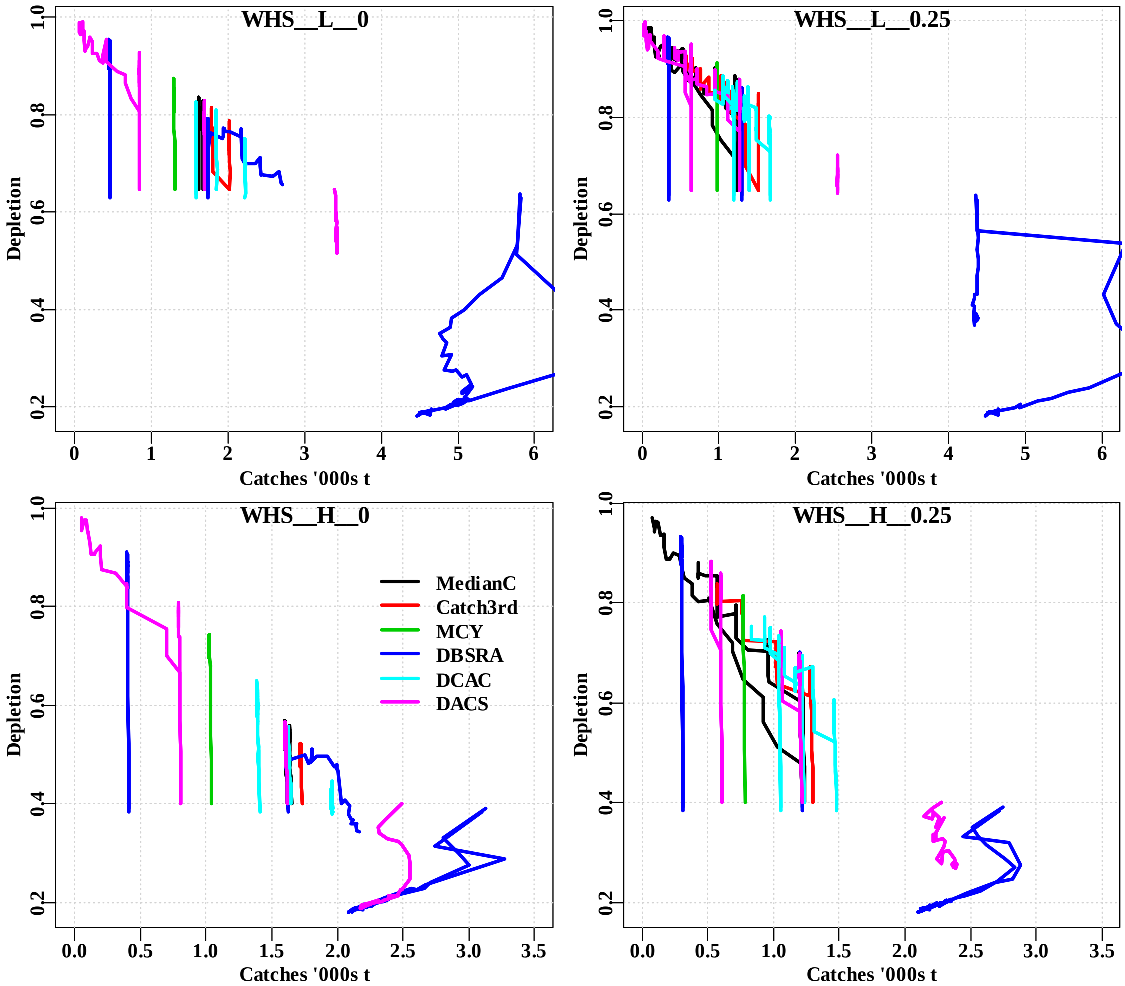

Figure 52. A phase plot of School Whiting depletion level against the respective catches for all scenarios within each of the four combinations of initial stock status and discount level. Vertical coloured lines represent constant catches. The legend applies to all panels. The thin black line represent the policies limit and target reference points.

Summarizing across all scenarios (**Figure 52**) the relationship between the catch taken and the related depletion level is very obvious. In School Whiting, the two DB-SRA scenarios where the stock is assumed to be at 78%B_0 at the start both exhibit extreme dynamics and strongly suggest, when combined with the findings in Flathead (**Figure 51**) that if DB-SRA or DACS are to be used the assumption about the current stock depletion level should never be optimistic and should only select being at the target of 48%B_0 with less risk attributed to the stock being below the target. Again combined with the findings from Flathead this finding can be generalized to conclude that the current stock status should never be assumed to be better than at the target. For School Whiting assuming the stock is at the TRP appears to be a reasonable option for all scenarios except when using DB-SRA with updating of the catch levels; and the outcomes suggest there are no improvements to catch, catch variability, or final depletion level from using a discount.

The results from any MSE tend to be voluminous and difficult to interpret. Attempting to consider **Figure 27** to **Figure 52** at once is not simple. It is however, possible to select out those HS and methods that lead to outcomes which are either intermediate between the LRP and TRP or above the TRP (**Table 11**), omitting those combinations that fail to maintain the stock above the LRP for the requisite 90% of the time.

Table 11. Summary of the outcomes from different methods used under different circumstances. Depletion is the actual stock depletion, FinalD is the final depletion level (median value T – at TargetRP, A – above, I - in-between LRP and TRP, and L – at LRP). AssumedD is the stock depletion level assumed for the scenario, and Fix or Dyn is whether the catch estimated was fixed at the start or was dynamically updated.

Species	Depletion	Discount	Method	FinalD	AssumedD	Fix or Dyn
Flathead	H	0	MCY	T	48	f
Flathead	H	0	DCAC	I	18	f & d
Flathead	H	0.25	DB-SRA	T	48	f & d
Flathead	H	0.25	3rd HighC	A	48	f
Flathead	H	0.25	DCAC	A	48	f
Flathead	H	0.25	DACS	A	48	f
Flathead	L	0	3yr MedC	A	48	f & d
Flathead	L	0	DB-SRA	A	48	f
Flathead	L	0.25	3rd HighC	A	48	f
Flathead	L	0.25	10yr MedC	A	48	f
Flathead	L	0.25	DCAC	A	48	f
Flathead	L	0.25	DACS	A	48	f
School Whiting	H	0	3rd HighC	A	48	f & d
School Whiting	H	0	MCY	A	48	f
School Whiting	H	0	DB-SRA	A	48	f
School Whiting	H	0	DB-SRA	I	48	d
School Whiting	H	0	DCAC	I	78	f & d
School Whiting	H	0.25	All	A	18 & 48	f & d
School Whiting	H	0.25	DCAC	A	78	f & d
School Whiting	H	0.25	DACS	I	78	f & d
School Whiting	L	0	DACS	A	48	f & d
School Whiting	L	0	DB-SRA	A	48	d
School Whiting	L	0	All others	A	18 & 48	f & d
School Whiting	L	0	All others	A	78	f & d
School Whiting	L	0.25	All	A	18 & 48	f & d
School Whiting	L	0.25	DB-SRA	I	78	f
School Whiting	L	0.25	All others	A	78	f & d

As long as the initial stock state is never assumed to be above the target RP, the DB-SRA and DCAC methods appear to be capable of producing catch estimates that maintain the intent of the Commonwealth HSP, even when only based on 10 years of data. At the same time, it often appears best to use a fixed estimate of sustainable catch although updating the estimates as more data becomes available would also appear to have benefits for some methods under some circumstances.

As with all stock assessments and fisheries management blindly following the dictates of any formal harvest strategy in the face of a weight of evidence that indicates a problem with whatever advice is being produced would be a risky strategy. Formal harvest strategies should always have escape clauses that allows for exceptional circumstances. This is especially the case when dealing with data-poor fisheries where any assumptions behind the methods used within the harvest strategy may only be weakly adhered to. This is not to say that rejecting the management advice from a harvest strategy should be simple or easy; such rejections should always be evidence based.

13.5 Application of Catch-Only Methods to Current Fisheries

13.5.1 Average and Median Catch Methods

Except when the state of the assessed stock was assumed to be only lightly depleted, the range of methods used all had some success in the MSE testing in that they managed to avoid the LRP in many cases. They often did this, however, by predicted relatively conservative catches. This may have been related to only using 10 or 3 years in their estimation and they may perform better with longer series. It does matter whether the time series is relatively stable or highly variable (Ministry of Primary Industries, 2014). If variable catches are usual then the longer the time series the better, but the use of the central tendency (mean or median) of a catch history as an estimate of sustainable catch would depend strongly on whether the fishery were developing or declining when the catch history was recorded. If on the way up the estimate may be an under-estimate but if on the way down it may over-estimate catches. The methods that use fixed estimates avoided the potential for a ratcheting down of catches that can occur in the strategies that include regular updating of the central tendency estimate. It is not the case that an allocated TAC will always be fully taken, especially with a by-product species that is not specifically targeted. If sustainable catch estimates are updated by using the mean or median of a time series that has an upper limit (a TAC) which is often not met, then the upper limit will automatically decline. Such catch estimates should be reviewed at five or ten year intervals in a weight-of-evidence context, especially of more information beyond catches has been collected, but otherwise the fixed methods have advantages over the dynamic or updated methods.

Methods that use a production model to mimic the stock dynamics have the advantage of being less vulnerable to such changes but also have the disadvantage that they require either estimates (guesses; expert opinion) of final depletion level or of the changes in depletion during the catches. Once again a weight or evidence approach through the whole RAG would be needed, and the plausibility of whatever level is selected would need to be defended.

13.5.2 DCAC

The application of the DCAC method is very rapid, even when using 10,000 replicates. For each species the distribution of sustainable catches was skewed to the left, which reflected the uncertainty that derives from the various assumptions made about the biology and the production model representing the stock dynamics. Nevertheless the estimates of sustainable catches were 2158t, 938t, and 1428t for Flathead, Jackass Morwong, and School Whiting respectively (**Figure 53**).

The catch histories for each species were relatively long with catches for Flathead and Jackass Morwong stretching back to 1915 and School Whiting back to 1980. In both Flathead and Jackass Morwong there has been high levels of contrast in catch levels through time but the two species are very different in that Flathead has had variable but large catches throughout its history whereas Jackass Morwong had relatively low catches up until 1949 then high catches to about 1990 and strongly reducing catches out to 2010. When one sequentially reduces the length of the time series of catches used by removing years from the start of each time series the effect on Flathead is relatively minor except for the spread of the outcome around the median values expanding. For Jackass Morwong however, the removal of the early catches soon leads to a maximum estimate of sustainable catch which then declines as the later much smaller catches begin to dominate the catches (**Figure 54**). This reflects the assumption that the change in depletion level over the fishing period remained at about 50%, if this were adjusted to something more appropriate (though difficult to estimate and the weakest part of this approach) hen the sustainable catch estimate would be expected to remain more stable.

Jackass Morwong was omitted from the MSE testing because an earlier stock assessment (Wayte, 2013) demonstrated that the average recruitment had declined two or three decades ago. Importantly the predicted sustainable catches from recent years (**Figure 54**) are greatly reduced which reflects this. It might be thought that the advent of the HSP, which led to catches being control more stringently so that over-fishing was constrained (Smith *et al*., 2014), might invalidate estimates based on recent catches. However, the HSP was only introduced in 2007 so its influence would not be very great. Of greater concern is that any method based on average catches is likely to generate estimates of sustainable catches that are biased high. As with all data-poor fisheries, in addition to application of such harvest strategies the weight of other evidence available should be consistent with the recommendations from the assessment method selected. The DCAC method certainly works well for some species and works well with longer time series of catches. With the three species considered the correction over the average catch for depletion was relatively minor being 137 t for Flathead, 110 t for Jackass Morwong, and 145 t for School Whiting from 1980

Table 12. The quantiles of the sustainable catch estimates from the DCAC with the average catch from the history of landings for each species. School Whiting was estimated using the full time series of catches from 1947 and then only from 1980 onwards; catches reported between 1947 – 1980 were only minor.

Quantile	Flathead	Jackass Morwong	School Whiting 47	School Whiting 80
0.025	2014.525	834.487	753.671	1268.282
0.05	2045.143	854.697	764.117	1302.302
0.5	2153.009	934.574	803.560	1439.002
0.95	2217.412	984.793	823.694	1514.271
0.975	2226.353	991.715	826.032	1523.272
AverageC	2290.357	1044.844	841.531	1584.376

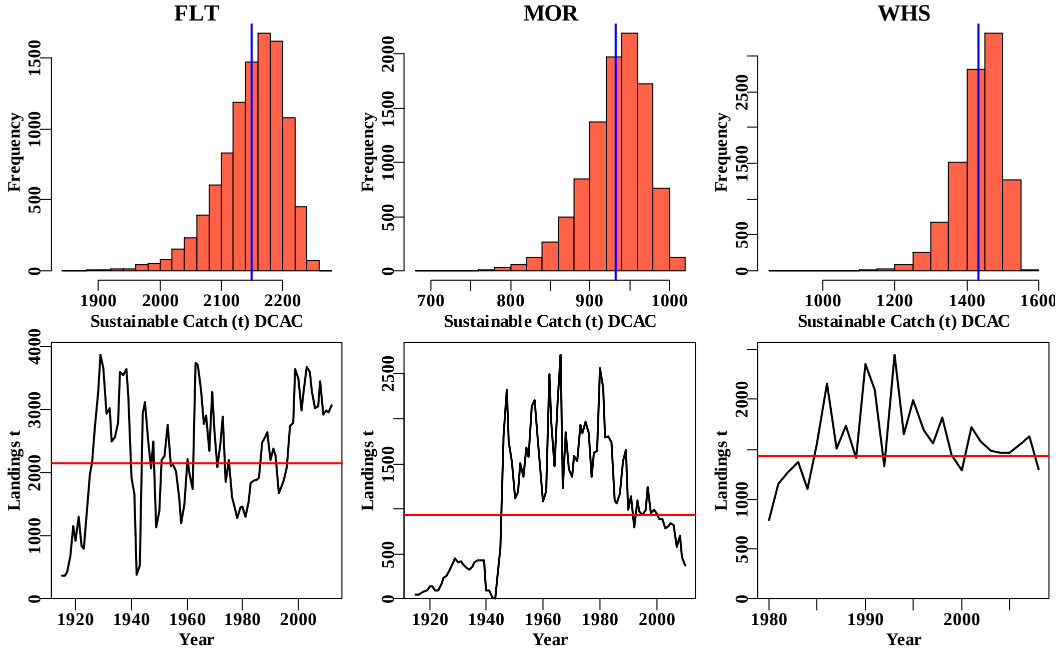

Figure 53. The outcome from 10,000 replicate estimate using the DCAC method on Flathead (FLT), Jackass Morwong (MOR), and School Whiting (WHS). The mean estimates for each species are reflected on the catch histories as red lines (2158t, 938t, and 1428t). The assumption was made that the stock was at the target biomass in all these assessments.

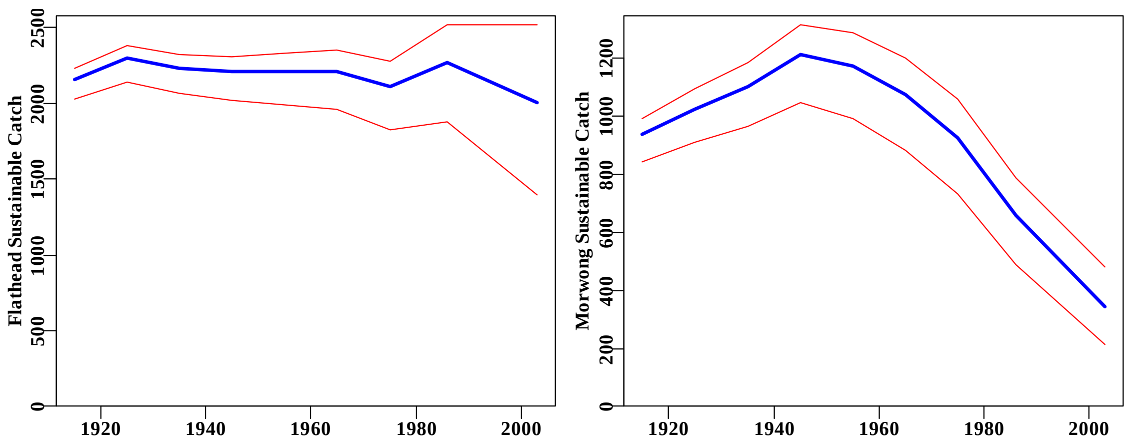

Figure 54. The effect on the estimate of sustainable catch of starting the time series of catches by dropping increments of 10 years from the start of each series for Flathead and Jackass Morwong. The blue lines are the median of the distributions of sustainable catch estimates while the red lines are the 95% confidence intervals.

As noted by MacCall (2009), the DCAC only provides an estimate of a "…moderately high yield that is likely to be sustainable, while having a low probability that the estimated yield level exceeds MSY…". Thus, this method could easily be included into the HSP as a Tier 5 method as long as the predicted catch level was accepted as a suitable proxy for the target fishing level. Being aimed at generally by-product species within the mixed species SESSF fishery, in line with other non-target species, a target of 40%B_0, acting as a proxy for MSY, may be appropriate (although such a decision would ultimately need to be made as a policy decision).

13.5.3 DB-SRA

The Depletion-Based Stock Reduction Analysis is more flexible than the DCAC in that it can provide estimates of MSY, B_{MSY}, F_{MSY} and depletion levels. These are dependent upon the production model used but nevertheless this enables the method to generate estimates that can be directly interpreted by the HSP. The application of the DB-SRA method is rather more time consuming although running 10,000 replicates to characterize a single species takes between 15 – 20 minutes (**Figure 55**).

With each of the species the final depletion level was assumed to be at 48%B_0 so the median in the final year would be expected to be centred close to this value. With Jackass Morwong this assumption of being on target with respect to the final depletion estimate is problematic as this would only be the case if a revised stock recruitment relationship were to be used. A regime shift occurred (Wayte, 2013) to lower the average recruitment and that means that the predicted rise following the early 1980s (**Figure 55**) is now no longer expected. The DB-SRA should have assumed the stock finished in a relatively depleted state and started the catch history following the regime shift. Of course, for a data-poor species such knowledge is unlikely to be available. More work on how to handle such species if they are data-poor is required. None of these catch-only methods should be applied blindly but rather only when taking into account all else known about a fishery. The use of a weight of evidence approach enables the assumptions made to be defensible and possibly testable given sufficient resources.

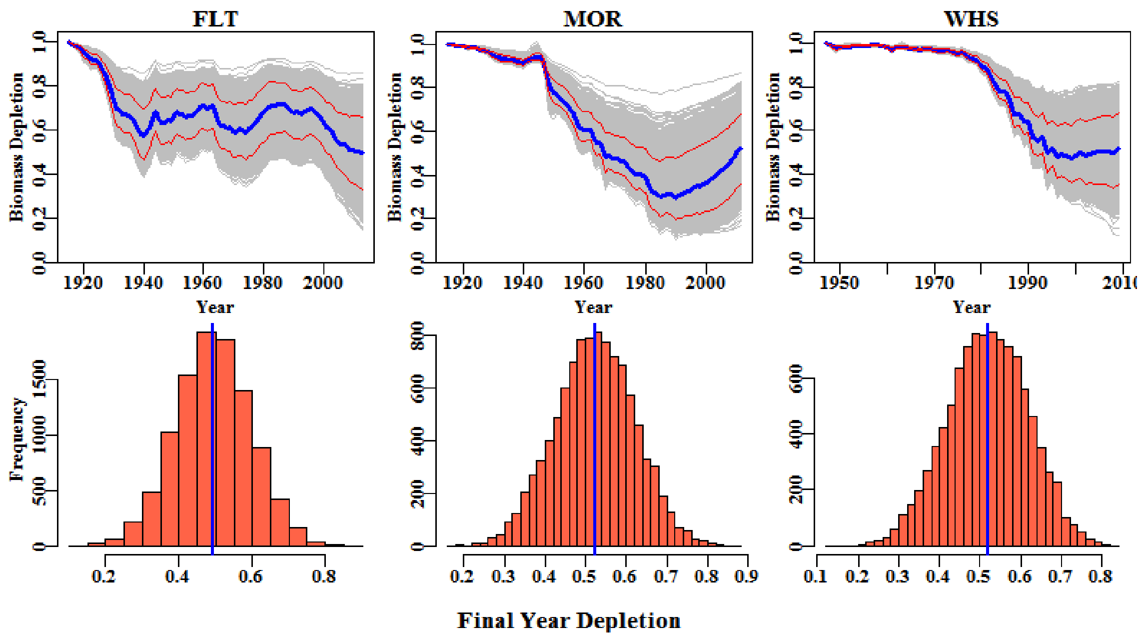

Figure 55. The spawning biomass depletion estimated using the DB-SRA methodology for Flathead, Jackass Morwong, and School Whiting. The blue lines are the median values and red lines are the 95% bounds on the spread of the replicates. The lower graphs represent the frequency distribution in the final year.

Outputs important for management include the MSY, F_{MSY}, and the depletion level, each with estimates of uncertainty included (**Figure 56**). The inclusion of the uncertainty estimates means that any harvest control rule developed for use with the DB-SRA assessment method can attempt to take into account the uncertainty included in the biological parameters and the catch time series used to estimate the management outputs.

As with the DCAC, MSE testing the use of only 10 years for the initial estimates within the DB-SRA provides a strong test of the utility of DB-SRA. In reality, for many of the SESSF species longer time series are available. By repeatedly running the DB-SRA on the Flathead data but sequentially removing 10 years at a time from the long time series (**Figure 57**) the effect of different lengths of time-series could be explored.

The effect of decreasing the length of the time series on the estimate of MSY is only minor up until about 1975 but it is in 1985 and onwards where first the spread of possible values increases so that smaller values make an appearance, but then in 1995 the mean drops from between 2800 and 2750t down to about 2550t with a shift downwards in the distribution (**Figure 58**; **Table 13**).

The analyses in the DB-SRA all assumed that the final depletion level in 2012 was 48% and the outcome from the analysis is sensitive to the selected depletion level (**Figure 59**). For example, when the depletion level in the final year is assumed to be 35%B_0 instead of 50% the MSY output has double the range of values (2168 – 3067 rather than 2344 – 4050) of the analysis at 50%B_0, with more difference between the upper bounds and the lower and a 260 t difference between the MSY estimates. The selection of the assumed depletion level is a decision that would need to be defended explicitly and agreed upon by the full RAG.

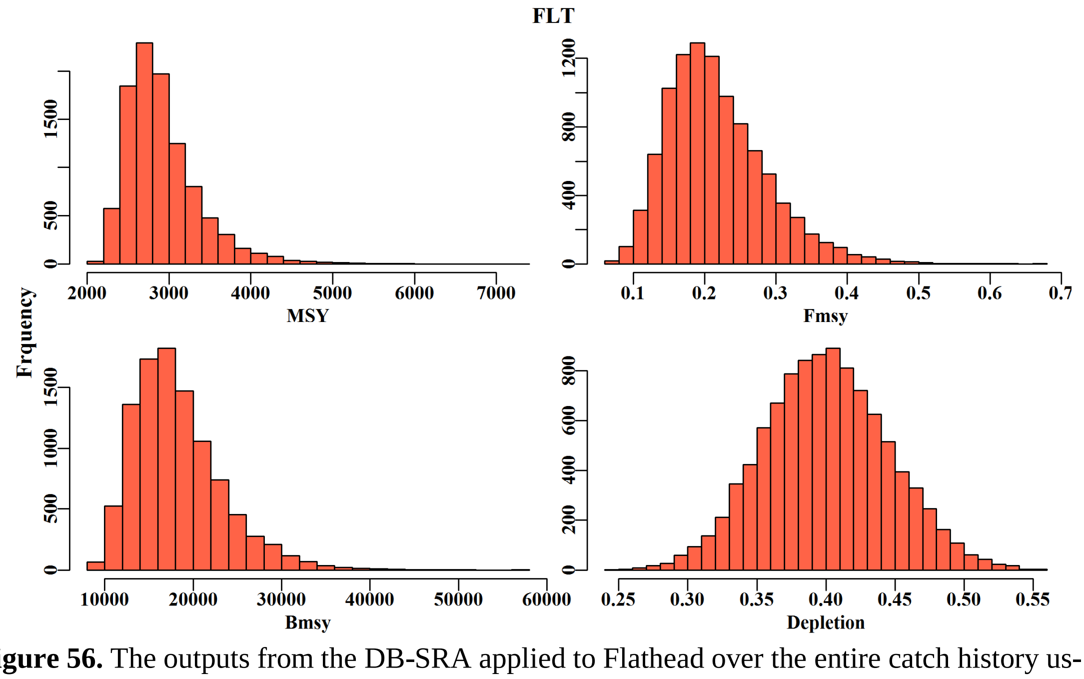

Figure 56. The outputs from the DB-SRA applied to Flathead over the entire catch history using 10,000 replicate runs of the Monte Carlo simulation.

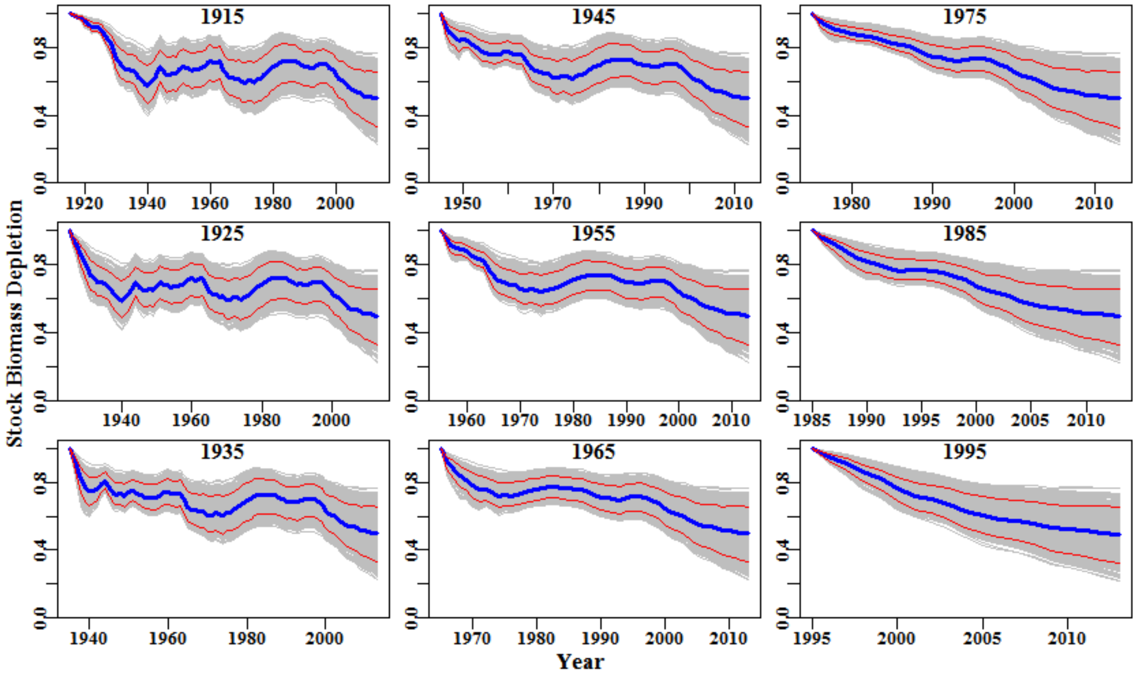

Figure 57. Predicted time series of stock depletion levels for Flathead using shorter and shorter time series of catches from 98 years down to 18 years. The numbers at the head of each graph represent the start year and the series continues up to 2012. Each represents 1000 replicates. The blue lines are the median depletions surrounded by the red 90% intervals. In reality, one would be changing the assumed initial depletion range , which would likely increase the spread of the final values.

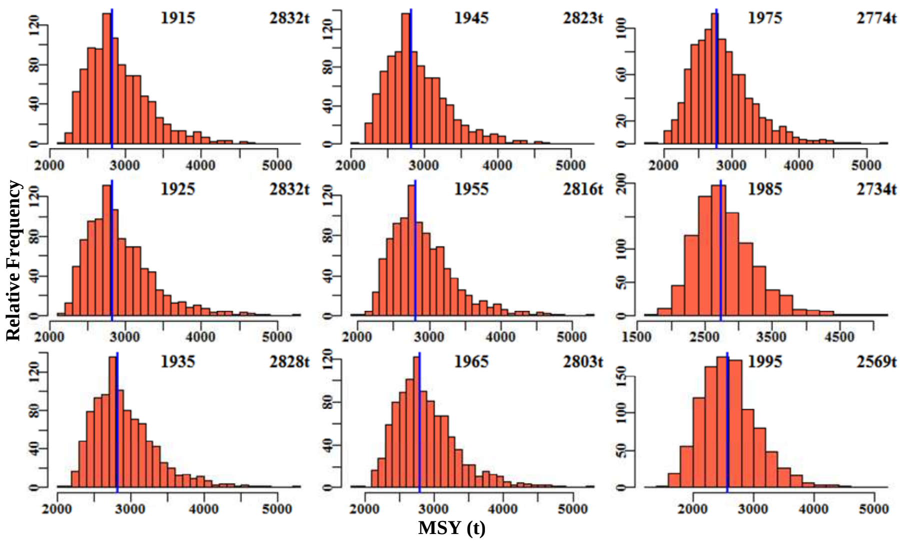

Figure 58. Estimates of MSY from the DB-SRA analysis using 1,000 replicates. The titles in each case represent the starting year of the catch time series, ending in 2012, and the MSY estimate, highlighted by the blue lines.

Table 13. The start year, number of years of catch data, and the quantiles of the estimates of the MSY. Note the shift to lower values in the last three rows and the difference relative to the average catches over the same period. In all cases except the last row the assumed final depletion was 48%; in the last row is the outcome from an assumption of 35%.

Start	Years	2.50%	5%	50%	95%	97.50%	AverageC
1915	98	2344.537	2382.149	2831.794	3776.656	4050.090	2290.357
1925	88	2342.034	2381.012	2831.601	3776.654	4048.268	2457.239
1935	78	2327.350	2373.559	2827.719	3776.633	4038.496	2404.397
1945	68	2306.799	2364.283	2823.057	3776.583	4031.961	2401.544
1955	58	2267.541	2350.947	2816.476	3775.969	3990.172	2434.862
1965	48	2234.215	2310.502	2802.807	3763.108	3983.085	2486.521
1975	38	2168.726	2255.357	2773.696	3725.866	3943.827	2440.500
1985	28	2067.754	2180.506	2733.887	3668.618	3917.452	2739.179
1995	18	1828.630	1934.611	2569.098	3442.084	3709.652	3051.000
1915	98	2168.842	2200.732	2426.595	2907.005	3067.365	2290.357

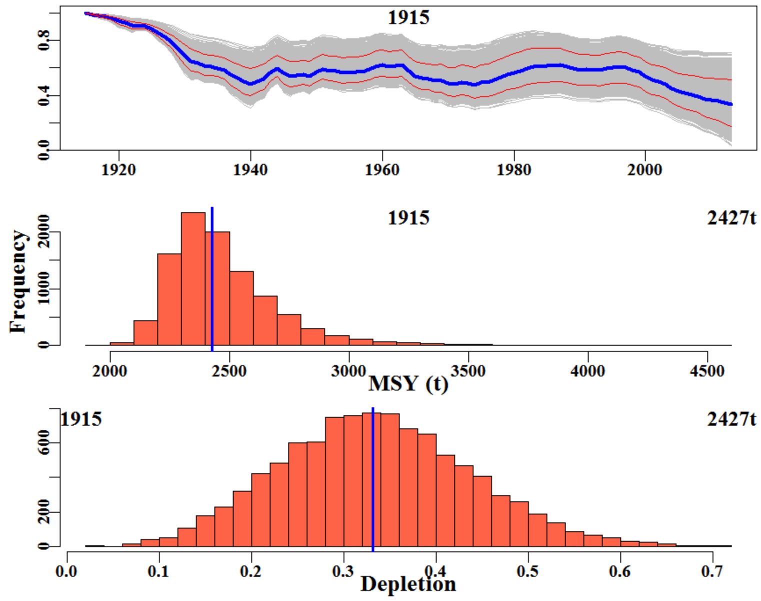

Figure 59. An example run of 10000 replicate runs the DB-SRA on Flathead from 1915 – 2012 with the final depletion set at 35%B_0 instead of 50%B_0. The median MSY is 2426 t, and median depletion is 0.331. The 10th percentile is at 0.2034, so the method meets the Limit Reference Point requirement.

Both the DCAC and the DB-SRA require an assumption about the level of depletion in the final year (or a particular year) but with data-poor species this will invariably be a very uncertain value. Despite this the DB-SRA methods holds some advantages over using simple average catches as even with only a few year's data and relatively high natural mortality rates the outcome differs from simply using the average of median catch (**Table 13**).

13.5.4 Further Alternative Tier 5 Methods

Alternative methods that are very new and were not included in the MSE include the Catch-MSY method by Martell and Froese (2014). This is very similar to the DB-SRA but rather than using the novel production model developed by Dick and MacCall (2011) Martell and Froese, use a simpler production model and conduct a Monte Carlo simulation on that production model's parameters. This leads to the production of a large number of potential stock trajectories and they include procedures for eliminating the implausible ones (stock goes extinct, biomass extends beyond the carrying capacity) and then calculating statistics of management interest from those remaining. They still require preliminary guesses as to the starting and the finishing depletion levels but these can be very wide and flat. They use the term 'prior' distribution and 'posterior' distribution to describe the effect of the analysis although it is not strictly a Bayesian analysis; because they are using a Bernoulli distribution (true or false; 1 or 0) there is not a multiplication or each prior with a likelihood but rather the prior is being thinned by rejection of implausible combinations of two parameters at a time. It is not necessarily the case that the most common combination of parameters that is not inconsistent with the data is the most likely. The production model is not being fitted to the data but rather the inverse of what does not fit it is being eliminated. Nevertheless, this approach holds promise and is already being used in a few places (Haddon, 2014b). The input requirements of the Catch-MSY method are fewer than those for the DCAC or the DB-SRA and so this approach would warrant further consideration.

14 Appendix: Further Discussion and Conclusions

14.1 Data-Poor Harvest Strategies

The current HSP recognizes that in a multi-species fishery it would not be possible to ensure that all species are being fished at their individual maximum economic yield level (the target reference point, TRP, or its proxy of 48%B_0). Here it is assumed that any species being considered for a Tier 5 assessment would not be a principle economic driver of the fishery involved and so it would be acceptable to adopt the minimum objective of maintaining the species above the limit reference point (LRP). With data-poor or data-limited species any of the assessment methods proposed could be justified initially by putting it in a weight-of-evidence appraisal of the state of the species concerned. This would both enable: 1) the data available to be reviewed, 2) to determine which methods and harvest strategies were feasible (**Figure 60**), and 3) determine what information was useful and how important the species actually was in the context of the multi-species fishery.

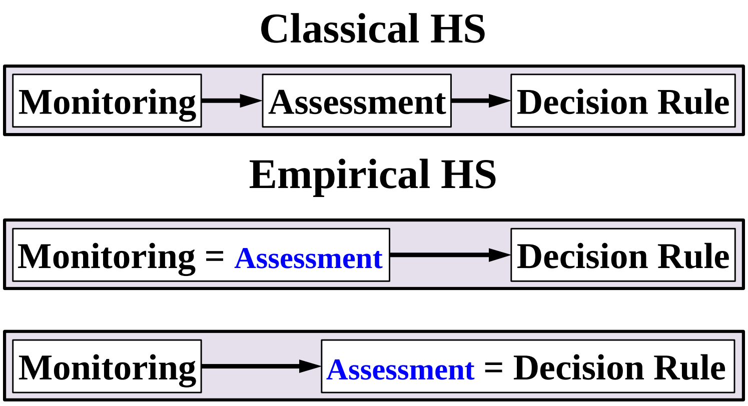

Figure 60. A comparison of the internal structure of a classically structured harvest strategy with those based on simple empirical rules. The classical HS can provide estimates of stock biomass or depletion or their proxies while the empirical HS treat some statistic from the fished stock as a performance measure which they can map directly onto a recommended catch.

The MSE work in this report has been considering different data-poor methods but also a default harvest strategy in each case, which was that the estimated sustainable catch was that which would be applied. This harvest strategy appears simpler than it would be in reality if it were implemented. Any new harvest strategy or method would be implemented in the context of the current management framework and processes. Thus, in the SESSF it is the RAGs for each sub-fishery who would review the available evidence and the fishery appraisal for the data-poor species being considered; they would then accept or reject the draft assessment. Within the SESSF the output from any accepted final version assessment is merely the recommended biological catch (the RBC). The management advisory committee with AFMA staff provide advice to the AFMA board as to whether the RBC needs modification, and it is the Board who provide the final advice on TACs to the Minister responsible for fisheries (**Figure 61**).

Generally, any TAC wold be expected to be smaller than or equal to the RBC, but it remains possible for the TAC to be larger than the RBC.

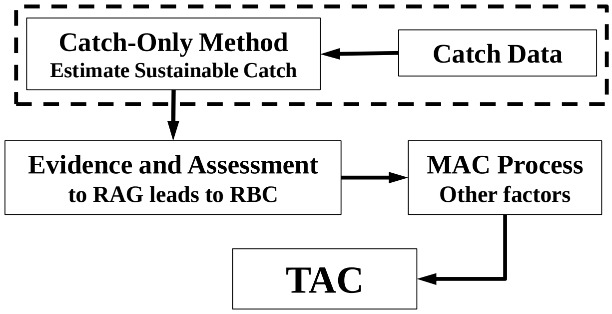

Figure 61. The process flow in the generation of a Total Allowable Catch in the SESSF. The assessment team work up a stock assessment and present that with its associated evidence to a RAG. Once the RAG has approved a draft, base case, assessment, a final assessment is made and this forms the basis of an RBC. This is considered by the respective MAC, who take into account socio-economic and possibly cultural/indigenous concerns to recommend a final TAC. The MSE testing only considered the processes within the dashed line.

14.2 The Appropriateness of a Selected Tier

The need for a higher tier than Tier 4 arose when it became clear that there were fisheries being assessed using Tier 3 and Tier 4 which were poorly suited to either approach. One obvious approach to determine the appropriateness of each assessment method when applied to a particular stock is to ensure that the assumptions behind each assessment method are all met or at least are not deviated from to any large extent. However, in the higher tiers (both 3 and 4) there have been instances where the assumptions have been known to have been compromised and yet the assessments are not necessarily rejected. In some cases a Tier 4 has been used when a Tier 3 only provided highly implausible predictions. But, in many cases where Tier 4 assessments have been known to have been invalid, the absence of an alternative that could provide catch level advice meant that the assessments were not rejected. The simplest example of this is where all the deep water Oreo species are assessed using the Tier 4 approach. The fundamental assumption behind the Tier 4 analyses (which use catch and standardized catch rate time series of data) is that catch rates reflect relative abundance of the stock and are representative of the whole stock. Neither of these assumptions are met, especially since the advent of the 700m closure. Oreo catch rates vary from extremely low to extremely high, depending upon whether the aggregations of fish are targeted or not (**Figure 62**).

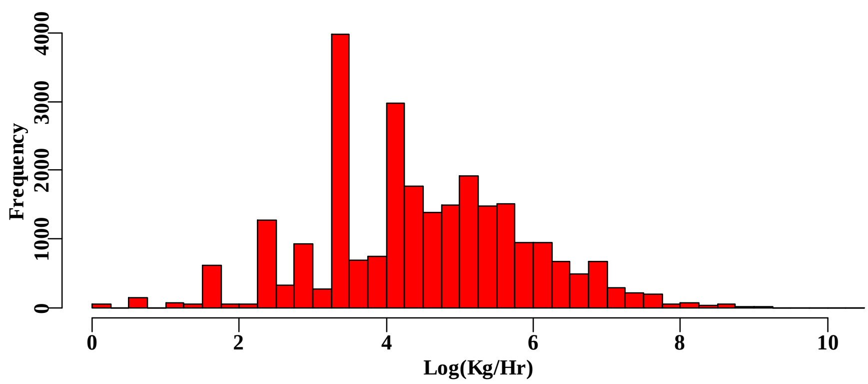

Figure 62. The log-transformed CPUE for the mixed Oreo category, to the end of 2013, which includes *Allocyttus verrucosus* (Warty Oreo), *Neocyttus rhomboidalis* (Spiky Oreodory), *Neocyttus psilorhynchus* (Rough Oreodory), *Allocyttus niger* (Black Oreodory) and a further mixed category (Oreo Dory). Note the spikiness of the lower levels of CPUE containing large numbers of records. The first five spikes relate to 5, 10, 15-20, 30, and 60 kg/hr.

Apart from being applied to a mixed species group (although ~97% tends to be Spikey Oreodory), the variation between catch rates between years is so great in some instances that it is biologically impossible for those catch rates to be reflecting the relative abundance of the stocks. With the 700m closure most of the Oreo habitat is now closed and the records that are obtained are in no way representative of the whole. Even a data-poor assessment method may not work for such species, where such a high proportion of the stock is protected from fishing mortality inside of closures. Some other means of allocating a TAC needs to be developed; none of the methods investigated here, even average catches, are applicable because the large closure has distorted the representation of the stock.

It is the case that such a large proportion is now no longer able to be fished because of the 700m closure that it would be valid to allocate a catch level that would not constrain catches anywhere and instead of assessing the stocks each year, merely monitoring catches to determine if they rise in the future and then decide whether more intervention is required. A set of trigger catch levels would provide the management tools necessary to defend the management (Dowling *et al*., 2015a,b).

There are no standard, routine methods, or formal criteria that can be applied to determine whether a fisheries stock assessment is appropriate which can be applied independently of the assessment and management process in which it is embedded. In the SESSF it is the Resource Assessment Groups (RAGs) that determine whether or not to accept a stock assessment, and this tends to be done on a weight-of-evidence approach that attempts to account for consistency through time, the relative quality of fit of the model to the data, and whether the model structure correctly represents the stock dynamics as far as they are known. Most often a stock assessment might be rejected on the basis of qualitative reviews of the match between the model structure used and what is known about the fished stock. In a Tier 1 assessment if two data streams are in conflict, with one implying things are improving and another implying things are declining, it would be more usual for one of the data streams to be rejected rather than the assessment. At very least the sensitivity of the assessment outcomes to including or excluding (or down-weighting) each data-stream would be examined.

While there are no formal criteria presently available, beyond the classical statistical fit criteria, it would undoubtedly be helpful for keeping processes open and understood, if such more formal criteria were developed. This is not suggesting that the current less formal review of the applicability of an assessment be discontinued, but rather that at least some more formal aspects be recognized and made part of the RAG's routine so as to make communication and understanding simpler.

14.3 The Effects of Sample Size, CV, and Bias on Assessments

14.3.1 Tier 3

The Tier 3 assessment method and harvest strategy appears capable of achieving the Target Reference Point of 48%B_0 for Flathead (and similar species) for all levels of age sample size even when starting from low or high levels of initial depletion, but this is the case only if there is no or only slightly positive sampling bias. If the sampling has a significant positive bias the outcomes can lead to missing the target in about 75% of occasions when the stock started below the target, but ended either at the target or just below the limit if the stock started well above the target. This was simply a reflection that the catches in the latter case were badly over-estimated.

With School Whiting (and similar species) the Tier 3 was only able to achieve the target or remain above it if there was negative bias. Positive bias led to depleted states and with the maximum positive bias the medians were effectively on the limit reference point. In all cases of different sample sizes and separately, with no bias the median depletion ended at or just below about 40% instead of the target of 48%. Once again, positive bias in the aging samples generated misleading outputs and undesirable management outcomes.

14.3.2 Tier 4

The effects on the Tier 4 applied to Flathead-like species, of the coefficient of variation (CV) of the CPUE and of bias tended to be highly exaggerated for the very high CVs of 0.6 and 0.9. These two levels led to time series of CPUE which were unlike any seen in the real fishery and so these levels should be ignored for Flathead.

With the remaining levels of CV, not surprisingly, as the CV increases so does the spread around the median levels of catch, the final depletion, and the probability of avoiding the LRP. The catch variation appeared to increase exponentially for CVs of 0.05, 0.15, and 0.45. The effect of positive levels of bias is to over-estimate the sustainable catches, which in turn leads to greater levels of depletion. If the stock starts already below the target then positive bias can force it down to or just below the LRP. This is of concern as 'effort creep' would lead to positive bias and has undoubtedly occurred with the advent of GPS and colour depth sounders, etc, for which there is no information that can be included in any of the CPUE standardizations.

With the School Whiting like species, the effects of increasing CV on the Tier 4 outcomes was to increase variation in catches, but this species is already highly variable from year to year so the higher values of CV have less of an impact on the outcomes. When the species starts already in a depleted state then the final median depletion level is always below the TRP; this is, however, related to the chance level of depletion related to the targets selected in the Tier 4 HCR. The Tier 4 HCR selects a reference period of years to identify the target catch and CPUE which drive the HCR. It is a mistake to believe these Tier 4 targets are at 48%B_0 rather than simply being a state of the fishery identified to be a good place to be in terms of sustainability and profitability; these targets are merely proxies and may in fact be above or below the equilibrium level represented by 48%B_0 (Haddon, 2014). Thus the outputs demonstrate that the method can generate consistent outcomes across a wide range of precision for both initially heavily and lightly depleted states (although with outcomes below the formal TRP and above the TRP respectively).

The effect of positive bias is very similar to that seen in the Flathead-like species. Strongly positive bias can lead to serious depletion and failure to avoid the LRP in the initially heavily depleted scenarios, although in the lightly depleted scenarios they all remain above the formal TRP.

The impact of positive bias is especially important as positive bias in CPUE could be brought about by improvements in technology and fishing practices. As it appears highly likely that such 'effort creep' will have occurred it would be valuable to further explore the possible impact of such positive biases in a more specific manner relating to the advent of events leading to increasing bias as a series of events across just a few years (e.g. the advent of GPS from 1990 – 1992). Having no way of taking such changes and resulting bias into account may be leading to overly optimistic views of each of the fisheries.

In Tier 1 assessments the effects of such bias would be detectable through the time series of CPUE or the ageing or length frequency data being inconsistent with each other. It is primarily the Tier 3 and Tier 4 methods that require further exploratory analyses.

14.4 Data-Poor Methods for Estimating Sustainable Catch

The objective of data-poor and data-limited methods is to estimate a practical level of yield that is likely to be sustainable (MacCall, 2009). By 'practical', MacCall means commercial yields rather than overly conservative yields.

The MSE testing highlights that a stock should never be assumed to be in only a lightly depleted state. A reasonable option if the assessment process (i.e. including the RAG's involvement) determines that a particular stock has only been lightly fished, is to assume 40%B_0 so as to avoid the risk of over-fishing. This is, however, a policy decision and all that can be done here is to point out that, obviously assuming an initial depletion of 40% (which would lead to lower predicted catches) would be more conservative and there would be a lower risk of over-fishing than selecting 48%B_0 as the assumed initial depletion. However, there would also be a higher risk of failing to take as much catch as would be sustainable. Given the MSE results, 40% should certainly avoid the stock declining below the LRP (even if the stock were really depleted to the LRP and it was the DCAC method; see **Figure 33** where DCAC only just meets the < 10% probability of being below the LRP).

14.4.1 Central Tendency of Catch Methods

Possible candidates for use in a new Tier 5 set of assessments would include the methods that involve a measure of the central tendency of catches such as the average or median catch (possibly the 3rd highest catch). Ideally, these average catches would be estimated from periods of stability within each fishery, but in reality, in Australia, such periods are not common. Such central tendency methods involve empirical harvest strategies where the estimated central tendency catch constitutes the sustainable catch estimate (the 'assessment' is the decision rule; **Figure 60**). The recommended sustainable catches would need to be presented in the context of a weight of evidence appraisal of whatever stock was being considered. Dowling et al (2015a,b) discuss the use of such catches in the context of a set of catch triggers where a set of catch levels are set that, if met by the fishery, trigger management actions that can vary from a simple review of events to the application of some simple assessment or update of the average catch applied. In the Commonwealth HSP within the SESSF this would entail setting a multi-year TAC that would be reviewed for a breakout each year and reviewed as to its level every few years.

Using a central tendency of catch estimate to set upper limits to catch before further management action, requires the assumption that the stock is currently in an acceptable state or that the catches already observed have not led to serious or undesirable levels of depletion. If the weight-of-evidence appraisal supports this assumption then a recommended biological catch can be made. Reasons for not using this approach include that the time series of catch data is not representative of the fishery or potential fishery (**Figure 63**), or that the catch data is too sporadic to obtain a representation of the fishable stock. Specific trigger catch levels could then be set (Dowling *e al*., 2015b). Whether a discount would be required would depend on the final decision rule adopted. In the MSE testing the particular central tendency of catch was used (mean or median) but some other quantile could be used. The 3rd highest catch usually proved to be as capable as the other central tendency methods at avoiding the LRP, so an average or median should be sufficiently conservative as long as the state of depletion is considered to be acceptably far from the LRP at the start.

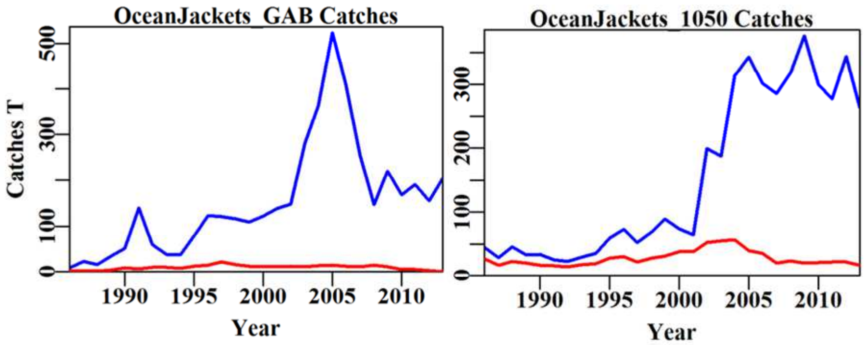

Figure 63. The catch history for Ocean Jacket (*Nelusetta ayraudi*) in the GAB and zones 10 – 50. The change in catches has been reported to reflect a major change in availability in the early 2000s, but the drop in catches in the late 2000s in the GAB is reported to be more about a lack of a market than not being available.

14.4.2 Model Supported Catch-Based Methods

The model supported catch-based methods include the Depletion-Corrected Average Catch (DCAC), the Depletion Adjusted Catch Scalar (DACS), and the Depletion-Based Stock Reduction Analysis (DB-SRA). Among these methods the DACS and DCAC are somewhat simpler to implement than DB-SRA. Each of these also has assumptions and input requirements beyond having estimates of natural mortality. Fortunately, these input requirements are not especially strict or onerous and even when relatively strong assumptions are made (such as restricting the initial and final depletion levels to values that would be conservative) these methods can still generate solutions. The advantage of these model supported methods is that whatever estimate of sustainable catch is derived it comes with an estimate of the uncertainty about the estimate (see **Figure 53** and **Figure 55**), so there is freedom in the harvest strategy to add further precaution if it is deemed necessary. This might depend on whether the RAG considered the catch time series used to be reliable. For example, earlier in the recent history of catches of blue and Silver Warehou (*Seriolella brama* and *S. punctate*) the two species were not distinguished. For example, "… in 1992 both species were lumped under a global TAC of 4000t, 2000t of which was allocated to the trawl sector. Separate TACs were established in 1993 to avoid issue of high-grading spotteds [Silver Warehou] in favour of blues." (Smith *et al*., 1994). Such potential flaws in the available catch data could be solved by eliminating the early data, although in the context of Blue Warehou, the early catches are verbally reported to have been large.

Assuming that no stock would be assumed to be initially well above the target (TRP), then no major consistent differences were observed between these three approaches.

The DB-SRA provides more information than the other two methods and so if it can be implemented this would be the method of choice. But at least the DCAC should also be run to ensure that the estimates are not significantly different. A comparison of at least two of methods should assist in discovering any unusual aspects of the available data as some reason would need to be found for any differences. The DB-SRA allocated a predicted depletion level although it doesn't necessarily hit the allocated value each replicate (see **Figure 57**). The trials run provide a spread of trajectories and the proportion that dip below the 20% depletion line in the final year would provide an indication of the relative risk of the predicted median MSY value of failing to meet the criteria of avoiding the LRP 90% of the time.

14.4.3 Alternative Data-Poor Assessment Methods

These model supported (or assisted) methods, as implemented by their respective authors, all assume that the full catch history is reliably known, which is not the case for many of the species in the SESSF. Alternative implementations that start within some selected range of initial depletion can solve this issue. Making such an implementation with DB-SRA would appear to be an evolutionary step that led to the development of the Catch-MSY method. While the DB-SRA provides useable and reasonable MSY values even when used with the most recent 28 years of data (see **Table 13**) it would clearly be sensible to explore the capacity, strengths, and weaknesses of this relatively new approach. MSE explorations of the method are on-going in the USA (pers comm Steve Martell) so such investigations are not necessarily required here.

The field of data-poor stock assessments and harvest strategies is receiving a great deal of attention world-wide (Dowling *et al*., 1915a,b). As new methods and alternative approaches are developed and reported these should be reviewed and considered for inclusion in the options available to the new Tier 5 category in the SESSF. The new Tier and its associated methods will have immediate value in providing substitute methods for species which are currently either not assessed or not assessed validly using the current tiers. However, much of the implementation and use of the new methods will also be dependent upon the content and requirements of the revised Commonwealth Harvest Strategy Policy that is due to be introduced this year (2015). Until it is known exactly which new species are going to require some form of assessment the exact range of data-poor methods that will be used cannot be known but this current work has identified an array of candidate methods that can be used with almost any species.

14.5 Communication of New Methods

The third objective for this project was to:

Produce presentations and explanatory documents for distribution across RAGs and MACs, describing the criteria and new Tier 5 harvest strategies.

This will be done ready for this year's round of RAG meetings. The first is the SESSF RAG data meeting in the first week of August 2015, then there are two further sets of meetings one likely to be in September and the other in October or November. Presentations will be made to the various separate RAGs that meeting during these multi-day meetings (SHELF, SLOPE, SHARK) and also at the GAB RAG, which tends only to meet once in October or November. Worked examples will be included, with details of the methodology, and how each method would be used in practice. In addition, electronic copies of the Final Report will be distributed once completed and published.

15 Appendix: Staff

In alphabetical order:

Assoc. Prof. Malcolm Haddon
Dr. Neil Klaer
Dr. Geoff Tuck
Dr. Sally Wayte

16 References

Anderson, S. C., Branch, T. A., Ricard, D., and Lotze, H. K. (2012) Assessing global marine fishery status with a revised dynamic catch-based method and stock-assessment reference points. – *ICES Journal of Marine Science*, **69**: 1491–1500.

Beddington, J.R. and J.G. Cooke (1983) The Potential Yield of Fish Stocks. *FAO Fisheries Technical Paper* **242:** 1-47.

Berkson, J., Barbieri, L., Cadrin, S., Cass-Calay, S., Crone, P., Dorn, P., Friess,C., Kobayashi, D., Miller, T.J., Patrick, W.S., Pautzke, S., Ralston, S.,Trianni, M., 2011. Calculating acceptable biological catch for stocks that have reliable catch data only. *NOAA Technical Memorandum, NMFS-SEFSC-616*. http://www.pifsc.noaa.gov/tech /NOAA Tech Memo SEFSC 616.pdf

Bentley N. (2014). Data and time poverty in fisheries estimation: potential approaches and solutions. *ICES Journal of Marine Science*, doi:10.1093/icesjms/fsu023

Bentley, N. and Langley, D. (2012) Feasible stock trajectories: a flexible and efficient sequential estimator for use in fisheries management procedures. *Canadian Journal of Fisheries and Aquatic Sciences*, **69**:161-177.

Butterworth, D.S. and M.O. Bergh (1993) The development of a management procedure for the South African Anchovy resource. P 83-99 *in* Smith, S.J, Hunt, J.J. & D Rivard (*eds*) Risk Evaluation and Biological Reference Points for Fisheries Management. *Canadian Special Publication of Fisheries and Aquatic Sciences* **120:** 1:442.

Caddy, J.F. and R. Mahon (1995) Reference Points for Fisheries Management *FAO Fisheries Technical Paper* **347** FAO, Rome. 83 p.

Caddy, J.F. and R. McGarvey (1996) Targets or limits for management of fisheries? *North American Journal of Fisheries Management* **16:** 479-487

Carruthers, T.R., Punt, A.E., Walters, C.J., MacCall, A., McAllister, M.K., Dick, E.J. and Cope, J. (2014) Evaluating methods for setting catch limits in data-limited fisheries. *Fisheries Research* **153**:48-68.

DAFF (2007) Commonwealth Fisheries Harvest Strategy. Policy and Guidelines. Department of Agriculture, Fisheries and Forestry. 55p.

Day, J. (2010) School Whiting (*Sillago flindersi*) stock assessment based on data up to 2008 pp 190 – 249 *in* Tuck, G.N. (ed.) 2010. Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2009. Part 1. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 334 p.

Day, J. (2012) School Whiting (*Sillago flindersi*): further exploration of fixed projected catches, potential indicators and alternative harvest strategy analyses. Pp 290 – 329 *in* Tuck, G.N. (ed.) 2012. Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2011. Part 1. Australian Fisheries Management Au-

thority and CSIRO Marine and Atmospheric Research, Hobart. 377 p.
Day, J. and N. Klaer (2014) Tiger Flathead (*Neoplatycephalus richardsoni*) stock assessment based on data up to 2012. Pp 174 - 232 *in* Tuck, G.N. (*ed*) (2014) *Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2013. Part 1.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 313p.
Dick, E.J. & MacCall, A.D. (2011) Depletion-based stock reduction analysis: a catch-based method for determining sustainable yields for data-poor fish stocks. *Fisheries Research*, **110** (2): 331–341.
DiCosimo, J. (*ed*) (2001) *Summary of the Gulf of Alaska Groundfish Management Plan*. North Pacific Management Council, Anchorage, Alaska 99501. 16p. Available at: http://www.npfmc.org/wp-content/PDFdocuments/fmp/GOA/GOASummary.pdf
Dowling, N. (2011) Management strategy evaluation testing of the management strategies used with North West slope trawl fisheries. RUSS. CSIRO, Hobart. 87p.
Dowling, N.A., Dichmont, C.M., Venables, W., Smith, A.D.M., Smith, D.C., Power, D. and Galeano, D. (2013) From low- to high-value fisheries: Is it possible to quantify the trade-off between management cost, risk and catch? *Marine Policy* **40**:41-52.
Dowling,N.A., Haddon,M., Smith,D.C., Dichmont,C.M., and A.D.M.Smith (2011a) *Harvest Strategies for Data-Poor Fisheries: A brief review of the Literature.* CSIRO, Hobart. 48 pp.
Dowling, N.A., Dichmont, C.M., Smith, A.D.M., Smith, D.C., and M. Haddon (2011b) *Guidelines on developing harvest strategies for data-poor fisheries.* CSIR0, Hobart. 32pp.
Dowling, N.A., Dichmont, C.D., Haddon, M., Smith, D.C., Smith, A.D.M. and K. Sainsbury (2015a) Guidelines for developing formal harvest strategies for data-poor species and fisheries. *Fisheries Research* dx.doi.org/10.1016/j.fishres.2014.09.013
Dowling, N.A., Dichmont, C.D., Haddon, M., Smith, D.C., Smith, A.D.M. and K. Sainsbury (2015b) Empirical harvest strategies for data-poor fisheries: A review of the literature. *Fisheries Research* dx.doi.org/10.1016/j.fishres.2014.11.005
FAO (1995) *Code of Conduct for Responsible Fisheries*. Food and Agricultural Organization of the United Nations, Rome. 41p.
FAO (1996) *Precautionary Approach to Capture Fisheries and Species Introductions*. FAO Technical Guidelines for Responsible Fisheries 2. Food and Agricultural Organization of the United Nations, Rome. 55p.
FAO (1997) *Fisheries Management.* FAO Technical Guidelines for Responsible Fisheries 4. Food and Agricultural Organization of the United Nations, Rome. 84p.
Fay, G. Punt, A.E. and A.D.M. Smith (2009) Operating Model Specification pp 125 – 133 *in* Wayte,S.E.(ed) (2009) *Evaluation of new harvest strategies for SESSF species.* CSIRO and AFMA. 137 p.
Fay, G., Little, L.R., Tuck, G.N., Haddon, M. and N.L. Klaer (2013) Maintaining risk equivalency among fishery harvest control rules in the SESSF. pp 1 – 29 *in* Tuck, G.N. (*ed*) (2013) *Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2012. Part 2: Tier 3 and Tier 4, catch rate standardizations and other work contributing to the assessment and management of the SESSF stocks in 2012.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 550p.
Flood, M, Stobutzki, I, Andrews, J, Ashby, C, Begg, G, Fletcher, R, Gardner, C, Georgeson, L, Hansen, S, Hartmann, K, Hone, P, Horvat, P, Maloney, L, McDon-

ald, B, Moore, A, Roelofs, A, Sainsbury, K, Saunders, T, Smith, T, Stewardson, C, Stewart, J & Wise, B (eds) (2014) *Status of key Australian fish stocks reports 2014*, Fisheries Research and Development Corporation, Canberra. 620 p.

Geromont, H.F. and D.S. Butterworth (2014) Generic management procedures for data-poor fisheries: forecasting with few data. *ICES Journal of Marine Science* doi:10.1093/icesjms/fst232 1:11

Haddon, M., Ziegler, P., Lyle, J., and Burch, P. 2005. Using a spatially structured model to assess the Tasmanian fishery for Banded Morwong (*Cheilodactylus spectabilis*). *Fisheries Assessment and Management in Data-Limited Situations* 21: 737-756.

Haddon, M. (2007) Fisheries and their management. Pp 515-532. *In Marine Ecology* (eds) S.D. Connell & B.M. Gillanders. Oxford University Press. 630 p.

Haddon, M. (2011) *Modelling and Quantitative Methods in Fisheries.* Chapman & Hall, CRC Press, 449 p.

Haddon, M. (2011b) *Management Strategy Evaluation testing of the Management Strategies used with South-Eastern Scallop Fisheries.* Report prepared for DAFF: part of the Reducing Uncertainty in Stock Status project. CSIRO Wealth from Oceans Flagship. 100 p.

Haddon, M. (*ed)* (2012) *Reducing Uncertainty in Stock Status: Harvest Strategy Testing, Evaluation, and Development. General Discussion and Summary.* CSIRO Marine and Atmospheric Research. 42 p.

Haddon, M. (2014) Tier 4 analyses in the SESSF, including deep water species. Data from 1986 – 2012. Pp 352 – 461 *in* Tuck, G.N. (*ed*) (2014) *Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2013. Part 1.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 313p.

Haddon, M. (2014b) Review of the Biomass-Augmented Catch-MSY Model for Pacific Island Coral Reef Ecosystem Resources. A review for the Center for Independent Experts, NOAA and NMFS, USA. 28p.

Haddon, M., Klaer, N., Smith, D., Dichmont, C. and T. Smith (2013) *Technical Reviews for the Commonwealth Harvest Strategy Policy*. FRDC 2012/225. CSIRO. Hobart. 69p. *in* Haddon, M., Dichmont, C., Klaer, N., Pascoe, S., Penney, A., Smith, T., Smith, D., Vieira, S. and P. Ward (2014) *Technical Reviews of Formal Harvest Strategies.* FRDC 2012/225. ABARES and CSIRO, Hobart. 29p and appendices.

Hilborn, R. (2010) Pretty good yield and exploited fisheries *Marine Policy* **34:**193-196.

Hilborn, R. and T.A. Branch (2013) Comment: Does catch reflect abundance? No, it is misleading. *Nature* **494**: 303-306

Hobday, A.J., Smith, A.D.M., Stobutzki, I.C., Bulman, C., Daley, R., Dambacher, J.M., Deng, R.A., Dowdney, J., Fuller, M., Furlani, D., Griffiths, S.P., Johnson, D., Kenyon, R., Knuckey, I.a., Ling, S.A., Pitcher, R., Sainsbury, K.J., Sporcic, M., Smith T., Turnbull, C., Walker, T.I., Wayte, S.E., Webb, H., Williams, A., Wise, B.S. and S. Zhou (2011) Ecological risk assessment for the effects of fishing. *Fisheries Research* **108**: 372 – 384

Kelly, C.J., and Codling, E.A. 2006. 'Cheap and dirty' fisheries science and management in the North Atlantic. *Fisheries Research* 79: 233-238.

Klaer, N.L. and Wayte, S.E. (2011 draft) Demersal MSE for trawl fish in the Southern and Eastern Scalefish and Shark Fishery and other like-species. Part of the RUSS Project.

Klaer, N. (2010) Tiger Flathead (Neoplatycephalus richardsoni) stock assessment based on data up to 2008. pp 164 – 189 *in* Tuck, G.N. (ed.) 2010. Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2009. Part 1. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 334 p.

Klaer, N. (2014) Yield, Total Mortality values and Tier 3 estimates for selected Shelf and Slope species in the SESSF 2013. Pp 7 – 56 *in* Tuck, G.N. (*ed*) (2014) *Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2013. Part 1.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 313p.

Klaer, N., Day, J., Fuller, M., Krusic-Golub, K. and J. Upston (2014) Data summary for the Southern and Eastern Scalefish and Shark Fishery: Logbook, Landings, and Observer Data to 2013. Australian Fisheries Management Authority and CSIRO Oceans and Atmosphere. 264 p.

Little, L.R., Wayte, S.E., Tuck, G.N., Smith, A.D.M., Klaer, N., Haddon, M., Punt, A.E., Thomson, R., Day, J. & Fuller, M. (2011) Development and evaluation of a cpue-based harvest control rule for the southern and eastern scalefish and shark fishery of Australia. *ICES Journal of Marine Science*, **68,** 1699-1705.

MacCall, A.D. 2009 Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. *ICES Journal of Marine Science,* **66**:2267-2271.

Martell, S. and Froese, R. (2013) A simple method for estimating MSY from catch and resilience. *Fish and Fisheries* **14**:504-514.

Ministry for Primary Industries (2014). Fisheries Assessment Plenary, May 2014: stock assessments and stock status. Compiled by the Fisheries Science Group, Ministry for Primary Industries, Wellington, New Zealand. 1381 p.

Newman, D., Berkson, J., and L. Suatoni (2015) Current methods for setting catch limits for data-limited fish stocks in the United States. *Fisheries Research* **164**: 86-93.

Pauly, D. (2013) Comment: Does catch reflect abundance? Yes, it is a crucial signal. *Nature* **494**: 303-305

Pikitch, E., Boersma, P.D., Boyd, I.L., Conover, D.O., Cury, P., Essington, T., Heppell, S.S., Houde, E.D., Mangel, M., Pauly, D., Plagányi, É., Sainsbury, K., and Steneck, R.S. (2012) *Little Fish, Big Impact: Managing a Crucial Link in Ocean Food Webs*. Lenfest Ocean Program. Washington, DC. 108 pp.

Plagányi, É.E., Skewes, T.D., Dowling, N.A. and M. Haddon (2013) Risk management tools for sustainable fisheries management under changing climate: a sea cucumber example. *Climate Change* **119:** 181-197.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A. and M. Haddon (2014) Management strategy evaluation: best practices. *Fish and Fisheries* DOI: 10.1111/faf.12104

Rosenberg, A.A., Fogarty, M.J., Cooper, A.B., Dickey-Collas, M., Fulton, E.A., Gutierrez, N.L., Hyde, K.J.W., Kleisner, K.M., Kristiansen, T., Longo, C., Minte-Vera, C.V., Minto, C., Mosqueira, I., Osio, G.C., Ovando, D., Selig, E.R., Thorston, J.T. and Ye, Y. (2014) Developing new approaches to global stock status assessment and fishery production potential of the seas. *FAO Fisheries and Aquaculture Circular No. 1086*. FIRF/C1086 (En). 175pp.

Sabater, Marlowe and Pierre Kleiber (2014) Improving Specifications of acceptable biological catches of data-poor reef fish stocks using a biomass-augmented catch-MSY approach. Western Pacific Regional Fishery Management Council, Honolulu, HI 96813 USA. NOAA Technical Memorandum NMFS-PIFSC-29 60p

Smith, D.C., McCoy, P., and J. Cottier (1994) *Assessment of the trawl and gillnet fisheries for warehous0.* Final Report to the Fisheries Research and Development Corporation. FRDC Project 90/11

Smith, A.D.M., Smith, D.C., Tuck, G.N., Klaer, N., Punt, A.E., Knuckey, I., Prince, J., Morison, A., Kloser, R., Haddon, M., Wayte, S., Day, J., Fay, G., Pribac, F., Fuller, M., Taylor, B., Little., R (2008) Experience in implementing harvest strategies in Australia's south-eastern fisheries. *Fisheries Research* **94:** 373-379.

Smith, A.D.M., Smith, D.C., Haddon, M., Knuckey, I., Sainsbury, K.J., Sloan, S. (2014) Implementing harvest strategies in Australia: five years on. *ICES Journal of Marine Science* **71:** 195-203.

Tuck, G.N. (*ed*) (2014) *Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2013. Part 1.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 313p.

Tukey, J.W. (1980) Sunset Salvo. *The American Statistician* **40**: 72-76

Vasconcellos, M., and Cochrane, K. 2005. Overview of world status of data-limited fisheries: Inferences from landings statistics. *Fisheries Assessment and Management in Data-Limited Situations* **21**: 1-20.

Wayte,S.E.(ed) (2009) *Evaluation of new harvest strategies for SESSF species.* CSIRO and AFMA. 137 p.

Wayte, S.E. (2013) Management implications of including a climate-induced recruitment shift in the stock assessment for Jackass Morwong (*Nemadactylus macropterus*) in south-eastern Australia. *Fisheries Research* **142**: 47 – 55

Wayte, S.E. and N.L. Klaer (2010) An effective harvest strategy using improved catch-curves. *Fisheries Research* **106**: 310 – 320

Wetzel, C.R., Punt, A.E. (2011) Model performance for the determination of appropriate harvest levels in the case of data-poor stocks. *Fisheries Research,* **110** (2): 342-355.

Zhou, S., Pascoe, S., Dowling, N., Haddon, M., Klaer, N., Larcombe, J., Smith, T., Thebaud, O., Viera, S. and Wayte, S. (2013) *Quantitatively defining biological and economic reference points in data poor fisheries.* Fisheries Research and Development Corporation Project No. 2010/044.

Zhou, S., Haddon, M., Smith, A.D.M., Dichmont, C.M. and D.C. Smith. (*Submitted*) An optimised catch-only assessment method for data-poor fisheries. *ICES Journal of Marine Science*

CONTACT US
t 1300 363 400
+61 3 9545 2176
e enquiries@csiro.au
w www.csiro.au

YOUR CSIRO
Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills for building prosperity, growth, health and sustainability. It serves governments, industries, business and communities across the nation.

FOR FURTHER INFORMATION
Division/Unit Name
Malcolm Haddon
t +61 3 6232 5097
e Malcolm.haddon@csiro.au
w www.csiro.au/