www.csiro.au

Operationalising the risk-cost-catch trade-off

Catherine M. Dichmont, Elizabeth Fulton, André E. Punt, Lorne R. Little, Natalie Dowling, Rebecca Gorton, Miriana Sporcic, David C. Smith, Malcolm Haddon, Neil Klaer

FRDC Project No 2012-202

2017

ISBN 978-1-4863-0708-1 (online)

Operationalising the risk-cost-catch trade-off

FRDC 2012-202

2017

Ownership of Intellectual property rights

This publication (and any information sourced from it) should be attributed to Catherine M. Dichmont, Elizabeth Fulton, André E. Punt, Lorne R. Little, Natalie Dowling, Rebecca Gorton, Miriana Sporcic, David C. Smith, Malcolm Haddon, Neil Klaer. 2017. *Operationalising the risk-cost-catch trade-off.* CSIRO Oceans and Atmosphere. Fisheries Research and Development Corporation Project 2012-202, Brisbane, October 138pp.

Creative Commons licence

Disclaimer

The Fisheries Research and Development Corporation plans, invests in and manages fisheries research and development throughout Australia. It is a statutory authority within the portfolio of the federal Minister for Agriculture, Fisheries and Forestry, jointly funded by the Australian Government and the fishing industry.

Researcher Contact Details

Name: Dr Catherine Dichmont
Address: Cathy Dichmont Consulting
47 Pioneer Rd, Sheldon, Qld, 4157
Phone: +61 419 950 076
Email: cathy.dichmont@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

Tables

Figures

Acknowledgments

We would like to thank several CSIRO (Drs Rik Buckworth, Cathy Bulman, Robert Campbell, Richard Hilary, Tim Skewes and Geoff Tuck), CEFAS (Dr Simon Jennings), NMFS (Dr Chris Legault), ABARES (Drs Mark Chambers, Ilona Stobutzki), anonymous and Guest Editor reviewers for their in-depth comments on draft versions of the various sections of the report.

Appendix F was co-written with AFMA with much input from AFMA staff. This input is greatly appreciated as it clarified some very important points.

The input from both the TTRAG and SESSFRAG at various stages of the project are much appreciated.

This work was funded by the Fisheries Research and Development Corporation Project 2012/202, CSIRO Oceans and Atmosphere and Cathy Dichmont Consulting.

Thanks enormously to Tonya van der Velde who helped with the collation of the various report components and getting the report to look really good.

Abbreviations

Abbreviation	Explanation
AFMA	Australian Fisheries Management Authority
DA	Department of Agriculture
MSE	Management Strategy Evaluation
SESSF	South and Eastern Shark and Shelfish Fishery
AtlantisRCC	Atlantis model applied to this project
HS	Harvest Strategy
HSP	Harvest Strategy Policy
USA	United States of America
ICES	International Council for the Exploration of the Sea
MSY	Maximum Sustainable Yield
MEY	Maximum Economic Yield
F	Fishing mortality
M	Natural mortality
B	Biomass

Executive Summary

Management of renewable resources such as fisheries can be complex given the range of species and habitats affected by fishing. This is true even if the focus is only on species that are directly targeted by fisheries. There is often a range of information sources and data quality available by fishery and species, with fisheries ranging from small scale, low value to large scale, high value. Fisheries throughout the world are undertaking increasingly complex science and monitoring in the face of higher demands for demonstrable scientific evidence across all aspects of Ecosystem Based Fisheries Management (EBFM) (economic, environmental, social). The problem, however, is that fishery management, including those for Australian Commonwealth fisheries, are facing these challenges with diminishing budgets. Many of the Australian Fisheries Management Authority's (AFMA's) managed fisheries have experienced large scale change in the past few years (e.g. fleet reductions through buy-backs, spatial Marine Protected Areas, Individual Tranferable Quotas), which have further complicated matters. This means that smaller fleets, in the context of cost recovery, have to support the industry component of science and information requirements. These costs can become onerous in some fisheries. The issue is how to trade the risks associated with managing a fishery, against the costs of management and the socio-economic benefit obtained from the catch?

Given that fisheries are often low in value, there is a real need to understand the trade-off between ecological and economic risk associated with harvesting; the benefits of harvesting, namely catch; and the costs associated with management. This relationship is known as the risk-cost-frontier and has been adopted conceptually by AFMA. Catch has been defined in terms of both its mean and variability. If, for example, fishing becomes increasingly aggressive, the overall mean catch may increase, at least in the short term, but so too may its inter-annual variability. Management costs are defined broadly in terms of the cost of information, the management decision process, management measures, and research and compliance. Finally, risk is used in the context of undesirable consequences to the target species, associated species, as well as to the economics of the fishery.

This work crosses the gap between concept and implementation. It is divided into several components that together provide insight into operationalising the risk-cost-catch trade-off:

1. an international review of tiered systems that attempts to address a core component of the risk-cost-catch trade-off, which is risk equivalency when using data-rich to data-limited assessment methods;
2. a principal component analysis of the existing SESSF tier system, and the expanded potential system developed by Dowling et al., (2013);
3. a 'first principle's approach to identifying the risk-cost-catch trade-off using market based economics (monetary economics);
4. Management Strategy Evolutions (MSE) of harvest strategies (HSs) based on those used for AFMA-managed fisheries and developed by Dowling et al., (2013), using the whole-of-system model Atlantis as previously applied to the South-east fisheries region (Fulton et al., 2014) – hereafter called AtlantisRCC;
5. tests of the buffer systems applied in the Australian Southern and Eastern Scalefish and Shark Fishery (SESSF) and USA groundfish fishery of the tier 1 to 4 harvest strategies as applied in the SESSF in 2015, again using AtlantisRCC; and
6. a section on how best to move towards operationalising the risk-cost-catch trade-off that forms the basis for further development, if required, for the new and present HSs by AFMA and/or the Department of Agriculture (DA).

Tier systems review

Several jurisdictions have developed hierarchical tier systems that categorise stocks based on, for example, the data available for assessment purposes and/or the extent to which the quantities on which management advice is based can be estimated. Four case studies (Australia's Southern and Eastern Scalefish and Shark Fishery, the USA west coast groundfish fishery, the USA Alaskan crab fishery, and EU fisheries) are used to contrast the types of hierarchical tier systems available, and to assess the extent to which each system constrains risk to be equivalent among the tiers (termed risk equivalency). Only the Australian system explicitly aims to achieve risk equivalency. However, this intent has not been fully operationalised. The review reveals that best practice is not to define tiers simply on data availability, but also on what the assessments based on those data are capable of estimating. In addition, clearly differentiating the quantification of uncertainty from how decision makers wish to address that uncertainty would simplify justification of buffers (the gap between the assessment-produced target catch or effort and the final management decision that accounts for uncertainty and risk). Risk equivalency can be achieved by using management strategy evaluation to select the values for control variables, which determine the buffer given the uncertainty associated with the assessment.

Principal Component Analysis

A principal component analysis was applied to the expanded Australian Commonwealth 8-tier system for fishery assessment and management to determine whether it adequately delineates across stocks according to data availability and quality. The presently implemented Australian tiers comprised four levels that were defined primarily according to the available stock assessment options, given data availability and quality. We asked fishery experts to score information quality for each of the main Australian Commonwealth species and/or fisheries. Multivariate analysis indicated that the eight tiers delineated between the extreme tier levels on the first principal component, although there was overlap for intermediate tiers. More generally, it is important that the aim of tier systems and the basis for tier delineations are explicitly defined given the increasing association of tiers with trade-offs between overfishing risk, management cost and catch.

Monetary economic analysis

Fisheries management operates in an environment characterized by multiple risks. These risks are often complementary, and can be traded off against each other. An important goal for managers is to develop strategies to minimize overall risk exposure. We illustrate a simple theoretical framework that quantifies a range of risks faced by fisheries management agencies in terms of expected budgetary expenditure. The analysis calculates the cost a management agency would be expected to incur from overfishing a stock, from being seen to overfish it, and from foregoing economic returns. These costs can be controlled by adjusting the biomass level targeted by management, or increasing expenditures for data collection to improve the precision of biomass estimates. The overall risk, expressed as the total expected cost to a management agency, depends strongly on the fisheries management objectives, and the emphasis between conservation and economic return. In general, a conservation-oriented objective would minimize risk by increasing target biomass levels, or expenditure on monitoring and assessment, while a profit-focused objective would seek to lower management costs by reducing expenditure on data collection and assessment. Biomass levels targeted by management depend on the variability associated with the stock biomass, but would typically minimize risk at intermediate levels associated with maximum yield. Profit-motivated fisheries would reduce targets to increase catches as the variability associated with a stock increases however, and the ability to make a meaningful estimate declines. The framework provides the basis for more extensive risk analyses, and serves as a simple lesson that the consequences of reducing the immediate costs associated with managing a fishery, can come with a concomitant increase in overall risk.

Atlantis analysis: all tiers

Harvest strategies have been applied to many data-rich fisheries, and are now increasingly being applied in data-limited situations. These have been evaluated using simulation frameworks, including Management Strategy Evaluation (MSE), but few studies have considered the full spectrum from data-rich to data-limited strategies, in the context of the risk-cost-catch trade-off. This involves evaluating whether the cost of implementing a harvest strategy, the risk to the resource and catch taken from the resource have been appropriately balanced, given the value of the resource. Harvest strategies implemented for Australian Commonwealth fisheries were placed in eight tiers, ranging from data-rich to data-limited, and their performance evaluated using a MSE based on a full end-to-end ecosystem model. Generally, the risk to the resource increased as fewer data were available, due to biases in the assessments and slow response times to unexpected declines in resource status. The most data-rich tiers maximize discounted catches and profits over a 45-year projection period. However, the opportunity cost response is variable, and shows that the benefit of short-term high catches have to be compensated by resource recovery in the long-term. On average, more data leads to improved management results in terms of risk of being overfished and not reaching a target, but this requires lower initial catches to recover the resources and lower short-term discounted profits.

Atlantis analysis: SESSF and buffers

Several fisheries jurisdictions are aiming to achieve risk equivalency (here defined as the probability of a stock being depleted below a limit reference point or not being maintained at a target reference point) irrespective of the stock assessment method used to provide management advice and the amount of data available. Risk equivalency is implicitly required under the USA Magnuson-Stevens Act, while in Australia it is an explicit component of the Australian Commonwealth Government’s Harvest Strategy Policy. Risk equivalency is well understood, but few fisheries have attempted to implement it. The Australian Southern and Eastern Scalefish and Shark Fishery (SESSF) is the only Australian fishery that has explicitly done so, albeit in a semi-arbitrary manner. Assessments and associated harvest strategies are placed into tiers from data-rich to data-limited. There are also meta-rules that control how much catch limits can change from one year to the next, and buffers by tier to achieve risk equivalency. Here, the SESSF tier system was evaluated in an ecosystem context using Management Strategy Evaluation. Two buffer systems were considered, the current SESSF system and a system inferred from how the Acceptable Biological Catches are set for the USA west coast groundfish fishery. Harvest strategies for all tiers were capable of moving productive stocks so their biomasses lay between the limit and target reference points. The USA buffer system was more conservative than the SESSF system, and achieved the fastest recovery for depleted stocks. The latter system led to slightly lower total catches, but was closest to achieving risk equivalency across the tiers. The USA buffer system led to biomass trajectories most similar to those when the system was managed so that biomass moves as rapidly as possible to its target reference point.

Suggested application

This section provides guiding text and advice on the next steps to apply the risk-cost-catch trade-off to AFMA-managed fisheries that target key commercial species (i.e. not byproduct or bycatch). It therefore relates to aspects of the 2007 Commonwealth Harvest Strategy Policy (HSP) (DAFF, 2007), through developing a set of guiding principles aimed at fisheries managers who have to implement the risk-cost-catch trade-off, while maintaining risk equivalency across tiers.

The terms "risk-cost-catch trade-off" and "risk equivalency" in this section are defined and discussed in relation to the HSP. It also address how a fisheries manager could best approach the goal of maintaining risk equivalency across tier levels for a specific fishery. Some of the controversies of this concept are also discussed.

The section divided into several sub-sections. Firstly, a background is provided to introduce the concept of risk equivalency and the HSP, the risk-cost-catch trade-off and the difference between variance and bias. It then moves to proposing a potential generic tier system (which needs to be socialised through the RAG and MAC system) and some guiding principles on how to address the risk-cost-catch trade-off in terms of risk equivalency. These are then assigned to a series of steps that a Resource Asssesment Goup or managers can use. The final section explains some of the difficulties with the risk-cost-catch trade-off that still may need to be addressed.

A basic philosophy is that a tier system (or an independent harvest strategy) should aim to achieve risk equivalency with respect to setting the TAC/TAE, irrespective of assessment and harvest strategy approach used, to achieve the target reference point and avoid the limit reference point, addressing associated uncertainties as stated in the Policy and Guidelines.

Implications

The risk-cost-catch trade-off is an important theoretical component of fisheries management. However, in most cases, risk is constrained by some biological bottom line such as a Limit Reference Point. This means that cost and catch can only be traded in the risk space where the resource will not become overfished.

In practice, the risk-cost-catch trade-off is complicated by (a) uncertainty (both variance and bias) in the assessment and associated harvest strategy and (b) the amount of social discounting of management costs and catch (profit). One of the largest factors is lost opportunity costs – the cost of not managing the resource well – which are often the hardest cost to calculate. One of the smallest costs is the cost of monitoring, assessing and managing the target species of the fishery.

The results are very species-specific, but generally the more data-limited fisheries and its assessment, the higher the bias (especially) and uncertainty of estimates of sustainable catch. The bias often results in slower response time to a biological conservation concern. Therefore, buffers should be applied by default, unless simulation testing (e.g. based on MSE) has shown otherwise.

Meta-rules that affect how much a TAC changes from one year to the next are complicated in that they can be context-specific, particularly related to where the species is in terms of stock status – if it was over-exploited but recovering, meta-rules slow TAC increases to the benefit of long-term profit and costs, whereas such rules slow the recovery of species that are declining.

The performance of data-limited assessments is highly dependent on their specific formulation, and on the appropriate setting of proxy reference points – and the latter can become a circular problem.

The current and extended tier systems are methods-based, which does not always relate to how well a specific management quantity is estimated. This has contributed to the species-specific results. However, this work has shown a way to operationalise the trade-off.

Keywords

Harvest strategies; data limited; data rich; risk equivalency, risk-cost-catch trade-off; Atlantis; buffers; tier systems

1 Introduction

Management of renewable resources such as fisheries can be complex given the range of species and habitats affected by fishing. This is true even if the focus is only on species that are directly targeted by fisheries. There is often a range of information sources and data quality available by fishery and species, with fisheries ranging from small scale, low value to large scale, high value (Sainsbury, 2005; Dowling et al., 2013). Thus, jurisdictions have had to tailor methods to manage each fishery according to their policies and legislative frameworks (e.g., Smith et al., 2007). Most policy frameworks include target and limit reference points (TRPs and LRPs), with the latter defined in terms of overfishing (a fishing mortality risk) and being overfished (a biomass risk). In addition to the risks management faces, are the costs incurred from managing the risk, as well as the corresponding benefits derived from the fishery in the form of catch: the risk-cost-catch trade-off (Sainsbury, 2005).

Implementing policies and frameworks, however, has resulted in different operational approaches to stock management, which includes data collection, assessment and decision procedures, often collectively called a harvest strategy (HS). As a result, there is usually, either directly or indirectly, some requirement that risk should remain similar regardless of data availability and the method used to manage a stock; i.e. some form of risk equivalency is required (Sainsbury, 2005).

Several jurisdictions have tried to address risk equivalency, most notably the European Union (EU), the USA, and Australia (Dichmont et al., 2015). The EU's Common Fisheries Policy (CFP) is the principal legal mechanism for managing fish stocks in EU waters, and regulates all aspects of fishing within the EU. The CFP has the overall objective of ensuring economically, environmentally and socially sustainable use of fishing and aquaculture resources (European Commission, 2013). In response to the range of data types and assessment methods available, the International Council for the Exploration of the Sea (ICES) has implemented a framework for data-limited stocks (ICES, 2012), which consists of six harvest strategy categories from data-rich (Category 1) to data-limited (Categories 2 to 6).

In the USA federal system, the risk-cost-catch trade-off and its implementation into a tier system for HSs is well defined for managers to achieve the goals of the Magnuson-Stevens Fishery Management and Conservation Act, per the National Standards (USA Doc, 2007). These standards recognise that there is a trade-off between conservation and utilization, but, as written, conservation takes precedence over minimizing impacts on fishing communities. Each region within the USA has a slightly different approach to addressing risk equivalency. For example, the Pacific Fisheries Management Council Scientific and Statistical Committee (PFMC SSC) places each stock into one of three 'categories' (and one of 11 sub-categories) depending on the method and reliability of the assessment. Management of Alaska's 10 crab stocks, on the other hand, which are jointly managed under federal and State jurisdiction, uses a tier system that includes five tiers depending on data availability and the ability to estimate key stock assessment parameters.

Australian Commonwealth (federal) fisheries are managed by the Australian Fisheries Management Authority (AFMA) under the Fisheries Management Act 1991, and in accordance with the Environment Protection and Biodiversity Conservation Act (1999). The Australian Fisheries Harvest Strategy Policy and Guidelines (DAFF, 2007) relate to key commercial species targeted by AFMA-managed fisheries. Eight classes of HSs are applied to most of the Australian Commonwealth-managed fisheries (Dowling et al., 2013). The original system of four categories (called "tiers" in Australia), on which this 8-tier system was based, was implemented in the Southern and Eastern Scalefish and Shark Fishery (SESSF) for several years (Anon, 2014). The TRP for Australian Commonwealth-managed fisheries is

B_{MEY}, the biomass corresponding to Maximum Economic Yield. The Australian Harvest Strategy Policy (Rayns, 2007) allows for the use of proxies for B_{MEY} (specifically, 1.2 × B_{MSY}, where the proxy for B_{MSY} is taken to be $0.4B_0$).

MSE is a simulation approach to explore the effects of alternative management options, including the potential trade-offs among the (pre-agreed and pre-specified) management objectives, taking into account various sources of uncertainty (e.g. uncertainty in the assessment, implementation error). This facilitates the identification of HSs that are robust to uncertainty, and that achieve desired trade-offs among the management objectives. MSE has been applied to several single and multispecies fisheries (Punt, 1992; De la Mare, 1996; Butterworth et al., 1997; Punt and Smith, 1999; Smith et al., 1999; Punt et al., 2002; Dichmont et al., 2006) and to ecosystems (Sainsbury et al., 2000; Fulton et al., 2014). Many of the HSs applied to SESSF stocks have been evaluated individually using MSE to ensure they conform to the limit reference point as defined in the Commonwealth Harvest Policy, and to compare the relative robustness of alternative tiers of assessment (Fay et al., 2009, 2011; Wayte and Klaer, 2010; Little et al., 2011; Klaer et al., 2012). However, the 8-tier framework developed by Dowling et al. (2013) has yet to be evaluated.

This report therefore uses existing data, economic theory and MSE to evaluate six of the tiers from the 8-tier system developed by Dowling et al. (2013) in terms of the risk-cost-catch trade-off. The enhanced tier system (noting that this system has not been implemented but is a description of all existing Commonwealth Harvest Strategies) was evaluated across a range of species types using the whole-of-system model Atlantis (Fulton et al., 2014). We aimed to determine how well a tier system of HSs conforms to the assumption of risk equivalency and to explore the overall risk-cost-catch trade-offs.

2 Objectives

1. Extend AtlantisSE to enable the full suite of Commonwealth fishery types (e.g. data poor) to be simulated.
2. Use this modelling platform to define the risk-cost-catch trade-off between target species at different information and Tier levels.
3. In close consultation with managers and industry, develop guiding context and concepts on the risk-cost-catch trade-off.

3 Methods

This body of work is divided into several components that together provide insight into operationalising the risk-cost-catch trade-off:

1. an international review of tiered systems used that attempt to address a core component of the risk-cost-catch trade-off which is risk equivalency when using data rich to data limited assessment methods, detailed in APPENDIX A: *Do management systems based on tiers of harvest control rules achieve risk equivalency? A comparison of four case studies*

2. a principal component analysis of the existing SESSF tier system, and the expanded potential system developed by Dowling et al. (2013), detailed in APPENDIX B: *Assessing a multilevel tier system: the role and implications of data quality and availability*

3. a first principles approach of developing the risk-cost-catch trade-off using market based economics (monetary economics), detailed in APPENDIX C: *Decision trade-offs for cost-constrained fisheries management*

4. MSEs of the HSs, based on those used by AFMA managed fisheries and Dowling et al. (2013), using the whole of system model of Atlantis as previously applied to the South-east fisheries region (Fulton et al., 2014) – hereafter called AtlantisRCC. APPENDIX D: *From data rich to data-limited harvest strategies – does more data mean better management?*

5. tests of the buffer systems applied in the SESSF and USA groundishery of the tier 1 to 4 harvest strategies as applied in the SESSF in 2015, again using AtlantisRCC. APPENDIX E: *Developing risk equivalent data-rich and data-limited harvest strategies*

6. an approach on how best to operationalise the risk-cost-catch trade-off that forms the basis for further development, if required, for the new and present HSs by AFMA and/or the DA. APPENDIX F: Suggested application

The Appendices contain the detailed methods and results, including the mathematical equations. The main sections therefore a) summarise the methods (“Methods”), b) highlight the key results (“Results”) and c) discussion points (“Discussion” and “Conclusions”). It also draws all the results together into d) “Implications” and e) “Recommendations”.

3.1 Tier systems review

The overarching policies governing fisheries management address the risk-cost-catch trade-off, but few jurisdictions operationalise it directly in their management systems. What all jurisdictions share, explicitly or otherwise, is an aim for risk equivalency (or reducing the risk when data are limited). Risk equivalency means having a common probability of stocks falling below the limit reference point. This can be achieved by adjusting catch limits downwards if these are determined using data limited methods. The extent of downward adjustment (to, for example, target levels of catch or effort) is variously referred to as the buffer or the discount factor. The parameters that determine the extent of downward adjustment will here be referred to as control variables.

This report addresses whether there is some element of risk being equivalent among tiers within tier-based management approaches, which is the expectation for the tier system in Australia (Smith et al., 2014). In Australia the expectation of risk equivalency is implemented so that each tier would include a buffer to reduce catch or effort, with the value of the control variables determining the buffer selected so that the risk of over-exploitation associated with managing under each tier is the same (Smith et al., 2014; Punt et al., 2014a).

Tier systems have only been formally developed and implemented in three jurisdictions globally (Australia's Commonwealth, the USA Federal system, and ICES). Four case studies describe how the trade-offs between risk, management costs, and catch can be addressed are considered. The cases include the well-established tier systems of Australia's SESSF, the USA Alaskan crab fisheries, the USA west coast groundfish fishery, and the recently-developed tier system of the European fisheries for which the ICES provides advice. The way assessment/management tiers in various jurisdictions are established has to be viewed in the appropriate policy context. This review provides the policy frameworks that underlie the tier systems for each case study and then outline how each policy structure has been implemented. Next, the extent to which the tier structures aim to achieve risk equivalency across tiers are evaluated, and finally, conclusions and recommendations are provided, with a focus on Australia, for achieving the expectation of risk equivalency are drawn.

3.2 Principal Component Analysis

The availability and quality of data for the main species, or species groups, under each of Australia's Commonwealth (Federal) fisheries, was scored according to the guidelines in Table 1. The scoring system of 0 to 3 (highest to lowest), was used to score the following seven types of data Table 1.

If CPUE data are available, it follows that so are some effort data. Data categories (6) and (7) of Table 1 ('landed catch' and 'effort') were intended to apply more to fisheries for which only either catch or effort data are available, and where these form the basis for "assessments". Such fisheries include multispecies fisheries where catch is not reported by species, or fisheries for which catch data are considered highly unreliable (such as the Australian Coral Sea Fishery Line, Trawl and Trap sub-fishery) (Dowling et al., 2008).

Table 1. Criteria used to score fishery data quality and availability

Score	0	1	2	3
Fishery-independent survey	Unbiased, low CVs	Unbiased, high CVS	Likely biased as indicators of trend, or poor spatial/temporal coverage	None
CPUE	Targeted Fishery, and standardized	Bycatch/non-targeted fishery and standardized, and/or issues with spatial structure	Available but perhaps not standardized, or poor spatial/temporal/fleet coverage	None
Length-frequency	Representative of the whole fishery	Representative of at least one fleet/part of the fishery	Some data available	None
Catch-at-age	Representative of the whole fishery, ageing error known	Representative of at least one fleet/part of the fishery	Some data available	None
Total catch (including discards)	Whole fishery covered, data reliable and/or observer effort covered > 50% of catch	Whole fishery covered; discard high and variable, and/or some uncertainty in reporting	Only landed catch; qualitative knowledge of bycatch, or high uncertainty in reporting, or poor spatial/temporal/fleet coverage	None
Landed catch	Well covered	Issues with stock identification, or catch uncertainty; discard high and variable	Issues with species identification, poor spatial/temporal/fleet coverage, and/or unreliable reporting	None
Effort	All sectors/fleets/participants covered	Multiple sectors with some not included, or not full coverage across fleets/participants	Poor spatial/temporal/fleet coverage, and/or unreliable reporting	None

Assignment of scores was undertaken by scientific experts for each fishery. All experts received the same explanatory brief, and were aware of the context for the analysis. Thus, scorings were standardised among fisheries to the best extent possible. For cross-validation purposes (and to test consistency of the scorers), the same experts were approached some months later to undertake an identical, repeated round of scorings. Tropical tuna species (Yellowfin Tuna (*Thunnus albacares*), Bigeye Tuna (*Thunnus obesus*), and Albacore (*Thunnus alalunga*) were excluded from the Australian Eastern Tuna and Billfish Fishery (ETBF), because assessment and harvest strategies for these species were no longer being applied at the time of analysis.

The expanded tier definitions (Dowling et al., 2013) provided in Table 2 were used to explore the extent to which quality and availability of the data are consistent with the tier level expectation. Using a principal components analysis (PCA), we examined whether the tier levels of a range of fisheries clustered according to the scoring for the seven data types describing data quality and availability.

Table 2. Harvest strategy tier levels (corresponding to an assessment and/or management framework), based on Dowling et al. (2013), and expanded from the 4-tier level system defined in Smith and Smith (2005) for the SESSF. Increasing tier numbers reflect an assumed increased risk of over-fishing. Note that, currently, no Australian stocks or species are assigned to tier 5, but this tier is included because it represents a level of data availability and an assessment intermediate in quality compared to tiers 4 and 6 (Dowling et al. 2013).

Tier	Tier description
0	Robust (in terms of associated low confidence intervals) assessment of fishing mortality (F) and biomass (B), based on fishery-dependent and -independent data
1	Robust assessment of F and B based on fishery-dependent data only
2	Less robust assessment of F and B, based on fishery-dependent and/or fishery-independent data
3	Empirical estimates of F based on size and/or age data
4	Empirical estimates of (a) trends in relative biomass based on catch-per-unit-effort (CPUE) data (b) within-season changes in relative biomass based on CPUE data (c) availability of relative biomass based on informal fishery-independent surveys
5	Empirical estimates of F based on the spatial distribution of effort relative to the distribution of the species
6	No estimate of biomass or F; management decisions based on fishery-dependent species-specific triggers
7	No estimate of biomass or F; management decisions based on fishery-dependent triggers for groups of species

3.3 Monetary economic analysis

A simple economic analysis was undertaken to test whether basic economic principles can be applied to the risk-cost-catch trade-off. The analysis is based on an objective function for the total expected management cost with several components (Field et al., 2004), including administrative costs, the cost of implementing management actions whether they are needed or not, the cost of being below a limit reference point but not recognizing it, the profit as a function of biomass, and the cost of data collection and assessment. The detailed mathematics are described in Appendix C.

The analyses are undertaken for different values of the underlying stock variability, the effect of increased assessment precision on management costs, and also under a range of weights attached to the cost function. The ratio between the different weights reflects the degree that profitability relative to conservation is emphasized as a management objective.

3.4 Atlantis analysis: all tiers

A full description of the model and methods are provided in Appendix D. An MSE is used to test different harvest strategies from data rich to data limited.

The MSE consists of an operating model and a set of candidate HSs. The success or otherwise of a HS depends on its three components: (i) data collection scheme; (ii) the assessment method; and (iii) the decision (or harvest control) rules to set, in the case of this study, the Total Allowable Catch (TAC). The end-to-end (or "whole of system") ecosystem model, Atlantis (Fulton et al., 2011; henceforth Atlantis-RCC [for "Atlantis-Risk-Cost-Catch"]) forms the basis for the operating model. Atlantis-RCC is based on Atlantis-SE ("Atlantis-South-East") (Fulton et al., 2014), which was originally developed to explore alternative management options for the SESSF. Atlantis-SE was chosen as the basis for Atlantis-RCC because it was developed for the ecosystem within which four of the tiers are applied.

Atlantis-RCC is a 3-D model, with model regions ("boxes") based on the physical and ecological properties of southeast Australia, the distribution of the water bodies, and the geomorphology of the area (summarised in IMCRA, 1998; Butler et al., 2001; Lyne and Hayes, 2005 and Fulton et al., 2007) (Figure 1).

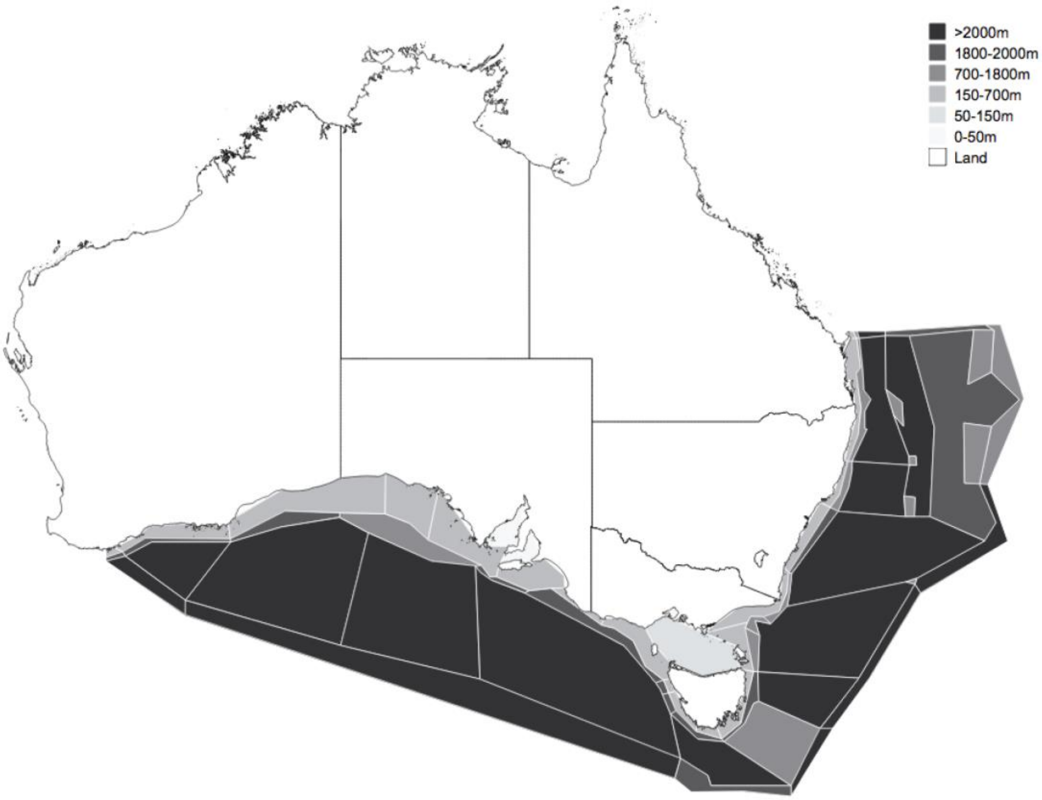


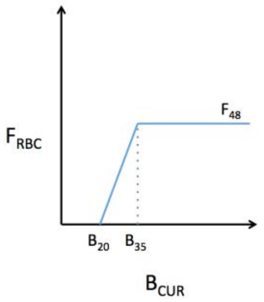
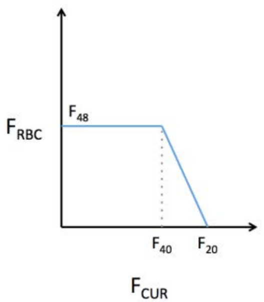
Figure 1. The domain of Atlantis-RCC, which contains 71 boxes with up to 5 vertical layers per box.

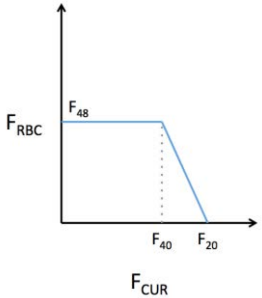
The physical environment for Atlantis-RCC includes ocean currents and water column properties such as temperature and salinity. Vertical and horizontal exchanges between spatial boxes, as well as temperature and salinity, were taken from the data-assimilated version of the global ocean model OFAM (Oke et al., 2005).[1]

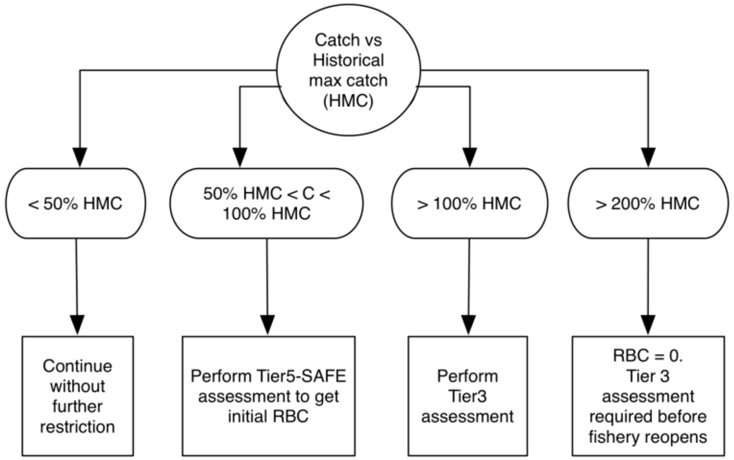
The tiers examined (Table 3) included those currently used in the SESSF (Smith et al., 2014), and updated versions of existing data-poor HSs that have been applied in other Australian federally-managed fisheries (Zhou et al., 2011; Dowling et al., 2008; Dowling et al. in review). The most data rich tier (tier 0) was not used for the SESSF and so it was omitted from the analyses. Similarly, tier 2, which was based on fitting population models that are more uncertain than tier 1 assessments, was omitted because there are no rules currently for how a stock would be assigned to this tier. Two variants of tier 5 were considered as there are many competing ways of empirically estimating fishing mortality in data-limited situations and both a surplus production-based method proposed for the SESSF (Haddon et al., 2015) and the SAFE method of Zhou et al. (2011), which is applied to many Australian bycatch species, were evaluated here.

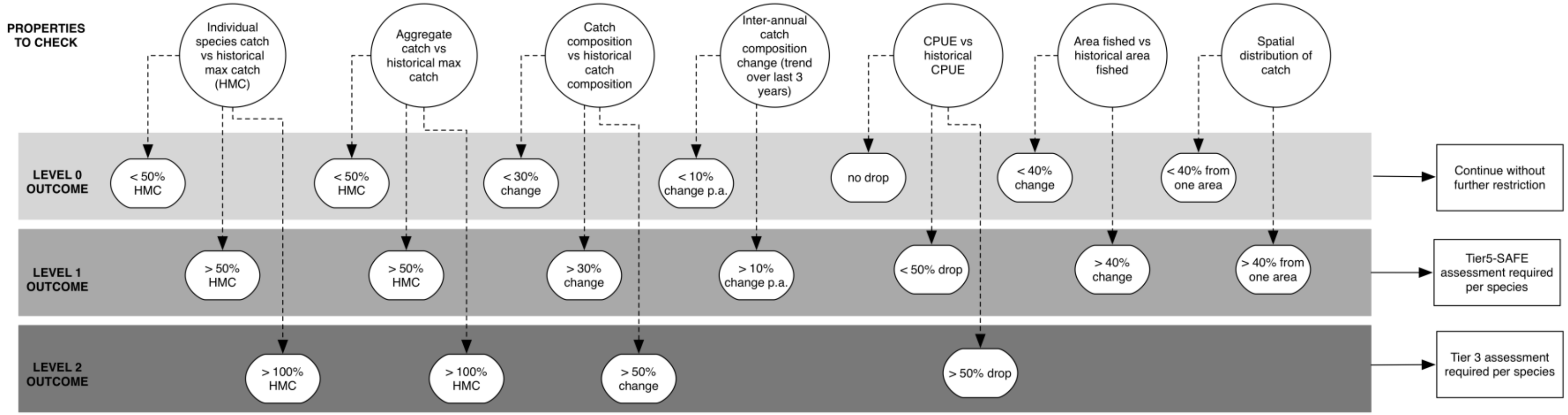
[1] The database used is available at http://www.bom.gov.au/bluelink/ and SPINUP6 from http://www.marine.csiro.au/ofam1/.

Table 3. Tiered assessment types and HCRs for setting Recommended Biological Catches (RBCs). Note currently no tier 2 exists for the SESSF so none was implemented here.

Tier	HS Graph	Rule
1		A full quantitative assessment provides estimates of spawning biomass (*B*) and depletion, which are used in a B_{20}:B_{35}:B_{48} (0.2B_0:0.35B_0:0.48B_0) "broken stick" HS to find the target fishing mortality (F_{TARG}). This F_{TARG} is then applied to the available biomass to calculate the RBC. For the purposes of the paper, the assessment was based on Stock Synthesis (Methot and Wetzel, 2013).
3		Catch curves are used to estimate F_{CUR}, and F_{20}, F_{40} and F_{48} are taken from the relationship between yield and fishing mortality. F_{RBC} is then determined from the HS, and the RBC is calculated using the equation: $RBC = \max\left(\frac{1-e^{-F_{RBC}}}{1-e^{-F_{CUR}}}, 3\right) C_{CUR}$ where C_{CUR} is the current catch. For the purposes of this paper, the approaches outlined in Wayte and Klaer (2010) are used to estimate F_{CUR} and the fishing mortality reference points.
4		The RBC from the Tier 4 HS is given by: $RBC = C_T \max\left(\frac{\overline{CPUE} - CPUE_L}{CPUE_T - CPUE_L}, 0\right)$ where C_T is a catch target, $CPUE_L$ is the limit CPUE, $\overline{CPUE}$ is the average CPUE over the most recent four years, and $CPUE_T$ is the target CPUE. The default catch and CPUE targets were the average for the simulated years 1996-2005 (Little et al., 2011). For some species (flathead, blue grenadier, blue warehou, redfish, pink ling and other shelf

Tier	HS Graph	Rule
		demersal fish) the reference period was set to the more conservative 1986 to 1996 period (this is in line with how the reference period can be tuned per species in reality, see the main text).
5		Tier 5 uses the average length of fish in the catch to determine F_{CUR} based on average length expected as a function of fishing mortality from a yield-per-recruit calculation (Haddon et al., 2015). This F_{CUR} is then used in Tier 3 harvest control rule.
5-SAFE		Tier 5-SAFE (Tier 5S) uses method for calculating fishing mortality for species *i* (F_i) as outlined in Zhou et al. (2011): $$F_i = \frac{q_i^h \cdot q_i^\lambda \cdot (1 - S_i) \sum_t a_{t,i} \cdot E_t}{A_i}$$ where q^h is habitat-dependent encounterability (parameterised using the relative habitat use and overlap defined for the stocks and fleets in the operating model), q^λ is size- and behaviour-dependent selectivity (also parameterised from the effort allocation model), *S* is the discard survival rate, a_t is the area covered in time step *t*, E_t is the effort applied in time step *t* and A_i is the area the species occupies. An aggregate annual *F* is provided by summing over all time steps fished in a year. The reference exploitation rates F_{20}, F_{40} and F_{48} are given by $F_{20} = 1.5\omega M$, $F_{40} = \omega M$ and $F_{48} = 0.8F_{40}$ with ω set to 0.91 for teleosts and 0.43 for chondrichthyans. Natural mortality, *M*, was estimated using the Jensen (1996) relationship: $M = 1.65 / t_{\mathrm{am}}$ where

Tier	HS Graph	Rule
		t_{am} is the age-at-maturity. F_i vs reference F is then used to determine the RBC and any further assessment actions.
6		The tier 6 HS is based on comparing annual total landed catch (C) against various triggers. HMC is the historical maximum annual total catch (Dowling et al., 2008).

Tier	HS Graph	Rule
7		Tier 7 HS uses a variety of triggers to determine changes in the HS. Tier 5-SAFE is the level 1 choice as this method was designed for data-poor species. Tier 3 was chosen as the level 2 choice because the data required for tier 3 are likely to be available (Dowling et al., 2008).

The tiers provide recommended biological catches (RBCs), which may be modified through, for example, the use of meta-rules, to determine TACs.

The meta-rules considered in this analysis were: (a) TACs may not change by more than 50% from one year to the next, and (b) the TAC is unchanged if the proposed change in TAC from one year to the next is 10% or less. One set of simulation experiments was run with the meta-rules active and one set with them disabled.

A sampling model was used to generate the following fishery-dependent data, for each stock and Atlantis region: (a) catch length- and age-composition data; (b) catch-per-unit-effort data (by vessel size-class and fishery sector); (c) landings data (and catch species composition) by vessel size-class and fishery sector, and (d) discard data. The data generation approach allowed for ageing error, measurement error, variation in catchability, and error when measuring discards, with error levels that were stock-specific (see Appendix D for further details). As stated above, since few fisheries in Australia have long-term independent survey data, and tier 0 was not being evaluated, no fishery-independent data were generated.

Atlantis-RCC represented all the major functional groups and species of fisheries or conservation interest in the southeast Australian ecosystem, including those within the SESSF. It was not feasible to explore the risk-cost-catch tradeoff for all these species and groups. Instead, a sub-set of ('treatment') species was selected for consideration in the MSE that was representative of a range of life histories.

The steps undertaken in each MSE simulation are given in Figure 2.

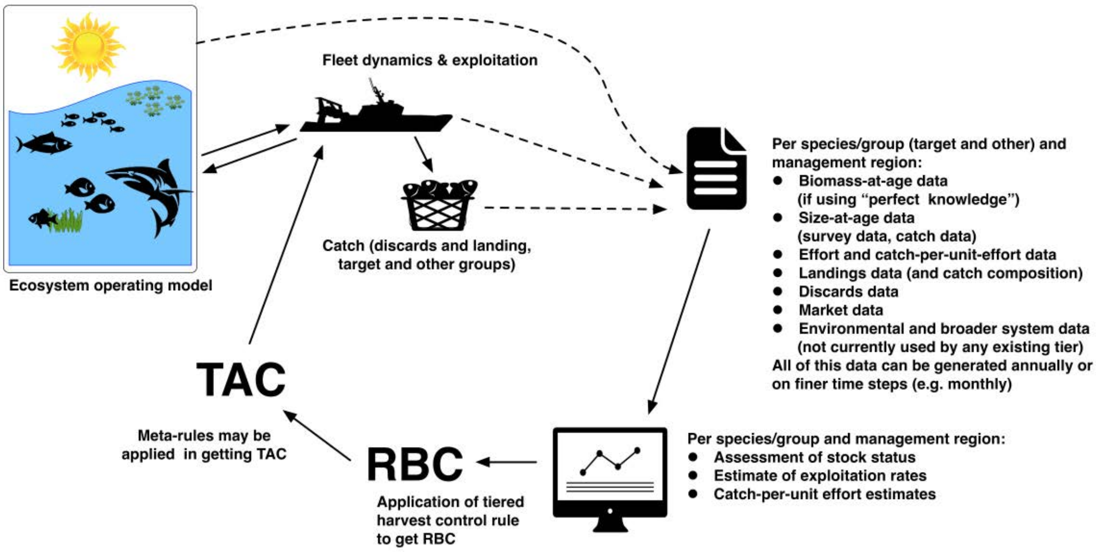


Figure 2. Schematic diagram of steps taken in each MSE simulation. TAC is the Total Allowable Catch and RBC is the Recommended Biological Catch.

Each simulation was 70 years (1980-2005, representing the historical period, and a 2005-2050 projection period). Twenty replicates were undertaken for each treatment species (i.e. those species assessed using one of the tiered HSs) for each tier. Computational speed precluded a larger set of replicates (20 replicates proved sufficient as the ecological components of Atlantis are deterministic and so the model only includes limited stochasticity; additional projections showed that increased numbers of simulations did not alter the results materially). The random deviates for effort allocation and observation error were the same for all simulations to ensure that the results of the projections were maximally comparable (i.e. the results are analogous to paired tests rather than independent tests and were compared in this manner). However, the small number of replicates means that estimates of intervals will be fairly imprecise and are hence considered primarily in a relative sense, i.e. between alternative HSs rather than as accurate precisions of actual intervals.

The actual TACs were applied between 1980 and 2005, although with implementation error due to the socioeconomic effort allocation model used in Atlantis-RCC. Once the projection began in 2005, the treatment species had their TACs set using the tier under consideration in that simulation, with the TACs for all other species set to the 2005 level, i.e. only one species was actively managed within a scenario using a single consistent (through time) HS.

It is not straightforward to handle non-treatment species when conducting a realistic test of the management system for a multispecies fishery. In theory, the fishing mortality rates for each species leading to a desired system state could be found and applied to the non-treatment species, but that is a non-trivial exercise for a complex food web and may still see some species fail to achieve the mandated biomass target level. Choosing any other fishing mortality level is as prone to driving over/underfishing of companion species as what was undertaken in the study. More importantly, while it may initially seem that the best test of the tiers would be to make treatment effects the only effect considered, such isolation isn't possible in a multispecies fishery, and the central reason for conducting an MSE was to test the strategies for robustness given the complexities inherent in the management and fishery system. It is important to know how robust the management system is to those conditions given the control rules are to achieve recovery and sustainability within the context of the multispecies fishery. Consequently, the treatment species were selected as they not only represented a range of life history types, but they also represented species with a range of influences on effort dynamics – including key target species (e.g. tiger flathead or blue grenadier), secondary species (e.g. blue warehou) and some bycatch species (e.g. gulper shark). Thus, we feel across the entire set of species tested the HS were tested in a realistic way for the type of multispecies fishery they will be used for in reality.

In total the HS simulations equated to 14 species × eight tiers × two meta-rules (with and without) = 224 scenarios. Two additional analyses were undertaken to create reference points or trajectories to facilitate interpretation of the output of the HS simulations:

1. No fishing on any group (to allow for the calculation of unfished biomass, B_0, by group).
2. Reference HS: Catches were selected using a "bang-bang" HS where all biomass over the target level is removed (Deroba and Bence, 2008) and there is perfect information about stock size. This HS eliminated targeted fishing for a species for N_1 years if $B < 0.48B_0$, and allowed for large catches for N_2 years if $B > 0.48B_0$. N_1 and N_2 were selected for each species iteratively: an analytical determination was not possible because the use of the dynamic effort allocation model led to implementation error. Elimination of targeted fishing was used because a complete reduction of fishing-related mortality on a species was not possible, owing to incidental catches. In addition, the effort allocation sub-model in Atlantis allows for non-compliance and a memory of past catch performance. Consequently, some catch began once stock rebuilding started and biomass exceeded approximately 0.25-0.3B_0. It follows that recovery for the reference HS is consequently (slightly) slowed compared to the true optimal catch. The discounted catches from the HS provided a reference given each species started from a different biomass relative to 0.4B_0.

The analyses involved assessing and managing stocks separately (see Table 3 and S1). However, the results by stock were combined into results by species for the purposes of summarizing performance. The projections assumed that assessments, and hence changes to RBCs and TACs, occurred annually. Twelve performance metrics were considered, classed into four broad categories: risk, economic, catch, and stock assessment model performance. Details of how each metric is calculated are given in Table 4 (which also provides a code for each performance measure to aid in reporting). Unless stated otherwise, the performance metrics were averages over simulations for each species and medians (with interquartile ranges) over species.

The three risk performance metrics are: (R1) the probability of being below the LRP (20% of the unfished spawning biomass, $0.2B_0$) during the last 30 years of the projection period, (R2) the probability of being below the TRP ($0.48B_0$) during the last 30 years of the projection period, and the time taken to reach a reference point threshold – (R3a) either to recover to $0.2B_0$, if overfished (i.e. starting the projection period with $B < 0.2B_0$), or (R3b) to drop to $0.48B_0$ if not fully exploited initially (i.e. starting the projection period with $B > 0.48B_0$). The choice of the last 30 years for the R1 and R2 performance metrics was selected to provide an opportunity for the stocks depleted at the end of the historical period to recover before comparative risk performance was assessed.

The four economic performance metrics are: (C4) opportunity costs, (C5) harvesting costs related to catching and handling the product (based on the fuel and market price models of Fulton et al. (2007)) and (C6) the cost of running the tier in terms of, for example, data collection and the assessment (using data from Dowling et al. (2013) and Tuck pers. commn). These metrics and estimated revenue were used to calculate the profitability performance metric (C5) and for the whole 45-year period as discounted values (Fulton et al., 2007). The discount rate was assumed to be 0.05, because this is similar to the discount rate often used in bio-economic models in Australia (e.g., Punt et al., 2011).

The discounted catch over the projection period relative to that expected under the Reference HS was also calculated as a performance metric (H8). Our calculation for discounted catch does not include the oft-used final term, set to the discounted sum of the catch in the last year for an infinite number of years. This is because the populations do not all reach equilibrium by the end of the projection period. The final harvest metric was catch variability through time (which can be important to industry economically and to markets who desire reliable and stable supplies).

Four assessment performance metrics were also computed. The first (A10) is the median (with inter-quartile ranges) estimated fishing mortality F_{year}, as determined by the tier's assessment (for tiers 1 to 5-SAFE), relative to the F_{MEY} (calculated from the operating model). The fishing mortality proxy for tier 4 was inferred by assuming the CPUE relative to the target CPUE is a proxy for F/F_{MEY}. This performance metric evaluates the direction in which the HS is likely to change fishing mortality; upwards if $F/F_{\text{MEY}} < 1$ and vice versa. The second assessment performance metric (A11) is the TACs set under the higher (more data-poor) tiers relative to that set by tier 1. Since it is not possible to compare the TACs between tiers over the full time period, as the biomass in each future depends on the past TACs, this metric was calculated only using the TAC in the first projection year, as this will be based on assessments of the same relative population biomass (and data) irrespective of the tier. An alternative metric was to use the actual catch rather than the TAC. The third assessment performance metric is the frequency that a more data-rich assessment is triggered in the two hierarchical HS (tiers 6 and 7) (A12). This metric is a count of the number of times the hierarchy's level 1 (SAFE) and level 2 (tier 3) triggers are activated. The final assessment performance metric (A13) is the response time, which is the number of years before the HS reacts to a change in biomass in the operating model where $TAC_{y}^{true} < 0.9TAC_{y-1}^{true}$. The true TAC is based on a tier 1 assessment with perfect information.

Table 4. Performance metrics. Note that R3, all the cost metrics, relative discounted catch, number of times hierarchical tiers are triggered and response time are calculated across all years of the projection period; whereas the other indices are calculated on shorter periods.

Code	Metric	How calculated	Notes
R1	Probability $B < B_{20}$	$\underset{s}{average}\left(N_{B<B_{20,s}}/30\right)$	$N_{B<B_{20},s}$ is the number of years for which spawning biomass is less than $0.2B_0$ during the last 30 years of the projection period for simulation *s*.
R2	Probability $B < B_{48}$	$\underset{s}{average}\left(N_{B<B_{48,s}}/30\right)$	$N_{B<B_{48},s}$ is the number of years for which spawning biomass is less than $0.48B_0$ during the last 30 years of the projection period for simulation *s*.
R3	Time to threshold	$\underset{s}{median}\left(T_{s,bx}\right)$	$T_{s,bx}$ is the median over simulations of the time taken to reach the threshold b_x for the first time during the projection period*.
C4	Opportunity costs	$\underset{s}{average}\left(\sum_{y} p_{y,s}\left(C_{b,y,s}-C_{y,s}\right)e^{-(y-2006)\delta}\right)$	*p* is the price per unit catch in year *y*, δ is the economic discount rate (0.05), $C_{y,s}$ is the catch in year *y* of simulation *s* and $C_{b,y}$ is the catch in the same year under the bang-bang control rule.
C5	Harvesting costs	$\underset{s}{average}\left(\sum_{y}\left(H_{cap,y,s}+H_{fix,y,s}+H_{ul,y,s}C_{y,s}+\left(H_{fu,y,s}+H_{g,y,s}\right)E_{y,s}+H_{fu,y,s}Z_{y,s}\right)e^{-(y-2006)\delta}\right)$	$H_{x,y,s}$ are the various operating cost components at time *y* in simulation *s*. H_{cap} are capital costs; H_{fix} are fixed costs; H_{ul} are unloading costs per unit catch; H_{fu} are fuel costs per unit of effort (or steaming time); H_g are gear-associated costs per unit effort. $C_{y,s}$ and $E_{y,s}$ are the catch and effort during year *y* of simulation *s*, $Z_{y,s}$ is steaming time.
C6	Management costs	$\underset{s}{average}\left(\sum_{y}\sum_{n} M_{n,y,s}e^{-(y-2006)\delta}\right)$	*M* are the various types of management costs, (*n* is the type of cost being assessment, administration, compliance associated with an assessment of tier *x*) for treatment species i.e. for those being assessed.
C7	Short-term discounted profits	$\underset{s}{average}\left(\sum_{y}\left(p_{y,s}C_{y,s}-K_{T,y,s}\right)e^{-(y-2006)\delta}\right)$	$K_{T,y,s}$ is the total costs for the fishery (harvesting, quota-related and management licence costs) in year *y* of simulation *s*; *p* is the price per unit catch

			and C is catch. δ is the economic discount rate (0.05).
H8	Relative Discounted Catch	$\frac{\sum_s \sum_y C_{y,s} e^{-(y-2006)\delta}}{\sum_s \sum_y C_{b,y,s} e^{-(y-2006)\delta}}$	$C_{y,s}$ is the catch in projection year y of simulation s, $C_{b,y,s}$ is the catch in projection year y of simulation s under the bang-bang control rule, and δ is the economic discount rate.
H9	Catch variability	$\underset{s}{average}\left(\frac{\mathrm{var}(C_s)}{\bar{C}_s}\right)$	var(C_s) is the variance in annual catch across the projection period in simulation s, and the denominator is the mean annual catch across the projection period in simulation s.
A10	F/F_{MEY}	$\underset{s}{average}\left(\frac{F_s}{F^o_{MEY,s}}\right)$	F is the HS estimate of fishing mortality (or its proxy) in the first year of the projection period for simulation s, and F^o_{MEY} is the fishing mortality rate for achieving 0.48B_0 (as defined by the operating model, *op*) in simulation s. This metric is based on first year of the simulation (see main text) and quantifies whether the HS believes there is a need for an increase in harvest rate – because the F is less than the target value
A11	Relative TAC bias	$\underset{s}{average}\left(\frac{TAC_{x,s}}{TAC_{1,s}}\right)$	$TAC_{x,s}$ is the TAC set under tier x in the first year of the projection in simulation s; and $TAC_{1,s}$ is the TAC set under tier 1. TAC is also replaced with actual catch as an alternative performance metric.
A12	Number times SAFE or Tier 3 triggered for tiers 6 and 7	$\underset{s}{average}\left(L_{x,s}\right)$	$L_{x,s}$ is the total number of trigger events for trigger level x over projection period in simulation s. This is also calculated for actual catches instead of TAC.
A13	Response time	$\underset{s}{average}\left(\underset{i}{average}\left(T_{R,i,s}\right)\right)$	Annually, the true state of the operating model was assessed with perfect knowledge and it as recorded if the biomass change was sufficient to require a >10% change in the TAC under tier 1 if

		there was perfect information. For each such event i in simulation s the response time for the HS $T_{R,i,s}$ was calculated as the time in years from when the event was recorded to when the HS changed the TAC in the correct direction.

3.5 Atlantis analysis: SESSF and tier buffers

This analysis used the same MSE as decsibed above but concentrated on the SESSF fishery harvest strategies. Here, the issue is to investigate and compare the role of buffers and meta-rules in a harvest strategy.

In the SESSF, the Total Allowable Commercial Catch (TACC) is calculated from the RBC by applying meta-rules and risk equivalency buffers. These “meta-rules” are used for all SESSF tiers, to prevent TACCs from changing by more than 50% from one year to the next. The SESSF also uses buffers to reduce the RBC for tiers 3 and 4, to attempt to achieve risk equivalency between these tiers and the data-rich tier 1. A tier 1 RBC is usually the output of an integrated stock assessment, typically Stock Synthesis (Methot and Wetzel, 2013). The buffers applied to all stocks managed in the SESSF are equal to 1.0 for tier 1 (no buffer), 0.95 for tier 3 and 0.85 for tier 4 (Stobutzki et al., 2001). These were set semi-arbitrarily by the SESSF with no formal evaluation.

The groundfish fishery off the USA west coast calculates a buffer between the Overfishing Level and the Acceptable Biological Catch, accounting for assessment uncertainty. The basis for the buffer for data–rich stocks is given in Ralston et al. (2011). Briefly, Ralston et al (2011) calculated the coefficient of variation (CV) of the spawning biomass estimates from a meta-analysis of groundfish stock assessments for the US west coast, and this CV is used in combination with a stated tolerance for the risk of fishing mortality exceeding the target levels to determine the buffer (mathematically decsribed further in Appendix E.

Ralston et al. (2011) explored three methods for calculating the CV, but found that the most effective approach was to pool residuals from all species across all available assessments (assuming equal weighting).

We computed a buffer by applying Ralston et al.’s (2011) pooled residual method to data-rich (US category 1) stocks. The final buffers for the SESSF and applying the USA system is provided in Table 5.

Table 5. Buffer system by species using the (inferred) US and Australian methods. The buffers for the USA system are based on percentiles of 0.4 for all tiers. Note that many species are split into east and west stocks for assessment purposes in the SESSF and that these stocks are list separately here.

Stock	Actual SESSF Tier	Assessment variance	USA system buffers Tier 1	Tier 3	Tier 4	Australian system buffers Tier 1	Tier 3	Tier 4
Blue Grenadier, *Macruronus novaezelandiae*[1]	1	-	0.920	0.870	0.820	1	0.95	0.85
Cascade Orange Roughy, *Hoplostethus atlanticus*	1	0.26	0.930	0.860	0.800	1	0.95	0.85
Deepwater Flathead, *Neoplatycephalus conatus*	1	0.25	0.940	0.870	0.790	1	0.95	0.85
Tiger Flathead, *Neoplatycephalus richardsoni*	1	0.27	0.925	0.880	0.830	1	0.95	0.85
Eastern Gemfish, *Rexea solandri*	1	0.24	0.940	0.890	0.800	1	0.95	0.85
Gummy Shark, *Mustelus antarcticus*[*]	1	-	0.920	0.870	0.820	1	0.95	0.85
Eastern Jackass Morwong, *Nemadactylus macropterus*	1	0.10	0.970	0.950	0.900	1	0.95	0.85
East Pink Ling, *Genypterus blacodes*	1	0.14	0.965	0.930	0.860	1	0.95	0.85
West Pink Ling, *Genypterus blacodes*	1	0.38	0.900	0.820	0.730	1	0.95	0.85
(School) Whiting, *Sillago* spp.	1	0.30	0.915	0.850	0.770	1	0.95	0.85
Bight Redfish, *Centroberyx gerrardi*	1	0.22	0.920	0.870	0.820	1	0.95	0.85
Eastern Blue Warehou, *Seriolella brama*	4	0.13	0.965	0.935	0.875	1	0.95	0.85
Western Blue Warehou, *Seriolella brama*	4	0.16	0.960	0.920	0.850	1	0.95	0.85
Blue-eye Trevalla, *Hyperoglyphe antarctica*[*]	4	-	0.920	0.870	0.820	1	0.95	0.85
Demersal Sharks[*]	5[†]	-	0.920	0.870	0.820	1	0.95	0.85
Gulper Shark, *Centrophorus* spp.[*]	5[†]	-	0.920	0.870	0.820	1	0.95	0.85
Generic demersal fish[*]	5[†]	-	0.920	0.870	0.820	1	0.95	0.85
Average		0.22	0.92	0.87	0.82	1	0.95	0.85

[1]: Insufficient assessments available to apply the Ralston et al. (2011) method, so the overall average outcome was used (as recommended by Ralston et al. (2011)).

Analyses involved assessing and managing stocks of a given species separately. However, for brevity's sake, results are summarised by combining results for all stocks of each species. The projections assume that assessments and hence changes to RBCs and TACCs occur annually.

Nine performance metrics were considered (some from Table 4 and a few in addition in

Table 6): (a) probability of being below the limit reference point (i.e. 20% of the unfished spawning biomass, $0.2B_0$) during the final 30 years of the 45-year projection period (R1), (b) probability of being below the target reference point ($0.48B_0$) during the last 30 years of the projection period (R2), (c) discounted catch relative to that expected under a reference HCR across the entire projection period (H8), (d) inter-annual variation in catch (across the entire projection period) (new1), (e) average spawning biomass over the last 30 years of the projection period relative to the average spawning biomass under the reference HCR during the same period, and (new2) (f) time taken to recover to $0.2B_0$, if overfished, or time taken to achieve $0.48B_0$ if not fully exploited initially (R3). There are four "time to threshold" indices: the time to increase to $0.2B_0$, the time to increase to $0.48B_0$, the time to decrease to $0.2B_0$, and the time to decrease to $0.48B_0$. Although only a subset of these indices will be meaningful for any one species groups – for example the time to decrease is meaningless if the biomass is less than $0.2B_0$ at the start of a simulation. However, the four indices are reported so that it is possible to determine if (a) an overfished species recovers to $0.2B_0$ (or beyond), (b) a stock that is not fully exploited has its biomass drop to $0.48B_0$ (or lower), and (c) a stock that is initially in the range $0.2B_0 < B < 0.48B_0$ remains in that range.Two new performance metrics were added, being new1 and new 2 below.

Table 6: New performance metrics used for the SESSF analysis

Metric	How calc	Notes
Annual Absolute Variation in Catch (new1)	$\underset{s}{average}\ \frac{1}{45}\sum_{y}\frac{\lvert C_{y,s}-C_{y-1,s}\rvert}{C_{y-1,s}}$	Calculated over the entire projection period.
Relative biomass (new2)	$\underset{s}{average}\left(\sum_{y=2021}^{2050}B_{y,s}\Big/\sum_{y=2021}^{2050}B_{b,y,s}\right)$	$B_{y,s}$ is the spawning biomass in year *y* of simulation *s*, and $B_{b,y,s}$ is the spawning biomass in year *y* of simulation *s* under the bang-bang HCR. Calculated for the final 30 years of the simulations.

Simulations were run to determine RBCs for each of the 14 species, for each of the tier 1, 3 and 4 harvest strategies (irrespective of their actual designated tier), for six options related to meta-rules and buffers. Each of the species was considered individually, in that the treatment species had their TACs set using the tier-meta-rule-buffer combinations, with the management rules and TACs for the remaining species were kept at their 2005 levels (i.e., only one species at a time was actively managed; simulations exploring the outcomes when multiple species were simultaneously actively managed will be reported elsewhere).

The six management options considered were: (a) no meta-rules or buffers (NM-NB), (b) with meta-rules, but no buffers (M-NB), (c) with meta-rules and the Australian buffers (M-AUB), (e) no meta-rules and the Australian buffers (NM-AUB), (e) with meta-rules and the inferred USA buffers (M-USB), and (f) no meta-rules and the inferred USA buffers (NM-USB).

Twenty simulations were run for each scenario and species, with each simulation including a 45-year projection period. The Atlantis ecosystem sub-model is deterministic. However, multiple simulations were still required as parts of the effort allocation model are stochastic, and there was additional stochasticity among the projections as a consequence of the sampling error associated with the data used to apply harvest strategies.

4 Results

4.1 Tier systems review

Several jurisdictions have developed hierarchical tier systems that categorise stocks based on, for example, the data available for assessment purposes and/or the extent to which the quantities on which management advice is based can be estimated. Four case studies (Australia's Southern and Eastern Scalefish and Shark Fishery, the USA west coast groundfish fishery, the USA Alaskan crab fishery, and EU fisheries) are used to contrast the types of hierarchical tier systems available, and to assess the extent to which each system constrains risk to be equivalent among the tiers (termed risk equivalency).

4.1.1 Policy structure of alternative tier structures

USA federal system

In the USA federal system, the risk-cost-catch trade-off and the implementation of this trade-off into a tier system for assessments and harvest control rules (HCRs) is well defined for managers, in order to achieve the goals of the Magnuson-Stevens Fishery Management and Conservation Act, as reflected by its National Standards (USA Doc, 2007). These standards recognise that there is a trade-off between conservation and utilization, but, as written, conservation takes precedence over minimizing impacts on fishing communities. The standards also state that management measures must be based on best available scientific information (National Standard 2) and this is achieved by having a well-structured and transparent peer-review system. In particular, recommendations for management decisions for federally-managed species in the USA are made by the Regional Fishery Management Councils based on scientific advice reviewed and approved by their specific Scientific and Statistical Committee (SSC).

Each Council has a slightly different structure for the development and review of stock assessments and hence the provision of scientific management advice. Nevertheless, the advice and decision-making frameworks are broadly similar among regions of the USA. The key scientific inputs into the decision making process for each stock are ideally:

- The Overfishing Level (OFL) – the catch that corresponds to a fishing mortality from a Maximum Sustainable Yield (F_{MSY}) control rule – for several regions around the US, F_{MSY} (or a proxy thereof) is the target fishing mortality rate.
- The Acceptable Biological Catch (ABC) – a catch that is less than or equal to the OFL to account for scientific uncertainty. The difference between the OFL and ABC for each stock is referred to as the 'buffer'.
- Whether the stock is subject to overfishing – variously defined as the catch exceeding the OFL or the fishing mortality rate exceeding that corresponding to the control rule for the OFL, F_{OFL}.
- Whether the stock is in an overfished state – defined as stock size being below the Minimum Stock Size Threshold (MSST). MSST is defined to be between B_{MSY} (the stock biomass corresponding to MSY) and $0.5B_{\mathrm{MSY}}$.

Councils use this information, along with other inputs, such as comments from the public on the impacts of various catch levels, to select an Annual Catch Level, ACL (which cannot exceed the ABC) and the optimum yield. The difference between ACL and the ABC can reflect a variety of policy considerations such as social, economic and other ecological factors, as well as increasing the rate at which overfished stocks rebuild to the target biomass and the uncertainty associated with implementing catch limits. The difference between the OFL and ABC is based on uncertainty that then forms the basis of the tier system.

Australian Commonwealth system

Australian Commonwealth fisheries are managed by the Australian Fisheries Management Authority (AFMA) under the Fisheries Management Act 1991 and in accordance with the Environment Protection and Biodiversity Conservation Act 1999. The Fisheries Harvest Strategy Policy and Guidelines (DAFF, 2007) relates to key commercial species targeted by AFMA-managed fisheries. The policy defines target and limit reference points, and, when these cannot be estimated, their proxy values. The target reference point for AFMA-managed fisheries is the Maximum Economic Yield (MEY), which is unusual because many other fisheries' jurisdictions use B_{MSY} as a target. No targeted fishing for a species is allowed when its stock size is at or below a limit reference point, which is usually set at 20% of the unexploited stock size. Risk is defined in the Harvest Policy, mainly for the limit reference point, in that management should "ensure that the stock stays above the limit biomass level at least 90% of the time". There is a tiered system to address data and assessment types and link these to risk equivalency among tiers.

European (and ICES) system

The European Union's (EU) Common Fisheries Policy (CFP) is the principal legal mechanism for managing fish stocks in EU waters, and regulates all aspects of fishing within the EU. The CFP has the overall objective of ensuring economically, environmentally and socially sustainable use of fishing and aquaculture resources (European Commission, 2013). The CFP was first implemented in 1983, and has undergone a number of reforms. The latest of these (which came into effect in January 2014) specifies that the management of all species that are subject to quotas should be on the basis of MSY by 2020 at the latest. It also introduces a phased discard ban to fully cover all EU fisheries by 2019, allows regionalisation of fisheries and management so as to better suit local conditions, and addresses fleet over-capacity by insisting that capacity matches fishing opportunities for member states. ICES is the intergovernmental body that develops science and advice on marine ecosystems and associated fisheries. In the context of this report, ICES co-ordinates and conducts stock assessments and provides stock status advice to the EU and other ICES member countries based on a Memorandum of Understanding.

ICES currently provides advice based on an MSY policy for those stocks without agreed, precautionary management plans but for which there is a basis for MSY advice, using a HCR that, in principle, should conform to the "hockey stick" (aka. "slope") rule (i.e. fishing mortality is a linear function of biomass, until a trigger level where fishing mortality remains constant independent of the biomass) with the trigger at F_{MSY} and B_{trigger} (the latter representing the lower bound of spawning biomass fluctuations around B_{MSY}). That is, the aim is ultimately to harvest these stocks at F_{MSY} (ICES, 2014a). Where there is an agreed, precautionary management plan, ICES will base its advice on such a plan, and where there is no basis for MSY advice or no agreed, precautionary management plan, ICES advice will be based on precautionary principles (ICES, 2014a). The tier system used by ICES (described later) is applied within this context, and implicitly attempts to use different harvest strategies (data type, assessment method and decision rules) to obtain the expectation of risk equivalency among tiers.

4.1.2 Implementation of the tier systems

United States west coast groundfish fishery

Over 90 species are included in the Pacific Coast Fishery Management Plan, including rockfishes, roundfishes, sharks and skates (PFMC, 2014a). The Pacific Fisheries Management Council Science Scientific and Statistical Committee (PFMC SSC) places each stock into one of three 'categories' (and one of 11 sub-categories) (Table 7) depending on the method and reliability of the assessment. Specifically, stocks assessed using data-rich methods may be considered data-moderate if, for example, the assessment is dated, the assessment did not estimate annual deviations in recruitment about the stock recruitment relationship, or the assessment results were sensitive to plausible changes to the assumptions. Of 144 stocks for which OFLs are available, 21 stocks are currently in category 1 (the most data-rich), 29 stocks are in category 2, and the remaining stocks (94) are in category 3 (the most data-limited) (PFMC, 2014a). Not all stocks are managed as individual species; several stocks are instead managed as complexes under the assumption the stocks within a complex have roughly the same productivity and vulnerability (PFMC, 2014a). Figure 3 shows the generic HCRs for setting OFLs and ABCs for USA west coast groundfish (noting that B_{MSY} cannot be estimated reliability for any west coast groundfish stock (PFMC, 2014b) so proxies are used instead).

Table 7. Tier assignments for USA west coast groundfish assessments (reproduced from PFMC [2014b]). Note: the USA uses the term "category" and the numerical order is reversed (but not the alphabetical sub-tiers order) compared to the Australian and ICES systems. A consistent terminology is used in this paper for ease of comparison.

Tier Number		Description	Example stock
Tier 1: Data rich	aa	Reliable compositional (age and/or size) data sufficient to resolve year-class strength and growth characteristics. Only fishery-dependent trend information available. Age/size structured assessment model.	Black rockfish (*Sebastes melanops;* Sebastidae)
	b	As in 1a, but trend information also available from surveys. Age/size structured assessment model.	Canary rockfish (*Sebastes pinniger;* Sebastidae)
	c	Age/size structured assessment model with reliable estimation of the stock-recruit relationship.	None
Tier 2: Data moderate	a	Natural mortality multiplied by survey biomass.	None
	b	Historical catches, fishery-dependent trend information only. An aggregate population model is fit to the available information.	None
	c	Historical catches, survey trend information, or at least one absolute abundance estimate. An aggregate population model is fit to the available information.	None
	d	Full age-structured assessment, but the results are substantially more uncertain than assessments used in the calculation of the σ used to compute the buffer for category 1 stocks (Ralston et al., 2001). Reasons for placing a stocks in this category include that assessment results are very sensitive to model and data assumptions, or that the assessment has not been updated for many years.	Shortbelly rockfish(*Sebastes jordani;* Sebastidae)
Tier 3: Data poor	a	No reliable catch history. No basis for establishing OFL.	Black and yellow rockfish (*Sebastes chrysomelas;* Sebastidae)

b	Reliable catch estimates only for recent years. OFL is the average catch during a period when the stock is considered to be stable and close to B_{MSY} equilibrium on the basis of expert judgment.	Pacific Grenadier (*Coryphaenoides acrolepis;* Macrouridae)
c	Reliable aggregate catches during period of fishery development and approximate values for natural mortality. Default analytical approach Depletion-Corrected Average Catch.	Cowcod south of Pt Conception (*Sebastes levis;* Sebastidae)
d	Reliable annual historical catches and approximate values for natural mortality and age at 50% maturity. Default analytical approach Depletion-based Stock-Reduction Analysis.	Kelp rockfish (*Sebastes atroviren;* Sebastidae)

The buffer between the OFL and the ABC reflects the extent of scientific uncertainty, which differs among categories. The buffer is a function of the extent of uncertainty and the risk tolerance given uncertainty (i.e., $p(ABC(\hat{OFL}) > OFL) = P^*$; Prager and Shertzer, 2010; Punt et al., 2012; Shertzer et al., 2008). The buffer is implemented for the west coast groundfish fishery by the PFMC SSC quantifying uncertainty in terms of coefficient of variation for the OFL (sigma) and the Council selecting the percentile of the resulting probability distribution at which to set the ABC (P^*). P^* is therefore the probability of overfishing occurring. For data-rich (category 1) stocks sigma is 0.36, unless the coefficient of variation of the estimated most-recent year biomass from the assessment exceeds this (PFMC, 2014a). This value is based on a meta-analysis of between-assessment variation in biomass estimates (Ralston et al., 2011). By contrast, the values for sigma for data-moderate and data-poor stocks are 0.72 and 1.44 respectively. The value of P^* is selected by the Council, which has agreed that it will not exceed 0.45 for any stock (legally, it must be 0.5 or less). The default choice for P^* is 0.45 for category 1 stocks and 0.4 for category 2 and 3 stocks. Therefore, for the USA west coast groundfish fishery, sigma and P^* are the control variables.

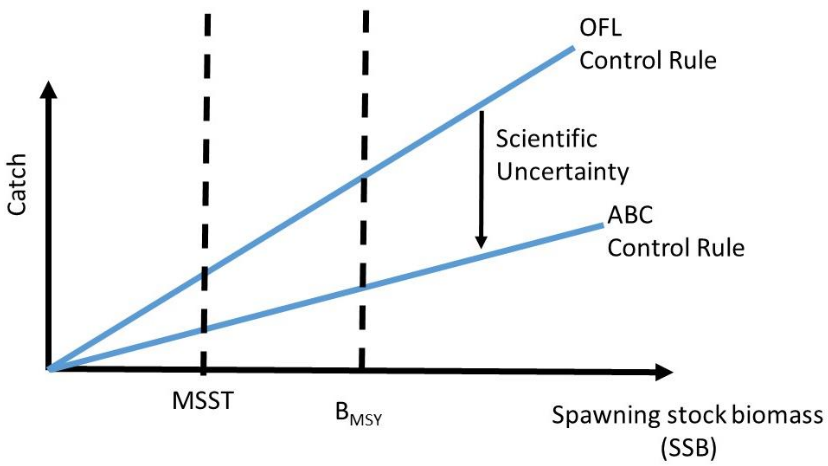


Figure 3. Harvest control rules for the USA west coast groundfish fishery. Terms are: Acceptable Biological Catch (ABC), Biomass (B) at Maximum Sustainable Yield (MSY), Current (CUR), Fishing mortality (F), Maximum Economic Yield (MEY), Minimum Stock Size Threshold (MSST), Overfishing Limit (OFL), Spawning Stock Biomass (SSB), Scientific Uncertainty is the total of P* and sigma as described in the text and unfished biomass (B_0).

United States Alaskan crab

Alaska's 10 crab stocks under federal jurisdiction are jointly managed by the North Pacific Fishery Management Council (NPFMC) and the Alaska Department of Fish and Game. The NPFMC makes recommendations for the OFL and the ABC for each crab stock, while the State of Alaska makes decisions on total allowable catches (TACs), which have to be less than the ABCs. OFLs are set using a tier system that includes five tiers depending on data availability and the ability to estimate key stock assessment parameters (Table 8).

Table 8. Tier assignments for the 10 Alaskan crab stocks under federal management (NPFMC, 2014). Note: some systems in the USA uses the term "category".

Tier number	Description	Example
1	Stocks with assessments which provide reliable estimates of biomass and B_{MSY}, and in which a probability density function (pdf) for F_{MSY} is estimated.	None
2	Stocks with assessments which provide reliable estimates of biomass and B_{MSY} and F_{MSY}.	None
3	Stocks where reliable estimates of the spawner-recruit relationship are not available, but biomass can be estimated reliably and proxies for F_{MSY} and B_{MSY} can be estimated.	Eastern Bering Sea snow crab (*Chionoecetes opilio*; Oregoniidae) Bristol Bay red king crab (*Paralithodes camtschaticus*; Lithodidae) Eastern Bering Sea Tanner crab (*Chionoecetes bairdi*; Oregoniidae)
4	Stocks for which reliable estimates of biomass are available (directly from surveys or from assessments), but insufficient biological data are available to estimate $F_{x\%}$.	Pribilof Islands red king crab (*Paralithodes camtschaticus*; Lithodidae) Pribilof Islands blue king crab (*Paralithodes platypus*; Lithodidae) St. Matthew Island blue king crab (*Paralithodes platypus*; Lithodidae) Norton Sound red king crab (*Paralithodes camtschaticus*; Lithodidae)
5	Stocks which have no reliable estimates of biomass and only historical catch data is available.	Aleutian Islands golden king crab (*Lithodes aequispinus*; Lithodidae) Pribilof Island golden king crab (*Lithodes aequispinus*; Lithodidae) Adak red king crab (*Paralithodes camtschaticus*; Lithodidae)

No crab stocks are currently in tiers 1 and 2 (although Punt et al. (2014b) suggest it may be possible to estimate stock-recruitment relationships for some of the stocks in tier 3). The most data-rich crab stocks are therefore in tier 3, with a few stocks in tier 4. Three stocks are currently in tier 5. Figure 5 shows the HCR used to set the OFL for Alaskan crab stocks in tiers 1 – 4. Note that unlike the OFL in the HCR for the west coast groundfish fishery (Figure 3), fishing mortality rate is reduced when stock size is less than B_{MSY}. The OFL for stocks in tier 5 is set to the average catch over a period considered by the NPFMC's Crab Plan Team to correspond to B_{MSY}.

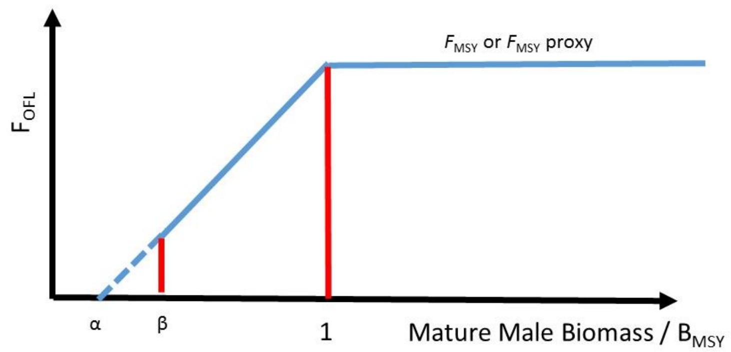


Figure 4. Harvest control rules for the USA Alaska crab fisheries. Terms are: Biomass (B) at Maximum Sustainable Yield (MSY), Fishing mortality (F), Overfishing Limit (OFL), Scientific Uncertainty is the total of P* and sigma (denoted σ) as described in the text and unfished biomass (B_0).

The buffer between the OFL and ABC is selected by the SSC of the NPFMC for each species based on suggestions from the NPFMC Crab Plan Team (e.g. NPFMC, 2014). The maximum possible ABC (MaxABC) is based on the P^* method for stocks in tiers 1 to 4, and is 90% of the OFL for stocks in tier 5. P^* is assumed to be 0.49, while the variance of the OFL is set based on the results of the stock assessments. The uncertainty levels in the assessments are considered unrealistically small (Punt et al., 2012), and consequently, the buffer between the OFL and the MaxABC is much smaller for Alaskan crab stocks than for west coast groundfish stocks. However, the buffer imposed by the NPFMC SSC is larger than that implied by the agreed P^* value. For example, for the 2011/12 to 2013/14 management years, the ratio of the ABC to OFL for tier 3 and 4 stocks has ranged between 0.43 and 0.90 (NPFMC, 2014). Moreover, the TACs set by the State of Alaska for these stocks are often substantially smaller than the ABCs set by the NPFMC because the Alaskan State harvest strategies are often more precautionary than the federal rules used to set OFLs. From 2011/12 to 2013/14 the TACs for tier 3 and 4 stocks has ranged between 0 and 0.804 of the ABC set by the NPFMC.

Australia's SESSF

The SESSF tier system uses the ability to produce a reliable assessment based on the available data to define tiers, which are used to inform management decisions through the setting of the Recommended Biological Catch (RBC) (Table 9). The type of data available defines the method used for assessment and the associated HCR in this system. Tier 1 equates to a robust age-and sex-structured assessment to assess stock status (see stocks listed in Table 9). It should be noted that, until recently, the SESSF did not have independent survey data and therefore such data did not influence the tier system. The generic structure of the HCR for tier 1 stocks is provided in Figure 5. The target biomass is the biomass corresponding to Maximum Economic Yield (B_{MEY}) or a proxy thereof (the default proxy is 1.2 B_{MSY}). B_{MSY} is only estimated for Tiger Flathead (*Neoplatycephalus richardsoni*; Platycephalidae) (Punt et al., 2014a). For the remaining SESSF stocks, B_{MSY} is based on a proxy (B_{MSY} = 0.4 B_0). The limit reference point (B_{LIM}) is 0.5B_{MSY} (or the proxy 0.2B_0). The target fishing mortality rate is set to that corresponding to F_{MEY} (or the proxy of F_{48}) and from 0.35B_0, this rate declines linearly with biomass to zero at the limit reference point.

Table 9. Tier assignments for the SESSF stocks under Australian Commonwealth management.

Tier Number	Description	Example stocks
1	Stocks with an available quantitative stock assessment, which provides estimates of biomass and fishing mortality in absolute terms and relative to agreed limit reference points.	Tiger Flathead (*Neoplatycephalus richardsoni*; Platycephalidae) Blue Grenadier (*Macruronus novaezelandiae*; Merlucciidae) Pink Ling (*Genypterus blacodes*; Ophidiidae) School Whiting (*Sillago flindersi*; Sillaginidae) Orange Roughy (*Hoplostethus atlanticus*; Trachichthyidae) Eastern Gemfish (*Rexea solandri*; Gempylidae) Silver Warehou (*Seriolella punctata*; Centrolophidae)
2	Stocks with a quantitative stock assessment but the assessment is considered less robust than a tier 1 assessment.	Not presently applied
3a (age-composition data)	Stocks for which age data are available for catch curve analysis (Wayte and Klaer, 2010).	Redfish (*Centroberyx gerrardi*; Berycidae) John Dory (*Zeus faber*; Zeidae) Mirror Dory (*Zenopsis nebulosus*; Zeidae)
3b (length-composition data)	Stocks for which age data are available for a length-based catch curve analysis (Klaer et al., 2012)	
4	Stocks for which catch-per-unit-effort is assumed to be a proxy for abundance (Little et al., 2011).	Silver Trevally (*Pseudocaranx dentex*; Carangidae) Royal Red Prawn (*Haliporoides sibogae*; Solenoceridae) Blue Warehou (*Seriolella brama*; Centrolophidae) Ribaldo (*Mora moro*; Moridae)

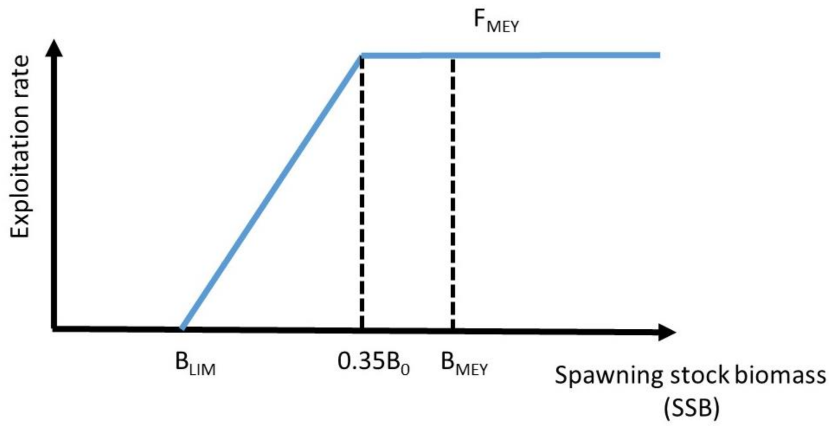


Figure 5. Harvest control rules for the tier 1 control rule for the Australian SESSF. Terms are: Biomass (B) at Maximum Economic Yield (MEY), Fishing mortality (F), Spawning Stock Biomass (SSB), and unfished biomass (B0).

Tier 2 is similar to tier 1, except that the stock assessment is less robust. To account for this reduced reliability, the default Tier 2 B_{MSY} proxy is 0.5 B_0, and the B_{MEY} proxy is 0.6 B_0 (Smith et al., 2008). However, at present no species are assigned to this tier. Tier 3 applies to stocks where age- (Wayte and Klaer, 2010) or length- (Klaer et al., 2012) structure data are available (Figure 6). Within tier 3, age- or length-based catch curve analyses are used to obtain the values required for the HCR (tiers 3a and 3b respectively in c). Tier 4 assessments use catch-per-unit-effort (CPUE) as a relative index of abundance (Little et al., 2011; see stocks listed in Table 9). Tier 3 is assumed to provide more robust assessment results than Tier 4, but this was not supported by results from MSEs (Fay et al., 2012; Little et al., 2014).

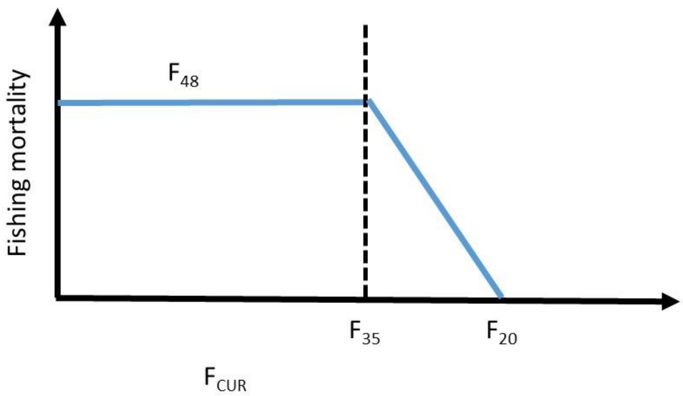


Figure 6. Harvest control rules for the tier 3 control rule for the Australian SESSF. Terms are: Current (CUR), Fishing mortality (F).

In 2010, RBCs for 12 SESSF stocks were placed in tier 1, 4 in tier 3, and 14 in tier 4. No stocks were placed in tiers 2 and 3b in 2010. A recent development in the fishery is the implementation of multiyear TACs. As a result, in 2014 two stocks were newly assessed using tier 1, 2 in tier 3 and 2 in tier 4. However, to date, there has been no formal testing (e.g. using management strategy evaluation – Smith 1993) to determine decision rules for multiyear TACs (Haddon et al., 2012).

The expectation of risk equivalency between tiers in the SESSF is generally achieved through the use of buffers in the HCR for each tier. Tier 1 is based on a robust assessment and therefore no buffer is applied. No buffer has been set for tier 2, due to it presently being unused. In tier 3, the RBC from the basic HCR is multiplied by 0.95 to account for the additional uncertainty associated with management advice based on catch curves (Smith et al., 2008). The buffer of 0.85 for tier 4 stocks is used to account for the additional uncertainty associated with management advice based on trends in catch-rate.

There are several meta-rules that further modify the outcomes from the HCRs (Stobutzki et al., 2011):

- The extent of precaution associated with the tier 3 and 4 HCRs (0.95 and 0.85) can be reduced if there is evidence that other management measures such as closed areas provide additional precaution for the stock.
- The maximum increases or decrease in TACs is limited to 50%.
- No change is made to the TAC if the recommended change in a TAC is less than 10% or 50 t, whichever is less, unless there is a long-term trend in RBC.

The Commonwealth Harvest Strategy Policy has a rebuilding requirement for stocks depleted below the limit reference point, but unlike the setting of RBCs, the process is not formalised but individual rebuilding strategies are developed for any such depleted species. In addition, there are several stocks for which there are both length/age-composition data and catch-rate data. In this case, the Resource Assessment Groups (RAGs) (Smith et al., 2001) use expert judgement to select whether to place the stock in tier 3 or 4.

Europe and ICES

In order to help policy makers move towards sustainable exploitation of all fish stocks, ICES implemented a framework in 2012 (the Data-Limited Stocks, or DLS, framework) to provide quantitative advice, not only for fish stocks for which there is sufficient data to conduct full analytical assessments, but also for those fish stocks considered data-limited. At the time, this represented a 6-fold increase in the number of data-limited stocks for which quantitative advice was provided (ICES, 2012). The framework relies on the principle that available information should be used, and that advice should follow a precautionary approach and be based on the same principles as applied to data-rich stocks. This framework therefore calls for the determination of exploitation relative to F_{MSY} and the consideration of stock trends, and provides a variety of methods that can be applied to facilitate this; the choice of method depends on the information and data available for a given stock. The ICES DLS framework is divided into six categories, with category 1 applied to the data-rich stocks and categories 2 to 6 to increasingly data-limited stocks (Table 10). A decision tree (Figure 7) determines under which category a stock will fall, with the availability of high-quality data and proxies decreasing, and the level of precaution applied increasing, from categories 1 to 6. Within each category, there are further groupings based on methods that can be applied within that category (ICES, 2012). Of the data-limited stocks for which advice was provided in 2014, the bulk fell within category 3 (which bases advice on fishery-dependent or -independent indices) and categories 5 and 6 (which bases advice on recent catch or landings data only) (Table 10; ICES, 2014b). In 2014, the most widely used DLS method essentially adjusts the previous year's catches (or average of previous years' catches, depending on whether a trend in catches is evident or not) (see Appendix A for mathematical details).

Table 10. Tier (referred to as categories by ICES) assignments for the stocks under ICES DLS framework (ICES 2012, 2014b).

Tier Number	Description	Example stocks
1	Stocks with quantitative assessments that provide present stock status and are also able to provide forecasts of stock status.	North Sea cod (*Gadus morhua*; Gadidae) Bay of Biscay sole (*Solea Solea*; Soleidae) *Nephrops* in North Minch
2	Stocks with analytical assessments and forecasts, but these can only be treated qualitatively.	Golden redfish in I and II (*Sebastes marinus*; Sebastidae) Eastern Channel plaice (*Pleuronectes platessa*; Pleuronectidae)
3	Stocks with reliable fishery-independent and -dependent indices that provide reliable indications of trends in stock metrics such as mortality, recruitment and biomass.	Irish Sea haddock (*Melanogrammus aeglefinus*; Gadidae) Greater forkbeard (*Phycis blennoides*; Phycidae) Southern Baltic flounder (*Platichthys flesus*; Pleuronectidae)
4	Stocks with only reliable catch data are available which allows MSY to be approximated.	Pollock in VI and VII (*Pollachius pollachius* and *P. virens*; Gadidae) *Nephrops* in Noup
5	Stocks for which only landings data are available.	Sea bass in VIIIc and IXa (*Dicentrarchus labrax*; Moronidae) Whiting in IIIa (*Merlangius merlangussense*; Gadidae)
6	Stocks for which there are negligible landings data and stocks caught in minor amounts as bycatch.	Rockall cod (*Gadus morhua*; Gadidae) Sandeel in VIa (*Ammodytes* spp.; Ammodytidae)

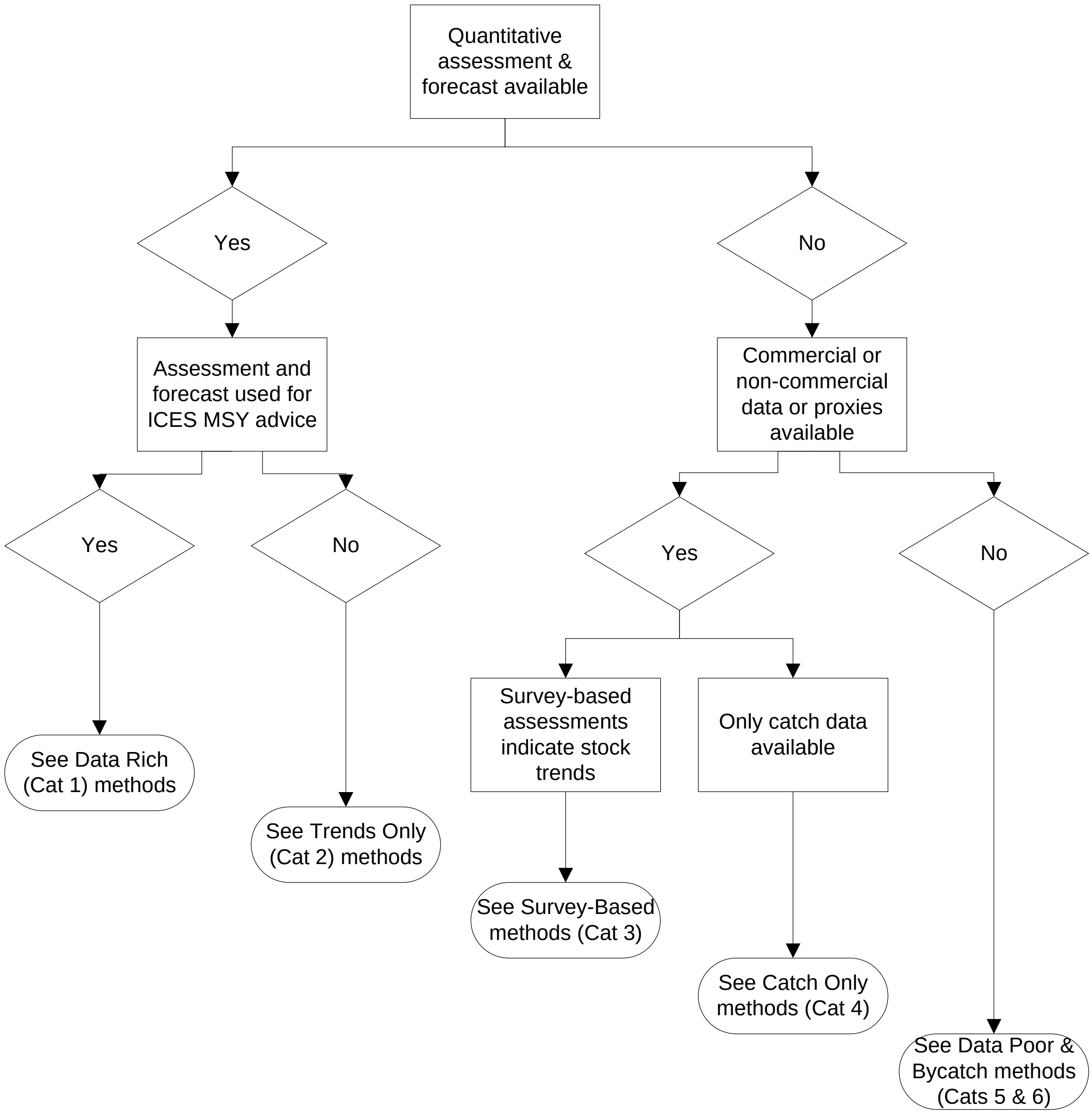


Figure 7. Overview of categories of ICES assessment types for data-rich (Category 1) and data-limited stocks (Categories: Cat 2-6). The availability of high quality data and proxies for the assessments decreases and the extent of precaution increases from left to right. Figure taken from ICES (2012) with permission.

The ICES DLS framework employs increasing precaution by first applying a change limit of ±20% relative to a previous catch or average of previous catches (category 2+), and then applying a precautionary margin of -20% (category 3+), although there are exceptions to the latter (ICES, 2012, 2014a). Further precaution is introduced by employing conservative proxies for F_{MSY} (e.g. $F_{0.1}$), where available. The precautionary margin is applied for those cases where it is likely that $F > F_{MSY}$ or when stock status relative to stock size or exploitation reference points is unknown. Exceptions to the latter are when expert judgement deems the stock not to be reproductively impaired and where there is evidence that stock size is increasing or exploitation has reduced significantly (ICES, 2014a). This approach is intended to move stocks in the direction of sustainable exploitation, given their biological characteristics and the level of uncertainty in the available information. Resultant advice is linked to a time frame that is compatible with a measurable response in the metrics used as a basis for the advice, implying that multi-annual constant catch advice could result (e.g. for three years) where the least information is available, unless important new information emerges justifying a revision of the advice (ICES, 2014a).

4.1.3 Risk equivalency and tier systems

The preceding section highlights that each jurisdiction has a different way to address the expectation of risk equivalency. Only the Australian tier system has an explicit assumption that the risk associated with all species should be equivalent irrespective of the data available (Smith et al., 2014). The Australian tier system also does not expect a fishery to move over time to tiers that are more data rich (especially if the value of the fishery is such that a data-limited tiers are more appropriate), as long as risk equivalency is addressed. In some other tier systems, moving up to more data-rich tiers is an explicit aim. For example, the USA west coast groundfish tier system has a lower $P*$ value for the more data-poor tiers and hence more precaution for more data-poor stocks. In contrast to the USA west coast groundfish fishery, the default $P*$ is the same for stocks in tiers 1 – 4 for the USA Alaskan crab fishery, implying an assumption of risk equivalency among tiers because $P*$ is a measure of the probability of overfishing, a key measure of risk for USA fisheries.

All of the tier systems have several control variables that determine the size of the buffer (e.g. sigma and $P*$ for the USA west coast groundfish fishery). Unfortunately, how the values for the control variables were selected is often unclear. For example, the default value for sigma in the USA west coast groundfish fishery is 0.36 for category 1 stocks, which is based on the meta-analysis of Ralston et al. (2011). In contrast, the values for sigma for category 2 and 3 stocks are simply multiples of 0.36 in the absence of a better basis to define scientific uncertainty.

In the USA system, a buffer is calculated based on the extent of scientific uncertainty, whereas the discount factors in the SESSF are essentially untested (Fay et al., 2012). The ICES system has a mix of approaches depending on the tier. Ideally, the values for the control variables should be selected to achieve desired policy goals. Tier systems can be tested using management strategy evaluation, ideally conducted with stakeholder involvement (Smith et al., 1999). The MSE would focus on the relationship between the values for the control variables and the performance of the management system. For example, the values of the control variables for a given stock could be selected so that risk is constant among tiers (Figure 8). The resulting trade-off would be between catch and monitoring cost, giving the decision makers the ability to select a monitoring strategy under the expectation of equal risk.

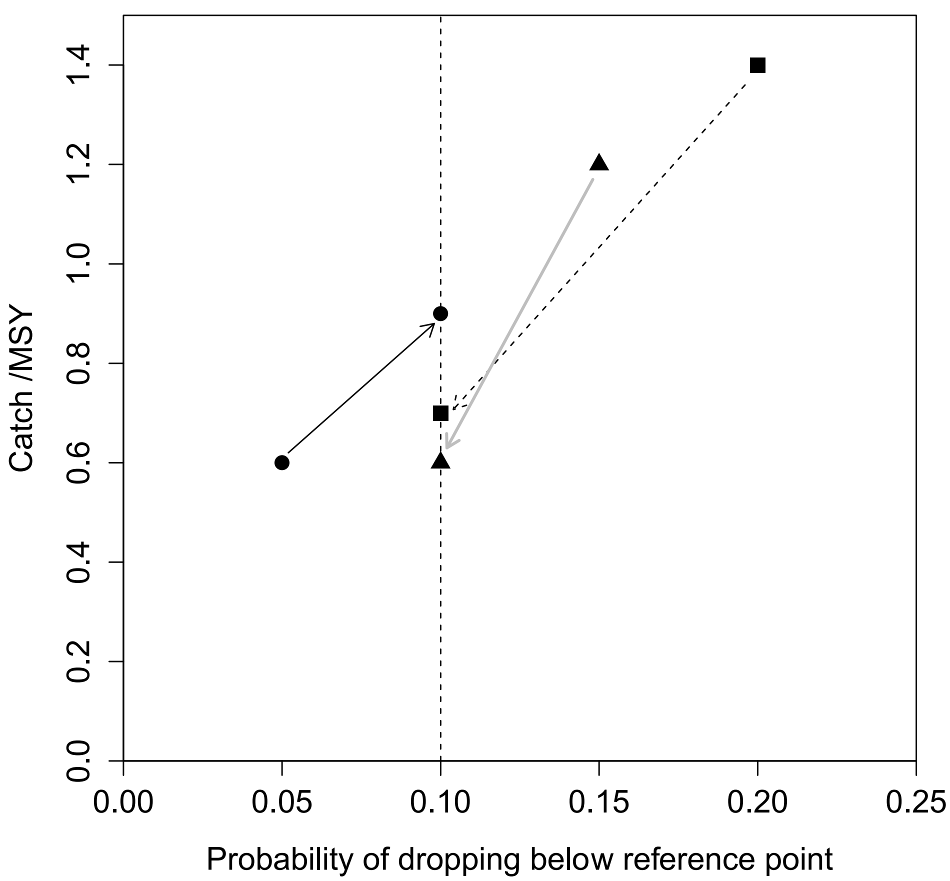


Figure 8. Three examples of the potential consequences of adjusting the control variables for each of three tier HCRs so that they achieve risk equivalency. The vertical dotted line is the Australian Commonwealth overfished risk level.

Punt et al. (2012) evaluated alternative choices for the buffer between the OFL and the ABC for a range of Alaskan crab stocks using an MSE-like process. They provided trade-offs between the probability of overfishing and catch, given uncertainty in stock size. The results were guided by members of the NPFMC Crab Plan Team and informed the choice by the NPFMC for its default values for P^* for crab stocks in tiers 1 to 4. An MSE is being planned to examine the consequences of the choices for σ of 0.72 and 1.44 for data-moderate and data-poor USA west coast groundfish stocks (C. Wetzel, NOAA, pers. commn). That MSE will consider the relationship between a key control variable (σ) and risk, given there is uncertainty regarding the true extent of uncertainty for most stocks of west coast groundfish.

MSE has been used extensively to evaluate the HCRs within each tier for the SESSF. For example, the performances of the tier 3a and 3b HCRs (Table 9) have been evaluated for both data-rich and data-poor fisheries based on the operating model developed by Fay et al. (2009) (Wayte and Klaer, 2010; Fay et al., 2011; Klaer et al., 2012). The performance of the Tier 4 HCR has also been evaluated using the same operating model (Little et al., 2011). These MSE analyses led to changes to the way the HCRs are applied for stocks in tiers 3 and 4. However, in relation to tier systems, this work has not attempted to evaluate the process of selecting tiers for individual stocks (which can be somewhat subjective), the values for the control variables which determine the buffers for each tier, nor has work been conducted to evaluate the process of deciding whether to set multi-year TACs and some of the “modifiers” such as whether a buffer should be applied, even though the decision about which tier a stock is placed in could have a marked impact on management outcomes. In general, the MSE analyses showed that the risk between tiers is very case-specific. Also, the use of the tier 3 HCR seems more likely to be a higher risk than use of tier 4, which runs counter to the current default buffers. This has been seen also in single a species treatment (Fay et al., 2012; Little et al., 2014).

ICES has conducted MSE simulation testing of methods proposed for the various DLS categories through a number of its groups; ICES (2014b) provides an overview of this simulation work. However, to date the focus has been on HCR performance, rather than to assess whether there is risk equivalency among tiers, although an MSE did consider whether the principle of increasing precaution down the tiers was achieved for a stock that was forced into more data-limited tiers by making less of its data available to the DLS framework. In this study, risk equivalency was found wanting (ICES, 2014b), indicating that further work was needed on the appropriate size of the precautionary margin for the different tiers. To date, the magnitude and duration of the precautionary margin has not been evaluated using MSE.

To practically demonstrate, using a more quantitative approach, the differences between the various systems, the SESSF species were placed into USA and ICES frameworks (Table 11) through comparison of data types, assessment methods and harvest control rules. The default buffers that would apply under these jurisdictions are also added. There is relatively strong concordance between the various systems, with most SESSF tier 1 stocks being assigned to high tiers within the USA and ICES systems. No stocks were assigned to the USA Alaskan crab tier 1 as none of the SESSF assessments fully quantify uncertainty using a probability density function. Two of the SESSF assessments (those for school whiting and orange roughy) would likely be placed in the USA west coast groundfish fishery category 2 owing to sensitivity to assumptions (school whiting) or limited data to estimate year-class strength (orange roughy). This is not unexpected given that the SESSF system is based on the assessment method applied and implemented, with less emphasis on its reliability. The SESSF tier 3 and 4 stocks would be assigned to tiers 3c/d under the USA west coast groundfish fishery system, tier 5 under the USA Alaskan crab system, and categories 3 and 4 under the ICES system.

There is some subjectivity associated with this assignment of stocks to tiers. For example, the USA west coast groundfish fishery system involves expert judgement regarding whether a full assessment should be assigned to tier 1 or tier 2. The groups responsible for assigning stocks to tiers in the USA and within ICES are scientists, whereas in the Australia system they include scientists, industry and managers (Smith et al., 2001)

Table 11 shows that the buffers that would be applied if the SESSF stocks were managed under the USA and ICES frameworks were largest for USA west coast groundfish fishery system; for some species the buffer would be more than 25% larger (stocks in tier 4). The buffers under the USA Alaska crab system are close to 1 for the assessments based on models (but given the way that system operates in reality, the buffers are usually 0.9 or smaller).

Table 11. Comparison of tier systems through placing the SESSF species into the different systems. CV refers to the CV of estimated biomass for the most recent year (used to calculate the buffer in the USA west coast groundfish fishery system). The values in parentheses are the default buffers that would apply to each stock.

SESSF species	CV	SESSF	USA west coast Groundfish	USA Alaska crab	ICES
Flathead	0.111	1 (1)	1a (0.956)	2 (0.997)	1 (1)
Blue grenadier	0.137	1 (1)	1b (0.956)	3 (0.997)	1 (1)
Ling[1]	0.225 (E)	1 (1)	1a (0.956)	3 (0.994)	1 (1)
	0.202 (W)	1 (1)	1a (0.956)	3 (0.995)	
School whiting	0.191	1 (1)	2d (0.833)	3 (0.995)	1 (1)
Orange roughy (east stock)	0.093	1 (1)	2d (0.833)	3 (0.998)	1 (1)
Eastern gemfish	0.306	1 (1)	1a (0.956)	3 (0.992)	1 (1)
Spotted warehou	0.096	1 (1)	1a (0.956)	3 (0.998)	1 (1)
Redfish	-	3 (0.95)	3c/d (0.694)	5 (0.9)	4 (0.8)
John dory	-	3 (0.95)	3c/d (0.694)	5 (0.9)	4(0.8)
Mirror dory	-	3 (0.95)	3c/d (0.694)	5 (0.9)	4(0.8)
Blue warehou	-	4 (0.85)	3/d (0.694)	5 (0.9)	3(0.8)
Royal red prawn	-	4 (0.85)	3a (0.694)	5 (0.9)	3 (0.8)
Silver trevally	-	4 (0.85)	3d (0.694)	5 (0.9)	3 (0.8)
Blue-eye trevalla	-	4 (0.85)	3c/d (0.833)	5 (0.9)	3 (0.8)
Ribaldo	-	4 (0.85)	3c/d (0.694)	5 (0.9)	3(0.8)
Eastern deepwater shark	-	4 (0.85)	3c/d (0.694)	5 (0.9)	3(0.8)

1 – there are two stocks of ling off southeast Australian (Whitten and Punt, 2014), east (E) and west (W)

4.2 Principal Component Analysis

The data scores for the Australian Commonwealth fisheries, for each main target species/species assemblage are shown in Appendix B, along with the currently assigned tier level for the stock. Tiers were assigned under the expanded 8-tier system. The SESSF species retained their original tier designations, with the exception of Blue Grenadier (*Macruronus novaezelandiae*) and Orange Roughy (*Hoplostethus atlanticus*). Both originally classified as SESSF tier 1 species under the Smith and Smith (2005) system, these were reassigned as tier 0 as per the definitions in Table 2. Currently, no Australian stocks or species are assigned to tier 5, but this tier is included because it represents a level of data availability and an assessment of intermediate quality compared to tiers 4 and 6 (Dowling et al., 2013)

PCA results detailed results are shown in Appendix B, and summarised below.

The data indicated that the more information-poor tiers (7, and, to some extent, 6) are located separately to the more information-rich tiers (0 and 1) in PCA space. A broad trend of increasing tier number from left to right along the first principal component axis is evident. The low tier numbers (0-1) were distinguished by low scorings associated with the first principal component that distinguished age (CA) and length frequencies (CL). The intermediate tiers (2-6) along the right showed a high degree of overlap.

There are several reasons for the overlap among tiers. Firstly, data alone may not determine or define the appropriate assessment. For example, although Common Banana Prawns (*Fenneropenaeus merguiensis*, a species in the Northern Prawn Fishery) have fishery independent survey data, and high quality catch, effort and CPUE data, their population dynamics are strongly environmentally driven, and the drivers are poorly understood. Hence they are classified as tier 4, as a more formal stock assessment is unable to be undertaken. More generally, the Northern Prawn Fishery has species within tiers 0, 2, and 4, and all of those have scores on PCA2 > 1.0 and PCA1 < 0.0. That is, they all group in the top left of Figure 10, which distorts tiers 0, 2, and 4.

There were three separate groups within tier 2, none of which were close to each other on the PCA plot (Figure 10): (i) Southern Bluefin Tuna had scores of 0 to 1 for each data type, (ii) Small Pelagic Fishery's blue mackerel, Australian sardine and redbait species had no catch data, and (iii) Northern Prawn Fishery's blue endeavour and red-legged banana prawn species (Figure B.1; Table B.3) had no, or poor quality, catch age- and length-composition data, whereas Southern Bluefin Tuna was scored at 0 and 1 for these data types, respectively (Table B.3). This illustrates that these data types alone do not explicitly characterise tier 2, because fisheries can still have tier 2 assessments even though there might not be length- or age-composition data (e.g. ageing is not currently possible for crustaceans), or catch data. On the basis of data availability and quality, therefore, tier 2 is not coherent; data is used differently to designate tier level between fisheries. The life-history characteristics of the species also affects the assessment method selected. Within the same tier, therefore, there is very different use of data.

Similar disparity occurred within the tier 0 and 6 species/species groups: the Heard Island and McDonald Island (HIMI) fishery (tier 0) and the Western Tuna and Billfish and Skipjack Tuna fishery (tier 6) had both moderate to high quality length- and age-composition data, whereas the Northern Prawn Fishery brown and grooved tiger prawns (tier 0), and Western Deepwater and North West Slope Trawl fisheries (tier 6) did not (Figure B.1; Table B.3). Alternatively, data may be available and yet are not necessarily used in, or appropriate for, the assessment for the species. For example, the tier 4 Bass Strait scallop fishery has fishery-independent survey data, but these data are not used in the context of a formal stock assessment, as the estimates of biomass are uncertain. The surveys are rather used to check on the location and relative size of scallop beds, as well as the proportion of under-sized and on condition of scallops prior to harvesting.

Also, data quality may be high per se, but not in the context of assessment requirements, relegating the species or fishery to a lower tier level. For example, effort data for blue endeavour prawns in the Northern Prawn Fishery is scored as high quality, (= 0), but in an assessment context, the quality of these data would be compromised because the effort is not targeted.

Additionally, the scorings are still relatively coarse. Ocean jackets, a SESSF tier 4 finfish species, and the tier 7 Coral Sea Hand Collection species assemblage, occurred in close proximity in the PCA bi-plot (Figure B.1, due to their identical scorings of 1 for total and landed catches, and effort, and 2 for CPUE (Table B.3). While an empirical assessment is able to be undertaken for the former, the latter, being species assemblages, often with opportunistic targeting and with more temporally sporadic time data series, is currently managed via a system of catch-based triggers.

Clustering by data type followed intuitive expectation: total catch and CPUE were tightly clustered, together with reported catch and effort, while age- and length-composition data were clustered together. Fishery-independent survey data clustered separately, but closer to catch and CPUE data.

Repeating the scoring exercise validated the original results: clear segregation between tiers 0 and 1, and tiers 6 and (especially) 7, and similar clustering by data types (results not shown).

4.3 Monetary economic analysis

4.3.1 Costs and assessment precision

Detailed results are shown in Appendix C. Management cost controlled by assessment variability (Figure 1) is shown based on data from eight stocks in the Australian Southern and Eastern Scalefish Fishery (SESSF). These costs included survey costs, and were fitted to a function (Appendix C). The relationship however, was strongly dependent on the assessment cost of mirror dory, which had a large CV (1.86) associated with the most recent biomass estimate, which was relatively inexpensive to determine (70,000 AUD). The relationship was also dependent on two stocks with relatively expensive surveys: orange roughy, *Hoplostethus atlanticus* (>300K AUD) and blue grenadier, *Macruronus novaezelandiae* (>100K AUD), but relatively precise biomass estimates (0.11 and 0.16 respectively).

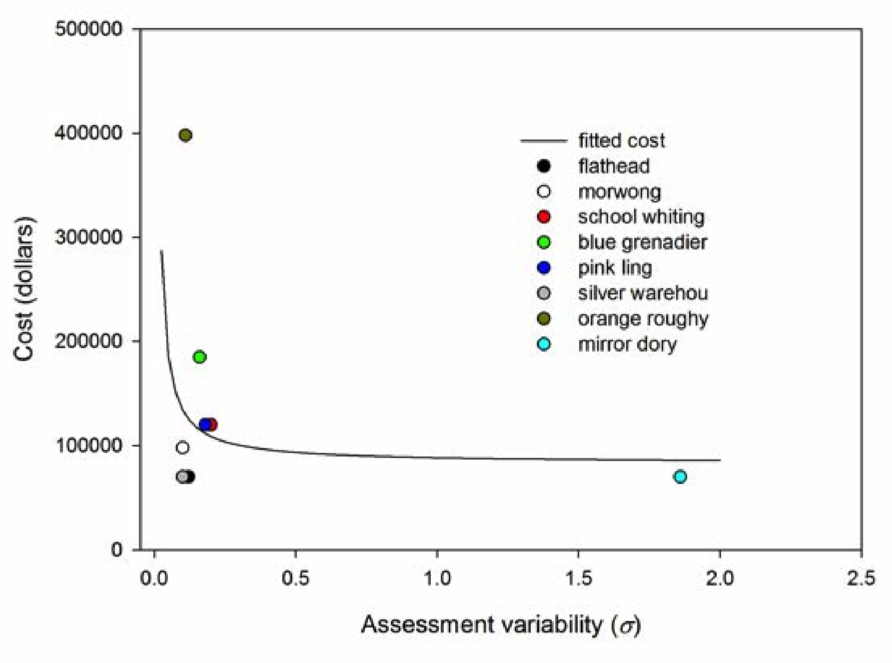


Figure 9. Fitted relation between stock assessment costs and the variability in estimates of biomass based on the most recent data from eight stocks in the Australian SESSF (Tiger Flathead: *Neoplatycephalus richardsoni*, Morwong: *Nemadactylus macropterus*, School Whiting: *Sillago flindersi*, Blue Grenadier: *Sillago flindersi*, Pink Ling: *Genypterus blacodes*, Silver Warehou: *Seriolella punctate*, Orange Roughy: *Hoplostethus atlanticus*, Mirror Dory: *Zenopsis nebulosa*).

4.3.2 Risk trade-offs

The probability that a stock is correctly estimated to be below a limit reference point of 20% is close to 1.0 when the true biomass is less than 20% and the assessment CV is low (Figure 10). This probability decreases as the assessment CV increases. In contrast, the probability is 0 when the true biomass is greater than the 20% limit reference point and the assessment variability is low, but converges to 0.5 as the assessment CV increases.

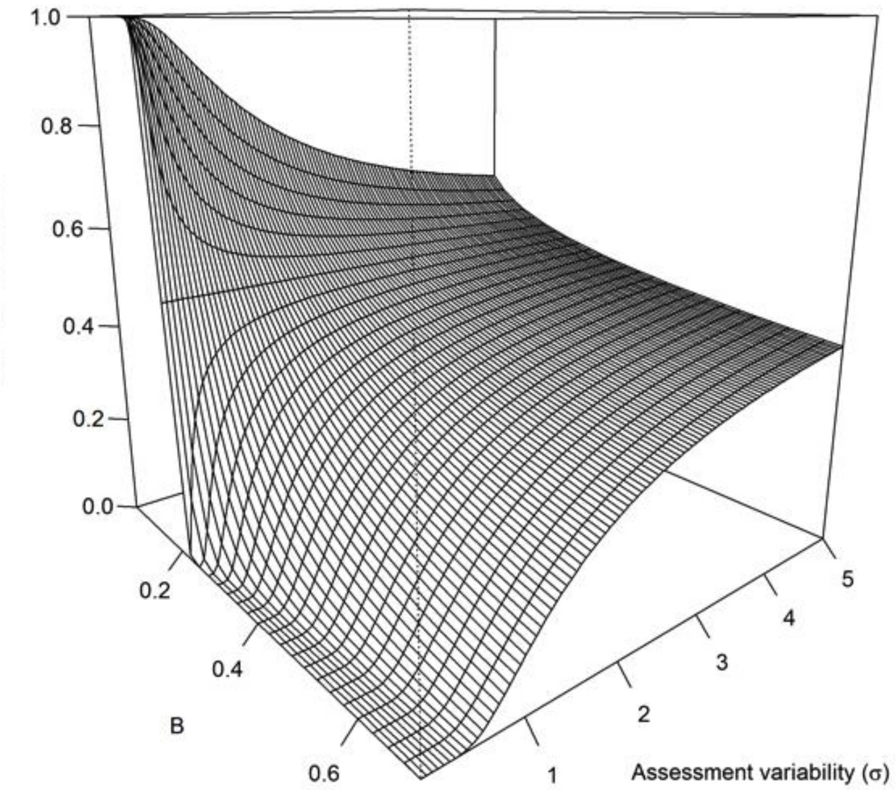


Figure 10. The probability that the estimated biomass is below a limit reference point of 20% of the unfished biomass as a function of the precision of the assessment variability and the true biomass (B).

The expected cost to management from incorrectly estimating the stock size as well as the expected cost of having to rebuild a stock are all influenced by the precision of the assessment (assuming it is unbiased), and the biomass target that management chooses. The expected cost of incorrectly claiming the stock has crossed the limit, when in fact is has not, is highest when the target biomass is close to the limit, and declines as the target biomass level increases past the limit. The expected cost of such an error also increases as the variability in the assessment estimate increases. Alternatively, there is little expected cost from a false positive error when the target is below the 20% limit, a situation which is not likely to result.

The expected cost of a false negative error; i.e. incorrectly concluding the stock is above the limit, decreases quickly as the target moves away from the limit. This cost also increases as the assessment variability increases. The expected cost associated with correctly estimating the stock to have crossed the limit also declines quickly as the target biomass increases away from the limit biomass.

The total expected cost, including all cost components under different assumed data collection and assessment cost functions, was lowest at intermediate target levels. This represents the best trade-off between the cost associated with crossing the 20% limit reference point, and the cost of forgoing profits.

When data collection and assessment cost are independent of assessment precision the global minimum cost occurs at the most precise assessment. However, this global minimum shifts to less precise assessments as the management cost of assessment precision increases however, because lower assessment variance increasingly adds to the cost. The highest costs on the cost surface thus shift away from those associated with low levels for the biomass target level to those associated with assessment precision.

The biomass target used by a management agency to minimize the total expected cost should, in general, increase as the assessment variability increases, with the underlying stock variability tending to reducing the effect. The effect of assessment variability in setting the optimal biomass target also declines as the profitability of the fishery dominates the objective function. When this happens, increasing variability in stock dynamics decreases the optimal target biomass toward the limit, as estimates of the underlying stock are obscured, and management priority is to maximize profit (through harvest).

When the weight on conservation increases, management should set higher biomass targets to minimize their total expected costs especially when the stock dynamics are noisy. The interaction effect between increasing both the profitability and conservation on the objective function indicates that profitability is dominant for the current parameter values.

In general, when true underlying stock variability is low, costs are minimized when biomass is targeted around 0.5. However, this target changes depending on the profit motive of management and the underlying stock variability. In fisheries where management places relatively low weight on profitability, the optimal target biomass that minimizes total expected costs, increases with increasing levels of true biomass variability to a maximum of 1.0. This occurs because at high levels of true underlying stock variability, management is so uncertain in terms of achieving desired outcomes, and the potential costs of overfishing so great that management should target a biomass of 1.0, and thus close the fishery. Increasing the weight on conservation decreases the point of true underlying stock variability at which the target biomass is set to 1.0, and thus management would close the fishery at lower levels of stock uncertainty.

As profitability dominates the objective function the optimal target biomass tends to decline with increasing levels of stock variability. This occurs because increasing stock variability obscures the ability to measure the stock status. When management is dominated by the profit motive, the result is simply to increase harvest by reducing the biomass target level. The interaction between the profit and conservation motives in the objective function indicates that the profitability is the dominant cost component.

4.4 Atlantis analysis: all tiers

4.4.1 Risk

There is a marked increase in R1, the median (across species) probability of falling below $0.2B_0$, when RBCs are based on tiers 5, 5SAFE, 6 and 7, irrespective of whether meta-rules are applied. The median (over species) risk of falling below the LRP for tiers 1, 3 and 4 is zero (Table 12a,b and Figure D.S4a,b). Apart from tier 1, most of the tiers do not recover the species to the TRP by the last 30 years of the projection period and thus the risk of being below the TRP (R2) is close to one. This is again regardless of whether meta-rules are applied. The R2 performance metric should be about 0.5 if a species fluctuates around the TRP during the last 30 years of the projection period. The tier 1 HS is closest to this value: the median (across species) probability of being below $0.48B_0$ is about 0.65 and 0.51 (without and with meta-rules respectively). That stated, the results are highly species-specific. For R1, the lower and upper quartiles of R1 and ~0 and ~1 in all cases (regardless of the application of meta-rules), though the lower quartile is substantially higher for tiers 6 and 7 than for the other tiers. For R2 the upper quartile is ~1 in all cases, as is the lower quartile for tiers 5 to 7, but it is ~0.5 in many cases for tiers 1 to 4. This indicates the breadth of the interspecies variability, but also shows that the more data poor tiers struggled to achieve species recovery to the TRP (and in some cases even the LRP) by the last 30 years of the projection period.

Table 12. Median over species (and inter quartile values across species (lower;upper)) of the risk, catch and economic performance metrics for all tiers (a) without meta-rules in use, (b) with meta-rules. Values in square brackets for management costs are the median costs when the more expensive higher level assessments in the hierarchy are triggered. The risk indices are calculated for the final 30 years of the projection period (to allow for recovery in those stocks that initially start in a depleted state), while all other indicators are calculated for the entire projection period. An entry of "NA" indicates that the threshold was not achieved during the projection. The costs are discounted at the same rate as the relative discounted catch; and as the management costs are a fixed price per tier there is no variation across species, except for tier 6 and 7 – where the value in brackets for the management costs for tiers 6 and 7 indicate the true total cost given the frequency with which the more intensive assessments were triggered from within these tiers.

(a)

Tier	Risk $B < 0.2B_0$ (R1)	Risk $B < 0.48B_0$ (R2)	Time Fish Down to *$0.48B_0$* (R3a)	Time Recover to *$0.2B_0$* (R3b)	Opportunity Cost ($ million) (C4)	Harvesting Costs ($ million) (C5)	Manage. costs ($ million) (C6)	Profits ($ million) (C7)	Relative discounted catch (H8)	Catch Variability (H9)
1	0.0 (0.0;0.04)	0.65 (0.53;0.97)	4 (2.25;6.5)	10 (3;NA)	170.1 (-368;1508.5)	1870.1 (55.6;11850)	4.9	275.1 (260.3;290.9)	0.83 (0.66;0.92)	158.45 (72.78;205.31)
3	0.0 (0.0;0.19)	0.94 (0.79;1.0)	4.25 (1.63;5.0)	31 (10.5;NA)	157.8 (-679.9;1894.1)	1719.7 (57.8;11840.2)	4.4	342.5 (301.5;452.2)	0.62 (0.5;0.83)	104.71 (74.9;168.3)
4	0.0 (0.0;0.16)	0.83 (0.66;1.0)	3 (2.25;4.5)	20 (11;NA)	296.7 (-628.8;4001.6)	1801.6 (49.3;7910.6)	3.6	162.1 (151.5;167.3)	0.61 (0.45;0.75)	78.4 (41.02;131.22)
5	0.08 (0.0;0.31)	1.0 (0.98;1.0)	4.25 (2.13;6.75)	23 (17;NA)	179.4 (-932.6.1;2412.9)	1758.6 (61.4;10121.7)	2.1	-45.4 (-50.9;56.6)	0.58 (0.41;0.71)	119.39 (70.78;249.25)
5SAFE	0.30 (0.03;0.69)	1.0 (0.99;1.0)	3.5 (0.5;5.75)	NA (28;NA)	324.5 (-802.7;2965.7)	1700.7 (57.5;9864.1)	0.8	-276.5 (-302.5;-47.85)	0.39 (0.21;0.59)	286.04 (97.42;637.95)
6	0.70 (0.12;1.0)	1.0 (0.97;1.0)	3.25 (2.13;4.75)	NA (32;NA)	189.9 (-461.1;3202.1)	1520.6 (51.6;8802.2)	0.51 [1.24]	-568.7 (-607.1;-273.5)	0.37 (0.24;0.58)	305.1 (166.62;673.62)
7	0.945 (0.21;1.0)	1.0 (1.0;1.0)	3.5 (1.25;5.75)	NA (43;NA)	334.7 (-445.9;3184.6)	1534.1 (60.8;10022.9)	0.46 [2.65]	-486.5 (-525.7;-188.6)	0.29 (0.13;0.41)	623.1 (216;1206.72)

(b)

Tier	Risk $B < 0.2B_0$ (R1)	Risk $B < 0.48B_0$ (R2)	Time Fish Down to $0.48B_0$ (R3a)	Time Recover to $0.2B_0$ (R3b)	Opportunity Cost ($ million) (C4)	Harvesting Costs ($ million) (C5)	Manage. costs ($ million) (C6)	Profits ($ million) (C7)	Relative discounted catch (H8)	Catch Variability (H9)
1	0.0 (0.0;0.12)	0.51 (0.44;0.99)	13.25 (7.75;24)	12.5 (6.5;NA)	210.7 (-358.8;1746.1)	1802.1 (64.4;14883.2)	4.9	57.7 (50.9;61.7)	0.81 (0.66;0.94)	29.68 (15.91;74.44)
3	0.0 (0.0;0.18)	1.0 (0.77;1.0)	7.75 (4.75;10.7)	13 (9;NA)	161.4 (-772.8;2025.8)	1675.1 (71.2;14949.5)	4.4	-195.1 (-210. 7;-74.8)	0.79 (0.42;0.91)	37.34 (12.87;83.63)
4	0.0 (0.0;0.21)	0.86 (0.43;1.0)	7.5 (5.13;11)	13.5 (13;NA)	253.3 (-654.3;3373.2)	1689.5 (62.3;8232.1)	3.6	-73.9 (-260.96;16.1)	0.74 (0.45;0.82)	27.17 (11.43;53.07)
5	0.06 (0.0;0.34)	1.0 (0.89;1.0)	5.25 (1.38;11)	16 (12;NA)	229.1 (-1159.2;2308.2)	1674.6 (69.5;14695.4)	2.1	-100.3 (-123.4;11.5)	0.7 (0.41;0.87)	52.54 (26.25;130.57)
5SAFE	0.19 (0.07;0.91)	1.0 (0.93;1.0)	5.25 (1.38;11)	NA (39;NA)	187.8 (-363.1;1720.5)	1660.8 (67.8;11924.7)	0.8	-281.7 (-297.3;-56.9)	0.59 (0.31;0.73)	184.49 (40.87;415.58)
6	0.48 (0.18;0.93)	1.0 (0.97;1.0)	5 (1.5;10)	NA (40;NA)	188.5 (-469.2;1931.7)	1451.3 (70.3;12360.1)	0.51 [1.68]	-335.1 (-371.5;-88.4)	0.55 (0.24;0.67)	252.09 (106.41;589.08)
7	0.73 (0.30;0.93)	1.0 (0.95;1.0)	4.5 (1.25;8.9)	NA	179.8 (-471.3;1817.8)	1573.7 (70.4;12298.8)	0.46 [3.23]	-367.4 (-394.4;-100.1)	0.54 (0.27;0.66)	229.66 (90.23;596.69)

Some of the differences between species are due to their initial biomass relative to B_0. The number of years it takes to fish down to $0.48B_0$ for those species initially above $0.48B_0$ (R3a in Table 12) highlights that the more data-limited tiers (without any meta-rules) tend to fish down the population slightly faster than tier 1. Including the meta-rules has the effect of extending this duration, and also separating the results among tiers. Only tiers 1, 3 and 4 generally recover species that are initially below $0.2B_0$ to $0.2\ B_0$ and higher (R3b in Table 12), with tier 3 (and an un-tuned tier 4) typically taking considerably longer than tier 1 (or a tuned tier 4). The meta-rules reduce the rate of recovery. Note that tier 3 is often less precautionary than tier 4 (especially tuned tier 4) and is therefore out of sequence if the tiers are meant to reflect degree of risk

By design, the meta-rules do not allow for exceedingly large changes to the TAC between years. However, the effect of these rules depends on stock status. They are more effective at recovering a resource using a data-limited HS because they reduce the large TACs these HSs tend to set. However, for a declining resource, restricting changes in TACs via meta-rules may inhibit recovery or even lead to further decline. The combination of these two dynamics leads to complicated effects of meta-rules temporally. While they can hinder the reversal of a stock decline in the short to medium term, in the longer term (once the stock decline has been realised and reversed) the use of meta-rules can see more conservative TACs (avoiding large and rapid increases), ultimately producing biomasses closer to $0.48B_0$.

The number of species that remain below $0.2B_0$ even towards the end of the projection increases for the more data-limited tiers (Figure 11). This increased risk by tier is due to a number of factors. First, the F (or its proxy) values estimated by tiers 3 to 5-SAFE relative to F_{MEY} are lower than for tier 1 (Figure 12); with the exception of the tuned tier 4, which can be more conservative than tier 1 (although F/F_{MEY} for tier 4 is based on CPUE rather than fishing mortality). Tiers 3-5 therefore tend to allow for more overfishing because TACs remain higher than they should be given the assessment underestimates past fishing mortality relative to the target level and trying to "correct" for thus. This is why the higher (more data-limited) tiers tend to set higher TACs at the start of the projection period compared to tier 1 (Figure 13). The cumulative impact would be large if this relative bias remains over the full time period, i.e. irrespective of stock size. There is no *a priori* reason for tiers 1 to 5-SAFE to have different relative biases relative to tier 1 over time, but tiers 6 and 7 might perform differently in terms of relative bias as biomass changes. This is because they depend on the level of conservatism set within in the first trigger levels, which are catch relative to the historical maximum catch (HMC) (tiers 6 and 7), and catch composition, CPUE area fished and spatial distribution trends (tier 7).

Tiers 6 and 7 perform the worst in terms of the risk performance metrics, because, as applied here, multiple trigger points need to be tripped before the TAC is changed. The triggers are not tripped annually, even for those species that are depleted at the start of the projection period. The level 1 trigger, requiring a tier 5-SAFE assessment, is typically tripped fewer than ten times during the 45 year projection period for either tier 6 or 7 (Figure 14). The level 2 trigger, which leads to the use of the tier 3 HS, differs between tiers 6 and 7 and the difference in performance between tier 6 and 7 is greater when the resource is being fished hard (based on species level results not presented). There is a larger difference in performance between tiers 6 and 7, with tier 7 triggering a tier 3 assessment more often over the projection period (due to a slightly longer response time leading to a poorer stock status before any trigger is activated and because the additional CPUE and composition metrics activate the level-2 trigger more often than the catch volume alone does). These HSs are based on the HMC so their performance is very species- and history-specific. Catch-based triggers rely very heavily on presupposing a level of catch that is precautionary; something that is less likely to be known for a data-limited fishery (as also found by Wiedenmann et al., 2013). Furthermore, the way these HSs were implemented in Atlantis-RCC optimistically assumes that the information required when a tier 5-SAFE or tier 3 assessment is triggered is available.

The final reason the data-poor tiers are riskier is that response time is influential in dictating risk (Figure 15). The best performance possible is two years, since it takes about two years to collect the data that highlights the issue, undertake the assessment and implement a lower TAC. In that regard, only tier 1 mostly acts within the minimum 2-year period. Tier 3 stands out compared to tiers 1 and 4 as having slow response times and therefore an increased risk. Tiers 5-SAFE to 7 take more than 4 years to identify implement a management change.

Overall, the more data-limited the tier, the higher the TACs in the short-term and hence the higher the risk. Meta-rules for an increasing stock help mitigate this risk by reducing unjustified TAC increases. The data-limited methods are slow to respond to potential reductions in stock size, because, relative to tier 1, they calculate more optimistic F/F_{MEY} values, and for tiers 6 and 7, their complexity renders them less sensitive to changes to which they should be responding.

4.4.2 Risk versus Catch

Risk and catch are often traded against each other, with the aim to manage a fishery sustainably by (at least) avoiding the target species falling below the LRP ($0.2B_0$) while remaining profitable (Little et al., 2016). However, the more data-limited HSs may fail to identify a long-term sustainable catch (Table 12), due to their inability to effectively estimate biomass or fishing mortality. This poor performance is expressed as lower stock status and consequently lower cumulative catches. Although there is a large range in the value of the performance metrics per tier due to species-specific differences (often due to different starting stock status and-TACs) (Figure 11), the median values show that the more data-limited the tier, the lower the discounted catch over the projection period (Table 12). The relative discounted catch is much higher for the data-limited tiers than for the data-rich tiers early in the projection period (not shown), as the species is being overfished. Ultimately, however the more degraded stock status means catches are lower than for the healthier stocks under the data-rich tiers and so the cumulative result is that the data-rich tiers lead to higher relative discounted catches.

4.4.3 Costs

A further consideration is the cost of implementing a specific tier. Although there is an increased risk of using more data-limited approaches, the question is whether moving to a tier 1 rule is prohibitively expensive, particularly for a small-scale fishery. Calculating costs is complex given that there are tangible (costs of management and fishing) and intangible (the opportunity costs associated with catch foregone because the fishery was not managed "optimally") costs (Little et al., 2016). In this study, the reference HS – a "bang-bang" HS, which takes all biomass above a target level, is the closest HS possible within the Atlantis-RCC modelling framework to an "optimal" HS. It should be noted that this reference HS does not consider economic or social costs of implementation (such as inter-annual volatility in the catch or jobs lost due to prolonged closures), and so is not a realistic HS that could be implemented. Despite this, the difference in opportunity cost (C4) for tier 1 is quite large, in the order of a year's value of the fishery or more (e.g. gross value of product of $72.3 million AUD in the 2013-14 financial year; Georgeson et al., 2015). Opportunity costs are variable among tiers but tend to be higher for the more data-limited tiers. Discounted profits vary by tier, with the largest for tier 3, then tier 1 and 4 with the remainder not being profitable and worsening with increasing tier number; i.e. as the probability or risk of overfishing, or being overfished increases. With meta-rules, only tier 1 is profitable. The management costs (C6), decrease for the more data-limited tiers, but are two orders of magnitude smaller than other costs. However, costs for tier 6 and 7 increase when higher tiers are triggered. Harvesting costs (C5) are more difficult to interpret. The highest of these are for tier 1 (because the higher stock size allows for greater catches, with associated higher handling and fishing costs) and then tier 3. Catch variability (H9) is highest for the data-limited tiers, which had the additional outcomes of seeing more vessels opting not to fish (for at least part of the year) under these data-poor tiers and the potential for greater undercatch of TACs.

Of the more data-rich tiers, Tier 3 has the lowest opportunity costs, highest profits, and lowest harvesting costs (c.f. tiers 1 and 4), and lower catch variability than Tier 1. However, this comes at the cost of poorer performance in terms of rebuilding to $0.2B_0$ for many species. This is a clear case of where those charged with balancing the trade-offs associated with the fishery would need to decide which criteria they weight more heavily – stock status and the risk of falling below limit reference points (a mandated concern under Australian policy) or economic performance. The need for frank discussion and careful attention is particularly important given the species specific performance of this tier.

4.4.4 Influence of history and ecology

The results per tier were often species-specific. Tier 1 had the potential to recover the majority of life histories, except those that were particularly depressed historically, to the point that even small incidental catches were enough to keep them depleted; e.g. due to very low rates of reproduction (gulper sharks) or a combination of technological and ecological interactions (gemfish). Tiers 3 and 4 are heavily influenced by the history of the fishery. Tier 3 was more sensitive to the ecology of the species than the other tiers, particularly if there are strong time-varying ecological interactions or shifts in productivity. Tier 5-SAFE often worked well for chondrichthyans, but was not as effective for teleosts, especially those with highly variable recruitment. In general, the more data-limited tiers can perform adequately if the stock is productive and large in absolute size. However, the initial catches need to be precautionary for these tiers to perform adequately if the stock is small or has low productivity.

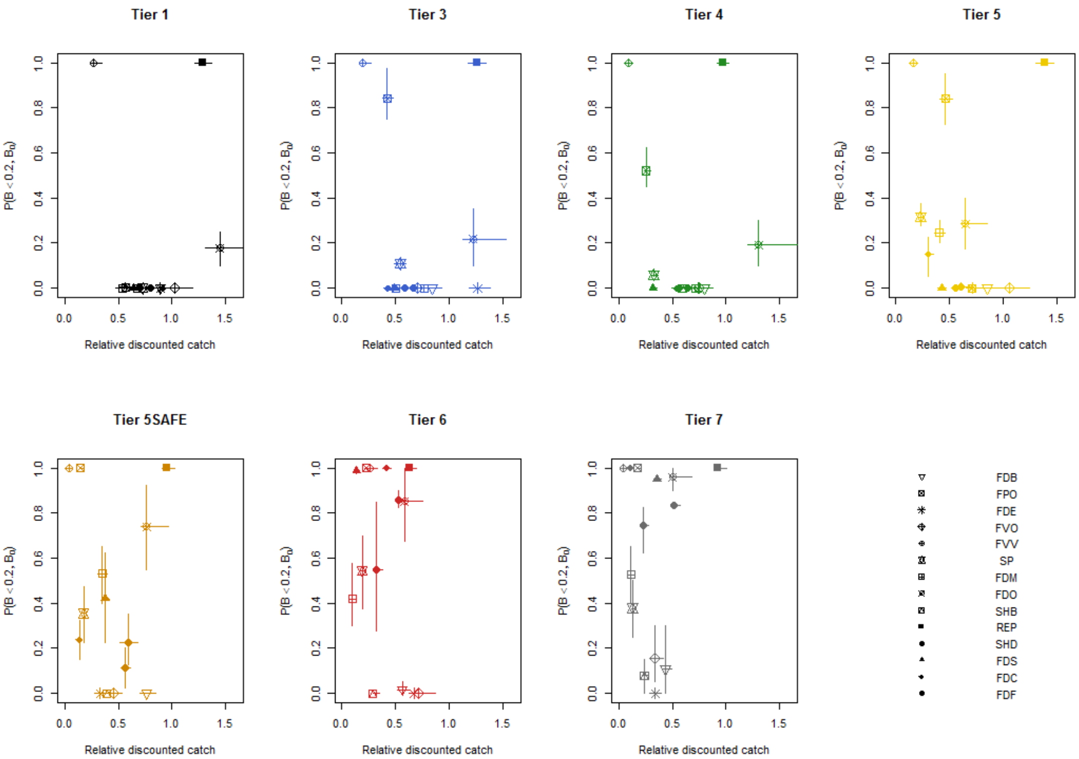


Figure 11. Median (across simulations) probability of being below 0.2B_0 against relative discounted catch with inter-quartile range values for each tier. Results without meta-rules. Each point is a species. FDB=Flathead; FPO=Morwong; FDE=Blue grenadier; FVO=Whiting; FVV=Gemfish; SP=Blue warehou; FDM=Redfish; FDO= Cascade orange roughy; SHB=Gummy shark; REP=Gulper shark; SHD=Demersal sharks; FDS=Generic demersal fish; FDC=Pink ling; FDF=Blue-eye trevalla.

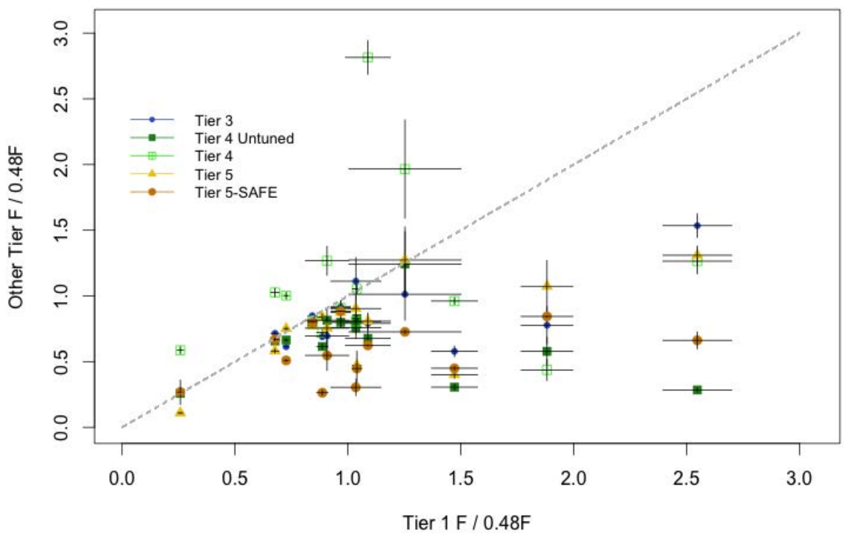


Figure 12. Median tier 1 F/F_{MEY} against that for tiers 3 to 7 for each species. F_{MEY} is the fishing mortality rate corresponding to Maximum Economic Yield (proxied here by the fishing mortality rate corresponding to 0.48B_0).

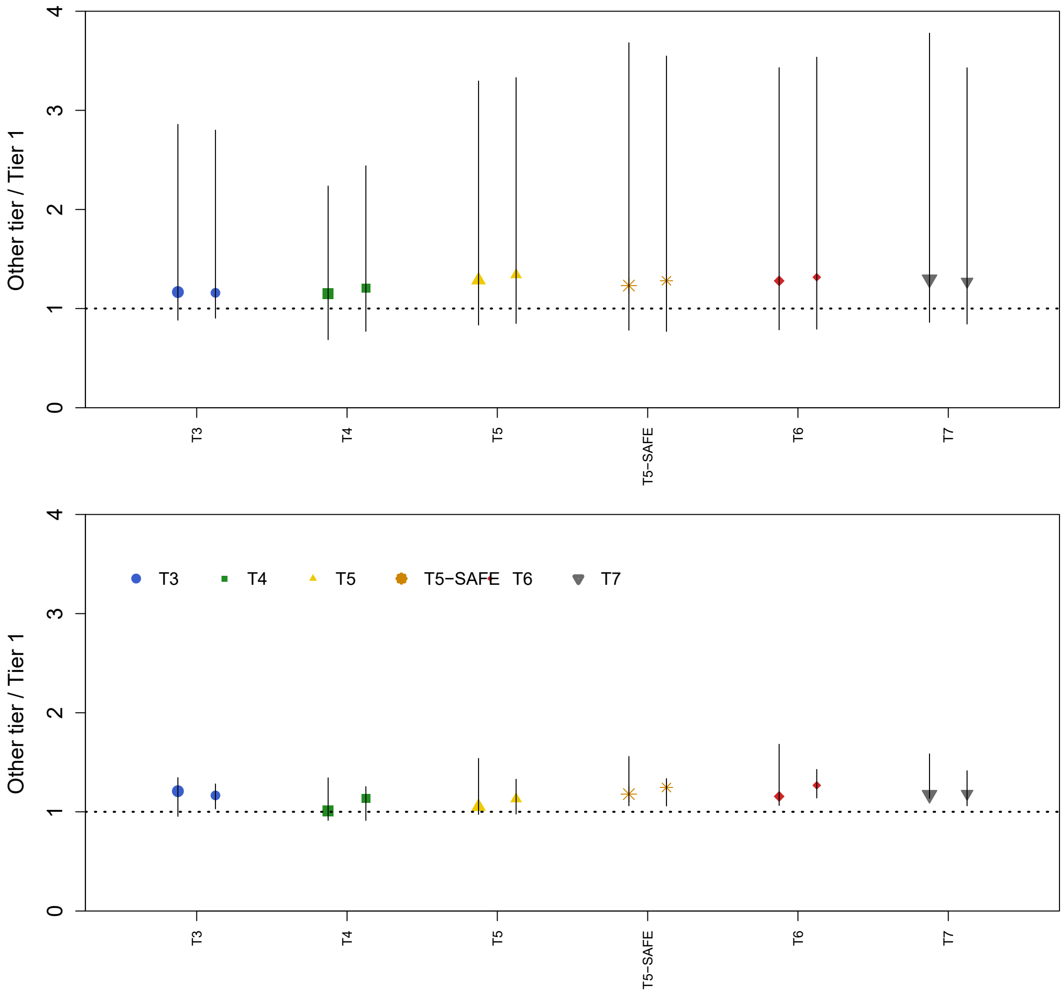


Figure 13. Median ratio (with inter-quartile range) over species of the TAC set in the first year of the projection period under each HS relative to that under the tier 1 HS (larger points) and the ratio of the resulting catches - i.e. the catches taken under those TACs (smaller points). Results are without (top) and with (bottom) meta-rules.

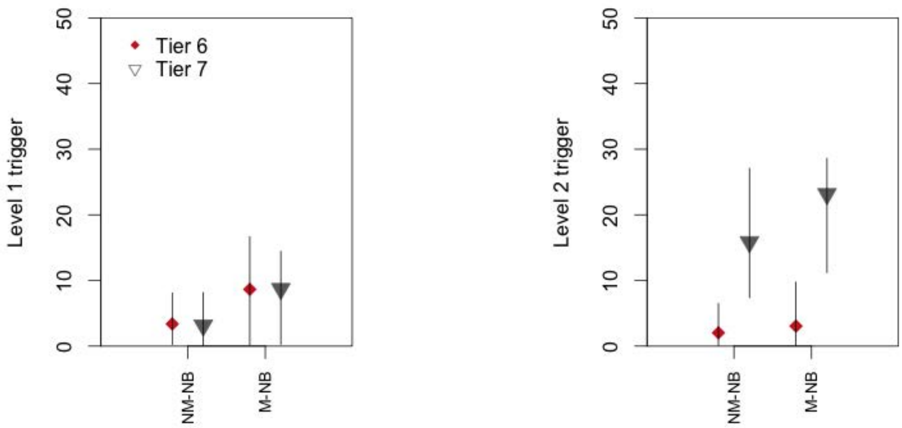


Figure 14. Median and inter-quartile range of the number of times the level 1 (SAFE) and level 2 (tier 3) assessments are triggered for tiers 6 and 7. 'NM' denotes without meta-rules and 'M' denotes with meta-rules.

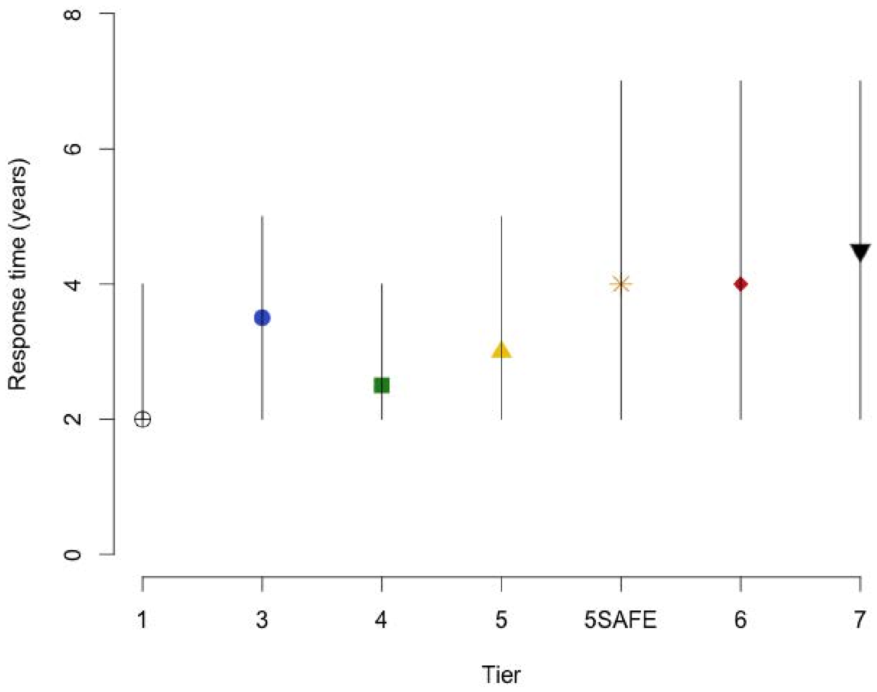


Figure 15. Median and inter-quartile range of the response time under the different tiers.

4.5 Atlantis analysis: SESSF and buffers

4.5.1 Variability among assessments for SESSF stocks

The calculated assessment variance values for the SESSF species varied between 0.1 and 0.38, with an average value of 0.22 and 0.24 if all species and stocks were aggregated (Table 2). These values are smaller than those for USA west coast stocks (cf., average value 0.337 and aggregate value 0.358; Ralston et al., 2011).

4.5.2 Comparing SESSF tiers: buffers and meta-rules

The bulk of the species groups in Table 12 were below the target spawner biomass of $0.48B_0$ in 2006, although some groups (such as blue grenadier, redfish and blue-eye trevalla) were well above $0.48B_0$ at this time (Figure 16). Catches of the groups that are below their biomass target levels are reduced substantially under the bang-bang HCR, while the catches of other species groups are increased substantially for a few years to drive the spawner biomass of the group downwards towards $0.48B_0$ (Figure 17). The reduction in catch for some groups (e.g., flathead) are for only a few years because these groups are productive and not depleted far below the target level. In contrast, catches of groups such as blue warehou, and particularly gemfish and gulper sharks, remain low for most of the 45-year projection period. All of the groups (except for gulper sharks) are at, or close to, $0.48B_0$ by the end of the projection period. Gulper sharks fail to recover (Figure 16), even though catches are reduced to very low levels (< 50 t annually). This is attributable to the impact of the fleet dynamics model that implies that fishing mortality is imposed on some groups (e.g., gemfish, blue warehou, morwong, pink ling and gulper sharks) even when there is no targeted fishery for them. In the case of gulper sharks, incidental catches are sufficient (in combination with time-varying predation mortality) to prevent this group from fully recovering even over 45 years. Similarly, incidental catches slow the recovery of other species; i.e. gemfish, morwong and blue warehou.

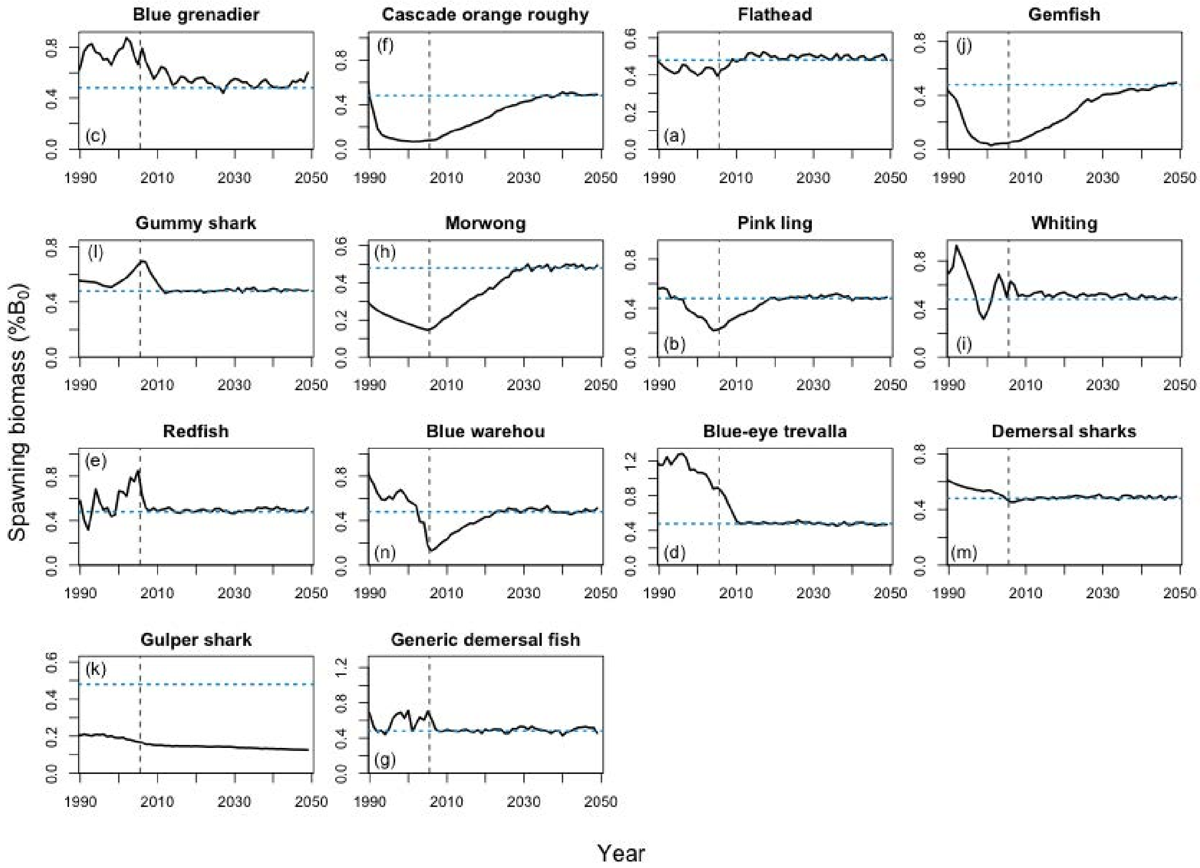


Figure 16. Average (over simulations) time-trajectories of spawning biomass by species / species-group within the SESSF. Catches after 2005 are based on the "bang-bang" HCR. The horizontal line denotes the target reference point.

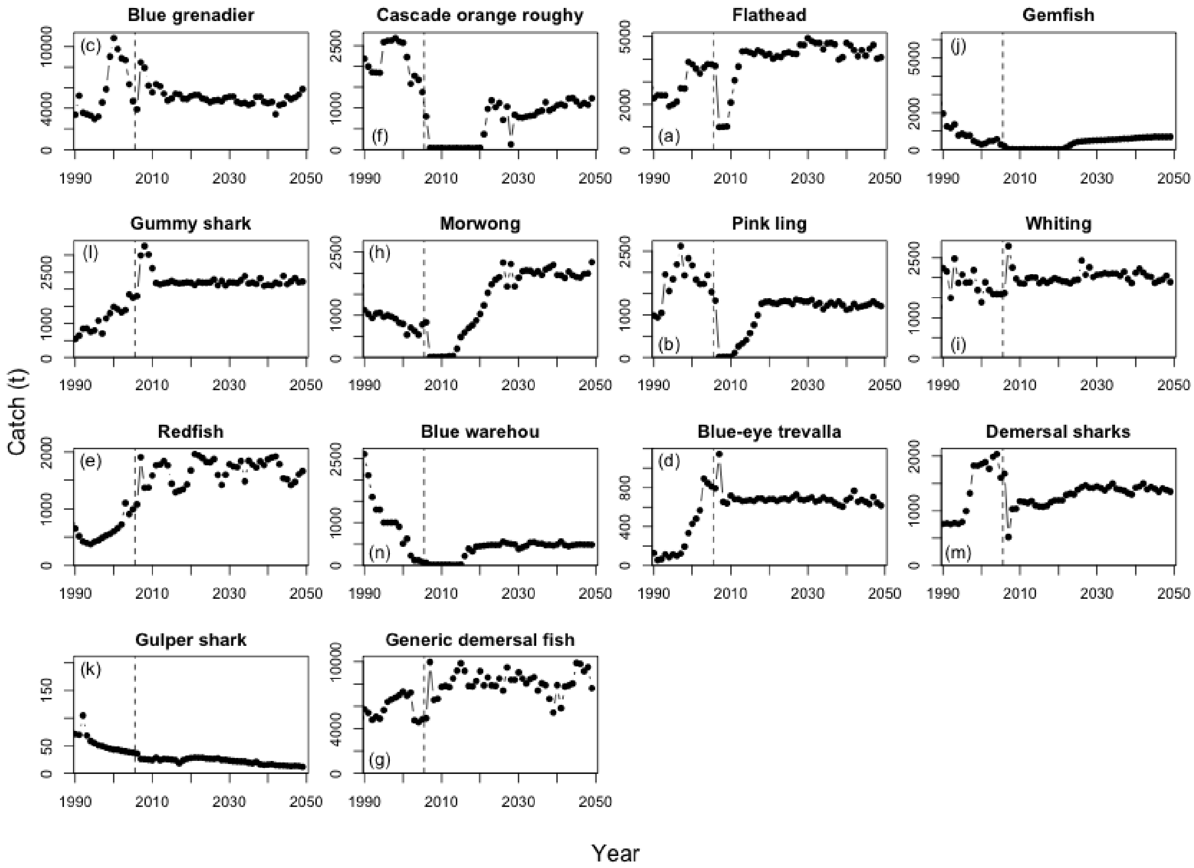


Figure 17. Average (over simulations) time-trajectories of catch by species / species-group within the SESSF. Catches after 2005 are based on the "bang-bang" HCR. The vertical line denotes the start of the projection period (2006).

The median (across species groups) probability of the spawning biomass being below $0.2B_0$ over the last 30 years of the projection period is zero (Figure 18a, b) for all options related to meta-rules and buffers, which is consistent with the requirement under the Australian Commonwealth Harvest Strategy Policy that dropping below this limit reference point should occur at most 10% of the time. However, at the species group level there were cases where this risk was exceeded, and there was considerable among-species variation in the probability of dropping below the limit reference point, with this probability being 100% for two species (gemfish and gulper shark) for all tiers, irrespective of whether meta-rules or buffers were implemented.

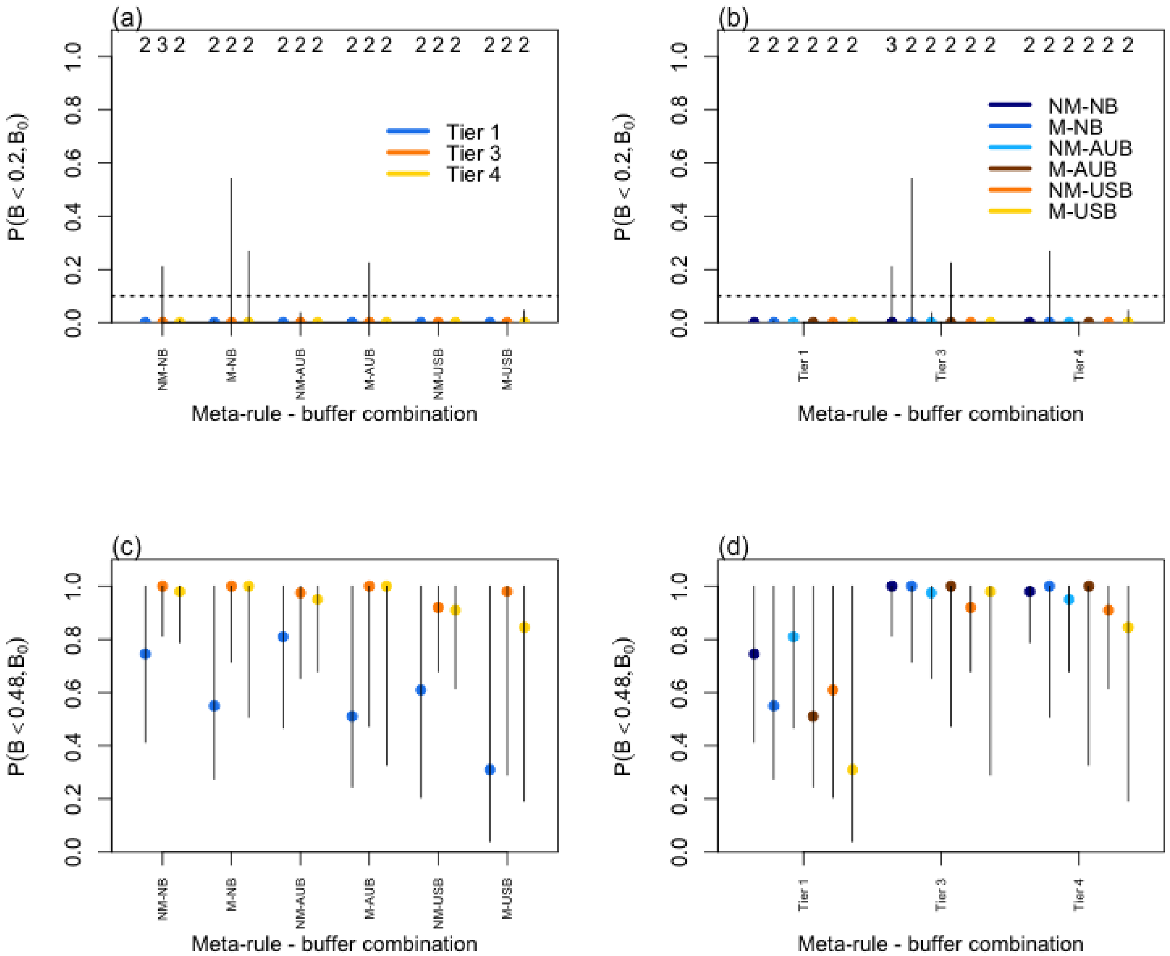
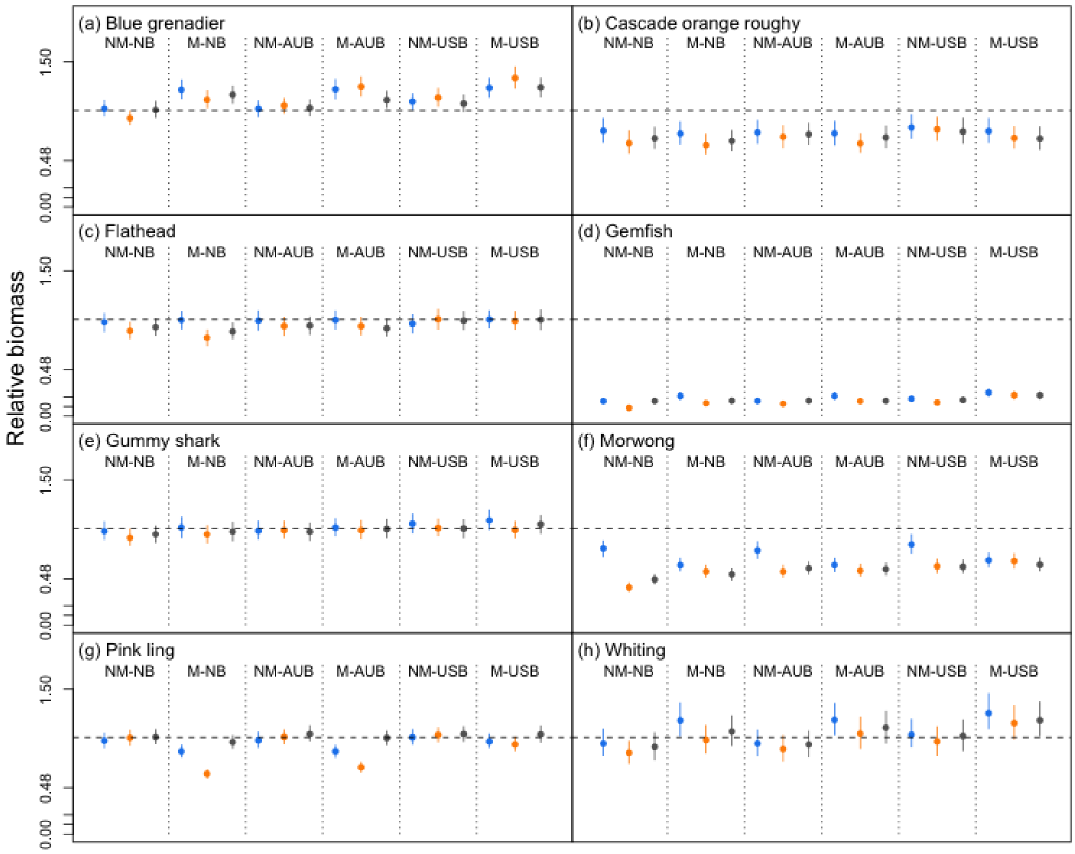


Figure 18. Upper panel: median over groups (and inter-quartile) probability of being below the limit reference point proxy of 0.2B_0 during the last 30 years of the projection period. Numbers across the top of each upper panel indicate the number of species for which the probability was 1.0. Results are presented by tier for each meta rule – buffer combination (left) and conversely, by each meta-rule for each tier (right). Lower panel: median over species groups (and inter-quartile) probability of being below the target reference point proxy of 0.48B_0 during the last 30 years of the projection period. Results are presented by tier for each meta rule – buffer combination (left) and by meta rule – buffer combination for each tier (right).

While the probability of being below the limit reference point is typically quite low over the medium- to long-term. The risk is higher in the short term (e.g., 15-20 years into the projection period, Figure E.S2), where the probability of being below the limit reference point can be as high as 0.6. Stocks managed using tier 3 HCR have the highest potential of being below the limit reference point in that short-term time frame (supplementary material, Figure E.S2b), irrespective of whether or not meta-rules are used. Tier 4 can also lead to several stocks having a probability in excess of 0.1 of being below the limit reference point in the short term. In more than three quarters of the tier-buffer combinations, the use of meta-rules did not reduce the probability of being below the limit reference point. Only the use of USA style buffers consistently reduced the short term risk of dropping below the limit reference points, particularly for tiers 3 and 4; it had less of an impact on the performance of tier 1.

The difference in performance of the tiers according to the presence or absence of meta-rules is also clear when examining the probability that a group is below the target reference point (Figure 18c,d). It may be expected that a stock being maintained at its target reference point would fluctuate slightly above and below its target reference point, so the probability of being below $0.48B_0$ should be roughly 0.5. Tier 1 is the only tier consistently approaching this goal (tending to higher biomasses than the target reference point if meta-rules are applied). Tier 4 can more readily achieve the goal in the medium to long term if buffers are used (it typically fails in the short-term, Figure E.S2c,d); although when using tuned reference periods tier 4 can overshoot (i.e., see biomass exceed $0.48B_0$) when buffers are also applied. Tier 3 only reliably achieves the target goal in the long term when USA style buffers are employed.

The medium- to long-term consequences of applying the tiers can be understood from the relative spawner biomass over the final five years of the projection period (Figure 19) and the time it takes for stocks to reach the limit or target reference points (Figure 20). Where there is any pattern to the long-term relative spawner biomass for a species, tier 1 almost always led to the highest median values, followed by tier 4 and then tier 3; with the exception of redfish where the tuned tier 4 results tend to lead to the highest relative spawner biomass. The implementation of meta-rules did not always lead to higher relative spawner biomass across all tiers. Rather, the impact was group-specific: higher relative biomass for groups such as blue grenadier, and whiting (Figure 19a, h), but lower values for other groups including pink ling, redfish and blue warehou (Figure 19g, i, j). By damping inter-annual variation, it is likely that meta-rules would improve performance for species such as whiting that are fast growing and short-lived (Day, 2010), as well as species with episodic recruitment such as blue grenadier (Tuck et al., 2014). The strong reduction in blue warehou biomass as a result of the meta-rule is potentially due to multispecies fishing effects (it is hard to avoid when fishing other species) and trophic effects, as blue warehou is susceptible to high levels of variability (spatially and temporally) in predation mortality and competition (e.g., with silver warehou) in Atlantis. Consequently, meta-rules constrained reductions in TACs in response to reductions in biomass beyond those anticipated from direct fishing effects. The impact of meta-rules on relative biomass for the remaining species depended on tier, or was minimal.



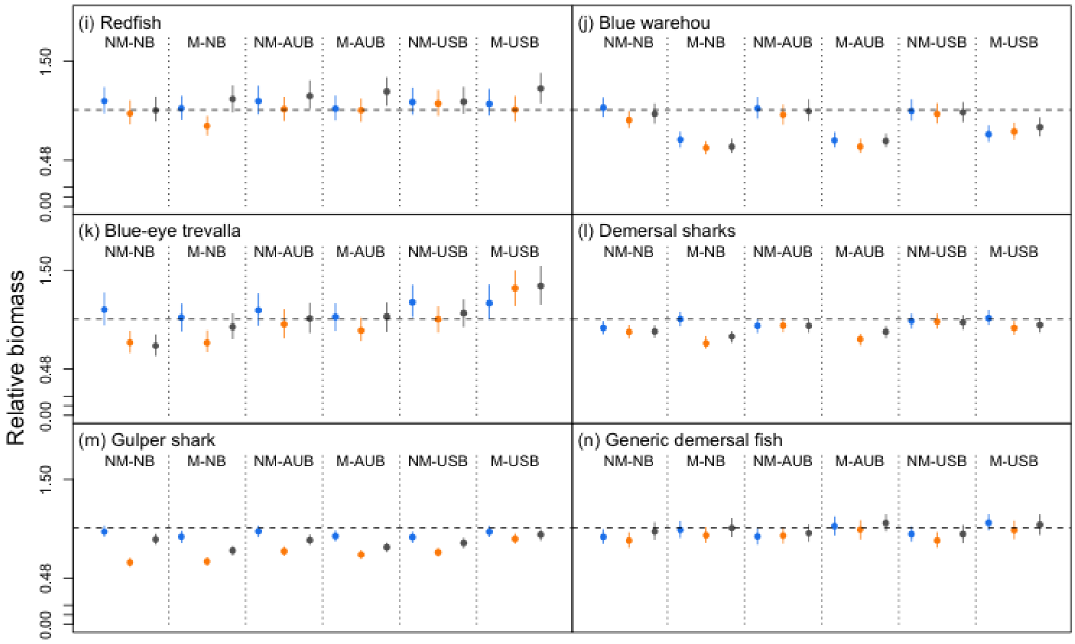


Figure 19. Average, minimum and maximum (over simulations) spawner biomass relative to that under the bang-bang HCR during the last 30 years of the projection period. The horizontal line indicates parity and the colours indicate tiers – tier 1: blue, tier 3; orange and tier 4: grey.

In contrast to the situation for the meta-rules, the application of any buffer with a value less than one led to higher median relative spawner biomass. The more conservative USA buffer system resulted in biomasses in each tier that were closest in value (typically within 5% of each other, as opposed to a >10% difference with the Australian buffers and potentially 20% or more with no buffers). Consequently, while the performance metrics in Figure 18 did not show perfect risk equivalency across all tiers for any buffer-meta-rule combination, the true state of the stock was close to equivalent when the buffers inferred from the USA west coast groundfish fishery were applied (especially with no meta-rules in place in the long term, although in the short term the risk equivalency was closest when meta-rules were in place; giving a sense of the complexity of the temporal dynamics).

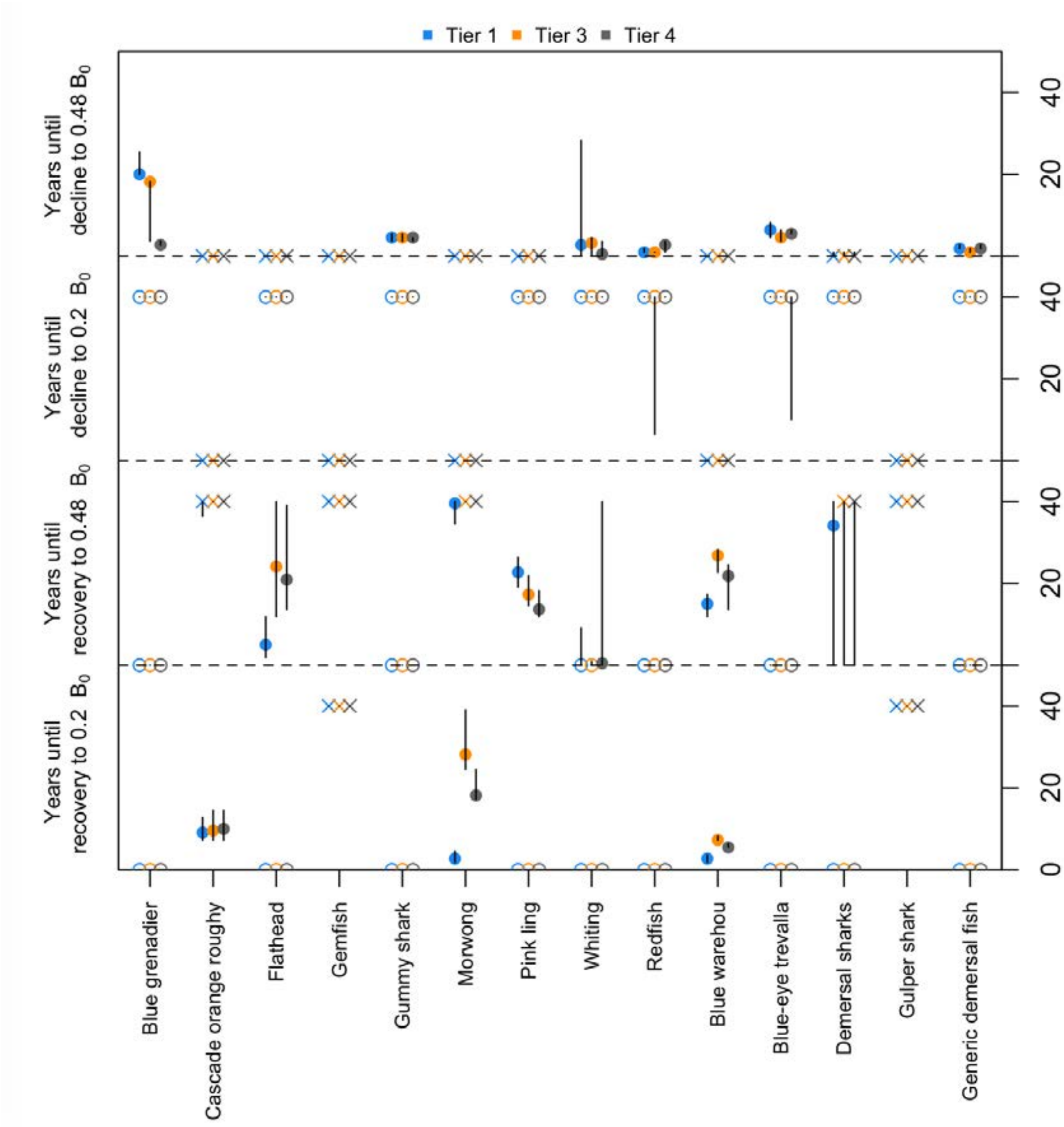


Figure 20a. No buffers; no meta-rules (NM-NB)

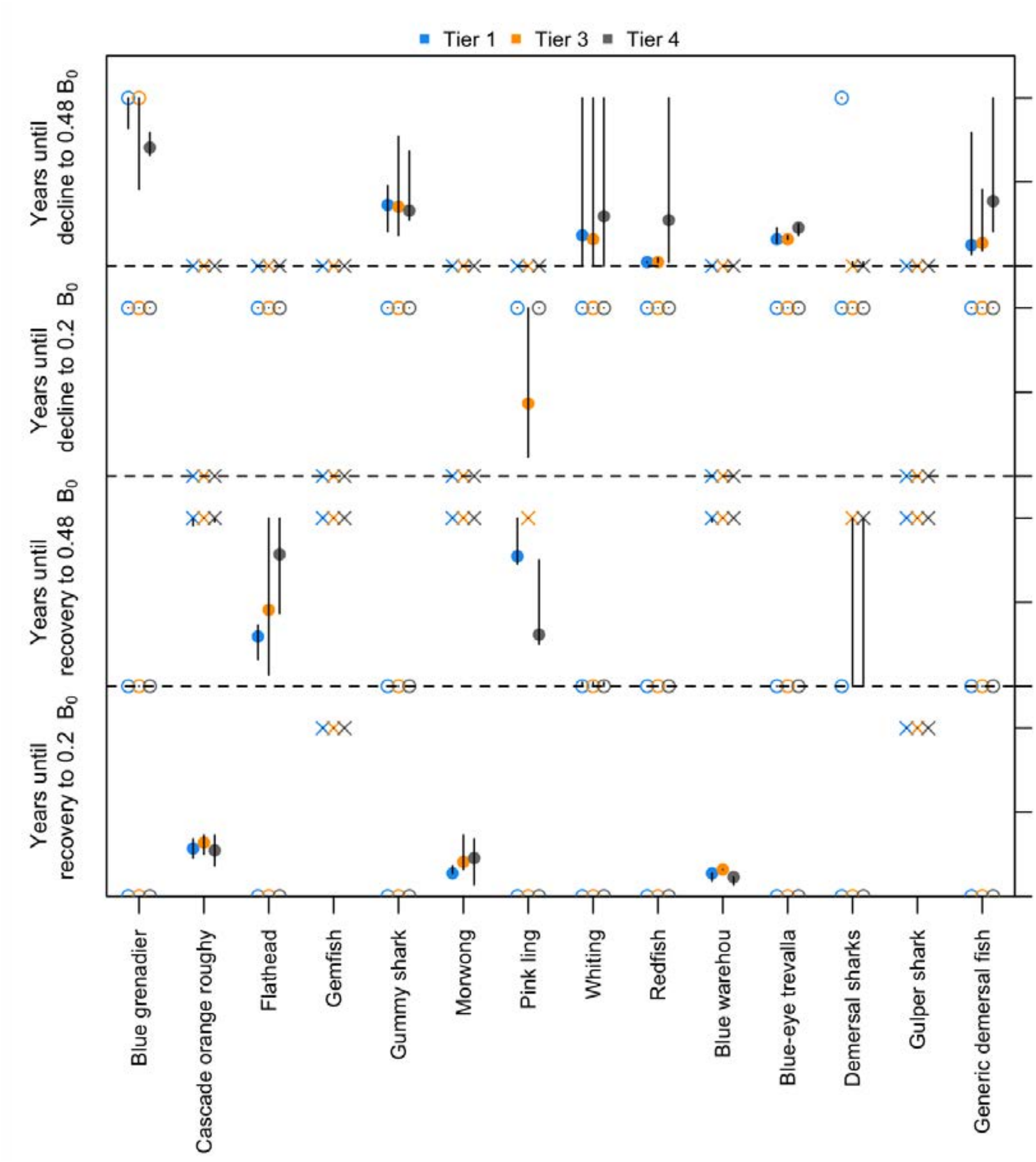


Figure 20b. With meta-rules and Australian buffers (M-AUB)

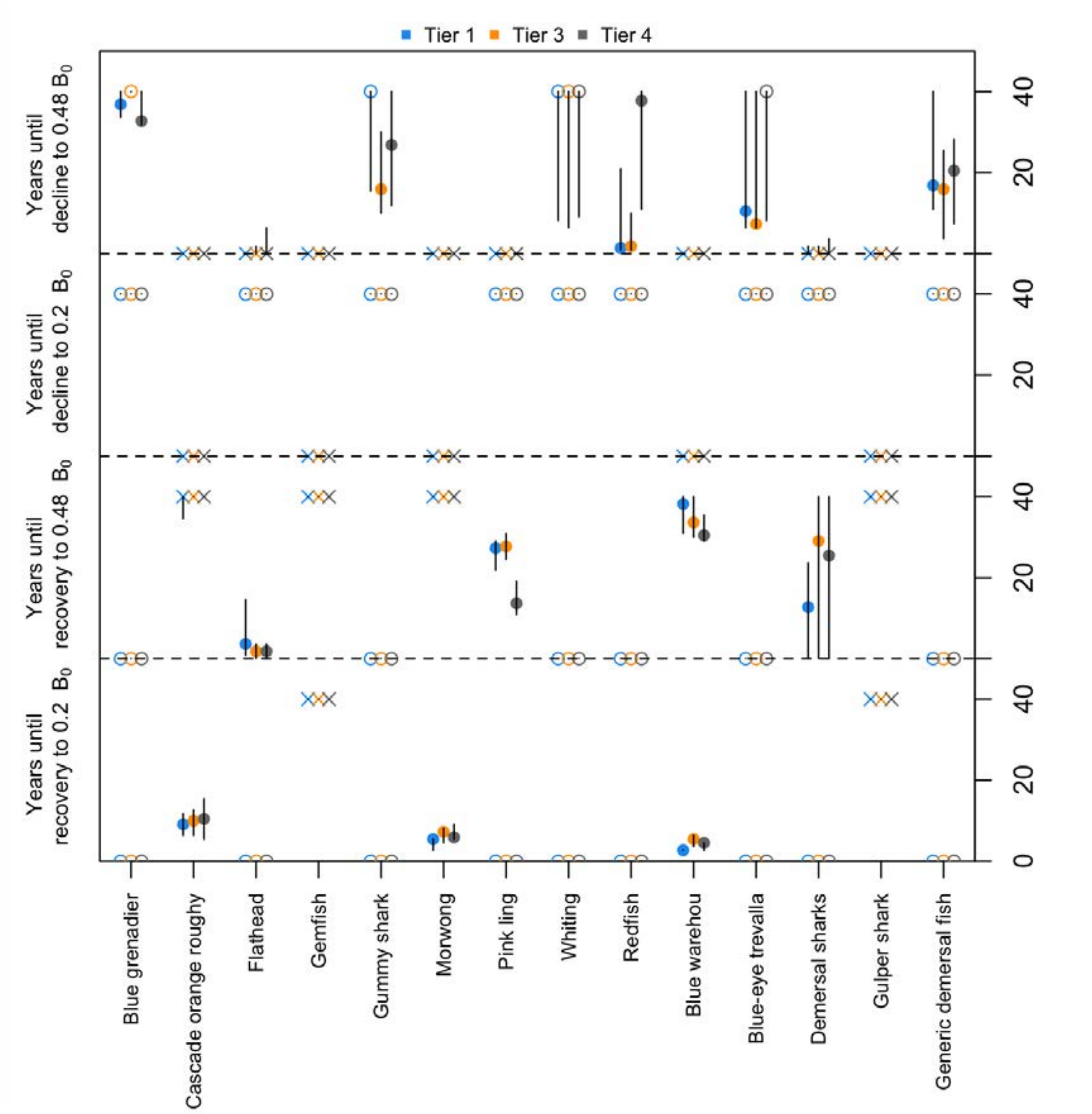


Figure 20c. With meta-rules and USA buffers (M-USB)

Figure 20. Number of years for a group initially above a reference point threshold to breach that threshold (upper two rows: thresholds of $0.48B_0$ and $0.2B_0$), and the number of years for a group initially below a threshold to recover to the threshold (lower two rows: thresholds of $0.2B_0$ and $0.48B_0$). A solid circle indicates the median time with the line indicating the minimum to maximum (over simulations) temporal spread. An open circle indicates that the group began above that threshold and never crossed it, while an "X" indicates that the group began below the threshold and never reached it. Thus, for the upper two panels an "X" indicates the test is irrelevant (as the group was initially below the threshold) and for the lower two panels an open circle indicates the test is irrelevant (as the group was initially above the threshold). The results of greatest concern are those in the lower panel with an "X" at 45, indicating that the group did not recover to $0.2B_0$ at any time during the projection period. Results are shown in (a) for "No buffers; no meta-rules" (NM-NB), in (b) for "With meta-rules and Australian buffers" (M-AUB); and in (c) for "With meta-rules and USA buffers" (M-USB).

Figure 20 shows that the majority of species pass through either the target or the limit reference point during the projection period, for at least one of the tiers. The failure of gulper sharks and gemfish to recover is clear in Figure 20 (they fail to reach the limit reference point before the end of the projection period). There was substantial overlap across tiers in results for the rest of the groups. However, tier 1 is often the "fastest route" to rebuild a depleted stock to the limit reference point. It is not as rapid as the bang-bang HCR, but is often only a few years slower. When using tuned reference periods Tier 4 (which is then typically highly conservative) has the potential to outperform tier 1 in terms of speed of recovery (approaching the bang-bang HCR for some species), but this is not the case for the untuned form of tier 4, underlining the potential sensitivity of application of this tier. Tier 4 can lead to a rapid fish down of lightly exploited stocks, but even so it also allows for a more rapid recovery of depleted stocks than tier 3. Tier 3 often did not perform as well as either tier 1 or 4, because it can allow excessive fishing, especially initially. Its performance is also the most variable across species and life history types of all the tiers. Applying meta-rules and either of the sets of buffers (Figure 20b,c) leads to faster rebuilding times for groups below the limit reference point and to fewer instances of groups being depleted to below $0.48B_0$ or $0.2B_0$. The results for the USA buffer system are noteworthy in that no stocks that were initially above $0.2B_0$ were depleted to be $0.2B_0$ or lower (Figure 18c).

Relative discounted catch indicates that the risks associated with the tier 3, for instance, do result in lower relative cumulative discounted catch (Figure 21a,b), with initially high catches reduced later during the projection due to the poorer stock status than achieved under tier 1. Similarly, the untuned tier 4 can lead to lower relative discounted catches long term, although when tuned its conservative nature can also lead to lower catches achieved than under the other tiers. The lower risk for tier 1 came at the cost of reduced relative catch, with catches more often than not remaining at or below those associated with the bang-bang HCR in the short term but growing to approach those of the bang-bang HCR in the long term. The use of meta-rules leads to slightly higher median relative catches for all tiers, especially for tier 3 (although this distinction is much less pronounced in the short term for tier 3, where the influence of meta-rules is effectively negligible).

The use of meta-rules restricted catch variability by preventing large changes in the RBC (Figure 21c,d). Annual variation in catches is less than 80% of that with no meta-rules: catch variation for tier 1 is consistently more than 25% lower, while that for tier 3 could be as much as 51% lower (or more in some extreme cases). However, the restrictions on catch variation reduced the rate at which catches could be reduced for stocks (such as morwong) that were initially depleted and in need of rebuilding, particularly for tier 1, which leads to the need for large reductions in TAC. However, the meta-rules can also slow prematurely large increases in catches, such as can be recommended under tier 3, facilitating rebuilding (for morwong relative spawner biomass can be 40% higher with meta-rules than without).

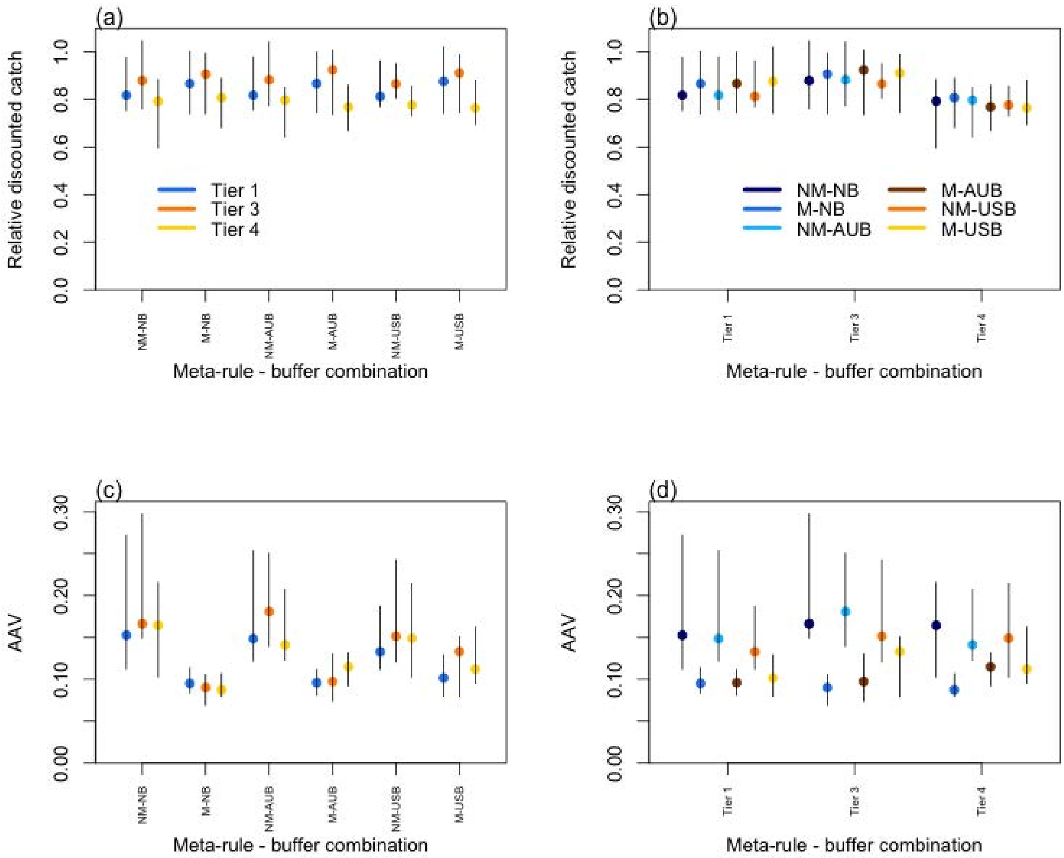


Figure 21. Median over groups (and inter-quartile range over groups) relative discounted catch (upper panels), and the median over groups (and inter-quartile) value of the AAV statistic (lower panels), calculated over the entire projection period. Results are by tier for each meta rule – buffer combination (left panels) and by meta rule – buffer combination for each tier (right panels).

5 Discussion

5.1 Tier systems review

The SESSF tier system evolved to become a means of providing RBCs for all quota stocks irrespective of the data available for assessment purposes, with buffers introduced with the intent of achieving risk equivalency. The SESSF system has since been expanded to include a broader range of species and fisheries beyond the SESSF (Dowling et al., 2013; Table 13). Although the original and expanded tier systems have been subjected to MSE testing, there is no explicit link between the buffers for individual species/stocks and the risk associated with their assessments. In particular, there is no concept in the Australian system that two assessments which use the same data types (that is, within the same tier) may have different associated risk, although that was part of the rationale for the (currently unused) tier 2. In contrast, the USA west coast groundfish fishery includes category 2d that allows stocks, for which the full age-structured assessment results are less than ideal, to be placed in a more data-poor tier, and hence assigned a larger buffer. Moreover, unlike the two USA systems, the Australian and ICES systems have no link between the estimated extent of uncertainty and the buffer size. In principle, the currently unused tier 2 in the SESSF could provide a way to distinguish between reliable and less reliable model-based assessments as was the original intent.

Table 13. Harvest strategy Tier levels from Dowling et al. (2013) with Australian Commonwealth fisheries examples.

Tier number	Description	Example
0	Robust assessment of F and B based on fishery dependent AND independent data	Northern Prawn Fishery tiger and endeavour prawns
1	Robust assessment of F and B based on fishery dependent data ONLY	SESSF Flathead
2	Assessment of F and B based on fishery dependent and/or fishery independent data	Not used
3	Empirical estimates of F based on size and/or age data	SESSF Redfish
4	Empirical estimates of (a) relative biomass based on fishery dependent and/or independent data, or (b) within season changes to relative biomass based on fishery dependent or independent data.	SESSF Blue-eye trevalla ETBF (size-based CPUE slope rules) Arrow Squid Fishery (within-season depletion analyses) Bass Strait Scallop Fishery (spatial management based on pre-season biomass surveys)
5	Empirical estimates of F based on spatial distribution of effort relative to species distribution	None
6	No estimate of biomass and F; use of fishery-dependent species-specific triggers	Western Deep Water Trawl Fishery, North-West Slope Trawl Fishery,

7	No estimate of biomass and F; use of fishery-dependent triggers for groups of species	Skipjack Tuna Coral Sea Fishery Hand Collection: Aquarium sub-sector Coral Sea Fishery: Line, Trawl and Trap sub-sectors

The tier systems include several common elements (Table 14). However, they also differ in some important ways. In particular, only the USA systems explicitly recognize implementation error which can be substantial in many fisheries, especially those with a large recreational component where enforcement of limits on catch and effort is difficult. In addition, the USA and ICES frameworks focus on the ability to estimate quantities needed to apply HCRs whereas the SESSF system is focused primarily on having the data needed to apply a HCR. Unlike the SESSF system, the other systems can assign species with survey data to more data-rich (lower buffer tiers). Finally, no jurisdiction has evaluated the entire tier system, including how species are assigned to tier, although the individual HCRs within the SESSF system have been subject to extensive simulation evaluation.

Table 14. Overview and comparison of different aspects of the four case studies

	USA west coast groundfish fishery	Alaskan crab	Australian SESSF	ICES
Clear definition of risk?	Yes, scientific and implementation risk.	Yes, scientific and implementation risk.	Yes, for Limit Reference Point risk but less clearly for the target reference point	Yes, for Limit Reference Point risk of data rich methods. Not explicitly for others
Basis of tier system	Ability of the assessment to estimate management quantities and reliability of the data types	Ability of the assessment to estimate management quantities and associated uncertainties	Ability to produce a reliable stock assessment	Ability of the assessment to estimate management quantities including forecast and reliability of the data types
Highest data type	Independent survey data	Independent survey data	Dependent data – no extra value for independent survey data	Independent survey data
Number of tiers	3	5	4	6
Presence of sub-tiers	Yes, at all tier levels	No	Yes, but only for tier 3	Yes for most tier levels
MSE tested?	Underway	Yes and does conform to the risk by tier assumption	Yes and does not always conform to the risk by tier assumptions	Yes and does not always conform to the risk by tier assumptions

All of the tier systems involve an element of expert judgement, the consequences of which for risk equivalency are unclear. For example, the choice of whether to assign a stock to tier 1a and 2d is a decision made by the PFMC SSC, while the NPFMC SSC can change the buffer for USA Alaskan crab from the default values. The Crab Plan Team provides the SSC with reasons (if there are any) for setting the ABC less than the OFL. However, at present there is no formal structure or tested basis for this rationale. In the SESSF system, the choice of whether to place a stock in tier 3 or 4 is made by the relevant assessment group. The elements of expert judgement within tier systems may help to achieve an expectation of risk equivalency, but potentially make the process somewhat subjective and hence impossible to simulation testing using MSE.

The tier system for the USA west coast groundfish fishery has been in place for eight years. A major advantage of that system is that there is a clear separation between the roles of the SSC (reviewing and approving assessments, assigning assessments to categories, and specifying sigma) and that of the Council (selecting the value of P^*, and choosing an ACL below the ABC). As such, management and policy is explicitly separated from the scientific advisory processes (Field et al., 2006). In contrast, it is impossible to separate the quantification of uncertainty from risk tolerance in the way the buffers for the SESSF and the ICES systems, as well as the buffer for tier 5 USA Alaskan crab stocks, are constructed.

The tier systems implement the expectation of risk equivalency through addressing uncertainty, but few discuss bias (i.e. that the expected value for biomass and hence the catch limit differs from the true values), although this certainly occurs in practice. For example, in the USA west coast system, account is taken of between-species variation in the CVs of biomass estimates, but bias is only addressed partially by assigning stocks for which the full assessment is 'less reliable' to category 2 (Table 8). The bias in the estimates from assessments may differ among species, and its direction and size may be impossible to pre-determine without case-specific simulation testing.

It is generally recognized that MSE is the best-practice approach for comparing management systems (Punt et al., in press). Of the various case studies, the SESSF is the furthest towards conducting MSE studies for all its tiers and for several species. Ideally, MSE could be conducted for all species and when each species is placed in each tier. This would permit decision makers the opportunity to select the level of monitoring to maximize return given a risk criterion. However, this will be computationally very intensive for systems (such as the USA west coast groundfish fishery and ICES systems) that have many species. A generic MSE system may be very useful here. In addition, having a different choice of control variables for each species would lead to a very unwieldy system. There would therefore be value in selecting values for control variables for groups of species, or applying simple rules such as that for the USA west coast groundfish fishery, which sets sigma to 0.36 for tier 1 stocks unless the estimate from the assessment is higher. The values for the control parameters would be set so that performance for some measure of risk, such as that the probability of stock staying above the limit biomass level at least 90% of the time, is the same among tiers. Figure 8 shows a generic example of the qualitative implications of this for the risk-cost-catch trade-off. In Figure 8, there are measures of risk for three possible tiers. The default HCRs for each tier lead to quite different measures of risk. Adjusting their control parameters so that risk is the same among tiers (symbols at the end of the arrows) allows a comparison of performance where the probability of dropping below the reference point (here set to 0.1) is equivalent among tiers. The choice of data to collect is then determined by the trade-off between cost and the amount of increased catch. In the context of Figure 8, for example, whether the increased cost associated with the tier corresponding to the diamond tier, compared to that associated with the square tier, is more than offset by the higher catch achieved by the diamond tier.

The results of this synthesis lead to several recommendations for how tier systems should be developed in the context of achieving risk equivalency:

- Tiers should not be defined simply on data availability, but also on the reliability of assessments used to estimate management quantities based on those data.
- The process for selecting control variables should clearly differentiate quantification of uncertainty, from how decision makers wish to address that uncertainty (risk

quantification versus risk tolerance and imposing a distinction between scientific uncertainty and additional uncertainty added by decisions makers).

- Risk equivalency is best tested by using MSE to select the values for control variables that determine the buffer given the uncertainty associated with the assessment.
- Basing values for control variables on a MSE analysis for each stock is ideal and recommended given potential biases and unknown consequences of additional rules within a harvest strategy (but may be computationally infeasible for regions with large numbers of stocks, or for stocks that are so data poor that a MSE is largely impossible to conduct).

5.2 Principal Component Analysis

The original four SESSF tiers were defined somewhat subjectively, and primarily in the context of assessment method, and data quality and availability at the time they were developed (Smith and Smith, 2005). Here, these attributes were captured by scoring data types according to their availability and quality. The scorings were applied to species across all Australian Commonwealth fisheries within an expanded 8-level tier system, and account for data availability by assigning zeros to those data types for which there is no information.

The PCA showed that the highest and lowest tier levels were clearly segregated by the first two principal components. The remaining, intermediate, tiers showed some level of ordered separation. Having expanded the original four tiers to eight, asking experts to consider information availability and quality yielded some independent, post-hoc evaluation of these criteria as tier delineators.

The clear delineation of the "data rich" tiers 0 and 1, and the more data-limited tier 7 along the first principal component axis, was more associated with the availability of catch and effort data than with that of fishery independent survey data. The importance of catch and effort data relative to that of fishery independent survey data is interesting, given that the latter are often considered the gold standard of high-quality stock assessments.

In interpreting the results, however, it should be noted that almost half of the Australian stocks/species (19 of 52) are assigned to a single tier (tier 4). Tiers 1 and 2 have 12 and 7 stocks, respectively, while tiers 0, 3, 6 and 7 each have 5 or fewer stocks. This unbalanced distribution of stocks within tiers could affect the ability to analytically delineate tiers on the basis of only data quality and availability. Additionally, other clustering techniques, such as a k-medoids approach, could render the relationships somewhat differently.

Despite the somewhat ordered trend by tier number along the first principal component axis, the extent of overlap, particularly for the intermediate tiers, highlights the possible ambiguities in defining data quality. Even if data quality could be explicitly defined, there remains a high propensity for overlap between tiers. For example, a tier 4 assessment (e.g., Little et al., 2011; Prince et al., 2011) may be based on high quality CPUE and/or length-composition information, as may a model-based tier 2 stock assessment. Australia's ETBF assessments are classified as tier 4 not because of a lack of data "quality", but because the mobility of the species is such that local fishing activities do not embrace the range of the stock.

More generally, the PCA demonstrated that many tiers are not consistent (in terms of data availability and quality) across fisheries. In this context, data availability and quality do not appear to reflect tiers. This is perhaps illustrated most strongly within the Northern Prawn Fishery, where blue endeavour prawns (*Metapenaeus endeavouri*) have identical scores to brown and grooved tiger prawns (*Penaeus esculentus* and *P. semisulcatus*), yet the tiger prawns are tier 0 and blue endeavour prawns are tier 2. Species biology and the value of the fishery are also relevant factors, in addition to data availability and quality, when determining the type of assessment and hence assigning a tier level to a species or fishery. Complementary arguments apply in the SESSF where many tier 4 species have some age and length information, but it is not used. If the tiers do not consistently reflect the data available then data availability and quality alone are not sufficient to identify which tier a species will fall into within Australian Commonwealth fisheries.

A broader issue is the need to firmly establish the intent or aim of tier systems. Historically, the Australian tier system originated within the SESSF (Smith et al., 2008), with the aim of providing recommendations of catch for all stocks managed using Total Allowable Catches, irrespective of the data available for assessment purposes. While precautionary adjustments to catch ("buffers" or "discount factors") have been set with the intent of achieving risk equivalency across the tiers, this was not, at least initially, a formal consideration (the need for risk equivalency was discussed in Smith and Smith (2005), but this was not formally adopted).

Finally, clarifying the definition of "risk" in the context of tier systems is still required. Tiers in the USA west coast groundfish fishery and the Alaskan crab fishery account *inter alia* for the ability of the assessment to estimate management quantities (Dichmont et al., 2016). Aligning overfishing risk with assessment certainty (and hence the ability to define a probability of overfishing) is arguably more direct and explicit than the assumption that data quality and availability are directly related to this risk. Nonetheless, our analysis provides some post-hoc justification for the Australian tier delineations in terms of whether the tiers discriminate stocks (given the way assessments are conducted and decision rules applied). Confronting an expanded 8-tier system, which embraces all Commonwealth species, with expert judgement of data quality in the context of assessment assumptions, provided an objective evaluation for defining tiers on this basis.

5.3 Monetary economic analysis

Fisheries management agencies are constrained by costs. These include not only the direct management costs of monitoring and assessing a stock, which are easily seen as a principle demand on an operating budget, but also the episodic costs of rebuilding stocks if this is needed. These episodic costs represent a risk, and management agencies that seek to minimize management costs in totality, must address such risk.

The risk to a management agency however, is not only that the stock will fall below a limit and require rebuilding efforts, but also that the stock will be perceived to fall below it. These can be quite different (Little et al., 2014). The cost of a stock that is correctly seen to have crossed a limit is related to the probability of correctly estimating the stock status. Our analysis showed that as assessment precision increases, or the management target ($\tilde{D}$) increases away from the lower limit declines, and with it the associated cost.

A stock may also be incorrectly perceived to have crossed a limit (false positive error), or it may have crossed a limit without being perceived to have done so (false negative error). These errors also have potential costs if a management agency unnecessarily spends money on rebuilding measures (PFMC, 2014), or unwittingly misses an opportunity to detect dangerously low stock levels. Failure to detect or act on the detection of dangerously low stock levels, may have contributed to the collapse of Northern Cod, *Gadus morhua* (Walters and Maguire, 1996). Our analysis showed how these costs might be affected by changing the target biomass, or by adjusting the precision of the biomass estimate through monitoring and assessment. Both of these options have consequences, however. Increasing the target biomass beyond that associated with maximum economic yield would likely reduce catches and thus economic production, while increasing the precision of the estimate of biomass, would likely increase the direct management costs of monitoring and assessment. We showed the aggregate trade-offs in the component costs of the cost function, and how they can be balanced to minimize the overall expected cost.

Implicit in this balancing is the weight placed on the different components, particularly the economic and the conservation imperatives. These weights will differ even between fisheries with similar management arrangements (e.g. Grafton et al., 2007), and ultimately will depend on the objective of management. For example, fisheries with a strong economic motivation, and sensitive to the risk of lost catches and economic returns, might focus more on setting the target biomass to minimize long-term expected costs, rather than efforts to increase assessment precision. In such fisheries, stocks with high variability are expected to be targeted at lower levels as more weight is placed on the risk of lost economic production than on being periodically overfished. Such fisheries would have little need for increasing the precision of the assessment as this would add to management costs, and so might forgo annual assessments and implement multi-year TACs (e.g. Smith et al., 2008). Alternatively, fisheries that place a high emphasis on the risk of overfishing a stock, such as those with Marine Stewardship Council accreditation (Gulbrandsen, 2009), would undoubtedly place greater emphasis on monitoring and assessment.

In general, the risk-catch-cost frontier posits that high risk, high catch stocks should also result in high cost fisheries management, and as management moves away from high risk and high catch conditions, the costs should correspondingly decline (Sainsbury, 2005). The exception to this occurs when a stock is already considered to be overfished. Such a fishery would typically have low or nil catches, but high risks, and high costs as rebuilding and recovery efforts are implemented (Dowling et al., 2013). We captured the risk-catch-cost relationship in a single scalar monetary value, with risk being represented by the expected cost associated with the terms; catch represented by the revenue, which was affected through the choice of management target, and cost represented by the operational management costs associated with monitoring and assessment, and affected through the choice of assessment precision.

The model and analysis was a relatively simple representation of fishery conditions, with little reliance on specific fishery data. Such an approach comes with drawbacks, and assumptions. First is that the model and results relied on comparative statics, and measuring the long-term effects of management policy on equilibrium conditions. Such an approach does not consider rate or path of the fishery to equilibrium, or even the possibility that equilibrium conditions may not be achieved (Anderson and Cavendish, 2001). For example, we assumed that fishery economic productivity was related to the equilibrium surplus production, irrespective of a current or initial state of a fishery. Thus, the results should not be applied to any specific fishery, but instead provide a guideline to management agencies that are considering reducing operational costs associated with monitoring and assessment. The general conclusion is that short-term action to reduce management costs could potentially have unintended consequences, by increasing the long-term costs of dealing with an over-fished stock, or an apparent over-fished stock.

Applying this analysis to a specific fishery would require a stochastic dynamic simulation approach that considers not only fishery-specific parameter values, but also the current stock state, or perceived stock state. Such an approach could calculate the present value of expected cost and thus explicitly consider time and path dependency, and relate revenue to a harvest level derived from a harvest control rule, which itself would depend on estimated biomass. Although, untangling these effects could be addressed using dynamic simulation models and management strategy evaluation (e.g. Cooke, 1999; Dichmont et al., 2008; Fay et al., 2011), the details of conducting the projections could obscure the general principles identified in the approach of this paper.

Another less obvious assumption is the independent relationship between the assessment variability, and both the true underlying stock variability as well as biomass. An alternative result might be that as biomass declines, the assessment variability does too. If this happens, it would become easier to determine whether the stock crosses the limit as it declined, and thus a lower target biomass might reduce risk. It would be correspondingly more difficult to detect when the stock crossed the threshold if assessment variability increased as the stock biomass declined, and the optimal risk strategy might be to set higher target levels.

Another important assumption we have made is that reducing assessment variability (σ) will cost more. The relationship we calculated was strongly influenced by three of the data points. This however, was a first pass; more detailed cost data are needed to explore this further.

The uncertainty associated with estimating stock status is also multi-faceted (Francis and Shotton, 1997; Fulton et al., 2011), because increased expenditure on monitoring might reduce observation uncertainty, but not the model uncertainty associated with assessment, for several reasons. First, mismatch between the life history or ecology of a stock and the assumed population dynamics might make obtaining an accurate, unbiased estimate of stock status difficult; or second, catch-per-unit-effort (CPUE) might be assumed to be linearly related to biomass, but instead exhibit hyper-stability, or third a stock might be assumed to be homogenously distributed in a stock assessment model, but in reality form a meta-population.

As a result, it might not always be possible to reduce σ with increased expenditure. Nevertheless, monitoring and assessment remain critical activities for managing fish stocks, because they provide information for setting control variables such as TAC or total allowable effort (TAE). In response to confidence in data and analysis used to estimate biomass, management agencies have started to invoke tiered level management, with the intention that TAC or TAE recommendations are tempered by the risk associated with the consequence of errors. In the Australian SESSF, a risk premium is attached to TAC recommendations resulting from catch curve or CPUE analysis to reflect the greater uncertainty, lower precision, and lower amounts of data used by these methods (Smith et al., 2014). Risk premiums have been found to be stock-specific (Fay et al., 2012), and methods to accurately represent the risk between tier levels are currently being developed (Little et al., 2014).

Whether the financial costs of either monitoring or assessment could be better used on other measures of protection and conservation (McDonald-Madden et al., 2010; Legg and Nagy, 2006) or are cost effective (Boyce et al., 2012) are important management questions. Trade-off analyses on marine ecosystems have been examined (Fulton et al., 2014), but have typically focused on the mean or expected cost outcome without consideration of the more extreme outcomes that could eventuate. The application of value at risk (VaR) approaches (Sethi et al., 2012) to measure the extremes of the cost distribution would provide a broader perspective than the focus on expected values used here. Nevertheless, management budgets are typically based on expectation, and by applying the rules presented here, a management agency can adjust either the biomass target, or their investment in data collection, to understand the larger picture in minimizing the combined risks and associated costs of fisheries management.

5.4 Atlantis analysis: all tiers

5.4.1 Comparison with other MSE studies

The results in terms of the R1 performance metric are similar to those from single species MSEs, where the tier 1 to 4 HSs perform well with respect to recovering a resource to above $0.2B_0$, although tier 3 performed worse that tier 4 (Fay et al., 2012; Little et al., 2014). The species-specific nature of their results was also highlighted by these previous studies.

Some species (e.g. gulper shark and gemfish) did not recover. There are several reasons for this: a) species are included in this study that started the projection period below the LRP, and their longevity and life history characteristics means that, even with good management, these species would not necessarily recover by the end of the projection period (e.g. gulper shark); b) environmental conditions and trophic interactions are included in Atlantis-RCC, but not in single-species MSEs, and c) implementation uncertainty (Fulton et al., 2011) is overtly accounted for through a multi-species fleet dynamics module and a quota trading module in Atlantis-RCC.

The various combinations of data, assessment methods and associated decision rules (i.e. HSs) to set a management measure, such as a TAC, in the EU framework were simulation tested for risk equivalency across categories 1 to 4 by ICES (2013). That work found that the impact of the meta-rules was similar to this paper's findings: the rules were useful when a resource was recovering or in a healthy condition, but tended to slow down action to stop a declining resource. As such, meta-rules are very context specific, especially relative to stock status and should be used with care (ICES, 2013) and an appreciation that the effects will vary through time and with stock status.

5.5 Atlantis analysis: SESSF and buffers

There is a need for buffers if the aim of the management system is to prevent stocks from being depleted to below the limit reference point. However, the analyses suggest that basing the size of the buffer on the uncertainty associated with the assessment (the USA west coast buffer approach) rather than on the assessment method applied (the Australian system) is more effective. The improved performance of the USA west coast system in terms of avoiding risk may, however, be more related to that the magnitude of the buffers (i.e., the HCRs are typically more conservative). Thus, the performance of the Australian approach may have matched that of the USA west coast approach if the buffer values were smaller (i.e., their ability to reduce the RBC was greater).

All systems led to low probabilities of stocks being below the limit reference points (except for gemfish and gulper shark for which rebuilding was essentially impossible in 45 years). Overall, however, none of the systems achieved complete risk equivalency. Although the USA west coast approach to setting buffers came close, appearing to be most able to achieve risk equivalency in relation to the probability of having half of the stocks above the target reference point across all tiers. This result also highlights that 'risk equivalency' relies on a definition of 'risk'. That is, achieving risk equivalency in terms of one performance metric (e.g., the LRP) will not necessarily lead to such achievement for other performance metrics (e.g., the TRP).

Constraints on catch variation (i.e. meta-rules) lead, as expected, to less variation in catch, even though the catch variation constraints are not particularly strict compared to those applied in other jurisdictions, such as South Africa (e.g., Plaganyi et al., 2007). The meta-rules will tend to lower the rate at which catches are reduced when stocks are in need of rebuilding, but will also reduce the instances of unrealistically large increases in catch. Furthermore, highly variable assessments (as a result of species variability or poor data quality) also lead to a larger effect of having meta-rules in place, as they damp the resulting variability in RBCs. Thus, the results of this study suggest that the benefits of meta-rules are likely case-specific.

The tier 1 harvest control rule outperformed the more data-poor harvest control rules in terms of allowing stocks to rebuild towards the limit and target reference points, albeit at a cost in terms of short-term catch (and variation in catch) – see supplementary material, Figure E.S3. Whether the reduced risk is worthwhile given the costs associated with conducting full stock assessments is beyond the scope of this paper, but should be taken into account when monitoring systems are developed.

5.6 Proposed application

This work highlights several questions – particularly which tier system to use, the buffers at each tier and how to develop a system that maintains species specific flexibility. Refer to Appendix F.

6 Conclusions

6.1 Tier systems review

The review reveals that best practice is not to define tiers simply on data availability, but also on what the assessments based on those data are capable of estimating. In addition, clearly differentiating the quantification of uncertainty from how decision makers wish to address that uncertainty would simplify justification of buffers (the gap between the assessment-produced target catch or effort and the final management decision that accounts for uncertainty and risk). Risk equivalency can be achieved by using management strategy evaluation to select the values for control variables, which determine the buffer given the uncertainty associated with the assessment.

6.2 Principal Component Analysis

Multivariate analysis indicated that the eight tiers delineated between the extreme tier levels on the first principal component, although there was overlap for intermediate tiers. More generally, it is important that the aim of tier systems and the basis for tier delineations are explicitly defined given the increasing association of tiers with trade-offs between overfishing risk, management cost and catch.

6.3 Monetary economic analysis

A conservation-oriented objective would minimize risk by increasing target biomass levels, or expenditure on monitoring and assessment, while a profit-focused objective would seek to lower management costs by reducing expenditure on data collection and assessment. Biomass levels targeted by management depends on the variability associated with the 26 stock biomass, but would typically minimize risk at intermediate levels associated with maximum yield. As the variability associated with a stock increases however, and the ability to make a meaningful estimate declines, profit-motivated fisheries would reduce targets to increase catches. The basic theory example provides the basis for the more extensive risk analyses in the MSE, and serves as a simple lesson that the consequences of reducing the immediate costs associated with managing a fishery, can come with a concomitant increase in overall risk.

6.4 Atlantis analysis: all tiers

The Australian multi-tier HS system does not, according to our results, achieve risk equivalency, especially for data-limited approaches. Profitability is lowest for the more data-limited approaches, despite lowest harvesting costs. The data-rich HSs benefit from decreases in opportunity costs, although there is not a consistent result among tiers. Tier 3 stands out by being less precautionary compared to tier 4 and therefore is out of sequence compared to tier 1 and 4 (if higher tier numbers are meant to designate more risky HSs). Profitability increases as more data-rich HSs are used. The meta-rules mitigate large unjustified TAC increases of the data-limited approaches; generally leading to a more precautionary approach when applied to a recovering stock, but not when applied to a declining one. As currently formulated/specified, the data-limited HSs are slow to respond to a stock decline or depletion; either because of their lack of conservatism when estimating key parameters (e.g. for tier 5 variants) and due to their complexity in the case of hierarchical approaches (for tiers 6 and 7). Overall, if risk, cost and catch are considered, the data–rich approaches would be favoured.

6.5 Atlantis analysis: SESSF and buffers

Harvest strategies for all tiers were capable of moving productive stocks so their biomasses lay between the limit and target reference points. The USA buffer system was more conservative than the SESSF system, and achieved the fastest recovery time for depleted stocks. This system led to slightly lower total catches, but was closest to achieving risk equivalency across the tiers. The USA buffer system led to biomass trajectories most similar to those when the system was managed so that biomass moves as rapidly as possible to its target reference point.

6.6 Proposed application

Appendix F provides the next step needed to operationalise the risk-cost-catch trade-off

7 References

AFMA 2009. Harvest Strategy Framework: for the Southern and eastern scalefish and shark fishery 2009 (Amended February 2014). Australian Fisheries Management Authority, Canberra, Australia.

AFMA 2010. AFMA cost recovery impact statement. Australian Fisheries Management Authority, Canberra. http://www.afma.gov.au/wp-content/uploads/2014/04/AFMA-Cost-Recovery-Impact-Statement-20101.pdf

Anon. 2014. Harvest strategy framework for the southern and eastern Scalefish and shark fishery. Report to Australian Fisheries Management Authority, Canberra. 25 pp. http://www.afma.gov.au/wp-content/uploads/2014/11/Harvest-Strategy-Framework-SESSF-Feb-2014.pdf (last opened 1 November 2015).

Anderson, D. and Cavendish W. 2001. Dynamic simulation and environmental policy analysis: beyond comparative statics and the environmental Kuznets curve. Oxford Economic Papers 53: 721-746.

Arnason, R., Hannesson, R. and W.E. Schrank, W.E. 2000. Cost of fisheries management: the cases of Iceland, Norway and Newfoundland. Marine Policy **24**, 233-243.

Bergh, M. O., and Butterworth, D. S. 1987. Towards rational harvesting of the South African anchovy considering survey imprecise and recruitment variability. South African Journal of Marine Science 5: 937–951.

Boyce, M. S., Baxter, P. W. J., and Possingham, H. P. 2012. Managing moose by the seat of your pants. Theoretical Population Biology 82: 340-347.

Butler A., Harris P., Lyne V., Heap A., Parslow V., and Porter-Smith R. 2001. An Interim Bioregionalisation for the continental slope and deeper waters of the South-East Marine Region of Australia. http://www.environment.gov.au/resource/interim-bioregionalisation-continental-slope-and-deeper-waters-south-east-marine-region (last accessed 1 September 2015).

Butterworth, D. S., Cochrane, K. L., and De Oliveira, J. A. A. 1997. Management procedures: a betterway to manage fisheries? The South African experience. American Fisheries Society Symposium Series, 20. In Global Trends: Fisheries Management, pp. 83–90. Eds by E. K. Pikitch, D.D. Huppert, and M.P. Sissenwine. American Fisheries Society, Seattle. 352 pp.

Butterworth, D.S., Bentley, N., De Oliveira, J.A.A., Donovan, G.P., Kell, L.T., Parma, A.M., Punt, A.E., Sainsbury, K.J., Smith, A.D.M., and Stokes, T.K. 2010. Purported flaws in management strategy evaluation: basic problems or misinterpretations? ICES J. Mar. Sci. 67, 567–574.

Clark, W. G. 1991. Groundfish exploitation rates based on life history parameters. Canadian Journal of Fisheries of Aquatic Sciences 48: 734-750.

Clark, W.G. 2002. $F_{35\%}$ revisited ten years later. North American Journal of Fisheries Management 22: 251–257.

Cooke, J.G. 1999. Improvement of fishery-management advice through simulation testing of harvest algorithms. ICES journal of Marine Science 56: 797-810.

Cox, A. 2000. Cost recovery in fisheries management: the Australian experience. IIFET 2000 Proceedings.

DAFF 2007. Commonwealth Fisheries Harvest Strategy Policy Guidelines. Australian Government Department of Agriculture, Fisheries and Forestry, Canberra, Australia, pp. 55. http://www.agriculture.gov.au/fisheries/domestic/harvest_strategy_policy

DAFF 2013. Report on the review of the Commonwealth Fisheries Harvest Strategy Policy and Guidelines, Department of Agriculture, Fisheries and Forestry, Canberra, Australia.

Day, J. 2010. School whiting (*Sillago flindersi*) stock assessment based on data up to 2008. In: G.N. Tuck [Ed.] Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2009, Part 1. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart, pp 190-249.

Day, J., Klaer, N., and Tuck, G. N. 2013. Silver Warehou (*Seriolella punctata*) stock assessment based on data up to 2013. In Stock Assessment for the Southern and Eastern Scalefish and Sharkfish: 2012, Part 1, Chapter 7, pp. 120–151. Ed. by G.N. Tuck. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 210 pp. http://www.afma.gov.au/wp-content/uploads/2010/06/SESSF-2013-Part-1-Final.pdf (last accessed 1 November 2015).

Day, J., and Klaer, N. 2014. Tiger flathead (*Neoplatycephalus richardsoni*) stock assessment based on data up to 2012. In Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2013, Part 1, Chapter 10, pp 174–232. Ed. by G.N. Tuck. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 210 pp.

De la Mare, W. K. 1996. Some recent developments in the management of marine living resources. In Frontiers of Population Ecology. pp. 599–616. Eds by R.B. Floyd, A.W. Shepherd, and P.J. De Barro. CSIRO Publishing, Melbourne, Australia.

Deroba, J. J., and Bence, J. J. 2008. A review of harvest policies: understanding relative performance of control rules. Fisheries Research 94: 210–223.

Dichmont , C.M. Fulton, E. Gorton, R. Sporcic, M. Little, L.R. Punt, A.E. Dowling, N., Haddon, M., Klaer, N., and Smith, D.C. From data rich to data-limited harvest strategies – does more data mean better management? ICES Journal of Marine Science; in prep.

Dichmont, C. M., Deng, A., Punt, A. E., Ellis, N., Venables, W. N., Kompas, T., Ye, Y., Zhou, S., and Bishop, J. 2008. Beyond biological performance measures in Management Strategy Evaluation: Bringing in economics and the effects of trawling on the benthos. Fisheries Research 94: 238–250.

Dichmont, C. M., Deng, R., Punt, A. E., Venables, W., and Haddon, M. 2006. Management strategies for short lived species: The case of Australia's Northern Prawn Fishery. 1. Accounting for multiple species, spatial structure and implementation uncertainty when evaluating risk. Fisheries Research, 82: 204–220.

Dichmont, C.M., Punt, A.E., Dowling, N., De Oliveira, J.A.A., Little, L.R., Sporcic, M., Fulton, E., Gorton, R., Klaer, N., Haddon, M., Smith, D.C., 2016. Is risk consistent across tier-based harvest control rule management systems? A comparison of four case studies. Fish and Fish. doi: 10.1111/faf.12142.

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., Punt, A. E., Little, L. R., Dichmont, C., Sporcic, M., Fulton, E .A., Gorton, R. J., Haddon, M., and Smith, D. C. 2016. Assessing a multilevel tier system: the role and implications of data quality and availability. Fisheries Research, 183, 588–593. http://dx.doi.org/10.1016/j.fishres.2016.05.001.

Dowling, N.A., Smith, D.C., Knuckey, I., Smith, A.D.M., Domaschenz, P., Patterson, H.M., and Whitelaw, W. 2008. Developing harvest strategies for low value and data-poor fisheries: case studies from three Australian fisheries. Fisheries Research, 94: 390-390.

European Commission. 2013. Regulation (EU) No 1380/2013 of the European Parliament and of the Council of 11 December 2013 on the Common Fisheries Policy, amending Council Regulations (EC) No 1954/2003 and (EC) No 1224/2009 and repealing Council Regulations (EC) No 2371/2002 and (EC) No 639/2004 and Council Decision 2004/585/EC. (OJ L354/22 28.12.2013, pp. 40. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:354:0022:0061:EN:PDF

European Commission. 2013. The Common Fisheries Policy (CFP). http://ec.europa.eu/fisheries/cfp/index_en.htm (last accessed 1 November 2015).

Fay, G., Punt, A .E., and Smith, A. D. M. 2009. Operating model specifications. In Evaluation of New Harvest Strategies for SESSF Species. pp. 125–133. Ed. by S.E. Wayte. CSIRO Marine and Atmospheric Research, Hobart and Australian Fisheries Management Authority, Canberra

Fay, G., Punt, A. E., and Smith, A. D. M. 2011. Impacts of spatial uncertainty on the performance of age structure-based harvest strategies for blue-eye trevalla (*Hyperoglyphe antarctica*). Fisheries Research, 110; 391–407.

Fay, G., Little, L. R., Tuck, G. N., Haddon, M., and Klaer, N. L. 2012. Maintaining risk equivalency among fishery harvest control rules in Southeast Australia. CSIRO Marine and Atmospheric Research, Hobart. 31 pp.

Field, J.C., Punt, A.E., Methot, R.D., and Thomson, C.J. 2006. Does MPA mean "Major Problem for Assessments?" Considering the consequences of place-based management systems. Fish and Fisheries **7**, 284–302.

Field, S. A., Tyre, A., Jonzén, N., Rhodes, J. R., and Possingham, H. P. 2004. Minimizing the cost of environmental management decisions by optimizing statistical thresholds. Ecology Letters 7: 669-675.

Francis, R. I. C. C., and Shotton, R. 1997. "Risk" in fisheries management: a review. Canadian Journal of Fisheries and Aquatic Sciences 54: 1699-1715.

Fulton E. A., Smith, A. D. M., and Smith, D. C. 2007. Alternative management strategies for Southeastern Australian Commonwealth Fisheries: Stage 2: Quantitative Management Strategy Evaluation. Report to the Australian Fisheries Management Authority and the Fisheries Research and Development Corporation. CSIRO Marine and Atmospheric Research. http://atlantis.cmar.csiro.au/www/en/atlantis/mainColumnParagraphs/02/text_files/file/AMS_Final_Report_v6.pdf (last accessed 3 November 2015).

Fulton E.A., Smith, A.D.M., and Smith, D.C., 2007. Alternative management strategies for Southeastern Australian Commonwealth Fisheries: Stage 2: Quantitative Management Strategy Evaluation. Report to the Australian Fisheries Management Authority and the Fisheries Research and Development Corporation. CSIRO Marine and Atmospheric Research.

Fulton, E. A. Smith A. D. M., Smith D. C., and Johnson P. 2014. An Integrated Approach Is Needed for Ecosystem Based Fisheries Management: Insights from Ecosystem-level Management Strategy Evaluation. PLoS One:e84242. DOI: 10.1371/journal.pone.0084242

Fulton, E. A., and Gorton, R. 2014. Adaptive Futures for SE Australian Fisheries & Aquaculture: Climate Adaptation Simulations. CSIRO, Australia. pp 309. http://frdc.com.au/research/Final_Reports/2010-023-DLD.pdf (last accessed 3 November 2015).

Fulton, E. A., Link, J., Kaplan, I. C., Johnson, P., Savina-Rolland, M., Ainsworth, C., Horne, P., Gorton, R., Gamble, R. J., Smith, T., and Smith D. 2011. Lessons in modelling and management of marine ecosystems: The Atlantis experience. Fish and Fisheries, 12: 171–188

Fulton, E. A., Smith, A. D. M., Smith, D. C., and van Putten, I .E. 2011. Human behavior: the key source of uncertainty in fisheries management. Fish and Fisheries 12, 2–17.

Fulton, E.A., and Gorton, R., 2014. Adaptive Futures for SE Australian Fisheries & Aquaculture: Climate Adaptation Simulations. CSIRO Climate Adaptation Flagship, Australia.

Georgeson, L., Ward, P., and Curtotti, R., 2015. Southern and Eastern Scalefish and Shark Fishery. In Fishery status reports 2015, Chapter 8, pp. 112-127. Ed. by H. Patterson, L. Georgeson, I. Stobutzki, and R. Curtotti. Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra. 487 pp. http://data.daff.gov.au/data/warehouse/9aam/fsrXXd9abm_/fsr15d9abm_20151030/08_FishStatus2015SthnEastnScalefishShark_1.0.0.pdf (last accessed 1 November 2015).

Grafton, R. Q., Kompas, T., McLoughlin, R., and Rayns, N. 2007. Benchmarking for fisheries governance. Marine Policy 31: 470-479.

Gulbrandsen, L. H. 2009. The emergence and effectiveness of the Marine Stewardship Council. Marine Policy 33: 654-660.

Haddon, M., Klaer, N., Smith, D.C., Dichmont, C.D. and Smith, A.D.M. 2012. Technical reviews for the Commonwealth harvest Strategy Policy. FRDC Report 2012/225. CSIRO, Hobart.

Haddon, M., Klaer, N., Wayte, S., and Tuck, G., 2015. Options for Tier 5 approaches in the SESSF and identification of when data support for harvest strategies are inappropriate. Report to Fisheries Research and Development Corporation. CSIRO Oceans and Atmosphere, Hobart. pp. 115.

Hilborn, R. 2010. Pretty Good Yield and exploited fishes. Marine Policy 34: 193-196.

ICES. 2012. ICES Implementation of Advice for Data-limited Stocks in 2012 in its 2012 Advice. ICES DLS Guidance Report 2012. ICES CM 2012/ACOM 68, 42 pp. http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2012/ADHOC/DLS%20Guidance%20Report%202012.pdf (last accessed 3 November 2015).

ICES. 2013. Report of the Workshop on the Development of Quantitative Assessment Methodologies based on LIFE-history traits, exploitation characteristics, and other key parameters for Data-limited Stocks (WKLIFE III). ICES CM 2013/ACOM:35. pp. 84. http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2013/WKLIFE3/Report%20WKILFE%20III.pdf. (last accessed 1 November 2015).

ICES. 2014a. Report of the ICES Advisory Committee 2013. ICES Advice, 2014. Book 1, Section 1.2. [Accessed online, 02/03/2015: http://www.ices.dk/sites/pub/Publication%20Reports/Advice/2014/2014/1.2_Advice_basis_2014.pdf]

ICES. 2014b. Report of the Workshop on the Development of Quantitative Assessment Methodologies based on LIFE-history traits, exploitation characteristics, and other relevant parameters for data-limited stocks (WKLIFE IV), 27–31 October 2014, Lisbon, Portugal. ICES CM 2014/ACOM:54. 223 pp.

Interim Marine and Coastal Regionalisation for Australia Technical Group (IMCRA) 1998. Interim Marine and Coastal Regionalisation for Australia: an ecosystem-based classification for marine and coastal environments. Version 3.3. Environment Australia, Commonwealth Department of the Environment: Canberra. http://www.environment.gov.au/resource/interim-marine-and-coastal-regionalisation-australia-version-33 (last accessed 1 September 2015).

Jensen, A.L., 1996. Beverton and Holt life history invariants result from optimal tradeoff of reproduction and survival. Canadian Journal of Fisheries and Aquatic Science, 53: 820–822.

Johnson P., Fulton E. A., Smith D. C., Jenkins G. P., and Barret N. 2011. The use of telescoping spatial scales to capture inshore to slope dynamics in marine ecosystem modelling. Natural Resource Modeling, 24: 335–364

Klaer, N. L. Wayte, S. E., and Fay, G. 2012. An evaluation of the performance of a harvest strategy that uses an average-length-based assessment method. Fisheries Research, 134-136: 42–51.

Knuckey, I., Harvey, E., and Koopman, M. 2009. Industry survey to obtain a relative abundance index for spawning eastern gemfish - traditional and innovative methods. AFMA Project R2006/830. Fishwell Consulting 103pp. http://www.fishwell.com.au/projects/project,id,169.aspx

Kompas, T., Dichmont, C.M., Punt, A.E., Deng, A., Che, T.N., Bishop, J., Gooday, P., Ye, Y.M., and Zhou, S. Maximizing profits and conserving stocks in the Australian Northern Prawn Fishery. Australian Agricultural and Resource Economics. 54:281–299; 2010

Kramer-Schadt S., Revilla E., Wiegand T., and Grimm, V., 2007. Patterns for parameters in simulation models. Ecological Modeling, 204: 553–556.

Legg, C. J., and Nagy, L. 2006. Why most conservation monitoring is, but need not be, a waste of time. Journal of Environmental Management 78: 194-199.

Link, J., Burnett, J., Kostovick, P., and J. Galbraith. 2008. Value-added sampling for fishery independent surveys: Don’t stop after you’re done counting and measuring. Fisheries Research 93, 229-233.

Little L.R., Punt, A. E., Mapstone, B. D., Pantus, F., Smith, A. D. M., Davies, C.R., and McDonald, A.D. 2007. ELFSim – A Model for Evaluating Management Options for Spatially-Structured Reef Fish Populations: An Illustration of the "Larval Subsidy" Effect. Ecological Modelling, 205: 381–396.

Little, L. R., Wayte, S. E., Tuck, G. N., Smith, A. D. M., Klaer, N., Haddon, M., Punt, A. E., Thomson, R., Day, J., and 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.

Little L. R., Parslow, J., Fay, G., Grafton, R. Q., Smith A. D. M., Punt, A. E., and Tuck G.N. 2014. Environmental derivatives, risk analysis, and conservation management. Conservation Letters 7: 196-207.

Little, L.R., Punt, A.E., Dichmont, C.M., Dowling, N., Smith, D.C., Fulton, E., Sporcic, M., and Gorton, R.J., 2015 Defining options for cost constrained fisheries management. ICES J. Marine Science. doi:10.1093/icesjms/fsv206

Lyne V., and Hayes D., 2005. Pelagic Regionalisation. National Oceans Office and CSIRO Marine and Atmospheric Research, Hobart. http://www.environment.gov.au/system/files/resources/b7d7587a-6330-41fe-8f97-cba74ad87306/files/nmb-pelagic-report.pdf (last accessed 1 September 2015).

Mapstone, B. D. 1995. Scalable decision rules for environmental impact studies. Ecological Applications 5: 401-410.

McDonald-Madden, E., Baxter, P. W. J., Fuller, R. A., Martin, T. G., Game, E. T., Montambault, J., and Possingham, H. P. 2010. Monitoring does not always count. Trends in Ecology and Evolution 25: 547-550.

McDonald, A. D., Smith, A. D. M., Punt, A. E., Tuck, G. N., and Davidson, A. J. 1997. Empirical evaluation of expected returns from research on stock structure for determination of total allowable catch. Natural Resource Modelling 10: 3-29.

Methot, R. D., and Wetzell, C. R. 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142: 86–99.

Ministry of Fisheries 2008. Harvest strategy standard for New Zealand fisheries. New Zealand Government. http://fs.fish.govt.nz/Page.aspx?pk=104

Morison, A. K., Knuckey, I. A., Simpfendorfer, C. A., and Buckworth, R.C. 2012. 2011 Stock Assessment Summaries for the South East Scalefish and Shark Fishery. http://www.afma.gov.au/wp-content/uploads/2010/07/2011-SESSF-Species-Summaries-March-2012.pdf. (last accessed 3 November 2015).

National Marine Fisheries Service and U.S. Fish and Wildlife Service 2005. Recovery Plan for the Gulf of Maine Distinct Population Segment of Atlantic Salmon (*Salmo salar*). National Marine Fisheries Service, Silver Spring, MD.

North Pacific Fishery Management Council (NPFMC). 2014. Stock Assessment and Fishery Evaluation Report for the king and Tanner crab fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, 605 W. 4th Avenue, #306, Anchorage, AK 99501.

Oke P. R., Schiller A., and Griffin D. A. B. 2005. Ensemble data assimilation for an eddy-resolving ocean model. Quarterly Journal of Royal Meteorological Society, 131: 3301–3311.

Pacific Fishery Management Council (PFMC). 2014a. Status of the Pacific Coast Groundfish Fishery: Stock Assessment and Fishery Evaluation. Pacific Fishery Management Council, 7700 Ambassador Place, Suite 101, Portland, OR 97220.

Pacific Fishery Management Council (PFMC). 2014b. Terms of Reference for the groundfish and coastal pelagic species stock assessment review process for 2015-2016. Pacific Fishery Management Council, 7700 Ambassador Place, Suite 101, Portland, OR 97220.

Pantus F. J., and Dennison W. C. 2005. Quantifying and Evaluating Ecosystem Health: A Case Study from Moreton Bay, Australia. Environmental Management, 36: 757–771.

Pascoe, S.; Punt, A.E.; Dichmont, C.M. Targeting ability and output controls in Australia's multi-species Northern Prawn Fishery. Eur Rev Agric Econ. 37:313–334; 2010

Plagányi, É.E., Rademeyer, R.A., Butterworth, D.S., Cunningham, C.L., Johnston, S.J., 2007. Making management procedures operational - innovations implemented in South Africa. ICES J. Mar. Sci. 64, 626–632.

Prager, M.H. and Shertzer, K.W. 2010. Deriving Acceptable Biological Catch from the Overfishing Limit: Implications for Assessment Models. North American Journal of Fisheries Management **30**, 289–294.

Prince, J.D., Dowling, N.A., Davies, C.R., Campbell, R.A., Kolody, D.S., 2011. A simple cost-effective and scale-less empirical approach to harvest strategies. ICES J. Mar. Sci. 68, 947–960.

Punt, A. E. 1992. Management procedures for Cape hake and baleen whale resources. BEP Report No 23. pp. 689.

Punt, A. E., and Smith, A. D. M. 1999. Harvest strategy evaluation for the eastern stock of gemfish (*Rexea solandri*). ICES Journal of Marine Science, 56: 860-875.

Punt, A.E., Smith, A.D.M., 1999. Management of long-lived marine resources: A comparison of feedback-control management procedures. p. 243-265. In: J.A. Musick [Ed.] Life in the slow lane: Ecology and conservation of long-lived marine animals. American Fisheries Society Symposium 23, Bethesda, MD.

Punt, A. E., Campbell, R., and Smith, A. D. M. 2001. Evaluating empirical indicators and reference points for fisheries management: Application to the broadbill swordfish fishery off Eastern Australia. Marine and Freshwater Research, 52: 819-832.

Punt, A. E., Smith, A. D. M., and Cui, G. 2002. Evaluation of management tools for Australia's South East Fishery. 3. Towards selecting appropriate harvest strategies. Marine and Freshwater Research, 53: 645–660.

Punt, A.E., Deng, R.A., Pascoe, S., Dichmont, C.M., Zhou, S., Plagányi, É.E., Hutton, T., Venables, W.N., Kenyon, R., and van der Velde, T. 2011. Calculating optimal effort and catch trajectories for multiple species modelled using a mix of size-structured, delay-difference and biomass dynamics models. Fisheries Research, 109: 201-211.

Punt, A.E., Siddeek, M.S.M., Garber-Yonts, B., Dalton, M., Rugolo, L., Stram, D., Turnock, B.J. and Zheng, J. 2012. Evaluating the impact of buffers to account for scientific uncertainty when setting TACs: Application to red king crab in Bristol Bay, Alaska. ICES J. Mar. Sci. 69, 624–634.

Punt, A. E., Smith, A. D. M., Smith, D. C., Tuck, G. N., and Klaer, N .L., 2014a. Selecting relative abundance proxies for B_{MSY} and B_{MEY}. ICES Journal of Marine Science 71: 469-483.

Punt, A.E., Szuwalski, C. and Stockhausen, W 2014b. An evaluation of stock–recruitment proxies and environmental change points for implementing the USA Sustainable Fisheries Act. Fisheries Research 157, 28–40.

Punt, A. E., Smith, D. C., Haddon, M., Russell, S., Tuck, G. N., and Ryan, T. 2015. Estimating the dynamics of spawning aggregations using biological and fisheries data. Marine and Freshwater Research 66: 1–15.

Punt, A. E., Butterworth, D. S., de Moor, C. L., De Oliveira, J. A. A., and Haddon, M. 2016. Management Strategy Evaluation: Best Practices. Fish and Fisheries, 17(2): 303–334. Doi: 10.1111/faf.12104.

Ralston, S., Punt, A. E., Hamel, O. S., DeVore, J. D., and Conser, R. J. 2011. A meta-analytic approach to quantifying scientific uncertainty in stock assessments. Fishery Bulletin 109: 217-231.

Ralston, S., Punt, A.E., Hamel, O.S., DeVore, J., Conser, R.J., 2011. An approach to quantifying scientific uncertainty in stock assessment. Fish. Bull. 109, 217–231.

Rayns, N. 2007. The Australian government's harvest strategy policy. ICES Journal of Marine Science 64: 596-598.

Sainsbury, K. J., Punt, A. E., and Smith, A. D. M. 2000. Design of operational management strategies for achieving fishery ecosystem objectives. ICES Journal of Marine Science, 57**:** 731–741.

Sainsbury K. 2005. Cost-effective management of uncertainty in fisheries. Outlook 2005. Canberra, A.C.T.: Australian Bureau of Agricultural and Resource Economics. http://data.daff.gov.au/data/warehouse/pe_abarebrs99001173/PC13024.pdf (last accessed 7 September 2015).

Savina M., Fulton E. A., Condie S., Forrest R., Scandol J., Astles K. and Gibbs P. 2008. Ecologically sustainable development of the regional marine and estuarine resources of NSW: Modelling of the NSW continental shelf ecosystem. CSIRO and NSW DPI Report. http://www.dpi.nsw.gov.au/research/areas/aquatic-ecosystems/outputs/2009/1196 (last accessed 7 September 2015).

Sethi, S.A. 2010. Risk management for fisheries. Fish and Fisheries 11: 341-365.

Sethi, S.A., Dalton, M., and Hilborn, R. 2012. Quantitative risk measures applied to Alaskan commercial fisheries. Canadian Journal of Fisheries and Aquatic Sciences 69, 1-12.

Shertzer, C.E., Prager, M.H. and Williams, E.H. 2008. A probability-based approach to setting annual catch levels. Fish. Bull. 206, 225–232.

Smith, A.D.M. 1993. Risks of over- and under-fishing new resources. In: Risk Evaluation and Biological Reference Points for Fisheries Management, 120. (eds. Smith, S.J., Hunt, J.J., Rivard, D.), Canadian Special Publication of Fisheries and Aquatic Sciences, pp. 261–267.

Smith, A. D. M., Sainsbury, K. J., and Stevens, R. A. 1999. Implementing effective fisheries-management systems—management strategy evaluation and the Australian partnership approach. ICES Journal of Marine Science, 56: 967–979.

Smith, D.C., Smith, A.D.M. and Punt, A.E. 2001. Approach and process for stock assessment in the south east fishery: A perspective. Marine Freshwater Research 52, 671–681.

Smith, A., Smith, D., 2005. A harvest strategy framework for the SESSF. Report to the Australian Fisheries Management Authority (AFMA), Canberra, June 2005, Available from CSIRO Oceans and Atmosphere, GPO Box 1508, Hobart, Tasmania, Australia. 7pp.

Smith, A. D. M., Fulton, E. J., Hobday, A. J., Smith, D. C., Shoulder, P. 2007. Scientific tools to support the practical implementation of ecosystem-based fisheries management. ICES Journal of Marine Science., 64(4): 633-639. doi: 10.1093/icesjms/fsm041.

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., and Little, L. 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.A., Sainsbury, K.J., and Sloan, S. R. 2014. Implementing harvest strategies in Australia: 5 years on. ICES Journal of Marine Science 71: 195-203.

Stobutzki, I., Vieira, S., Ward, P., and Noriega, R., 2011. Southern and Eastern Scalefish and Shark Fishery overview. In: Fishery Status Reports 2010: Stautus of fish stocks and fisheries managed by the Australian Government (eds Woodhams, J., Stobutzzki, I., Vieier, S., Curtotti, R., and Begg, G.A.) Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra.

Tuck, G. N. 2014. Stock assessment of blue grenadier (*Macruronus novaezelandiae*) based on data up to 2012. In Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2013, Part 1. pp 61–115. Ed. By G.N. Tuck. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research. 210 pp. http://www.afma.gov.au/wp-content/uploads/2010/06/SESSF-2013-Part-1-Final.pdf (last accessed 3 November 2015).

USA Department of Commerce (USA Doc). (2007) Magnuson-Stevens Fishery Conservation and Management Act as Amended Through January 12, 2007. http://www.nmfs.noaa.gov/sfa/magact/MSA_Amended_2007%20.pdf (last accessed 18 July 2014).

Walters, C., and Maguire, J.-J. 1996. Lessons for stock assessment from the northern cod collapse. Reviews in Fish Biology and Fisheries 6: 125-137.

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

Whitten, A.E. and Punt, A.E. 2014. Stock Assessment of pink ling (*Genypterus blacodes*) using data up to 2012. p. 116-142. In: Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2013, Part 1. (ed. G.N. Tuck). Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart.

Woodhams, J., Stobutzki, I., Vieira, S., Curtotti, R. and Begg, G.A. 2011. Fishery Status reports 2010: status of fish stocks and fisheries managed by the Australian Government, Australian Bureau of Agriculture and Resource Economics and Sciences, Canberra, Australia.

Zhou, S., Smith, A. D. M., and Fuller, M., 2011. Quantitative ecological risk assessment for fishing effects on diverse data-poor non-target species in a multi-sector and multi-gear fishery. Fisheries Research, 112: 168–178.

8 Implications

The risk-cost-catch trade-off is an important theoretical component of fisheries management. However:

- In most cases, risk is constrained by some biological bottom line such as a Limit Reference Point. This means that cost and catch can only be traded in the risk space where the resource will not become overfished.
- In practice, the risk-cost-catch trade-off is complicated by
    - Uncertainty (BOTH variance and bias) in the assessment and associated harvest strategy; and
    - The amount of social discounting of management costs and catch (profit).
- One of the biggest costs is lost opportunity costs – the cost of not managing the resource well – which is often the hardest costs to calculate.
- One of the smallest costs is the cost of monitoring, assessing and managing the target species of the fishery.
- Results are very species specific, but generally the more data limited the harvest strategy the higher the bias (especially) and uncertainty. The bias often results in slower response time to a biological issue. Therefore, buffers should be applied as a default, unless simulation testing (such as Management Strategy Evaluation) have shown otherwise.
- Meta-rules that affect how much a TAC changes from one year to the next are complicated in that they can be context specific, particularly related to where the species sits in terms of stock status – if it was over-exploited but recovering, meta-rules slow TAC increases to the benefit of long-term profit and costs, where it slows the recovery of species that are declining.
- The performance of data-limited assessments is highly dependent on their specific articulation, and on the appropriate setting of proxy reference points - and the latter is a circular problem.
- The current tier systems are methods-based, and these do not always relate to how well a specific management quantity is estimated. This has contributed to the species-specific results.

9 Recommendations

1. The tier system should not be based solely on the assessment method or available data, but should instead be based on how well that method performs for that species in current circumstances. This would be more similar to some of the USA tier systems,
2. Undertaking the most quantitative assessment and associated Harvest Strategy possible has benefits in terms of improved management performance through:
   a) reduced risk of missing the Target Reference Point and/or of being below the Limit Reference Point;
   b) improved likelihood of the recovery for depleted stocks in the short to medium term;
   c) improved ecosystem signature; and
   d) reduced opportunity costs in medium to long term,
3. It is important to consider lost opportunity costs, because the MSE tests showed that choosing not to undertake the more data intensive tiers (if these were available) resulted in orders of magnitude greater lost opportunity costs than the cost of doing the assessment,
4. Buffers (aka. discount factors) based on the Ralston method (i.e. explicitly taking assessment uncertainty into account, with a buffer in place for tier 1) have an improved chance of meeting management objectives (though at the cost of some lost catch in the short term).
5. Some incentive for the inclusion of fishery independent data, which is considered the global gold standard of stock assessment, could be considered through a smaller or zero buffer for tier 0,
6. Use meta-rules with care as their benefits and limitations are asymmetric (i.e. they can delay recovery even if they slow depletion),
7. Details matter and the default use of methods must be done with extreme care. This was highlighted when a standard tier 4 harvest strategy was compared to the species specific tuned tier 4,
8. Bias is as important as the variance,
9. The less data intensive tiers may work, but they need to be implemented with care as their complexity and insensitivity means they may take a significant time to detect that a stock status is declining or in poor health,
10. SAFE is a useful method, but like many "data poor" methods must itself be used in a sophisticated way. For example, modelling results suggest that it may work well for chondrichthyans, but be more challenged for short lived and/or highly variable teleosts; and
11. One size does not fit all:
    a) System and species specific information should be used where ever possible to make sure what is being employed makes sense.
    b) A single tier structure will not suit all species within the one fishery, let alone all species across all fisheries,
    c) The proposed buffers may need to be modified for fishery or species specific circumstances, but rigorous methods must be used to substantiate these changes,

d) MSE testing of harvest strategies is a desirable means of allowing for individual treatment of harvest strategies.

9.1 Further development

1. An Appendix on how best to start operationalising the risk-cost-catch trade-off forms the basis for further development, if required, for the new and present HSs by AFMA and/or the DA. At present, it is aligned to the existing HSP and would need to be adapted to the new HSP, when finalised.
2. A new tier system - with extensive input from stakeholders and in the context of the new Harvest Strategy Policy (therefore beyond the scope of this project) - is required that places all the Harvest Strategies together into a combined set. The Dowling et al. (2011) approach is a good approach, but tends to still concentrate on data type and assessment method. That of the USA crab fishery showed that concentrating on what the Harvest Strategy can estimate in terms of the Harvest Policy requirements works much better, especially when adding new methods into the tier system.
3. Methods to apply a buffer system to a broad range of harvest strategy tiers are required. The best approach is that of Ralston et al. (2011), but this is more applicable to the high and medium information approaches.

10 Extension and Adoption

The project team presented to the South-east Shark and Scale Fishery Resource Assessment Group and to the Tropical Tuna Resource Assessment Group its findings on several occassions during the project. It also presented to several Policy discussions fora within AFMA and DA. Appendix F was co-written with AFMA staff and with much input from several AFMA meetings that included a broad group of AFMA staff. Many of the findings are being input into the management of the SESSF and also into the new draft Commonwealth Harvesy Strategy Policy. The work was also presented at the 2016 World Fisheries Congress. All of the major components have also been published in peer reviewed journals.

11 Project materials developed

11.1 Journal publications

1. Dichmont, C.M., Punt, A.E., Dowling, N., De Oliveira, J.A.A., Little, L.R., Sporcic, M., Fulton, E., Gorton, R., Klaer, N., Haddon, M., Smith, D.C. 2016. Is risk consistent across tier-based harvest control rule management systems? A comparison of four case studies. Fish and Fisheries. 17(3): 731-747. doi: 10.1111/faf.12142.
2. Little, R.L., Punt, A.E., Dichmont, C.M., Dowling, N., Smith, D.S., Fulton, E.A. 2016. Decision trade-offs for cost-constrained fisheries management. ICES Journal of Marine Science. 73(2): 494-502. doi: 10.1093/icesjms/fsv206
3. Dowling, N.A., Punt, A.E., Little, L.R., Dichmont, C.M., Smith, D.C., Haddon, M., Sporcic, M., Fulton, E., Gorton, R.J. 2016. Assessing a multilevel tier system: the role and implications of data quality and availability. Fisheries Research, 183, 588–593. http://dx.doi.org/10.1016/j.fishres.2016.05.001.
4. Dichmont, C.M., Fulton, E., Gorton, R., Sporcic, M., Dowling, N., Little, R.L., Punt, A.E., Haddon, M., Smith, D.S. submitted. From data rich to data-limited harvest strategies – does more data mean better management? ICES J. Mar. Sci.
5. Fulton, E.A., Punt, A.E., Dichmont, C.M., Gorton, R., Sprocic, M., Dowling, N., little, L.R. Haddon, M., Klaer, N., Smith, D.C. 2016. Developing risk equivalent data-rich and data-limited harvest strategies. Fisheries Research. 183: 574-587. doi: 10.1016/j.fishres.2016.07.004.
6. Dichmont, C.M., Fulton, E., Gorton, R., Sporcic, M., Dowling, N., Little, R.L., Punt, A.E., Haddon, M., Smith, D.S. accepted. From data rich to data-limited harvest strategies – does more data mean better management? ICES J. Mar. Sci.

11.2 Conference presentations

7. Dichmont, C.M., Fulton, E., Gorton, R., Sporcic, M., Dowling, N., Little, R.L., Punt, A.E., Haddon, M., Smith, D.S. accepted. From data rich to data-limited harvest strategies – does more data mean better management? World Fisheries Congress, 23-27 May 2016, Busan, South Korea.
8. Invited keynote: Dichmont, C.M., Fulton, E., Gorton, R., Sporcic, M., Dowling, N., Little, R.L., Punt, A.E., Haddon, M., Smith, D.S. accepted. From data rich to data-limited harvest strategies – does more data mean better management? DRuMFISH Data-limited Workshop, 6-8 September, Barcelona, Spain,

Appendices

Appendix A. Do management systems based on tiers of harvest control rules achieve risk equivalency? A comparison of four case studies

Catherine M. Dichmont[1], André E. Punt[2,3], Natalie Dowling[2], José A.A. De Oliveira[4], Lorne R. Little[2], Miriana Sporcic[2], Elizabeth Fulton[2], Rebecca Gorton[2], Neil Klaer[2], Malcolm Haddon[2], David C. Smith[2]

[1]*CSIRO Oceans and Atmosphere, Ecosciences Precinct, Dutton Park, QLD, 4750, Australia*

[2]*CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart, TAS, 7001, Australia*

[3] *School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, WA 98195, USA*

[4]*CEFAS Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk, NR33 0HT, United Kingdom*

Corresponding authors details: Dichmont, Cathy; 41 Boggo Road, Dutton Park, 4750, Queensland, Australia; Phone: +61 7 3833 5629; *Cathy.dichmont@csiro.au*

Published in Dichmont, C.M., Punt, A.E., Dowling, N., De Oliveira, J.A.A., Little, L.R., Sporcic, M., Fulton, E., Gorton, R., Klaer, N., Haddon, M., Smith, D.C. 2016. Is risk consistent across tier-based harvest control rule management systems? A comparison of four case studies. Fish and Fisheries. 17(3): 731-747. doi: 10.1111/faf.12142.

Abstract

There can be substantial differences in data quality and quantity among fished species. Consequently, the quality and type of assessments can also vary substantially. However, all species, especially those that are targeted, need to be managed. Several jurisdictions have developed hierarchical tier systems that categorise stocks based on, for example, the data available for assessment purposes and/or the extent to which quantities on which management advice is based can be estimated. Four case studies (Australia's Southern and Eastern Scalefish and Shark Fishery, the USA west coast groundfish fishery, the USA Alaskan crab fishery, and EU fisheries) are used to contrast the types of hierarchical tier systems available, and to assess the extent to which each system constrains risk to be equivalent among the tiers (termed risk equivalency). Only the Australian system explicitly aims to achieve risk equivalency. However, this intent has not been fully operationalised. Our review reveals that best practice is not to define tiers simply on data availability, but also on what the assessments based on those data are capable of estimating. In addition, clearly differentiating the quantification of uncertainty from how decision makers wish to address that uncertainty would simplify justification of buffers (the gap between the assessment-produced target catch or effort and the final management decision that accounts for uncertainty and risk). Risk equivalency can be achieved by using management strategy evaluation to select the values for control variables, which determine the buffer given the uncertainty associated with the assessment.

Keywords

Data limited, Data rich, Fisheries management, Management Strategy Evaluation, Precaution, Risk trade-offs.

Introduction

In resource management, risk can be defined in several ways, but is often (as here) taken to be the probability of a resource being overfished i.e. falling below a biomass limit reference point. This risk for a given stock is directly related to its productivity and the amount of catch taken from it. Therefore, fishery managers are faced with the challenge of managing fisheries at biologically sustainable levels, while maximising economic returns to the community and minimising management costs. This trade-off has been described as the risk-cost-catch frontier (Sainsbury, 2005; Dowling et al., 2013). Effectively, this means trading off cost or catch for a given level of risk. However, the effect of management and thus management costs on the risk of overfishing is neither clear nor well documented. The belief is that as management activity and costs increase, the risk associated with being overfished declines, but long-term harvests may increase more slowly than the rate at which costs increase (Sainsbury, 2005).

Management activities and costs pertaining to harvest or management strategies comprise three components: monitoring and data collection, assessment or analysis of the data, and the decision or control exerted by management based on the analysis. The first component is costly, as it includes collection of fishery-dependent and/or -independent data, length/weight and length/age data, life history information, or data on broader ecosystem effects (Link et al., 2008). A range of methods to analyse monitored data are available, each with different costs; typically, assessments that structure populations by age, sex or length are considered the most costly, while mean length or catch-per-unit-effort (CPUE) based assessments, catch curves or simple empirical triggers are progressively less so.

A consequence of using different monitoring and assessment methods is varying levels of confidence in the scientific advice for decision making. For example, in Australia's Southern and Eastern Scalefish and Shark Fishery (SESSF), estimates of stock abundance, in both absolute terms and relative to reference points, from statistical catch-at-age stock assessments, are assumed to be more accurate and precise than estimates from catch curves or CPUE (Smith et al., 2008) because they use a broader suite of data and life history information. In contrast, CPUE and catch curve analyses are thought to make more risky assumptions, and thus use of their results is anticipated to result in a higher probability of overfishing (Smith et al., 2014). The need to provide management advice for all stocks (DAFF, 2007), irrespective of the amount of data available, combined with the variability in confidence among assessment methods has led to the development of a hierarchy of tiered assessments, with the more data-limited tiers assumed to be riskier. The management decisions that result (usually target levels of catch or effort) from higher risk, but presumably less costly, assessments, are often adjusted to reflect greater caution, thus acknowledging the higher risk inherently associated with data-poor assessments, and encouraging data collection efforts that would move a stock to higher tiers.

Management costs associated with obtaining the information needed to assess and manage fisheries, conduct research, run the management decision process and ensure compliance are generally thought to be low compared to the benefits conveyed (Dowling et al., 2013). In Australia, the cost associated with managing Commonwealth fisheries was around 7% of the value of the fisheries in 1998-99 (Cox, 2000), but was about 4-5% in the past few years (AFMA, pers. commn). The issue is complicated, however, by who pays. Cost recovery is commonly argued on the grounds that the beneficiary should pay, and it is fisheries policy in several countries or jurisdictions that costs are passed to industry through levies or cost-recovery programs (Arnason et al., 2000). In such cases, there is commonly pressure to reduce research and management costs. Alternatively, where the public is seen as the beneficiary of fisheries management, and the information is thus considered to contribute to the public good management costs are viewed as a public cost (Arnason et al., 2000), although pressure to reduce costs may occur indirectly. There are also options which combine both cost recovery and government funding (public good) as in Australia (AFMA, 2010), whereas in the USA and Europe, costs are generally borne by government, except that in some USA fisheries there is also a levy applied and industry can contribute voluntarily to surveys.

The overarching policies governing fisheries management explicitly or implicitly address the risk-cost-catch trade-off, but few jurisdictions operationalise it directly in their management systems. What all jurisdictions share, explicitly or otherwise, is an aim for risk equivalency (or reducing the risk when data are poorer). Risk equivalency means having a common probability of stocks falling below the limit reference point. This can be achieved by adjusting catch limits downwards if these are determined using data limited methods. The extent of downward adjustment (to, for example, target levels of catch or effort) is variously referred to as the buffer or the discount factor. The parameters that determine the extent of downward adjustment will here be referred to as control variables.

In this paper, we address whether there is some element of risk being equivalent among tiers within tier-based management approaches, which is the expectation for the tier system in Australia (Smith et al., 2014). In Australia the expectation of risk equivalency is implemented so that each tier would include a buffer to reduce catch or effort, with the value of the control variables determining the buffer selected so that the risk of over-exploitation associated with managing under each tier is the same (Smith et al., 2014; Punt et al., 2014a).

Tier systems have only been formally developed and implemented in three jurisdictions globally (Australia's Commonwealth, the USA Federal system, and ICES). We consider four case studies to describe how the trade-offs between risk, management costs, and catch can be addressed. The cases include the well-established tier systems of Australia's SESSF, the USA Alaskan crab fisheries, the USA west coast groundfish fishery, and the recently developed tier system of the European fisheries for which the International Council for the Exploration of the Sea (ICES) provides advice. The way assessment/management tiers in various jurisdictions are established has to be viewed in the appropriate policy context. We therefore first provide the policy frameworks that underlie the tier systems for each case study (Section 2) and then outline how each policy structure has been implemented (Section 3). Section 4 evaluates the extent to which the tier structures aim to achieve risk equivalency across tiers, and Section 5 draws conclusions and provides recommendations, with a focus on Australia, for achieving the expectation of risk equivalency.

Policy structure of alternative tier structures

USA federal system

In the USA federal system, the risk-cost-catch trade-off and the implementation of this trade-off into a tier system for assessments and harvest control rules (HCRs) is well defined for managers, in order to achieve the goals of the Magnuson-Stevens Fishery Management and Conservation Act, as reflected by its National Standards (USA Doc, 2007). These standards recognise that there is a trade-off between conservation and utilization, but, as written, conservation takes precedence over minimizing impacts on fishing communities. The standards also state that management measures must be based on best available scientific information (National Standard 2) and this is achieved by having a well-structured and transparent peer-review system. In particular, recommendations for management decisions for federally-managed species in the USA are made by the Regional Fishery Management Councils based on scientific advice reviewed and approved by their specific Scientific and Statistical Committee (SSC).

Each Council has a slightly different structure for the development and review of stock assessments and hence the provision of scientific management advice. Nevertheless, the advice and decision-making frameworks are broadly similar among regions of the USA. The key scientific inputs into the decision making process for each stock are ideally:

- The Overfishing Level (OFL) – the catch that corresponds to a fishing mortality from a Maximum Sustainable Yield (F_{MSY}) control rule – for several regions around the US, F_{MSY} (or a proxy thereof) is the target fishing mortality rate.
- The Acceptable Biological Catch (ABC) – a catch that is less than or equal to the OFL to account for scientific uncertainty. The difference between the OFL and ABC for each stock is referred to as the 'buffer'.
- Whether the stock is subject to overfishing – variously defined as the catch exceeding the OFL or the fishing mortality rate exceeding that corresponding to the control rule for the OFL, F_{OFL}.
- Whether the stock is in an overfished state – defined as stock size being below the Minimum Stock Size Threshold (MSST). MSST is defined to be between B_{MSY} (the stock biomass corresponding to MSY) and $0.5B_{MSY}$.

Councils use this information, along with other inputs, such as comments from the public on the impacts of various catch levels, to select an Annual Catch Level, ACL (which cannot exceed the ABC) and the optimum yield. The difference between ACL and the ABC can reflect a variety of policy considerations such as social, economic and other ecological factors, as well as increasing the rate at which overfished stocks rebuild to the target biomass and the uncertainty associated with implementing catch limits. The difference between the OFL and ABC is based on uncertainty that then forms the basis of the tier system.

Australian Commonwealth system

Australian Commonwealth fisheries are managed by the Australian Fisheries Management Authority (AFMA) under the Fisheries Management Act 1991 and in accordance with the Environment Protection and Biodiversity Conservation Act 1999. The Fisheries Harvest Strategy Policy and Guidelines (DAFF, 2007) relates to key commercial species targeted by AFMA-managed fisheries. The policy defines target and limit reference points, and, when these cannot be estimated, their proxy values. The target reference point for AFMA-managed fisheries is the Maximum Economic Yield (MEY), which is unusual because many other fisheries' jurisdictions use B_{MSY} as a target. No targeted fishing for a species is allowed when its stock size is at or below a limit reference point, which is usually set at 20% of the unexploited stock size. Risk is defined in the Harvest Policy, mainly for the limit reference point, in that management should "ensure that the stock stays above the limit biomass level at least 90% of the time". There is a tiered system to address data and assessment types and link these to risk equivalency among tiers.

European (and ICES) system

The European Union's (EU) Common Fisheries Policy (CFP) is the principal legal mechanism for managing fish stocks in EU waters, and regulates all aspects of fishing within the EU. The CFP has the overall objective of ensuring economically, environmentally and socially sustainable use of fishing and aquaculture resources (European Commission, 2013). The CFP was first implemented in 1983, and has undergone a number of reforms. The latest of these (which came into effect in January 2014) specifies that the management of all species that are subject to quotas should be on the basis of MSY by 2020 at the latest. It also introduces a phased discard ban to fully cover all EU fisheries by 2019, allows regionalisation of fisheries and management so as to better suit local conditions, and addresses fleet over-capacity by insisting that capacity matches fishing opportunities for member states. ICES is the intergovernmental body that develops science and advice on marine ecosystems and associated fisheries. In the context of this paper, ICES co-ordinates and conducts stock assessments and provides stock status advice to the EU and other ICES member countries based on a Memorandum of Understanding.

ICES currently provides advice based on an MSY policy for those stocks without agreed, precautionary management plans but for which there is a basis for MSY advice, using a HCR that, in principle, should conform to the hockey stick (aka. slope) rule (i.e. fishing mortality is a linear function of biomass, until a trigger level above which it is constant and independent of the biomass) with the trigger at F_{MSY} and $B_{trigger}$ (the latter representing the lower bound of spawning biomass fluctuations around B_{MSY}). That is, the aim is ultimately to harvest these stocks at F_{MSY} (ICES, 2014a). Where there is an agreed, precautionary management plan, ICES will base its advice on such a plan, and where there is no basis for MSY advice or no agreed, precautionary management plan, ICES advice will be based on precautionary principles (ICES, 2014a). The tier system used by ICES (described later) is applied within this context, and implicitly attempts to use different harvest strategies (data type, assessment method and decision rules) to obtain the expectation of risk equivalency among tiers.

Implementation of the tier systems

United States west coast groundfish fishery

Over 90 species are included in the Pacific Coast Fishery Management Plan, including rockfishes, roundfishes, sharks and skates (PFMC, 2014a).

The Pacific Fisheries Management Council Science Scientific and Statistical Committee (PFMC SSC) places each stock into one of three 'categories' (and one of 11 sub-categories) (Table A.1) depending on the method and reliability of the assessment. Specifically, stocks assessed using data-rich methods may be considered data-moderate if, for example, the assessment is dated, the assessment did not estimate annual deviations in recruitment about the stock recruitment relationship, or the assessment results were sensitive to plausible changes to the assumptions. Of 144 stocks for which OFLs are available, 21 stocks are currently in category 1 (the most data-rich), 29 stocks are in category 2, and the remaining stocks (94) are in category 3 (the most data-limited) (PFMC, 2014a). Not all stocks are managed as individual species; several stocks are instead managed as complexes under the assumption the stocks within a complex have roughly the same productivity and vulnerability (PFMC, 2014a)). Fig. A.1a shows the generic HCRs for setting OFLs and ABCs for USA west coast groundfish (noting that B_{MSY} cannot be estimated reliability for any west coast groundfish stock (PFMC, 2014b) so proxies are used instead).

Table A. 1 Tier assignments for USA west coast groundfish assessments (reproduced from PFMC [2014b]). Note: the USA uses the term "category" and the numerical order is reversed (but not the alphabetical sub-tiers order) compared to the Australian and ICES systems. A consistent terminology is used in this paper for ease of comparison.

Tier Number		Description	Example stock
Tier 1: Data rich	aa	Reliable compositional (age and/or size) data sufficient to resolve year-class strength and growth characteristics. Only fishery-dependent trend information available. Age/size structured assessment model.	Black rockfish (*Sebastes melanops;* Sebastidae)
	b	As in 1a, but trend information also available from surveys. Age/size structured assessment model.	Canary rockfish (*Sebastes pinniger;* Sebastidae)
	c	Age/size structured assessment model with reliable estimation of the stock-recruit relationship.	None
Tier 2: Data moderate	a	Natural mortality multiplied by survey biomass.	None
	b	Historical catches, fishery-dependent trend information only. An aggregate population model is fit to the available information.	None
	c	Historical catches, survey trend information, or at least one absolute abundance estimate. An aggregate population model is fit to the available information.	None
	d	Full age-structured assessment, but the results are substantially more uncertain than assessments used in the calculation of the σ used to compute the buffer for category 1 stocks (Ralston et al., 2001). Reasons for placing a stocks in this category include that assessment results are very sensitive to model and data assumptions, or that the assessment has not been updated for many years.	Shortbelly rockfish(*Sebastes jordani;* Sebastidae)
Tier 3: Data poor	a	No reliable catch history. No basis for establishing OFL.	Black and yellow rockfish (*Sebastes chrysomelas;* Sebastidae)
	b	Reliable catch estimates only for recent years. OFL is the average catch during a period when the stock is considered to be stable and close to B_{MSY} equilibrium on the basis of expert judgment.	Pacific Grenadier (*Coryphaenoides acrolepis;* Macrouridae)
	c	Reliable aggregate catches during period of fishery development and approximate values for natural mortality. Default analytical approach Depletion-Corrected Average Catch.	Cowcod south of Pt Conception (*Sebastes levis;* Sebastidae)
	d	Reliable annual historical catches and approximate values for natural mortality and age at 50% maturity. Default analytical approach Depletion-based Stock-Reduction Analysis.	Kelp rockfish (*Sebastes atroviren;* Sebastidae)

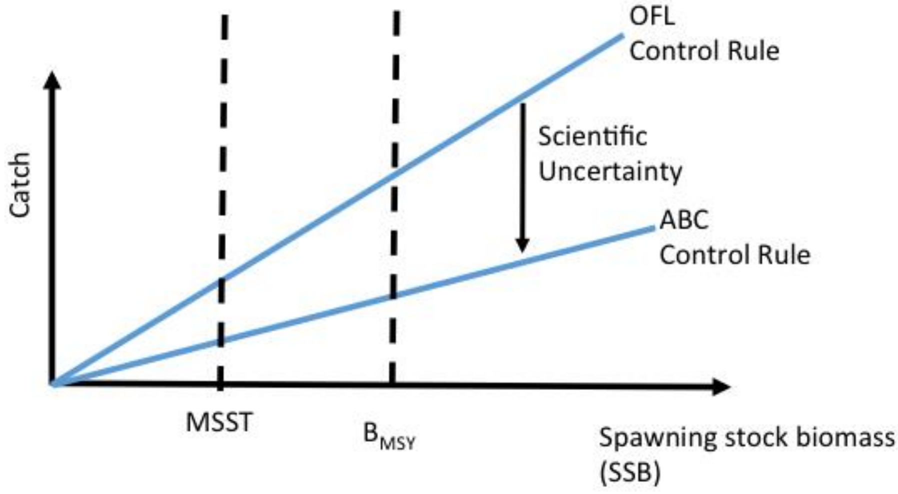


Figure A.1(a)

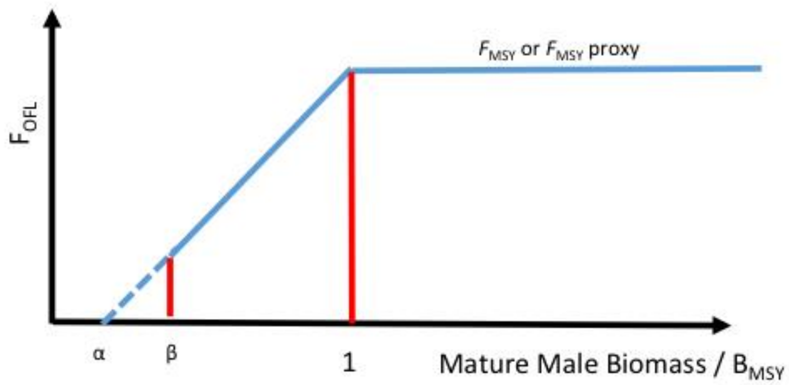


Figure A.1(b)

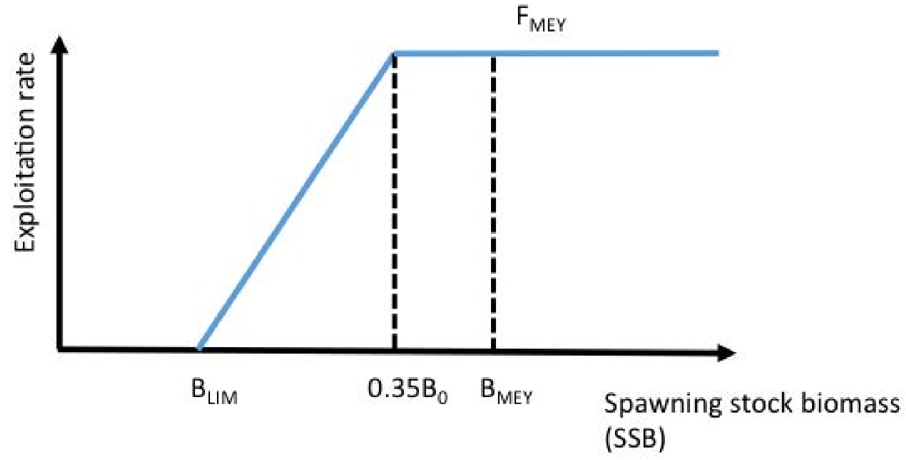


Figure A.1©

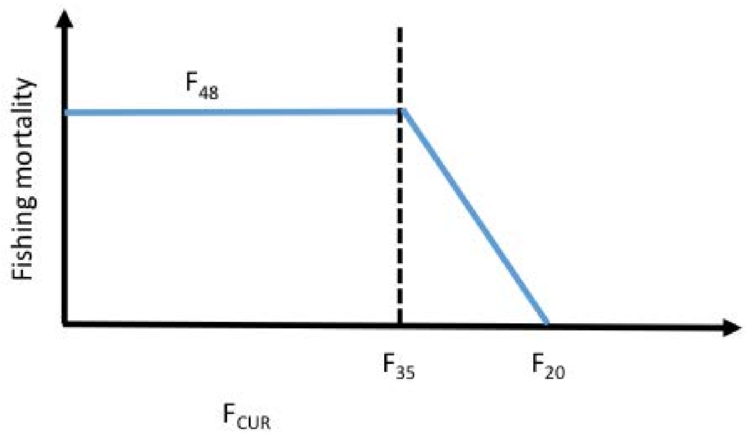


Figure A.1(d)

Figure A. 1a: Harvest control rules for the USA west coast groundfish fishery; b) Alaska crab fisheries; c) The tier 1 control rule for the Australian SESSF; d) The tier 3 control rule for the Australian SESSF. Terms are: Acceptable Biological Catch (ABC), Biomass (B) at Maximum Sustainable Yield (MSY), Current (CUR), Fishing mortality (F), Maximum Economic Yield (MEY), Minimum Stock Size Threshold (MSST), Overfishing Limit (OFL), Spawning Stock Biomass (SSB), Scientific Uncertainty is the total of P* and σ as described in the text and unfished biomass (B0).

The buffer between the OFL and the ABC reflects the extent of scientific uncertainty, which differs among categories. The buffer is a function of the extent of uncertainty and the risk tolerance given uncertainty (i.e., $p(ABC(\hat{OFL}) > OFL) = P^*$; Prager and Shertzer, 2010; Punt et al., 2012; Shertzer et al., 2008).

This is implemented for the west coast groundfish fishery by the PFMC SSC quantifying uncertainty in terms of coefficient of variation for the OFL (σ) and the Council selecting the percentile of the resulting probability distribution at which to set the ABC (P^*). P^* is therefore the probability of overfishing occurring. For data-rich (category 1) stocks σ is 0.36, unless the coefficient of variation of the estimated most-recent year biomass from the assessment exceeds this (PFMC, 2014a). This value is based on a meta-analysis of between-assessment variation in biomass estimates (Ralston et al., 2011). By contrast, the values for σ for data-moderate and data-poor stocks are 0.72 and 1.44 respectively. The value of P^* is selected by the Council, which has agreed that it will not exceed 0.45 for any stock (legally, it must be 0.5 or less). The default choice for P^* is 0.45 for category 1 stocks and 0.4 for category 2 and 3 stocks. Therefore, for the USA west coast groundfish fishery, σ and P^* are the control variables.

United States Alaskan crab

Alaska's 10 crab stocks under federal jurisdiction are jointly managed by the North Pacific Fishery Management Council (NPFMC) and the Alaska Department of Fish and Game. The NPFMC makes recommendations for the OFL and the ABC for each crab stock, while the State of Alaska makes decisions on total allowable catches (TACs), which have to be less than the ABCs. OFLs are set using a tier system that includes five tiers depending on data availability and the ability to estimate key stock assessment parameters (Table A.2).

Table A. 2 Tier assignments for the 10 Alaskan crab stocks under federal management (NPFMC, 2014). Note: the USA uses the term "category".

Tier number	Description	Example
1	Stocks with assessments which provide reliable estimates of biomass and B_{MSY}, and in which a probability density function (pdf) for F_{MSY} is estimated.	None
2	Stocks with assessments which provide reliable estimates of biomass and B_{MSY} and F_{MSY}.	None
3	Stocks where reliable estimates of the spawner-recruit relationship are not available, but biomass can be estimated reliably and proxies for F_{MSY} and B_{MSY} can be estimated.	Eastern Bering Sea snow crab (*Chionoecetes opilio*; Oregoniidae) Bristol Bay red king crab (*Paralithodes camtschaticus*; Lithodidae) Eastern Bering Sea Tanner crab (*Chionoecetes bairdi*; Oregoniidae)
4	Stocks for which reliable estimates of biomass are available (directly from surveys or from assessments), but insufficient biological data are available to estimate $F_{x\%}$.	Pribilof Islands red king crab (*Paralithodes camtschaticus*; Lithodidae) Pribilof Islands blue king crab (*Paralithodes platypus*; Lithodidae) St. Matthew Island blue king crab (*Paralithodes platypus*; Lithodidae) Norton Sound red king crab (*Paralithodes camtschaticus*; Lithodidae)
5	Stocks which have no reliable estimates of biomass and only historical catch data is available.	Aleutian Islands golden king crab (*Lithodes aequispinus*; Lithodidae) Pribilof Island golden king crab (*Lithodes aequispinus*; Lithodidae) Adak red king crab (*Paralithodes camtschaticus*; Lithodidae)

No crab stocks are currently in tiers 1 and 2 (although Punt et al. [2014b] suggest it may be possible to estimate stock-recruitment relationships for some of the stocks in tier 3). The most data-rich crab stocks are therefore in tier 3, with a few stocks in tier 4. Three stocks are currently in tier 5.

Fig. A.1b shows the HCR used to set the OFL for Alaskan crab stocks in tiers 1 – 4. Note that unlike the OFL in the HCR for the west coast groundfish fishery (Fig. A.1a), fishing mortality rate is reduced when stock size is less than B_{MSY}. The OFL for stocks in tier 5 is set to the average catch over a period considered by the NPFMC's Crab Plan Team to correspond to B_{MSY}.

The buffer between the OFL and ABC is selected by the SSC of the NPFMC for each species based on suggestions from the NPFMC Crab Plan Team (e.g. NPFMC, 2014). The maximum possible ABC (MaxABC) is based on the P^* method (see Section 3.1) for stocks in tiers 1 to 4, and is 90% of the OFL for stocks in tier 5. P^* is assumed to be 0.49, while the variance of the OFL is set based on the results of the stock assessments. The uncertainty levels in the assessments are considered unrealistically small (Punt et al., 2012), and consequently, the buffer between the OFL and the MaxABC is much smaller for Alaskan crab stocks than for west coast groundfish stocks. However, the buffer imposed by the NPFMC SSC is larger than that implied by the agreed P^* value. For example, for the 2011/12 to 2013/14 management years, the ratio of the ABC to OFL for tier 3 and 4 stocks has ranged between 0.43 and 0.90 (NPFMC, 2014). Moreover, the TACs set by the State of Alaska for these stocks are often substantially smaller than the ABCs set by the NPFMC because the Alaskan State harvest strategies are often more precautionary than the federal rules used to set OFLs. From 2011/12 to 2013/14 the TACs for tier 3 and 4 stocks has ranged between 0 and 0.804 of the ABC set by the NPFMC.

Australia's SESSF

The SESSF tier system uses the ability to produce a reliable assessment based on the available data to define tiers, which are used to inform management decisions through the setting of the Recommended Biological Catch, RBC (Table A.3). The type of data available defines the method used for assessment and the associated HCR in this system. Tier 1 equates to a robust age-and sex-structured assessment to assess stock status (see stocks listed in Table A.3). It should be noted that, until recently, the SESSF did not have independent survey data and therefore such data did not influence the tier system. The generic structure of the HCR for tier 1 stocks is provided in Fig. A.1c. The target biomass is the biomass corresponding to Maximum Economic Yield (B_{MEY}) or a proxy thereof (the default proxy is 1.2 B_{MSY}). B_{MSY} is only estimated for tiger flathead (*Neoplatycephalus richardsoni*; Platycephalidae) (Punt et al., 2014a). For the remaining SESSF stocks, B_{MSY} is based on a proxy ($B_{\mathrm{MSY}} = 0.4\ B_0$). The limit reference point ($\mathrm{B_{LIM}}$) is $0.5B_{\mathrm{MSY}}$ (or the proxy $0.2B_0$). The target fishing mortality rate is set to that corresponding to F_{MEY} (or the proxy of F_{48}) and from $0.35B_0$, this rate declines linearly with biomass to zero at the limit reference point.

Tier 2 is similar to tier 1, except that the stock assessment is less robust. To account for this reduced reliability, the default Tier 2 B_{MSY} proxy is 0.5 B_0, and the B_{MEY} proxy is 0.6 B_0 (Smith et al., 2008). However, at present no species are assigned to this tier. Tier 3 applies to stocks where age (Wayte and Klaer, 2010) or length (Klaer et al., 2012) structure data are available (Fig. A.1d). Within tier 3, age- or length-based catch curve analyses are used to obtain the values required for the HCR (tiers 3a and 3b respectively in Table A.3). Tier 4 assessments use catch-per-unit-effort (CPUE) as a relative index of abundance (Little et al. 2011; see stocks listed in Table A.3). Tier 3 is assumed to provide more robust assessment results than Tier 4, but this was not supported by results from management strategy evaluations (Fay et al., 2012; Little et al. 2014).

Table A. 3 Tier assignments for the SESSF stocks under Australian Commonwealth management.

Tier Number	Description	Example stocks
1	Stocks with an available quantitative stock assessment, which provides estimates of biomass and fishing mortality in absolute terms and relative to agreed limit reference points.	Flathead (*Neoplatycephalus richardsoni*; Platycephalidae) Blue grenadier (*Macruronus novaezelandiae*; Merlucciidae) Pink ling (*Genypterus blacodes*; Ophidiidae) School whiting (*Sillago flindersi*; Sillaginidae) Orange roughy (*Hoplostethus atlanticus*; Trachichthyidae) Eastern gemfish (*Rexea solandri*; Gempylidae) Silver warehou (*Seriolella punctata*; Centrolophidae)
2	Stocks with a quantitative stock assessment but the assessment is considered less robust than a tier 1 assessment.	Not presently applied
3a (age-composition data)	Stocks for which age data are available for catch curve analysis (Wayte and Klaer, 2010).	Redfish (*Centroberyx gerrardi*; Berycidae) John Dory (*Zeus faber*; Zeidae) Mirror Dory (*Zenopsis nebulosus*; Zeidae)
3b (length-composition data)	Stocks for which age data are available for a length-based catch curve analysis (Klaer et al., 2012)	
4	Stocks for which catch-per-unit-effort is assumed to be a proxy for abundance (Little et al., 2011).	Silver trevally (*Pseudocaranx dentex*; Carangidae) Royal red prawn (*Haliporoides sibogae*; Solenoceridae) Blue warehou (*Seriolella brama*; Centrolophidae) Ribaldo (*Mora moro*; Moridae)

In 2010, RBCs for 12 SESSF stocks were placed in tier 1, 4 in tier 3, and 14 in tier 4. No stocks were placed in tiers 2 and 3b in 2010. A recent development in the fishery is the implementation of multiyear TACs. As a result, in 2014 two stocks were newly assessed using tier 1, 2 in tier 3 and 2 in tier 4. However, to date, there has been no formal testing (e.g. using management strategy evaluation, MSE – Smith 1993) to determine decision rules for multiyear TACs (Haddon et al., 2012).

The expectation of risk equivalency between tiers in the SESSF is generally achieved through the use of buffers in the HCR for each tier. Tier 1 is based on a robust assessment and therefore no buffer is applied. No buffer has been set for tier 2, due to it presently being unused. In tier 3, the RBC from the basic HCR is multiplied by 0.95 to account for the additional uncertainty associated with management advice based on catch curves (Smith et al., 2008). The buffer of 0.85 for tier 4 stocks is used to account for the additional uncertainty associated with management advice based on trends in catch-rate.

There are several meta-rules that further modify the outcomes from the HCRs (Stobutzki et al., 2011):

1. The extent of precaution associated with the tier 3 and 4 HCRs (0.95 and 0.85) can be reduced if there is evidence that other management measures such as closed areas provide additional precaution for the stock.
2. The maximum increases or decrease in TACs is limited to 50%.
3. No change is made to the TAC if the recommended change in a TAC is less than 10% or 50 t, whichever is less, unless there is a long-term trend in RBC.
4.

The Commonwealth Harvest Strategy Policy has a rebuilding requirement for stocks depleted below the limit reference point, but unlike the setting of RBCs, the process is not formalised but individual rebuilding strategies are developed for any such depleted species. In addition, there are several stocks for which there are both length/age-composition data and catch-rate data. In this case, the Resource Assessment Groups (RAGs) (Smith et al., 2001) use expert judgement to select whether to place the stock in tier 3 or 4.

Europe and ICES

In order to help policy makers move towards sustainable exploitation of all fish stocks, ICES implemented a framework in 2012 (the Data-Limited Stocks, or DLS, framework) to provide quantitative advice, not only for fish stocks for which there is sufficient data to conduct full analytical assessments, but also for those fish stocks considered data-limited. At the time, this represented a 6-fold increase in the number of data-limited stocks for which quantitative advice was provided (ICES, 2012). The framework relies on the principle that available information should be used, and that advice should follow a precautionary approach and be based on the same principles as applied to data-rich stocks. This framework therefore calls for the determination of exploitation relative to F_{MSY} and the consideration of stock trends, and provides a variety of methods that can be applied to facilitate this; the choice of method depends on the information and data available for a given stock. The ICES DLS framework is divided into six categories, with category 1 applied to the data-rich stocks and categories 2 to 6 to increasingly data-limited stocks (Table A.4). A decision tree (Fig. A.2) determines under which category a stock will fall, with the availability of high-quality data and proxies decreasing, and the level of precaution applied increasing, from categories 1 to 6. Within each category, there are further groupings based on methods that can be applied within that category (ICES, 2012). Of the data-limited stocks for which advice was provided in 2014, the bulk fell within category 3 (which bases advice on fishery-dependent or -independent indices) and categories 5 and 6 (which bases advice on recent catch or landings data only) (Table A.4; ICES, 2014b). In 2014, the most widely used DLS method essentially adjusts the previous year's catches (or average of previous years' catches, depending on whether a trend in catches is evident or not), $C_{y\text{-}1}$, on the basis of a ratio of average stock size indices (e.g. age-aggregated survey biomass indices) as follows:

$$C_{y+1} = C_{y-1} \sum_{i=y-x}^{y-1} I_i / x \Bigg/ \sum_{i=y-z}^{y-x-1} I_i / (z-x) \tag{1}$$

where I is the age-aggregated index, x is typically 2, and z is typically 5.

Table A. 4 Tier (referred to as categories by ICES) assignments for the stocks under ICES DLS framework (ICES 2012, 2014b).

Tier Number	Description	Example stocks
1	Stocks with quantitative assessments that provide present stock status and are also able to provide forecasts of stock status.	North Sea cod (*Gadus morhua*; Gadidae) Bay of Biscay sole (*Solea Solea*; Soleidae) *Nephrops* in North Minch
2	Stocks with analytical assessments and forecasts, but these can only be treated qualitatively.	Golden redfish in I and II (*Sebastes marinus*; Sebastidae) Eastern Channel plaice (*Pleuronectes platessa*; Pleuronectidae)
3	Stocks with reliable fishery-independent and -dependent indices that provide reliable indications of trends in stock metrics such as mortality, recruitment and biomass.	Irish Sea haddock (*Melanogrammus aeglefinus*; Gadidae) Greater forkbeard (*Phycis blennoides*; Phycidae) Southern Baltic flounder (*Platichthys flesus*; Pleuronectidae)
4	Stocks with only reliable catch data are available which allows MSY to be approximated.	Pollock in VI and VII (*Pollachius pollachius* and *P. virens*; Gadidae) *Nephrops* in Noup
5	Stocks for which only landings data are available.	Sea bass in VIIIc and IXa (*Dicentrarchus labrax*; Moronidae) Whiting in IIIa (*Merlangius merlangus*; Gadidae)
6	Stocks for which there are negligible landings data and stocks caught in minor amounts as bycatch.	Rockall cod (*Gadus morhua*; Gadidae) Sandeel in VIa (*Ammodytes* spp.; Ammodytidae)

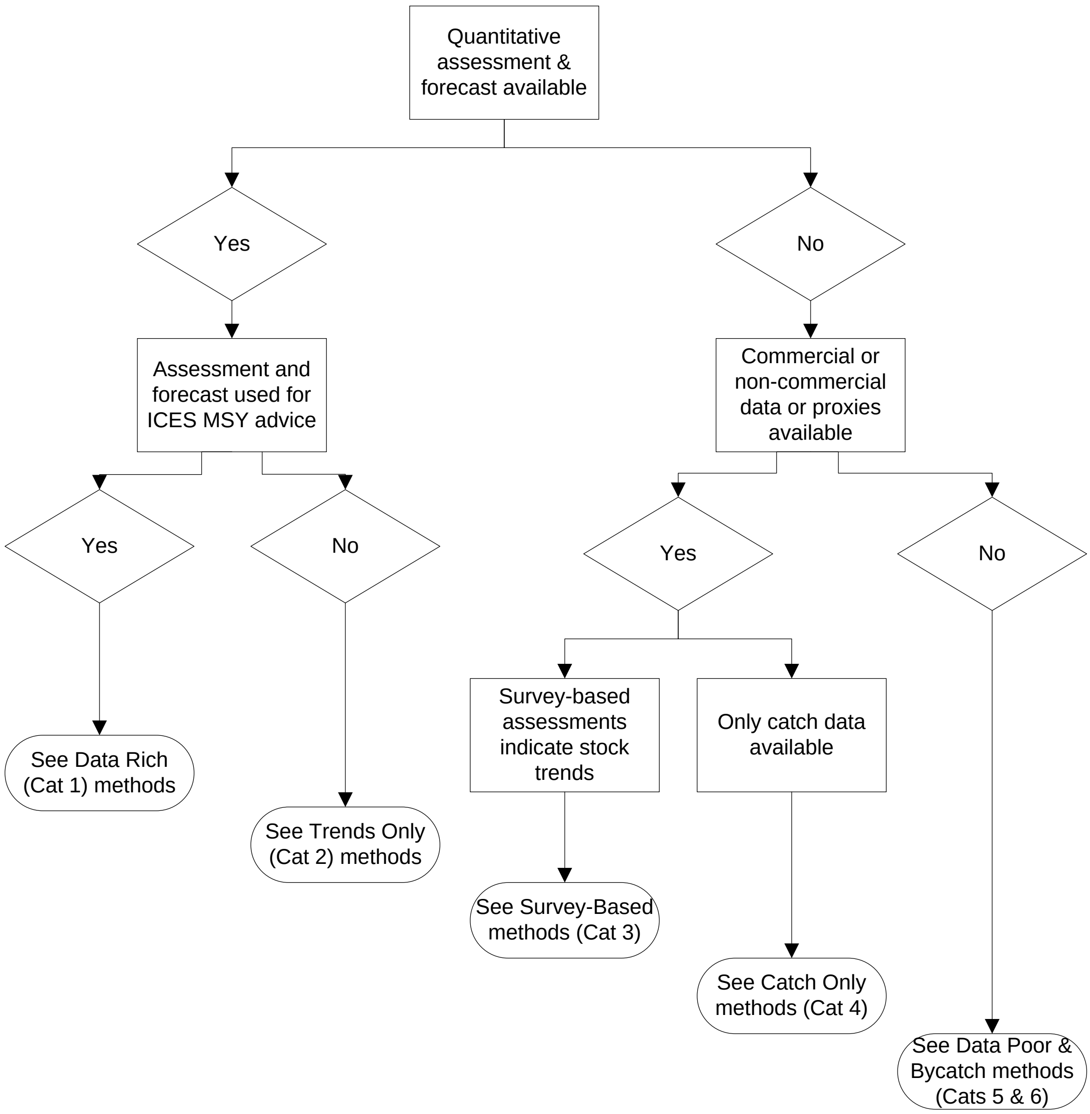


Figure A. 2 Overview of categories of ICES assessment types for data-rich (Category 1) and data-limited stocks (Categories: Cat 2-6). The availability of high quality data and proxies for the assessments decreases and the extent of precaution increases from left to right. Figure taken from ICES (2012) with permission.

The ICES DLS framework employs increasing precaution by first applying a change limit of ±20% relative to a previous catch or average of previous catches (category 2+), and then applying a precautionary margin of -20% (category 3+), although there are exceptions to the latter (ICES, 2012, 2014a). Further precaution is introduced by employing conservative proxies for F_{MSY} (e.g. $F_{0.1}$), where available. The precautionary margin is applied for those cases where it is likely that $F > F_{\mathrm{MSY}}$ or when stock status relative to stock size or exploitation reference points is unknown. Exceptions to the latter are when expert judgement deems the stock not to be reproductively impaired and where there is evidence that stock size is increasing or exploitation has reduced significantly (ICES, 2014a). This approach is intended to move stocks in the direction of sustainable exploitation, given their biological characteristics and the level of uncertainty in the available information. Resultant advice is linked to a time frame that is compatible with a measurable response in the metrics used as a basis for the advice, implying that multi-annual constant catch advice could result (e.g. for three years) where the least information is available, unless important new information emerges justifying a revision of the advice (ICES, 2014a).

Risk equivalency and tier systems

The preceding section highlights that each jurisdiction has a different way to address the expectation of risk equivalency. Only the Australian tier system has an explicit assumption that the risk associated with all species should be equivalent irrespective of the data available (Smith et al., 2014). The Australian tier system also does not expect a fishery to move over time to tiers that are more data rich (especially if the value of the fishery is such that a data-limited tiers are more appropriate), as long as risk equivalency is addressed. In some other tier systems, moving up to more data-rich tiers is an explicit aim. For example, the USA west coast groundfish tier system has a lower $P*$ value for the more data-poor tiers and hence more precaution for more data-poor stocks. In contrast to the USA west coast groundfish fishery, the default $P*$ is the same for stocks in tiers 1 – 4 for the USA Alaskan crab fishery, implying an assumption of risk equivalency among tiers because $P*$ is a measure of the probability of overfishing, a key measure of risk for USA fisheries.

All of the tier systems have several control variables that determine the size of the buffer (e.g. σ and $P*$ for the USA west coast groundfish fishery). Unfortunately, how the values for the control variables were selected is often unclear. For example, the default value for σ in the USA west coast groundfish fishery is 0.36 for category 1 stocks, which is based on the meta-analysis of Ralston et al. (2011). In contrast, the values for σ for category 2 and 3 stocks are simply multiples of 0.36 in the absence of a better basis to define scientific uncertainty.

In the USA system, a buffer is calculated based on the extent of scientific uncertainty, whereas the discount factors in the SESSF are essentially untested (Fay et al., 2012). The ICES system has a mix of approaches depending on the tier. Ideally, the values for the control variables should be selected to achieve desired policy goals. Tier systems can be tested using management strategy evaluation, ideally conducted with stakeholder involvement (Smith et al., 1999). The MSE would focus on the relationship between the values for the control variables and the performance of the management system. For example, the values of the control variables for a given stock could be selected so that risk is constant among tiers (Fig. A.3). The resulting trade-off would be between catch and monitoring cost, giving the decision makers the ability to select a monitoring strategy under the expectation of equal risk.

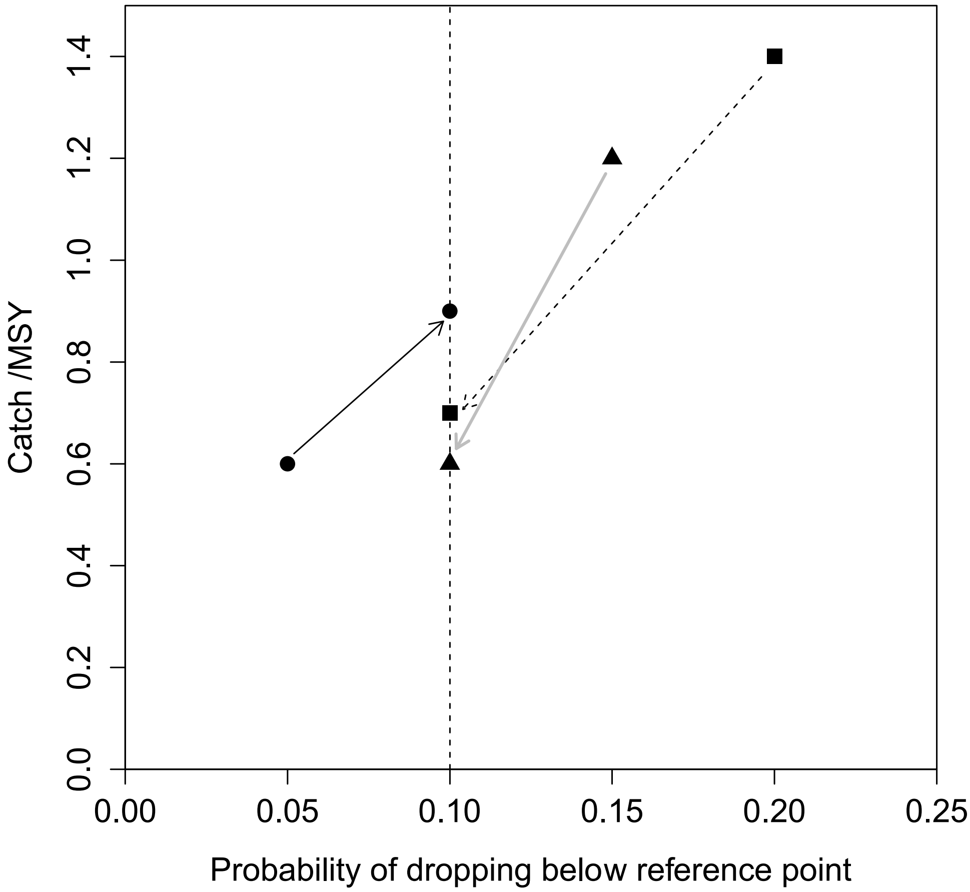


Figure A. 3 Three examples of the potential consequences of adjusting the control variables for each of three tier HCRs so that they achieve risk equivalency. The vertical dotted line is the Australian Commonwealth overfished risk level.

Punt et al. (2012) evaluated alternative choices for the buffer between the OFL and the ABC for a range of Alaskan crab stocks using an MSE-like process. They provided trade-offs between the probability of overfishing and catch given uncertainty in stock size. The results were guided by members of the NPFMC Crab Plan Team and informed the choice by the NPFMC for its default values for $P*$ for crab stocks in tiers 1 – 4. An MSE is being planned to examine the consequences of the choices for σ of 0.72 and 1.44 for data-moderate and data-poor USA west coast groundfish stocks (C. Wetzel, NOAA, pers. commn). That MSE will consider the relationship between a key control variable (σ) and risk, given there is uncertainty regarding the true extent of uncertainty for most stocks of west coast groundfish.

MSE has been used extensively to evaluate the HCRs within each tier for the SESSF. For example, the performances of the tier 3a and 3b HCRs (Table A.3) have been evaluated for both data-rich and data-poor fisheries based on the operating model developed by Fay et al. (2009) (Wayte and Klaer, 2010; Fay et al., 2011; Klaer et al., 2012). The performance of the Tier 4 HCR has also been evaluated using the same operating model (Little et al., 2011). These MSE analyses led to changes to the way the HCRs are applied for stocks in tiers 3 and 4. However, in relation to tier systems, this work has not attempted to evaluate the process of selecting tiers for individual stocks (which can be somewhat subjective), the values for the control variables which determine the buffers for each tier, nor has work been conducted to evaluate the process of deciding whether to set multi-year TACs and some of the “modifiers” such as whether a buffer should be applied, even though the decision about which tier a stock is placed in could have a marked impact on management outcomes. In general, the MSE analyses showed that the risk between tiers is very case-specific.

Also, the use of the tier 3 HCR seems more likely to be a higher risk than use of tier 4, which runs counter to the current default buffers. This has been seen also in single a species treatment (Fay et al. 2012; Little et al. 2014).

ICES has conducted MSE simulation testing of methods proposed for the various DLS categories through a number of its groups; ICES (2014b) provides an overview of this simulation work. However, to date the focus has been on HCR performance, rather than to assess whether there is risk equivalency among tiers, although an MSE did consider whether the principle of increasing precaution down the tiers was achieved for a stock that was forced into more data-limited tiers by making less of its data available to the DLS framework. In this study, risk equivalency was found wanting (ICES, 2014b), indicating that further work was needed on the appropriate size of the precautionary margin for the different tiers. To date, the magnitude and duration of the precautionary margin has not been evaluated using MSE.

Application to the SESSF

To practically demonstrate, using a more quantitative approach, the differences between the various systems, the SESSF species were placed into USA and ICES frameworks (Table A.5) through comparison of data types, assessment methods and harvest control rules. The default buffers that would apply under these jurisdictions are also added. There is relatively strong concordance between the various systems, with most SESSF tier 1 stocks being assigned to high tiers within the USA and ICES systems. No stocks were assigned to the USA Alaskan crab tier 1 as none of the SESSF assessments fully quantify uncertainty using a probability density function. Two of the SESSF assessments (those for school whiting and orange roughy) would likely be placed in the USA west coast groundfish fishery category 2 owing to sensitivity to assumptions (school whiting) or limited data to estimate year-class strength (orange roughy). This is not unexpected given that the SESSF system is based on the assessment method applied and implemented, with less emphasis on its reliability. The SESSF tier 3 and 4 stocks would be assigned to tiers 3c/d under the USA west coast groundfish fishery system, tier 5 under the USA Alaskan crab system, and categories 3 and 4 under the ICES system.

Table A. 5 Comparison of tier systems through placing the SESSF species into the different systems. CV refers to the CV of estimated biomass for the most recent year (used to calculate the buffer in the USA west coast groundfish fishery system). The values in parentheses are the default buffers that would apply to each stock.

SESSF species	CV	SESSF	USA west coast Groundfish	USA Alaska crab	ICES
Flathead	0.111	1 (1)	1a (0.956)	2 (0.997)	1 (1)
Blue grenadier	0.137	1 (1)	1b (0.956)	3 (0.997)	1 (1)
Ling[1]	0.225 (E)	1 (1)	1a (0.956)	3 (0.994)	1 (1)
	0.202 (W)	1 (1)	1a (0.956)	3 (0.995)	
School whiting	0.191	1 (1)	2d (0.833)	3 (0.995)	1 (1)
Orange roughy (east stock)	0.093	1 (1)	2d (0.833)	3 (0.998)	1 (1)
Eastern gemfish	0.306	1 (1)	1a (0.956)	3 (0.992)	1 (1)
Spotted warehou	0.096	1 (1)	1a (0.956)	3 (0.998)	1 (1)
Redfish	-	3 (0.95)	3c/d (0.694)	5 (0.9)	4 (0.8)
John dory	-	3 (0.95)	3c/d (0.694)	5 (0.9)	4(0.8)
Mirror dory	-	3 (0.95)	3c/d (0.694)	5 (0.9)	4(0.8)
Blue warehou	-	4 (0.85)	3/d (0.694)	5 (0.9)	3(0.8)
Royal red prawn	-	4 (0.85)	3a (0.694)	5 (0.9)	3 (0.8)
Silver trevally	-	4 (0.85)	3d (0.694)	5 (0.9)	3 (0.8)
Blue-eye trevalla	-	4 (0.85)	3c/d (0.833)	5 (0.9)	3 (0.8)
Ribaldo	-	4 (0.85)	3c/d (0.694)	5 (0.9)	3(0.8)
Eastern deepwater shark	-	4 (0.85)	3c/d (0.694)	5 (0.9)	3(0.8)

1 – there are two stocks of ling off southeast Australian (Whitten and Punt, 2014), east (E) and west (W

There is some subjectivity associated with this assignment of stocks to tiers. For example, the USA west coast groundfish fishery system involves expert judgement regarding whether a full assessment should be assigned to tier 1 or tier 2 (Table A.1). The groups responsible for assigning stocks to tiers in the USA and within ICES are scientists, whereas in the Australia system they include scientists, industry and managers (Smith et al. 2001)

Table A.5 shows that the buffers that would be applied if the SESSF stocks were managed under the USA and ICES frameworks were largest for USA west coast groundfish fishery system; for some species the buffer would be more than 25% larger (stocks in tier 4). The buffers under the USA Alaska crab system are close to 1 for the assessments based on models (but given the way that system operates in reality, the buffers are usually 0.9 or smaller; see Section 3.2).

Discussion and synthesis

The SESSF tier system evolved to become a means of providing RBCs for all quota stocks irrespective of the data available for assessment purposes, with buffers introduced with the intent of achieving risk equivalency. The SESSF system has since been expanded to include a broader range of species and fisheries beyond the SESSF (Dowling et al., 2013; Table A.6). Although the original and expanded tier systems have been subjected to MSE testing, there is no explicit link between the buffers for individual species/stocks and the risk associated with their assessments. In particular, there is no concept in the Australian system that two assessments which use the same data types (that is, within the same tier) may have different associated risk, although that was part of the rationale for the (currently unused) tier 2. In contrast, the USA west coast groundfish fishery includes category 2d that allows stocks, for which the full age-structured assessment results are less than ideal, to be placed in a more data-poor tier, and hence assigned a larger buffer. Moreover, unlike the two USA systems, the Australian and ICES systems have no link between the estimated extent of uncertainty and the buffer size. In principle, the currently unused tier 2 in the SESSF could provide a way to distinguish between reliable and less reliable model-based assessments as was the original intent.

Table A. 6 Harvest strategy Tier levels from Dowling et al. (2013).

Tier number	Description	Example
0	Robust assessment of F and B based on fishery dependent AND independent data	Northern Prawn Fishery tiger and endeavour prawns
1	Robust assessment of F and B based on fishery dependent data ONLY	SESSF Flathead
2	Assessment of F and B based on fishery dependent and/or fishery independent data	Not used
3	Empirical estimates of F based on size and/or age data	SESSF Redfish
4	Empirical estimates of (a) relative biomass based on fishery dependent and/or independent data, or (b) within season changes to relative biomass based on fishery dependent or independent data.	SESSF Blue-eye trevalla ETBF (size-based CPUE slope rules) Arrow Squid Fishery (within-season depletion analyses) Bass Strait Scallop Fishery (spatial management based on pre-season biomass surveys)
5	Empirical estimates of F based on spatial distribution of effort relative to species distribution	None
6	No estimate of biomass and F; use of fishery-dependent species-specific triggers	Western Deep Water Trawl Fishery, North-West Slope Trawl Fishery, Skipjack Tuna
7	No estimate of biomass and F; use of fishery-dependent triggers for groups of species	Coral Sea Fishery Hand Collection: Aquarium sub-sector Coral Sea Fishery: Line, Trawl and Trap sub-sectors

The tier systems include several common elements (Table A.7). However, they also differ in some important ways. In particular, only the USA systems explicitly recognize implementation error which can be substantial in many fisheries, especially those with a large recreational component where enforcement of limits on catch and effort is difficult. In addition, the USA and ICES frameworks focus on the ability to estimate quantities needed to apply HCRs whereas the SESSF system is focused primarily on having the data needed to apply a HCR. Unlike the SESSF system, the other systems can assign species with survey data to more data-rich (lower buffer tiers). Finally, no jurisdiction has evaluated the entire tier system, including how species are assigned to tier, although the individual HCRs within the SESSF system have been subject to extensive simulation evaluation.

Table A. 7 Overview and comparison of different aspects of the four case studies

	USA west coast groundfish fishery	Alaskan crab	Australian SESSF	ICES
Clear definition of risk?	Yes, scientific and implementation risk.	Yes, scientific and implementation risk.	Yes, for Limit Reference Point risk but less clearly for the target reference point	Yes, for Limit Reference Point risk of data rich methods. Not explicitly for others
Basis of tier system	Ability of the assessment to estimate management quantities and reliability of the data types	Ability of the assessment to estimate management quantities and associated uncertainties	Ability to produce a reliable stock assessment	Ability of the assessment to estimate management quantities including forecast and reliability of the data types
Highest data type	Independent survey data	Independent survey data	Dependent data – no extra value for independent survey data	Independent survey data
Number of tiers	3	5	4	6
Presence of sub-tiers	Yes, at all tier levels	No	Yes, but only for tier 3	Yes for most tier levels
MSE tested?	Underway	Yes and does conform to the risk by tier assumption	Yes and does not always conform to the risk by tier assumptions	Yes and does not always conform to the risk by tier assumptions

All of the tier systems involve an element of expert judgement, the consequences of which for risk equivalency are unclear. For example, the choice of whether to assign a stock to tier 1a and 2d is a decision made by the PFMC SSC, while the NPFMC SSC can change the buffer for USA Alaskan crab from the default values. The Crab Plan Team provides the SSC with reasons (if there are any) for setting the ABC less than the OFL. However, at present there is no formal structure or tested basis for this rationale. In the SESSF system, the choice of whether to place a stock in tier 3 or 4 is made by the relevant assessment group. The elements of expert judgement within tier systems may help to achieve an expectation of risk equivalency, but potentially make the process somewhat subjective and hence impossible to simulation test using MSE.

The tier system for the USA west coast groundfish fishery has been in place for eight years. A major advantage of that system is that there is a clear separation between the roles of the SSC (reviewing and approving assessments, assigning assessments to categories, and specifying σ) and that of the Council (selecting the value of P^*, and choosing an ACL below the ABC). As such, management and policy is explicitly separated from the scientific advisory processes (Field et al., 2006). In contrast, it is impossible to separate the quantification of uncertainty from risk tolerance in the way the buffers for the SESSF and the ICES systems, as well as the buffer for tier 5 USA Alaskan crab stocks, are constructed.

The tier systems implement the expectation of risk equivalency through addressing uncertainty, but few discuss bias (i.e. that the expected value for biomass and hence the catch limit differs from the true values), although this certainly occurs in practice. For example, in the USA west coast system, account is taken of between-species variation in the CVs of biomass estimates, but bias is only addressed partially by assigning stocks for which the full assessment is 'less reliable' to category 2 (Table A.2). The bias in the estimates from assessments may differ among species, and its direction and size may be impossible to pre-determine without case-specific simulation testing.

It is generally recognized that MSE is the best-practice approach for comparing management systems (Punt et al., in press). Of the various case studies, the SESSF is the furthest towards conducting MSE studies for all its tiers and for several species. Ideally, MSE could be conducted for all species and when each species is placed in each tier. This would permit decision makers the opportunity to select the level of monitoring to maximize return given a risk criterion. However, this will be computationally very intensive for systems (such as the USA west coast groundfish fishery and ICES systems; Tables A.2 and A.4) that have many species. A generic MSE system may be very useful here. In addition, having a different choice of control variables for each species would lead to a very unwieldy system. There would therefore be value in selecting values for control variables for groups of species, or applying simple rules such as that for the USA west coast groundfish fishery, which sets σ to 0.36 for tier 1 stocks unless the estimate from the assessment is higher. The values for the control parameters would be set so that performance for some measure of risk, such as that the probability of stock staying above the limit biomass level at least 90% of the time, is the same among tiers. Fig. A.3 shows a generic example of the qualitative implications of this for the risk-cost-catch trade-off. In Fig. A.3, there are measures of risk for three possible tiers. The default HCRs for each tier lead to quite different measures of risk. Adjusting their control parameters so that risk is the same among tiers (symbols at the end of the arrows) allows a comparison of performance where the probability of dropping below the reference point (here set to 0.1) is equivalent among tiers. The choice of data to collect is then determined by the trade-off between cost and the amount of increased catch. In the context of Fig. A.3, for example, whether the increased cost associated with the tier corresponding to the diamond tier, compared to that associated with the square tier, is more than offset by the higher catch achieved by the diamond tier.

The results of this synthesis lead to several recommendations for how tier systems should be developed in the context of achieving risk equivalency:

- Tiers should not be defined simply on data availability, but also on the reliability of assessments used to estimate management quantities based on those data.
- The process for selecting control variables should clearly differentiate quantification of uncertainty, from how decision makers wish to address that uncertainty (risk quantification versus risk tolerance and imposing a distinction between scientific uncertainty and additional uncertainty added by decisions makers).
- Risk equivalency is best tested by using MSE to select the values for control variables that determine the buffer given the uncertainty associated with the assessment.
- Basing values for control variables on a MSE analysis for each stock is ideal and recommended given potential biases and unknown consequences of additional rules within a harvest strategy (but may be computationally infeasible for regions with large numbers of stocks, or for stocks that are so data poor that a MSE is largely impossible to conduct).
-

Acknowledgements

We would like to thank Drs Ilona Stobutzki (ABARES) and Simon Jennings (CEFAS), and two anonymous reviewers for their in depth comments on draft versions of the paper. Also input from the various SESSF RAGs and AFMA members that contributed to several iterations of the work. This work was funded by the Fisheries Research and Development Corporation Project 2012/202 and CSIRO Oceans and Atmosphere Flagship.

References

AFMA. (2010) *AFMA cost recovery impact statement.* Australian Fisheries Management Authority, Canberra. http://www.afma.gov.au/wp-content/uploads/2014/04/AFMA-Cost-Recovery-Impact-Statement-20101.pdf

Arnason, R., Hannesson, R. and W.E. Schrank, W.E. (2000) Cost of fisheries management: the cases of Iceland, Norway and Newfoundland. *Marine Policy* **24**, 233-243.

Cox, A. (2000) Cost recovery in fisheries management: the Australian experience. *IIFET 2000 Proceedings.*

DAFF. (2007) *Commonwealth Fisheries Harvest Strategy Policy Guidelines*. Australian Government Department of Agriculture, Fisheries and Forestry, Canberra, Australia, pp. 55. http://www.agriculture.gov.au/fisheries/domestic/harvest_strategy_policy

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

European Commission. (2013). Regulation (EU) No 1380/2013 of the European Parliament and of the Council of 11 December 2013 on the Common Fisheries Policy, amending Council Regulations (EC) No 1954/2003 and (EC) No 1224/2009 and repealing Council Regulations (EC) No 2371/2002 and (EC) No 639/2004 and Council Decision 2004/585/EC. (OJ L354/22 28.12.2013, pp. 40. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:354:0022:0061:EN:PDF

Fay, G., Punt, A.E. and Smith, A.D.M. (2009) Operating model specifications. In: *Evaluation of New Harvest Strategies for SESSF Species.* (ed. Wayte, S.E.) CSIRO Marine and Atmospheric Research, Hobart and Australian Fisheries Management Authority, Canberra pp. 125-133.

Fay, G., Punt, A.E. and Smith, A.D.M. (2011) Impacts of spatial uncertainty on the performance of age structure-based harvest strategies for blue-eye trevalla (*Hyperoglyphe antarctica*). *Fisheries Research* **110**, 391–407.

Fay, G., Little, L.R. and Tuck, G. (2012) Maintaining risk equivalency among fishery harvest control rules: testing methods for precautionary fisheries management in Southeast Australia. CSIRO Marine and Atmospheric Research Report.

Field, J.C., Punt, A.E., Methot, R.D. and Thomson, C.J. (2006) Does MPA mean “Major Problem for Assessments?” Considering the consequences of place-based management systems. *Fish and Fisheries* **7**, 284–302.

Haddon, M., Klaer, N., Smith, D.C., Dichmont, C.D. and Smith, A.D.M. (2012) Technical reviews for the Commonwealth harvest Strategy Policy. *FRDC Report 2012/225*. CSIRO, Hobart.

ICES. (2012) ICES Implementation of Advice for Data-limited Stocks in 2012 in its 2012 Advice. ICES DLS Guidance Report 2012. *ICES CM 2012/ACOM 68*, 42 pp.

ICES. (2014a.) Report of the ICES Advisory Committee 2013. *ICES Advice, 2014*. Book 1, Section 1.2. [Accessed online, 02/03/2015: http://www.ices.dk/sites/pub/Publication%20Reports/Advice/2014/2014/1.2_Advice_basis_2014.pdf]

ICES. (2014b.) Report of the Workshop on the Development of Quantitative Assessment Methodologies based on LIFE-history traits, exploitation characteristics, and other relevant parameters for data-limited stocks (WKLIFE IV), 27–31 October 2014, Lisbon, Portugal. *ICES CM 2014/ACOM:54*. 223 pp.

Klaer, N.L., Wayte, S.E. and Fay, G. (2012) An evaluation of the performance of a harvest strategy that uses an average-length-based assessment method. *Fisheries Research* **134-136**, 42-51.

Little, L.R., Parslow, J., Fay, G., Grafton, R.Q., Smith, A.D.M., Punt, A.E., and Tuck, G.N. (2014) Environmental Derivatives, Risk Analysis, and Conservation Management. *Conservation Letters* **7**, 196-207.

Little, L.R., Wayte, S.E., Tuck, G.N. et al. (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.

Link, J., Burnett, J., Kostovick, P., and J. Galbraith. (2008) Value-added sampling for fishery independent surveys: Don't stop after you're done counting and measuring. *Fisheries Research* **93**, 229-233.

North Pacific Fishery Management Council (NPFMC). (2014) Stock Assessment and Fishery Evaluation Report for the king and Tanner crab fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, 605 W. 4th Avenue, #306, Anchorage, AK 99501.

Pacific Fishery Management Council (PFMC). (2014a) *Status of the Pacific Coast Groundfish Fishery: Stock Assessment and Fishery Evaluation.* Pacific Fishery Management Council, 7700 Ambassador Place, Suite 101, Portland, OR 97220.

Pacific Fishery Management Council (PFMC). (2014b) *Terms of Reference for the groundfish and coastal pelagic species stock assessment review process for 2015-2016.* Pacific Fishery Management Council, 7700 Ambassador Place, Suite 101, Portland, OR 97220.

Prager, M.H. and Shertzer, K.W. (2010) Deriving Acceptable Biological Catch from the Overfishing Limit: Implications for Assessment Models. *North American Journal of Fisheries Management* **30**, 289–294.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A. and Haddon, M. (in press) Management Strategy Evaluation: Best Practices. *Fish and Fisheries* **00**, 00-00.

Punt, A.E., Siddeek, M.S.M., Garber-Yonts, B., Dalton, M., Rugolo, L., Stram, D., Turnock, B.J. and Zheng, J. (2012) Evaluating the impact of buffers to account for scientific uncertainty when setting TACs: Application to red king crab in Bristol Bay, Alaska. *ICES Journal of Marine Science* **69**, 624–634.

Punt, A.E., Smith, A.D.M., Smith, D.C., Tuck, G.N. and Klaer, N.L. (2014a.) Selecting relative abundance proxies for BMSY and BMEY. *ICES Journal of Marine Science* **71**, 469–483.

Punt, A.E., Szuwalski, C. and Stockhausen, W (2014b) An evaluation of stock–recruitment proxies and environmental change points for implementing the USA Sustainable Fisheries Act. *Fisheries Research* **157**, 28–40.

Ralston, S., Punt, A.E., Hamel, O.S., DeVore, J. and Conser, R.J. (2011) An approach to quantifying scientific uncertainty in stock assessment. *Fishery Bulletin* **109**, 217–231.

Sainsbury, K. (2005) Cost-effective management of uncertainty in fisheries. *National Outlook Conference 2005.* Australian Bureau of Agricultural and Resource Economics, Canberra. [Accessed online, 20/05/2015: http://data.daff.gov.au/data/warehouse/pe_abarebrs99001173/PC13024.pdf]

Shertzer, C.E., Prager, M.H. and Williams, E.H. (2008) A probability-based approach to setting annual catch levels. *Fishery Bulletin* **106**, 225–232.

Smith, A.D.M., (1993) Risks of over- and under-fishing new resources. In: *Risk Evaluation and Biological Reference Points for Fisheries Management, 120.* (eds. Smith, S.J., Hunt, J.J., Rivard, D.), Canadian Special Publication of Fisheries and Aquatic Sciences, pp. 261–267.

Smith, A.D.M., Sainsbury, K.J., and Stevens, R.A. (1999) Implementing effective fisheries management systems—management strategy evaluation and the Australian partnership approach. *ICES Journal of Marine Science* **56**, 967–979.

Smith, D.C., Smith, A.D.M. and Punt, A.E. (2001). Approach and process for stock assessment in the south east fishery: A perspective. *Marine Freshwater Research* **52**, 671–681.

Smith, A.D.M., Smith, D.C., Tuck, G.N., et al. (2008) Experience in implementing harvest strategies in Australia's south-eastern fisheries. *Fisheries Research* **94**, 373–379.

Smith, A.D., Smith, D.C., Haddon, M., Knuckey, I.A., Sainsbury, K.J. and Sloan, S.R. (2014) Implementing harvest strategies in Australia: 5 years on. *ICES Journal of Marine Science* 71, 195–205.

Stobutzki, I., Vieira, S., Ward, P. and Noriega, R. (2011) Southern and Eastern Scalefish and Shark Fishery overview. In: *Fishery Status Reports 2010: Stautus of fish stocks and fisheries managed by the Australian Government* (eds Woodhams, J., Stobutzzki, I., Vieier, S., Curtotti, R., and Begg, G.A.) Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra.

USA Department of Commerce (USA Doc). (2007) *Magnuson-Stevens Fishery Conservation and Management Act as Amended Through January 12, 2007.* http://www.nmfs.noaa.gov/sfa/magact/MSA_Amended_2007%20.pdf (last accessed 18 July 2014).

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

Whitten, A.E. and Punt, A.E. (2014) Stock Assessment of pink ling (*Genypterus blacodes*) using data up to 2012. p. 116-142. In: *Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2013, Part 1.* (ed. G.N. Tuck). Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart.

Appendix B. Assessing a multilevel tier system: the role and implications of data quality and availability

Natalie A. Dowling[1], André E. Punt[1,2], L. Richard Little[1], Catherine M. Dichmont[3], David C. Smith[1], Malcolm Haddon[1], Miriana Sporcic[1], Elizabeth A. Fulton[1], Rebecca J. Gorton[1]

[1]*CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart, Tasmania, Australia*

[2]*School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195-5020, USA*

[3]*CSIRO Oceans and Atmosphere, Ecosciences Precinct, Dutton Park, Queensland, 4750, Australia*

Published in Dowling, N.A., Punt, A.E., Little, L.R., Dichmont, C.M., Smith, D.C., Haddon, M., Sporcic, M., Fulton, E., Gorton, R.J. 2016. Assessing a multilevel tier system: the role and implications of data quality and availability. Fisheries Research. 183: 588–593. doi: 10.1016/j.fishres.2016.05.001

Abstract

Tier systems for fisheries assessment and management are widely used, but defined differently by jurisdiction. A principal component analysis was applied to the expanded Australian Commonwealth 8-tier system for fishery assessment and management to determine whether it adequately delineates across stocks according to data availability and quality. The original Australian tiers comprised four levels that were defined primarily according to the available stock assessment options, given the data availability and quality. We asked fishery experts to score information quality for each of the main Australian Commonwealth species and/or fisheries. Multivariate analysis indicated that the eight tiers delineated between the extreme tier levels on the first principal component, although there was overlap for intermediate tiers. More generally, it is important that the aim of tier systems and the basis for tier delineations are explicitly defined given the increasing association of tiers with trade-offs between overfishing risk, management cost and catch.

Keywords: fishery assessment; risk-cost-catch; risk equivalency; tier systems

Corresponding author: N.A. Dowling

a. *Email*: Natalie.Dowling@csiro.au
b. *Phone*: 61 36232 5148
c. *Fax* : 61 36232 5012

Highlights

- Tier systems can delineate stocks in terms of data availability and quality.
- Classification of species within Australia's tier system is only broadly consistent with data availability and quality.
- There is overlap in data quality among the intermediate tiers.

Introduction

Successful fisheries management requires a trade-off between maintaining fisheries at biologically sustainable levels, while ensuring economic sustainability, and effective but minimal management costs. Described as the “risk-cost-catch frontier” (Sainsbury, 2005; Dowling et al., 2013), there is an inherent assumption that the greater the risk of overfishing, the more precautionary management should be, and a trade-off follows that attempts to balance risk against the cost of management and catch levels (Little et al., 2015).

Across the United States, Europe and Australia, the risk-cost-catch trade-off is intended to be encapsulated, either implicitly (USA, Europe – International Council for the Exploration of the Seas (ICES)), or explicitly (Australia) using “tier” systems of assessment and management (e.g., PFMC, 2014; ICES, 2012; Dowling et al., 2013, Smith et al., 2008, 2014; Dichmont et al., 2016). The higher the tier level (number), i.e. the more data limited, the greater the uncertainty and risk of overfishing, and so, presumably, the more conservative the recommended biological catch. The assumption is that there is greater risk of over-exploitation when data are poorer or fewer data are available, or if a formal stock assessment has not been undertaken.

The basis of current tier delineations vary by jurisdiction (Dichmont et al., 2016). For the USA west coast groundfish fishery and the Alaskan crab fishery, 3 or 5-level tier systems (NPFMC, 2014; PFMC, 2014) are delineated according to data availability, the ability to estimate quantities used in decision rules, and the perceived reliability of the resulting estimates of management-related quantities. The ICES system (ICES, 2012) is based on the ability of an assessment to estimate management quantities and data reliability.

In Australia, a four-level tier system was originally developed for the Southern and Eastern Scalefish and Shark Fishery (SESSF), primarily according to data quality and availability, given the available assessment options concordant with these (Smith and Smith, 2005). Tiers were based on the ability to produce a reliable assessment from the available data, which in turn defined the assessment method and the associated harvest control rule (HCR). The SESSF tier system attempted to apply a “discount factor” on TAC of each tier to equalise risk (Smith et al., 2014), although this requirement for risk equivalency evolved over time and during its use. The aim was to acknowledge the amount and quality of data used for assessment, and the supposedly lower certainty associated with data-limited assessments; this was consistent with the requirements of the Australian Commonwealth Harvest Strategy Policy (DAFF, 2007). Implicit in this was the assumption that the cost of collecting more data at a higher-level tier, would be offset by increased assessment certainty and thus less precaution in the recommended biological catches.

The question remains as to the most appropriate criteria by which assessment and management tiers should be delineated. Dichmont et al. (2016) recommended that tiers should not be defined simply on data and associated assessment availability, but also on the reliability of the stock assessments. A stock assessment could be considered to be “unreliable” when there is considerable sensitivity to changing some of its assumptions, when some key parameters (such as deviations about the stock-recruitment relationship) are not estimated, or when there are obvious retrospective patterns.

Dowling et al. (2013) describe an (unofficial) 8-level tier system based on the four tiers originally defined for the SESSF (Smith et al., 2008), while adding the remaining Australian Commonwealth harvest strategies into these and the four additional tiers. This new ordered tier system (numbered from 0 to 7, with the original SESSF tiers corresponding to tiers 1-4) attempts to accommodate more data-limited species and fisheries, with the aim of embracing a broader range of species and situations using existing Australian Commonwealth fishery harvest strategies. Here, we examine this expanded (8-level) tier system to determine the factors responsible for allocating Australian stocks into their respective tier level. This is valuable, as the assessment and decision rules beyond those developed for the SESSF have never before been evaluated as a unified system. This analysis forms an adjunct to Dichmont et al (2016), and to a Management Strategy Evaluation (MSE) currently being undertaken to evaluate the risk-cost-catch frontier across tiers for Australian Commonwealth fisheries.

Methods

The availability and quality of data for the main species, or species groups, under each of Australia's Commonwealth (Federal) fisheries, was scored according to the guidelines in Table B.1. The scoring system of 0 to 3 (highest to lowest), was used to score the following seven types of data: (1) Fishery-independent survey(s) (FIS) index, (2) catch-per-unit-effort (CPUE), (3) catch length-composition (CL), (4) catch age-composition (CA), (5) total landed catch including estimates of discards, illegal catch, etc. (totC), (6) landed catch (C), and (7) effort (E).

Table B 1 Criteria used to score quality and availability

Score	0	1	2	3
Fishery-independent survey	Unbiased, low CVs	Unbiased, high CVS	Likely biased as indicators of trend, or poor spatial/temporal coverage	None
CPUE	Targeted Fishery, and standardized	Bycatch/non-targeted fishery and standardized, and/or issues with spatial structure	Available but perhaps not standardized, or poor spatial/temporal/fleet coverage	None
Length-frequency	Representative of the whole fishery	Representative of at least one fleet/part of the fishery	Some data available	None
Catch-at-age	Representative of the whole fishery, ageing error known	Representative of at least one fleet/part of the fishery	Some data available	None
Total catch (including discards)	Whole fishery covered, data reliable and/or observer effort covered > 50% of catch	Whole fishery covered; discard high and variable, and/or some uncertainty in reporting	Only landed catch; qualitative knowledge of bycatch, or high uncertainty in reporting, or poor spatial/temporal/fleet coverage	None
Landed catch	Well covered	Issues with stock identification, or catch uncertainty; discard high and variable	Issues with species identification, poor spatial/temporal/fleet coverage, and/or unreliable reporting	None
Effort	All sectors/fleets/participants covered	Multiple sectors with some not included, or not full coverage across fleets/participants	Poor spatial/temporal/fleet coverage, and/or unreliable reporting	None

If CPUE data are available, it follows that so are some effort data. Data categories (6) and (7) of Table B.1 ('landed catch' and 'effort') were intended to apply more to fisheries for which only either catch or effort data are available, and where these form the basis for "assessments". Such fisheries include multispecies fisheries where catch is not reported by species, or fisheries for which catch data are considered highly unreliable (such as the Australian Coral Sea Fishery Line, Trawl and Trap sub-fishery) (Dowling et al., 2008).

Assignment of scores was undertaken by scientific experts for each fishery. All experts received the same explanatory brief, and were aware of the context for the analysis. Thus, scorings were standardised among fisheries to the extent possible. For cross-validation purposes (and to test consistency of the scorers), the same experts were approached some months later to undertake an identical, repeated round of scorings.

Tropical tuna species (Yellowfin Tuna, Bigeye Tuna, and Albacore) were excluded from the Australian Eastern Tuna and Billfish Fishery (ETBF), because assessment and harvest strategies for these species were still being considered at the time of analysis.

The expanded tier definitions (Dowling et al., 2013) provided in Table B.2 were used to explore the extent to which quality and availability of the data are consistent with the tier level expectation. Using a principal components analysis (PCA), we examined whether the tier levels of a range of fisheries clustered according to the scoring for the seven data types describing data quality and availability.

Table B 2 Harvest strategy tier levels (corresponding to an assessment and/or management framework), based on Dowling et al. (2013), and expanded from the 4-tier level system defined in Smith and Smith (2005) for the SESSF. Increasing tier numbers reflect an assumed increased risk of over-fishing. Note that, currently, no Australian stocks or species are assigned to tier 5, but this tier is included because it represents a level of data availability and an assessment intermediate in quality compared to tiers 4 and 6 (Dowling et al. 2013).

Tier	Tier description
0	Robust (in terms of associated low confidence intervals) assessment of fishing mortality (F) and biomass (B), based on fishery-dependent and -independent data
1	Robust assessment of F and B based on fishery-dependent data only
2	Less robust assessment of F and B, based on fishery-dependent and/or fishery-independent data
3	Empirical estimates of F based on size and/or age data
4	Empirical estimates of (a) trends in relative biomass based on catch-per-unit-effort (CPUE) data (b) within-season changes in relative biomass based on CPUE data (c) availability of relative biomass based on informal fishery-independent surveys
5	Empirical estimates of F based on the spatial distribution of effort relative to the distribution of the species
6	No estimate of biomass or F; management decisions based on fishery-dependent species-specific triggers
7	No estimate of biomass or F; management decisions based on fishery-dependent triggers for groups of species

Results

The data scores for the Australian Commonwealth fisheries, for each main target species/species assemblage are shown in Table B.3, along with the currently assigned tier level for the stock. Tiers were assigned under the expanded 8-tier system. The SESSF species retained their original tier designations, with the exception of blue grenadier (*Macruronus novaezelandiae*) and orange roughy (*Hoplostethus atlanticus*). Both originally classified as SESSF tier 1 species under the Smith and Smith (2005) system, these were reassigned as tier 0 as per the definitions in Table B.2. Currently, no Australian stocks or species are assigned to tier 5, but this tier is included because it represents a level of data availability and an assessment of intermediate quality compared to tiers 4 and 6 (Dowling et al., 2013).

Table B 3 Data scores and tiers (as per the expanded 8-level tier system) for the Australian Commonwealth fisheries, for each main target species/species groups. The data scoring system is listed in Table B.1 and tiers are described in Table B.2. Data categories are: fishery-independent surveys (FIS); catch-per-unit-effort (CPUE); catch length-composition (CL); catch age-composition (CA); total catch including discards and illegal catches (TotC); reported landed catch (C) and effort (E) respectively.

Fishery (bold) with Species/Species Assemblage (common and scientific names)	FIS	CPUE	LF	CA	totC	C	E	Tier
Southern and Eastern Scalefish and Shark Fishery								
Flathead (5 species)	2	0	0	0	0	0	1	1
Blue grenadier (*Macruronus novaezelandiae*)	0	0	0	0	0	0	0	0
Blue-eye trevalla (*Hyperoglyphe antarctica*)	2	1	1	2	0	0	1	4
Redfish (*Centroberyz affinis*)	2	1	0	0	1	1	1	1
Pink ling (*Genypterus blacodes*)	1	0	0	0	0	0	0	1
School whiting (*Sillago flindersi*)	2	1	1	1	1	1	1	1
Orange roughy (*Hoplostethus atlanticus*)	0	2	0	0	0	0	0	0
Eastern gemfish (Rexea solandri)	2	0	0	0	1	0	0	1
Silver trevallly (*Pseudocaranx georgianus*)	2	1	1	2	1	1	1	4
Spotted warehou (*Seriolella punctata*)	2	0	0	0	0	0	1	1
Royal red prawn (*Haliporoides sibogae*)	3	2	3	3	1	2	2	4
John Dory (*Zeus faber*)	2	1	2	3	1	0	1	3
Blue warehou (*Seriolella brama*)	2	1	2	2	1	0	1	4
Mirror dory (*Zenopsis nebulosus*)	2	1	2	2	1	1	1	3
Ribaldo (*Mora moro*)	2	2	1	2	2	2	2	4
Eastern deepwater sharks (18 species)	2	1	3	3	0	2	1	4
Western deepwater sharks (18 species)	2	0	3	3	0	2	0	4
Jackass morwong (*Nemadactylus macropterus*)	3	0	0	0	1	1	1	1
Western gemfish (*Rexea solandri*)	3	1	1	1	1	1	1	4
Offshore ocean perch (*Helicolenus barathri*)	3	2	2	3	2	2	2	4
Inshore ocean perch (*Helicolenus percoides*)	3	2	2	3	2	2	2	4
Ocean jackets (*Nelusetta ayraudi*)	3	2	3	3	1	1	1	4
Deepwater flathead (*Neoplatycephalus conatus*)	1	0	0	0	0	0	0	1
Bight redfish (*Centroberyx gerrardi*)	1	0	0	0	0	0	0	1
Mixed oreo dories (6 species)	3	2	3	3	2	2	2	4
School shark (*Galeorhinus galeus*)	3	1	0	2	1	1	1	1
Gummy shark (*Mustelus antarcticus*)	3	0	0	0	0	0	0	1
Elephant fish (*Callorhinchus milii, Harriotta haeckeli, H. raleighana*)	3	1	0	3	1	1	1	4
Sawshark (*Pristiophorus cirratus, P. nudipinnis, P. peroniensis*)	3	1	0	3	1	1	1	4
Small Pelagic Fishery								
Jack mackerels (*Trachurus declivis, T. murphyi)*	1	3	1	3	1	3	1	2
Blue mackerels (*Scomber australasicus*)	1	3	1	3	1	3	1	2
Australian sardines (*Sardinops sagax*)	1	3	1	3	1	3	1	2
Redbait (*Emmelichthys nitidus*)	1	3	1	3	1	3	1	2
Arrow Squid Fishery (jig only)								
Arrow squid (*Nototodarus gouldi*)	3	1	3	3	1	0	0	4
Northern Prawn Fishery								
Brown tiger prawns (*Penaeus esculentus*)	0	0	2	3	0	0	0	0
Grooved tiger prawns (*Penaeus semisulcatus*)	0	0	2	3	0	0	0	0
Blue endeavour prawns (*Metapenaeus endeavouri*)	0	0	2	3	0	0	0	2
Red-legged banana prawns (*Fenneropenaeus indicus*)	3	0	3	3	0	0	0	2
Common banana prawns (*Fenneropenaeus merguiensis*)	1	0	3	3	0	0	0	4
Eastern Tuna and Billfish Fishery								
Broadbill swordfish (*Xiphias gladius*)	3	0	0	1	1	0	0	4
Striped marlin (*Tetrapturus audax*)	3	0	0	1	1	0	0	4
Western Tuna and Billfish Fishery								
Tuna and billfish (6 species)	3	0	1	1	1	0	1	6
Skipjack Tuna Fishery								
Skipjack tuna (*Katsuwonus pelamis*)	3	2	1	3	1	1	1	6
Southern Bluefin Tuna Fishery								
Southern Bluefin Tuna (*Thunnus maccoyii*)	0	1	1	1	1	1	0	2
Western Deepwater Trawl Fishery								
Mixed finfish	3	2	3	3	2	0	0	6
North-West Slope Trawl Fishery								
Scampi (*Metanephrops australiensis, M. boschmai, M. veluntinus*)	3	0	1	3	0	0	0	6
Deepwater prawns (6 species*)*	3	2	3	3	2	0	0	6
Bass Strait Scallop Fishery								
Scallops (*Pecten fumatusi)*	2	2	0	3	2	1	0	4
Coral Sea Fishery: Hand Collection								
Lobster, trochus, beche de mer, aquarium finfish species	3	2	3	3	1	1	1	7
Coral Sea Fishery: Line, Trawl, Trap								
Mixed finfish	3	2	3	3	2	1	1	7
Heard Island and McDonald Island Fishery								
Mackerel icefish (*Champsocephalus gunnari*), Patagonian toothfish (*Dissostichus eleginoides*)	0	0	0	0	0	0	0	0
Macquarie Island								
Patagonian toothfish (*Dissostichus eleginoides*)	3	3	0	0	0	0	0	1

PCA results (Fig. B.1) showed that the first two principal components explained approximately 65% of the variance. The bi-plot and principal component scores for these two components (Table B.4) indicated that principal component 1 related mainly to effort data (E), landings (CDR), CPUE and total catch (totC), and weakly to a survey (FIS), whereas principal component 2 related mainly to length frequencies (CL) and to a lesser extent to age frequencies (CA).

Table B 4 Principal component scores for each data type (factor), scaled proportional to eigenvalues

	Principal component 1	Principal component 2
FIS	0.7391	-0.0676
CPUE	1.2786	-0.3678
CL	0.8905	1.2661
CA	1.1788	0.9538
totC	1.2656	-0.3069
CDR	1.2451	-0.4764
E	1.2174	-0.616

The data indicated that the more information-poor tiers (7, and, to some extent, 6) are located separately to the more information-rich tiers (0 and 1) in PCA space (Fig. B.1). A broad trend of increasing tier number from left to right along the first principal component axis is evident. The low tier numbers (0-1) were distinguished by low scorings associated with the first principal component that distinguished age (CA) and length frequencies (CL). The intermediate tiers (2-6) along the right showed a high degree of overlap (Fig. B.1).

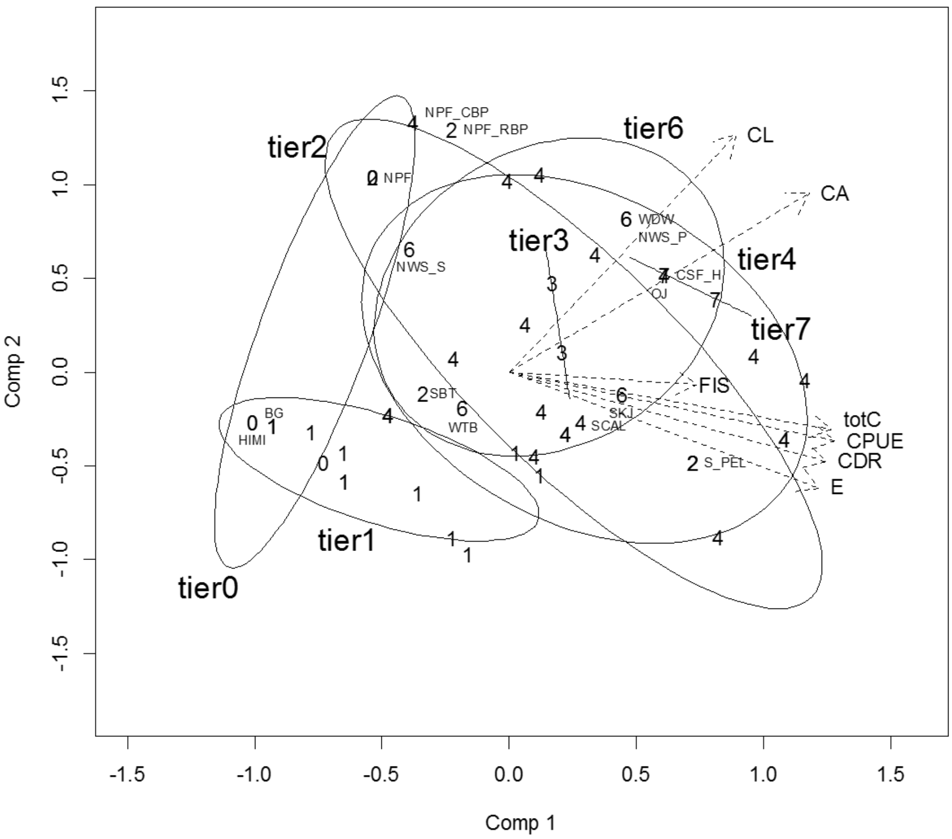


Figure B 1 PCA bi-plot of component 2 versus component 1 for all main Commonwealth species. The vectors represent each data type and the numbers correspond to the assessment tier to which the species/fishery is assigned. Ellipses (representing 75% confidence intervals) (95% confidence intervals showed much overlap and did not allow for clear interpretation) surround the points for each tier. Data type vectors are abbreviated as fishery-independent surveys (FIS); catch-per-unit-effort (CPUE); catch length-composition (CL); catch age-composition (CA); total catch including discards and illegal catches (TotC); reported landed catch (C) and effort (E). The centres of the ellipses along principal component 1 are in the order of tiers 0, 1, 6, 3, 2, 4, 7, with 2, 3, 4, and 6 being nearly identical. Fisheries or species to which specific reference is made in the text have been labelled adjacent to their tier number: "OJ" = Ocean Jackets, "NPF" = Northern Prawn Fishery Brown and Grooved Tiger Prawns and Blue Endeavour Prawns, "WTB" = Western Tuna and Billfish Fishery, "SKJ" = Skipjack Fishery, "SBT" = Southern Bluefin Tuna Fishery, "WDW" = Western Deepwater Trawl Fishery, "NWS_S" = North West Slope Trawl Fishery - scampi, "NWS_P" = North West Slope Trawl Fishery – deepwater prawns, "CSF_H" = Coral Sea Fishery: Hand Collection, "HIMI" = Heard Island and MacDonald Island Fishery, "SCAL" = Bass Strait Scallops, "BG" = Blue grenadier, "S_PEL" = Small Pelagics (all species), "NPF_RBP" = Northern Prawn Fishery Red-legged Banana Prawn, "NPF_CBP" = Northern Prawn Fishery Common Banana Prawn.

There are several reasons for the overlap among tiers. Firstly, data alone may not determine or define the appropriate assessment. For example, although common banana prawns (*Fenneropenaeus merguiensis*, a species in the Northern Prawn Fishery) have fishery independent survey data, and high quality catch, effort and CPUE data, their population dynamics are strongly environmentally driven, and the drivers are poorly understood. Hence they are classified as tier 4, as a more formal stock assessment is unable to be undertaken. More generally, the Northern Prawn Fishery has species within tiers 0, 2, and 4, and all of those have scores on PCA2 > 1.0 and PCA1 < 0.0. That is, they all group in the top left of Figure 1, which distorts tiers 0, 2, and 4.

There were three separate groups within tier 2, none of which were close to each other on the PCA plot (Figure B.1): (i) Southern Bluefin Tuna (bottom left on the PCA plot) had scores of 0 to 1 for each data type, (ii) Small Pelagic Fishery's blue mackerel, Australian sardine and redbait species (bottom right on PCA plot) had no catch data, and (iii) Northern Prawn Fishery's blue endeavour and red-legged banana prawn species (top left on PCA plot) (Fig. B.1; Table B.3) had no, or poor quality, catch age- and length-composition data, whereas Southern Bluefin Tuna was scored at 0 and 1 for these data types, respectively (Table B.3). This illustrates that these data types alone do not explicitly characterise tier 2, because fisheries can still have tier 2 assessments even though there might not be length- or age-composition data (e.g. aging is not currently possible for crustaceans), or catch data. On the basis of data availability and quality, therefore, tier 2 is not coherent; data is used differently to designate tier level between fisheries. The life-history characteristics of the species also affects the assessment method selected. Within the same tier, therefore, there is very different use of data.

Similar disparity occurred within the tier 0 and 6 species/species groups: the Heard Island and McDonald Island (HIMI) fishery (tier 0) and the Western Tuna and Billfish and Skipjack Tuna fishery (tier 6) had both moderate to high quality length- and age-composition data, whereas the Northern Prawn Fishery brown and grooved tiger prawns (tier 0), and Western Deepwater and North West Slope Trawl fisheries (tier 6) did not (Fig. B.1; Table B.3). Alternatively, data may be available and yet are not necessarily used in, or appropriate for, the assessment for the species. For example, the tier 4 Bass Strait scallop fishery has fishery-independent survey data, but these data are not used in the context of a formal stock assessment, as the estimates of biomass are uncertain. The surveys are rather used to check on the location and relative size of scallop beds, as well as the proportion of under-sized and on condition of scallops prior to harvesting.

Also, data quality may be high per se, but not in the context of assessment requirements, relegating the species or fishery to a lower tier level. For example, effort data for blue endeavour prawns in the Northern Prawn Fishery is scored as high quality, (= 0), but in an assessment context, the quality of these data would be compromised because the effort is not targeted.

Additionally, the scorings are still relatively coarse. Ocean jackets, a SESSF tier 4 finfish species, and the tier 7 Coral Sea Hand Collection species assemblage, occurred in close proximity in the PCA bi-plot (Fig. B.1), due to their identical scorings of 1 for total and landed catches, and effort, and 2 for CPUE (Table B.3). While an empirical assessment is able to be undertaken for the former, the latter, being species assemblages, often with opportunistic targeting and with more temporally sporadic time data series, is currently managed via a system of catch-based triggers.

Clustering by data type followed intuitive expectation: total catch and CPUE were tightly clustered, together with reported catch and effort, while age- and length-composition data were clustered together. Fishery-independent survey data clustered separately, but closer to catch and CPUE data. Repeating the scoring exercise validated the original results: clear segregation between tiers 0 and 1, and tiers 6 and (especially) 7, and similar clustering by data types (results not shown).

Discussion

The original four SESSF tiers were defined somewhat subjectively, and primarily in the context of assessment method, and data quality and availability at the time they were developed (Smith and Smith, 2005). Here, these attributes were captured by scoring data types according to their availability and quality. The scorings were applied to species across all Australian Commonwealth fisheries within an expanded 8-level tier system, and account for data availability by assigning zeros to those data types for which there is no information.

The PCA showed that the highest and lowest tier levels were clearly segregated by the first two principal components. The remaining, intermediate, tiers showed some level of ordered separation. Having expanded the original four tiers to eight, asking experts to consider information availability and quality yielded some independent, post-hoc evaluation of these criteria as tier delineators.

The clear delineation of the “data rich” tiers 0 and 1, and the more data-limited tier 7 along the first principal component axis, was more associated with the availability of catch and effort data than with that of fishery independent survey data. The importance of catch and effort data relative to that of fishery independent survey data is interesting, given that the latter are often considered the gold standard of high-quality stock assessments.

In interpreting the results, however, it should be noted that almost half of the Australian stocks/species (19 of 52) are assigned to a single tier (tier 4). Tiers 1 and 2 have 12 and 7 stocks, respectively, while tiers 0, 3, 6 and 7 each have 5 or fewer stocks. This unbalanced distribution of stocks within tiers could affect the ability to analytically delineate tiers on the basis of only data quality and availability. Additionally, other clustering techniques, such as a k-medoids approach, could render the relationships somewhat differently

Despite the somewhat ordered trend by tier number along the first principal component axis, the extent of overlap, particularly for the intermediate tiers, highlights the possible ambiguities in defining data quality. Even if data quality could be explicitly defined, there remains a high propensity for overlap between tiers. For example, a tier 4 assessment (e.g., Little et al., 2011; Prince et al., 2011) may be based on high quality CPUE and/or length-composition information, as may a model-based tier 2 stock assessment. Australia's ETBF assessments are classified as tier 4 not because of a lack of data "quality", but because the mobility of the species is such that local fishing activities do not embrace the range of the stock.

More generally, the PCA (Fig. B.1) demonstrated that many tiers are not consistent (in terms of data availability and quality) across fisheries. In this context, data availability and quality do not appear to reflect tiers. This is perhaps illustrated most strongly within the Northern Prawn Fishery, where blue endeavour prawns (*Metapenaeus endeavouri*) have identical scores to brown and grooved tiger prawns (*Penaeus esculentus* and *P. semisulcatus*), yet the tiger prawns are tier 0 and blue endeavour prawns are tier 2. Species biology and the value of the fishery are also relevant factors, in addition to data availability and quality, when determining the type of assessment and hence assigning a tier level to a species or fishery. Complementary arguments apply in the SESSF where many tier 4 species have some age and length information, but it is not used. If the tiers do not consistently reflect the data available then data availability and quality alone are not sufficient to identify which tier a species will fall into within Australian Commonwealth fisheries.

A broader issue is the need to firmly establish the intent or aim of tier systems. Historically, the Australian tier system originated within the SESSF (Smith et al., 2008), with the aim of providing recommendations of catch for all stocks managed using Total Allowable Catches, irrespective of the data available for assessment purposes. While precautionary adjustments to catch ("buffers" or "discount factors") have been set with the intent of achieving risk equivalency across the tiers, this was not, at least initially, a formal consideration (the need for risk equivalency was discussed in Smith and Smith (2005), but this was not formally adopted).

Finally, clarifying the definition of "risk" in the context of tier systems is still required. Tiers in the USA west coast groundfish fishery and the Alaskan crab fishery account *inter alia* for the ability of the assessment to estimate management quantities (Dichmont et al., 2016). Aligning overfishing risk with assessment certainty (and hence the ability to define a probability of overfishing) is arguably more direct and explicit than the assumption that data quality and availability are directly related to this risk. Nonetheless, our analysis provides some post-hoc justification for the Australian tier delineations in terms of whether the tiers discriminate stocks (given the way assessments are conducted and decision rules applied). Confronting an expanded 8-tier system, which embraces all Commonwealth species, with expert judgement of data quality in the context of assessment assumptions, provided an objective evaluation for defining tiers on this basis.

Acknowledgements

Rik Buckworth, Cathy Bulman, Robert Campbell, Mark Chambers, Richard Hilary, Tim Skewes and Geoff Tuck are thanked for their assistance with data scoring. This work was funded by the Fisheries Research and Development Corporation Project 2012/202 and CSIRO Oceans and Atmosphere

Flagship. Chris Legault, an anonymous reviewer, and the Guest Editor are thanked for their comments on an earlier draft of this paper.

References

DAFF. (2007) *Commonwealth Fisheries Harvest Strategy Policy Guidelines*. Australian Government Department of Agriculture, Fisheries and Forestry, Canberra, Australia, pp. 55. http://www.agriculture.gov.au/fisheries/domestic/harvest_strategy_policy

Dichmont, C.M., Punt, A.E., Dowling, N., De Oliveira, J.A.A., Little, L.R., Sporcic, M., Fulton, E., Gorton, R., Klaer, N., Haddon, M., Smith, D.C., 2016. Is risk consistent across tier-based harvest control rule management systems? A comparison of four case studies. Fish and Fish. doi: 10.1111/faf.12142.

Dowling, N.A., Smith, D.C., Knuckey, I., Smith, A.D.M., Domaschenz, P., Patterson, H.M., Whitelaw, W., 2008. Developing harvest strategies for low-value and data-poor fisheries: case studies from three Australian fisheries. Fish. Res. 94,380–390.

Dowling, N.A., Dichmont, C.M., Venables, W., Smith, A.D.M., Smith, D.C., Power, D., Galeano, D., 2013. From low- to high-value fisheries: Is it possible to quantify the trade-off between management cost, risk and catch? Mar. Pol. 40, 41–52

ICES. 2012. ICES Implementation of Advice for Data-limited Stocks in 2012 in its 2012 Advice. ICES CM 2012/ACOM 68. 42 pp.

Little, L.R., Punt, A.E., Dichmont, C.M., Dowling, N., Smith, D.C., Fulton, E., Sporcic, M., Gorton, R.J., 2015 Defining options for cost constrained fisheries management. ICES J. Marine Science. doi:10.1093/icesjms/fsv206

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 J. Mar. Sci. 68, 1699–1705.

North Pacific Fishery Management Council (NPFMC). 2014. Stock Assessment and Fishery Evaluation Report for the king and Tanner crab fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, 605 W. 4th Avenue, #306, Anchorage, AK 99501.

Pacific Fishery Management Council (PFMC). 2014. Status of the Pacific Coast Groundfish Fishery: Stock Assessment and Fishery Evaluation. Pacific Fishery Management Council, 7700 Ambassador Place, Suite 101, Portland, OR 97220.

Prince, J.D., Dowling, N.A., Davies, C.R., Campbell, R.A., Kolody, D.S., 2011. A simple cost-effective and scale-less empirical approach to harvest strategies. ICES J. Mar. Sci. 68, 947–960.

Sainsbury K., 2005. Cost-effective management of uncertainty in fisheries. Outlook 2005. Canberra, A.C.T.: Australian Bureau of Agricultural and Resource Economics. [Accessed online, 20/05/2015: http://data.daff.gov.au/data/warehouse/pe_abarebrs99001173/PC13024.pdf]

Smith, A., Smith, D., 2005. A harvest strategy framework for the SESSF. Report to the Australian Fisheries Management Authority (AFMA), Canberra, June 2005, Available from CSIRO Oceans and Atmosphere, GPO Box 1508, Hobart, Tasmania, Australia. 7pp.

Smith, A.D.M., Smith, D., Tuck, G., 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, L.R., 2008. Experience in implementing harvest strategies in Australia’s south-eastern fisheries. Fish. Res. 94, 373–379.

Smith, A.D.M., Smith, D.C., Haddon, M., Knuckey, I.A., Sainsbury, K.J., Sloan, S.R., 2014. Implementing harvest strategies in Australia: 5 years on. ICES J. Mar. Sci. 71, 195–20

Appendix C. Decision trade-offs for cost-constrained fisheries management

L. Richard Little[1], André E. Punt[1,2], Catherine M. Dichmont[3], Natalie Dowling[1], David C. Smith[1], Elizabeth A. Fulton[1], Miriana Sporcic[1], Rebecca J. Gorton[1]

[1] *CSIRO Oceans and Atmosphere, GPO Box 1538, Hobart TAS 7000, Australia*

[2] *School of Aquatic and Fisheries Sciences, Box 355020, University of Washington, Seattle WA, USA*

[3] *CSIRO Oceans and Atmosphere, Ecoscience Precinct, Dutton Park QLD 4102, Australia*

Published in Little, R.L., Punt, A.E., Dichmont, C.M., Dowling, N., Smith, D.S., Fulton, E.A. 2016. Decision trade-offs for cost-constrained fisheries management. ICES Journal of Marine Science. 73 (2): 494-502. doi: 10.1093/icesjms/fsv206

Abstract

Fisheries management operates in an environment characterized by multiple risks. These risks are often complementary, and can be traded off against each other. An important goal for managers is to develop strategies to minimize the overall risk exposure at minimal cost. We show a simple model that quantifies a range of risks faced by fisheries management agencies in terms of long-term expected budgetary expenditure. The model includes not only the cost a management agency would be expected to incur from overfishing a stock, or from being seen to overfish it, but also the social cost incurred from not achieving its objectives, such as the opportunity cost of foregoing catches and economic returns. These costs can be controlled by adjusting the biomass level targeted by management, or increasing expenditures for data collection to improve the precision of biomass estimates. The overall risk, expressed as the long-term total expected cost to a management agency, depends strongly on the fisheries management objectives, and the emphasis on conservation or economic motives. In general, management under a conservation-oriented objective would reduce risk either by increasing target biomass levels, or expenditure on monitoring and assessment, while a catch-focused objective would seek to lower management costs by reducing expenditure on data collection and assessment. Increased natural stock variability affects the risk of overfishing, and long-term expected costs as the ability to make a meaningful estimate of biomass declines. Management of catch-focused fisheries would reduce the biomass target as stock variability increases, because the benefit of catches are seen to outweigh the cost, or risk of being overfished. The model provides the basis for more extensive risk analyses, and serves as a simple lesson that the consequences of reducing the short-term costs associated with managing a fishery, can come with a concomitant increase in overall risk.

Key words: risk-cost-catch trade-off; false positive estimation error; false negative estimation error; fisheries management cost; risk equivalency

Introduction

Fisheries management is the art of balancing risk with cost (Dowling *et al*., 2013). The risk faced is the failure to achieve management objectives, either explicitly stated or not, relating to whether a stock is seen to be at, or on track to, a target, and above a lower limit. The cost is seen not only as the operational resources used to monitor and support management decisions, but also the expected funds that are required to rebuild a stock that has crossed the lower limit, the social costs of forgone profit or catches from an over-fished stock, and even the intangible cost to management of being seen as ineffective or worse.

A significant cost to management is the operational support required for monitoring and assessment activities, which feed into decision-making processes, often in the form of a harvest control rule (Deroba and Bence, 2008; Punt *et al.*, 2014). Assessments on which management decisions are based however, vary. Ralston *et al*. (2011) for example, characterized the variability of estimated biomass in assessments of US west coast groundfish at 36%, expressed as a coefficient of variation. The nature of the variability associated with an assessment is complicated because a stock could be below a lower limit and not be detected as such, with potential for fishery collapse, or it could be above the limit and erroneously perceived to be below it (Mapstone, 1995). The cost of the latter error can be substantial. Widow rockfish, *Sebastes entomelas*, off the US west coast, was erroneously assessed to be below the overfished level, which resulted in closure of the directed midwater trawl fishery (PFMC, 2014). Even if a stock is correctly estimated to be below a limit, a rebuilding strategy is usually required, along with the financial resources for administrative and scientific support to implement it (Knuckey et al. 2009). Recovery costs of Atlantic salmon, *Salmo salar*, in the United States for example, have been estimated to be at least USD$36M (National Marine Fisheries Service and the U.S. Fish and Wildlife Service, 2005).

In general, risk management involves avoiding, retaining or transferring risk (Sethi 2010). Fisheries management typically retains risk, but controls or manages it through monitoring efforts that feed into decision processes that use management tools such as Total Allowable Catches (TACs), and target biomasses. In principle, adopting a higher target biomass (i.e. a higher target reference point), can reduce the risk of crossing a limit, and the associated management costs of doing so, because the probability that a stock is below, or is perceived to be below a limit, decreases with increasing target level. However, such an action would affect the sustainable yield and economic returns of the fishery, which from a single species perspective, are considered highest at an intermediate level of biomass (Hilborn, 2010; Punt *et al*., 2014), thus providing an incentive to target these biomass levels. Although in practice maximum sustained catches can be only identified with uncertainty (Clark 1991, 1993; Punt *et al*. 2014), a fairly broad range of biomasses can be targeted that would result in close to maximum catch (Hilborn, 2010).

Alternatively, increased monitoring and assessment can reduce the risk of crossing a limit. This has been well documented in single stock contexts (Bergh and Butterworth, 1987; McDonald *et al.*, 1997). The cost effectiveness of fisheries monitoring and assessment however, has not been very well addressed, but it is coming under increasing attention as fisheries management budgets are scrutinized.

Increasingly, the three-way relationship between the risk of over-fishing a stock, the catches that are derived from it, and the costs of management is being seen as a risk-catch-cost frontier (Sainsbury, 2005). We suggest that given an accepted or allowed amount of risk (e.g. DAFF 2013), there may be risk equivalent decision trade-offs between management costs and catch, whereby the level of risk is maintained albeit in potentially different ways.

Here we distil this trade-off down to a single combined quantitative monetary cost measure, composed of 1) the *risk* of over-fishing represented as a long-term expected cost, 2) *catches*, or more accurately, the lack thereof, represented as an opportunity cost, and 3) *management costs* associated with monitoring and assessment represented as the short-term operational requirements. We take an alternative approach to calculating optimal monitoring costs (Field *et al.,* 2004), by showing trade-offs associated with monitoring and assessment. This provides a quantitative theoretical method that could be used to guide managers who wish, or are required, to minimize the total expected management costs, through a trade-off between investing in monitoring and assessment, and adjusting the target biomass.

We consider assessment variability and also 'management uncertainty' that relates to the ability of management to sustain stocks at chosen target biomass. This uncertainty is partially related to uncertainty in the estimate of current biomass levels, and also to 'implementation uncertainty' whereby the selected regulations may not achieve the expected goals (Francis and Shotton, 1997; Dichmont *et al.*, 2008; Fulton *et al.*, 2011). Finally, and consistent with most past work that has considered how to select target biomass levels (e.g., Clark, 1991, 2002; Punt *et al.*, 2014), we ignore transient effects and focus on the choice of targets even though the costs of moving a stock towards a target level, including implementing a rebuilding plan, can be substantial.

Methods

The analysis is based on an objective function representing the long-term total expected management cost, and consisting of several components (Field *et al.*, 2004), including administrative costs, the cost of implementing management actions whether they are needed or not, the cost of being below a lower limit but not recognizing it, the catch as a function of biomass, and the cost of data collection and assessment. The total expected cost thus, is given by:

$$E[cost] = \alpha R + \beta V + \gamma R - \rho P + M \tag{1}$$

where α is the probability of estimating the stock to be below the limit when this is not the case, β is the probability of estimating the stock to be above the limit when this is not the case, and γ is the probability of correctly estimating the stock to be below the limit reference point (Figure C.1). R represents the cost of management action based on the perception that the stock is below the limit. This cost could include the opportunity cost of foregone profit or revenue from closing or severely constraining the fishery, and the cost of management in the form of additional surveys and monitoring of the stock. It also reflects the reduced fishing opportunities for 'healthy' species that co-occur with the stock deemed to be below the limit. V represents the 'cost' to the fishery that is presumed if the stock unknowingly declines below the limit. This intangible cost would be difficult to measure, but can be used to represent the weight of a conservation objective associated with overfishing. P represents the revenue as catches or sustainable yield, ρ represents a conversion factor between catch and economic return (e.g. price), and M represents the cost of assessment and data collection, which we consider is related to the precision with which the stock status is estimated (i.e. assessment error).

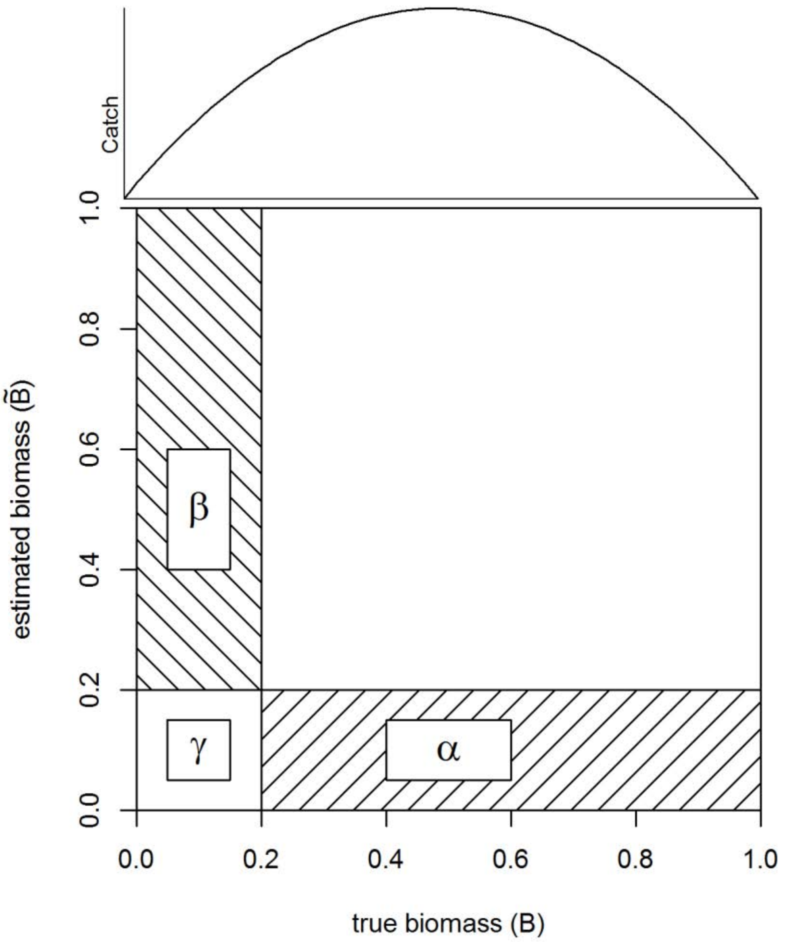


Figure C. 1 Bottom panel: Conceptual overview of the risk associated with estimating stock status when the limit reference point is set to 20% of the unfished level. The area depicted by α represents the probability of erroneously estimating the biomass to be below 0.2; the area depicted by β represents the probability of erroneously estimating the biomass to be above 0.2; and the area depicted by γ represents the probability of correctly estimating the biomass to be below 0.2. Top panel: The assumed relation between catch or surplus production and (true) stock biomass, which for this example was specified by a logistic function.

The probabilities α, β and γ depend on the value for the limit (τ), the true biomass (B) and the assessment error (σ). Under the assumption that the assessment is unbiased and assessment outcomes are lognormally distributed about the true biomass (B), the probability that the estimated biomass $\tilde{B}$ is below the limit τ is:

(2)

$$P\left(\tilde{B} < \tau \middle| B, \sigma\right) = \int_0^\tau \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \ln B)^2}{2\sigma^2}} dx$$

where σ is the assessment variability, defined as the standard deviation of the logarithm of estimated biomass. The values for α, β, and γ are then defined as follows:

$$\alpha(\tau, B, \sigma) = \begin{cases} 0 & \text{if } B \leq \tau \\ P(\tilde{B} < \tau | B, \sigma) & \text{otherwise} \end{cases} \quad (3a)$$

$$\beta(\tau, B, \sigma) = \begin{cases} 1 - P(\tilde{B} < \tau | B, \sigma) & \text{if } B \leq \tau \\ 0 & \text{otherwise} \end{cases} \quad (3b)$$

$$\gamma(\tau, B, \sigma) = \begin{cases} P(\tilde{B} < \tau | B, \sigma) & \text{if } B \leq \tau \\ 0 & \text{otherwise} \end{cases} \quad (3c)$$

The assessment and monitoring cost, $M,$ is assumed to be inversely proportional to the assessment variability σ, i.e.:

$$M = \frac{\delta}{\sigma} + K \quad (4)$$

The coefficient δ represents the effect of σ on management costs M; at low values of δ, management costs M are affected less by σ, and more by K, the fixed costs of managing the resource, which would include meetings to discuss management arrangements, and processing data.

The relation between management cost, M, and assessment variability, σ is illustrated in Figure C.2, based on data from eight stocks in the Australian Southern and Eastern Scalefish Fishery (SESSF). The SESSF is a multispecies and multi-gear fishery that provides fresh seafood to the major fish markets in Sydney and Melbourne, Australia (Smith *et al.* 2014). In 2010, it had a gross value of production of 89 million AUD (Woodhams *et al.* 2011), and management involves over 30 stocks under a tiered system of harvest control rules corresponding to data availability and type of assessment (Smith *et al.* 2014). Dedicated acoustic surveys have been conducted for orange roughy, *Hoplostethus atlanticus*, which was perceived to be over-fished, but the fishery has recently re-opened. Blue grenadier, *Macruronus novaezelandiae*, an economically important but episodic species, has an industry-based acoustic survey to estimate stock status (Punt *et al.* 2015). An industry-based fishery-independent survey has been conducted for other key target species since 2008 to provide a measure of relative abundance for estimating stock status (AFMA 2009).

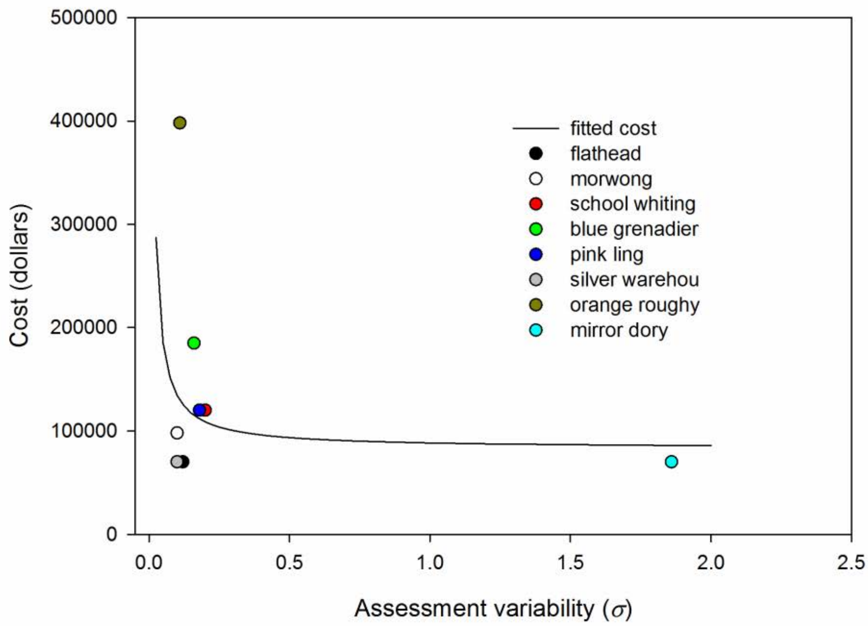


Figure C. 2 Fitted relation between stock assessment costs and the variability in estimates of biomass based on the most recent data from eight stocks in the Australian SESSF (flathead: *Neoplatycephalus richardsoni*, morwong: *Nemadactylus macropterus*, school whiting: *Sillago flindersi*, blue grenadier *Macruronus*

***novaezelandiae*: pink ling: *Genypterus blacodes*, silver warehou: *Seriolella punctate*, orange roughy: *Hoplostethus atlanticus*, mirror dory: *Zenopsis nebulosa*).**

We included survey costs into the SESSF management costs, and fitted a function of the form as in Equation 4, to give a value of δ equal to 204 AUD and K equal to 83,381 AUD. The relationship, was strongly dependent on three points: the two stocks with dedicated surveys: orange roughy (>300K AUD) and blue grenadier (>100K AUD) and relatively precise biomass estimates (CVs of 0.11 and 0.16 respectively), and mirror dory, which had a large CV (1.86) associated with the most recent biomass estimate, which was relatively inexpensive to determine (70,000 AUD).

Economic production ρP, is assumed to be related to sustainable yield P and specified as a logistic function of biomass (Figure C.1, top panel), in such a way that Maximum Economic Yield (MEY) is achieved at 0.5 of the unfished level, and scaled to unity, i.e.:

$$P = 4B(1 - B) \quad (5)$$

Figure C.1, bottom panel, shows the relationship between α, β and γ for the case where the limit, τ, is 20% of the unfished biomass (a common reference point used in management of major fisheries; e.g. Rayns, 2007; Ministry of Fisheries, 2008). Equation 1 is henceforth expressed relative to R to reduce the number of parameters, giving:

$$E[cost] = \alpha + \beta V/R + \gamma - \rho P/R + M/R \quad (6)$$

The true biomass (B) will vary over time given environmental uncertainty, assessment error and the ability to implement regulations. Consequently, if management targets a particular biomass level, $\widetilde{D}$, it will achieve it with uncertainty. Biomass thus is assumed to vary around the target $\widetilde{D}$ according to a lognormal distribution, with variability defined by $\tilde{\sigma}$, the measure of the true underlying biomass variability. The long-term expected cost $E\left[cost\middle|\widetilde{D},\tilde{\sigma},\sigma,\delta\right]$ given the choice of target $\widetilde{D}$, the environmental variability $\tilde{\sigma}$, effect of management costs δ, and the assessment error σ is:

(7)

$$E\left[cost\middle|\widetilde{D},\tilde{\sigma},\sigma,\delta\right] = \int_0^\infty \frac{1}{B\sqrt{2\pi}\tilde{\sigma}} e^{\frac{(lnB - ln\widetilde{D})^2}{2\tilde{\sigma}^2}} E[cost|B,\sigma,\delta]dB$$

$$= \int_0^\infty \frac{1}{B\sqrt{2\pi}\tilde{\sigma}} e^{\frac{(lnB - ln\widetilde{D})^2}{2\tilde{\sigma}^2}} \left[\alpha\left(\widetilde{D},B,\sigma\right) + \beta\left(\widetilde{D},B,\sigma\right)\frac{V}{R} + \gamma\left(\widetilde{D},B,\sigma\right) + M(\sigma,\delta) + \rho P(B)\right]dB$$

The target biomass $\widetilde{D}$ and assessment variability σ represent management controls that can be used to adjust $E\left[cost\middle|\widetilde{D},\tilde{\sigma},\sigma,\delta\right]$. It is assumed that assessment variability (σ) and biomass variability ($\tilde{\sigma}$) are independent of each other, and that assessment variability is directly controlled by monitoring and data collection i.e. through management costs M. For example, σ would be smaller as management expenditure increases according to Equation 4.

The long-term expected cost thus is shown as being controlled by σ through its effect on M, and by the management target ($\widetilde{D}$) through its complementary effects on catches and the risk of over-fishing. We show the long-term expected cost, $E\left[cost\middle|\widetilde{D},\tilde{\sigma},\sigma,\delta\right]$, for different values of the underlying stock variability $\tilde{\sigma}$, the effect of increased assessment precision on management costs, δ, and also under a range of weights attached to the cost function, ρ/R and V/R. The range of ρ/R and V/R selected assigns different weight to each component of the cost function, $E\left[cost\middle|\widetilde{D},\tilde{\sigma},\sigma,\delta\right]$. Because the relation among δ, ρ, V and R in Equation 6 is likely to be finely balanced for a particular fishery, and we did not have access to the full range of data needed to apply the analysis to a specific fishery such as the SESSF, our analysis shows general principles behind the risk-cost-catch trade-off, and does not pertain to a specific example fishery.

Results

Risks associated with overfishing

The probability that a stock is correctly estimated to be below a limit of 20% is close to 1.0 when the true biomass is less than 20% and the assessment σ is low (Figure C.3). This probability decreases as σ increases. In contrast, the probability is 0 when the true biomass is greater than the 20% limit and the assessment variability is low, but increases to 0.5 as the σ increases.

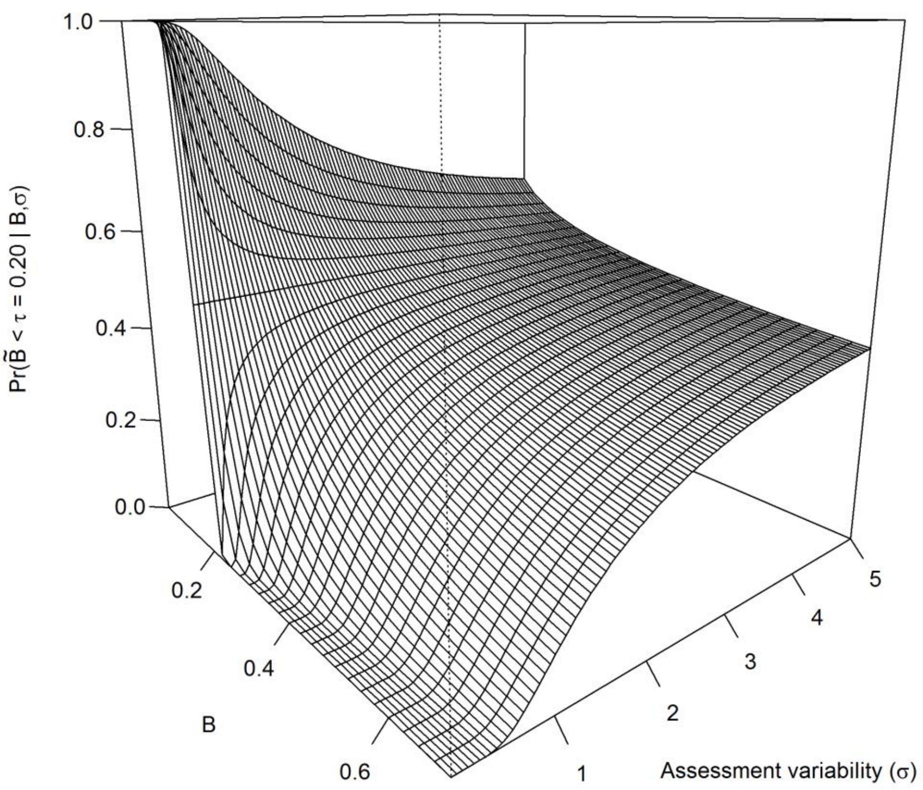


Figure C. 3 The probability that the estimated biomass ($\tilde{B}$) is below a limit reference point, τ of 20% of the unfished biomass as a function of the precision of the assessment (σ) and the true biomass (B).

The risk or expected cost to management from incorrectly estimating the stock size (α and $\frac{V}{R}\beta$), as well as the expected cost of having to rebuild a stock (γ) are all influenced by the precision of the assessment (assuming it is unbiased), and the biomass target that management chooses (Figure C.4; top panels). The expected cost of incorrectly claiming the stock has crossed the limit, when in fact it has not, (i.e. α , or the false positive error rate) is highest when the target biomass is close to the limit, and declines as the target biomass level increases past the limit (Figure C.4; top left panel). The expected cost of such an error also increases as the variability in the assessment estimate increases. There is little expected cost from a false positive error when the target is below the 20% limit, a situation which is not likely to result.

The expected cost of a false negative error ($\beta \frac{V}{R}$), i.e. incorrectly concluding the stock is above the limit is highest just below the limit, and when the assessment variability is high (Figure C.4; top middle panel). It decreases both as the target moves away from the limit, and as the assessment variability decreases. The expected cost associated with correctly estimating the stock to have crossed the limit (γ) also declines quickly as the target biomass increases away from the limit biomass (Figure C.4; top right panel). A complementary effect between $\beta \frac{V}{R}$ and γ is seen when the cost associated with γ increases, but $\beta \frac{V}{R}$ decreases, as the assessment variability declines.

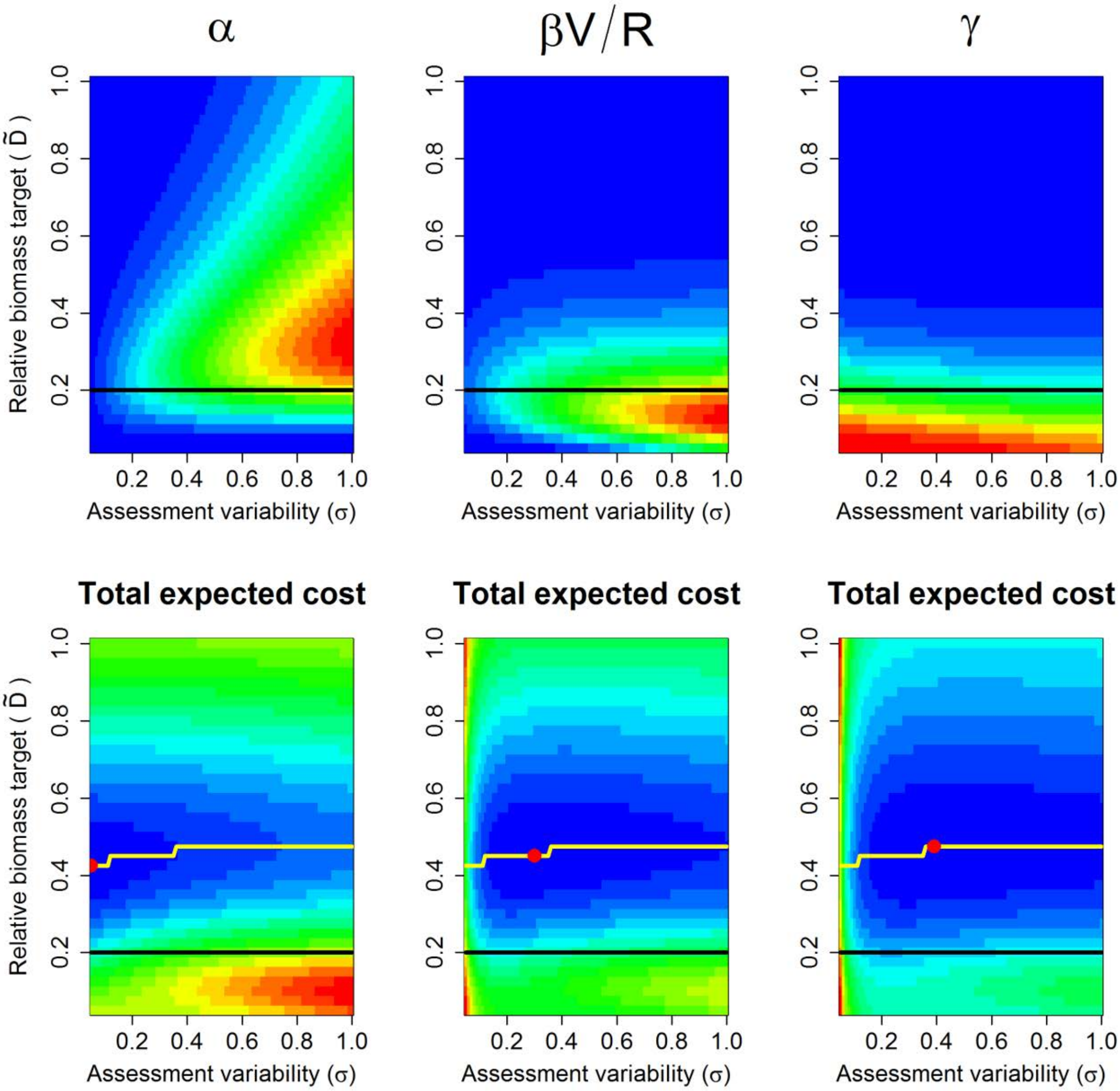


Figure C. 4 Top panels: Expected cost of components α, β, γ, and total expected cost as a function of relative biomass target and assessment variability (σ) under condition $V/R = 5$, $\rho/R = 1$ and $\tilde{\sigma} = 0.5$. Bottom panels: The expected cost totalled across all components in EQ 2 for three choices of δ (0, 0.1 and 0.2), and for $K = 0$. In the lower panels, red points indicate the global surface minimum, yellow lines the target biomass associated with the minimum cost under a given assessment variability, and the black lines the 20% limit reference point. Warm colours are higher associated costs.

Total expected cost

The combined total expected cost across all components of Equation 6 is lowest at intermediate target biomass levels (yellow line in Figure C.4 lower panels). This line represents the best trade-off between the cost associated with crossing the 20% limit (α , $\frac{V}{R}$, and γ), and the social cost of forgoing catch or revenue ($\rho P/R$; Equation 6). These results also show that management should increase the target biomass as assessment variability increases to minimize total expected cost. A global minimum point occurs on these cost surfaces (red points in Figure C.4; lower panels), representing the optimal combination of assessment variability (σ) and biomass target ($\widetilde{D}$) that results in the global minimum cost. Any deviation from such a point, in a management context, would ultimately result in higher long-term expected costs.

The global minimum cost occurs at the most precise assessment (i.e., at $\sigma = 0$, the red point in Figure C.4 bottom left panel for $\delta = 0$) when data collection and assessment cost (M) are independent of assessment precision (σ). However, this global minimum shifts to less precise assessments as assessment costs increase (i.e. for $\delta > 0$), because a lower σ increasingly adds to the cost through M (Figure C.4 bottom middle, bottom right panels). The highest costs on the cost surface (red areas, Figure C.4, bottom panels), thus shift away from those associated with low biomass targets (horizontal red area, Figure C.4 bottom left panel) to those associated with high assessment precision, i.e. low σ (thin vertical red band on left of Figure C.4 bottom centre and right panels).

The biomass target used by a management agency to minimize the total expected cost would, in general, increase as the assessment variability increases, with the underlying stock variability ($\tilde{\sigma}$) tending to reduce the effect (Figure C.5; top left panel). The effect of assessment variability declines as the focus on revenue dominates the objective function (through ρ/R, Figure C.5; top right panel). When this happens, increasing variability in stock dynamics ($\tilde{\sigma}$) decreases the optimal target biomass toward the limit, as estimates of the underlying stock are obscured, and the management objective is to maximize production through harvest.

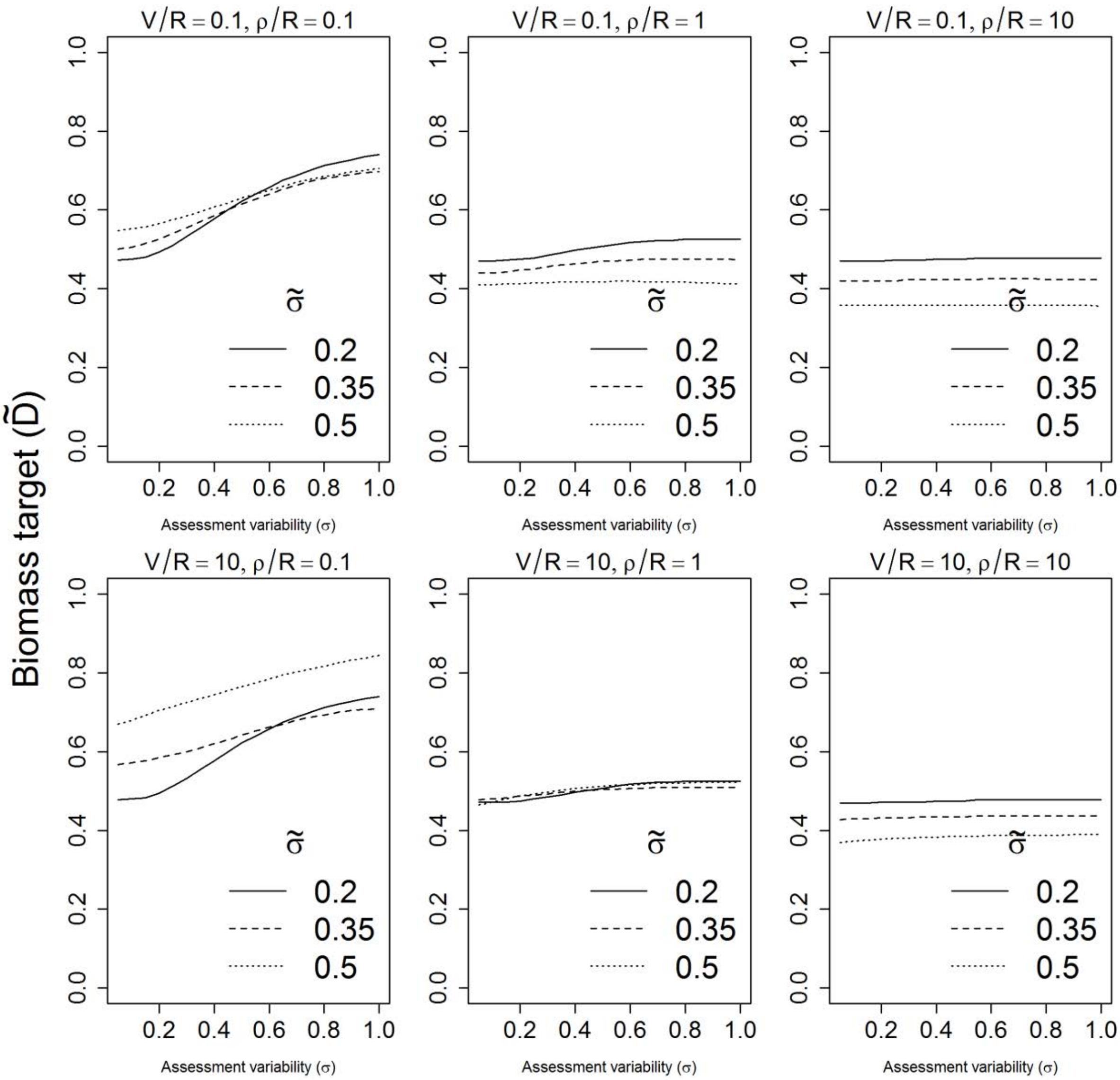


Figure C. 5 The target biomass that corresponds to the minimum expected management cost as a function of assessment variability, under different weights attached to fishery revenue (ρ/R), and stock conservation (V/R), i.e. ensuring the stock does not unknowingly drop below the limit threshold, for δ=0.2, for $K = 0$, and $\tilde{\sigma}$ (uncertainty in stock dynamics) values of 0.20, 0.35, and 0.50.

Management would set higher biomass targets to minimize total expected costs when the weight on conservation increases, i.e. as V/R increases (Figure C.5 bottom left panel) especially when the stock dynamics are noisy ($\tilde{\sigma} = 0.5$). The interaction between increasing both revenue and conservation (Figure C.5 bottom right panel) indicates that the sensitivity to the conservation cost weighting (from $V/R = 0.1$ to 10) is low when the economic weighting is strong ($\rho/R = 10$).

In general, costs are minimized when the biomass target is close to 0.5 and the true underlying stock variability, $\tilde{\sigma}$, is low. However, this target changes depending on the economic motive of management and the underlying stock variability. The target biomass that minimizes total expected costs increases with increasing levels of true biomass variability, $\tilde{\sigma}$ to a maximum of 1.0 in fisheries where management places relatively low weight on revenue (Figure C.6, left panels). This occurs because management would target a biomass of 1.0, and thus close the fishery (Figure C.6 left panels) at high levels of $\tilde{\sigma}$, given the high uncertainty in terms of achieving desired outcomes, and the potential costs of overfishing. Increasing the weight on conservation (i.e. increasing *V/R* from higher to lower panels in Figure C.6) decreases the value of stock uncertainty ($\tilde{\sigma}$) that closes the fishery (i.e. by setting the target biomass at 1.0).

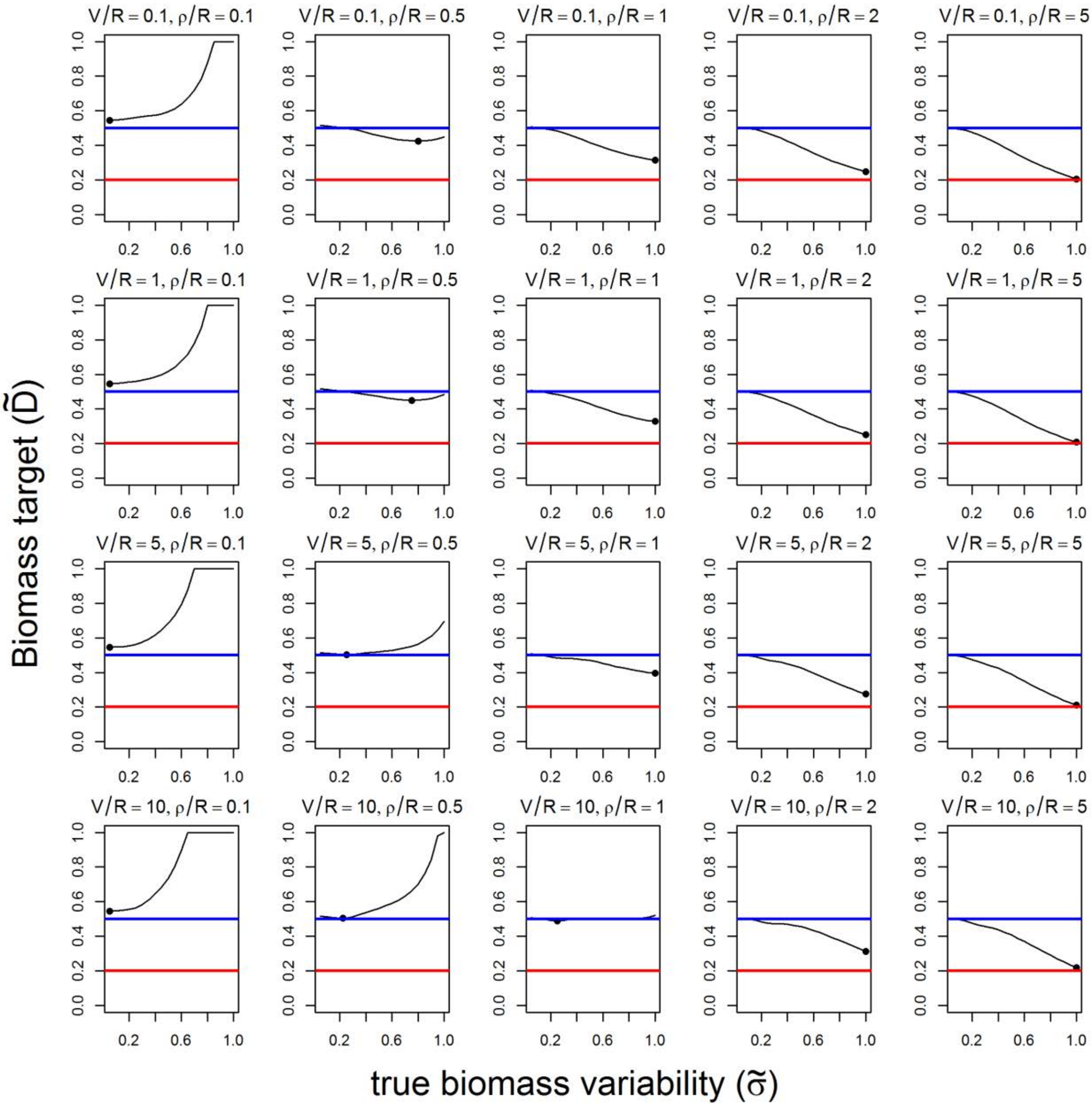


Figure C. 6 The target biomass $\tilde{D}$ that is associated with the minimum expected cost $E[cost|\tilde{D},\tilde{\sigma},\sigma]$, as a function of the true underlying stock variability ($\tilde{\sigma}$) (black lines). Panels show different weights attached to fishery revenue (ρ/R) and stock conservation (V/R), i.e. ensuring the stock does not unknowingly drop below the limit threshold, with σ =35% and δ=0.2, $K = 0$. The blue and red solid lines indicate the biomasses corresponding to the maximum revenue and the limit reference point. The solid dots indicate the minimum point on the curve.

The optimal target biomass tends to decline with increasing levels of stock variability ($\tilde{\sigma}$) as revenue dominates the objective function (i.e. ρ/R increases from 0.1 to 10, Figure C.6, left to right panels) This occurs because increasing stock variability obscures the ability to measure stock status. When management is dominated by an economic motive, the result is simply to increase harvest by reducing the biomass target level. The interaction between the revenue and conservation in the objective function again indicates that the sensitivity to the conservation cost weighting (from $V/R = 0.1$ to 10) is low when the economic weighting is strong ($\rho/R = 10$, Figure C.6; bottom right).

Discussion

Fisheries management agencies are constrained by costs. These include not only the direct management costs of monitoring and assessing a stock, which are easily seen as a principle demand on an operating budget, but also the episodic costs of rebuilding stocks if this is needed. These episodic costs represent a risk, and management agencies that seek to minimize management costs in totality, must address such risk.

The risk to a management agency however, is not only that the stock will fall below a limit and require rebuilding efforts, but also that the stock will be perceived to fall below it. These can be quite different (Little *et al*., 2014). The cost of a stock that is correctly seen to have crossed a limit is related to the probability of correctly estimating the stock status, γ. Our analysis showed that as assessment precision (σ) increases, or the management target ($\widetilde{D}$) increases away from the lower limit, γ declines, and with it the associated cost.

A stock may also be incorrectly perceived to have crossed a limit (false positive error), or it may have crossed a limit without being perceived to have done so (false negative error). These errors also have potential costs if a management agency unnecessarily spends money on rebuilding measures (PFMC, 2014), or unwittingly misses an opportunity to detect dangerously low stock levels. Failure to detect or act on the detection of dangerously low stock levels, may have contributed to the collapse of northern cod, *Gadus morhua* (Walters and Maguire, 1996). Our analysis showed how these costs might be affected by changing the target biomass ($\widetilde{D}$), or by adjusting the precision of the biomass estimate (σ) through monitoring and assessment. Both of these options have consequences, however. Increasing the target biomass beyond that associated with maximum economic yield would likely reduce catches and thus economic production, while increasing the precision of the estimate of biomass, would likely increase the direct management costs of monitoring and assessment. We showed the aggregate trade-offs in the component costs of the cost function (Equation 1), and how they can be balanced to minimize the overall expected cost.

Implicit in this balancing is the weight placed on the different components, particularly the economic and the conservation imperatives. These weights will differ even between fisheries with similar management arrangements (e.g. Grafton *et al*., 2007), and ultimately will depend on the objective of management. For example, fisheries with a strong economic motivation, and sensitive to the risk of lost catches and economic returns, would place a high value on ρ/R, relative to V/R, and thus might focus more on setting the target biomass to minimize long-term expected costs, rather than efforts to increase assessment precision. In such fisheries, stocks with high variability ($\tilde{\sigma}$) are expected to be targeted at lower levels as more weight is placed on the risk of lost economic production than on being periodically overfished (Figure C.5, top right panel). Such fisheries would have little need for increasing the precision of the assessment (σ) as this would add to management costs, and so might forgo annual assessments and implement multi-year TACs (e.g. Smith *et al*., 2008). Alternatively, fisheries that place a high emphasis on the risk of overfishing a stock, such as those with Marine Stewardship Council accreditation (Gulbrandsen, 2009), would undoubtedly place greater emphasis on monitoring and assessment.

In general, the risk-catch-cost frontier posits that high risk, high catch stocks should also result in high cost fisheries management, and as management moves away from high risk and high catch conditions, the costs should correspondingly decline (Sainsbury, 2005). The exception to this occurs when a stock is already considered to be overfished. Such a fishery would typically have low or nil catches, but high risks, and high costs as rebuilding and recovery efforts are implemented (Dowling *et al.*, 2013). We captured the risk-catch-cost relationship in a single scalar monetary value $E\left[cost\middle|\widetilde{D},\tilde{\sigma},\sigma,\delta\right]$, with risk being represented by the expected cost associated with the terms α, β, and γ; catch represented by the revenue associated with the term P, which was affected through the choice of management target, $\widetilde{D}$;

and cost represented by the operational management costs associated with monitoring and assessment, M and affected through the choice of assessment precision, σ.

The model and analysis was a relatively simple representation of fishery conditions, with little reliance on specific fishery data. Such an approach comes with drawbacks, and assumptions. First is that the model and results relied on comparative statics, and measuring the long-term effects of management policy on equilibrium conditions. Such an approach does not consider rate or path of the fishery to equilibrium, or even the possibility that equilibrium conditions may not be achieved (Anderson and Cavendish, 2001). For example, we assumed that fishery economic productivity was related to the equilibrium surplus production, irrespective of a current or initial state of a fishery. Thus, the results should not be applied to any specific fishery, but instead provide a guideline to management agencies that are considering reducing operational costs associated with monitoring and assessment. The general conclusion is that short-term action to reduce management costs (M) could potentially have unintended consequences, by increasing the long-term costs of dealing with an over-fished stock, or an apparent over-fished stock.

Applying this analysis to a specific fishery would require a stochastic dynamic simulation approach that considers not only fishery-specific parameter values, but also the current stock state, or perceived stock state. Such an approach could calculate the present value of expected cost $E[cost]$ and thus explicitly consider time and path dependency, and relate revenue to a harvest level derived from a harvest control rule, which itself would depend on estimated biomass. Although, untangling these effects could be addressed using dynamic simulation models and management strategy evaluation (e.g. Cooke, 1999; Dichmont *et al.* 2008; Fay *et al.*, 2011), the details of conducting the projections could obscure the general principles identified in the approach of this paper.

Another less obvious assumption is the independent relationship between the assessment variability σ, and both the true underlying stock variability, $\tilde{\sigma}$ as well as biomass B. An alternative result might be that as biomass declines, the assessment variability does too. If this happens, it would become easier to determine whether the stock crosses the limit as it declined, and thus a lower target biomass might reduce risk. It would be correspondingly more difficult to detect when the stock crossed the threshold if assessment variability increased as the stock biomass declined, and the optimal risk strategy might be to set higher target levels.

Another important assumption we have made is that reducing assessment variability σ will cost more. While we based this assumption on Figure C.2, the relationship we calculated was strongly influenced by three of the data points. This however, was a first pass; more detailed cost data are needed to explore this further.

The uncertainty associated with estimating stock status σ is also multi-faceted (Francis and Shotton, 1997; Fulton *et al.* 2011), because increased expenditure on monitoring might reduce *observation uncertainty*, but not the *model uncertainty* associated with assessment, for several reasons. First, mismatch between the life history or ecology of a stock and the assumed population dynamics might make obtaining an accurate, unbiased estimate of stock status difficult; or second, catch-per-unit-effort (CPUE) might be assumed to be linearly related to biomass, but instead exhibit hyper-stability, or third a stock might be assumed to be homogenously distributed in a stock assessment model, but in reality form a meta-population.

As a result, it might not always be possible to reduce σ with increased expenditure. Nevertheless, monitoring and assessment remain critical activities for managing fish stocks, because they provide information for setting control variables such as TAC or total allowable effort (TAE). In response to confidence in data and analysis used to estimate biomass, management agencies have started to invoke tiered level management, with the intention that TAC or TAE recommendations are tempered by the risk associated with the consequence of errors. In the Australian SESSF, a risk premium is attached to TAC recommendations resulting from catch curve or CPUE analysis to reflect the greater uncertainty, lower precision, and lower amounts of data used by these methods (Smith *et al*., 2014). Risk premiums have been found to be stock-specific (Fay *et al*., 2012), and methods to accurately represent the risk between tier levels are currently being developed (Little *et al*., 2014).

Whether the financial costs of either *monitoring* or *assessment* could be better used on other measures of protection and conservation (McDonald-Madden *et al*., 2010; Legg and Nagy, 2006) or are cost effective (Boyce *et al*., 2012) are important management questions. Trade-off analyses on marine ecosystems have been examined (Fulton *et al*., 2014), but have typically focused on the mean or expected cost outcome without consideration of the more extreme outcomes that could eventuate. The application of value at risk (VaR) approaches (Sethi *et al*., 2012) to measure the extremes of the cost distribution would provide a broader perspective than the focus on expected values used here. Nevertheless, management budgets are typically based on expectation, and by applying the rules presented here, a management agency can adjust either the biomass target, or their investment in data collection, to understand the larger picture in minimizing the combined risks and associated costs of fisheries management.

Acknowledgements

This research was supported by CSIRO and the Fisheries Research and Development Corporation. We acknowledge and appreciate the comments of two anonymous reviewers and the editor on an earlier draft of this paper.

References

AFMA 2009. Harvest Strategy Framework: for the Southern and eastern scalefish and shark fishery 2009 (Amended February 2014). Australian Fisheries Management Authority, Canberra, Australia.

Anderson, D. and Cavendish W. 2001. Dynamic simulation and environmental policy analysis: beyond comparative statics and the environmental Kuznets curve. Oxford Economic Papers 53: 721-746.

Bergh, M. O., and Butterworth, D. S. 1987. Towards rational harvesting of the South African anchovy considering survey imprecise and recruitment variability. South African Journal of Marine Science 5: 937–951.

Boyce, M. S., Baxter, P. W. J., and Possingham, H. P. 2012. Managing moose by the seat of your pants. Theoretical Population Biology 82: 340-347.

Clark, W. G. 1991. Groundfish exploitation rates based on life history parameters. Canadian Journal of Fisheries of Aquatic Sciences 48: 734-750.

Clark, W. G. 2002. $F_{35\%}$ revisited ten years later. North American Journal of Fisheries Management 22: 251–257.

Cooke, J.G. 1999. Improvement of fishery-management advice through simulation testing of harvest algorithms. ICES journal of Marine Science 56: 797-810.

DAFF 2013. Report on the review of the Commonwealth Fisheries Harvest Strategy Policy and Guidelines, Department of Agriculture, Fisheries and Forestry, Canberra, Australia.

Deroba, J. J., and Bence, J. J. 2008. A review of harvest policies: understanding relative performance of control rules. Fisheries Research 94: 210–223.

Dichmont, C. M., Deng, A., Punt, A. E., Ellis, N., Venables, W. N., Kompas, T., Ye, Y., Zhou, S., and Bishop, J. 2008. Beyond biological performance measures in Management

Strategy Evaluation: Bringing in economics and the effects of trawling on the benthos. Fisheries Research 94: 238–250.

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 costs, risk and catch? Marine Policy 40: 41-52.

Fay, G., Punt, A.E., and Smith A.D.M. 2011. Impacts of spatial uncertainty on performance of age structure-based harvest strategies for blue eye trevalla (*Hyperoglyphe antarctica*). Fisheries Research 110: 391-407.

Fay, G., Little, L. R., Tuck, G. N., Haddon, M., and Klaer, N. L. 2012. Maintaining risk equivalency among fishery harvest control rules in Southeast Australia. CSIRO Marine and Atmospheric Research, Hobart. 31 pp.

Field, S. A., Tyre, A., Jonzén, N., Rhodes, J. R., and Possingham, H. P. 2004. Minimizing the cost of environmental management decisions by optimizing statistical thresholds. Ecology Letters 7: 669-675.

Francis, R. I. C. C., and Shotton, R. 1997. “Risk” in fisheries management: a review. Canadian Journal of Fisheries and Aquatic Sciences 54: 1699-1715.

Fulton, E. A., Smith, A. D. M., Smith, D. C., and van Putten, I .E. 2011. Human behavior: the key source of uncertainty in fisheries management. Fish and Fisheries 12, 2–17.

Fulton, E. A. Smith A. D. M., Smith D. C., and Johnson P. 2014. An Integrated Approach Is Needed for Ecosystem Based Fisheries Management: Insights from Ecosystem-level Management Strategy Evaluation. PLoS One:e84242. DOI: 10.1371/journal.pone.0084242

Grafton, R. Q., Kompas, T., McLoughlin, R., and Rayns, N. 2007. Benchmarking for fisheries governance. Marine Policy 31: 470-479.

Gulbrandsen, L. H. 2009. The emergence and effectiveness of the Marine Stewardship Council. Marine Policy 33: 654-660.

Hilborn, R. 2010. Pretty Good Yield and exploited fishes. Marine Policy 34: 193-196.

Knuckey, I., Harvey, E., and Koopman, M. 2009. Industry survey to obtain a relative abundance index for spawning eastern gemfish - traditional and innovative methods. AFMA Project R2006/830. Fishwell Consulting 103pp. http://www.fishwell.com.au/projects/project,id,169.aspx

Legg, C. J., and Nagy, L. 2006. Why most conservation monitoring is, but need not be, a waste of time. Journal of Environmental Management 78: 194-199.

Little L. R., Parslow, J., Fay, G., Grafton, R. Q., Smith A. D. M., Punt, A. E., and Tuck G.N. 2014. Environmental derivatives, risk analysis, and conservation management. Conservation Letters 7: 196-207.

Mapstone, B. D. 1995. Scalable decision rules for environmental impact studies. Ecological Applications 5: 401-410.

McDonald, A. D., Smith, A. D. M., Punt, A. E., Tuck, G. N., and Davidson, A. J. 1997. Empirical evaluation of expected returns from research on stock structure for determination of total allowable catch. Natural Resource Modelling 10: 3-29.

McDonald-Madden, E., Baxter, P. W. J., Fuller, R. A., Martin, T. G., Game, E. T., Montambault, J., and Possingham, H. P. 2010. Monitoring does not always count. Trends in Ecology and Evolution 25: 547-550.

Ministry of Fisheries 2008. Harvest strategy standard for New Zealand fisheries. New Zealand Government. http://fs.fish.govt.nz/Page.aspx?pk=104

National Marine Fisheries Service and U.S. Fish and Wildlife Service 2005. Recovery Plan for the Gulf of Maine Distinct Population Segment of Atlantic Salmon (*Salmo salar*). National Marine Fisheries Service, Silver Spring, MD.

Pacific Fishery Management Council (PFMC). 2014. Status of the Pacific Coast Groundfish Fishery: Stock Assessment and Fishery Evaluation. Pacific Fishery Management Council.

Punt, A. E., Smith, A. D. M., Smith, D. C., Tuck, G. N., and Klaer, N .L., 2014. Selecting relative abundance proxies for B_{MSY} and B_{MEY}. ICES Journal of Marine Science 71: 469-483.

Punt, A. E., Smith, D. C., Haddon, M., Russell, S., Tuck, G. N., and Ryan, T. 2015. Estimating the dynamics of spawning aggregations using biological and fisheries data. Marine and Freshwater Research 66: 1–15.

Ralston, S., Punt, A. E., Hamel, O. S., DeVore, J. D., and Conser, R. J. 2011. A meta-analytic approach to quantifying scientific uncertainty in stock assessments. Fishery Bulletin 109: 217-231.

Rayns, N. 2007. The Australian government's harvest strategy policy. ICES Journal of Marine Science 64: 596-598.

Sainsbury, K. 2005. Cost-effective management of uncertainty in fisheries. Presented at ABARE Outlook 2005 Conference, March 2005. http://data.daff.gov.au/data/warehouse/pe_abarebrs99001173/PC13024.pdf

Sethi, S.A. 2010. Risk management for fisheries. Fish and Fisheries 11: 341-365.

Sethi, S.A., Dalton, M., and Hilborn, R. 2012. Quantitative risk measures applied to Alaskan commercial fisheries. Canadian Journal of Fisheries and Aquatic Sciences 69, 1-12.

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., and Little, L. 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.A., Sainsbury, K.J., and Sloan, S. R. 2014. Implementing harvest strategies in Australia: 5 years on. ICES Journal of Marine Science 71: 195-203.

Woodhams, J., Stobutzki, I., Vieira, S., Curtotti, R. and Begg, G.A. 2011. Fishery Status reports 2010: status of fish stocks and fisheries managed by the Australian Government, Australian Bureau of Agriculture and Resource Economics and Sciences, Canberra, Australia.

Walters, C., and Maguire, J.-J. 1996. Lessons for stock assessment from the northern cod collapse. Reviews in Fish Biology and Fisheries 6: 125-137

Appendix D. From data rich to data-limited harvest strategies – does more data mean better management?

Catherine M. Dichmont[1,2], Elizabeth A. Fulton[3], Rebecca Gorton[3], Miriana Sporcic[3], Lorne R. Little[3], André E. Punt[3,4], Natalie Dowling[3], Malcolm Haddon[3], Neil Klaer[3], David C. Smith[3]

[1]CSIRO Oceans and Atmosphere, Ecosciences Precinct, Dutton Park, QLD, 4750, Australia

[2]Cathy Dichmont Consulting, 47 Pioneer Road, Sheldon, Redlands, QLD, 4157, Australia

[3]CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart, TAS, 7001, Australia

[4]School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, WA 98195, USA

Corresponding authors details: Dichmont, Cathy; 41 Boggo Road, Dutton Park, 4750, Queensland, Australia; Phone: +61 7 3833 5925; Cathy.dichmont@csiro.au

Published in Dichmont, C.M., Fulton, E., Gorton, R., Sporcic, M., Dowling, N., Little, R.L., Punt, A.E., Haddon, M., Smith, D.S. 2016. From data rich to data-limited harvest strategies – does more data mean better management? ICES J. Mar. Sci. ICES Journal of Marine Science, doi:10.1093/icesjms/fsw199

Abstract

Harvest strategies have been applied to many data-rich fisheries, and are now increasingly being applied in data-limited situations. These have been evaluated using simulation frameworks, including Management Strategy Evaluation (MSE), but few studies have considered the full spectrum from data-rich to data-limited strategies, in the context of the risk-cost-catch trade-off. This involves evaluating whether the cost of implementing a harvest strategy, the risk to the resource and catch taken from the resource have been appropriately balanced, given the value of the resource. Harvest strategies implemented for Australian Commonwealth fisheries were placed in eight tiers, ranging from data-rich to data-limited, and their performance evaluated using a MSE based on a full end-to-end ecosystem model. Generally, the risk to the resource increased as fewer data were available, due to biases in the assessments and slow response times to unexpected declines in resource status. The most data-rich tiers maximize discounted catches and profits over a 45-year projection period. However, the opportunity costs response is variable, and shows that the benefit of short-term high catches have to be compensated by resource recovery in the long term. On average, more data leads to improved management results in terms of risk of being overfished and not reaching a target, but this requires lower initial catches to recover the resources and lower short-term discounted profits.

Keywords

Risk-cost-catch, harvest strategies, data limited, data rich, risk equivalency, Atlantis

Introduction

Management of renewable resources such as fisheries can be complex given the range of species and habitats affected by fishing. This is true even if the focus is only on species that are directly targeted by fisheries. There is often a range of information sources and data quality available by fishery and species, with fisheries ranging from small scale, low value to large scale, high value (Sainsbury 2005; Dowling *et al*., 2013). Thus, jurisdictions have had to tailor methods to manage each fishery according to their policies and legislative frameworks (e.g., Smith *et al*., 2007). Most policy frameworks include target and limit reference points (TRPs and LRPs), with the latter defined in terms of overfishing (a fishing mortality risk) and being overfished (a biomass risk). In addition to the risks management faces, are the costs incurred from managing the risk, as well as the corresponding benefits derived from the fishery in the form of catch: the risk-cost-catch trade-off (Sainsbury, 2005).

Implementing policies and frameworks, however, has resulted in different operational approaches to stock management, which includes data collection, assessment and decision procedures, often collectively called a harvest strategy (HS). As a result, there is usually, either directly or indirectly, some requirement that risk should remain similar regardless of data availability and the method used to manage a stock; i.e. some form of risk equivalency is required (Sainsbury, 2005).

Several jurisdictions have tried to address risk equivalency, most notably the European Union (EU), USA, and Australia (Dichmont *et al*., 2015). The EU's Common Fisheries Policy (CFP) is the principal legal mechanism for managing fish stocks in EU waters, and regulates all aspects of fishing within the EU. The CFP has the overall objective of ensuring economically, environmentally and socially sustainable use of fishing and aquaculture resources (European Commission, 2013). In response to the range of data types and assessment methods available, the International Council for the Exploration of the Sea (ICES) has implemented a framework for data-limited stocks (ICES, 2012), which consists of six harvest strategy categories from data-rich (Category 1) to data-limited (Categories 2 to 6).

In the USA federal system, the risk-cost-catch trade-off and its implementation into a tier system for HSs is well defined for managers to achieve the goals of the Magnuson-Stevens Fishery Management and Conservation Act, per the National Standards (USA Doc, 2007). These standards recognise that there is a trade-off between conservation and utilization, but, as written, conservation takes precedence over minimizing impacts on fishing communities. Each region within the US has a slightly different approach to addressing risk equivalency. For example, the Pacific Fisheries Management Council Scientific and Statistical Committee (PFMC SSC) places each stock into one of three 'categories' (and one of 11 sub-categories) depending on the method and reliability of the assessment. Management of Alaska's 10 crab stocks, on the other hand, which are jointly managed under federal and State jurisdiction, uses a tier system that includes five tiers depending on data availability and the ability to estimate key stock assessment parameters.

Australian Commonwealth (federal) fisheries are managed by the Australian Fisheries Management Authority (AFMA) under the Fisheries Management Act 1991, and in accordance with the Environment Protection and Biodiversity Conservation Act (1999). The Australian Fisheries Harvest Strategy Policy and Guidelines (DAFF, 2007) relates to key commercial species targeted by AFMA-managed fisheries. Eight classes of HSs are applied to most of the Australian Commonwealth-managed fisheries (Dowling *et al*., 2013). The original system of four categories (called "tiers" in Australia), on which this 8-tier system was based, was implemented in the Southern and Eastern Scalefish and Shark Fishery (SESSF) for several years (Anon, 2014). The Target Reference Point (TRP) for Australian Commonwealth-managed fisheries is B_{MEY}, the biomass corresponding to Maximum Economic Yield. The Australian Harvest Strategy Policy (Rayns, 2007) allows for the use of proxies for B_{MEY} (specifically, 1.2 × B_{MSY}, where the proxy for B_{MSY} is taken to be $0.4B_0$ and thus the proxy for B_{MEY} is $0.48B_0$).

Management strategy evaluation (MSE) is a simulation approach to explore the effects of alternative management options, including the potential trade-offs among the (pre-agreed and pre-specified) management objectives, taking into account various sources of uncertainty (e.g. uncertainty in the assessment, implementation error). This facilitates the identification of HSs that are robust to uncertainty, and that achieve desired trade-offs among the management objectives. MSE has been applied to several single and multispecies fisheries (Punt, 1992; De la Mare, 1996; Butterworth *et al*., 1997; Punt and Smith, 1999; Smith *et al*., 1999; Punt *et al*., 2002; Dichmont *et al*., 2006) and to ecosystems (Sainsbury *et al*., 2000; Fulton *et al*., 2014). Many of the HSs applied to SESSF stocks have been evaluated individually using MSE to ensure they conform to the limit reference point as defined in the Commonwealth Harvest Policy, and to compare the relative robustness of alternative tiers of assessment (Fay *et al*., 2009, 2011; Wayte and Klaer, 2010; Little *et al*., 2011; Klaer et al., 2012). However, the 8-tier framework developed by Dowling *et al.* (2013) has yet to be evaluated.

This paper therefore uses MSE to evaluate six of the tiers from the 8-tier system of Dowling *et al.* (2013) in terms of the risk-cost-catch trade-off, in the context of the species groups in the SESSF (reasons for the omission of two of the tiers is provided below). The tier system was evaluated across a range of species types using the ecosystem model Atlantis (Fulton *et al*., 2014). We aimed to determine how well a tier system of HSs conforms to the assumption of risk equivalency and to explore the overall risk-cost-catch trade-offs.

Methods

Overview

The MSE consists of an operating model and a set of candidate HSs (Punt *et al*., 2016). The operating model ("Atlantis operating model" below) represents "reality" and includes several virtual resources. It therefore describes the biology and environment of the system that is being managed; and the reaction of the fishery to management actions through fleet movements and, in some cases such as this study, catch quota trading. The HSs ("Harvest Strategies") determine the management response through pre-defined decision rules based on the data provided from the operating model, an assessment method, and a decision rule. Importantly, HSs are "unaware" of what the operating model contains, except that which is provided by the modelled monitoring module that samples the operating model ("Data generation"). The success or otherwise of a HS depends on its three components: data collection scheme, assessment method and decision (or harvest control) rules to set, in the case of this study, the Total Allowable Catch (TAC). The end-to-end (or "whole of system") ecosystem model, Atlantis (Fulton *et al*., 2011; henceforth Atlantis-RCC [for "Atlantis-Risk-Cost-Catch"]) forms the basis for the operating model. Atlantis-RCC is based on Atlantis-SE ("Atlantis-South-East") (Fulton *et al*., 2007, Savina *et al*., 2008, Johnson *et al*., 2011, Fulton *et al*., 2014), which was originally developed to explore alternative management options for the SESSF.

A summary of the structure of the base Atlantis-RCC model is given below. This is followed by a description of the steps used in the evaluation of each management strategy.

Atlantis operating model

Atlantis-RCC is a 3-D model, with 71 model regions ("boxes") based on the physical and ecological properties of southeast Australia, the distribution of the water bodies, and the geomorphology of the area (summarised in IMCRA, 1998; Butler *et al*., 2001; Lyne and Hayes, 2005 and Fulton *et al*., 2007) (Figure D.1). Each of these boxes has a single sediment layer and up to five water column layers (dictated by total depth, with shallower boxes having fewer layers).

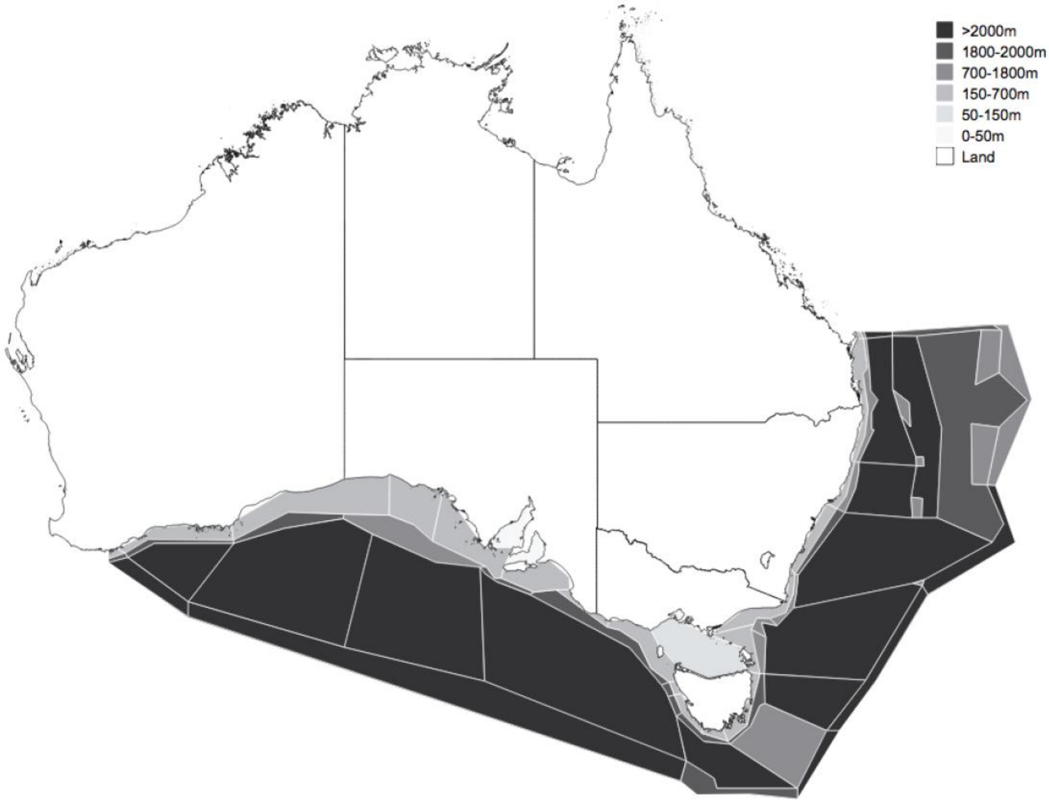


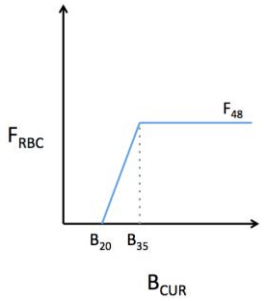
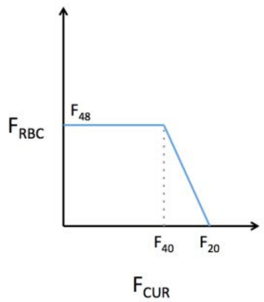
Figure D.1 Map of the model domain of Atlantis-RCC (this matches the domain of Atlantis-SE; Fulton *et al.* 2007, 2014).

The physical environment for Atlantis-RCC includes ocean currents and water column properties such as temperature and salinity. Vertical and horizontal exchanges between spatial boxes, as well as temperature and salinity, were taken from the data-assimilated version of the global ocean model OFAM (Oke et al., 2005).[2]

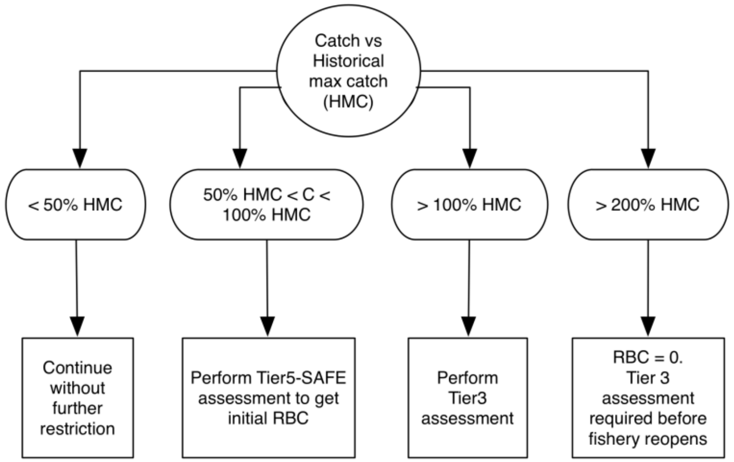
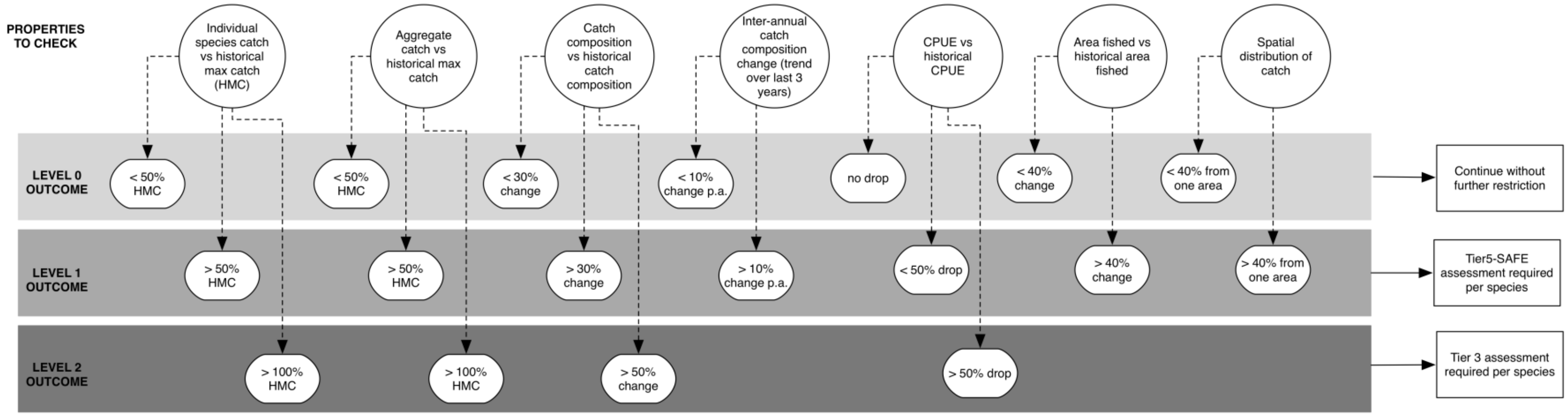
Atlantis-RCC includes the food web as described in Fulton *et al.* (2007, 2014) (Table D.S1). It was initialised to commence in 1980 and run with the fishery catch and known environmental drivers until 2005. Pattern-oriented modelling (Fulton *et al.*, 2007, Kramer-Schadt *et al.*, 2007) was used to calibrate Atlantis-RCC to historical biological and catch data for each group and all spatial areas. The calibration data were constructed from available observational data (Fulton *et al.*, 2007, 2014) and reported fisheries statistics. Projections in which the harvest strategies were applied began in 2005 and ran to 2050, using the OFAM reanalysis of environmental conditions until 2014. When the historical reanalysis was exhausted, the pattern of variance in the environmental conditions were looped (from the beginning of the time series) to complete the projection period (trends in the conditions were maintained in-line with that found in long-term climate projects, see Fulton and Gorton (2014) for further details).

[2] The database used is available at http://www.bom.gov.au/bluelink/ and SPINUP6 from http://www.marine.csiro.au/ofam1/.

Table D. 1 Tiered assessment types and harvest strategies (HSs) for setting recommended biological catches (RBCs). Note currently no tier 2 exists for the SESSF so none was implemented here.

Tier	HS Graph	Rule
1		A full quantitative assessment provides estimates of spawning biomass (*B*) and depletion, which are used in a B_{20}:B_{35}:B_{48} (0.2B_0:0.35B_0:0.48B_0) "broken stick" HS to find the target fishing mortality (F_{TARG}). This F_{TARG} is then applied to the available biomass to calculate the RBC. For the purposes of the paper, the assessment was based on Stock Synthesis (Methot and Wetzel, 2013).
3		Catch curves are used to estimate F_{CUR}, and F_{20}, F_{40} and F_{48} are taken from the relationship between yield and fishing mortality. F_{RBC} is then determined from the HS, and the RBC is calculated using the equation: $$RBC = \max\left(\frac{1-e^{-F_{\mathrm{RBC}}}}{1-e^{-F_{\mathrm{CUR}}}}, 3\right) C_{\mathrm{CUR}}$$ where C_{CUR} is the current catch. For the purposes of this paper, the approaches outlined in Wayte and Klaer (2010) are used to estimate F_{CUR} and the fishing mortality reference points.
4		The RBC from the Tier 4 HS is given by: $$RBC = C_{\mathrm{T}} \max\left(\frac{\overline{\mathrm{CPUE}}-\mathrm{CPUE_L}}{\mathrm{CPUE_T}-\mathrm{CPUE_L}}, 0\right)$$ where C_T is a catch target, $\mathrm{CPUE_L}$ is the limit CPUE, $\overline{\mathrm{CPUE}}$ is the average CPUE over the most recent four years, and $\mathrm{CPUE_T}$ is the target CPUE. The default catch and CPUE targets were the average for the simulated years 1996-2005 (Little *et al*., 2011). For some species (flathead, blue grenadier, blue warehou, redfish, pink ling and other shelf demersal fish) the reference period was set to the more conservative 1986-1996 period (this is in line with how the reference period can be tuned per species in reality, see the main text).

Tier	HS Graph	Rule
5		Tier 5 uses the average length of fish in the catch to determine F_{CUR} based on average length expected as a function of fishing mortality from a yield-per-recruit calculation (Haddon *et al*., 2015). This F_{CUR} is then used in Tier 3 harvest control rule.
5 SAFE		Tier 5SAFE (Tier 5S) uses method for calculating fishing mortality for species *i* (F_i) as outlined in Zhou *et al* (2011): $$F_i = \frac{q_i^h \cdot q_i^l \cdot (1 - S_i) \sum_t a_{t,i} \cdot E_t}{A_i}$$ where q^h is habitat-dependent encounterability (parameterised using the relative habitat use and overlap defined for the stocks and fleets in the operating model), q^{λ} is size- and behaviour-dependent selectivity (also parameterised from the effort allocation model), *S* is the discard survival rate, a_t is the area covered in time step *t*, E_t is the effort applied in time step *t* and A_i is the area the species occupies. An aggregate annual *F* is provided by summing over all time steps fished in a year. The reference exploitation rates F_{20}, F_{40} and F_{48} are given by $F_{20} = 1.5\omega M$, $F_{40} = \omega M$ and $F_{48} = 0.8 F_{40}$ with ω set to 0.91 for teleosts and 0.43 for chondrichthyans. *N*atural mortality, *M*, was estimated using the Jensen (1996) relationship: $M = 1.65 / t_{\mathrm{am}}$ where t_{am} is the age-at-maturity. F_{i} vs reference *F* is then used to determine the RBC and any further assessment actions.

Tier	HS Graph	Rule
6		The tier 6 HS is based on comparing annual total landed catch (C) against various triggers. HMC is the historical maximum annual total catch (Dowling *et al.*, 2008).
7		

Tier	HS Graph	Rule
		Tier 7 HS uses a variety of triggers to determine changes in the HS. Tier 5-SAFE is the level 1 choice as this method was designed for data-poor species. Tier 3 was chosen as the level 2 choice because the data required for tier 3 are likely to be available (Dowling *et al.*, 2008).

Atlantis-SE (Fulton *et al*., 2007) was used as a starting point in developing Atlantis-RCC. As Atlantis-SE was calibrated focusing on 2005 biomass levels, thus the first step in the redevelopment was to set initial (1980) biomasses. In all cases this involved finding a measure of the level of depletion from 1980 to 2005 and then dividing the 2005 biomass values in the most recent (updated) version of Atlantis-SE (Fulton and Gorton, 2014) to give the 1980s biomass values (e.g. if the relative depletion was 50% then the 1980s biomass estimate was twice the 2005 biomass value). For assessed species, the relative depletion was taken from Morison *et al*. (2012) and for all other (non-assessed) groups, an earlier variant of Atlantis-SE (which had been run through the period 1980-2005) was used to provide the estimate of relative depletion.

One set of biological parameter values (i.e. values for non-predation mortality rates, physiological, consumption and growth rates, habitat preferences, movement rates etc.) per species group was used across the model domain, unless the group was defined as having multiple stocks, in which case fecundity, background mortality and diet connection strength varied among stocks (Table D.S1).

Size-at-age for each vertebrate group in Atlantis-SE varied through time and among locations, but Atlantis-RCC also includes multiple growth "morphs" (c.f. Punt *et al*., 2001; Methot and Wetzel, 2013) for all main SESSF species (Table D.S1). Three morphs were used in the simulations presented here, though a small number of projections with tens morphs were also conducted to check that the results were robust to the number of morphs. Each morph follows a different growth trajectory so that there is variation in size-at-age within a cohort at each location. This is more realistic than how size-structure was represented in earlier Atlantis models (where there was a single growth morph per species), and more in line with previous evaluations of harvest strategies (e.g., Little *et al*., 2007). It also means that the size-composition data generated using the operating model were more realistic.

Recruitment (settlers) was governed by a Beverton-Holt style stock-recruitment curve, except for sharks, seabirds and marine mammals, for which a linear function was used. Tiger flathead *Neoplatycephalus richardsoni*, blue grenadier *Macruronus novaezelandiae* and silver (spotted) warehou *Seriolella punctata* appear to exhibit cyclic dynamics (Day and Klaer, 2014; Tuck, 2014; Day *et al*., 2013). This behaviour was mimicked in Atlantis-RCC by applying deviations to the predictions from the stock-recruitment curve based on the outcomes from the single-species age-structured assessment models for those species. This cyclic variation was in addition to any variation produced by trophodynamic interactions and environmental conditioning of, for example, movement, survival, reproduction, or physiology. Cyclic variation for these groups was extended into the future by assuming that the pattern of deviations about the stock-recruitment relationship estimated for the past applied into the future (this was done by pre-generating the time series of deviations, with magnitude and period matching the historical pattern, and applying those via forcing files in all simulations).

The diet connections between the biological groups identify potential food web pathways. Predation mortality was related to the maximum potential availability of each prey to each potential predator. The realised rate of predation was then conditioned on the level of physical contact (i.e. spatial overlap within a box given habitat preferences and patchiness), habitat (refugia) state, and gape limitation (i.e. size of the mouth versus size of the prey given the feeding mode of the predator). The diet connections of Atlantis-SE were changed to include a smoother gape limitation function (Figure D.S1), this was required to make sure the realised diets matched those from real world observations once the multiple growth morphs had been implemented.

Parameters were set to achieve a stable ecosystem, under constant fishing pressure, within the range of biomass values reported for these groups in the literature. Parameter tuning targeted the most sensitive species-specific parameters (previously identified through sensitivity and factor analyses by Pantus and Dennison (2005) and Fulton *et al*. (2007)) – typically the diet availabilities, growth and consumption rates, background natural mortalities (especially for the highest trophic levels), fecundity levels and the steepness of the stock-recruitment curve. This tuning began with the most uncertain parameters and adjusted these according to the following criteria: 1) the predicted spatial distributions and time series of biomasses, age structure, realised diet composition, and catches, must approximate the shape, magnitude and variability of observed time series across the majority of boxes; 2) observed catches and discards must be sustained without rendering any model group extinct; and 3) rate parameters must not to be adjusted beyond bounds reported in the literature without expert advice from researchers active in the region.

The socio-economic effort allocation model of Atlantis-SE (Fulton *et al*., 2007) was unchanged; as such, allowance was made for multi-species targeting. The effort allocation model was driven by two main components – a quota trading module and a fleet effort space-time dynamic module (Fulton *et al*., 2007; van Putten *et al*., 2013). Together, these determine which suite of species was targeted, where, and at what time of year. As such, a TAC may be under- or over-caught (though the degree of potential over catch is constrained, as it is in reality). All prices and cost structures used in the effort allocation model (or used to calculate the economic performance measures) were sourced as detailed in Fulton *et al*. (2007), except for the management costs which were based on the latest management cost breakdown from the SESSF (Geoff Tuck, CSIRO, *pers. com.*) and from data used in Dowling *et al*. (2013).

Stochasticity was applied to the data generated for use by the assessment model (described below) and to parts of the effort allocation model, but not to the population or food web dynamics within the operating model.

Harvest Strategies

The tiers examined (Table D.1) included those currently used in the SESSF (Smith *et al*., 2014), and updated versions of existing data-poor HSs that have been applied in other Australian federally-managed fisheries (Zhou *et al*., 2011; Dowling *et al*., 2008; Dowling *et al*., 2016). Tiers 1 to 5 are straightforward application of the HS, given the estimates of fishing mortality, biomass, and catch (whichever is required by the tier) to give the recommended biological catches (RBC). Tiers 6 and 7 required following the flow charts to see what action must be taken. In the case of tier 7 the highest RBC across all the checks was applied.

The most data rich tier (tier 0) was not used for the SESSF and so it was omitted from the analyses. Similarly, tier 2, which was based on fitting population models that are more uncertain than tier 1 assessments, was omitted because there are no rules currently for how a stock would be assigned to this tier. Two variants of tier 5 were considered, however, as there were many competing ways of empirically estimating fishing mortality in data-limited situations and both a surplus production-based method proposed for the SESSF (Haddon *et al.,* 2015) and the SAFE method of Zhou *et al.* (2011), which is applied to many Australian bycatch species, were evaluated here.

The tiers provide RBCs, which may be modified through, for example, the use of meta-rules, to determine TACs. The meta-rules considered in this paper were: (a) TACs may not change by more than 50% from one year to the next, and (b) the TAC is unchanged if the proposed change in TAC from one year to the next is 10% or less. One set of simulation experiments was run with the meta-rules active and one set with them disabled (i.e. without meta-rules). Additional tier buffers that increase the gap between the RBC and the TAC based on tier and therefore data availability are also usually applied within the SESSF (Anon, 2014), but the purpose of this paper was to evaluate the value of information so these were ignored.

Data generation

A sampling model was used to generate the following fishery-dependent data, for each stock and Atlantis region: (a) catch length- and age-composition data; (b) catch-per-unit-effort data (by vessel size-class and fishery sector); (c) landings data (and catch species composition) by vessel size-class and fishery sector, and (d) discard data. The data generation approach allowed for ageing error, measurement error, variation in catchability, and error when measuring discards, with error levels that were stock-specific (Table D.S2). As stated above, since few fisheries in Australia have long-term independent survey data, and tier 0 was not being evaluated, no fishery-independent data were generated.

Table D. 2 . Performance metrics. Note that R3, all the cost metrics, relative discounted catch, number of times hierarchical tiers are triggered and response time are calculated across all years of the projection period; whereas the other indices are calculated on shorter periods.

Code	Metric	How calculated	Notes
R1	Probability $B < B_{20}$	$\underset{s}{average}\left(N_{B<B_{20},s}/30\right)$	$N_{B<B_{20},s}$ is the number of years for which spawning biomass is less than *0.2B*$_0$ during the last 30 years of the projection period for simulation *s.*
R2	Probability $B < B_{48}$	$\underset{s}{average}\left(N_{B<B_{48},s}/30\right)$	$N_{B<B_{48},s}$ is the number of years for which spawning biomass is less than *0.48B*$_0$ during the last 30 years of the projection period for simulation *s*.
R3	Time to threshold	$\underset{s}{median}\left(T_{s,bx}\right)$	$T_{s,bx}$ is the median over simulations of the time taken to reach the threshold b_x for the first time during the projection period*.
C4	Opportunity costs	$\underset{s}{average}\left(\sum_{y} p_{y,s}\left(C_{b,y,s}-C_{y,s}\right)e^{-(y-2006)\delta}\right)$	*p* is the price per unit catch in year *y*, δ is the economic discount rate (0.05), $C_{y,s}$ is the catch in year *y* of simulation *s* and $C_{b,y}$ is the catch in the same year under the bang-bang control rule.
C5	Harvesting costs	$\underset{s}{average}\left(\sum_{y}\left(H_{cap,y,s}+H_{fix,y,s}+H_{ul,y,s}C_{y,s}+\left(H_{fu,y,s}+H_{g,y,s}\right)E_{y,s}+H_{fu,y,s}Z_{y,s}\right)e^{-(y-2006)\delta}\right)$	$H_{x,y,s}$ are the various operating cost components at time *y* in simulation *s*. H_{cap} are capital costs; H_{fix} are fixed costs; H_{ul} are unloading costs per unit catch; H_{fu} are fuel costs per unit of effort (or steaming time); H_g are gear-associated costs per unit effort. $C_{y,s}$ and $E_{y,s}$ are the catch and effort during year *y* of simulation *s*, $Z_{y,s}$ is steaming time.
C6	Management costs	$\underset{s}{average}\left(\sum_{y}\sum_{n} M_{n,y,s}e^{-(y-2006)\delta}\right)$	*M* are the various types of management costs, (*n* is the type of cost being assessment, administration, compliance associated with an assessment of tier *x*) for treatment species i.e. for those being assessed.
C7	Short-term discounted profits	$\underset{s}{average}\left(\sum_{y}\left(p_{y,s}C_{y,s}-K_{T,y,s}\right)e^{-(y-2006)\delta}\right)$	$K_{T,y,s}$ is the total costs for the fishery (harvesting, quota-related and management licence costs) in year *y* of simulation *s*; *p* is the price per unit catch

			and C is catch. δ is the economic discount rate (0.05).
H8	Relative Discounted Catch	$\frac{\sum_s \sum_y C_{y,s} e^{-(y-2006)\delta}}{\sum_s \sum_y C_{b,y,s} e^{-(y-2006)\delta}}$	$C_{y,s}$ is the catch in projection year *y* of simulation *s*, $C_{\mathrm{b},y,s}$ is the catch in projection year *y* of simulation *s* under the bang-bang control rule, and δ is the economic discount rate.
H9	Catch variability	$\underset{s}{average}\left(\frac{\mathrm{var}(C_s)}{\bar{C}_s}\right)$	var(C_s) is the variance in annual catch across the projection period in simulation *s*, and the denominator is the mean annual catch across the projection period in simulation *s*.
A10	F/F_{MEY}	$\underset{s}{average}\left(\frac{F_s}{F^o_{MEY,s}}\right)$	*F* is the HS estimate of fishing mortality (or its proxy) in the first year of the projection period for simulation *s,* and F^o_{MEY} is the fishing mortality rate for achieving $0.48B_0$ (as defined by the operating model, *op*) in simulation *s*. This metric is based on first year of the simulation (see main text) and quantifies whether the HS believes there is a need for an increase in harvest rate – because the *F* is less than the target value
A11	Relative TAC bias	$\underset{s}{average}\left(\frac{TAC_{x,s}}{TAC_{1,s}}\right)$	$TAC_{x,s}$ is the TAC set under tier *x* in the first year of the projection in simulation *s*; and $TAC_{1,s}$ is the TAC set under tier 1. TAC is also replaced with actual catch as an alternative performance metric.
A12	Number times SAFE or Tier 3 triggered for tiers 6 and 7	$\underset{s}{average}(L_{x,s})$	$L_{x,s}$ is the total number of trigger events for trigger level *x* over projection period in simulation *s*. This is also calculated for actual catches instead of TAC.
A13	Response time	$\underset{s}{average}\left(\underset{i}{average}(T_{R,i,s})\right)$	Annually, the true state of the operating model was assessed with perfect knowledge and it as recorded if the biomass change was sufficient to require a >10% change in the TAC under tier 1 if

	there was perfect information. For each such event *i* in simulation *s* the response time for the HS $T_{R,i,s}$ was calculated as the time in years from when the event was recorded to when the HS changed the TAC in the correct direction.

* There are four "time to threshold" indices: the time to increase to $0.2B_0$, the time to increase to $0.48B_0$, the time to decrease to $0.2B_0$, and the time to decrease to $0.48B_0$. Although only a subset of these indices will be meaningful for any one species group – for example the time to decrease is meaningless if the biomass is less than $0.2B_0$ at the start of a simulation. However, the four indices are reported so that it is possible to determine if (a) an overfished species recovers to $0.2B_0$ (or beyond), (b) a stock that is not fully exploited has its biomass drop to $0.48B_0$ (or lower), and (c) a stock that is initially in the

Data were generated for each 12-hour Atlantis time-step and aggregated to trip, month, and year. Aggregate annual data were used for assessment purposes, as is common when applying the HSs in reality. Vessel level catch and effort data were aggregated for the tier 1 assessments based on the gear used (or season for blue grenadier – spawning versus non-spawning), so that fleet-specific parameters (e.g. selectivity) could be estimated in the assessment (as is the case in reality). The data from the fleet that took the majority of the catch during the final five years of the historical period (i.e. immediately before the projection period began) were used when applying the tier 3 and 4 harvest control rules. Alternative options were explored (e.g. data pooled across all fleets or using the fleet that caught the most in any five-year period), but the choice of option made little difference to the general results.

Tier 4 can also be sensitive to the selection of the reference period used. While the period 1996 – 2005 is the default (i.e. final 10 years of the historical period), for a number of species in the SESSF this has been modified based on the history of exploitation. Simulations with different reference periods were used to explore the influence of the reference period for those species with modified Tier 4 reference periods (flathead, blue grenadier, blue warehou, redfish, pink ling and other shelf demersal fish). In all cases a 1986-1996 reference period was amongst the most conservative and so that period was used here for the tier 4 cases. An analysis in which the reference period was set to the default was also run ("Tier 4 Untuned").

Three parameterisations of size-at-age per group were used. An additional smoother was applied when calculating age-length keys for the assessments, if the size-frequency distributions were not smooth (Figure D.S2). In these cases, normally distributed noise was added to size-at-age to provide smoother length-frequency distributions that better approximate observations.

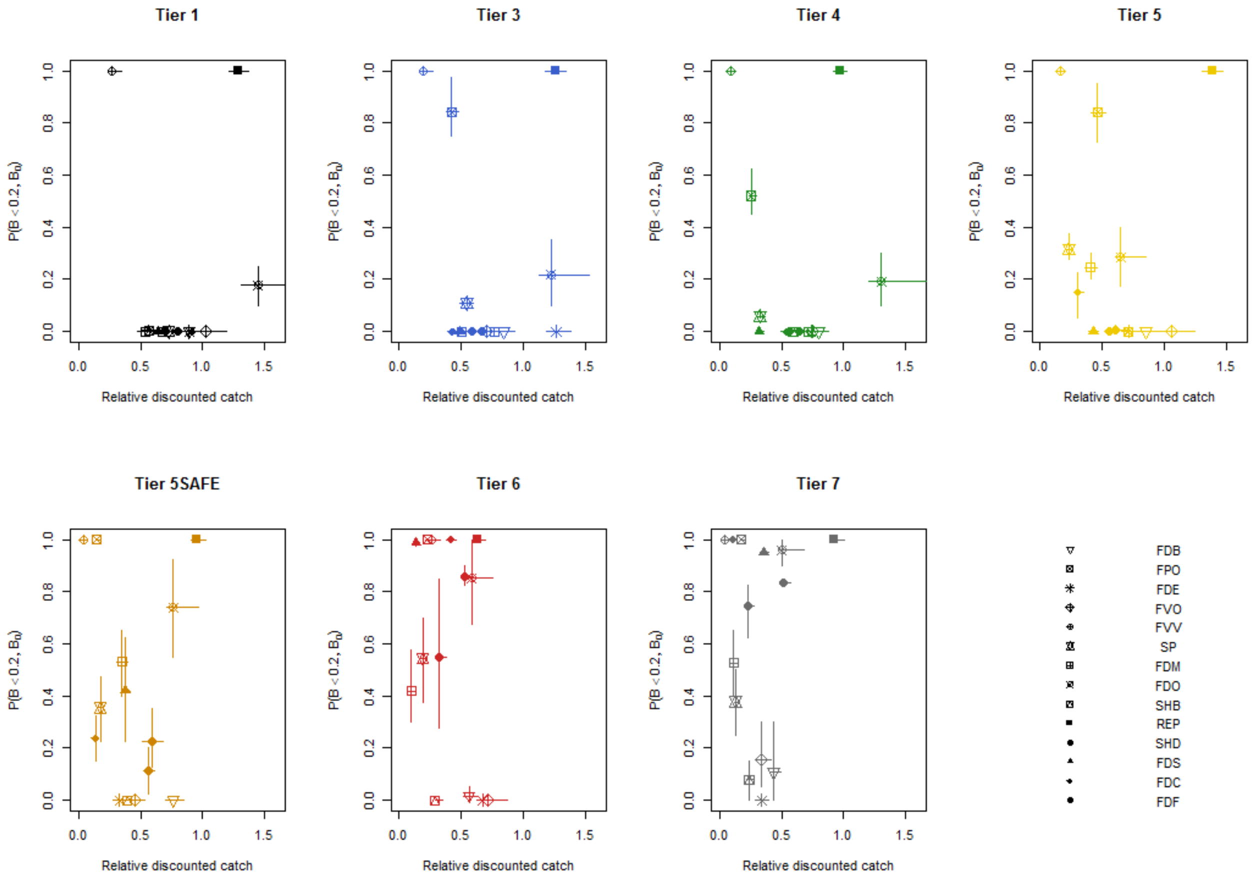


Figure D. 2 Median (across simulations) probability of being below 0.2B_0 against relative discounted catch with inter-quartile range values for each tier. Results without meta-rules. Each point is a species. FDB=Flathead; FPO=Morwong; FDE=Blue grenadier; FVO=Whiting; FVV=Gemfish; SP=Blue warehou; FDM=Redfish; FDO= Cascade orange roughy; SHB=Gummy shark; REP=Gulper shark; SHD=Demersal sharks; FDS=Generic demersal fish; FDC=Pink ling; FDF=Blue-eye trevalla.

Simulation experiments

Atlantis-RCC represented all the major functional groups and species of fisheries or conservation interest in the southeast Australian ecosystem, including those within the SESSF. It was not feasible to explore the risk-cost-catch tradeoff for all these species and groups. Instead, a sub-set of ('treatment') species was selected for consideration in the MSE that was representative of a range of life histories (Table D.S2, which also provides the criteria used to select species).

The steps undertaken in each MSE simulation are given in Figure D.S3. Each simulation was 70 years (1980-2005, representing the historical period, and a 2005-2050 projection period). Twenty replicates were undertaken for each treatment species (i.e. those species assessed using one of the tiered HSs) for each tier. Computational speed precluded a larger set of replicates (20 replicates proved sufficient as the ecological components of Atlantis are deterministic and so the model only includes limited stochasticity; additional projections showed that increased numbers of simulations did not alter the results materially). The random deviates for effort allocation and observation error were the same for all simulations to ensure that the results of the projections were maximally comparable (i.e. the results are analogous to paired tests rather than independent tests and were compared in this manner). However, the small number of replicates means that estimates of intervals will be fairly imprecise and are hence considered primarily in a relative sense, i.e. between alternative HSs rather than as accurate precisions of actual intervals.

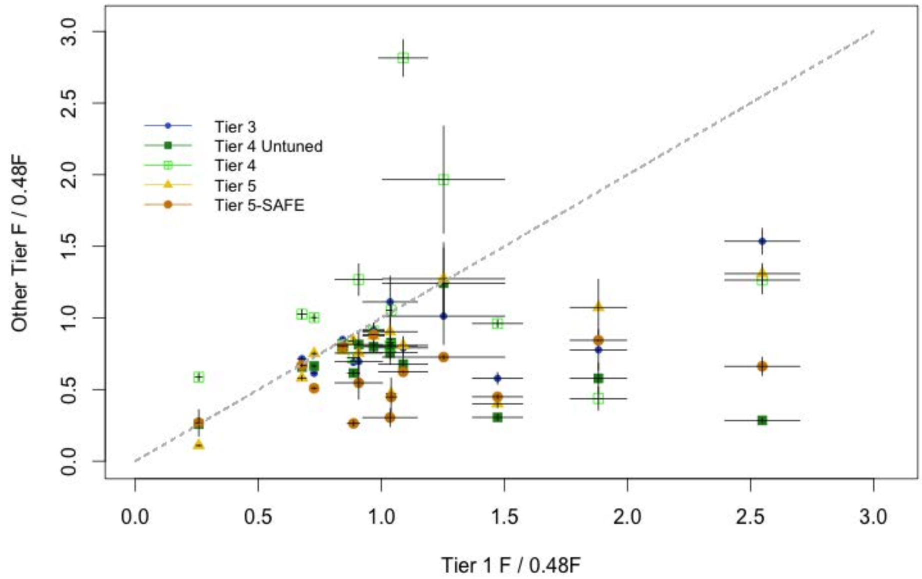


Figure D. 3 Median tier 1 F/F_{MEY} against that for tiers 3 to 7 for each species. F_{MEY} is the fishing mortality rate corresponding to Maximum Economic Yield (proxied here by the fishing mortality rate corresponding to 0.48B_0).

The actual TACs were applied between 1980 and 2005, although with implementation error due to the socioeconomic effort allocation model used in Atlantis-RCC. Once the projection began in 2005, the treatment species had their TACs set using the tier under consideration in that simulation, with the TACs for all other species set to the 2005 level, i.e. only one species was actively managed within a scenario using a single consistent (through time) HS.

It is not straightforward to handle non-treatment species when conducting a realistic test of the management system for a multispecies fishery. In theory, the fishing mortality rates for each species leading to a desired system state could be found and applied to the non-treatment species, but that is a non-trivial exercise for a complex food web and may still see some species fail to achieve the mandated biomass target level. Choosing any other fishing mortality level is as prone to driving over/underfishing of companion species as what was undertaken in the study. More importantly, while it may initially seem that the best test of the tiers would be to make treatment effects the only effect considered, such isolation isn't possible in a multispecies fishery, and the central reason for conducting an MSE was to test the strategies for robustness given the complexities inherent in the management and fishery system. It is important to know how robust the management system is to those conditions given the control rules are to achieve recovery and sustainability within the context of the multispecies fishery. Consequently, the treatment species were selected as they not only represented a range of life history types, but they also represented species with a range of influences on effort dynamics – including key target species (e.g. tiger flathead or blue grenadier), secondary species (e.g. blue warehou) and some bycatch species (e.g. gulper shark). Thus, we feel across the entire set of species tested the HS were tested in a realistic way for the type of multispecies fishery they will be used for in reality.

In total the HS simulations equated to 14 species X eight tiers X two meta-rules (with and without) = 224 scenarios. Two additional analyses were undertaken to create reference points or trajectories to facilitate interpretation of the output of the HS simulations:No fishing on any group (to allow for the calculation of unfished biomass, B_0, by group).

Reference HS: Catches were selected using a "bang-bang" HS where all biomass over the target level is removed (Deroba and Bence, 2008) and there is perfect information about stock size. This HS eliminated targeted fishing for a species for N_1 years if $B < 0.48B_0$, and allowed for large catches for N_2 years if $B > 0.48B_0$. N_1 and N_2 were selected for each species iteratively: an analytical determination was not possible because the use of the dynamic effort allocation model led to implementation error. Elimination of targeted fishing was used because a complete reduction of fishing-related mortality on a species was not possible, owing to incidental catches. In addition, the effort allocation sub-model in Atlantis allows for non-compliance and a memory of past catch performance. Consequently, some catch began once stock rebuilding started and biomass exceeded approximately 0.25-0.3B_0. It follows that recovery for the reference HS is consequently (slightly) slowed compared to the true optimal catch. The discounted catches from the HS provided a reference given each species started from a different biomass relative to 0.4B_0.

Performance metrics

The analyses involved assessing and managing stocks separately (see Tables D.1 and D.S1). However, the results by stock were combined into results by species for the purposes of summarizing performance. The projections assumed that assessments and hence changes to RBCs and TACs occurred annually. Twelve performance metrics were considered, classed into four broad categories: risk, economic, catch, and stock assessment model performance. Details of how each metric is calculated are given in Table D.2 (which also provides a code for each performance measure to aid in reporting). Unless stated otherwise, the performance metrics were averages over simulations for each species and medians (with interquartile ranges) over species.

The three risk performance metrics are: (R1) the probability of being below the LRP (20% of the unfished spawning biomass, 0.2B_0) during the last 30 years of the projection period, (R2) the probability of being below the TRP (0.48B_0) during the last 30 years of the projection period, and the time taken to reach a reference point threshold – (R3a) either to recover to 0.2B_0, if overfished (i.e. starting the projection period with $B < 0.2B_0$), or (R3b) to drop to 0.48B_0 if not fully exploited initially (i.e. starting the projection period with $B > 0.48B_0$). The choice of the last 30 years for the R1 and R2 performance metrics was selected to provide an opportunity for the stocks depleted at the end of the historical period to recover before comparative risk performance was assessed.

The four economic performance metrics are: (C4) opportunity costs, (C5) harvesting costs related to catching and handling the product (based on the fuel and market price models of Fulton *et al.* (2007)) and (C6) the cost of running the tier in terms of, for example, data collection and the assessment (using data from Dowling *et al*. (2013) and Tuck pers. commn). These metrics and estimated revenue were used to calculate the profitability performance metric (C5) and for the whole 45-year period as discounted values (Fulton *et al.*, 2007). The discount rate was assumed to be 0.05, because this is similar to the discount rate often used in bio-economic models in Australia (e.g., Punt *et al*., 2011).

The discounted catch over the projection period relative to that expected under the Reference HS was also calculated as a performance metric (H8). Our calculation for discounted catch does not include the oft-used final term, set to the discounted sum of the catch in the last year for an infinite number of years. This is because the populations do not all reach equilibrium by the end of the projection period. The final harvest metric was catch variability through time (which can be important to industry economically and to markets who desire reliable and stable supplies).

Four assessment performance metrics were also computed. The first (A10) is the median (with inter-quartile ranges) estimated fishing mortality F_{year}, as determined by the tier's assessment (for tiers 1 to 5-SAFE), relative to the F_{MEY} (calculated from the operating model). The fishing mortality proxy for tier 4 was inferred by assuming the CPUE relative to the target CPUE is a proxy for F/F_{MEY}. This performance metric evaluates the direction in which the HS is likely to change fishing mortality; upwards if $F/F_{\mathrm{MEY}} < 1$ and vice versa. The second assessment performance metric (A11) is the TACs set under the higher (more data-poor) tiers relative to that set by tier 1. Since it is not possible to compare the TACs between tiers over the full time period, as the biomass in each future depends on the past TACs, this metric was calculated only using the TAC in the first projection year, as this will be based on assessments of the same relative population biomass (and data) irrespective of the tier. An alternative metric was to use the actual catch rather than the TAC. The third assessment performance metric is the frequency with which a more data-rich assessment is triggered in the two hierarchical HS (tiers 6 and 7) (A12). This metric is a count of the number of times the hierarchy's level 1 (SAFE) and level 2 (tier 3) triggers are activated. The final assessment performance metric (A13) is the response time, which is the number of years before the HS reacts to a change in biomass in the operating model where $TAC_y^{true} < 0.9TAC_{y-1}^{true}$. The true TAC is based on a tier 1 assessment with perfect information.

Results and Discussion

Risk

There is a marked increase in R1, the median (across species) probability of falling below 0.2B0, when RBCs are based on tiers 5, 5SAFE, 6 and 7, irrespective of whether meta-rules are applied. The median (over species) risk of falling below the 0.2B0 for tiers 1, 3 and 4 is zero (Tables D.3a,b and Online Figure D.S4a,b). Apart from tier 1, most of the tiers do not recover the species to the TRP (0.48B0) by the last 30 years of the projection period, and thus the risk of being below 0.48B0 (R2) is close to one. This is again regardless of whether meta-rules are applied. The R2 performance metric should be about 0.5 if a species fluctuates around 0.48B0 during the last 30 years of the projection period. The tier 1 HS is closest to this value: the median (across species) probability of being below 0.48B0 is about 0.65 and 0.51 (without and with meta-rules respectively). That stated, the results are highly species-specific. For example, for R2 the upper (across species) quartile is approximately 1 in all cases, as is the lower quartile for tiers 5-7, but it is about 0.5 in many cases for tiers 1-4. This indicates the breadth of the interspecies variability, but also shows that the more data limited tiers struggled to achieve species recovery to the TRP (and in some cases even the LRP) by the last 30 years of the projection period.

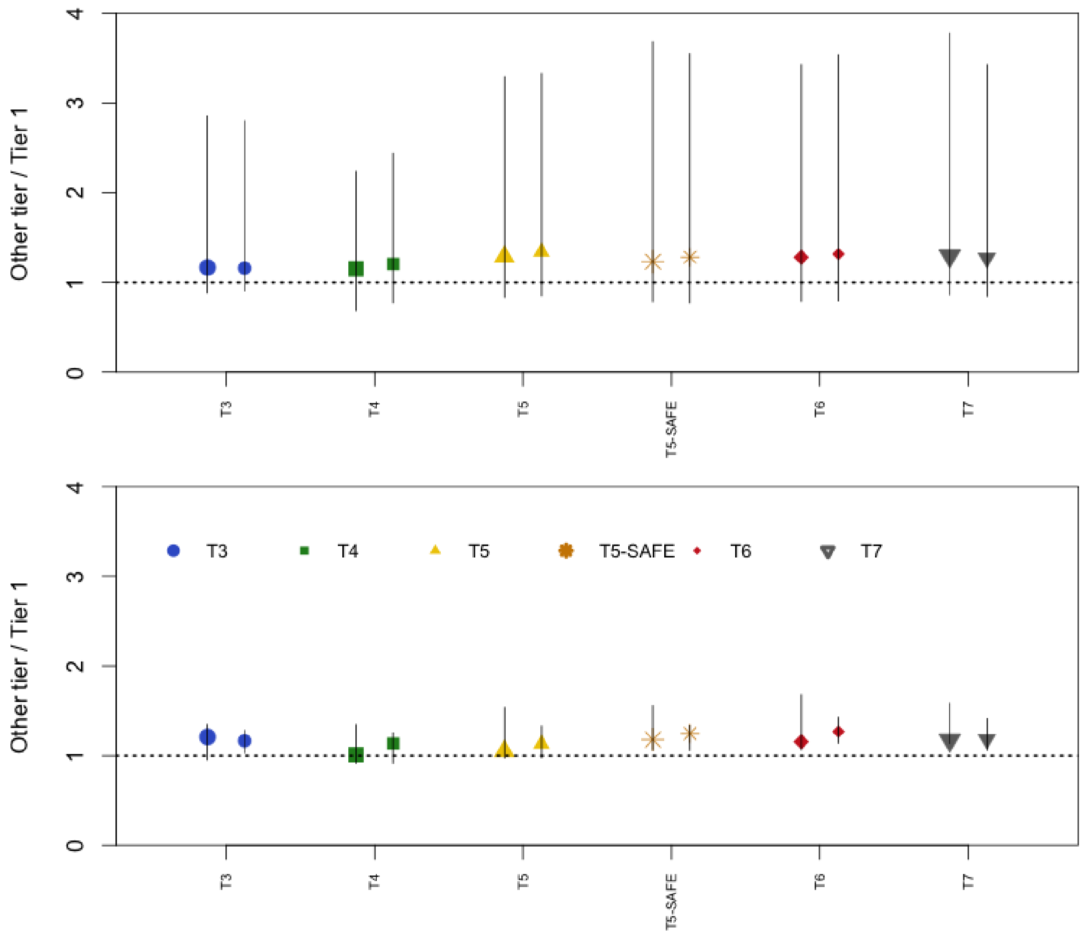


Figure D 4 Median ratio (with inter-quartile range) over species of the TAC set in the first year of the projection period under each HS relative to that under the tier 1 HS (larger points) and the ratio of the resulting catches - i.e. the catches taken under those TACs (smaller points). Results are without (top) and with (bottom) meta-rules.

Table D. 3 **Median over species (and inter quartile values across species (lower;upper)) of the risk, catch and economic performance metrics for all tiers (a) without meta-rules in use, (b) with meta-rules. Values in square brackets for management costs are the median costs when the more expensive higher level assessments in the hierarchy are triggered. The risk indices are calculated for the final 30 years of the projection period (to allow for recovery in those stocks that initially start in a depleted state), while all other indicators are calculated for the entire projection period. An entry of "NA" indicates that the threshold was not achieved during the projection. The costs are discounted at the same rate as the relative discounted catch; and as the management costs are a fixed price per tier there is no variation across species, except for tier 6 and 7 – where the value in brackets for the management costs for tiers 6 and 7 indicate the true total cost given the frequency with which the more intensive assessments were triggered from within these tiers). All $ values are AUD.**

(a)

Tier	Risk $B < 0.2B_0$ (R1)	Risk $B < 0.48B_0$ (R2)	Time Fish Down to *$0.48B_0$* (R3a)	Time Recover to *$0.2B_0$* (R3b)	Opportunity Cost ($ million) (C4)	Harvesting Costs ($ million) (C5)	Manage. costs ($ million) (C6)	Profits ($ million) (C7)	Relative discounted catch (H8)	Catch Variability (H9)
1	0.0 (0.0;0.04)	0.65 (0.53;0.97)	4 (2.25;6.5)	10 (3;NA)	170.1 (-368;1508.5)	1870.1 (55.6;11850)	4.9	275.1 (260.3;290.9)	0.83 (0.66;0.92)	158.45 (72.78;205.31)
3	0.0 (0.0;0.19)	0.94 (0.79;1.0)	4.25 (1.63;5.0)	31 (10.5;NA)	157.8 (-679.9;1894.1)	1719.7 (57.8;11840.2)	4.4	342.5 (301.5;452.2)	0.62 (0.5;0.83)	104.71 (74.9;168.3)
4	0.0 (0.0;0.16)	0.83 (0.66;1.0)	3 (2.25;4.5)	20 (11;NA)	296.7 (-628.8;4001.6)	1801.6 (49.3;7910.6)	3.6	162.1 (151.5;167.3)	0.61 (0.45;0.75)	78.4 (41.02;131.22)
5	0.08 (0.0;0.31)	1.0 (0.98;1.0)	4.25 (2.13;6.75)	23 (17;NA)	179.4 (-932.6.1;2412.9)	1758.6 (61.4;10121.7)	2.1	-45.4 (-50.9;56.6)	0.58 (0.41;0.71)	119.39 (70.78;249.25)
5SAFE	0.30 (0.03;0.69)	1.0 (0.99;1.0)	3.5 (0.5;5.75)	NA (28;NA)	324.5 (-802.7;2965.7)	1700.7 (57.5;9864.1)	0.8	-276.5 (-302.5;-47.85)	0.39 (0.21;0.59)	286.04 (97.42;637.95)
6	0.70 (0.12;1.0)	1.0 (0.97;1.0)	3.25 (2.13;4.75)	NA (32;NA)	189.9 (-461.1;3202.1)	1520.6 (51.6;8802.2)	0.51 [1.24]	-568.7 (-607.1;-273.5)	0.37 (0.24;0.58)	305.1 (166.62;673.62)
7	0.945 (0.21;1.0)	1.0 (1.0;1.0)	3.5 (1.25;5.75)	NA (43;NA)	334.7 (-445.9;3184.6)	1534.1 (60.8;10022.9)	0.46 [2.65]	-486.5 (-525.7;-188.6)	0.29 (0.13;0.41)	623.1 (216;1206.72)

(b)

Tier	Risk $B < 0.2B_0$ (R1)	Risk $B < 0.48B_0$ (R2)	Time Fish Down to *$0.48B_0$* (R3a)	Time Recover to *$0.2B_0$* (R3b)	Opportunity Cost ($ million) (C4)	Harvesting Costs ($ million) (C5)	Manage. costs ($ million) (C6)	Profits ($ million) (C7)	Relative discounted catch (H8)	Catch Variability (H9)
1	0.0 (0.0;0.12)	0.51 (0.44;0.99)	13.25 (7.75;24)	12.5 (6.5;NA)	210.7 (-358.8;1746.1)	1802.1 (64.4;14883.2)	4.9	57.7 (50.9;61.7)	0.81 (0.66;0.94)	29.68 (15.91;74.44)
3	0.0 (0.0;0.18)	1.0 (0.77;1.0)	7.75 (4.75;10.7)	13 (9;NA)	161.4 (-772.8;2025.8)	1675.1 (71.2;14949.5)	4.4	-195.1 (-210. 7;-74.8)	0.79 (0.42;0.91)	37.34 (12.87;83.63)
4	0.0 (0.0;0.21)	0.86 (0.43;1.0)	7.5 (5.13;11)	13.5 (13;NA)	253.3 (-654.3;3373.2)	1689.5 (62.3;8232.1)	3.6	-73.9 (-260.96;16.1)	0.74 (0.45;0.82)	27.17 (11.43;53.07)
5	0.06 (0.0;0.34)	1.0 (0.89;1.0)	5.25 (1.38;11)	16 (12;NA)	229.1 (-1159.2;2308.2)	1674.6 (69.5;14695.4)	2.1	-100.3 (-123.4;11.5)	0.7 (0.41;0.87)	52.54 (26.25;130.57)
5SAFE	0.19 (0.07;0.91)	1.0 (0.93;1.0)	5.25 (1.38;11)	NA (39;NA)	187.8 (-363.1;1720.5)	1660.8 (67.8;11924.7)	0.8	-281.7 (-297.3;-56.9)	0.59 (0.31;0.73)	184.49 (40.87;415.58)
6	0.48 (0.18;0.93)	1.0 (0.97;1.0)	5 (1.5;10)	NA (40;NA)	188.5 (-469.2;1931.7)	1451.3 (70.3;12360.1)	0.51 [1.68]	-335.1 (-371.5;-88.4)	0.55 (0.24;0.67)	252.09 (106.41;589.08)
7	0.73 (0.30;0.93)	1.0 (0.95;1.0)	4.5 (1.25;8.9)	NA	179.8 (-471.3;1817.8)	1573.7 (70.4;12298.8)	0.46 [3.23]	-367.4 (-394.4;-100.1)	0.54 (0.27;0.66)	229.66 (90.23;596.69)

Some of the differences between species are due to their initial biomass relative to B_0. The number of years it takes to fish down to $0.48B_0$ for those species initially above $0.48B_0$ (R3a in Table D.3) highlights that the more data-limited tiers (without any meta-rules) tend to fish down the population slightly faster than tier 1. Including the meta-rules has the effect of extending this duration, and also separating the results among tiers. Only tiers 1, 3 and 4 generally recover species that are initially below $0.2B_0$ to $0.2\ B_0$ and higher (R3b in Table D.3), with tier 3 (and an un-tuned tier 4, results not shown here) typically taking considerably longer than tier 1 (or a tuned tier 4). The meta-rules reduce the rate of recovery. Note that tier 3 is often less precautionary than tier 4 (especially tuned tier 4) and is therefore out of sequence if the tiers are meant to reflect degree of risk.

By design, the meta-rules do not allow for exceedingly large changes to the TAC between years. However, the effect of these rules depends on stock status. They are more effective at recovering a resource using a data-limited HS because they reduce the large TACs these HSs tend to set. However, for a declining resource, restricting changes in TACs via meta-rules may inhibit recovery or even lead to further decline. The combination of these two dynamics leads to complicated effects of meta-rules temporally. While they can hinder the reversal of a stock decline in the short to medium term, in the longer term (once the stock decline has been realised and reversed) the use of meta-rules can see more conservative TACs (avoiding large and rapid increases), ultimately producing biomasses closer to $0.48B_0$.

The number of species that remain below $0.2B_0$ even towards the end of the projection increases for the more data-limited tiers (Figure D.2). This increased risk by tier is due to a number of factors. First, the F (or its proxy) values estimated by tiers 3 to 5-SAFE relative to F_{MEY} are lower than for tier 1 (Figure D.3) – with the exception of the tuned tier 4, which can be more conservative than tier 1 (although F/F_{MEY} for tier 4 is based on CPUE rather than fishing mortality). Tiers 3-5 therefore tend allow for more overfishing because TACs remain higher than they should be given the assessment underestimates past fishing mortality relative to the target level and trying to "correct" for thus. This is why the higher (more data-limited) tiers tend to set higher TACs at the start of the projection period compared to tier 1 (Figure D.4). The cumulative impact would be large if this relative bias remains over the full time period, i.e. irrespective of stock size. There is no *a priori* reason for tiers 1 to 5-SAFE to have different relative biases relative to tier 1 over time, but tiers 6 and 7 might perform differently in terms of relative bias as biomass changes. This is because they depend on the level of conservatism set within in the first trigger levels, which are catch relative to the historical maximum catch (HMC) (tiers 6 and 7), and catch composition, CPUE area fished and spatial distribution trends (tier 7).

Tiers 6 and 7 perform the worst in terms of the risk performance metrics, because, as applied here, multiple trigger points need to be tripped before the TAC is changed. The triggers are not tripped annually, even for those species that are depleted at the start of the projection period. The level 1 trigger, requiring a tier 5-SAFE assessment, is typically tripped fewer than ten times during the 45 year projection period (Figure D.5) for either tier 6 or 7. The level 2 trigger, which leads to the use of the tier 3 HS, differs between tiers 6 and 7 and the difference in performance between tier 6 and 7 is greater when the resource is being fished hard (based on species level results not presented). There is a larger difference in performance between tiers 6 and 7, with tier 7 triggering a tier 3 assessment more often over the projection period (due to a slightly longer response time leading to a poorer stock status before any trigger is activated and because the additional CPUE and composition metrics activate the level-2 trigger more often than the catch volume alone does). These HSs are based on the HMC so their performance is very species- and history-specific. Catch-based triggers rely very heavily on presupposing a level of catch that is precautionary – something that is less likely to be known for a data-limited fishery (as also found by Wiedenmann et al. 2013). Furthermore, the way these HSs were implemented in Atlantis-RCC optimistically assumes that the information required when a tier 5-SAFE or tier 3 assessment is triggered is available.

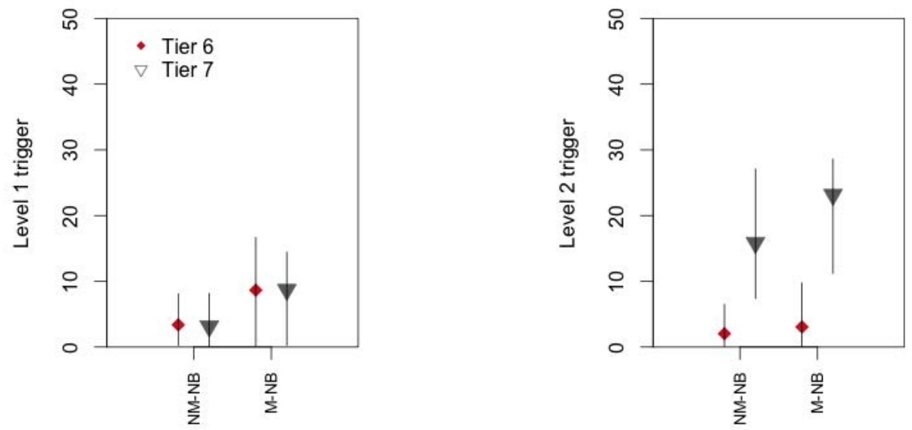


Figure D. 5 Median and inter-quartile range of the number of times the level 1 (SAFE) and level 2 (tier 3) assessments are triggered for tiers 6 and 7. 'NM' denotes without meta-rules and 'M' denotes with meta-rules.

The final reason the data-poor tiers are riskier is that response time is influential in dictating risk (Figure D.6). The best performance possible is two years, since it takes about two years to collect the data that highlights the issue, undertake the assessment and implement a lower TAC. In that regard, only tier 1 mostly acts within the minimum 2-year period. Tier 3 stands out compared to tiers 1 and 4 as having slow response times and therefore an increased risk. Tiers 5-SAFE to 7 take more than four years to identify implement a management change.

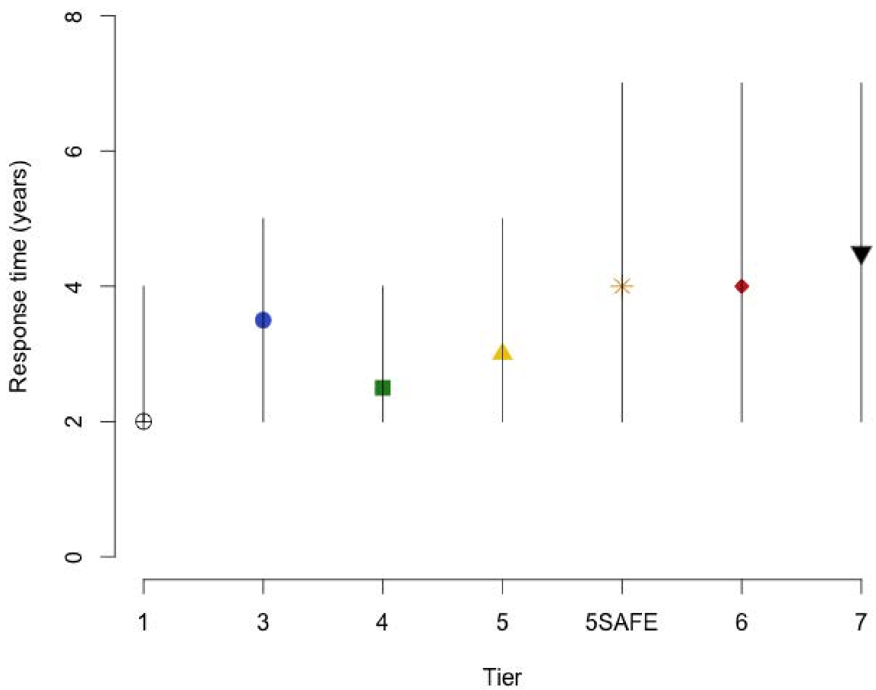


Figure D. 6 Median and inter-quartile range of the response time under the different tiers.

Overall, the more data-limited the tier, the higher the TACs in the short-term and hence the higher the risk. Meta-rules for an increasing stock help mitigate this risk by reducing unjustified TAC increases. The data-limited methods are slow to respond to potential reductions in stock size, because, relative to tier 1, they calculate more optimistic F/F_{MEY} values, and for tiers 6 and 7, their complexity renders them less sensitive to changes to which they should be responding.

Risk versus Catch

Risk and catch are often traded against each other, with the aim to manage a fishery sustainably by (at least) avoiding the target species falling below the LRP ($0.2B_0$) while remaining profitable (Little *et al*., 2016). However, the more data-limited HSs may fail to identify a long-term sustainable catch (Table D.3a), due to their inability to effectively estimate biomass or fishing mortality. This poor performance is expressed as lower stock status and consequently lower cumulative catches. Although there is a large range in the value of the performance metrics per tier due to species-specific differences (often due to different starting stock status and-TACs) (Figure D.2), the median values show that the more data-limited the tier, the lower the discounted catch over the projection period (Table D.3). The relative discounted catch is much higher for the data-limited tiers than for the data-rich tiers early in the projection period (not shown), as the species is being overfished. Ultimately, however the more degraded stock status means catches are lower than for the healthier stocks under the data-rich tiers and so the cumulative result is that the data-rich tiers lead to higher relative discounted catches.

Costs

A further consideration is the cost of implementing a specific tier. Although there is an increased risk of using more data-limited approaches, the question is whether moving to a tier 1 rule is prohibitively expensive, particularly for a small-scale fishery. Calculating costs is complex given that there are tangible (costs of management and fishing) and intangible (the opportunity costs associated with catch foregone because the fishery was not managed "optimally") costs (Little *et al*., 2016). In this paper, the reference HS – a "bang-bang" HS, which takes all biomass above a target level - is the closest HS possible within the Atlantis-RCC modelling framework to an "optimal" HS. It should be noted that this reference HS does not consider economic or social costs of implementation (such as inter-annual volatility in the catch or jobs lost due to prolonged closures), and so is not a realistic HS that could be implemented. Despite this, the difference in opportunity cost (C4) for tier 1 is quite large – in the order of a year's value of the fishery or more (e.g. gross value of product of $72.3 million AUD in the 2013-14 financial year; Georgeson *et al*. 2015). Opportunity costs are variable among tiers but tend to be higher for the more data-limited tiers. Discounted profits vary by tier, with the largest for tier 3, then tier 1 and 4 with the remainder not being profitable and worsening with increasing tier number, i.e. as the probability or risk of overfishing, or being overfished increases. With meta-rules, only tier 1 is profitable. The management costs (C6), decrease for the more data-limited tiers, but are two orders of magnitude smaller than other costs. However, costs for tier 6 and 7 increase when higher tiers are triggered. Harvesting costs (C5) are more difficult to interpret. The highest of these are for tier 1 (because the higher stock size allows for greater catches, with associated higher handling and fishing costs) and then tier 3. Catch variability (H9) is highest for the data-limited tiers, which had the additional outcomes of seeing more vessels opting not to fish (for at least part of the year) under these data-poor tiers and the potential for greater undercatch of TACs.

Of the more data-rich tiers, Tier 3 has the lowest opportunity costs, highest profits, and lowest harvesting costs (c.f. tiers 1 and 4), and lower catch variability than tier 1. However, this comes at the cost of poorer performance in terms of rebuilding to $0.2B_0$ for many species. This is a clear case of where those charged with balancing the trade-offs associated with the fishery would need to decide which criteria they weight more heavily – stock status and the risk of falling below limit reference points (a mandated concern under Australian policy) or economic performance. The need for frank discussion and careful attention is particularly important given the species specific performance of this tier.

Influence of history and ecology

The results per tier were often species-specific. Tier 1 had the potential to recover the majority of life histories, except those that were particularly depressed historically, to the point that even small incidental catches were enough to keep them depleted – due to very low rates of reproduction (gulper sharks) or a combination of technological and ecological interactions (gemfish). Tiers 3 and 4 are heavily influenced by the history of the fishery. Tier 3 was more sensitive to the ecology of the species than the other tiers, particularly if there are strong time-varying ecological interactions or shifts in productivity. Tier5-SAFE often worked well for chondrichthyans, but was not as effective for teleosts, especially those with highly variable recruitment. In general, the more data-limited tiers can perform adequately if the stock is productive and large in absolute size. However, the initial catches need to be precautionary for these tiers to perform adequately if the stock is small or has low productivity.

Comparison with other MSE studies

The results in terms of the R1 performance metric are similar to those from single-species MSEs, where the tier 1 to 4 HSs perform well with respect to recovering a resource to above $0.2B_0$, although tier 3 performed worse that tier 4 (Fay *et al.*, 2012; Little *et al.*, 2014). The species-specific nature of their results was also highlighted by these previous studies.

Some species (e.g. gulper shark and gemfish) did not recover. There are several reasons for this: a) species are included in this study that started the projection period below $0.2B_0$, and their longevity and life history characteristics mean that, even with good management, these species would not necessarily recover by the end of the projection period (e.g. gulper shark); b) environmental conditions and trophic interactions are included in Atlantis-RCC, but not in single-species MSEs, and c) implementation uncertainty (Fulton *et al*., 2011), including the potential for a low level of illegal activity, is overtly accounted for through a multi-species fleet dynamics module and a quota trading module in Atlantis-RCC.

The various combinations of data, assessment methods and associated decision rules (i.e. HSs) to set a management measure, such as a TAC, in the EU framework were simulation tested for risk equivalency across categories 1 to 4 by ICES (2013). That work found that the impact of the meta-rules was similar to this paper's findings: the rules were useful when a resource was recovering or in a healthy condition, but tended to slow down action to stop a declining resource. As such, meta-rules are very context specific, especially relative to stock status and should be used with care (ICES, 2013), and there needs to be an appreciation that the effects will vary through time and with stock status.

Caveats

While the work presented here attempted to be comprehensive in terms of including the kinds of processes and data imperfections that a real world harvest strategy must contend with, it remains a model-based study. While a broad range of demersal life histories was considered (spanning the majority of those assessed in the SESSF), the range is not exhaustive, and caution will need to be applied around life histories not considered here (such as herbivores, both short-lived and sedentary invertebrates, as well as forage fish). Furthermore, the ecological components of Atlantis are deterministic, and so the model only includes limited stochasticity. In addition, while ecosystem models such as Atlantis are highly uncertain, the technical difficulty in applying the multiple size-at-age trajectories and smooth feeding window meant that it was not possible to find multiple suitable parameterisations. Consequently, the usual practice of running Atlantis with multiple alternative parameterisations was not possible in this instance and so the results should be treated with some caution, as a different parameterisation could lead to somewhat different outcomes. That stated, we do have considerable confidence in the robustness of our results given their general agreement with single-species MSE testing of many of tiers 1-4 (Wayte and Klaer, 2010; Fay *et al*., 2011; Little *et al.*, 2011; Klaer *et al*., 2012).

While stochastic ecosystem models (e.g. OSMOSE; Shin and Cury, 2004) are now beginning to be used more often, deterministic models have historically been the norm. This has meant that the majority of previous ecosystem model based studies have utilized a single iteration (e.g. Kaplan *et al.*, 2014), with only a limited number of studies including multiple iterations or parameterizations (Fulton *et al.*, 2014). Our study was constrained in its capacity to use multiple parameterisations, but was able to include multiple iterations. Previous experience with Atlantis, which is at heart a deterministic model, has shown that a small number of iterations is sufficient for capturing the general patterns of response. Consequently, only 20 iterations were undertaken due to the extensive computer time (several months) required given the need to link stock assessments and a full end-to-end model in a single framework. As noted in the text, we did conduct additional simulations to assess how sensitive results were to the number of replicates and found this to be minimal. There was also very little between-replicate variation, with the largest instantaneous range in biomass for any one species across its treatment simulations being 10%. This gave us confidence that our use of 20 iterations was not misleading in terms of patterns of response. However, the small number of replicates does limit the extent to which risk based measures can be interpreted as fully representing the statistical population of possible values. Nevertheless, even though the precision of percentiles will be limited, a strength of methodology here is that the simulations are paired (i.e. the same random numbers are applied for all comparable scenarios), which substantially increases the ability to compare options in a relative sense.

Conclusion

The multi-tier HS system does not, according to our results, achieve risk equivalency, especially for data-limited approaches. Profitability is lowest for the more data-limited approaches, despite these approaches having the lowest harvesting costs. The data-rich HSs benefit from decreases in opportunity costs, although there is not a consistent result among tiers. Tier 3 stands out by being less precautionary compared to tier 4 and therefore is out of sequence compared to tier 1 and 4 (if higher tier numbers are meant to designate more risky HSs). Profitability increases as more data-rich HSs are used. The meta-rules mitigate large unjustified TAC increases of the data-limited approaches; generally leading to a more precautionary approach when applied to a recovering stock, but not when applied to a declining one. As currently formulated/specified, the data-limited HSs are slow to respond to a stock decline or depletion; either because of their lack of conservatism when estimating key parameters (e.g. for tier 5 variants) and due to their complexity in the case of hierarchical approaches (for tiers 6 and 7). Overall, if risk, cost and catch are considered, the data–rich approaches would be favoured.

Acknowledgements

We would like to thank two anonymous reviewers and the CSIRO internal reviewers for their comments on draft versions of the paper, as well as the various SESSF RAGs and AFMA members who provided input. This work was funded by Fisheries Research and Development Corporation Project 2012/202 and CSIRO Oceans and Atmosphere Flagship.

References

Anon. 2014. Harvest strategy framework for the southern and eastern Scalefish and shark fishery. Report to Australian Fisheries Management Authority, Canberra. 25 pp. http://www.afma.gov.au/wp-content/uploads/2014/11/Harvest-Strategy-Framework-SESSF-Feb-2014.pdf (last opened 1 November 2015).

Butler A., Harris P., Lyne V., Heap A., Parslow V., and Porter-Smith R. 2001. An Interim Bioregionalisation for the continental slope and deeper waters of the South-East Marine Region of

Australia. http://www.environment.gov.au/resource/interim-bioregionalisation-continental-slope-and-deeper-waters-south-east-marine-region (last accessed 1 September 2015).

Butterworth, D. S., Cochrane, K. L., and De Oliveira, J. A. A. 1997. Management procedures: a betterway to manage fisheries? The South African experience. American Fisheries Society Symposium Series, 20. *In* Global Trends: Fisheries Management, pp. 83–90. Eds by E. K. Pikitch, D.D. Huppert, and M.P. Sissenwine. American Fisheries Society, Seattle. 352 pp.

DAFF. 2007. Commonwealth Fisheries Harvest Strategy Policy Guidelines. Australian Government Department of Agriculture, Fisheries and Forestry, Canberra, Australia, pp. 55. http://www.agriculture.gov.au/fisheries/domestic/harvest_strategy_policy

Deroba, J. J., and Bence, J. R. 2008. A review of harvest policies: Understanding relative performance of control rules. Fisheries Research, 94: 210–223.

Day, J., and Klaer, N. 2014. Tiger flathead (*Neoplatycephalus richardsoni*) stock assessment based on data up to 2012. *In* Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2013, Part 1, Chapter 10, pp 174–232. Ed. by G.N. Tuck. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 210 pp.

Day, J., Klaer, N., and Tuck, G. N. 2013. Silver Warehou (*seriolella punctata)* stock assessment based on data up to 2013. *In* Stock Assessment for the Southern and Eastern Scalefish and Sharkfish: 2012, Part 1, Chapter 7, pp. 120–151. Ed. by G.N. Tuck. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 210 pp. http://www.afma.gov.au/wp-content/uploads/2010/06/SESSF-2013-Part-1-Final.pdf (last accessed 1 November 2015).

De la Mare, W. K. 1996. Some recent developments in the management of marine living resources. *In* Frontiers of Population Ecology. pp. 599–616. Eds by R.B. Floyd, A.W. Shepherd, and P.J. De Barro. CSIRO Publishing, Melbourne, Australia.

Dichmont, C.M., Punt, A.E., Dowling, N., De Oliveira, J.A.A., Little, L.R., Sporcic, M., Fulton, E., Gorton, R., Klaer, N., Haddon, M., and Smith, D.C. 2015. Is risk consistent across tier-based harvest control rule management systems? A comparison of four case studies. Fish and Fisheries. doi: 10.1111/faf.12142.

Dichmont, C. M., Deng, R., Punt, A. E., Venables, W., and Haddon, M. 2006. Management strategies for short lived species: The case of Australia's Northern Prawn Fishery. 1. Accounting for multiple species, spatial structure and implementation uncertainty when evaluating risk. Fisheries Research, 82: 204–220.

Dowling, N. A., Punt, A. E., Little, L. R., Dichmont, C., Sporcic, M., Fulton, E .A., Gorton, R. J., Haddon, M., and Smith, D. C. 2016. Assessing a multilevel tier system: the role and implications of data quality and availability. Fisheries Research, 183, 588–593. http://dx.doi.org/10.1016/j.fishres.2016.05.001.

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., Smith, D.C., Knuckey, I., Smith, A.D.M., Domaschenz, P., Patterson, H.M., and Whitelaw, W. 2008. Developing harvest strategies for low value and data-poor fisheries: case studies from three Australian fisheries. Fisheries Research, 94: 390-390.

European Commission. 2013. The Common Fisheries Policy (CFP). http://ec.europa.eu/fisheries/cfp/index_en.htm (last accessed 1 November 2015).

Fay, G., Little, L.R., Tuck, G.N., Haddon, M., and Klaer, N. 2012. Maintaining risk equivalency among fishery harvest control rules in Southeast Australia. Report to the Australian Fisheries Management Authority, Hobart.

Fay, G., Punt, A .E., and Smith, A. D. M. 2009. Operating model specifications. *In* Evaluation of New Harvest Strategies for SESSF Species. pp. 125–133. Ed. by S.E. Wayte. CSIRO Marine and Atmospheric Research, Hobart and Australian Fisheries Management Authority, Canberra

Fay, G., Punt, A. E., and Smith, A. D. M. 2011. Impacts of spatial uncertainty on the performance of age structure-based harvest strategies for blue-eye trevalla (*Hyperoglyphe antarctica*). Fisheries Research, 110; 391–407.

Fulton, E. A., and Gorton, R. 2014. Adaptive Futures for SE Australian Fisheries & Aquaculture: Climate Adaptation Simulations. CSIRO, Australia. pp 309. http://frdc.com.au/research/Final_Reports/2010-023-DLD.pdf (last accessed 3 November 2015).

Fulton E. A., Smith, A. D. M., and Smith, D. C. 2007. Alternative management strategies for Southeastern Australian Commonwealth Fisheries: Stage 2: Quantitative Management Strategy Evaluation. Report to the Australian Fisheries Management Authority and the Fisheries Research and Development Corporation. CSIRO Marine and Atmospheric Research. http://atlantis.cmar.csiro.au/www/en/atlantis/mainColumnParagraphs/02/text_files/file/AMS_Final_Report_v6.pdf (last accessed 3 November 2015).

Fulton, E. A**.**, Link, J., Kaplan, I. C., Johnson, P., Savina-Rolland, M., Ainsworth, C., Horne, P., Gorton, R., Gamble, R. J., Smith, T., and Smith D. 2011. Lessons in modelling and management of marine ecosystems: The Atlantis experience. Fish and Fisheries, 12: 171–188

Fulton, E. A., Smith, A. D. M., Smith, D. C., and Johnson, P. 2014. An integrated approach is needed for Ecosystem Based Fisheries Management: Insights from ecosystem-level management strategy evaluation. PLoS One.

Georgeson, L., Ward, P., and Curtotti, R., 2015. Southern and Eastern Scalefish and Shark Fishery. *In* Fishery status reports 2015, Chapter 8, pp. 112-127. Ed. by H. Patterson, L. Georgeson, I. Stobutzki, and R. Curtotti. Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra. 487 pp. http://data.daff.gov.au/data/warehouse/9aam/fsrXXd9abm_/fsr15d9abm_20151030/08_FishStatus2015SthnEastnScalefishShark_1.0.0.pdf (last accessed 1 November 2015).

Haddon, M., Klaer, N., Wayte, S., and Tuck, G., 2015. Options for Tier 5 approaches in the SESSF and identification of when data support for harvest strategies are inappropriate. Report to Fisheries Research and Development Corporation. CSIRO Oceans and Atmosphere, Hobart. pp. 115.

ICES. 2012. ICES Implementation of Advice for Data-limited Stocks in 2012 in its 2012 Advice. ICES DLS Guidance Report 2012. ICES CM 2012/ACOM 68, 42 pp. http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2012/ADHOC/DLS%20Guidance%20Report%202012.pdf (last accessed 3 November 2015).

ICES. 2013. Report of the Workshop on the Development of Quantitative Assessment Methodologies based on LIFE-history traits, exploitation characteristics, and other key parameters for Data-limited Stocks (WKLIFE III). ICES CM 2013/ACOM:35. pp. 84. http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2013/WKLIFE3/Report%20WKILFE%20III.pdf. (last accessed 1 November 2015).

Interim Marine and Coastal Regionalisation for Australia Technical Group (IMCRA) 1998. Interim Marine and Coastal Regionalisation for Australia: an ecosystem-based classification for marine and coastal environments. Version 3.3. Environment Australia, Commonwealth Department of the Environment: Canberra. http://www.environment.gov.au/resource/interim-marine-and-coastal-regionalisation-australia-version-33 (last accessed 1 September 2015).

Jensen, A.L., 1996. Beverton and Holt life history invariants result from optimal tradeoff of reproduction and survival. Canadian Journal of Fisheries and Aquatic Science, 53: 820–822.

Johnson P., Fulton E. A., Smith D. C., Jenkins G. P., and Barret N. 2011. The use of telescoping spatial scales to capture inshore to slope dynamics in marine ecosystem modelling. Natural Resource Modeling, 24: 335–364

Kaplan, I.C., Holland, D.S., and Fulton, E.A. 2014. Finding the accelerator and brake in an individual quota fishery: linking ecology, economics, and fleet dynamics of US West Coast trawl fisheries. ICES Journal of Marine Science. 71: 308–319.

Klaer, N. L., Wayte, S. E., and Fay, G. 2012. An evaluation of the performance of a harvest strategy that uses an average-length-based assessment method. Fisheries Research, 134-136: 42–51.

Kramer-Schadt S., Revilla E., Wiegand T., and Grimm, V., 2007. Patterns for parameters in simulation models. Ecological Modeling, 204: 553–556.

Little, L. R., Punt, A. E., Dichmont, C. M., Dowling, N., Smith, D. C., Fulton, E., Sporcic, M., and Gorton, R. J. 2016. Decision trade-offs for cost constrained fisheries management. ICES J. Marine Science, 73: 464-502.

Little, L. R., Parslow, J., Fay, G., Grafton, R.Q., Smith, A.D.M., Punt, A.E., and Tuck, G.N. 2014. Environmental derivatives, risk analysis and conservation management. Conservation Letters, 7: 196-207.

Little L. R., Punt, A. E., Mapstone, B. D., Pantus, F., Smith, A. D. M., Davies, C.R., and McDonald, A.D. 2007. ELFSim – A Model for Evaluating Management Options for Spatially-Structured Reef Fish Populations: An Illustration of the "Larval Subsidy" Effect. Ecological Modelling, 205: 381–396.

Little, L. R., Wayte, S. E., Tuck, G. N., Smith, A. D. M., Klaer, N., Haddon, M., Punt, A. E., Thomson, R., Day, J., and 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.

Lyne V., and Hayes D. 2005. Pelagic Regionalisation. National Oceans Office and CSIRO Marine and Atmospheric Research, Hobart. http://www.environment.gov.au/system/files/resources/b7d7587a-6330-41fe-8f97-cba74ad87306/files/nmb-pelagic-report.pdf (last accessed 1 September 2015).

Methot, R. D., and Wetzell, C. R. 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142: 86–99.

Morison, A. K., Knuckey, I. A., Simpfendorfer, C. A., and Buckworth, R.C. 2012. 2011 Stock Assessment Summaries for the South East Scalefish and Shark Fishery. http://www.afma.gov.au/wp-content/uploads/2010/07/2011-SESSF-Species-Summaries-March-2012.pdf. (last accessed 3 November 2015).

Oke P. R., Schiller A., and Griffin D. A. B. 2005. Ensemble data assimilation for an eddy-resolving ocean model. Quarterly Journal of Royal Meteorological Society, 131: 3301–3311.

Pantus F. J., and Dennison W. C. 2005. Quantifying and Evaluating Ecosystem Health: A Case Study from Moreton Bay, Australia. Environmental Management, 36: 757–771.

Punt, A.E., Deng, R.A., Pascoe, S., Dichmont, C.M., Zhou, S., Plagányi, É.E., Hutton, T., Venables, W.N., Kenyon, R., and van der Velde, T. 2011. Calculating optimal effort and catch trajectories for multiple species modelled using a mix of size-structured, delay-difference and biomass dynamics models. Fisheries Research, 109: 201-211.

Punt, A. E. 1992. Management procedures for Cape hake and baleen whale resources. BEP Report No 23. pp. 689.

Punt, A. E., and Smith, A. D. M. 1999. Harvest strategy evaluation for the eastern stock of gemfish (*Rexea solandri*). ICES Journal of Marine Science, 56: 860-875.

Punt, A. E., Butterworth, D. S., de Moor, C. L., De Oliveira, J. A. A., and Haddon, M. 2016. Management Strategy Evaluation: Best Practices. Fish and Fisheries, 17: 303-334.

Punt, A. E., Smith, A. D. M., and Cui, G. 2002. Evaluation of management tools for Australia's South East Fishery. 3. Towards selecting appropriate harvest strategies. Marine and Freshwater Research, 53: 645–660.

Punt, A. E., Campbell, R., and Smith, A. D. M. 2001. Evaluating empirical indicators and reference points for fisheries management: Application to the broadbill swordfish fishery off Eastern Australia. Marine and Freshwater Research, 52: 819-832.

Rayns, N. 2007. The Australian government's harvest strategy policy. ICES Journal of Marine Science, 64: 596–598.

Sainsbury K. 2005. Cost-effective management of uncertainty in fisheries. Outlook 2005. Canberra, A.C.T.: Australian Bureau of Agricultural and Resource Economics. http://data.daff.gov.au/data/warehouse/pe_abarebrs99001173/PC13024.pdf (last accessed 7 September 2015).

Sainsbury, K. J., Punt, A. E., and Smith, A. D. M. 2000. Design of operational management strategies for achieving fishery ecosystem objectives. ICES Journal of Marine Science, 57**:** 731–741.

Savina M., Fulton E. A., Condie S., Forrest R., Scandol J., Astles K. and Gibbs P. 2008. Ecologically sustainable development of the regional marine and estuarine resources of NSW: Modelling of the NSW continental shelf ecosystem. CSIRO and NSW DPI Report. http://www.dpi.nsw.gov.au/research/areas/aquatic-ecosystems/outputs/2009/1196 (last accessed 7 September 2015).

Shin, Y.-J. and Cury, P., 2004. Using an individual-based model of fish assemblages to study the response of size spectra to changes in fishing. Canadian Journal of Fisheries and Aquatic Sciences, 61: 414–431.

Smith, A. D. M., Sainsbury, K. J., and Stevens, R. A. 1999. Implementing effective fisheries-management systems—management strategy evaluation and the Australian partnership approach. ICES Journal of Marine Science, 56: 967–979.

Smith, A. D. M., Fulton, E. J., Hobday, A. J., Smith, D. C., Shoulder, P. 2007. Scientific tools to support the practical implementation of ecosystem-based fisheries management. ICES Journal of Marine Science., 64(4): 633-639. doi: 10.1093/icesjms/fsm041.

Smith, A. D. M., Smith, D. C., Haddon, M., Knuckey, I. A., Sainsbury, K. J., and Sloan, S. R. 2014. Implementing harvest strategies in Australia: 5 years on. ICES Journal of Marine Science, 71: 195–203.

Tuck, G. N. 2014. Stock assessment of blue grenadier (*Macruronus novaezelandiae*) based on data up to 2012. *In* Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2013, Part 1. pp 61–115. Ed. By G.N. Tuck. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research. 210 pp. http://www.afma.gov.au/wp-content/uploads/2010/06/SESSF-2013-Part-1-Final.pdf (last accessed 3 November 2015).

USA Department of Commerce (USA Doc). 2007. Magnuson-Stevens Fishery Conservation and Management Act as Amended Through January 12, 2007. http://www.nmfs.noaa.gov/sfa/magact/MSA_Amended_2007%20.pdf (last accessed 18 July 2014).

van Putten, I. E., Gorton, R .J., Fulton E. A. and Thebaud, O. 2013. The role of behavioural flexibility in a whole of ecosystem model. ICES Journal Marine Science, 70: 150–163.

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

Wiedenmannm J. Wilberg, M. J., and Miller, T. J. (2013. An evaluation of harvest control rules for data-poor fisheries. North American Journal of Fisheries Management, 33: 845-860.

Zhou, S., Smith, A. D. M., and Fuller, M., 2011. Quantitative ecological risk assessment for fishing effects on diverse data-poor non-target species in a multi-sector and multi-gear fishery. Fisheries Research, 112: 168–178.

Supplementary Materials

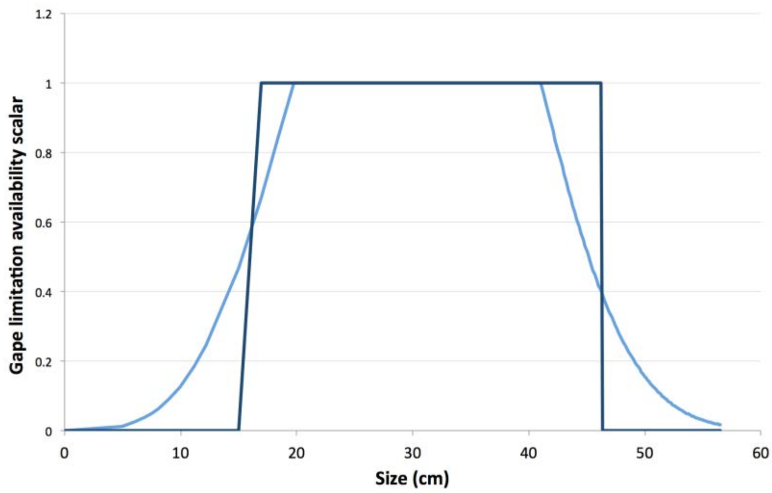
Table D. S1. Trophic groups included in Atlantis-RCC. Species in bold have been modified to allow for multiple size-at-age growth curves (see description in the Size-at-Age section of main text for details). The species marked with an asterisk have multiple stocks.

Model Component	Group Composition
Pelagic invertebrates	
Large phytoplankton	Diatoms
Small phytoplankton	Picophytoplankton
Small zooplankton	Heterotrophic flagellates
Mesozooplankton	Copepods
Large zooplankton	Krill and chaetognaths
Gelatinous zooplankton	Salps (pryosomes), coelenterates
Pelagic bacteria	Pelagic attached and free-living bacteria
Squid	*Sepioteuthis australis*, *Notodarus gouldi*
Benthic invertebrates	
Sediment bacteria	Aerobic and anaerobic bacteria
Carnivorous infauna	Polychaetes
Deposit feeders	Holothurians, echinoderms, burrowing bivalves
Deep water filter feeders	Sponges, corals, crinoids, bivalves
Shallow water filter feeders	Mussels, oysters, sponges, corals
Scallops	*Pecten fumatus*
Herbivorous grazers	Urchins, *Haliotis laevigata*, *Haliotis rubra*, gastropods
Deep water megazoobenthos	Crustacea, asteroids, molluscs
Shallow water megazoobenthos	Stomatopods, octopus, seastar, gastropod, and non-commercial crustaceans
Rock lobster	*Jasus edwardsii*, *Jasus verreauxi*
Meiobenthos	Meiobenthos
Macroalgae	Kelp
Seagrass	Seagrass
Prawns	*Haliporoides sibogae*
Giant crab	*Pseudocarcinus gigas*
Fin-fish	
Small pelagics*	Sardinops, sprat, *Engraulis*
Redbait	Emmelichthyidae (*Emmelichthys nitidus*)
Mackerel*	*Trachurus declivis*, *Scomber australisicus*
Migratory mesopelagics	Myctophids
Non-migratory mesopelagics	Sternophychids, cyclothene (lightfish)
School whiting*	*Sillago*
Shallow water piscivores	*Arripis*, *Thyrsites atu*, *Seriola*, leatherjackets
Blue warehou*	*Seriolella brama*
Spotted warehou	*Seriolella punctata*
Tuna and billfish*	*Thunnus*, *Makaira*, *Tetrapturus*, *Xiphias*
Gemfish*	*Rexea solandri*
Shallow water demersal fish*	Flounder, *Pagrus auratus*, Labridae, *Chelidonichthys kumu*, *Pterygotrigla*, *Sillaginoides punctata*, *Zeus faber*
Flathead*	*Neoplatycephalus richardsoni*, *Platycephalus*
Redfish*	*Centroberyx*
Morwong*	*Nemadactylus*
Pink ling*	*Genypterus blacodes*
Blue grenadier	*Macruronus novaezelandiae*
Blue-eye trevalla	*Hyperoglyphe antarctica*
Ribaldo	*Mora moro*
Orange roughy*	*Hoplostethus atlanticus*
Dories and oreos*	Oreosomatidae, Macrouridae, *Zenopsis*
Cardinalfish	Cardinalfish
Sharks	

Model Component	Group Composition
Gummy shark*	Mustelus antarcticus
School shark*	Galeorhinus galeus
Demersal sharks	*Heterodontus portusjacksoni*, Scyliorhinidae, Orectolobidae
Pelagic sharks	*Prionace glauca*, *Isurus oxyrunchus*, *Carcharodon carcharias*, *Carcharhinus*
Dogfish	Squalidae
Gulper sharks	Centrophorus
Skates and rays	Rajidae, Dasyatidae
Top predators	
Seabirds	Albatross, shearwater, gulls, terns, gannets
Seals	*Arctocephalus pusillus doriferus*, *Arctocephalus forsteri*
Sea lion	*Neophoca cinerea*
Dolphins	Delphinidae
Orcas	*Orcinus orca*
Baleen whales	*Megaptera novaeangliae*, *Balaenoptera*, *Eubalaena australis*

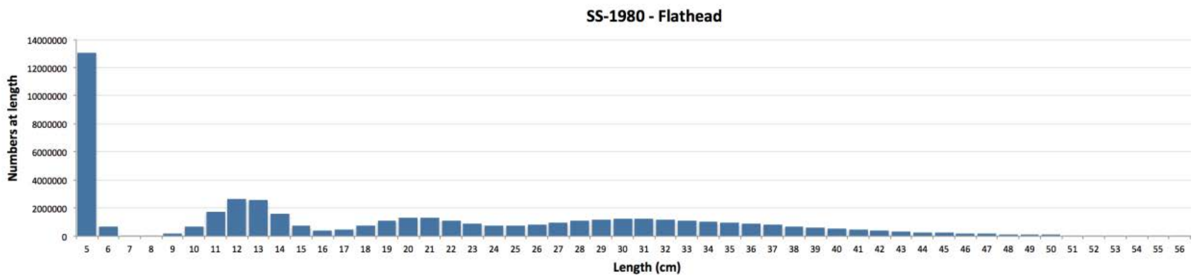
Table D.S2. Table of the existing management tiers and number fleets used in the tier 1 assessment species criteria by species. Species selection criteria by species also included. For the ecological characteristics P = piscivores, O = omnivore. Note 5S stands for 5-SAFE. The number of fleets used in the tier 1 assessments indicates the number per management stock, where east (E) and west (W) represents the split here between Redfish and Bight redfish. + Morwong have had tier 1 assessments in the past. Stock status here reflects real world status versus target and limit reference points (especially during the period spanning the end of the historical / beginning of the projection period). The errors used in the fleet specific data distributions are assumed to be unbiased and log-normally distributed, while the aging and sizing errors are also unbiased but normally distributed.

Criteria	Tiger flathead	Jackass morwong	School whiting	Gemfish	Blue warehou	Blue grenadier	Redfish‡	Orange roughy	Blue-eye trevalla	Pink ling	Gummy shark	Demersal sharks	Demersal fish	Gulper sharks
Existing management														
SESSF Tier applied in reality	1	1+	1	1	4	1	3	1	4	1	1	4	5S	5S
Number fleets used in tier 1 assessment	4	3E 1W	3	4E 2W	3E 1W	2	1E 1W	1	1	2E 2W	3	3	1	2
Species selection criteria														
Longevity														
Long							X	X						X
Medium	X	X		X	X	X			X	X	X		X	
Short			X											
Recruitment														
Random	X	X	X											
Periodic						X								
Regime Shift														
Shifted productivity		X												
Geography														
Inner shelf (reef or coastal)													X	
Shelf	X	X	X		X						X	X	X	
Slope or deepwater				X		X		X	X	X		X		X
Great Australian Bight				X			X						X	
Stock Status														
Healthy	X		X			X	X		X		X		X	
Borderline		X						X		X		X		
Overfished				X	X							X		X
Ecological characteristics														
Predator species	O	O	O	P	O	P	O	P	P	P	O	P	O	P
Prey species			X		X								X	
Not strongly connected								X						
Chondrichthyan											X	X		X
Data generation parameters														
Aging error relative standard error	0.103	0.055	0.073	0.1	0.05	0.103	0.1	0.08	0.1	0.1	0.1	0.05	0.055	0.05
Length measurement relative standard error	0.103	0.011	0.073	0.12	0.05	0.123	0.1	0.08	0.1	0.1	0.1	0.05	0.011	0.05
Fleet specific CPUE CV	0.5 0.2 0.05 0.4	0.1 for all	0.3 for all	0.2 for all	0.1 for all	0.1 for all	0.1 for all	0.1	0.2	0.1 for all	0.1 for all	0.1 for all	0.1	0.25 for all
Fleet specific landings CV	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1	0.1	0.1 for all	0.1 for all	0.1 for all	0.1	0.1 for all
Fleet specific discards CV	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1	0.1l	0.1 for all	0.1 for all	0.1 for all	0.1	0.1 for all

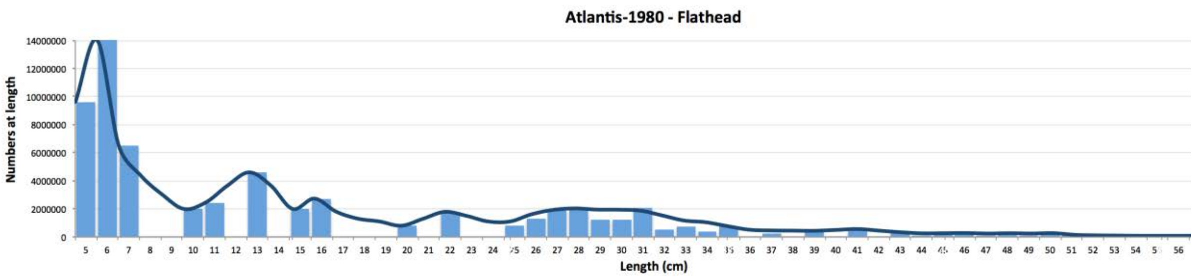


Online Figure D.S1: Comparison of different gape limitation functions - original Atlantis-SE (dark blue) and smooth curve of Atlantis-RCC (lighter blue).

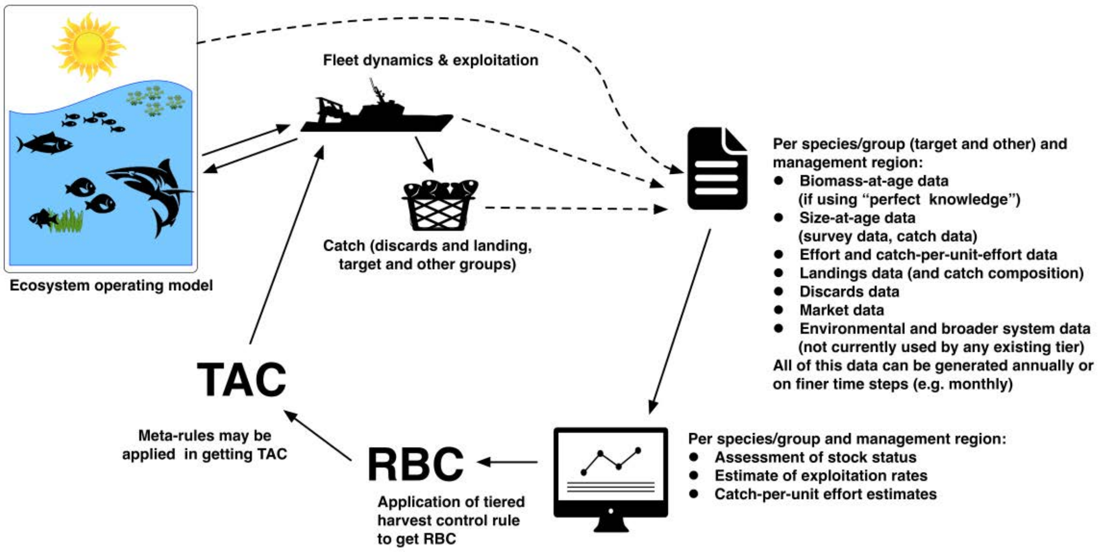
(a)



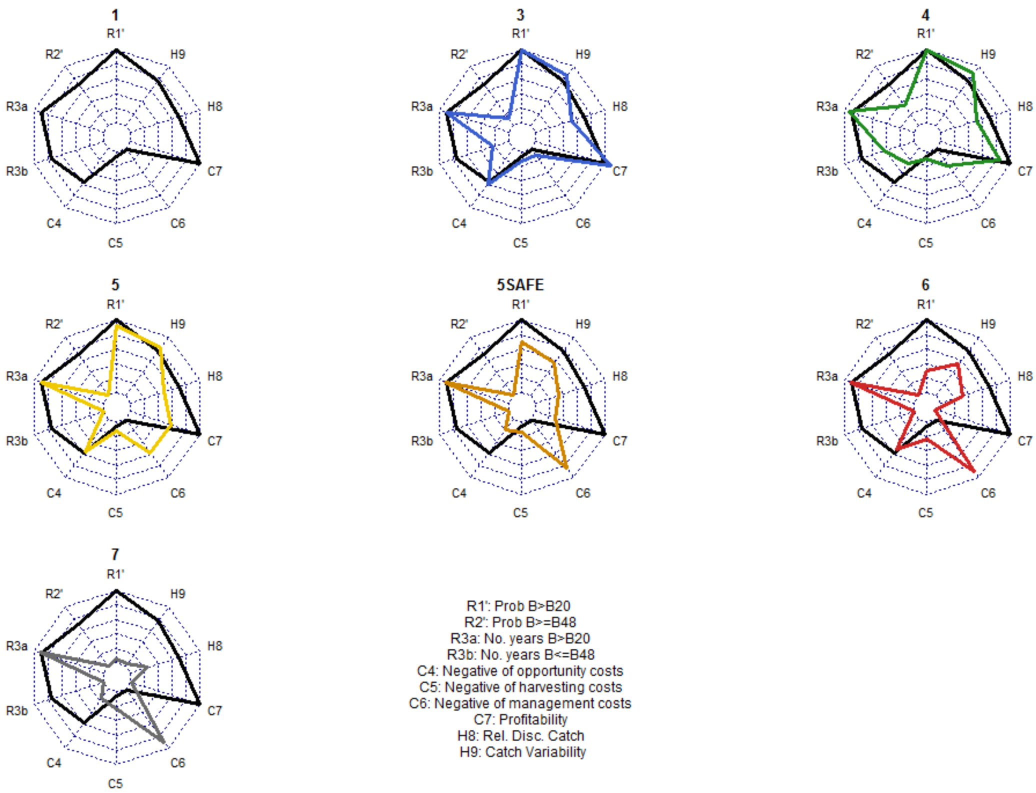
(b)



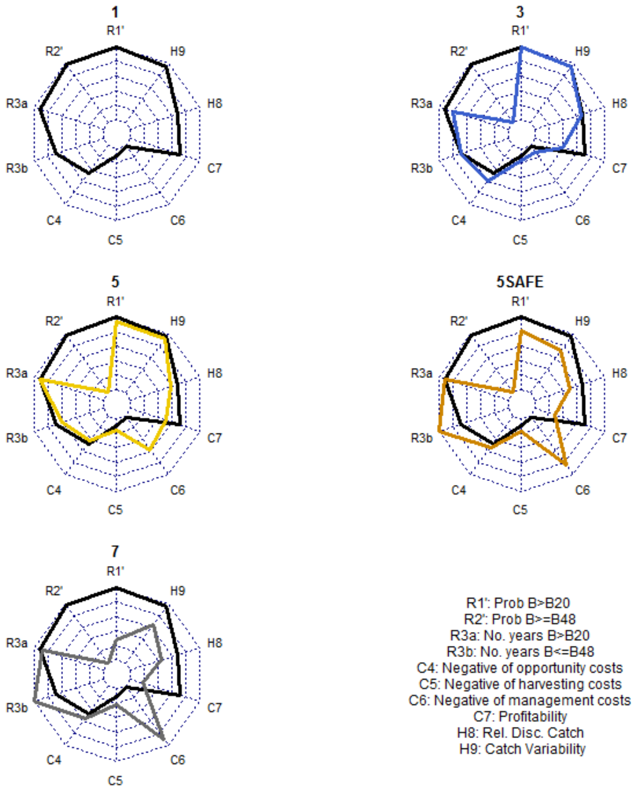
Online Figure D.S2: Example length frequency plots for Tiger Flathead in (a) SS3 and (b) Atlantis-RCC. By adding a smoother to the Atlantis structure something approximating the SS3 structure can be achieved so size-based assessment rules can be applied more easily (an example of the resulting smooth line is provided in the lower plot).



Online Figure D.S3. Schematic diagram of steps taken in each MSE simulation. TAC is the Total Allowable Catch and RBC is the Recommended Biological Catch.



Online Figure D. S4a: Performance metrics for the no meta-rules case for each tier. Note to make sure that "best performance" is towards the outside of the plot for all indicators the values given for R1' and R2' are (1-R1) and (1-R2) respectively. The black line represents the performance of tier 1 in all plots – provided for reference. The maximum and minimum values for R1'=(1, 0);   R2'=(0.5, 0), R3a=(40, 0), R3b=(40, 0), R4=(0,-400), C5=(0,-2000), C6=(0,-5.0), C7=(280,-500), H8=(1.2,0), H9=(0,-650))



Online Figure D.S4b: Performance metrics for the meta-rules case for each tier. Note to make sure that “best performance” is towards the outside of the plot for all indicators the values given for R1’ and R2’ are (1-R1) and (1-R2) respectively. The black line represents the performance of tier 1 in all plots – provided for reference.

Appendix E. Developing risk equivalent data-rich and data-limited harvest strategies

Elizabeth A. Fulton[1], André E. Punt[1,2], Catherine M. Dichmont[3], Rebecca Gorton[1], Miriana Sporcic[1], Natalie Dowling[1], L. Richard Little[1], Malcolm Haddon[1], Neil Klaer[1], David C. Smith[1]

[1]CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart, TAS, 7001, Australia

[2] School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, WA 98195, USA

[3]CSIRO Oceans and Atmosphere, Ecosciences Precinct, Dutton Park, QLD, 4750, Australia

Published in Fulton, E.A., Punt, A.E., Dichmont, C.M., Gorton, R., Sprocic, M., Dowling, N., little, L.R. Haddon, M., Klaer, N., Smith, D.C. 2016. Developing risk equivalent data-rich and data-limited harvest strategies. Fisheries Research. 183: 574-587. doi: 10.1016/j.fishres.2016.07.004.

Abstract

Several fisheries jurisdictions are aiming to achieve risk equivalency (here defined as the probability of a stock being depleted below a limit reference point or not being maintained at a target reference point) irrespective of the stock assessment method used to provide management advice and the amount of data available. Risk equivalency is implicitly required under the USA Magnuson-Stevens Act, while in Australia it is an explicit component of the Australian Commonwealth Government's Harvest Strategy Policy. Risk equivalency is well understood, but few fisheries have attempted to implement it. The Australian Southern and Eastern Scalefish and Shark Fishery (SESSF) is the only Australian fishery that has explicitly done so, albeit in a semi-arbitrary manner. Assessments and associated harvest strategies are placed into tiers from data-rich to data-limited. There are also meta-rules that control how much catch limits can change from one year to the next, and buffers by tier to achieve risk equivalency. Here, the SESSF tier system was evaluated in an ecosystem context using Management Strategy Evaluation. Two buffer systems were considered, the current SESSF system and a system inferred from how the Acceptable Biological Catches are set for the USA west coast groundfish fishery. Harvest strategies for all tiers were capable of moving productive stocks so their biomasses lay between the limit and target reference points. The USA buffer system was more conservative than the SESSF system, and achieved the fastest recovery for depleted stocks. The latter system led to slightly lower total catches, but was closest to achieving risk equivalency across the tiers. The USA buffer system led to biomass trajectories most similar to those when the system was managed so that biomass moves as rapidly as possible to its target reference point.

Keywords: Harvest strategies; data limited; risk equivalency

Corresponding author: Elizabeth Fulton

d. *Email*: Beth.Fulton@csiro.au
e. *Phone*: +61 3 6232 5018
f. *Fax* : +61 3 6232 5222

Highlights

1. Buffers help achieve risk equivalency and avoid overfishing stocks.
2. Buffer based on the uncertainty associated with the assessment are most effective.

Introduction

Over the last few decades, jurisdictions such as Australia, New Zealand, Canada, and the USA have implemented a process of managing target species using harvest strategies (Butterworth 2007; Smith et al., 2013): that is, a system of monitoring, assessment and harvest control rules that are used to determine management actions for a fishery. The need to make such recommendations for many stocks with differing levels of data availability has led to the development of tier systems in which species are categorized from data-rich to data-poor, with harvest strategies developed for each category of species. Tier systems are used in some USA federal fisheries, the Australian Southern and Eastern Scalefish and Shark Fishery Fishery (SESSF), and by the International Council for the Exploration of the Sea (ICES) (see review in Dichmont et al., in press).

Risk equivalency' is defined in the Australian Commonwealth Harvest Policy (HSP) as "ensur(ing) that the stock stays above the limit biomass level at least 90% of the time" (DAFF, 2007). The SESSF is the only Australian fishery that has formally placed harvest control rules (HCRs) and their associated assessment methods into tiers. The tiers arose due to the multi-species nature of the SESSF and the large number (34) of Total Allowable Catches (TACs) that are set within this fishery. Within each tier, a set of "buffers" or "discount factors" are used to attempt to equalize risk between tiers (Fay et al., 2012). These buffers are applied to the assessment-produced target catch or effort to account for uncertainty in the assessment and hence the Recommended Biological Catch (RBC). The intention is to reduce the final TAC determined from high-risk data-poor HCRs to reflect the bias and uncertainty associated with the assessment method and HCR being applied. Similarly, in the USA, the buffer between the Overfishing Limit and the Allowable Biological Catch reflects the extent of scientific uncertainty and differs between the tiers and species. The buffer is a function of the extent of assessment uncertainty and the risk tolerance given uncertainty (e.g. Prager and Shertzer, 2010; Punt et al., 2012; Shertzer et al., 2008).

The application of buffers is one means of trying to account for uncertainty associated with the HCR. An additional common tool used in Australia and internationally is Management Strategy Evaluation, MSE (Butterworth, 2007; Punt et al., 2016). The IWC (e.g., Punt and Donovan 2007), South Africa (e.g., Plagányi et al., 2007; De Oliveira and Butterworth 2004), Australia (e.g., Wayte and Klaer 2010; Dichmont and Brown 2010; Fay et al., 2011; Klaer et al., 2012), USA (e.g., Punt et al., 2012; Hurtado-Ferro and Punt, 2014) and ICES (e.g. ICES, 2013, 2014) have all used MSE to try and ensure that their HCRs are robust to model, assessment and implementation uncertainty. Dankel et al. (2015) go further, including uncertainty in the HCR itself.

Many of the HCRs and their associated assessment methods (i.e., harvest strategies) that define a tier for the SESSF have been tested using MSE. For example, MSE was used to evaluate several 'data-rich harvest strategies' for the eastern Australian gemfish stock, *Rexea solandri* (Punt and Smith, 1999). Results from that evaluation helped form the basis of the SESSF tier 1 HCR. MSE has also been used to evaluate an average-length-based HCR, defined as the SESSF tier 3 HCR, which performed well for demersal trawl species exhibiting reasonably high productivity (Klaer et al., 2012). Variants of the tier 3 HCR have also been compared/evaluated, which showed that appropriate values for the control parameters (of the HCR) were species-specific, and related to parameters such as the steepness of the stock-recruitment relationship and natural mortality (Fay et al., 2011). In addition, work in the SESSF has also showed that the performance of each HCR varies among stocks. However, to date no MSE analyses has included the currently implemented buffers (Fay et al., 2012; Little et al., 2014).

The first four tiers (or aspects of them) in the recently developed ICES tier (termed "categories") system were evaluated using MSEs (ICES, 2013, 2014; STECF, 2015), determining performance for alternative life histories and stock status (e.g., well managed or over exploited). The choice of buffer size for the USA tier system for the Bering Sea and Aleutian Islands crab stocks has also been evaluated, assuming a range of life histories and information content (Punt et al., 2012). The results were used by the North Pacific Fishery Management Council to establish default buffers for its more data-rich tiers.

The majority of these MSEs performed their evaluations in a single species context (with the exception of STECF [2015], which used Ecopath with Ecosim to provide some long-term perspectives in a multi-model comparison of the implications of alternative fishing mortality levels). In addition, most MSEs did not evaluate their tier system with candidate risk buffers (except Punt et al., 2012). Considering the performance of HCRs across a range of species life history types and within a multi-species or ecosystem context still remains relatively rare. In this paper, we use an ecosystem model that was modified for the SESSF to evaluate its four-tier system for a range of representative species, with the emphasis on evaluating the efficacy of existing buffer systems in the context of achieving risk equivalency. Analyses consider the SESSF buffers as well as a set of buffers inferred from how buffers are set for the USA west coast groundfish fishery. The effect of constraints on the extent of permitted inter-annual change in RBCs is also evaluated.

Methods

Operating model

At the core of an MSE is the operating model, which describes the dynamics of the system of interest. This is then sampled (in much the same way the real world is sampled) using a sampling model (detailed below).

The end-to-end ecosystem model, Atlantis for South Eastern Australia (Atlantis-SE; Fulton et al., 2014) formed the operating model for the MSE outlined here. It was modified (and henceforth referred to as Atlantis-RCC) to generate more realistic (smoother) size-composition data. Atlantis-RCC is a 3-D box model: regions (Fig. E.1) are based on the (i) physical and (ii) ecological properties, and (iii) distribution of the water bodies and geomorphology of south eastern Australia (summarised in IMCRA, 1998; Butler et al., 2001; Lyne and Hayes, 2005 and Fulton et al., 2007). The maximum modelled depth is 1800m (waters deeper than this are treated as an open boundary).

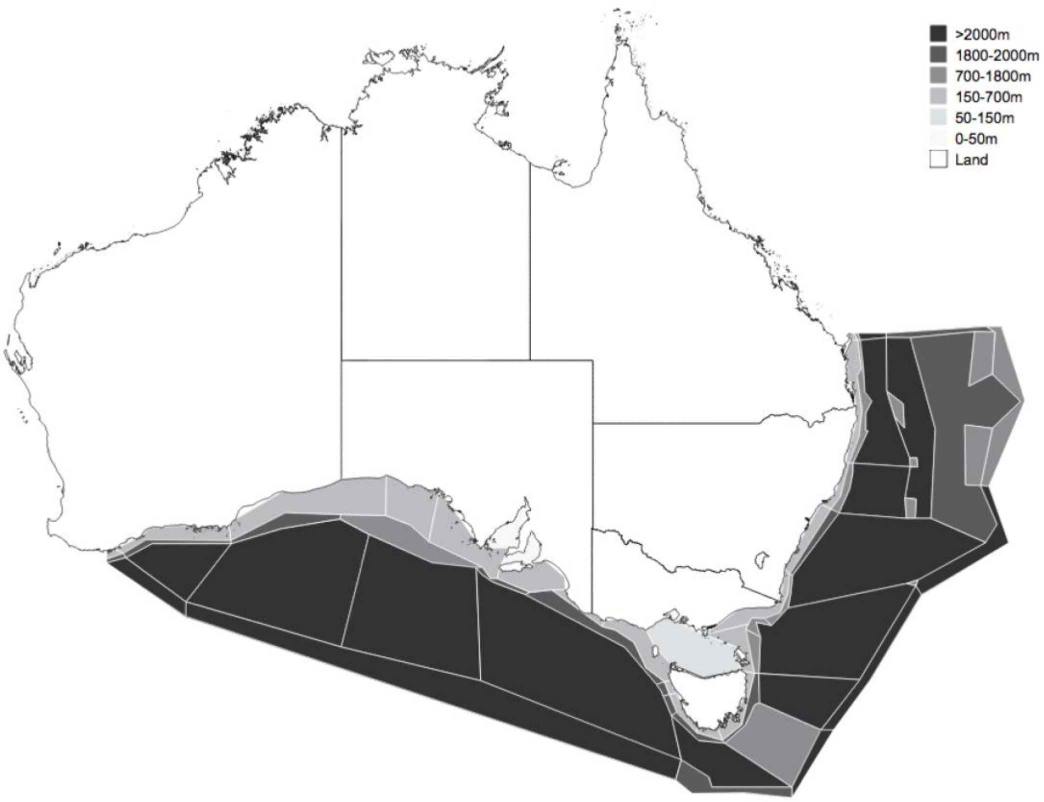


Figure E. 1 The domain of Atlantis-RCC, which contains 71 boxes with up to 5 vertical layers per box.

The physical environment for Atlantis-RCC included ocean currents and water column properties (e.g. temperature and salinity). Vertical and horizontal exchanges between boxes, as well as physical properties such as temperature and salinity, were taken from the data-assimilated version of the global ocean model OFAM (Oke et al., 2005).[3]

Atlantis-RCC includes the food web described by Fulton et al. (2007, 2014) (summarised in Table E.S1). It was initialised in 1980 and run under historical fishery catches and known environmental drivers until 2005, and pre-specified harvest strategies applied thereafter. To do this a new set of initial conditions was needed, as previous updating of Atlantis-SE and Atlantis-RCC had primarily been for the period post 2000. A two-step processes was undertaken to create the new initial conditions. For species with existing assessments, the levels of depletion identified by Morison et al. (2012) were used to infer the the 1980 biomass form the 2005 biomass estimates from the most recent (updated) version of Atlantis-SE (Fulton et al., 2007). For all other groups, century long term historical simulations (1910 – 2010) run using an earlier version of Atlantis-SE (Fulton et al., 2007) were used to calculate the relative (simulated) biomass in 1980 versus 2005. This scalar was then applied to the 2005 biomasses from the most recent (updated) version of Atlantis-SE (Fulton and Gorton, 2014) to get the 1980s biomasses to use with Atlantis-RCC.

One set of biological parameter values (e.g., values for non-predation mortality rates, physiological, consumption and growth rates, habitat preferences, movement rates) was used for a species (or functional group), unless the species (or functional group) was assumed to have multiple stocks – in which case fecundity, background mortality and diet connection strength varied among stocks (supplementary material, Table E.S1).

[3] The database used is available at http://www.bom.gov.au/bluelink/ and SPINUP6 from http://www.marine.csiro.au/ofam1/.

The single size-at-age for each vertebrate group in Atlantis-SE varies through time and among locations, based on available prey and resulting realised growth. Nevertheless, the tracking of "average individuals" typically used in Atlantis was very coarse compared to that of single-species stock assessments applied in Australia. As a result, it was causing problems when trying to apply tier 1 assessment methods to data simulated using Atlantis. Consequently, multiple growth "morphs" were used in Atlantis-RCC (c.f., Punt et al., 2001; Methot and Wetzel, 2013) for all main SESSF species (Table E.S1). Each morph followed a different growth trajectory so that there was variation in size-at-age within a cohort at each location. This is more realistic and more in line with previous evaluations of harvest strategies (e.g., Little et al., 2007).

Recruitment was based on a Beverton-Holt-style stock-recruitment relationship, except for sharks, seabirds and marine mammals, for which a linear function is used. Tiger flathead (*Neoplatycephalus richardsoni*), blue grenadier (*Macruronus novaezelandiae*) and silver warehou (*Seriolella punctata*) appear to exhibit cyclic recruitment dynamics (Day and Klaer, 2014; Tuck, 2014; Day et al., 2013). This behaviour was mimicked in Atlantis-RCC by applying deviations to the predictions from the stock-recruitment curve based on the outcomes from the single-species age-structured assessment models for those species. This cyclic variation was additional to any variation produced by trophodynamic interactions and environmental conditioning of, for example, movement, survival, reproduction, physiology. Cyclic variation for these groups was extended into the future by assuming that the pattern of deviations about the stock-recruitment relationship estimated for the past applied into the future.

Diet connections between groups identify potential food web pathways. Predation mortality is related to the maximum potential availability of each prey to each potential predator. The realised rate of predation is then conditioned on the level of physical contact (i.e., spatial overlap within a box given habitat preferences and patchiness), habitat (refugia) state and gape limitation (i.e., size of the mouth versus size of the prey given the feeding mode of the predator). Diet connections of Atlantis-SE were changed to include a smoother gape limitation function (supplementary material, Fig. E.S1), in line with the smoother size-at-age distributions, so that the size-composition data generated using the operating model were directly comparable with those used in assessments.

Atlantis-RCC was calibrated to historical biological and catch data for each functional group across all areas based on pattern-oriented modelling (Fulton et al., 2007; Kramer-Schadt et al., 2007). The data used to calibrate the model were constructed from available observational data (Fulton et al., 2007, 2014) and reported fisheries statistics. During calibration, parameters were adjusted to achieve a stable ecosystem, under constant fishing pressure, within the range of reported biomass estimates for these functional groups. The calibration was constrained to the most sensitive parameters (previously identified through sensitivity and factor analyses by Pantus and Dennison, 2005 and Fulton et al., 2007) – typically diet availabilities, growth and consumption rates, background natural mortalities (especially for highest trophic levels), fecundity levels and steepness of the stock-recruitment curve. This tuning began with the most uncertain of these parameters, which were adjusted to ensure that: 1) the shape of the predicted time series of biomasses, age structure, realised diet composition, and catches reflected their observed time series in the majority of boxes; 2) observed catches and discards were sustained without any functional group becoming extinct; and 3) rate parameters (e.g. growth or consumption rates) were not adjusted beyond bounds reported in literature without expert advice.

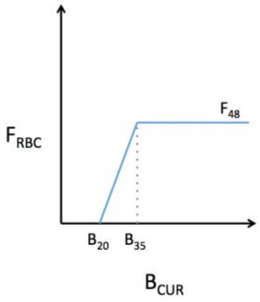
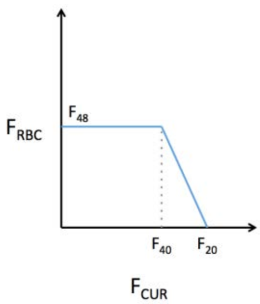
The same socio-economic effort allocation model used in Atlantis-SE was used in this study (Atlantis-RCC), which allows for multi-species targeting (Fulton et al., 2007). This socio-economic effort allocation model comprises two main components – a quota trading module and a fleet effort space-time dynamic module (Fulton et al., 2007). Together, these determine which suite of species is targeted, where and at what time of year. A hierarchical decision model is used to reflect how the decision making process is influenced on annual, monthly and trip scales (van Putten et al., 2013). The key detail for representing the reference HCR used in this model (described below) is that the trip-level effort allocation includes a decision based on available quota, costs and expected catch and discards associated with each potential fishing location (given historical catches, season, environmental conditions, market prices, species preference weightings and recent reported catch history or information sharing within the fleet(s)). The objective is to maximise returns while simultaneously minimising bycatch (all subject to available quota). Targeted fishing of a species can then be eliminated by setting quota to zero, which effectively the catch preference weights to zero and sees the initiation of the bycatch minimisation terms in the effort allocation decision steps.

Harvest strategies

The SESSF uses three harvest strategy tiers (tiers 1, 3, 4)[4] that reflect data availability and quality as well as the assessment method used and harvest control rule applied (Table E.1). Available data are analysed using a model-based stock assessment or a more empirical assessment (e.g., estimates of fishing mortality from a catch curve or a catch rate analyses). Resulting estimates of stock status are evaluated relative to target and limit reference point (TRP and LRP) proxies, to set a Recommended Biological Catch (RBC). The Total Allowable Commercial Catch (referred to as TAC here) is calculated from the RBC by applying meta-rules and risk equivalency buffers. In the SESSF, additional catch constraints, referred to as "meta-rules", are used for all tiers to prevent TACs from changing by more than 50% from one year to the next. The SESSF also uses buffers to reduce the RBC for tiers 3 and 4, to attempt to achieve risk equivalency between these tiers and the data-rich tier 1. A tier 1 RBC is usually the output of an integrated stock assessment, typically Stock Synthesis (Methot and Wetzel, 2013). The buffers applied to all stocks managed in the SESSF are equal to 1.0 for tier 1 (no buffer), 0.95 for tier 3 and 0.85 for tier 4 (Stobutzki et al., 2011). These were set semi-arbitrarily by the SESSF with no formal evaluation.

[4] Tier 2 is based on fitting population models that are more uncertain than tier 1 assessments, but this tier is not currently used for any stock. Moreover, how a stock would be assigned to this tier has never been specified. Tier 2 is thus omitted from this paper.

Table E. 1 The SESSF harvest control rules by tier.

Tier	HCR Graph	Rule
1		A full quantitative assessment provides estimates of spawning biomass (B) and depletion, which are used in a B_{20}:B_{35}:B_{48} (0.2B_0:0.35B_0:0.48B_0) "broken stick" HCR to select the target fishing mortality (F_{TARG}). This F_{TARG} (F_{48} is the F that achieves 0.48B_0) is then applied to the available biomass to calculate the RBC. For the purposes of the paper, the assessment was based on Stock Synthesis (Methot and Wetzel, 2013). Note that the reference points of 0.2B_0, 0.35B_0 and 0.48B_0 were chosen by the SESSF in compliance with Australia's harvest strategy policy.
3		Catch curves are used to estimate F_{CUR}, and F_{20}, F_{40} and F_{48} are taken from the relationship between yield and fishing mortality. F_{RBC} is then determined from the HCR, and the RBC is calculated using the equation: $$RBC = \max\left(\frac{1-e^{-F_{RBC}}}{1-e^{-F_{CUR}}}, 3\right) C_{CUR}$$ where C_{CUR} is the current catch. For the purposes of this paper, the approaches outlined in Wayte and Klaer (2010) are used to estimate F_{CUR} and the fishing mortality reference points.
4		The RBC from the Tier 4 HCR is given by: $$RBC = C_T \max\left(\frac{\overline{CPUE}-CPUE_L}{CPUE_T-CPUE_L}, 0\right)$$ where C_T is a catch target, $CPUE_L$ is the limit CPUE, $\overline{CPUE}$ is the average CPUE over the most recent four years, and $CPUE_T$ is the target CPUE. Both catch and CPUE targets were taken as the average for the simulated years 1996-2005 ("untuned" case). In reality the time period used is species-specific (often 1986-1996) and so "tuned" cases of this HCR were also tested (set to 1986-1996; see main text for further details).

One system that has a more transparent approach to determining buffers is the USA west coast groundfish fishery, which calculates a buffer between the Overfishing Level and the Acceptable Biological Catch, accounting for assessment uncertainty. The basis for the buffer for data–rich stocks is given in Ralston et al. (2011). Briefly, Ralston et al (2011) calculated the standard deviation in log-space (σ) of the spawner biomass estimates from a meta-analysis of groundfish stock assessments for the USA west coast, and this σ is used in combination with a stated tolerance for the risk of fishing mortality exceeding the target levels to determine the buffer.

Ralston et al. (2011) explored three methods for calculating the σ, but found that the most effective approach was to pool residuals from all species across all available assessments (assuming equal weighting). This method involves assessing variability among assessments for a stock in a given year (e.g., the estimated biomass in year x in the assessment done in 2010, 2011, 2012 and so on). This is done by taking the time series from the different assessments and using the squared deviations from the annual mean (over assessments) estimates of spawner log-biomass:

$$\overline{\ln[B_t]} = \frac{1}{n_t}\sum_i \ln[B_{i,t}] \qquad (1)$$

where n_t (≥2) is the number of available assessment estimates for the biomass of the stock in year t and $B_{i,t}$ is the biomass estimate of the stock from assessment i in year t. The measure of among-assessment variation is then given by σ:

$$\sigma = \sqrt{\frac{1}{\sum_t n_t - 1} \sum_t \sum_i (\ln[B_{i,t}] - \overline{\ln[B_t]})^2} \qquad (2)$$

The buffer is a cumulative percentile of a lognormal distribution with (log scale) mean of 0 and σ, as computed using the deviations from the mean method. We evaluated the same three methods from Ralston et al. (2011), and also found that the "deviations from the mean" method was most appropriate, so only results based on this method are reported here. The estimated σ for stocks in USA category 2 (generally less reliable model-based assessments), and USA category 3 stocks (stocks with no model-based assessment or very unreliable assessments) are respectively 2x and 4x the category 1 σ (PFMC, 2014). This method was applied to all available quantitative assessments for SESSF species, with the SESSF tiers aligned to the USA west coast categories by Dichmont et al (2016). However, unlike Ralston et al. (2011) who used the the pooled σ estimates for the USA stocks, we used the σ estimated for each stock separately (Table E.2), only using pooled values as a default if multiple assessments are actually not available for the species being considered in the simulations. The cumulative percentile was set to 0.4 for all tiers for consistency (this may be modified with consultation with the managers in the SESSF, as it was the case in Ralston et al. (2011) who chose to use 0.4 for their tiers 2 and 3, but a higher cumulative percentile of 0.45 for tier 1).

Table E. 2 Buffer system by species using the (inferred) US and Australian methods. The buffers for the USA system are based on percentiles of 0.4 for all tiers. Note that many species are split into east and west stocks for assessment purposes in the SESSF and that these stocks are list separately here.

Stock	Actual SESSF Tier	Assessment σ	USA system buffers Tier 1	Tier 3	Tier 4	Australian system buffers Tier 1	Tier 3	Tier 4
Blue grenadier, *Macruronus novaezelandiae*[1]	1	-	0.920	0.870	0.820	1	0.95	0.85
Cascade orange roughy, *Hoplostethus atlanticus*	1	0.26	0.930	0.860	0.800	1	0.95	0.85
Deepwater flathead, Neoplatycephalus conatus	1	0.25	0.940	0.870	0.790	1	0.95	0.85
Tiger flathead, Neoplatycephalus richardsoni	1	0.27	0.925	0.880	0.830	1	0.95	0.85
Eastern gemfish, *Rexea solandri*	1	0.24	0.940	0.890	0.800	1	0.95	0.85
Gummy shark, *Mustelus antarcticus**	1	-	0.920	0.870	0.820	1	0.95	0.85
Eastern jackass morwong, *Nemadactylus macropterus*	1	0.10	0.970	0.950	0.900	1	0.95	0.85
East pink ling, *Genypterus blacodes*	1	0.14	0.965	0.930	0.860	1	0.95	0.85
West pink ling, *Genypterus blacodes*	1	0.38	0.900	0.820	0.730	1	0.95	0.85
(School) Whiting, *Sillago spp*	1	0.30	0.915	0.850	0.770	1	0.95	0.85
Bight redfish, *Centroberyx gerrardi*	1	0.22	0.920	0.870	0.820	1	0.95	0.85
Eastern blue warehou, *Seriolella brama*	4	0.13	0.965	0.935	0.875	1	0.95	0.85
Western blue warehou, *Seriolella brama*	4	0.16	0.960	0.920	0.850	1	0.95	0.85
Blue-eye trevalla, *Hyperoglyphe Antarctica**	4	-	0.920	0.870	0.820	1	0.95	0.85
Demersal sharks*	5†	-	0.920	0.870	0.820	1	0.95	0.85
Gulper shark, *Centrophorus* spp*	5†	-	0.920	0.870	0.820	1	0.95	0.85
Generic demersal fish*	5†	-	0.920	0.870	0.820	1	0.95	0.85
Average		0.22	0.92	0.87	0.82	1	0.95	0.85

* Insufficient assessments available to apply the Ralston et al (2011) method, so the overall average outcome was used (as recommended by Ralston et al (2011)).

† Tier 5 is not considered in this paper as it is only recently implemented, but as it is implemented in reality in the SESSF it is noted here for completeness. It uses a length-based assessment method and interested readers should consult Haddon et al. (2015).

Data generation

A sampling model is used to generate observer and other fishery-dependent data types for each stock and Atlantis region: (a) catch length- and age-compositions; (b) catch-per-unit-effort data (by vessel size-class and fishery sector); (c) landings data (and catch species composition) by vessel size-class and fishery sector, and (d) discard data. This data generation process allows for ageing error, measurement error, variation in catchability, and error when measuring discards, with error levels that are stock-specific (supplementary material, Table E.S2).

Data were generated for each 12 hour Atlantis time-step and aggregated to trip, month, and year. Annual data were used for assessment purposes. CPUE data for each vessel were aggregated for the tier 1 assessments based on the gear used (or season for Blue grenadier – spawning versus non-spawning), so that fleet-specific parameters (e.g., selectivity) could be estimated in the assessment model (as is done in reality). Data from the fleet that had the highest catch in the final five years of the historical period were used when applying the tier 3 and 4 HCRs. Alternative options were explored (e.g., pooling across all fleets, or using the fleet that had the highest catch in any five-year period), but this made little difference to the general results. Tier 4 can also be sensitive to the selection of the reference period used. While 1996 – 2005 is the default period (i.e., final 10 years of the historical period), for a number of species this has been changed based on the history of exploitation. 'Sliding window' scenarios (where analyses were conducted with different reference periods) were used to explore the influence of the reference period for those species with modified Tier 4 reference periods in the SESSF (Flathead, Blue grenadier, Blue warehou, Redfish, Pink ling and other generic (shelf) demersal fish). In all cases the reference period 1986-1996 was amongst the most conservative and so that period was used here for the "Tuned-Tier 4" cases.

Performance metrics

Analyses involved assessing and managing stocks (see Tables E.2 and E.S1) of a given species individually. However, for brevity's sake, results are summarised by combining results for all stocks of each species. The projections assume that assessments and hence changes to RBCs and TACs occur annually.

Nine performance metrics were considered: (a) probability of being below the limit reference point (i.e., 20% of the unfished spawner biomass, $0.2B_0$) during the final 30 years of a 45-year (i.e., 2006 to 2050) projection period, (b) probability of being below the target reference point ($0.48B_0$) during the last 30 years of the projection period, (c) discounted catch relative to that expected under a reference HCR (defined below) across the entire projection period, (d) inter-annual variation in catch (across the entire projection period), (e) average spawner biomass over the last 30 years of the projection period relative to the average spawner biomass under the reference HCR during the same period, and (f) time taken to recover to $0.2B_0$, if overfished, or time taken to achieve $0.48B_0$ if below $0.48B_0$ initially (Table E.3). The discount rate used when computing the discounted catch was assumed to be 0.05yr^{-1} to reflect opportunity costs of the catch relative to other investments (e.g., Punt et al., 2011). The last 30 years were selected when calculating performance metrics (a), (b) and (e) to enable most stocks depleted at the end of the historical period to recover before assessing comparative risk performance.

Table E. 3 Performance metrics

Metric	How calculated	Notes
Probability $B < 0.2B_0$	$\underset{s}{average}\left(N_{B<B_{20,s}}/30\right)$	$N_{B<B_{20,s}}$ is the number of years for which spawning biomass is less than $0.2B_0$ during the last 30 years of the projection period for simulation *s*.
Probability $B < 0.48B_0$	$\underset{s}{average}\left(N_{B<B_{48,s}}/30\right)$	$N_{B<B_{48,s}}$ is the number of years for which spawning biomass is less than $0.48B_0$ during the last 30 years of the projection period for simulation *s*.
Relative Discounted Catch	$\frac{\sum_s \sum_y C_{y,s} e^{-(y-2006)\delta}}{\sum_s \sum_y C_{b,y,s} e^{-(y-2006)\delta}}$	$C_{y,s}$ is the catch in projection year *y* of simulation *s* (for the entire projection period), $C_{b,y,s}$ is the catch during projection year *y* of simulation *s* under the bang-bang HCR, and δ is the discount rate (0.05).
Annual Absolute Variation in Catch	$\underset{s}{average}\ \frac{1}{45}\sum_{y}\frac{\lvert C_{y,s}-C_{y-1,s}\rvert}{C_{y-1,s}}$	Calculated over the entire projection period.
Relative biomass	$\underset{s}{average}\left(\sum_{y=2021}^{2050} B_{y,s} \Big/ \sum_{y=2021}^{2050} B_{b,y,s}\right)$	$B_{y,s}$ is the spawning biomass in year *y* of simulation *s*, and $B_{b,y,s}$ is the spawning biomass in year *y* of simulation *s* under the bang-bang HCR. Calculated for the final 30 years of the simulations.
Time to threshold	$\underset{s}{\text{median}}\left(T_{s,bx}\right)$	$T_{s,bx}$ is the time taken to reach the threshold b_x for the first time in the projection period*.

* There are four "time to threshold" indices: the time to increase to $0.2B_0$, the time to increase to $0.48B_0$, the time to decrease to $0.2B_0$, and the time to decrease to $0.48B_0$. Although only a subset of these indices will be meaningful for any one species group – for example the time to decrease is meaningless if the biomass is less than $0.2B_0$ at the start of a simulation. However, the four indices are reported so that it is possible to determine if (a) an overfished species recovers to $0.2B_0$ (or beyond), (b) a stock that is not fully exploited has its biomass drop to $0.48B_0$ (or lower), and (c) a stock that is initially in the range $0.2B_0 < B < 0.48B_0$ remains in that range.

The target reference point for Australian Commonwealth managed fisheries is B_{MEY}, the biomass corresponding to Maximum Economic Yield. The Australian Harvest Strategy Policy allows for the use of proxies for B_{MEY} ($1.2 \times B_{MSY}$), where the proxy for B_{MSY} is taken to be $0.4B_0$ (Rayns, 2007).

The stock status during the historical period (2005) in the simulations varied across species and dominated any results if statistics were made relative to any fixed temporal point. Consequently, a reference trajectory was needed that would reflect a "best possible" bound. Here, the reference (albeit unrealistic) HCR was a "bang-bang" HCR (Deroba and Bence, 2008) based on perfect information about stock size. This HCR assumes perfect knowledge of the stock size (i.e., biomass used in the HCR is taken directly from the operating model with no error added by the sampling model). The HCR eliminated targeted fishing for a species for

N_1 years if $B < 0.48B_0$, and allowed for large catches for N_2 years if $B > 0.48B_0$. N_1 and N_2 were selected for each species iteratively (starting with 1 year and incrementing by a single year per simulation until the fastest approach to $0.48B_0$ was found). An analytical determination was not possible because the use of the dynamic effort allocation model led to implementation error. Elimination of targeted fishing in Atlantis is achieved by modifying quota and weighting terms used in the effort allocation decision sub-model. Focus was on the elimination of targeted fishing because a complete reduction of fishing-related mortality on a species was not possible, owing to incidental catches. In addition, the effort allocation sub-model in Atlantis allows for non-compliance and a memory of past catch performance. Consequently, some catch began once stock rebuilding started and biomass exceeded approximately 0.25-0.3B_0. Atlantis also does not yet have the capacity to mimic the fine-scale targeting fishers achieve in reality, so attempting to enforce a strict closure leads to the small landed catch under a targeting ban rather than a zero take. Consequently, a small "bycatch" take occurred when targeted fishing was eliminated. It follows that recovery for the reference harvest strategy is consequently (slightly) slowed.

The performance metric for discounted catch does not include the discounted sum of the catch in the last year for an infinite number of years. This is primarily because the populations do not all stabilise even after the harvest strategies have been applied for 45 years, particularly for tiers 3 and 4. Species such as Redfish and Blue grenadier showed high inter-annual variation, and a number of other species (e.g., cascade orange roughy) has still not recovered after 45 years. A small number of test simulations run for longer time periods (>100 years) indicated that trying to run the model for longer periods to achieve "equilibrium" (as is the standard economic method) was not a viable solution as it was computationally expensive and often did not lead to increased stability due to ongoing environmentally- or trophodynamically-driven variation.

Simulation experiments

Simulations were run to determine RBCs for each of the 14 species in Table E.2, for each of the tier 1, 3 and 4 harvest strategies (irrespective of their actual designated tier), for six options related to meta-rules and buffers. Each of the species was considered individually, in that the treatment species had their TACs set using the tier-meta-rule-buffer combinations, with the management rules and TACs for the remaining species were kept at their 2005 levels (i.e., only one species at a time was actively managed; simulations exploring the outcomes when multiple species were simultaneously actively managed will be reported elsewhere).

The six management options considered were: (a) no meta-rules or buffers (NM-NB), (b) with meta-rules, but no buffers (M-NB), (c) with meta-rules and the Australian buffers (M-AUB), (e) no meta-rules and the Australian buffers (NM-AUB), (e) with meta-rules and the inferred USA buffers (M-USB), and (f) no meta-rules and the inferred USA buffers (NM-USB).

Twenty simulations were run for each scenario and species, with each simulation including a 45-year projection period. The Atlantis ecosystem sub-model is deterministic. However, multiple simulations were still required as parts of the effort allocation model are stochastic, and there was additional stochasticity among the projections as a consequence of the sampling error associated with the data used to apply harvest strategies.

Results and Discussion

Variability among assessments for SESSF stocks

The calculated σ values for the SESSF species varied between 0.1 and 0.38, with an average value of 0.22 and 0.24 if all species and stocks were aggregated (Table E.2). These values are smaller than those for USA west coast stocks (cf., average value 0.337 and aggregate value 0.358; Ralston et al., 2011).

Comparing tiers and associated rules: risk, catch, and catch variation

The bulk of the species groups in Table E.2 were below the target spawner biomass of $0.48B_0$ in 2006, although some groups (such as Blue grenadier, redfish and Blue-eye trevalla) were well above $0.48B_0$ at this time (Fig. E.2). Catches of the groups that are below their biomass target levels are reduced substantially under the bang-bang HCR, while the catches of other species groups are increased substantially for a few years to drive the spawner biomass of the group downwards towards $0.48B_0$ (Fig. E.3). The reduction in catch for some groups (e.g., flathead) are for only a few years because these groups are productive and not depleted far below the target level. In contrast, catches of groups such as Blue warehou, and particularly gemfish and gulper sharks, remain low for most of the 45-year projection period. All of the groups (except for gulper sharks) are at, or close to, $0.48B_0$ by the end of the projection period. Gulper sharks fail to recover (Fig. E.2), even though catches are reduced to very low levels (<50t annually). This is attributable to the impact of the fleet dynamics model that implies that fishing mortality is imposed on some groups (e.g., Gemfish, Blue warehou, Morwong, Pink ling and Gulper sharks) even when there is no targeted fishery for them. In the case of gulper sharks, incidental catches are sufficient (in combination with time-varying predation mortality) to prevent this group from fully recovering even over 45 years. Similarly, incidental catches slow the recovery of other species – such as gemfish, morwong and Blue warehou.

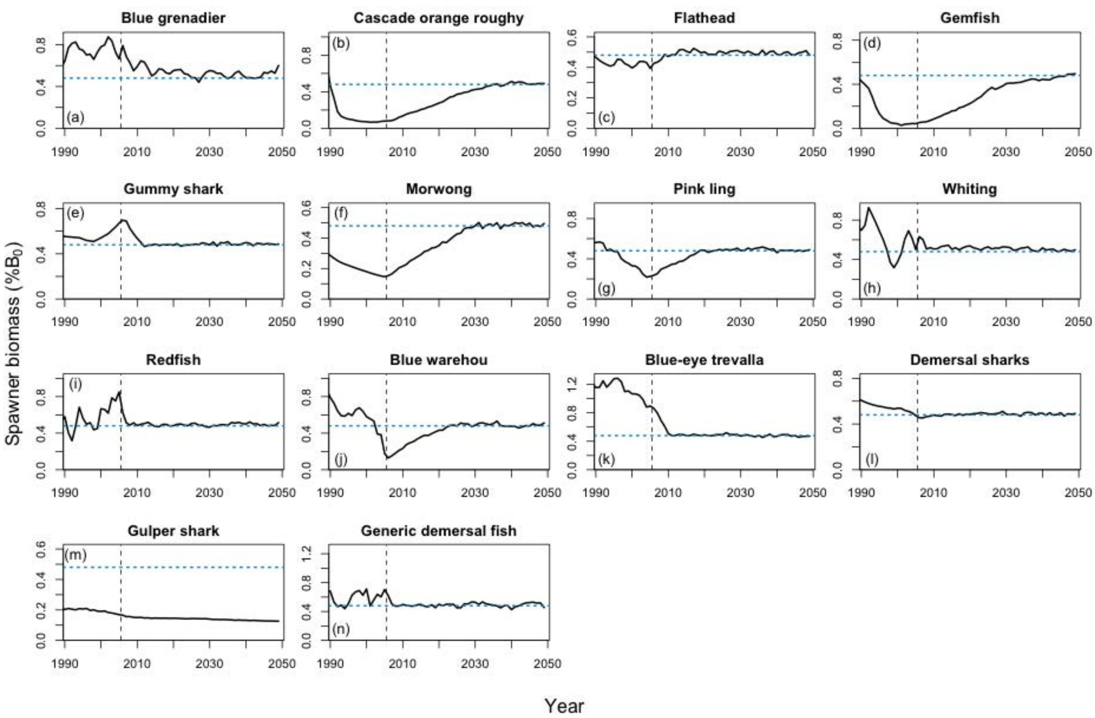


Figure E. 2 Average (over simulations) time-trajectories of spawner biomass by species / species-group. Catches after 2005 are based on the "bang-bang" HCR. The horizontal line denotes the target reference point. The vertical line marks the end of the historical period.

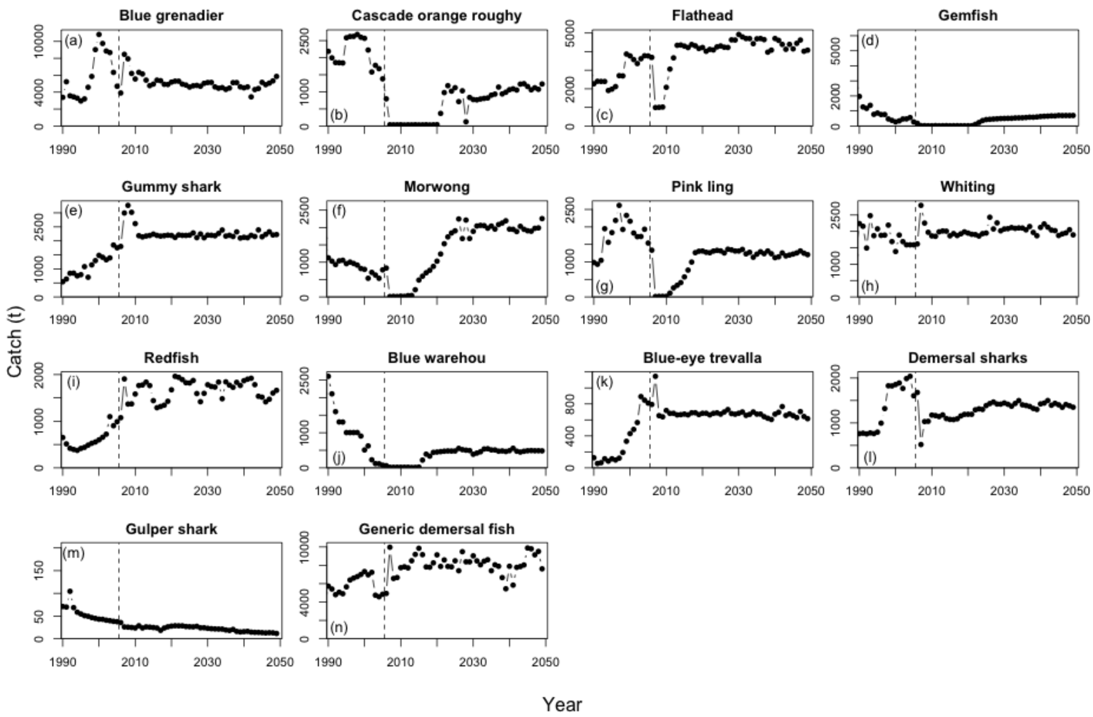


Figure E. 3 Average (over simulations) time-trajectories of catch by species / species-group. Catches after 2005 are based on the "bang-bang" HCR. The vertical line denotes the start of the projection period (2006).

The median (across species groups) probability of the spawner biomass being below $0.2B_0$ over the last 30 years of the projection period is zero (Figs. E.4a, b) for all options related to meta-rules and buffers, which is consistent with the requirement under the Australian Commonwealth Harvest Strategy Policy that dropping below this limit reference point should occur at most 10% of the time. However, at the species group level there were cases where this risk was exceeded, and there was considerable among-species variation in the probability of dropping below the limit reference point, with this probability being 100% for two species (gemfish and gulper shark) for all tiers, irrespective of whether meta-rules or buffers were implemented.

While the probability of being below the limit reference point is typically quite low over the medium- to long-term. The risk is higher in the short term (e.g., 15-20 years into the projection period, Fig. E.S2), where the probability of being below the limit reference point can be as high as 0.6. Stocks managed using tier 3 HCR have the highest potential of being below the limit reference point in that short-term time frame (supplementary material, Fig. E.S2b), irrespective of whether or not meta-rules are used. Tier 4 can also lead to several stocks having a probability in excess of 0.1 of being below the limit reference point in the short term. In more than three quarters of the tier-buffer combinations, the use of meta-rules did not reduce the probability of being below the limit reference point. Only the use of USA style buffers consistently reduced the short term risk of dropping below the limit reference points - particularly for tiers 3 and 4; it had less of an impact on the performance of tier 1.

The difference in performance of the tiers according to the presence or absence of meta-rules is also clear when examining the probability that a group is below the target reference point (Figs 4c,d). It may be expected that a stock being maintained at its target reference point would fluctuate slightly above and below its target reference point, so the probability of being below $0.48B_0$ should be roughly 0.5. Tier 1 is the only tier consistently approaching this goal (tending to higher biomasses than the target reference point if meta-rules are applied). Tier 4 can more readily achieve the goal in the medium to long term if buffers are used (it typically fails in the short-term, Fig E.S2c,d); although when using tuned reference periods tier 4 can overshoot (i.e., see biomass exceed $0.48B_0$) when buffers are also applied. Tier 3 only reliably achieves the target goal in the long term when USA style buffers are employed.

The medium- to long-term consequences of applying the tiers can be understood from the relative spawner biomass over the final five years of the projection period (Fig. E.5) and the time it takes for stocks to reach the limit or target reference points (Fig. E.6). Where there is any pattern to the long-term relative spawner biomass for a species, tier 1 almost always led to the highest median values, followed by tier 4 and then tier 3; with the exception of redfish where the tuned tier 4 results tend to lead to the highest relative spawner biomass. The implementation of meta-rules did not always lead to higher relative spawner biomass across all tiers. Rather, the impact was group-specific: higher relative biomass for groups such as Blue grenadier, and Whiting (Figs E.5a, h), but lower values for other groups including Pink ling, Redfish and Blue warehou (Figs E.5g, i, j). By damping inter-annual variation, it is likely that meta-rules would improve performance for species such as Whiting that are fast growing and short-lived (Day, 2010), as well as species with episodic recruitment such as Blue grenadier (Tuck et al., 2014). The strong reduction in Blue warehou biomass as a result of the meta-rule is potentially due to multispecies fishing effects (it is hard to avoid when fishing other species) and trophic effects, as Blue warehou is susceptible to high levels of variability (spatially and temporally) in predation mortality and competition (e.g., with Silver warehou) in Atlantis. Consequently, meta-rules constrained reductions in TACs in response to reductions in biomass beyond those anticipated from direct fishing effects. The impact of meta-rules on relative biomass for the remaining species depended on tier, or was minimal.

In contrast to the situation for the meta-rules, the application of any buffer with a value less than one led to higher median relative spawner biomass. The more conservative USA buffer system resulted in biomasses in each tier that were closest in value (typically within 5% of each other, as opposed to a >10% difference with the Australian buffers and potentially 20% or more with no buffers). Consequently, while the performance metrics in Fig. E.4 did not show perfect risk equivalency across all tiers for any buffer-meta-rule combination, the true state of the stock was close to equivalent when the buffers inferred from the USA west coast groundfish fishery were applied (especially with no meta-rules in place in the long term, although in the short term the risk equivalency was closest when meta-rules were in place; giving a sense of the complexity of the temporal dynamics).

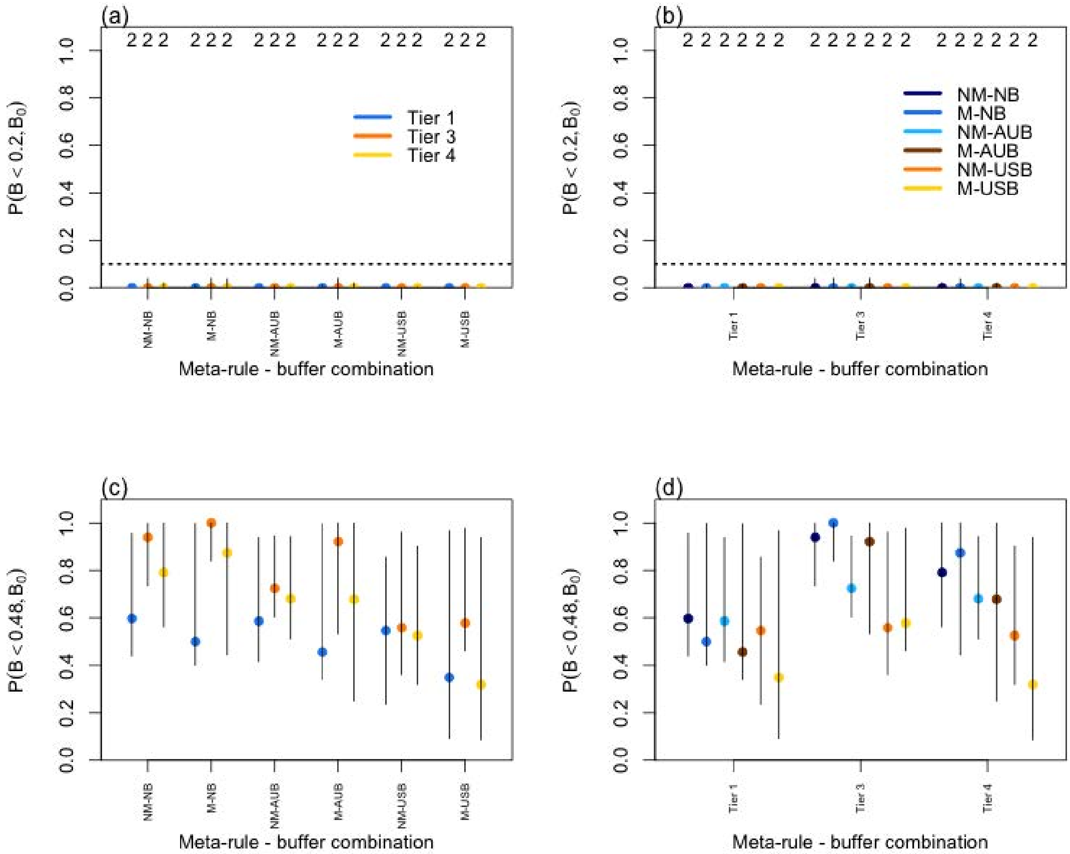


Figure E. 4 Upper panel: median over groups (and inter-quartile range over groups) probability of being below the limit reference point proxy of 0.2B_0 during the last 30 years of the projection period. Numbers across the top of each upper panel indicate the number of species for which the probability was 1.0. Lower panel: median over species groups (and inter-quartile) probability of being below the target reference point proxy of 0.48B_0 during the last 30 years of the projection period. Results are presented by tier for each meta rule – buffer combination (left) and by meta rule – buffer combination for each tier (right).

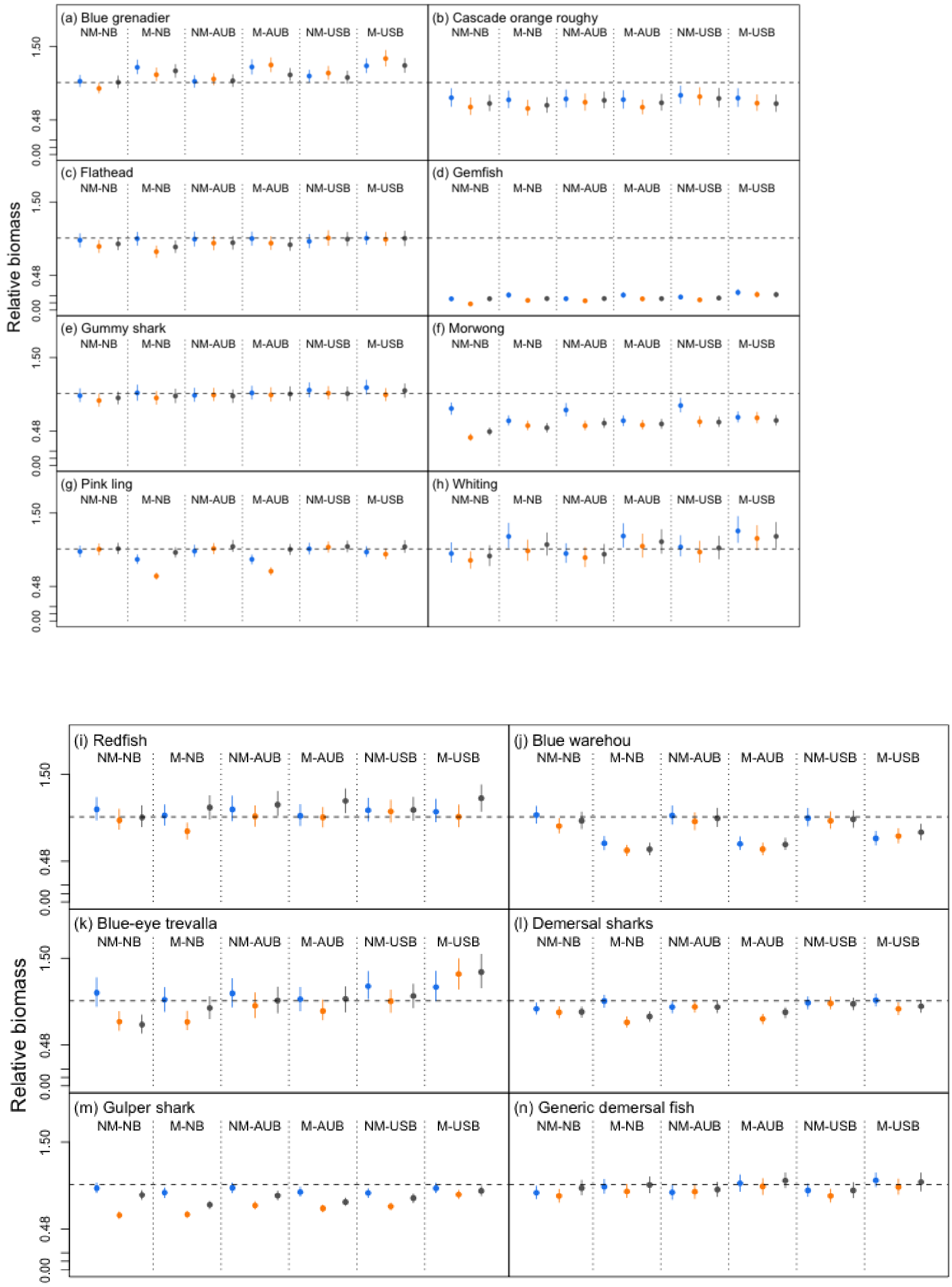


Figure E. 5 Average, minimum and maximum (over simulations) spawner biomass relative to that under the bang-bang HCR during the last 30 years of the projection period. The horizontal line indicates parity and the colours indicate tiers – tier 1: blue, tier 3; orange and tier 4: grey.

Fig. E.6 shows that the majority of species pass through either the target or the limit reference point during the projection period, for at least one of the tiers. The failure of gulper sharks and gemfish to recover is clear in Fig. E.6 (they fail to reach the limit reference point before the end of the projection period). There was substantial overlap across tiers in results for the rest of the groups. However, tier 1 is often the "fastest route" to rebuild a depleted stock to the limit reference point. It is not as rapid as the bang-bang HCR, but is often only a few years slower. When using tuned reference periods Tier 4 (which is then typically highly conservative) has the potential to outperform tier 1 in terms of speed of recovery (approaching the bang-bang HCR for some species), but this is not the case for the untuned form of tier 4, underlining the potential sensitivity of application of this tier. Tier 4 can lead to a rapid fish down of lightly exploited stocks, but even so it also allows for a more rapid recovery of depleted stocks than tier 3. Tier 3 often did not perform as well as either tier 1 or 4, because it can allow excessive fishing, especially initially. Its performance is also the most variable across species and life history types of all the tiers. Applying meta-rules and either of the sets of buffers (Figs E.6b,c) leads to faster rebuilding times for groups below the limit reference point and to fewer instances of groups being depleted to below $0.48B_0$ or $0.2B_0$. The results for the USA buffer system are noteworthy in that no stocks that were initially above $0.2B_0$ were depleted to be $0.2B_0$ or lower (Fig. E.4c)

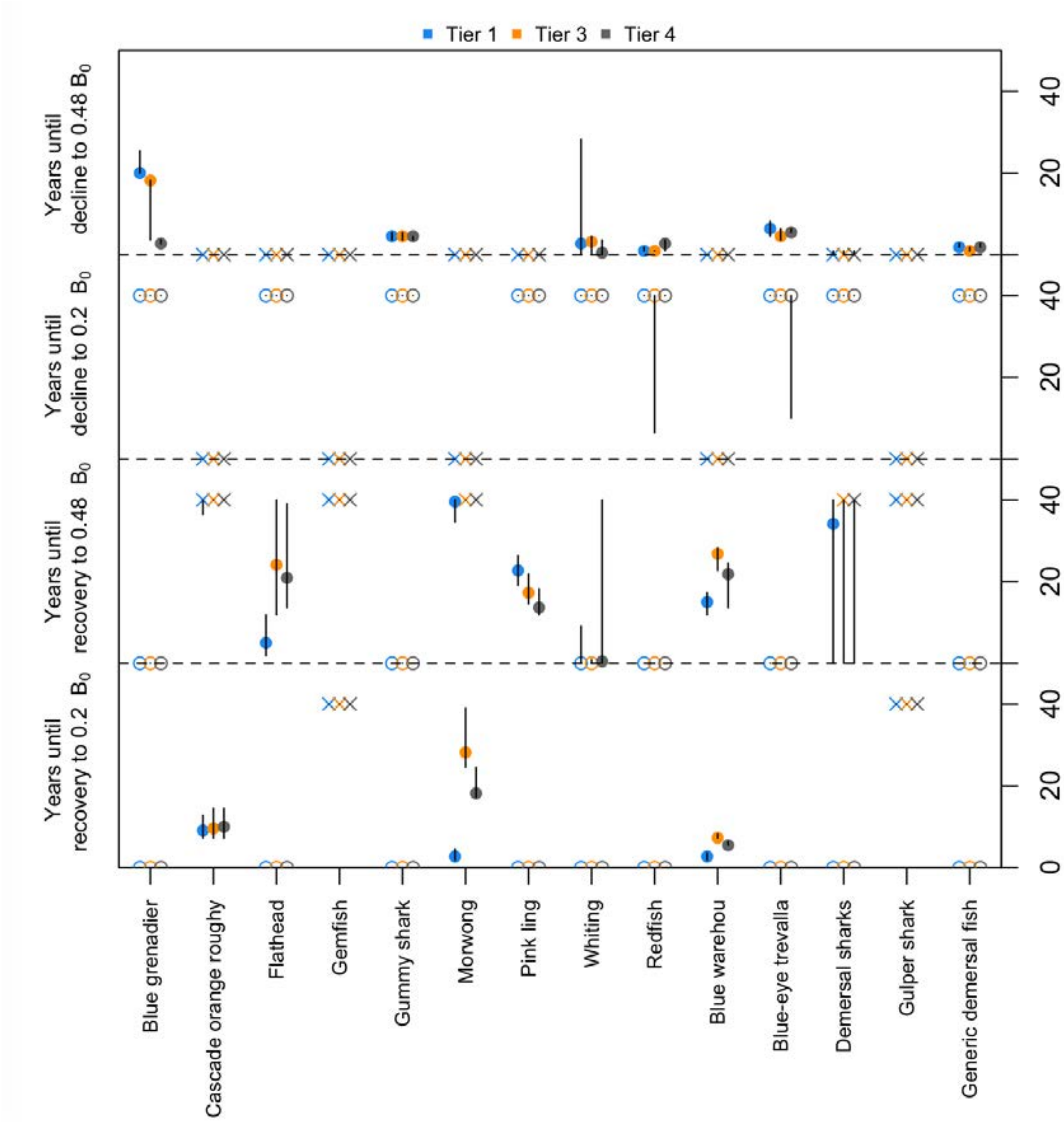


Figure E.6a. No buffers; no meta-rules (NM-NB)

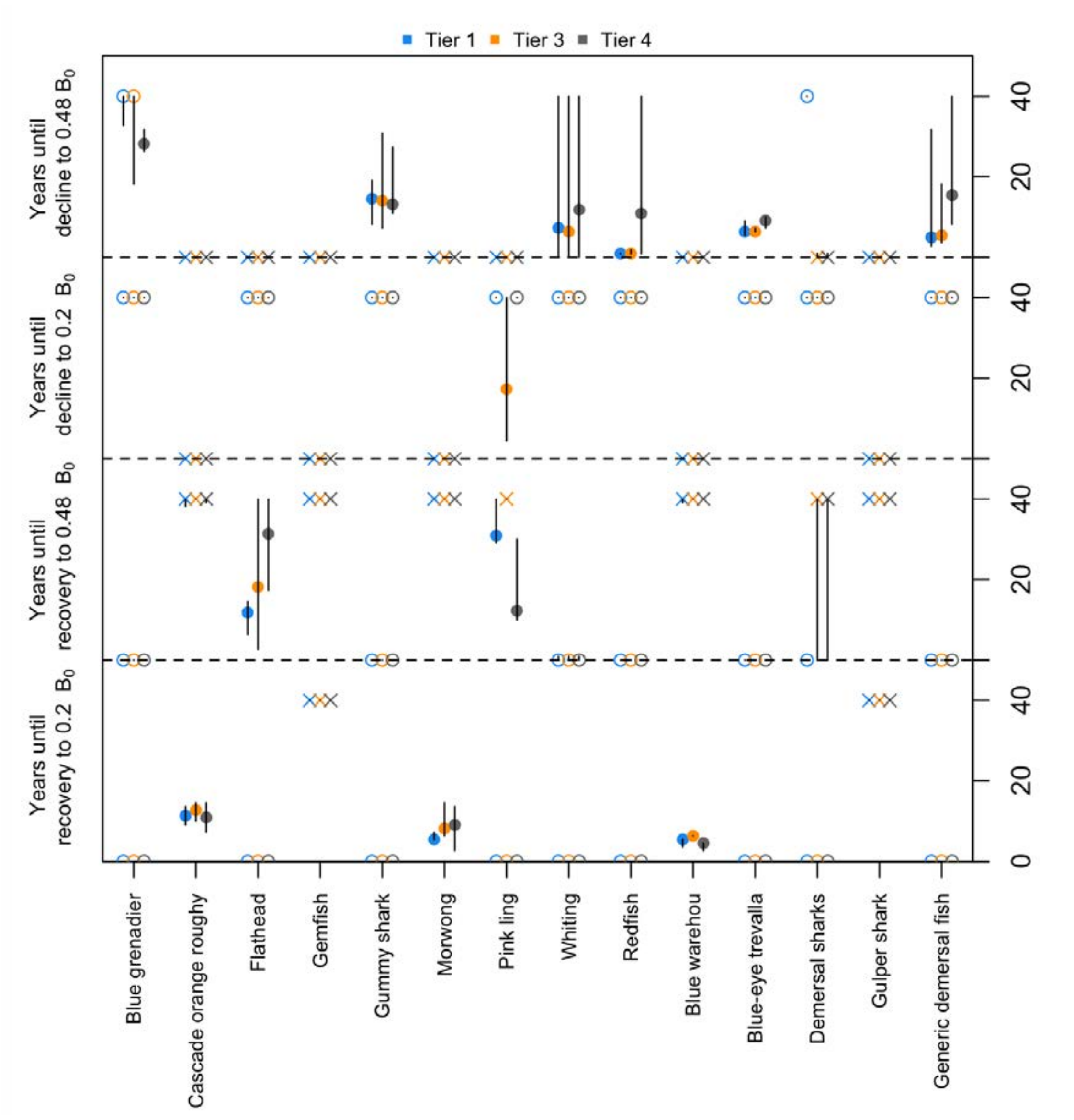


Figure E.6b. With meta-rules and Australian buffers (M-AUB)

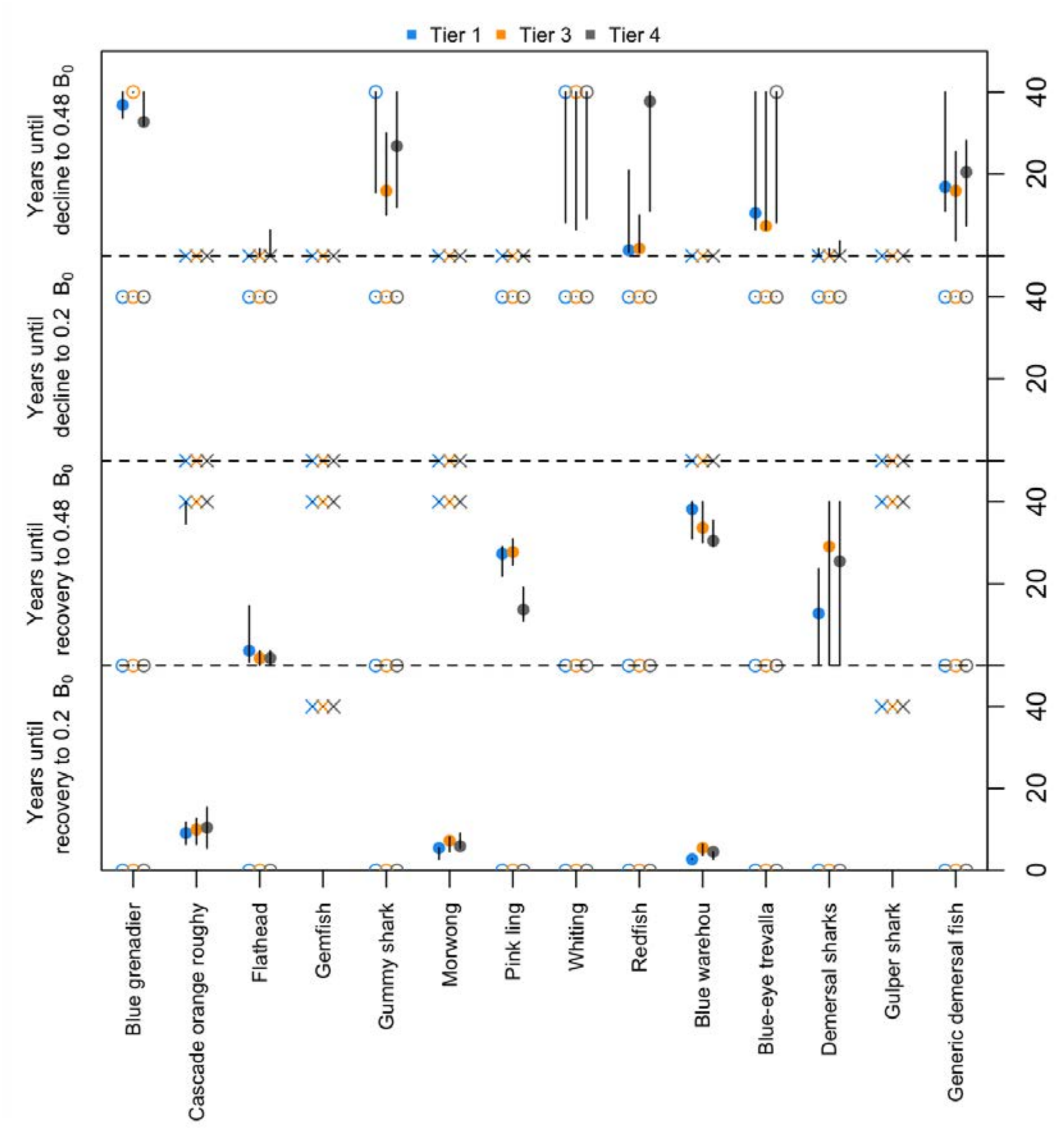


Figure E.6c. With meta-rules and USA buffers (M-USB)

Figure E.6. Number of years for a group initially above a reference point threshold to breach that threshold (upper two rows: thresholds of 0.48B0 and 0.2B0), and the number of years for a group initially below a threshold to recover to the threshold (lower two rows: thresholds of 0.2B0 and 0.48B0). A solid circle indicates the median time with the line indicating the minimum to maximum (over simulations) temporal spread. An open circle indicates that the group began above that threshold and never crossed it, while an "X" indicates that the group began below the threshold and never reached it. Thus, for the upper two panels an "X" indicates the test is irrelevant (as the group was initially below the threshold) and for the lower two panels an open circle indicates the test is irrelevant (as the group was initially above the threshold). The results of greatest concern are those in the lower panel with an "X" at 45, indicating that the group did not recover to 0.2B0 at any time during the projection period. Results are shown in (a) for "No buffers; no meta-rules" (NM-NB), in (b) for "With meta-rules and Australian buffers" (M-AUB); and in (c) for "With meta-rules and USA buffers" (M-USB).

Relative discounted catch indicates that the risks associated with the tier 3, for instance, do result in lower relative cumulative discounted catch (Figs E.7a,b), with initially high catches reduced later during the projection due to the poorer stock status than achieved under tier 1.

Similarly, the untuned tier 4 can lead to lower relative discounted catches long term, although when tuned its conservative nature can also lead to lower catches achieved than under the other tiers. The lower risk for tier 1 came at the cost of reduced relative catch, with catches more often than not remaining at or below those associated with the bang-bang HCR in the short term but growing to approach those of the bang-bang HCR in the long term. The use of meta-rules leads to slightly higher median relative catches for all tiers, especially for tier 3 (although this distinction is much less pronounced in the short term for tier 3, where the influence of meta-rules is effectively negligible).

The use of meta-rules restricted catch variability by preventing large changes in the RBC (Figs E.7c,d). Annual variation in catches is less than 80% of that with no meta-rules: catch variation for tier 1 is consistently more than 25% lower, while that for tier 3 could be as much as 51% lower (or more in some extreme cases). However, the restrictions on catch variation reduced the rate at which catches could be reduced for stocks (such as morwong) that were initially depleted and in need of rebuilding, particularly for tier 1, which leads to the need for large reductions in TAC. However, the meta-rules can also slow prematurely large increases in catches, such as can be recommended under tier 3, facilitating rebuilding (for morwong relative spawner biomass can be 40% higher with meta-rules than without).

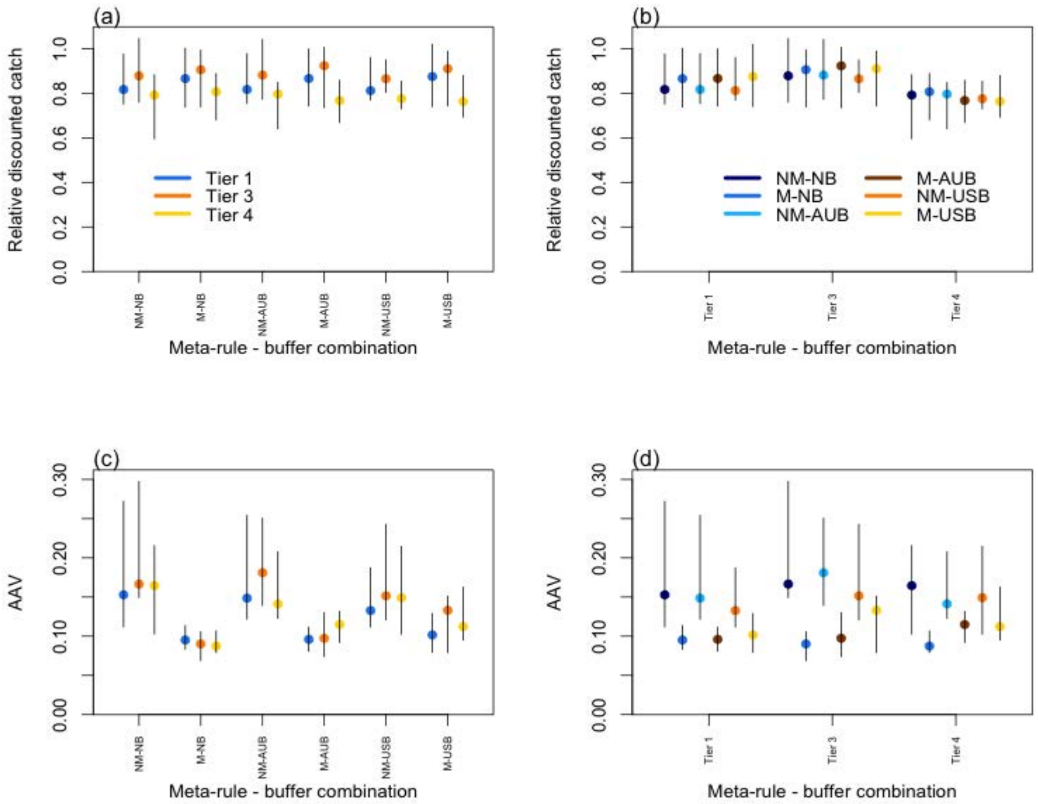


Figure E. 7 Number of years for a group initially above a reference point threshold to breach that threshold (upper two rows: thresholds of 0.48B_0 and 0.2B_0), and the number of years for a group initially below a threshold to recover to the threshold (lower two rows: thresholds of 0.2B_0 and 0.48B_0). A solid circle indicates the median time with the line indicating the minimum to maximum (over simulations) temporal spread. An open circle indicates that the group began above that threshold and never crossed it, while an "X" indicates that the group began below the threshold and never reached it. Thus, for the upper two panels an "X" indicates the test is irrelevant (as the group was initially below the threshold) and for the lower two panels an open circle indicates the test is irrelevant (as the group was initially above the threshold). The results of greatest concern are those in the lower panel with an "X" at 45, indicating that the group did not recover to 0.2B_0 at any time during the projection period. Results are shown in (a) for "No buffers; no meta-rules" (NM-NB), in (b) for "With meta-rules and Australian buffers" (M-AUB); and in (c) for "With meta-rules and USA buffers" (M-USB).

Caveats

The inability to close the fishery and the lack of recovery for gemfish and gulper sharks (Fig. E.6) highlight that harvest strategies that perform adequately in single-species situations can perform poorly in an ecosystem and socioecological context. The confidence in the robustness of our results is that they are largely in agreement with those from single species MSEs for the SESSF (Wayte and Klaer, 2010; Fay et al., 2011; Little et al., 2011; Klaer et al., 2012). Nevertheless, there are limitations to our analysis. The ecological components of Atlantis are deterministic and the effort allocation sub-model only includes limited stochasticity. Moreover, as an ecosystem model, Atlantis is highly uncertain and usually has multiple plausible parameterisations. The technical difficulty in applying the multiple size-at-age trajectories for some groups and the smooth feeding window implementation of the model meant that it was not possible to find multiple suitable parameterisations and, as a result, this work was undertaken with a single parameterisation (the one that best fits available data). As such, the results should be treated with some caution, as there is the possibility that an alternative parameterisation could lead to somewhat different outcomes. Undertaking a sensitivity analysis and repeating the work using other model frameworks (either multi-species or other ecosystem modelling platforms) would increase confidence in the robustness of the results. Multi-model comparisons have previously proven fruitful in the exploration of HCR for STECF (2015) and have been used to good effect in exploring the implications of the depletion of forage fish (Smith et al., 2011) and for exploring alternative fisheries and management options (e.g. Fulton and Smith 2014; Forrest et al., 2015; Smith et al., 2015; Jacobsen et al., 2016).

Final remarks and conclusions

There is a need for buffers if the aim of the management system is to prevent stocks from being depleted to below the limit reference point. However, the analyses suggest that basing the size of the buffer on the uncertainty associated with the assessment (the USA west coast buffer approach) rather than on the assessment method applied (the Australian system) is more effective. The improved performance of the USA west coast system in terms of avoiding risk may, however, be more related to that the magnitude of the buffers (Table E.2) (i.e., the HCRs are typically more conservative). Thus, the performance of the Australian approach may have matched that of the USA west coast approach if the buffer values were smaller (i.e., their ability to reduce the RBC was greater).

All systems led to low probabilities of stocks being below the limit reference points (except for gemfish and gulper shark for which rebuilding was essentially impossible in 45 years). Overall, however, none of the systems achieved complete risk equivalency. Although the USA west coast approach to setting buffers came close, appearing to be most able to achieve risk equivalency in relation to the probability of having half of the stocks above the target reference point across all tiers. This result also highlights that ‘risk equivalency’ relies on a definition of ‘risk’. That is, achieving risk equivalency in terms of one performance metric (e.g., the LRP) will not necessarily lead to such achievement for other performance metrics (e.g., the TRP) (see Figs E.4 and S2).

Constraints on catch variation (i.e. meta-rules) lead, as expected, to less variation in catch, even though the catch variation constraints are not particularly strict compared to those applied in other jurisdictions, such as South Africa (e.g., Plaganyi et al., 2007). The meta-rules will tend to lower the rate at which catches are reduced when stocks are in need of rebuilding, but will also reduce the instances of unrealistically large increases in catch. Furthermore, highly variable assessments (as a result of species variability or poor data quality) also lead to a larger effect of having meta-rules in place, as they damp the resulting variability in RBCs. Thus, the results of this study suggest that the benefits of meta-rules are likely case-specific.

The tier 1 harvest control rule outperformed the more data-poor harvest control rules in terms of allowing stocks to rebuild towards the limit and target reference points, albeit at a cost in terms of short-term catch (and variation in catch) – See supplementary material, Fig. E.S3. Whether the reduced risk is worthwhile given the costs associated with conducting full stock assessments is beyond the scope of this paper, but should be taken into account when monitoring systems are developed.

Acknowledgements

We would like to thank CSIRO internal reviewers for their comments on draft versions of the paper, as well as the various SESSF RAGs and AFMA members who provided input. This work was funded by Fisheries Research and Development Corporation Project 2012/202 and CSIRO Oceans and Atmosphere. In addition, we would like to thank the reviewers who provided many useful suggestions which have strengthened and clarified the manuscript.

References

Butler A., Harris P., Lyne V., Heap A., Parslow V., Porter-Smith R., 2001. An Interim Bioregionalisation for the continental slope and deeper waters of the South-East Marine Region of Australia. http://www.environment.gov.au/resource/interim-bioregionalisation-continental-slope-and-deeper-waters-south-east-marine-region (last accessed 1 September 2015).

Butterworth, D.S 2007. Why a management procedure approach? Some positives and negatives ICES J. Mar. Sci. 64, 613–617.

DAFF. 2007. *Commonwealth Fisheries Harvest Strategy Policy Guidelines*. Australian Government Department of Agriculture, Fisheries and Forestry, Canberra, Australia, pp. 55. http://www.agriculture.gov.au/fisheries/domestic/harvest_strategy_policy

Dankel, D.J., Vølstad, J.H., Aanes, S., 2015. Communicating uncertainty in quota advice: a case for confidence interval harvest control rules (CI-HCRs) for fisheries. Can. J. Fish. Aquat. Sci. 73: 1-9.

Day, J. 2010. School whiting (*Sillago flindersi*) stock assessment based on data up to 2008. In: G.N. Tuck [Ed.] *Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2009, Part 1.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart, pp 190-249.

Day, J., Klaer, N., 2014. Tiger flathead (*Neoplatycephalus richardsoni*) stock assessment based on data up to 2012. In: G.N. Tuck [Ed.] *Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2013, Part 1.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research. Chapter 10, pp 174–232.

Day, J., Klaer, N., Tuck, G.N., 2013. Silver Warehou (*seriolella punctata)* stockassessment based on data up to 20113. In: G.N. Tuck [Ed.] *Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2012, Part 1.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research. Chapter 7, pp 120–151.

Deroba, J.J., Bence, J.R., 2008. A review of harvest policies: Understanding relative performance of control rules. Fish. Res. 94, 210–223.

De Oliveira, J.A.A., Butterworth, D.S. 2004. Developing and refining a joint management procedure for the multispecies South African pelagic fishery. ICES J. Mar. Sci. 61: 1432–1442.

Dichmont, C.M., Brown, I.W. 2010. A case study in successful management of a data-poor fishery using simple decision rules: the Queensland spanner crab fishery. Marine and Coastal Fisheries: Dyn. Manage. Ecosys. Sci. 2: 1–13.

Dichmont, C.M., Punt, A.E., Dowling, N., De Oliveira, J.A.A., Little, L.R., Sporcic, M., Fulton, E., Gorton, R., Klaer, N., Haddon, M., Smith, D.C., 2015. Is risk consistent across tier-based harvest control rule management systems? A comparison of four case studies. Fish. And Fish. DOI: 10.1111/faf.12142

Fay, G., Punt, A.E., Smith, A.D.M., 2011. Impacts of spatial uncertainty on the performance of age structure-based harvest strategies for blue-eye trevalla (*Hyperoglyphe antarctica*). Fish. Res. 110, 391–407.

Fay, G., Little, L.R., Tuck, G.N., Haddon, M., Klaer, N. 2012. Maintaining risk equivalency among fishery harvest control rules in Southeast Australia. Report to the Australian Fisheries Management Authority.

Forrest, R.E., Savina, M., Fulton, E.A., Pitcher, T.J. 2015. Do marine ecosystem models give consistent policy evaluations? A comparison of Atlantis and Ecosim. Fish. Res. 167: 293–312

Fulton, E.A., Gorton, R., 2014. Adaptive Futures for SE Australian Fisheries & Aquaculture: Climate Adaptation Simulations. CSIRO Climate Adaptation Flagship, Australia.

Fulton, E.A., Smith, A.D.M., 2004. Lessons learnt from the comparison of three ecosystem models for Port Phillip Bay, Australia. African J. Mar. Sci. 26: 219 – 243

Fulton E.A., Smith, A.D.M., Smith, D.C., 2007. Alternative management strategies for Southeastern Australian Commonwealth Fisheries: Stage 2: Quantitative Management Strategy Evaluation. Report to the Australian Fisheries Management Authority and the Fisheries Research and Development Corporation. CSIRO Marine and Atmospheric Research.

Fulton, E.A., Smith, A.D.M., Smith, D.C., Johnson, P., 2014. An integrated approach is needed for Ecosystem Based Fisheries Management: Insights from ecosystem-level management strategy evaluation. *PLoS One*. 9, e84242.

ICES. 2013. Report of the Workshop on the Development of Quantitative Assessment Methodologies based on LIFE-history traits, exploitation characteristics, and other key parameters for Data-limited Stocks (WKLIFE III), 28 October–1 November 2013, Copenhagen, Denmark. ICES CM 2013/ACOM:35. 98 pp.

ICES. 2014. Report of the Workshop on the Development of Quantitative Assessment Methodologies based on LIFE-history traits, exploitation characteristics, and other relevant parameters for data-limited stocks (WKLIFE IV), 27–31 October 2014, Lisbon, Portugal. ICES CM 2014/ACOM:54. 223 pp.

Interim Marine and Coastal Regionalisation for Australia Technical Group (IMCRA) 1998. Interim Marine and Coastal Regionalisation for Australia: an ecosystem-based classification for marine and coastal environments. Version 3.3. Environment Australia, Commonwealth Department of the Environment: Canberra. http://www.environment.gov.au/resource/interim-marine-and-coastal-regionalisation-australia-version-33 (last accessed 1 September 2015).

Haddon, M., Klaer, N., Wayte, S., Tuck, G., 2015. Options for Tier 5 approaches in the SESSF and identification of when data support for harvest strategies are inappropriate. Report to Fisheries Research and Development Corporation. CSIRO Oceans and Atmosphere, Hobart. pp. 115.

Hurtado-Ferro, F., Punt, A.E. 2014. Revised Analyses Related to Pacific Sardine Harvest Parameters. Pacific Fishery Management Council, Portland, OR.

Jacobsen, N.S., Essington, T.E., Andersen, K.H. 2016. Comparing model predictions for ecosystem-based management. Can. J. Fish. Aquat. Sci. 73: 666-676, 10.1139/cjfas-2014-0561

Klaer, N.L., Wayte, S.E., Fay, G., 2012. An evaluation of the performance of a harvest strategy that uses an average-length-based assessment method. Fish. Res. 134-136, 42-51.

Kramer-Schadt S., Revilla E., Wiegand T., Grimm, V., 2007. Patterns for parameters in simulation models. Ecol. Model. 204: 553–556.

Little L.R., Punt, A.E., Mapstone, B.D., Pantus, F., Smith, A.D.M., Davies, C.R., McDonald, A.D., 2007. ELFSim – A Model for Evaluating Management Options for Spatially-Structured Reef Fish Populations: An Illustration of the "Larval Subsidy" Effect. Ecol. Model. 205, 381–396.

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 J. Mar. Sci. 68, 1699–1705.

Little, L.R., Parslow, J., Fay, G., Grafton, R.Q., Smith, A.D.M., Punt, A.E., Tuck, G.N. 2014. Environmental derivatives, risk analysis and conservation management. Cons. Lett. 7, 196-207

Lyne V., Hayes D., 2005. Pelagic Regionalisation. National Oceans Office and CSIRO Marine and Atmospheric Research, Hobart. http://www.environment.gov.au/system/files/resources/b7d7587a-6330-41fe-8f97-cba74ad87306/files/nmb-pelagic-report.pdf (last accessed 1 September 2015).

Methot, R.D., Wetzell, C.R., 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142, 86–99.

Morison, A.K., Knuckey, I.A., Simpfendorfer, C.A., Buckworth, R.C., 2012. 2011 Stock Assessment Summaries for the South East Scalefish and Shark Fishery. http://www.afma.gov.au/wp-content/uploads/2010/07/2011-SESSF-Species-Summaries-March-2012.pdf (last accessed 1 September 2015).

Oke P.R., Schiller A., Griffin D.A.B., 2005. Ensemble data assimilation for an eddy-resolving ocean model. Quart. J. Roy. Met. Soc. 131, 3301–3311.

Pacific Fishery Management Council (PFMC). 2014. Status of the Pacific Coast Groundfish Fishery: Stock Assessment and Fishery Evaluation. Pacific Fishery Management Council, 7700 Ambassador Place, Suite 101, Portland, OR 97220.

Pantus F.J., Dennison W.C., 2005. Quantifying and Evaluating Ecosystem Health: A Case Study from Moreton Bay, Australia. Enviro. Manage. 36, 757–771.

Plagányi, É.E., Rademeyer, R.A., Butterworth, D.S., Cunningham, C.L., Johnston, S.J., 2007. Making management procedures operational - innovations implemented in South Africa. ICES J. Mar. Sci. 64, 626–632.

Prager, M.H., Shertzer, K.W., 2010. Deriving Acceptable Biological Catch from the Overfishing Limit: Implications for Assessment Models. N. Am. J. Fish. Man. 30, 289–294.

Punt, A.E., Smith, A.D.M., 1999. Management of long-lived marine resources: A comparison of feedback-control management procedures. p. 243-265. In: J.A. Musick [Ed.] Life in the slow lane: Ecology and conservation of long-lived marine animals. *American Fisheries Society Symposium* 23, Bethesda, MD.

Punt, A.E., Donovan, G.P., 2007. Developing management procedures that are robust to uncertainty: lessons from the International Whaling Commission. ICES J. Mar. Sci. 64, 603–612.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2016. Management Strategy Evaluation: Best Practices. Fish and Fish. 17, 303–334.

Punt, A.E., Campbell, R., Smith, A.D.M., 2001. Evaluating empirical indicators and reference points for fisheries management: Application to the broadbill swordfish fishery off Eastern Australia. Mar. Freshw. Res. 52, 819–832.

Punt, A.E., Deng, R., Pascoe, S., Dichmont, C.M., Zhou, S., Plaganyi, E.E, Hutton, T., Venables, W.N., Kenyon, R., van der Velde, T., 2011. Calculating optimal effort and catch trajectories for multiple species modeled using a mix of size-structured, delay-difference and biomass dynamics models. Fish. Res. 101, 201–211.

Punt, A.E., Siddeek, M.S.M., Garber-Yonts, B., Dalton, M., Rugolo, L., Stram, D., Turnock, B.J., Zheng, J., 2012. Evaluating the impact of buffers to account for scientific uncertainty when setting TACs: Application to red king crab in Bristol Bay, Alaska. ICES J. Mar. Sci. 69, 624–634.

Ralston, S., Punt, A.E., Hamel, O.S., DeVore, J., Conser, R.J., 2011. An approach to quantifying scientific uncertainty in stock assessment. Fish. Bull. 109, 217–231.

Rayns, N., 2007. The Australian government's harvest strategy policy. ICES J. Mar. Sci. 64, 596–598.

Scientific, Technical and Economic Committee for Fisheries (STECF) – Evaluation of management plans: Evaluation of the multi-annual plan for the North Sea demersal stocks (STECF-15-04). 2015. Publications Office of the European Union, Luxembourg. 153pp.
Shertzer, C.E., Prager, M.H. Williams, E.H., 2008. A probability-based approach to setting annual catch levels. Fish. Bull. 206, 225–232.

Smith, A.D.M., Brown, C.J., Bulman, C.M., Fulton, E.A., Johnson, P., Kaplan, I.C., Lozano-Montes, H., Mackinson, S., Marzloff, M., Shannon, L.J., Shin, Y.-J., Tam, J., 2011. Impacts of fishing low trophic level species on marine ecosystems. Science, 333: 1147-1150

Smith, A.D.M., Smith, D.C., Haddon, M., Knuckey, I.A., Sainsbury, K.J., Sloan, S.R., 2013. Implementing harvest strategies in Australia: 5 years on. ICES J. Mar. Sci. 71, 195-203.

Smith, M.D., Fulton, E.A., Day, R.W., Shannon, L.J. Shin, Y.-J., 2015: Ecosystem modelling in the southern Benguela: comparisons of Atlantis, Ecopath with Ecosim, and OSMOSE under fishing scenarios. African J. Mar. Sci. doi: 10.2989/1814232X.2015.1013501.

Stobutzki, I., Vieira, S., Ward, P., Noriega, R., 2011. Southern and Eastern Scalefish and Shark Fishery overview. In: *Fishery Status Reports 2010: Stautus of fish stocks and fisheries managed by the Australian Government* (eds Woodhams, J., Stobutzzki, I., Vieier, S., Curtotti, R., and Begg, G.A.) Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra.
Tuck, G.N., 2014. Stock assessment of Blue grenadier (*Macruronus novaezelandiae*) based on data up to 2012. In: G.N. Tuck [Ed.] *Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2013, Part 1.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research. Chapter 6, pp 61–115.

van Putten, I. E., Gorton, R. J., Fulton, E. A., Thebaud, O. 2013. The role of behavioural flexibility in a whole of ecosystem model. ICES J. Mar. Sci. 70: 150–163.

Wayte S., Klaer N., 2010. An effective harvest strategy using improved catch-curves. Fish. Res. 106, 310–320.

Supplementary Materials – Figure

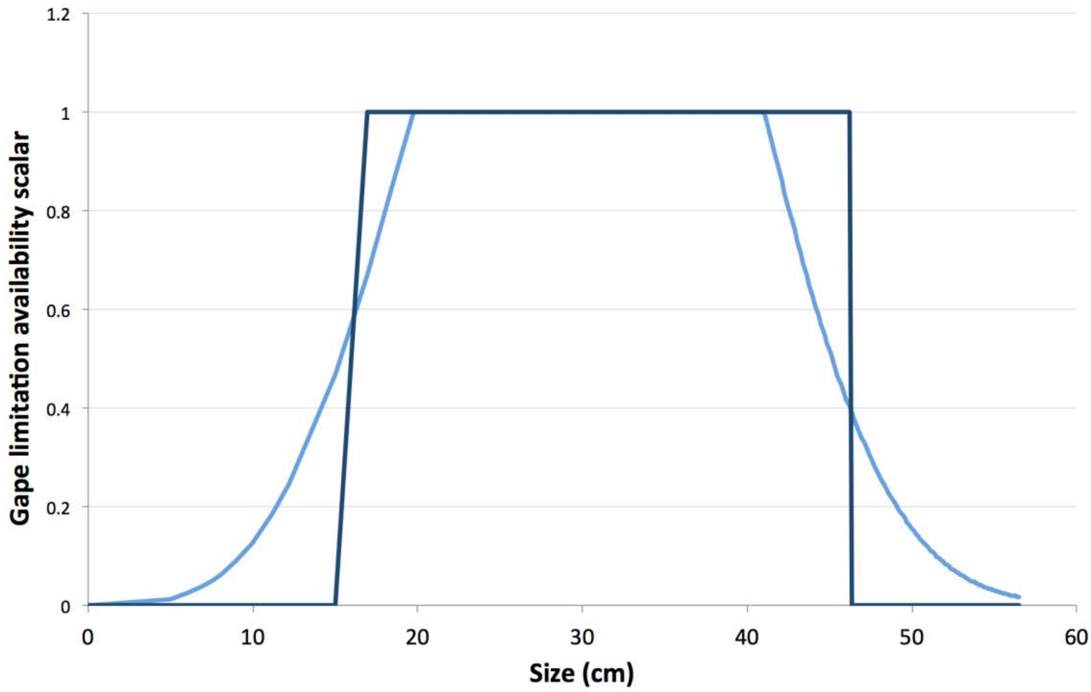


Figure E.S. 1Comparison of gape limitation functions: original Atlantis-SE (dark blue) and smooth curve of Atlantis-RCC (lighter blue).

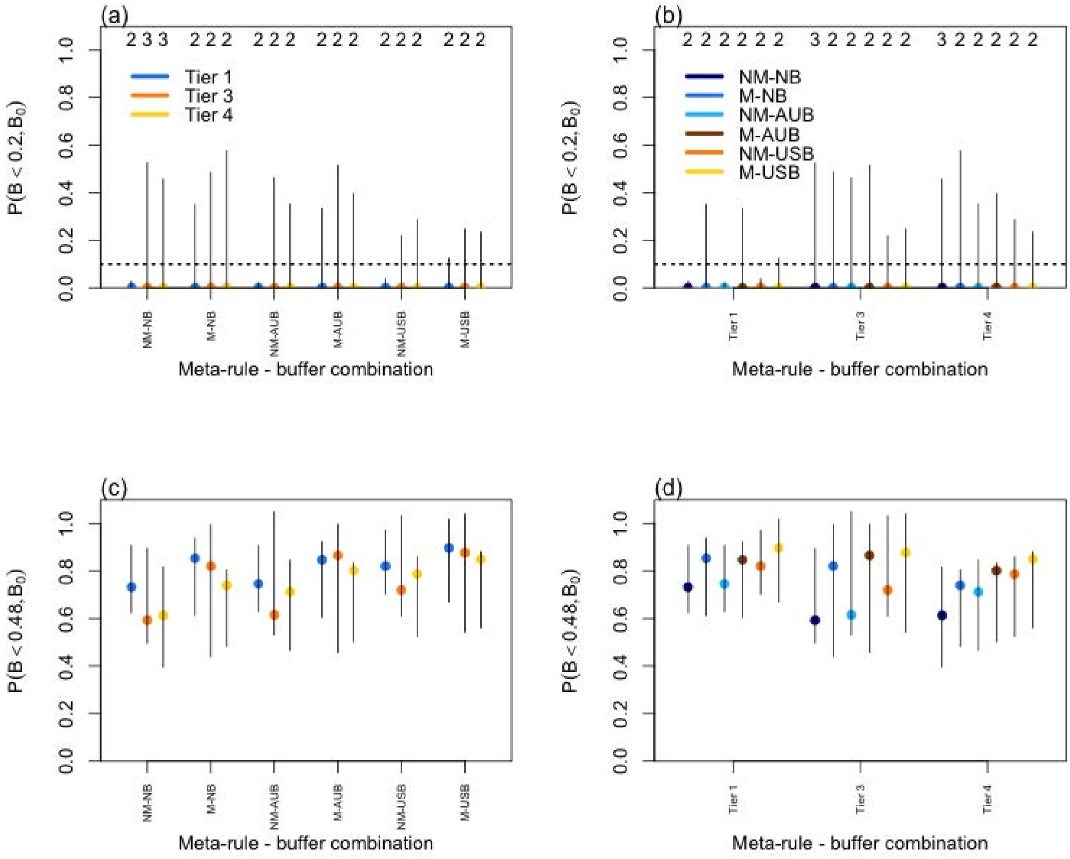


Figure E.S. 2 As for Figure 4, but calculated for the short term only (during years 10-20 of the projection period; this period was chosen to give historically depleted stocks some opportunity to recover while still highlighting shorter term risks than those conveyed in the Figure E.4).

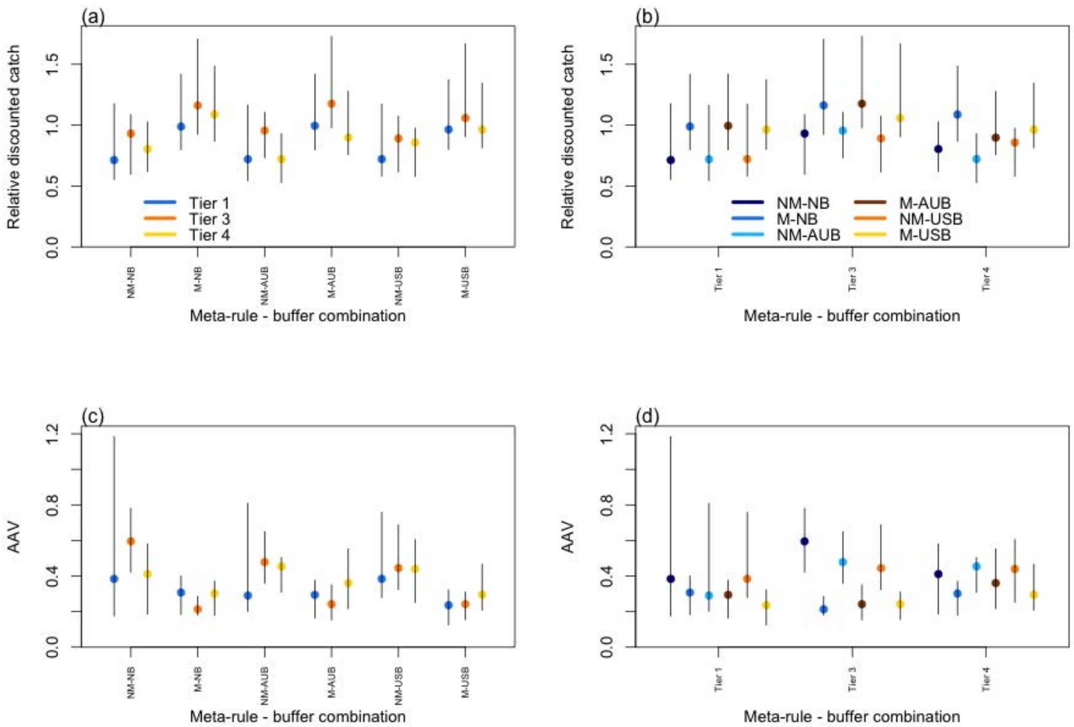


Figure E.S. 3 As for Figure E.7, but calculated for the short term only (during the first 20 years of the projection period).

Supplementary Materials – Table

Table E.S. 1 Trophic groups used in Atlantis-RCC. Groups in bold have been modified to allow for multiple size-at-age growth curves (see Section 2.1 for details). Groups marked with an asterisk have multiple stocks.

Model Component	Group Composition
Pelagic invertebrates	
Large phytoplankton	Diatoms
Small phytoplankton	Picophytoplankton
Small zooplankton	Heterotrophic flagellates
Mesozooplankton	Copepods
Large zooplankton	Krill and chaetognaths
Gelatinous zooplankton	Salps (pryosomes), coelenterates
Pelagic bacteria	Pelagic attached and free-living bacteria
Squid	*Sepioteuthis australis*, *Notodarus gouldi*
Benthic invertebrates	
Sediment bacteria	Aerobic and anaerobic bacteria
Carnivorous infauna	Polychaetes
Deposit feeders	Holothurians, echinoderms, burrowing bivalves
Deep water filter feeders	Sponges, corals, crinoids, bivalves
Shallow water filter feeders	Mussels, oysters, sponges, corals
Scallops	*Pecten fumatus*
Herbivorous grazers	Urchins, *Haliotis laevigata*, *Haliotis rubra*, gastropods
Deep water megazoobenthos	Crustacea, asteroids, molluscs
Shallow water megazoobenthos	Stomatopods, octopus, seastar, gastropod, and non-commercial crustaceans
Rock lobster	*Jasus edwardsii*, *Jasus verreauxi*
Meiobenthos	Meiobenthos
Macroalgae	Kelp
Seagrass	Seagrass
Prawns	*Haliporoides sibogae*
Giant crab	*Pseudocarcinus gigas*
Fin-fish	
Small pelagics*	Sardinops, sprat, *Engraulis*
Redbait	Emmelichthyidae (*Emmelichthys nitidus*)
Mackerel*	*Trachurus declivis*, *Scomber australisicus*
Migratory mesopelagics	Myctophids
Non-migratory mesopelagics	Sternophychids, cyclothene (lightfish)
(School) Whiting*	*Sillago* spp
Shallow water piscivores	*Arripis*, *Thyrsites atu*, *Seriola*, leatherjackets
Blue warehou*	*Seriolella brama*
Silver warehou	*Seriolella punctata*
Tuna and billfish*	*Thunnus*, *Makaira*, *Tetrapturus*, *Xiphias*
Gemfish*	*Rexea solandri*
Shallow water demersal fish*	Flounder, *Pagrus auratus*, Labridae, *Chelidonichthys kumu*, *Pterygotrigla*, *Sillaginoides punctata*, *Zeus faber*
Flathead*	*Neoplatycephalus richardsoni*, *Platycephalus* spp
Redfish*	*Centroberyx*
Morwong*	*Nemadactylus*
Pink ling*	*Genypterus blacodes*
Blue grenadier	*Macruronus novaezelandiae*
Blue-eye trevalla	*Hyperoglyphe antarctica*
Ribaldo	*Mora moro*
Orange roughy*	*Hoplostethus atlanticus*

Model Component	Group Composition
Dories and oreos*	Oreosomatidae, Macrouridae, *Zenopsis*
Cardinalfish	*Epigonus*, *Apogonops anomalus* and other Apogonidae and Dinolestidae
Sharks	
Gummy shark*	Mustelus antarcticus
School shark*	Galeorhinus galeus
Demersal sharks	*Heterodontus portusjacksoni*, Scyliorhinidae, Orectolobidae
Pelagic sharks	*Prionace glauca*, *Isurus oxyrunchus*, *Carcharodon carcharias*, *Carcharhinus*
Dogfish	Squalidae
Gulper sharks	Centrophorus
Skates and rays	Rajidae, Dasyatidae
Top predators	
Seabirds	Diomedeidae, *Ardenna*, Laridae, Sternidae, *Morus*
Seals	*Arctocephalus pusillus doriferus*, *Arctocephalus forsteri*
Sea lion	*Neophoca cinerea*
Dolphins	Delphinidae
Orcas	*Orcinus orca*
Baleen whales	*Megaptera novaeangliae*, Balaenoptera, *Eubalaena australis*

Table E.S. 2 Specifications for the stock assessments of the 14 treatment species and how the data were generated. Note 5S stands for 5-SAFE. “E” and “W” denote fleets (and assessments for Redfish) east and west of 1400E respectively.

Criteria	Blue grenadier	Orange roughy	Tiger flathead	Gemfish	Gummy shark	Morwong	Pink ling	Whiting	Redfish[‡]	Blue warehou	Blue-eye trevalla	Demersal sharks	Gulper sharks	Generic demersal fish
Management Specification														
Actual SESSF Tier	1	1	1	1	1	1[+]	1	1	3 E 1 W	4	4	4	5S	5S
Number fleets used in tier 1 assessment	2	1	4	4E 2W	3	3E 1W	2E 2W	3	1E 1W	3E 1W	1	3	2	1
Data generation parameters														
Aging error standard error	0.103	0.08	0.103	0.1	0.1	0.055	0.1	0.073	0.1	0.05	0.1	0.05	0.05	0.055
Length measurement standard error	0.123	0.08	0.103	0.12	0.1	0.011	0.1	0.073	0.1	0.05	0.1	0.05	0.05	0.011
Fleet specific coefficients of variation for CPUE	0.1 for all	0.1	0.5 0.2 0.05 0.4	0.2 for all	0.1 for all	0.1 for all	0.1 for all	0.3 for all	0.1 for all	0.1 for all	0.2	0.1 for all	0.25 for all	0.1
Fleet-specific coefficient of variation for landings	0.1 for all	0.1	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1	0.1 for all	0.1 for all	0.1
Fleet-specific coefficient of variation for discards	0.3 for all	0.1	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1 for all	0.1	0.1 for all	0.1 for all	0.1

‡ East and West represents the split here between Redfish and Bight redfish

Appendix F. Suggested application

Summary

This Appendix is aimed at key target commercial species, not byproduct or bycatch. However, some of the logic in this Appendix also can apply to byproduct and bycatch species. It addresses aspects of the 2007 Commonwealth Harvest Strategy Policy (HSP) (DAFF, 2007), which is also aimed at key target commercial species, by providing advice aimed at fisheries managers to implement the risk-cost-catch trade-off, while maintaining risk equivalency across tiers. It is a source of resource materials for other documents being developed e.g. the new Commonwealth Harvest Strategy Guideline (FRDC 2016-234) and AFMAs Ecological Risk Management Framework.

The Appendix is divided into several sections. Firstly, a background is provided that introduces the concept of the risk-cost-catch trade-off, a definition of risk and the difference between variance and bias. The Appendix then moves to proposing a generic system of Categories and some guiding principles are provided on how to address the risk-cost-catch trade-off in terms of risk equivalency. These guiding principles are then operationalised through a series of steps that could be adapted for use by a RAG or managers. The final section considers two fishery contexts wherein difficulties exist in applying with the risk-cost-catch trade-off.

A basic philosophy of the Appendix is that a hierarchical system (or an independent harvest strategy) should aim for risk equivalency with respect to setting the TAC/TAE - irrespective of assessment and harvest strategy approach used - to achieve the target reference point while avoiding the limit reference point. This should be undertaken in the context of addressing associated uncertainties, as stated in the Policy and Guidelines.

Background

Risk-cost-catch trade-off

The management of renewable resources such as fisheries can be complex given the range of species and habitats that are affected by fishing within a jurisdiction. This is true even if one narrows the focus to species that are directly targeted by fisheries. Given that fisheries managed by AFMA range from small scale and low value to large scale and high value, there is often a range of information sources and information quality available for each fishery and species (Dowling et al., 2013). Consequently, a range of methods have been developed (Smith et al., 2007) to manage each fishery while still conforming to the Australian Commonwealth's policies and legislative frameworks (or, at least, to its intent). The Harvest Strategy Policy (HSP) defines target and limit reference points (and proxies if these cannot be estimated). There are both fishing mortality- and biomass-based limit reference point so ideally it is possible to determine whether a species is being overfished (a biomass risk) or overfishing is occurring (a fishing mortality risk). Furthermore, the HSP requires that risk should remain similarly defined no matter which assessment method or harvest strategy is used – specifically, that a species remains above the limit reference point ≥90% of the time.

Fisheries managers need to trade-off between the cost of management, the risk to the resource and the catch benefit that a fishery gains from exploitation. This is known as the risk-cost-catch trade-off (Sainsbury, 2005) ((Figure F.1). The trade-offs between risk and cost, and between risk and catch are often unknown and likely to be non-linear (red surface in Figure F.1This non-linearity reflects the complex behaviour of both the resource and the economics of fishing. Although there is a trade-off of risk against catch and cost, the reality is that the HSP clearly provides some bottom lines with regard to risk of overfishing and a resource being overfished. Therefore, we are interested in the trade-off of catch and cost for a given risk (blue surface). The resultant trade-off curve can be seen in Figure F.1.

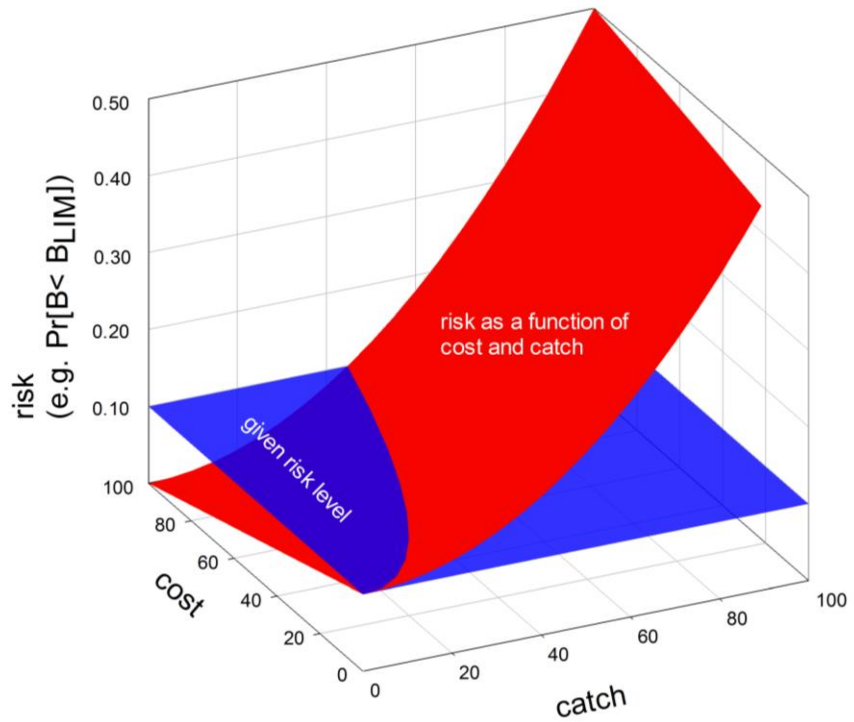


Figure F.1: Graphic of the risk-cost-catch trade-off from Sainsbury (2005).

Which risk?

The HSP addresses three major forms of risk – (i) the risk of falling below the biological LRP (known as being "overfished"), (ii) the risk of the fishing mortality being greater than the fishing mortality (F) Limit Reference Point (LRP) ("overfishing") and (iii) the risk of not reaching the Target Reference Point (TRP) over time. *In this Appendix, risk equivalency applies to both the risk of falling below the LRP and the risk of not reaching the TRP.* In the HSP (DAFF, 2007), overfishing and overfished are defined as:

Overfished: a fish stock with a biomass below the biomass limit reference point.

Overfishing: A stock is experiencing too much fishing and the removal rate from the stock is unsustainable.

The following specifications and exceptions apply:

- "Fishing mortality (F) exceeds the limit reference point (F_{LIM}). When stock levels are at, or above, B_{MSY}, F_{MSY} will be the default level for F_{LIM}.

- Fishing mortality in excess of F_{LIM} will not be defined as overfishing if a formal 'fish down' or similar strategy is in place for a stock and the stock remains above the target level (B_{TARG}).
- When the stock is less than B_{MSY} but greater than B_{LIM}, F_{LIM} will decrease in proportion to the level of biomass relative to B_{MSY}.
- At these stock levels, fishing mortality in excess of the target reference point (F_{TARG}) but less than F_{LIM} may also be defined as overfishing depending on the harvest strategy in place and/or recent trends in biomass levels.
- Any fishing mortality will be defined as overfishing if the stock level is below B_{LIM}, unless fishing mortality is below the level that will allow the stock to recover within a period of 10 years plus one mean generation time, or three times the mean generation time, whichever is less." (DAFF, 2007)

Reference points are a critical part of the HSP. However, their use relies on the ability to assess or estimate stock status, typically biomass. Assessment or estimation of an underlying stock size is, however, invariably incorrect to some degree. The nature of this "error" or "uncertainty" in the stock status can be related to the precision of the estimate, or amount the estimate could be seen to vary around the mean (i.e. the variance), and/or it can be related to the accuracy of the estimate, and how far off the estimate would consistently be from the true underlying value (i.e. bias).

A note on bias and variance

The two major sources of uncertainty that relate to fitting a model to data are variance and bias.

Variance is a statistical measure of how far each value is in a dataset from its mean; essentially the degree of noise in the data or model output. A model estimate with low variance is referred to as being a "highly precise" estimate.

Bias in statistics (in the context of this Appendix) relates to the expected difference between the estimated value and the true underlying population value. A low bias in a model estimate is referred to as being a "highly accurate" estimate.

Since an assessment and associated harvest strategies can have variance and/or bias (usually both), the **general assumption is that variance and bias increases as more data limited approaches are applied** (Figure F.2).

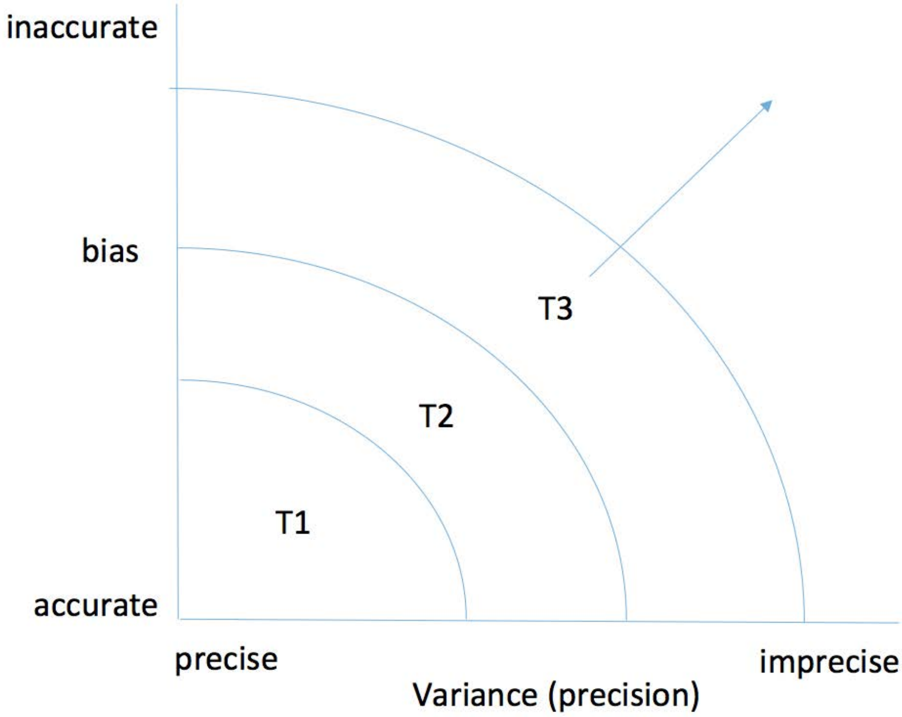


Figure F.2: Assumption of increasing bias and variances with tier. The reality is more complicated and may need individual specifications

Buffers or discount factors

Hierarchical tiered harvest strategies are the accepted means of approaching the issue of producing management advice under data or estimation constraints. Buffers are the gap between the harvest strategy-produced recommended management control (usually in terms of catch or effort) and the final management decision that accounts for risk under uncertainty (Figure F.3). If one trades catch against cost for a specific risk, there is a risk equivalent TAC (as a default; alternative controls are possible) that acknowledges assessment uncertainty. If the best estimates from the assessments are without bias and uncertainty (i.e. one has perfect information) one can implement the TAC that is calculated by this harvest strategy without buffer. However, since it is more precautionary to assume that a more data limited tier is positively biased and has higher variance, then the buffer should address assessment uncertainty (Figure F.4, lower arrow).

This approach is supported by studies that show that the more data-limited a fishery is, the poorer the performance of a stock assessment and harvest strategy in terms of risk (see review in Dichmont et al. (2015)). This was also demonstrated using the Atlantis model (Fulton et al., 2014) which tested most of the Commonwealth's applied harvest strategies (Dichmont et al., submitted; Dichmont et al., this report). The simulation tests found that many of the data limited harvest strategies did not conform to the HSP, because they exceeded the maximum specified risk of the resource being overfished (Dichmont et al. (in prep.)). Importantly however, the results were species-specific, highlighting that broad generic guidelines should be tested on a species by species basis.

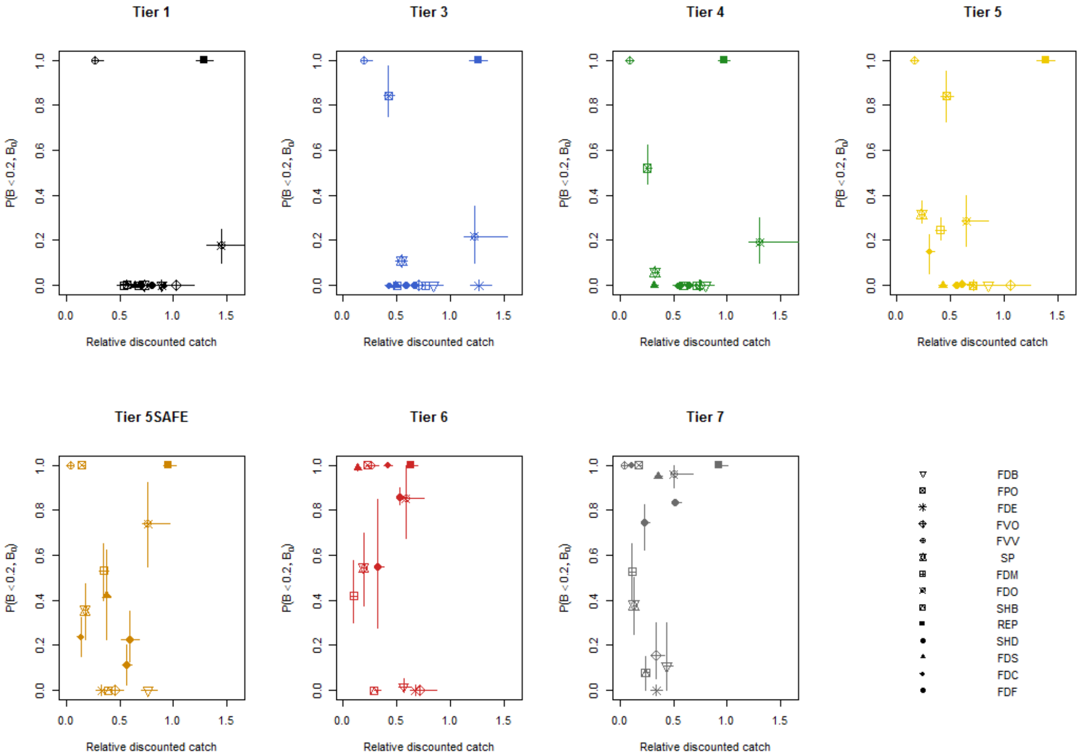


Figure F.3 Median (across simulations) probability of being below 0.2B0 against relative discounted catch with inter-quartile range values for each tier from Dowling et al. (2013). Results without meta-rules or any buffers. Each point is a species. FDB=Flathead; FPO=Morwong; FDE=Blue grenadier; FVO=Whiting; FVV=Gemfish; SP=Blue warehou; FDM=Redfish; FDO= Cascade orange roughy; SHB=Gummy shark; REP=Gulper shark; SHD=Demersal sharks; FDS=Generic demersal fish; FDC=Pink ling; FDF=Blue-eye trevalla. (see Dichmont et al., 2016).

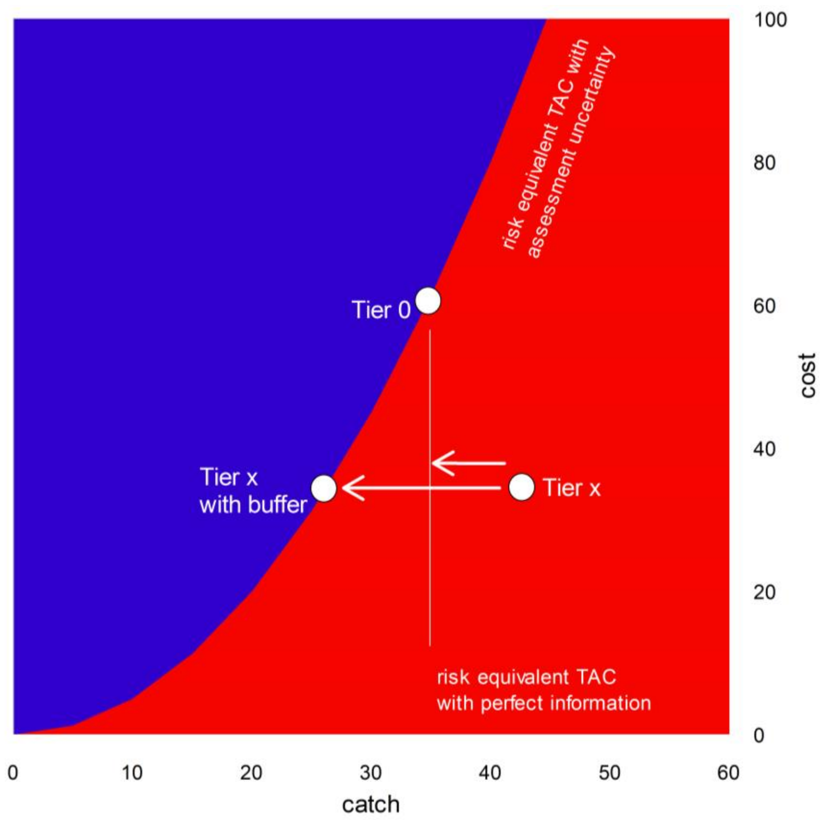


Figure F.4: The role of the buffer in the context of the cost and catch trade-off to obtain risk equivalency. Tier 0 here is a perfect information assessment without variance or bias, whereas Tier x is biased and has higher variance than Tier 0. Without considering variance and therefore only considering bias, the buffer would be the upper arrow, whereas consideration of bias and variance would place (lower arrow) on the risk equivalent line.

Guiding principles

The species-specific nature of the details behind some steps in these guiding principles means that highly detailed and prescriptive guidelines are inappropriate. Where general steps are feasible these have been noted, however the steps where species specific considerations are required have also been highlighted.

The following guiding principles should apply when developing and implementing tiered harvest strategies in accordance with the Australian HSP:

1. A hierarchical system of Categories (see below) should be used for setting the Total Allowable Catch or Total Allowable Effort unit, with the aim of achieving **risk equivalency** irrespective of the assessment and harvest strategy approach used. This risk equivalency should aim to achieve the **target reference point (on average) and avoid the limit reference point, addressing associated uncertainties as stated in the HSP**.
2. When defining the Categories to be used, the Category should be based on its relative uncertainty, in terms of both **bias and variance** in the harvest strategy, in order to account for the data-limited context. Data-limited categories should be assumed to have **higher variance and <u>positive</u> bias** in the assessment and RBC. A proposed 8 tier system is recommended if no simulation tested alternative exists for the species/fishery system in question (see below).

3. **Bias and variance** in a category **should be addressed** in one of two ways. **Either by** developing a harvest strategy that, through **simulation testing (such as Management Strategy Evaluation)**, is shown to conform to the Policy, **or by adding buffers** so that the uncertainty is addressed. A set of default buffers is suggested if no species-specific values exist/can be estimated.
4. The risk-cost-catch trade-off **should be implemented on a case-by-case (i.e. species-by-species) basis** as there is not one universally appropriate set of buffer values that can be applied across all stocks with equal cost and risk implications. Default or generic values may be used in the absence of species specific information, but simulation testing for a specific species or life history is recommended if unnecessary risk or cost (in terms of foregone catch) is to be avoided.
5. The more data-limited categories should invest in data collection in preparation for a time when a trigger is reached and a more data rich assessment may be needed (**data "banking" for the future**).

These general guiding principles can be implemented by following the recommended steps outlined below.

Recommended steps

This section guides the user through the process of operationalising the risk-cost-catch frontier within their harvest strategy, given the above guiding principles. The steps are outlined in Figure F.5 and discussed in more detail below.

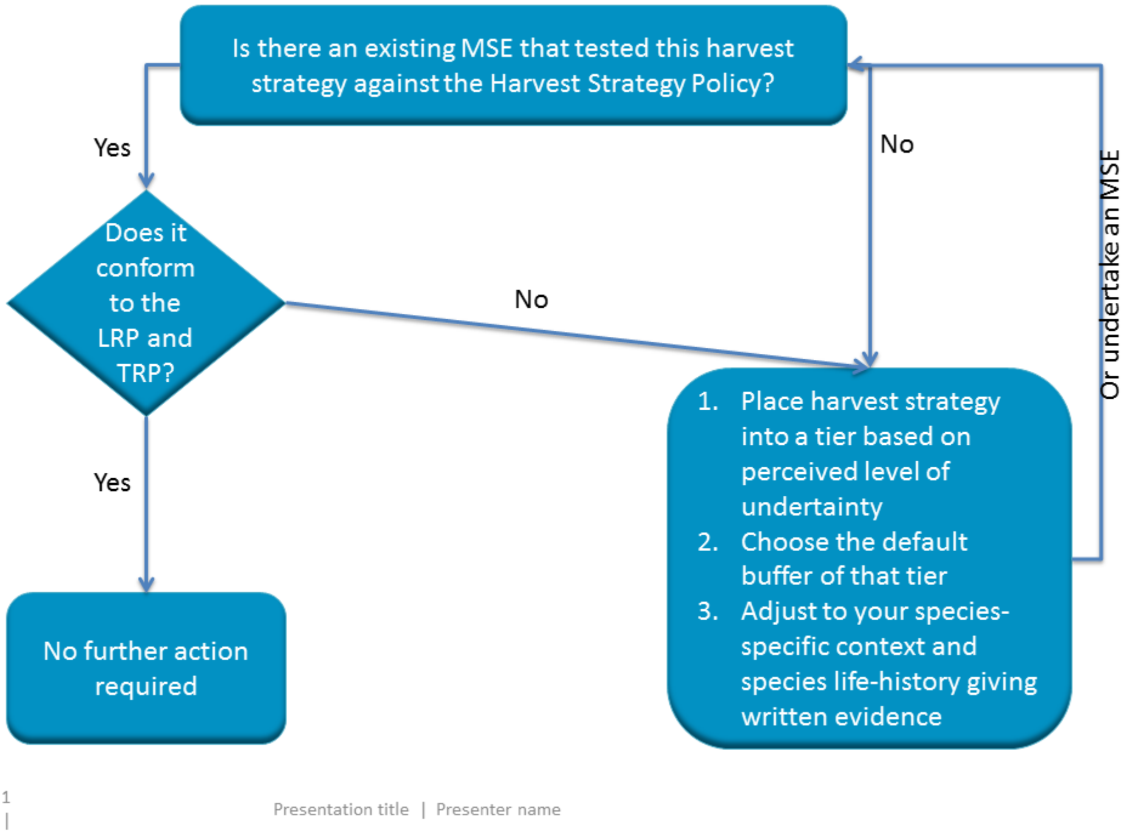


Figure F.5: Flow diagram explaining the different options for achieving risk equivalency for a specific harvest strategy. LRP and TRP are limit and target reference points.

As stated under guiding principle 3, risk equivalency (both (a) avoiding the limit reference point, and (b) achieving and maintaining the target reference point) can be achieved:

- directly through MSE testing to confirm that the harvest strategy conforms to the HSP; or via
- the use of buffers, which seek to account for bias and variance, as well as the fishery-specific context and species life history. Where possible these buffers should be directly estimated for the species/life history.

If an MSE has been undertaken that confirms that the harvest strategy conforms to both the limit and target reference point components of the HSP, then then there is **no need for Category placement or addition of a buffer**. If there is no MSE, or the MSE only tested conformation to only one of the reference points (target or limit) but not both then **there are three main steps to operationalising risk equivalency**:

1. ***Identify the appropriate Category level*** (See "Identify the Category level") for the assessment being used in conjunction with the harvest strategy – this helps determine where the harvest strategy fits in terms of data limitation, with the implication that as the harvest strategy becomes more data-limited it becomes less costly to implement, but risk is likely to also increase;

2. ***Choose the** default **buffers*** for the assessment Category being used in conjunction with the harvest strategy (see "Determine buffers"). Below we provide a basis for these. These buffers will help compensate for the potential increased risk of a more data limited harvest strategy by reducing the recommended catch;
3. If required, adjust the buffers using a scientifically-credible approach based on, for example, the fishery-specific context and species life-history. The defaults are based on simulation testing under specific conditions, so there may be specific examples why risk is less (e.g. short-lived species) or more (e.g. long-lived species);

Note that a fishery can at any stage opt to test the harvest strategy in terms of both the target **and** limit reference point using an MSE, thereby terminating the need for the use of Categories or buffers.

The operational steps are expanded on below.

Identify the Category level

The proposed harvest strategy system uses 8 categories of assessment (Table F.1.) with four levels of buffers per category Table F.2). This table still needs further adaptation to accommodate the new Harvest Strategy Policy when it is released.

Table F.1. Proposed new harvest strategy assessment Category levels (corresponding to an assessment and/or management framework), revised from those defined in SESSF harvest strategy, Dowling et al. (2013) and adjusted to include results from the simulation tests undertaken. Increasing Category numbers reflect an assumed increased risk of over-fishing due to information limitations.

Category	Description
0	Robust assessment of fishing mortality (*F*) and biomass (*B*), based on fishery-dependent and -independent data. If not MSE tested this still requires a buffer.
1	Robust assessment of *F* and *B* based on fishery-dependent data only
2	Less robust assessment of *F* and *B*, based on fishery-dependent and/or fishery-independent data
3	Empirical estimates of a) trends in relative biomass based on catch-per-unit-effort (CPUE) data b) within-season changes in relative biomass based on CPUE data c) availability of relative biomass based on informal fishery-independent surveys
4	Empirical estimates of *F* based on size and/or age data
5	Empirical estimates of *F* based on the spatial distribution of effort relative to the distribution of the species
6	No estimate of biomass or *F*; management decisions based on fishery-dependent species-specific triggers
7	No estimate of biomass or *F*; management decisions based on fishery-dependent triggers and/or indices for groups of species

These categories have been informed by hierarchical harvest strategy systems implemented in a range of fisheries in Australia and worldwide (a full description of which can be found in Dichmont et al. (2015)). This proposed system, includes a Category 0, the highest quality harvest strategy available, which is assumed to include fisheries independent information as a means of reducing both bias and variance as sources of assessment uncertainty. This would be the only category where there would in effect be no buffer – i.e. where there was no discounting of the RBC. The other Categories are based on their relative potential levels of uncertainty, with the higher the category number the more uncertain that assessment is assumed to be.

Determine buffers

These should attempt to account for both assessment bias and variance. The default buffers using Ralston et al. (2011) could be applied as a first cut to each Category. Species-specific estimates are recommended, however, because life history specifics may render a generic buffer inappropriate – either because it is not sufficient to achieve the HSP requirements or that it is overly precautionary, resulting in a cost associated with lost catch (opportunity).

Under the above Category system, Category 0 would be considered the benchmark. It combines fisheries dependent and independent data, and it is therefore assumed that bias and variance are minimised, such that no buffer is required.[5]

Table F.2: Category number and related buffers

Category	Default buffer
0 (data richest)	1[A]
1 and 2 (data rich)	0.91[B]
3, 4 and 5 (data moderate)	0.87 (Category 3); 0.82 (Category 4 and 5);[C]
5 (SAFE), 6 and 7 (data limited)	0.68

A. The Ralston et al. (2011) method places a buffer on all assessments even the data richest so this Category having no buffer does not conform and needs discussion
B. This is the average value of Ralston et al. (2011) approach from the SESSF where the percentile was set to 0.45. Other fisheries may need to calculate the relevant amount based on their assessments (if possible). Individual species values can also be used.
C. Both of these buffer values use the Ralston et al. (2011) approach, but with the percentile set to 0.4. These percentile values need further discussion.

The recommended method of estimating the buffers involves calculating the coefficient of variation (CV) of the spawning biomass estimates from a meta-analysis of all assessments performed for that species. This CV is the used in combination with a stated tolerance for the risk of fishing mortality exceeding the target levels to determine the buffer.

Where assessments are not available for a species then aggregate values can be used. These aggregate values are calculated by using the same method but across all target species assessments for the region of interest (the US west coast in the original work by Ralston et al. (2011), and the SESSF in the work presented in the main report). Ralston et al. (2011) explored three methods for calculating the CV, but found that the most effective approach was to pool residuals from all species across all available assessments (assuming equal weighting). If sufficient assessments exist, this aggregate estimation could be done for classes of life histories rather than all species together. Based on example analyses to date, this would require multiple assessments be collated for more than 10 species or stocks per life history being considered.

Buffers could be further adjusted, given a fishery-specific context, and/or the species life history. A very credible reason for using a smaller buffer than the default would need to be provided, preferably from simulation testing. The following should be considered:

- The species-specific values of Ralston et al. (2011) as opposed to the pooled values across all species.

[5] This assumption is based on simulation testing and is why these Categories are re-ordered in comparison with past/existing SESSF tier systems, such as that used in the SESSF. For example, SESSF tiers 4 and 3 now fall into Categories 3 and 4 respectively, this is because the simulation results have shown that the SESSF tier 3 (Empirical estimates of F based on size and/or age data) is, on average, less precautionary than SESSF tier 4 (which is based on CPUE) (Figure).

- Harvest control rule performance, particularly for more data-limited harvest strategies, can be predicated on the perceived current status of the fishery, and on historical precedence. For example, if the stock is likely to be below the Limit Reference Point, more precaution needs to be applied whereas if the species is likely to be above the Target Reference Point then less precaution may be acceptable.
- If historical catches are used as proxy reference points, and the stock has been heavily fished, much more precaution would be needed. This is especially true if using catch only or catch based trigger harvest strategies, as these are typically developed in the context of detecting problems/changes associated with fishery expansion from a low-effort/low-impact starting point. Such trigger based strategies are risky when used in the context of fisheries that have been heavily fishing down the stock.
- Harvest control rule performance can also depend on life history. For example, slow-growing, less productive stocks can take longer to respond to control rules, and empirical-based assessments may be slow to detect changes and respond accordingly. Thus the risk profile down the categories may vary with life history. Arguing for a less precautionary buffer for a long or medium lived species would therefore need strong scientific backing.
- The current and past MSE work have both indicated that, for some species, more data-limited (e.g. category 3) assessments perform as well as, if not better (in terms of data requirements and costs of running) than category 1 assessments. For such cases, buffers may actually be changed to conform to the MSE results.
- F-based methods, if applied to a species that has been historically overfished and where management has reduced fishing mortality (F) such that overfishing is not occurring, do not provide biomass-based information. They therefore may not describe the short-term biomass based risk of being overfished.

At any stage, the harvest strategy should ideally be tested using MSE to validate the performance of the buffer size and/or the harvest strategy. The latter would not be obliged to use buffers, provided the harvest strategy conforms to the Policy's statements regarding the target and limit reference point.

Further challenges

Highly variable stocks

Highly variable stocks present a challenge for a HSP. Fishing mortality may have to be zero to meet the HSP requirements to maintain a stock above $0.2B_0$ (with a probability greater than 0.9) and for its median of stock biomass to be at or above the target roughly 50% of the time. For example Figure F.6 and Figure F.7 show that even if we can measure or estimate a stock perfectly, and manage it to a target of $48\%B_0$, a stock that varies according to either of the black distributions would achieve the HSP objective, but the stock represented by the blue lines would not.

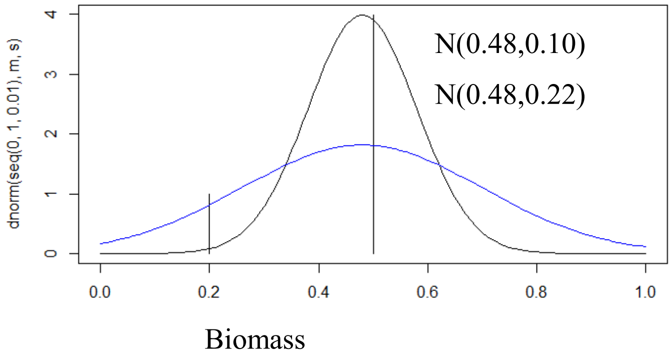


Figure F.6 Stock biomass distribution (Gaussian), black achieves the HSP: with P(B>Btarg)=0.5, and P(B>Blim)<0.1; blue does not achieve the HSP: with P(B>Btarg)=0.5, and P(B>Blim)>0.1.

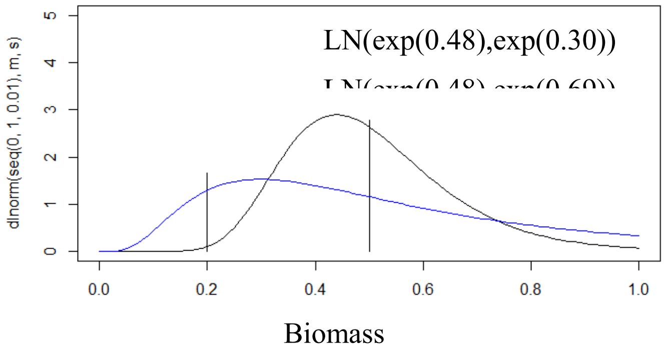


Figure F.7 Stock biomass distribution (Log-normal), black achieves HSP: with P(B>Btarg)=0.5, and P(B>Blim)<0.1; blue does not achieve the HSP: with P(B>Btarg)=0.5, and P(B>Blim)>0.1.

The manner in which these stocks should be managed with respect to the HSP reference points needs additional consideration based on the specifics of the variability. If that information is still being collected for the species then existing good practice for highly variable species (such the use of input controls in the Northern Prawn Fishery using MSE tested multi-species assessments and associated harvest strategies) should be implemented in the interim while further understanding, exploration and simulation testing is performed to ensure overall performance requirements are met.

Multi-species fisheries

Several studies have highlighted the difficulties of managing multi-species fisheries both with respect to the TRP of Maximum Economic Yield and their ecosystem interactions (see an overview paper (Rindorf et al., 2016). All agree that multi-species fisheries management is difficult. It is highly unlikely that each target species in a fishery could be maintained at the same relative target biomass proxy point (e.g. B_{48}) as this implies an improbable ecosystem structure. Furthermore, the concept of Maximum Economic Yield (B_{MEY}) is fishery-wide yet most assessments are species specific and not multi-species. The Northern Prawn Fishery is, as yet, the only Australian managed fishery that has estimated the biomass at MEY for three species combined (2 primary and 1 secondary), which means that the B_{MEY} relative to the unfished stock size is different for each (Dichmont et al., 2008; Kompas et al., 2010; Pascoe et al., 2010). These multispecies MEY estimates technically require full bioeconomic models be implemented for the full fished system – as the trajectory is as important as the end point in this case. However, this is not a straightforward endeavour and simpler, but still multispecies or ecosystem, simulations that span the major features of the fishery are required to provide relative ratios for targets across species. This is a conundrum facing many fisheries and the use of multi-species models to inform single-species assessments are on the increase, but robust solutions have yet to fully identified.

If the multispecies MEY has not been estimated and work is underway to determine feasible biomass targets then in the interim, it is recommended to ease the risk equivalency requirement across species.

References

DAFF. 2007. Commonwealth Fisheries Harvest Strategy: Policy and Guidelines.

Dichmont, C. M., Deng, A., Punt, A. E., Ellis, N., Venables, W. N., Kompas, T., Ye, Y., et al. 2008. Beyond biological performance measures in management strategy evaluation

Bringing in economics and the effects of trawling on the benthos. Fisheries Research, 94: 238–250.

Dichmont , C. M., Fulton, E., Gorton, R., Sporcic, M., Little, L. R., Punt, A. E., Dowling, N., et al. in prep. From data rich to data-limited harvest strategies – does more data mean better management? ICES Journal of Marine Science.

Dichmont, C. M., Punt, A. E., Dowling, N., De Oliveira, J. A. A., Little, L. R., Sporcic, M., Fulton, E., et al. 2015. Is risk consistent across tier-based harvest control rule management systems? A comparison of four case-studies. Fish and Fisheries: n/a-n/a.

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.

Fulton, E. A., Smith, A. D. M., Smith, D. C., and Johnson, P. 2014. An Integrated Approach Is Needed for Ecosystem Based Fisheries Management: Insights from Ecosystem-Level Management Strategy Evaluation. PLOS One, 9: e84242.

Kompas, T., Dichmont, C. M., Punt, A. E., Deng, A., Che, T. N., Bishop, J., Gooday, P., et al. 2010. Maximizing profits and conserving stocks in the Australian Northern Prawn Fishery. Australian Journal of Agricultural and Resource Economics, 54: 281–299.

Pascoe, S., Punt, A. E., and Dichmont, C. M. 2010. Targeting ability and output controls in Australia's multi-species Northern Prawn Fishery. European Review of Agricultural Economics, 37: 313–334.

Ralston, S., Punt, A. E., Hamel, O. S., deVore, J. D., and Conser, R. 2011. A meta-analytic approach to quantifying scientific uncertainty in stock assessments. Fisheries Bulletin, 109: 217-231.

Rindorf, A., Dichmont, C. M., Levin, P. S., Mace, P., Pascoe, S., Prellezo, R., Punt, A. E., et al. 2016. Food for thought: pretty good multispecies yield. ICES Journal of Marine Science: Journal du Conseil: fsw071.

Sainsbury, K. 2005. Cost-effective management of uncertainty in fisheries. *In* Outlook 2005. Australian Bureau of Agricultural and Resource Economics, Canberra, A.C.T.

Smith, A. D. M., Fulton, E. J., Hobday, A. J., Smith, D. C., and Shoulder, P. 2007. Scientific tools to support the practical implementation of ecosystem-based fisheries management. ICES Journal of Marine Science, 64: 633-639.

CONTACT US

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

FOR FURTHER INFORMATION

Dr Catherine Dichmont
t +61 419 950 076
e cathydichmont@gmail.com
w www.dichmont.weebly.com

AT CSIRO WE SHAPE THE FUTURE

We do this by using science to solve real issues. Our research makes a difference to industry, people and the planet.

As Australia's national science agency we've been pushing the edge of what's possible for over 85 years. Today we have more than 5,000 talented people working out of 50-plus centres in Australia and internationally. Our people work closely with industry and communities to leave a lasting legacy. Collectively, our innovation and excellence places us in the top ten applied research agencies in the world.

WE ASK, WE SEEK AND WE SOLVE