Stock Assessment Integration: a review

Dichmont, C.M., Deng, R.A., Punt, A.E. and Little, L.R.

2018

FRDC Project No **2014-039**

ISBN **978-1-4863-0997-9**

FRDC report 2014-039.

OWNERSHIP OF INTELLECTUAL PROPERTY RIGHTS

CREATIVE COMMONS LICENCE

DISCLAIMER

RESEARCHER CONTACT DETAILS

Name: C.M. Dichmont

Address: Headsail Dve

Banksia Beach, Qld

Phone: 0419950076

Fax:

Email: cathydichmont@gmail.com

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

Contents .... **iii**

Acknowledgments .... **vi**

Executive Summary .... **vii**

Keywords ix

Introduction **1**

Background 1

Need 2

Objectives **3**

Chapter 1. A Review of Stock Assessment Packages in the United States **4**

Introduction 4

Overview of packages 6

Output statistics, projections, and quantification of uncertainty 15

Software issues 16

Discussion 17

Advantages and disadvantages of packages 18

Beyond the US 20

Peer-review and packages 20

Next steps 21

Final comments 22

References 24

Chapter 2. How many of Australia's stock assessments can be conducted using stock assessment packages? **35**

Introduction 35

Materials and Methods 36

Survey questionnaires 36

Data analysis 37

Results 42

Representativeness of the data set 42
Synthesis of assessments 46
Which assessments can be conducted using stock assessment packages? 52
Discussion 55
References 59
Supplementary Tables 64
Supplementary table 2.2 References 69
Chapter 3. Present and future stock assessment needs 71
Introduction 71
Methods 71
Results and discussion 72
Conclusions and recommendations 79
References 81
Supplementary material: Interview survey questions 83
Questions regarding Stock Assessment strategy, capacity and needs 83
Ethics statement 83
Background 83
Definitions 84
Stock assessment strategic vision 84
Stock assessment business case 84
Stock assessment capacity 84
Stock assessment needs 85
Chapter 4. International stock assessment review processes 86
Introduction 86
Conclusions and Recommendations 87
Regional summaries 88
USA: summary of presentation “National Standard on Scientific Information for Marine Fisheries Management in the United States” (Presenter: Bill Michaels) 88
USA: “Center for Independent Experts – NOAA Fisheries’ Peer Review Program” (Presenter: Stephen Brown) 88

USA – "Peer Review of High Profile Benchmark Stock Assessments in the New England and Mid-Atlantic Regions: SAW/SARC Process" (Presenter: James Weinberg) 89

USA: "The Importance of Peer Review for Sedar Stock Assessments" (Presenter: Julia Byrd) 90

Canada – "Peer Review and Science Advice in Canada" (Presenter: Jake Rice) 91

EU – "The Northeast Atlantic Experience with Peer Review of Fish Stock Assessments" (Presenter: Henrik Sperholt) 92

Australia – "Improving Fisheries Science and Management Via Peer Review: An Australian Perspective" (Presenter: Cathy Dichmont) 93

New Zealand – "Science Peer Review: What, When, Where, and How Should We Review for Cost-Effectiveness?" (Presenter: Martin Cryer) 94

Tuna RFMOs – "Peer Review in Regional Fisheries Management Organizations, the Example of Iccat" (Presenter: David Die) 95

Antarctic – "Peer-Review of Fisheries Management in the Commission for the Conservation of Marine Living Resources (CCAMLR)" (Presenter: Alistair Dunn) 96

Marine Stewardship Council – "The Role of Peer Review in Providing Assurance in MSC Certification" (Presenter: David Agnew) 97

South Africa (not presented but added for interest) 98

Conference special session Program 99

References 100

Implications 101

Recommendations 102

Further development 105

Extension and Adoption 106

Project coverage 106

ACKNOWLEDGMENTS

The authors would like to thank the Australian, NZ and USA respondents to our surveys, and other interactions.

In Australia and NZ, these were Patrick Cordue, Ainslie Denham, Alistair Dunn, Andre Punt, Malcolm Haddon, Klaas Hartmann, Athol Whitten, Belinda Barnes, C. T. T. Edwards, Eva Plaganyi, Geoff Tuck, Jemery Day, Judy Upston, Marco Kienzle, Michael O'Neill, Neil Klaer, P.L. Cordue, Rich Little, Richard Hillary, Rick McGarvey, Rik Buckworth, Robin Thompson, Roy Deng, Sally Wayte, Shijie Zhou, Simon Conron, Thor Saunders.

In the USA, these were Jon. Brodziak, Yi-Jay Chang, Jason M. Cope, James N. Ianelli, Christopher M. Legault, Richard D. Methot, Jr, Clay E. Porch, Michael H. Prager, Kyle Shertzer, Nick Davies, Beth Babcock (U. Miami), Murdoch McAllister (UBC), Paul Nitschke (NOAA, NEFSC), Erik Williams (NOAA, SEFSC), and E.J. Dick (NOAA, SWFSC). Melissa Haltuch (NWFSC) and Mike Wilberg (University of Maryland) are thanked for comments on an earlier version of the USA paper.

Individuals from each jurisdiction that provides stock assessment expertise were interviewed. Comments and summaries, and other input for Chapter 3 were also provided by this group. Their honest and well thought through responses to the survey interview was greatly appreciated. Thanks to Thor Saunders, Mick O'Neill, Geoff Liggins, Paul Hamer, Harry Gorfine, Mellissa Schubert, Klaas Haartman, Craig Mundy, Stephen Mayfield, Rick McGarvey, Brent Wise, Norm Hall, Ainsley Denham, Alex Hesp, Geoff Tuck, Éva Plagányi.

CMD, RAD, and AEP were funded by the Fisheries Research and Development Corporation, Cathy Dichmont Consulting and CSIRO. AEP was partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative agreement No. NA10OAR4320148.

EXECUTIVE SUMMARY

Stock assessments provide scientific advice in support of fisheries decision making. They involve fitting population dynamics models to fishery and monitoring data to provide estimates of time-trajectories of biomass and fishing mortality in absolute terms and relative to biological reference points such as B_{MSY} (the biomass corresponding to Maximum Sustainability Yield, MSY) and F_{MSY} (the fishing mortality rate corresponding to MSY), along with measures of uncertainty. Some stock assessments are conducted using software developed for a specific stock or group of stocks. However, increasingly, stock assessments are being conducted using packages developed for application to several taxa and across multiple regions. We reviewed the range of packages used to conduct assessments of fish and invertebrate stocks in the United States because these assessments tend to have common goals, and need to provide similar outputs for decision making. Sixteen packages were considered, five based on surplus production models ("A Stock Production Model Incorporating Covariates"; "Bayesian Surplus Production Model-1"; "Bayesian Surplus Production Model-2"; "Depletion-Based Stock Reduction Analysis"; "Extended Depletion-Based Stock Reduction Analysis"), one based on a delay-difference model ("Collie-Sissenwine Analysis"), and the remainder based on age-structured models ("Assessment Method for Alaska", "Age Structured Assessment Program", "Beaufort Assessment Model", "MULRIFAN-CL", "Statistical catch-at-length", "Stock Synthesis", "Simple Stock Synthesis", "Extended Stock Synthesis", "Virtual Population Analysis", "VPA-2BOX", "). Most of the packages that are freely available for use by analysts in the US and around the world, have been evaluated using simulations, and can form the basis for forecasts. The packages differ in the types of data they can use and their ease of use. This report highlights the benefits and disadvantages of stock assessment packages in terms of allowing analysts to explore many assessment configurations and facilitating the peer-review of assessments. It also highlights the disadvantages associated with the use of packages for conducting assessments. Packages with the most options and greatest flexibility are the most difficult to use, and see the greatest development of auxiliary tools to facilitate their use.

Most of the stock assessments conducted in the USA and in New Zealand are based on packages that have been developed for generic use, and are well documented, and tested using simulation. However, this is not the case for assessments conducted in Australia and many other countries where most assessments are bespoke. The second chapter of this report reviews many of the model-based stock assessments for Australian fisheries to evaluate how many assessments could have been conducted using the publicly-available stock assessment packages used widely in the USA and New Zealand. The 76 model-based assessments reflect 37% of the 2013 catch recorded in Australia's Status for Key Australian Fish Stocks Reports (or 34% of the total catch in 2013). It is the opinion of the authors, that all but 18 (or 24 if full rather than approximate age-size-structured models need to be used) of the stock assessments could have been conducted using stock assessment packages used in the United States and New Zealand. Adoption and use of packages for more stocks in Australia should increase the

likelihood that results are based on correctly-coded models whose estimation performance is widely understood, reduce the time needed to conduct assessments, and speed up the peer-review process. The availability of training, manuals, and example data sets for stock assessment packages should partially address their additional complexity. Additional benefits, in terms of numbers of assessed stocks, could be gained if Australian stock assessment scientists develop a forum to collaborate and share methods. These results are applicable to many other jurisdictions that undertake stock assessments.

In Chapter 3, a semi-structured set of phone interviews with stock assessment scientists from each jurisdiction in Australia that provides stock assessment expertise was undertaken. The interview questions asked how each jurisdiction's business case (if there was one) linked with the appropriate fisheries policies; the stock assessment team's capacity and resourcing; their view on the role of stock assessment packages; any impediments to moving to packages if they require or wish to; and their medium term overall stock assessment needs and priorities. All eight jurisdictions were interviewed. Responses and discussions during the interview, lasting about two hours, were extensive and fruitful. Interviewees were provided summaries of their respective interviews, as well as this chapter for input and comment. Summaries of key points were provided in the results, but mainly showed a link between interviewees' comments and their own internal stock assessment capacity and investment. For example, there was generally an inverse relationship between a jurisdiction's in-house stock assessment capacity and investment, and its willingness to transition to packages. There was also concern expressed about succession within the stock assessment community – with few young, well trained stock assessment scientists being available, and observations that positions often took a long time to be filled. In many jurisdictions' opinions, universities were not generally seen as a good source of new stock assessment scientists with various reasons provided. The great need for more streamlining and automation of data entry, extraction, delivery, assessment and reporting was highlighted by all. Only some of the jurisdictions were successful in moving to the full streamlining process, although many others had succeeded in aspects thereof. However, sharing of tools and approaches was seen as a priority. Indeed, isolation of assessors were highlighted by all interviewees and several recommendations in this regard are provided including more sharing of libraries, more shared training and forums for sharing breakthroughs. The first few topics of these proposed forums are also provided. The review process is also seen as lacking with several options discussed including a Guideline to best practice assessment processes and setting up peer review panels. The highest and most commonly mentioned priority need was for more and better quality data. This was not a request for more and better data to develop more data-rich assessments, but rather a need for a) more or even some data for data-limited fisheries so that a data-limited assessment could be undertaken, b) more recreational data, where paucity of such data is an impediment for assessing even quite valuable fisheries, and c) better quality control of existing processes for data collection. In all, more than 10 recommendations are provided from these interviews.

In Chapter 4, an international review of stock assessment peer review was undertaken. The reason for this is that the above analyses showed that most of the Australian stock assessments were published as external reports that had gone through some form of internal review. Fewer than half of the stock assessment methods were published in journals. There was a general view that the input from groups such as Resource Assessment Groups (if they existed at all) and the review process of scientific journals was insufficient to fully capture the range of assessment work, and review a stock assessment. Resource Assessment Groups often lack the expertise to technically review a stock assessment (Penney et al. 2016). A small desktop study on different international stock assessment peer review processes was undertaken. This included information from a special session entitled "National and International Perspectives on Improving Science and Management through peer review – symposium" which was initiated as part of the American Fisheries Society Conference in Tampa (August 18-24, 2017). Presentations on the peer review systems in the USA; Northeast Regional Stock Assessment Workshop, Canada, New Zealand, Australia, EU particularly ICES, tuna fisheries, the Antarctic and Marine Stewardship Council. South Africa was also added as several Australian stock assessment analysts have been involved in this peer review system. The desktop review has found that three broad peer review systems can be found internationally – a) an independent in-person panel of experts; b) an independent desktop review, and c) a Working Group (WG) (or Resource Assessment Group (RAG)) system. Apart from urgent ad hoc reviews, undertaking regular independent panel-based reviews is probably too costly for Australia, especially under the current cost recovery model. However, if a panel model is considered, the process undertaken in South Africa where a fairly consistent group of individuals forms the panel and annually undertakes multiple nominated assessment reviews over a week, could be the best approach in terms of costs and benefits. Another option is for the panel to review multiple species or generic issues. If a panel approach is not supported, then the following options could be considered: 1. For all jurisdictions: FRDC facilitates an in-depth desktop review process linked to key controversial and/or valuable species SAFS chapters, where the reviewers comment on key science behind the chapter, including the stock assessment (if any), and 2. Commonwealth: AFMA to set up a cross-RAG review system for the Commonwealth fisheries where each RAG reviews the other RAGs' assessments. To facilitate more consistency between assessments for review by the public and experts, a more standard system of assessment documentation including Appendices of model settings would be beneficial.

KEYWORDS

CASAL, Fishing mortality, freeware, funding, packages, priorities, population dynamics, reference points, review, stock assessment, Stock Synthesis, uncertainty, peer review

INTRODUCTION

BACKGROUND

Fisheries stock assessment modelling is undertaken throughout Australia's jurisdictions. It usually uses a set of tools that have been developed in various programming languages and for a specific fishery. Furthermore, most of the tools used are developed by experts with high modelling capacity, yet the need for stock assessment more often outstrips the resource (financial and human) constraints. There have been many moves to address this issue through the use of generic tools and/or sharing of people resources among agencies. In other parts of the world, and to some extent Australia, a move towards cost savings has resulted in generic toolkits that are freely available, but are peer reviewed and tested in the literature.

To further address the issue of resourcing, consolidation of which and when a fishery or resource is assessed is undertaken. Data-limited fisheries tend to be under-resourced in terms of being assessed, yet cumulatively they may be an important contribution to some jurisdiction's or local region's revenue. Data-poor methods are on the increase, but are competing for resources with the more valuable data-rich fisheries. Multi-year assessments are common in Queensland, for example whereas in the Commonwealth the norm is still annual assessments. However, the stock assessment domain would benefit from a sharing of modelling status and a strategic view of the next steps. This should include creating an inventory of tools available and being used by the different jurisdictions, how much they are written for a specific case study, and whether more generic approaches can apply. Furthermore, a review of international generic approaches and learning is added to this strategic overview to investigate what the resource implication of such an approach are and the risk and rewards.

This project reviewed the stock assessment program within Australia in the context of international learnings, while still maintaining the need for Australian specific solutions and building on existing methods especially those that have resonance with fishers and managers. It worked with the jurisdictions to undertake a strategic review of stock assessment in Australia and provided recommendations for a possible set of investments models.

NEED

Stock assessment is a set of tools and methods generally used to assess the status of wild capture fisheries stocks. They range from complex statistical and mathematical models, to simple, almost back of the envelope, methods. They are used to predict population size, quantify the impact of fisheries on the population and in some jurisdictions, provide key outputs needed in harvest strategies. There is a diverse range of methods in a field where practitioners have tended to produce home-grown tools in their favourite code languages (R, FORTRAN, C++, Visual Basic, ADMB etc.). The use of a specific model or method is often historical rather based on an objective evaluation of options e.g. the risk-cost-catch framework (see Method references). In recent years changes have occurred allowing some shift away from previous approaches:

- More off-the-shelf methods, with a diverse range of flexible features, have become available and some uptake has occurred e.g. Stock Synthesis (SS) (http://nft.nefsc.noaa.gov/SS3.html)
- There has been some convergence of language tools using the open source model (e.g. ADMB, Gnu and R)
- Stock assessment tool kits have become freely available e.g. the NOAA fisheries toolbox (http://nft.nefsc.noaa.gov/index.html)

Although the use of home-grown tools is not in itself an issue, it does not always allow for synergies and more cost-effective practices. For example, it has become standard practice in the USA to have a model developed and maintained by a team, have it independently tested and then made available as an off-the-shelf GUI driven tool. Many stock assessment scientists now use these tools. In Europe, ICES also tends to use standard approaches.

There is a real need for a more strategic view of the framework Australia should adopt in the present climate of:

- increasing cost and financial constraints;
- decreasing capability resources; and
- a greater need to rapidly assess not only data rich but also data poor fisheries.

OBJECTIVES

1. Review existing stock assessment tools/methods used in Australia.
2. Review Australian stock assessment needs, model developer and user capacity.
3. Review methods used and reviews undertaken elsewhere in the world.
4. Assess the relative merits of off-the-shelf versus case-specific assessments.
5. With input from the different jurisdictions, provide recommendations for a possible set of investment models.

CHAPTER 1. A REVIEW OF STOCK ASSESSMENT PACKAGES IN THE UNITED STATES[1]

INTRODUCTION

Management of many of the world's marine fish and invertebrate stocks is supported by scientific advice (Mace et al., 2001). Many of the fisheries in countries such as the United States (US), Canada, Australia, and New Zealand, as well as the fisheries in European waters, base this advice on estimates of abundance and fishing mortality derived from stock assessments. These estimates are typically expressed both in absolute terms and relative to reference points. They often are used as the basis for applying harvest control rules or evaluating proposed regulatory measures (such as total allowable catches or limits on fishing effort) that will frame management actions. The value of this technical approach to fishery management is exemplified in Worm et al. (2009).

The comprehensiveness of stock assessments, and the number of data sources, is continuing to increase. The assessments conducted in the 1970s and 1980s that estimated biomass and fishing mortality[2] were typically based on Virtual Population Analyses (VPA; Laurec and Shepherd, 1983; Pope and Shepherd, 1982; Shepherd, 1999) or surplus production (biomass dynamics) models (Schnute, 1977; Butterworth and Andrew, 1984). These methods relied on estimates of catch and effort, and catch-at-age data in the case of VPA. The trend in stock assessments since 1982 has been towards 'integrated' or 'statistical' methods that separate the development of the model of the population dynamics from that of the model that relates the observations to the predictions of the population dynamics model (e.g. Fournier and Archibald, 1982; Deriso et al., 1985; Methot, 1990). In contrast to the earlier approaches, "integrated" assessment methods allow more and diverse data sets to be included in assessments (Maunder and Punt, 2013). These methods also allow uncertainty to be quantified using frequentist and Bayesian methods (Magnusson et al., 2013; Stewart et al., 2013), and they allow much more complex population dynamics models to form the basis for stock assessments. Few assessments explicitly allow for trophic interactions to determine management actions (but see Plaganyi et al., 2014). However,

---

[1] Published in *Fisheries Research* with authors: Catherine M. Dichmont, Roy Deng, Andre E. Punt, Jon. Brodziak, Yi-Jay Chang, Jason M. Cope, James N. Ianelli, Christopher M. Legault, Richard D. Methot, Jr, Clay E. Porch, Michael H. Prager, and Kyle Shertzer

[2] Other methods such as yield-per-recruit analysis were, and continue to be, applied but do not provide estimates of biomass and fishing mortality.

increasingly the assessments used to provide scientific advice allow for spatial structure (e.g. Punt et al., 2000; Thorson and Wetzel, 2015) and analyse data for multiple stocks (e.g. Punt and Kennedy, 1997; Porch et al., 2001; McKenzie, 2015) and even multiple species (e.g. Dichmont et al., 2003) simultaneously.

The increasing complexity of stock assessments has not come without cost. In particular, the technical skills required to apply modern stock assessment methods have increased substantially, not only because analysts need to choose which model they wish to use, but because, with multiple data sources, it is necessary to weight each data source as well as each data point within each data source, and the results of stock assessments may be sensitive to how data are weighted (Richards, 1991; Francis, 2011). In addition, there may be many options within any one modelling package. Unfortunately, the US currently has an insufficient number of suitably qualified scientists capable of conducting stock assessments (DoC, 2008). The lack of assessment analysts has been best documented in the US, but is a worldwide issue. Effective use of existing stock assessment tools requires training, and in the US at least, the key federal agency supporting stock assessments (NMFS) has invested in supporting faculty at universities to increase the number of scientists who are capable of conducting stock assessments (e.g. Berkson et al., 2009) as well as in funding for students studying population dynamics.

The importance of stock assessments as the basis for fisheries management advice has meant that the need for an effective peer-review process has increased almost in concert with the complexity of the stock assessment methods used to provide advice. Affected members of fishing communities cannot be expected to understand the technical details of assessment methods, so they demand an open review process to assure the correctness of the results. Peer review relates to many aspects of a stock assessment, including the choice of the assessment method, selection and weighting of data, and selection of options when applying the assessment method. In several cases, multiple alternative model configurations are available, and the peer review process leads to a selection of which, if any, of those configurations should form the basis for management advice. The peer review of stock assessments is made more challenging if the reviewers are not familiar with the assessment method, including whether it performs as expected.

One consequence of increasing stock assessment complexity is a trend towards the use of stock assessment packages. For example, in 1999 all assessments for fish stocks in Australia's Southern and Eastern Scalefish and Shark Fishery were based on models coded by the assessment analyst and designed for a unique stock. However, in 2015, all but one of these assessments were conducted using a stock assessment package (Stock Synthesis) developed in the US (Methot and Wetzel, 2013). This trend is evident in several other fisheries management jurisdictions, because the use of packages reduces the time it takes to conduct a stock assessment and can benefit from improvements based on repeated use and past simulation testing. However, the use of packages is not without disadvantages. This Chapter provides a summary of sixteen stock assessment packages developed for application to fish and invertebrate stocks under federal management in the US, and discusses the advantages and disadvantages of the use of packages for stock assessments. A package for the purpose of this Chapter is software designed to be applied to more than one species / stock and is freely available. We focus on the US because assessments for stocks in the US tend to have common goals and need to provide similar outputs for decision making.

OVERVIEW OF PACKAGES

We summarised the characteristics of the sixteen packages considered in this Chapter (Table 1.1). Most of the packages were developed by the US National Marine Fisheries Service (NMFS), while some of the packages (e.g. BSP1, and MULTIFAN-CL) are collaborations between NMFS scientists and researchers at universities and other management agencies. One of the packages (BSP2) was developed solely by university researchers. The packages have been used to conduct assessments for both state and federally-managed fisheries in the US, as well as for highly migratory species (HMS) such as tunas, billfishes and sharks. Several of the packages have been used in countries outside of the US, such as Canada, Chile, Argentina and Australia.

Table 1.1 excludes some packages developed for the US that are not well documented or have not been used in recent years to provide management advice, such as Coleraine (Hilborn et al., 2000), A-SCALA (Maunder and Watters, 2003a), and CAB (Cope et al., 2003), although Coleraine formed the basis for the simulation study conducted by Magnusson and Hilborn (2007), and CAB formed the basis for the simulation study conducted by Little et al. (2015).

The packages evolved in each region of the US given the nature of the available data and the needs and requests of the managers. Thus, the packages tend to have a regional focus for application (see Figure 1.1 for the areas managed by the seven US Regional Fisheries Management Councils), with most packages used exclusively in a single region of the United States. Four of the packages (BSP1, BSP2, BAM, VPA-2BOX, and SS) are used in multiple regions (including HMS), although only Stock Synthesis is used widely in multiple regions and for multiple taxa.

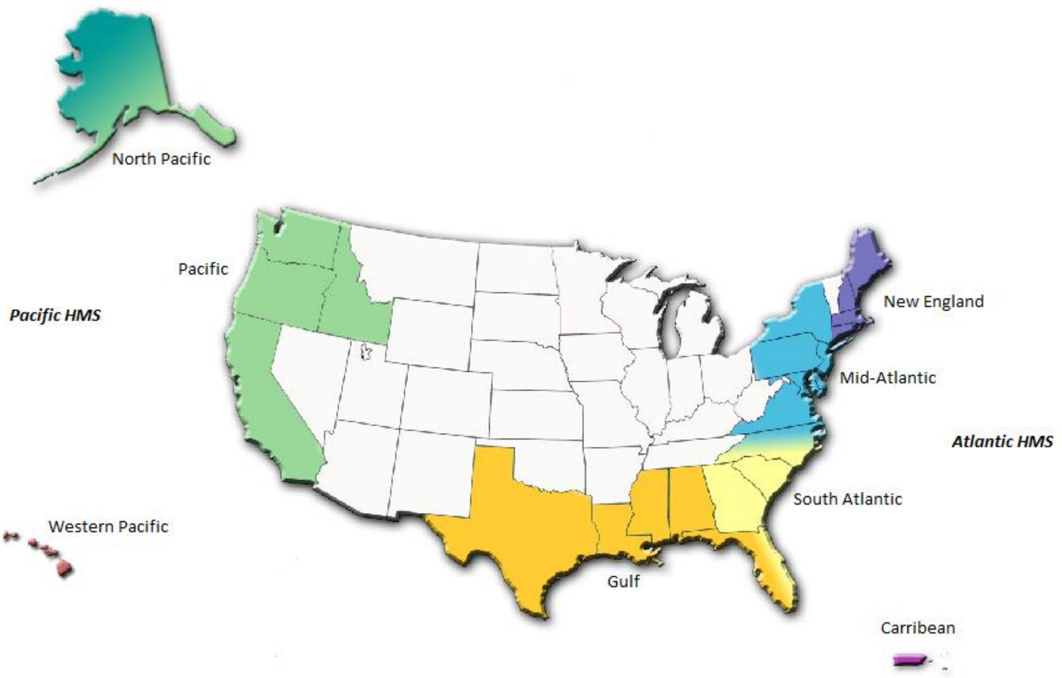

Figure 1.1. The locations of the seven US Regional Fishery Councils (New England, Mid-Atlantis, South Atlantic, Gulf, Caribbean, Pacific, North Pacific, and Western Pacific) . HMS fisheries assessed by US scientists are indicated in bold font.

The packages are based on a variety of population dynamics models. Five of the packages (ASPIC, BSP1, BSP2, DB-SRA, and XDB-SRA) are based on surplus production models, and one of the packages (CSA) is based on a delay-difference model. The remaining packages are based on age-structured population dynamics models. The population models on which three of the packages (SS, SSS and XSSS) are based track both age and sex, although SSS and XSSS are based on the same population dynamics model as SS (Cope, 2013). SS and SSS are linked in Tables 1.1, 1.2 and 1.3 because they use SS as an estimation engine and add methods for quantifying uncertainty in data-poor situations. Only SS (and SSS and XSSS), MULTIFAN-CL, and VPA2-BOX include the ability to represent stocks spatially. None of the packages allow multiple stocks to interact beyond movement, although an extension of AMAK, MALMAK (Kinzey and Punt, 2009) allows multiple species that interact through predation to be assessed simultaneously, and SS has been used with a predator entering into the population dynamics as a pseudo-fishing fleet (Livingston and Methot, 1998).

Catch data are used by all of the packages, with index data (either standardised catch-per-unit-of-effort [CPUE] or survey index data) used in 14 of the 16 packages. The two packages that do not use index data (DB-SRA and SSS) were developed for data-poor situations where the only available information is a time-series of catches. All of the packages can handle situations in which index values are not available for all years of the assessment period. The surplus production models and delay-difference model use only index data, with all but one of the remaining packages able to use data on the age-composition of the catches. Several of the packages can make use of catch and survey size-composition data even though size-dependent mortality is not explicitly modelled in any of the packages, except through the "morphs" feature of Stock Synthesis (Taylor and Methot, 2013) that can capture some aspects of size-selective mortality. The packages without spatial structure tend to represent spatial differences in catch age and size-structure using fleets (the 'areas-as-fleets" approach; Waterhouse et al., 2014; Punt et al., 2014a), although this can lead to biased results (Punt et al., 2015a). Stock Synthesis is the only package that can use conditional age-at-length data to estimate growth curves within the assessment (He et al., In press).

All of the packages have been peer-reviewed in several ways. These include publication in the peer-reviewed literature (Table 1.1) as well as review during stock assessment review meetings. However, most packages are under development to a greater or lesser extent, which means that the reviewed version of the package may not be the one currently available and currently being used.

The performance of most of the packages has been explored using simulation, including comprehensive simulation exercises in which scientists analysing the data sets are unaware of how the data sets were generated so as to avoid overly optimistic results (e.g. NRC, 1998; Carruthers et al., 2014; Deroba et al., 2015). Simulation testing can be in the form of "self tests" where the model being tested is used to generate the data sets or when the model used to generate the data differs from the model being tested (Deroba et al., 2015). "Self tests" are necessary, but insufficient, to fully test a package because they fail to fully explore model mis-specification. However, being unable to provide unbiased estimates of management-related quantities under self-tests suggests that either some of the parameters of the model are confounded, the estimator is statistically inconsistent (as may occur with many asymptotically-unbiased methods when the ratio of data to parameters is low), or there are errors with the software implementation of the model equations. The performance of Stock Synthesis has been explored to the greatest extent, while some of the packages (e.g. BSP1 and BSP2) have only been subject to limited testing.

The evaluation of Stock Synthesis has been facilitated by the development of an R package that can generate pseudo data sets (ss3sim) (Anderson et al., 2014). This package has been the basis for a series of papers evaluating the performance of aspects of Stock Synthesis as well as matters relating to best practices for stock assessment (Table 1.4).

Table 1.1: Overall summaries of the 16 packages considered in this Chapter

Package Name	Primary Developer(s)	US regions of use	Data for use (in addition to catch)	Population dynamic structure	Uncertainty Quantification	Tested using simulation	Primary references
Assessment Method for Alaska (AMAK)	Jim Ianelli	Alaska	Age, Index	Age	Asymptotic, Bayesian	SWG (2010), Kinzey (2010)	Anon (2015)
Age Structured Assessment Procedure (ASAP)	Chis Legault	Northeast, Mid-Atlantic, Southeast	Age, Index	Age	Asymptotic, Bayesian	Brooks et al. (2008)	Legault and Restrepo (1998)
A Stock Production Model Incorporating Covariates (ASPIC)	Michael Prager	HMS	Index	Production model	Bootstrap	Prager et al. (1996); Prager (2002); Williams and Prager (2002)	Prager (1992, 1994, 2002)
Beaufort Assessment Model (BAM)	Erik Williams, Kyle Shertzer	Gulf of Mexico, Southeast, Mid-Atlantic	Age, Size, Index, Discards	Age	Monte Carlo, Bootstrap	Unpublished research	Williams and Shertzer (2015)
Bayesian Surplus Production Model-1 (BSP1)	Jon Brodziak, Yi-Jay Chang	HMS, Northeast, Pacific Islands	Index	Production model	Bayesian	Chang et al. (2014)	Brodziak and Ishimura (2011); Brodziak et al. (2014)
Bayesian Surplus Production Model-2 (BSP2)	Mudoch McAllister, Beth Babcock	HMS, Southeast	Index	Production model	Bayesian SIR algorithm	Unpublished research	McAllister (2014)
Collie-Sissenwine Analysis (CSA)	Jeremy Collie, Michael Sissenwine	Northeast	Index	Delay Difference	Asymptotic and bootstrap with limited Bayesian capability	Mesnil (2003)	Collie and Sissenwine (1983)
Depletion-Based Stock	EJ Dick; Alec MacCall	West coast	None	Production model	Bayesian+	Carruthers et al. (2014)	Dick and MacCall (2011)

Reduction Analysis (DB-SRA)							
MULTIFAN-CL	David Fournier	HMS	Age, Length, Tagging, Index	Age	Asymptotic	Labelle (2005)	Fournier et al. (1998)
Statistical Catch-At-Length (SCALE)	Paul Nitschke	Northeast, Mid-Atlantic	Age, Index	Age	Asymptotic and Bayesian	Brooks et al. (2008)	NOAA Fisheries Toolbox
Stock Synthesis (SS)	Richard Methot	All	Age, Length, Conditional age-at-length, Index, Discards	Age-Size*, Sex, Area	Asymptotic, bootstrap, Bayesian	See below	Methot (1990); Methot and Wetzel (2013)
Simple Stock Synthesis (SSS)	Jason Cope	West coast	None	Age-Size*, Sex Area	Monte Carlo	Cope (2013)	Cope (2013)
Extended Simple Stock Synthesis (XSSS)	Jason Cope; Chantel Wetzel	West coast	Index	Age-Size*, Sex Area	Adaptive Importance Sampling	Wetzel and Punt (In press)	Cope et al. (2015, in press); Wetzel and Punt (In press)
Virtual Population Analysis (VPA)	Many	Northeast	Age, Index	Age	Bootstrap	Brooks et al. (2008)	Gavaris (1988); Conser and Powers (1990)
VPA-2Box	Clay Porch	HMS, Southeast	Age, Index, Tagging	Age	Asymptotic, bootstrap	Porch et al. (1998); Porch (1995), unpublished research	Porch et al. (1995); Porch and Turner (1999); Restrepo and Porch (2000); Porch et al. (2001); Walter and Porch (2012)
Extended Depletion-Based Stock Reduction Analysis (XDB-SRA)	EJ Dick; Alec MacCall, Maria DeYoreo	West coast	Index	Production model	Bayesian	Wetzel and Punt (In press)	Cope et al. (2015);Wetzel and Punt (In press)

+ No index data so the posteriors are priors implied for model output; * Partially through the use of the platoon feature of Stock Synthesis.

Table 1.2. Outputs and projection capability

Package Name	Biomass reference points	Fishing mortality reference points	Catch-based projections	Fishing mortality-based projections	Notes
AMAK	B_{MSY}	F_{MSY}	Yes	Yes	Projection alternatives based on "future" recruitment estimates (REML w/ σ_R)
ASAP	B_{MSY}	F_{MSY} $F_{0.1}$[a], $F_{x\%}$[b]	Yes	Yes	Reference points calculated internally; projections based on the AgePro software
ASPIC	B_0, B_{MSY}, $B_{0.1}$	F_{MSY}, $F_{0.1}$	Yes	Yes	
BAM	B_{MSY}	F_{MSY}, $F_{x\%}$	Yes	Yes	Separate projection routine
BSP1	B_0,B_{MSY}	F_{MSY}	Yes	Yes	Separate projection routine
BSP2	B_0, B_{MSY}	F_{MSY}	Yes	Yes	
CSA	-	-	No	No	
DB-SRA	-	-	Yes	No	
MULTIFAN-CL	B_0, B_{MSY}	F_{MSY}	Yes	Yes (Effort)	
SCALE	-	-	Yes	Yes	Reference points calculated externally; projections based on the AgePro software
SS	B_0, B_{MSY}	F_{MSY}, $F_{x\%}$	Yes	Yes	Many projection options are available
SSS	-	-	No		
XSSS	B_0, B_{MSY}	F_{MSY}	Yes	Yes	
VPA	B_{MSY}	F_{MSY}, $F_{x\%}$	Yes	Yes	Reference points calculated externally; projections based on the AgePro software
VPA-2Box	B_{MSY}, $B_{x\%}$ $B_{0.1}$, B_{Max}	F_{MSY}, $F_{x\%}$, $F_{0.1}$, F_{Max}	Yes	Yes	Projections based on PRO-2BOX
XDB-SRA	B_0, B_{MSY}	F_{MSY}	Yes	No	

a: the fishing mortality rate at which the derivative of yield-per-recruit is 10% of that at F=0; b – the fishing mortality rate at which spawning biomass-per-recruit is reduced to x% of its unfished level.

Table 1.3. Software and maintenance

Package Name	Executable available from	Source code available from	GUI	Primary language	Training available?	Manual and test data sets	Diagnostics produced automatically?	Annual time commitment to maintenance and development (Months)
AMAK	From author; https://github.com/NMFS-toolbox/AMAK	https://github.com/NMFS-toolbox/AMAK	Yes	ADMB	No	No	Plots of fits to data; effective sample sizes; residual variances	1
ASAP	NOAA Fisheries Toolbox	NOAA Fisheries Toolbox	Yes	ADMB	Yes	Yes	Plots of fits to data; effective sample sizes; residuals, residual time-series; multipliers for multinomial distributions	1
ASPIC	www.mhprager.com	Form author	Yes	Fortran95	Yes	Yes	Goodness-of-fit information, fits to data, residual time-series, plots, convergence diagnostics	0.25
BAM	From author	From author	No	ADMB	No	Yes	Output is handled through an R package (FishGraph) which provides diagnostics and output graphics and analyses	0.5
BSP1	No	From author	Yes	WINBugs and R	No	No	DIC, RMSE, Standardized CPUE Residuals, Posteriors, various MCMC diagnostics	1.5
BSP2	From Author	From author	No	Visual Basic	Yes	No	MPD fit to index data; convergence diagnostics; residual variances; plot of mpd estimates of process error deviates by year	consultancy
CSA	NOAA Fisheries Toolbox	NEFSC	Yes	ADMB	No	Yes	Residuals	0.1
DB-SRA	N/A	Github	Yes	R	No	No	Posteriors for model outputs only	0.25

MULTIFAN-CL	http://www.multifan-cl.org/	No	No	C++	Yes	Yes	Residuals; catchability and effort deviates; retrospective analysis (R package R4MFCL)	18
SCALE	NOAA Fisheries Toolbox	NOAA Fisheries Toolbox	Yes	ADMB	Yes	Yes	Fits to catch and index data; effective sample sizes, residuals; multipliers for multinomial distributions.	0.5
SS	NOAA Fisheries Toolbox	From author	Yes	ADMB	Yes	Yes	R package R4ss	15[a]
SSS	https://github.com/shcaba/SSS	No	No	R, ADMB	Yes	No	Posteriors for model outputs only	0.25
XSSS	No	No	No	R, ADMB	Yes	No	Predictive checks and posteriors	0.25
VPA	NOAA Fisheries Toolbox	NEFSC	Yes	C	Yes	Yes	Fits to index data	0.5
VPA-2Box	http://www.iccat.int/en/AssessCatalog.htm	http://www.iccat.int/en/AssessCatalog.htm	Yes	FORTRAN	Yes	Yes	Fits to index and mark-recapture data	0.25
XDB-SRA	N/A	Github	No	R	No	No	Predictive checks and posteriors	0.25

a - NOAA Fisheries Toolbox: http://nft.nefsc.noaa.gov/; b – does not include the time dedicated by non-NMFS researchers to the development of r4ss and the testing of SS using simulation

1 – Interactive GUI for running BSP-1 is available using the WinBUGS14 GUI, available at http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/

Table 1.4. Examples of where Stock Synthesis has formed the basis for a simulation study to evaluate its ability to estimate quantities of management interest.

Reference	General topic
Bence et al. (1993)	Selectivity estimation
Sampson (1993)	Selectivity estimation
Sampson and Yin (1998)	General estimation issues
Helu et al (2000)	Selectivity estimation
Yin and Sampson (2004)	General estimation issues
Stewart and Piner (2007)	Estimation of ageing bias
Schirripa et al. (2009)	Climate drivers of recruitment
Cope and Punt (2011)	Spatial structure and catch trends
Garrison et al. (2011)	Spatial structure and MPAs
Lee et al. (2011)	Natural mortality estimation
Piner et al. (2011)	Steepness and selectivity misspecification
Methot and Taylor (2011)	Stock-recruitment bias
Wetzel and Punt (2011)	Sample sizes length and age samples
Lee et al. (2012)	Steepness estimation
Maunder (2012)	Stock–recruitment relationship and management reference points
Stewart et al. (2013)	Comparison of maximum likelihood and Bayesian methods
Wayte (2013)	Performance under feedback control: Recruitment shifts
Crone and Valero (2014)	Selectivity estimation
Hurtado-Ferro et al. (2014)	Spatial structure and selectivity
Hurtado-Ferro et al. (2015)	Retrospective patterns
Johnson et al. (2015)	Natural mortality estimation
McGilliard et al. (2015)	Spatial structure and MPAs
Ono et al. (2015)	Sample sizes length and age samples
Punt et al. (2015a)	Spatial assessment configurations
He et al. (in press)	Growth estimation

OUTPUT STATISTICS, PROJECTIONS, AND QUANTIFICATION OF UNCERTAINTY

All of the packages are able to estimate time-trajectories of biomass and fishing mortality. Most of the packages based on age-structured population dynamics models are able to represent multiple fleets, each with a different selectivity pattern, and hence estimate the biomass available to each fleet as well as fleet-specific trends in fishing mortality.

All of the packages, except CSA, the two methods designed for data-poor situations (DB-SRA and SSS), and SCALE, provide reference points used for management decision making (or in the case of the VPA package can be linked to software that allows reference points to be computed) (Table 1.2). The most common reference points computed using the packages are the biomass associated with maximum sustainable yield (B_{MSY}) and the corresponding fishing mortality (F_{MSY}). The choice of these reference points is not unexpected because, under the guidelines associated with the US Magnuson-Stevens Fishery Conservation and Management Act (MSA)[3], "overfishing" can be defined as fishing mortality exceeding F_{MSY}. A stock is defined as being "overfished" when its biomass is reduced below the Minimum Stock Size Threshold (MSST), where MSST is defined in terms of B_{MSY} or a proxy for B_{MSY} such as 40% of the unfished biomass, 0.4 B_0 (Punt et al., 2014b). Given that B_{MSY} is often estimated using proxy methods, generally some fraction of B_0, several of the packages also estimate B_0.

All of the packages except CSA have a projection capability. Most of the packages have an inbuilt ability to conduct projections. Three of the packages (ASAP, SCALE, and VPA) conduct projections using a separate piece of software AgePro (http://nft.nefsc.noaa.gov/AGEPRO.html) while the VPA-2BOX package conducts projections using the PRO-2BOX software (http://www.iccat.es/en/AssessCatalog.htm), and ASPIC provides projections through ASPICP. All of the packages that can conduct projections can explore the consequences of future time series of catches, often by fleet. Fewer packages can conduct projections based on specifying future time-trajectories of fishing mortality or effort (Table 1.2).

All of the packages are able to quantify uncertainty, although the approaches differ among packages. The most common way of exploring uncertainty is through the use of sensitivity tests, in which some of the assumptions of the assessment and / or data sources are changed, and using retrospective analysis (Butterworth, 1981; Mohn, 1999). Sensitivity tests can be conducted fairly straightforwardly for those packages with Graphical User Interfaces (GUIs). The "statistical" uncertainty associated with model outputs quantifies the

---

[3] http://www.fisheries.noaa.gov/sfa/laws_policies/national_standards/index.html

uncertainty reflected in the fit of the model to the data, conditional on the model being correct. The packages variously compute "statistical" uncertainty using asymptotic methods, bootstrapping, and Bayesian approaches.

SOFTWARE ISSUES

The packages are generally available, many through the NOAA Fisheries Toolbox (NFT)[4] (ASAP, ASPIC, CSA, SCALE, VPA, VPA-2BOX). Executables, Graphical User Interfaces (when they exist), and the source code for some of the packages are available directly from the authors or from public code locations (Table 1.3). The availability of the source code for packages such as ASAP and AMAK increases the ability of peer reviewers to fully understand the assumptions and technical aspects of the methods underlying the package. Graphical User Interfaces have been developed for most of the packages, and some form of training and an instruction manual is available for most of packages.

Effective use of the packages relies on being able to evaluate model fit using diagnostic statistics and plots, ideally automatically. All of the packages provide diagnostics ranging from tables of residuals and model outputs to the extensive R package r4ss (Taylor et al., 2014) that is capable of producing HTML and PDF files with key outputs and diagnostics, and is continually under development. Similarly, BAM generates R output using the package ADMB2R (Martin et al., 2006), and from that R output, detailed automated graphics using the R code in Fishgraph (Prager et al., 2015). ASPIC generates R output similarly, and several analysts have developed R code for automatically graphing from that output (MH Prager, pers. obs.). The availability of diagnostics packages has spurred research into diagnostics, as the time spent developing diagnostics will be used by a wide community of assessment analysts.

The primary cost associated with the packages is the initial development time, but all of the packages require ongoing modification, development and maintenance (Table 1.3). While the time commitment to maintain some individual packages is relatively minor (less than a month for several packages), maintaining all of the packages is a substantial investment for the stock assessment community that in many cases is considered a core activity for the developers.

---

[4] http://nft.nefsc.noaa.gov/

DISCUSSION

Within the US, the bulk of the assessments that estimate time-trends in fishing mortality and biomass on which the Report to Congress[5] is based are now conducted using stock assessment packages. This is also the case in Europe and New Zealand (see below). However, as shown in Table 1.1, the choice of package is very region-specific. This reflects several factors, including the locations where the authors of the packages are based, the history of how assessments are conducted in regions around the US, and the quirks of data availability specific to certain regions and key stocks. For example, the use of "VPA-type" methods has been, and continues to be, common on the east coast of the US (and Canada) because one of the key assessment packages (ADAPT; Gavaris, 1988) was developed there. In contrast, "integrated methods" are used almost to the exclusion of other methods on the US (and Canadian) west coasts and Alaska due to the fact that several initial developments of the integrated approach occurred on the west coast of North America (e.g. Fournier and Archibald, 1982; Deriso et al., 1985; Methot, 1990), probably increasing the rate of adoption of the approach. The use of integrated methods (particular ASAP, BAM and SS) is becoming more common on the east coast of the US.

The choice of a package also reflects data availability. For example, packages such as SS appear to perform best when there are estimates of catch going back to the start of the fishery, and this is the case for many of the fisheries along the west coast of North America and Alaska. In contrast, the fisheries along the U.S. and Canadian east coast tend to be much older and estimates of the historical catch, when available, are often imprecise (e.g. Rosenberg et al., 2005). Moreover, catch age-composition data, which are required for VPA, have been available for decades for several of the key fish species off the US east coast. In contrast, the collection of age-composition data for species off the west coast of the US only started more recently, was often sporadic, and has been subject to revisions to ageing methods and substantial differences in age-estimates among labs (e.g., Beamish, 1979; Stawitz et al., 2015).

The lack of reliable age-composition data for many highly migratory species (e.g., billfishes and sharks) is reflected by the still widespread use of surplus production methods for these species (Table 1.1). While VPA-type methods are also employed for some HMS, the required catch age-compositions are often derived from length-frequency data by use of "cohort slicing" algorithms, which tend to perform poorly for older, slower growing fish (ICCAT, 2011; Kell and Kell, 2011). The increased recognition of the uncertainties associated with cohort slicing, along with improvements in the size and age information for some

---

[5] http://www.nmfs.noaa.gov/sfa/fisheries_eco/status_of_fisheries/

species (and often tagging data), is leading to more widespread adoption of integrated methods (even for fairly data-poor groups such as billfishes; Punt et al. 2015b).

Within the US, the only region that still predominantly uses stock-specific software developed by individual assessment analysts is Alaska, and even in that region there is a trend towards use of packages, in particular Stock Synthesis and AMAK.

The overall trend throughout the US is towards integrated methods, in particular because they can more easily use a wide variety of data sources, they provide estimates of uncertainty fairly straightforwardly using Bayesian methods (and hence can more easily include prior information; Stewart et al., 2013), and because they can "scale" from data-poor to data-rich situations (Cope, 2013[6]).

ADVANTAGES AND DISADVANTAGES OF PACKAGES

The advantages of the use of packages rather than user-developed software includes: (a) substantially-increased flexibility to explore alternative assessment configurations, if the options to be explored are implemented in the package, (b) ease of peer-reviewing (if the peer-reviewers are familiar with the package) because reviewers can request and see results displayed 'on the fly' and can examine the input files and hence possibly detect errors in the way the model is specified, (c) increased confidence that the assessment is correctly coded and tested, (d) the availability of tools to explore uncertainty and summarize model fits to data, (e) increased collaboration amongst assessments scientists among regions of the US (and the world) on generic questions related to stock assessment practice given a common software platform is being used, (f) faster development time for an assessment, (g) increased ability for a new analyst to take over a stock assessment because they are familiar with the basis for the assessment, and (h) a large user base to facilitate further development and improvement (and to detect errors). Many of the simulation studies to evaluate the performance of methods of stock assessments have been based on packages (see Table 1.4 for such analyses based on SS). This is because the use of packages allows more assessment options to be explored rapidly and without the need to code the assessment configurations from scratch (e.g. different ways to implement spatial models; Punt et al. 2015a). In addition, by basing simulation studies on an existing package, there is a much higher chance that the results of the simulation will be considered relevant by assessment analysts.

The use of packages for stock assessment purposes is not without problems. The largest problems with the use of packages are: (a) difficulties implementing new options and ideas,

[6] Punt (1994) referred to an age-structured integrated analysis model with deterministic recruitment and fitted only to index data as an "age structured production model" to reflect that, although the model was mathematically complex, it was effectively no more complex than a surplus production model.

(b) errors in the package may impact several assessments, and (c) the possibility that the package is treated as a "black box" and used with insufficient thought by inexperienced or untrained assessment analysts. An advantage of stock-specific software for conducting assessments is that a suitably skilled analyst can include aspects of the assessment that are unique to the problem at hand. For example, the assessment of gummy shark (*Mustelus antarcticus*) off Australia is still based on a stock-specific assessment method (Pribac et al., 2005) because (to date) no package exists that simultaneously allows for multiple stocks, sharing of parameters among stocks, tagging data and a non-linear relationship between stock biomass and CPUE. Thus, the move to packages does not preclude the need for some stock-specific assessment tools. In principle, all of the packages can be updated to reflect conditions that are unique to a given stock (or set of stocks). However, updating packages tends to be more difficult than updating a piece of software developed by an individual assessment analyst. In addition, developers and users of packages tend to expect that new additions to packages are tested before they are made generally available. For packages with many users and potentially many suggested changes (e.g. Stock Synthesis), there is a need for careful version control. Thankfully, version control packages (such as GitHub https://github.com/) are now available to assist with this.

The "black box" problem should not be under-stated. Most of the packages reviewed will provide an answer that will look correct, and in package with many options, the possibility of mistakes by non-experts and experts alike remains a concern. The packages differ in the types of data they can use and their complexity. Packages with the most options and greatest flexibility (such as Stock Synthesis and MULTIFAN-CL) are the most difficult to use and hence likely most susceptible to mis-use. The large number of model specifications available in the most sophisticated packages can also lead to mistakes, even from experts. In some cases, poor documentation contributes to this problem. While a "terms of reference" guideline for assessments, multiple assessment authors for one assessment, along with peer-review of resulting assessments, can help, the importance of adequate training of assessment analysts cannot be over-emphasized. As new features are added to packages, training must be ongoing. Development of best practice guidelines for the use of stock assessment packages, particularly those that have many options, will also help to ensure that the "black box" problem is avoided. The Center for Advancement of Population Assessment Methodology (CAPAM) has helped in this regard by holding workshops (to date on selectivity and growth) that aim to produce best practice guidelines (e.g. Crone et al., 2013).

There can be substantial cost associated with the development and maintenance of packages. However, compared to the other costs of conducting assessments such as data collection, staff time to conduct assessments (and potentially coding, testing and documenting the assessment), and the cost of conducting a peer review, the cost of these packages is relatively minor. If cost-benefit analyses are to be conducted for the development (and use) of stock assessment packages, the costs need to be placed in the context of the total cost of stock assessment.

BEYOND THE US

This study has focused on the US because of the relatively well-specified goals and objectives for US assessments. Several other jurisdictions have developed packages. For example, International Council for the Exploration of the Sea (ICES) assessments are conducted using a different set of methods and hence stock assessment packages. The SAM (Nielsen and Berg, 2014) method is a flexible platform which, unlike most integrated analysis approaches, formally integrates out the random effects (rather than adopting an errors-in-variables estimation approach). A random effects model allows the parameters that quantify the extent to which parameters vary over time to be quantified and avoids the statistical inconsistency that occurs when the number of parameters increases with the number of years of data. Other packages used in the ICES arena are SURBA (Needle, 2003), XSA (Shepherd, 1999), and ICA (Patterson and Melvin, 1996).

Most of the assessments in New Zealand are conducted using CASAL (Bull et al., 2005). This package includes many of the features of Stock Synthesis. However, to date, the performance of this package has not been explored using simulation and it has not been widely used outside of New Zealand.

The most obvious exception to the use of packages is South Africa where the integrated analysis approach has been adopted for all key species, but where each assessment is conducted using stock-specific software products (although coded using AD Model Builder (Fournier et al., 2012), similarly to many stock assessment packages in the US). The primary reason for this is that each stock under assessment has unique features (e.g. multiple species assessed as one species, complex spatial structure, etc).

PEER-REVIEW AND PACKAGES

In the US, the Magnuson-Stevens Act National Standard Guidelines require that management decisions are based on the best scientific information available (U.S. Doc, 2007). Each Regional Fishery Management Council must have a peer-review process, and the 2006 reauthorization of the MSA stipulated that the Scientific and Statistical Committee for each Council can be involved in that process. Conducting and reviewing a “benchmark” assessment (Methot, 2015) is a substantial exercise that involves *inter alia* reviewing data inputs, selecting a model and assessment platform, identifying one or more best assessment configurations, evaluating the plausibility and reasonableness of the results, and providing an overall summary of the assessment, including characterising uncertainty (see, for example, PFMC 2014).

The use of packages for conducting stock assessments provides a number of key benefits to reviewers: (a) an assessment platform that has been tested and for which performance, at least for a range of situations, is known, (b) improved documentation, (c) a GUI or a fairly simple to use input-output system allowing reviewers to check that the input has been specified as documented and also potentially to conduct sensitivity analyses themselves, and (d) a quick way to create diagnostics that should be understandable by the reviewers. In the experience of the authors of this Chapter, the availability of packages (rather than user-written software products) can substantially reduce the time needed to review an assessment.

NEXT STEPS

The ability to assess multiple species simultaneously is a desirable feature of an assessment package as it then becomes possible to share quantities such as trends in fishing mortality among stocks (c.f., Punt et al. 2011; Thorson et al., 2014), although only VPA2-BOX can do this and only approximately. None of the packages allow for multispecies interactions. In addition, very few of the packages can currently represent spatial structure explicitly (most can, however, represent spatial dynamics using the areas-as-fleets approach, though caution should be used in this approach (Cope and Punt, 2011; Punt et al., 2015a)). Of the packages, only Stock Synthesis can allow most of the parameters (including growth and selectivity) to change over time and for those parameters to be driven as a function of environmental conditions. However, best practices for including environmental drivers of model parameters into an assessment is still an active area of research (e.g. Maunder and Watters, 2003b; Haltuch and Punt, 2011).

The most obvious gap in the packages used in the US is the lack of packages to conduct assessments for animals that cannot be aged and for which the primary data source is size-compositions. Hence, it may be most appropriate to base the dynamics on a size- rather than an age-structured population dynamics model. Several methods for conducting assessments of such stocks have been developed (Punt et al., 2013), and assessments for several stocks in the US are based on such models (e.g. Jacobson, 2010; Zheng and Siddeek, 2014; Turnock and Rugulo, 2014), but to date none of those assessments have been coded in a package that is generally available. Packages based on size-structured models (e.g. the Generic Size-structured Stock Assessment Model for Alaskan Crab Stocks [GMACS][7]) are under development, but have not been used for any assessments to date.

All of the packages considered in this Chapter have been evaluated to some extent using simulation. However, the effort needed to conduct the simulation evaluations of estimation

[7] https://github.com/seacode/gmacs

performance and Management Strategy Evaluation more generally can be substantial. The development of generic tools to support simulation evaluation (particularly of new packages and new options within existing packages) appears to be a key need. The work of Anderson et al. (2014) is an important start in this direction, but ss3sim is tailored to only one package (SS) and can only explore a limited set of options. This package also does not have an automatic way to evaluate feedback control management strategies. The POPSIM package[8] can be used to create inputs for VPA, CSA, ASPIC and ASAP, but this package is also somewhat limited in the scenarios that can be explored.

FINAL COMMENTS

The development of packages to conduct stock assessments has led to an increased ability to conduct assessments using methods that have been evaluated, have been used by several analysts, and are generally well-documented. The assessments conducted using such packages are less likely to be subject to coding errors, meaning that the peer-review process can focus on evaluating whether the chosen assessment configuration is valid and justifiable. This review process is facilitated by the availability of standard diagnostics for evaluating model fit and performance. The evolution of assessment packages also benefits from repeated and varied uses and users. The packages reviewed in this Chapter are based (to a greater or lesser extent) on flexible models, allowing the assessment analysts the ability to rapidly explore the sensitivity of assessment results to alternative model assumptions and data set choices, in contrast to the past when even slight changes to a model configuration would have led to (usually overnight) changes to the model code.

Assessment packages are, of course, not without problems, including ensuring that changes to the package are made such that the package grows with the needs of the assessment community, and that there is confidence that all of the options perform as expected. Another challenge is the production of clear, concise documentation, which must be kept up to date as the corresponding package evolves. The reliance on a small number of developers who ensure that the packages are maintained is a long-term challenge for the use of packages for stock assessment, especially as several of the authors of the packages in Table 1.1 have retired (or are close to retirement age). The need for a good understanding of the assumptions underlying the models in a package can never be eliminated, but by using packages with known properties, the possibility of detecting errors is increased.

As a next generation of models is developed, it seems advantageous to collaboratively undertake this development using a team with expertise in population dynamics, statistics, programming, and graphical interfaces. Some aspects of such an approach is underway in

[8] http://nft.nefsc.noaa.gov/POPSIM-A.html; http://nft.nefsc.noaa.gov/POPSIM-B.html

the development of the GMACS size-based model (https://github.com/seacode/gmacs) and is envisioned for a next generation of Stock Synthesis.

It is clear that the benefits of the investment by the NMFS in the development of packages more than offsets the associated costs, and we expect packages to continue to be a core component of stock assessment science. There is value in following in the footsteps of other disciplines such as weather forecasting that have developed a rigorous process for research development, testing against a testbed, and then moving into operational status. The future for packages in the US should see packages that allow scientists to analyse multiple stocks simultaneously and account for predation and environmental effects. However, the danger will then be (as it sometimes is now) that modelling frameworks can be far more advanced than the data on which any assessment could be based!

REFERENCES

Anderson, S.C., Monnahan, C.C., Johnson, K.F., Ono, K., Valero, J., 2014. ss3sim: an R package for fisheries stock assessment simulation with Stock Synthesis. Plos One 9, e92725.

Anon. 2015. Assessment Model for Alaska Description of GUI and Instructions. Available from https://github.com/NMFS-toolbox/AMAK/blob/master/docs/AMAK%20Documentation.pdf (last accessed 6 August 2015)

Beamish, R.J., 1979. Differences in the age of Pacific hake (*Meluccius productus*) using whole ololiths and sections of otoliths. J. Fish. Res. Board. Can. 36, 141–151.

Bence, J.R., Gordoa, A., Hightower, J.E., 1993. Influence of age-selective surveys on the reliability of stock synthesis assessments. Can. J. Fish. Aquat. Sci. 50, 827–840.

Berkson, J., Hunt, K.M., Whitehead, J.C., Murie, D.J., Kwak, T.J., Boreman, J., 2009. Is there a shortage of fisheries stock assessment scientists? Fisheries 34, 217–219.

Brodziak, J., Ishimura, G., 2011. Development of Bayesian surplus production models for assessing the North Pacific swordfish population. Fish. Sci. 77, 22–34

Brodziak, J., Yau, A., O'Malley, J., Andrews, A., Humphreys, R., DeMartini, E., Pan, M., Parke, M., Fletcher, E., 2014. Stock assessment update for the main Hawaiian Islands Deep 7 bottomfish complex through 2013 with projected annual catch limits through 2016. U.S. Dep. Commer., NOAA Tech. Memo., NOAA-TM-NMFS-PIFSC-42, 61 pp .

Brooks, L., Legault, C., Nitschke, P., O'Brien, L., Sosebee, K., Rago, P., Seaver, A., 2008. Evaluation of NMFS Toolbox Assessment Models on Simulated Groundfish Data Sets. http://www.nefsc.noaa.gov/program_review/background2014/TOR1Legault.pdf (last accessed 8 August 2015).)

Bull, B., Francis, R.I.C.C., Dunn, A., McKenzie, A., Gilbert, D.J., Smith, M.H., 2005. CASAL (C++ algorithmic stock assessment laboratory): CASAL user manualv2.07-2005/08/21. NIWA Technical Report 127.

Butterworth, D.S., 1981. The value of catch-statistics-based management techniques for heavily fished pelagic stocks with special reference to the recent decline of the southwest African pilchard stock. In Applied Operations Research in Fishing, pp. 441–464. Ed. by Haley . Springer, New York, US.

Butterworth, D.S., Andrew, P.A., 1984. Dynamics catch-effort models for the hake stocks ICSEAF Divisions 1.3-2.2 Colln scient. Pap. Int. Commn SEL Atl. Fish. 11, 29–58.

Carruthers, T.R., Punt, A.E., Walters, C.J., MacCall, A., McAllister, M.K., Dick, E.J., Cope, J., 2014. Evaluating methods for setting catch limits in data-limited fisheries. Fish. Res. 153, 48–68.

Chang, Y-J., Brodziak, J., O'Malley, J., Lee, H-H, DiNardo, G., Sun, C-L. 2014. Model selection and multimodel inference for fishery surplus production models: A case study for Pacific blue and striped marlins. Fish. Res. 166, 129–139

Collie, J.S., Sissenwine, M.P., 1983. Estimating population size from relative abundance data measured with error. Can. J. Fish. Aquat. Sci. 40, 1871–1879.

Conser, R., Powers, J.E., 1990. Extensions of the ADAPT VPA tuning method designed to facilitate assessment work on tuna and swordfish stocks. Col. Vol. Sci. Pap. ICCAT. 32, 461-467.

Cope, J.M., 2013. Implementing a statistical catch-at-age model (Stock Synthesis) as a tool for deriving overfishing limits in data-limited situations. Fish. Res. 142, 3–14.

Cope, J.M., Punt, A.E., 2011. Reconciling stock assessment and management scales under conditions of spatially varying catch histories. Fish. Res. 107, 22-38.

Crone, P.R., Valero, J.L., 2014. Evaluation of length- vs. age-composition data and associated selectivity assumptions used in stock assessments based on robustness of derived management quantities. Fish. Res. 158, 165–171.

Cope, J., Dick, E.J., MacCall, A., Monk, M., Soper, B., Wetzel, C., 2015. Data-Moderate Stock Assessments for Brown, China Copper, Sharpchin, Stripetail, and Yellowtail Rockfishes and English and Rex Soles in 2013. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220. 298 pp. http://www.pcouncil.org/wp-content/uploads/Data-Moderate_Assessments_2013_FINAL_160116.pdf.

Cope, J.M., Piner, K., Minte-Vera, C.V., Punt, A.E., 2003. Status and Future Prospects for the Cabezon (*Scorpaenichthys marmoratus*) as Assessed in 2003. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220. 158pp,

Cope, J.M., Thorson, J.T., Wetzel, C.R., DeVore, J., In press. Evaluating a prior on relative stock status using simplified age-structured models. Fish. Res. 00, 00-00.

Crone, P.R., Maunder, M.N. Valero, J.L. McDaniel, J.D. and B.X. Semmens (Editors). 2013. Selectivity: theory, estimation, and application in fishery stock assessment models. Workshop Series Report 1. Center for the Advancement of Population

Assessment Methodology (CAPAM). NOAA/IATTC/SIO, 8901 La Jolla Shores Dr., La Jolla, CA 92037. 46 p

Department of Commerce (DoC). 2008. The shortage in the number of individuals with post-baccalaureate degrees in subjects related to fishery science. NOAA Technical Memorandum NMFS-F/SPO-91. 84pp.

Deriso, R.B., Quinn II, T.J., Neal, P.R., 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42, 815–824.

Deroba, J.J., Butterworth, D.S., Methot, R.D. Jr., De Oliveira, J.A.A., Fernandez, C., Nielsen, A., Cadrin, S.X., Dickey-Collas, M., Legault, C.M., Ianelli, J., Valero, J.L., Needle, C.L., O'Malley, J.M., Chang, Y-J., Thompson, G.G., Canales, C., Swain, D.P., Miller, D.C.M., Hintzen, N.T., Bertignac, M., Ibaibarriaga, L., Silva, A., Murta, A., Kell, L.T., de Moor, C.L., Parma, A.M., Dichmont, C.M., Restrepo, V.R., Ye, Y., Jardim, E., Spencer, P.D., Hanselman, D.H., Blaylock, J., Mood, M., Hulson, P.-J.F., 2015. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES J. Mar. Sci. 72, 19–30.

Dichmont, C.M., Punt, A.E., Deng, A., Dell, Q., Venables, W., Dichmont, C.M., Punt, A.E., Deng, A., Dell, Q., Venables, W., 2003. Application of a weekly delay-difference model to commercial catch and effort data in Australia's Northern Prawn Fishery. Fish. Res. 65, 335–350.

Dick, E.J., MacCall, A.D., 2011. Depletion-Based Stock Reduction Analysis: A catch-based method for determining sustainable yields for data-poor fish stocks. Fish. Res. 110, 331–341.

Fournier, D., Archibald, C.P., 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39, 1195–1207.

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

Fournier, D.A., Skaug, H.G., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., Sibert, J., 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27, 233–249.

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124–1138.

Garrison, T.M., Hamel, O.S., Punt, A.E., 2011. Can data collected from marine protected areas improve estimates of life-history parameters? Can. J. Fish. Aquat. Sci. 68, 1761-1777.

Gavaris, S., 1988. An Adaptive framework for the estimation of population size. CAFSAC Res. Doc. 88/29. 12pp.

Haltuch, M.A., Punt, A.E., 2011. The promises and pitfalls of including decadal scale climate forcing of recruitment in groundfish stock assessment. Can. J. Fish. Aquat. Sci. 68, 912–926.

He, X., Field, J.C., Pearson, D.E., Lefebvre, L.S., In press. Age sample sizes and their effects on growth estimation and stock assessment outputs: Three case studies from U.S. West Coast fisheries. Fish. Res. 00, 00–00.

Helu, S.L., Sampson, D.B., Yin, Y., 2000. Application of statistical model selection criteria to the Stock Synthesis assessment program. Can. J. Fish. Aquat. Sci. 57, 1784–1793.

Hilborn, R., Maunder, M.N., Parma, A., Ernst, B., Paynes, J., Starr, P.J., 2000. Documentation for a general age-structured Bayesian stock assessment model: code named Coleraine. Fisheries Research Institute, University of Washington, Seattle, Washington (FRI/UW 00/01).

Hurtado-Ferro, F., Punt, A.E., Hill, K.T., 2014. Use of multiple selectivity patterns as a proxy for spatial structure. Fish. Res. 158, 102–115.

Hurtado Ferro, F, Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R.R., McGillaird, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vertpre, K.A., Whitten A.R., Punt, A.E., 2015. Looking in the rearview mirror: bias and retrospective patterns in integrated, agestructured stock assessment models. ICES J. Mar. Sci. 82, 99–110.

ICCAT, 2011. Report of the 2010 Mediterranean swordfish stock assessment meeting. Coll. Vol. Sci. Pap. ICCAT 66, 1405–1470.

Jacobson, L. 2010. Appendix B11: Technical documentation for the CASA length structured stock assessment model. In "Northeast Fisheries Science Center (NEFSC). 2010. 50th Northeast Regional Stock Assessment Workshop (50th SAW) Assessment Report. Northeast Fish Sci Cent Ref Doc. 10-17. http://www.nefsc.noaa.gov/publications/crd/crd1017/pdfs/appb11.pdf (last accessed 25 August 2015)

Johnson, K.F., Monnahan, C.C., McGillaird, C., Vertpre, K., Anderson, S.C., Cunningham, C., Hurtado-Ferro, F., Licandeo, R., Muradian, M., Ono, K., Szuwalski, C., Valero, J., Whitten A., Punt, A.E., 2015. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. ICES J. Mar. Sci. 72, 137–150.

Kell, L.T., Kell, A., 2011. A comparison of age slicing and statistical age estimation for Mediterranean swordfish (*Xiphias gladius*). Coll. Vol. Sci. Pap. ICCAT 66, 1522–1534.

Kinzey, D., 2010. Multispecies stock assessment with predator-prey interactions. PhD thesis, University of Washington.

Kinzey, D., Punt, A.E., 2009. Multispecies and single-species age-structured models of fish population dynamics: Comparing parameter estimates. Nat. Res. Mod. 22, 67–104.

Labelle, M., 2005. Testing the MULTIFAN-CL assessment model using simulated tuna fisheries data. Fish. Res. 71, 311–334.

Laurec, A., and Shepherd, J.G., 1983. On the analysis of catch and effort data. J. Cons. Int. Explor. Mer. 41, 81–84.

Lee, H.H., Maunder, M.N., Piner, K.R., Methot, R.D., 2012. Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? Fish. Res. 113, 254–261.

Lee, H.H., Maunder, M.N., Piner, K.R., Methot R.D., 2011. Estimating natural mortality within a fisheries stock assessment model: An evaluation using simulation analysis based on twelve stock assessments. Fish. Res. 109, 89–94.

Legault, C.M., Restrepo, V.R., 1998. A flexible forward age-structured assessment program. Col. Vol. Sci. Paper. ICCAT 49, 246–253.

Little, L.R., Kerrigan, B., Thébaud, O., Campbell, A., Innes, J., Cameron, D., Norman-López, A., Mapstone, B.D., Punt, A.E., Hatfield, B., Tickell, S., Kung, J., Slade, S., Leigh, G., O'Neil, M., Tobin, A., 2015. Evaluating candidate monitoring strategies, assessment procedures and harvest control rules in the spatially complex Queensland Coral Reef Fin-fish Fishery. Report of FRDC Project 2011/030. 198pp.

Livingston, P.A., Methot, R.D., 1998. Incorporation of predation into a population assessment model of eastern Bering Sea walleye pollock. In Fishery Stock Assessment Models edited by F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Zhang. Alaska Sea Grant AK-SG-98-01. pp 663-678.

Mace, P.M., Bartoo, N.W., Hollowed, A,B., Kleiber, P., Methot, R.D., Murawski, S.A., Powers, J.E., Scott, G.P., 2001. National Marine Fisheries Service Stock Assessment Improvement Plan. Report of the NMFS National Task Force for Improving Fish Stock Assessments. NOAA Technical Memorandum NMFS-F SPO-56. 76pp.

Magnusson, A., Hilborn, R., 2007. What makes fisheries data informative? Fish and Fish. 8, 337–358.

Magnusson, A., Punt, A.E., Hilborn, R., 2013. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. Fish and Fish. 14, 325–342.

Martin, J.L., Prager, M.H., Stephens, A., 2006. User's guide to ADMB2R: A set of AD Model Builder output routines compatible with the R statistics language. NOAA Technical Memorandum NMFS-SEFSC-546.

Maunder, M., 2012. Evaluating the stock–recruitment relationship and management reference points: Application to summer flounder (*Paralichthys dentatus*) in the U.S. mid-Atlantic. Fish. Res. 125-126, 20–26.

Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142, 61–74.

Maunder, M.N., Watters, G.M., 2003a. A-SCALA: an age-structured statistical catch-at-length analysis for assessing tuna stocks in the Eastern Pacific Ocean. Inter-Am. Trop. Tuna Comm. Bull. 22, 433–582.

Maunder, M.N., Watters, G.M., 2003b. A general framework for integrating environmental time series into stock assessment models: model description, simulation testing, and example. Fish Bull. 101, 89–99.

McAllister, M.K., 2014. A generalized Bayesian surplus production stock assessment software (BSP2). Col. Vol. Sci. Pap. ICCAT. 70, 1725–1757.

McGilliard, C.R., Punt, A.E., Methot, R.D. Jr, Hilborn, R., 2015. Accounting for marine reserves using spatial stock assessments. Can. J. Fish. Aquat. Sci. 72, 262–280.

McKenzie, A., 2015. Assessment of hoki (*Macruronus novaezelandiae*) in 2014. NZ Fish Ass Rep. 2015/09. 68pp.

Mesnil, B., 2003. The Catch-Survey Analysis (CSA) method of fish stock assessment: an evaluation using simulated data. Fish. Res. 6, 193–212

Methot, R.D., 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. Bull. Int. North Pacific Fish. Comm. 50, 259–277.

Methot, R.D., Jr (Ed) 2015. Prioritizing Fish Stock Assessments. NOAA Technical Memorandum NMFS-F/SPO-152. 31pp.

Methot, R.D. Jr., Taylor, I.G., 2011. Adjusting for Bias due to Variability of Estimated Recruitments in Fishery Assessment Models. Can. J. Fish. Aquat. Sci. 68, 1744–1760.

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.

Mohn R., 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56, 473–488.

National Research Council (NRC). 1998. Improving Fish Stock Assessments. The National Academies Press, Washington, DC.

Needle, C.L., 2003. Survey-based assessments with SURBA. Working Document to the ICES Working Group on Methods of Fish Stock Assessment, Copenhagen, February 2003.

Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res. 158, 96–101.

Ono, K., Licandeo, R., Muradian, M.L., Cunningham, C.J., Anderson, S.C., Hurtado-Ferro, F., Johnson, K.F., McGilliard, C.R., Monnahan, C.C., Szuwalski, C.S., Valero, J., VertPre, K.A., Whitten, A.R., Punt, A.E., 2015. The importance of length and age composition data in statistical age-structured models for marine species. ICES J. Mar. Sci. 72, 31–43.

Pacific Fishery Management Council (PFMC). 2014. Terms of Reference for the goundfish and coastal pelagic species stock assessment review process for 2015-2016. http://www.pcouncil.org/wp-content/uploads/Stock_Assessment_ToR_2015-16.pdf (last accessed 8 August 2015)

Patterson, K.R., Melvin, G.D., 1996. Integrated catch at age analysis version 1.2. Scottish Fisheries Research Report, 56. Aberdeen: FRS. 66pp. http://www.gov.scot/Uploads/Documents/No%2058.pdf

Piner, K.R., Lee, H.H., Maunder, M.N., Methot, R.D., 2011. A Simulation-Based Method to Determine Model Misspecification: Examples Using Natural Mortality and Population Dynamics Models. Mar. Coast. Fish. 3, 336–343.

Plagányi, É.E., Punt, A.E., Hillary, R., Morello, E.B., Thébaud, O., Hutton, T., Pillans, R.D., Thorson, J.T., Fulton, E.A., Smith, A.D.M., Smith, F., Bayliss, P., Haywood, M., Lyne, V., Rothlisberg, P.C., 2014. Models of intermediate complexity for ecosystem assessment to support tactical management decisions in fisheries and conservation. Fish and Fish. 15, 1–22.

Pope, J.G., Shepherd, J.G., 1982. A simple method for the consistent interpretation of catch-at-age data. J. Cons. Int. Explor. Mer. 40, 176–184.

Porch, C.E., 1995. A two-area VPA with discrete mixing: Can we discriminate between mixing rates given the present condition of the data? Col. Vol. Sci. Pap. ICCAT 44, 198–208.

Porch, C.E., Turner, S.C., 1999. Virtual population analyses of Atlantic bluefin tuna with alternative models of trans-Atlantic migration. Col. Vol. Sci. Pap. ICCAT 49, 291–305.

Porch, C., Kleiber, P., Turner, S., Sibert, J., Bailey, R., Cort, J., 1998. The efficacy of VPA models in the presence of complicated movement patterns. ICCAT Collect. Vol. Sci. Pap. ICCAT 50, 591–622.

Porch, C.E., Restrepo, V.R., Turner, S.C., Scott, G.P., 1995. Virtual population analyses of Atlantic bluefin tuna incorporating movement and tagging data. Col. Vol. Sci. Pap. ICCAT 44, 183–190.

Porch, C.E., Turner, S.C., Powers, J.E., 2001. Virtual population analyses of Atlantic bluefin tuna with alternative models of transatlantic migration: 1970-1997. Col. Vol. Sci. Pap. ICCAT 52, 1022–1045.

Prager, M.H., 1992. ASPIC: A Surplus-Production Model Incorporating Covariates. Col. Vol. Sci. Pap. ICCAT. 28, 218–229.

Prager, M.H., 1994. A suite of extensions to a nonequilibrium surplus–production model. Fish. Bull. 92, 374–389.

Prager, M.H., 2002. Comparison of logistic and generalized surplus-production models applied to swordfish, Xiphias gladius, in the north Atlantic Ocean. Fish. Res. 58, 41–57.

Prager, M.H., Goodyear, C.P., Scott, G.P., 1996. Application of a Surplus Production Model to a Swordfish-Like Simulated Stock with Time-Changing Gear Selectivity. Trans. Am. Fish. Soc. 125, 729–740.

Prager, M., Williams, E., Shertzer, K., Cheshire, R., Purcell, K. 2015. User's guide to FishGraph: A set of R graphics functions for fishery stock assessment. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SEFSC-XXX.

Pribac, F., Punt, A.E., Walker, T.I., Taylor, B.L., 2005. Using length, age and tagging data in a stock assessment of a length selective fishery for gummy shark *(Mustelus antarcticus)*. J. Northw. Atl. Fish. Sci. 35, 267-290.

Punt, A.E., 1994. Assessments of the stocks of Cape hake *Merluccius* spp. off South Africa. S. Afr J. mar. Sci. 14, 159–186.

Punt, A.E., Kennedy, R.B., 1997. Population modelling of Tasmanian rock lobster, *Jasus edwardsii*, resources. Mar. Freshw. Res. 48, 967–980.

Punt, A.E., Haddon, M., Tuck, G.N., 2015a. Which assessment configurations perform best in the face of spatial heterogeneity in fishing mortality, growth and recruitment: A case study based on pink ling in Australia. Fish Res. 168, 85–99.

Punt, A.E., Haung, T-C., Maunder, M.N., 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES J. Mar. Sci. 70, 16-33.

Punt, A.E., Hurtado-Ferro, F., Whitten, A.R., 2014a. Model selection for selectivity in fisheries stock assessment. Fish. Res. 158, 124–134.

Punt, A.E., Pribac, F., Walker, T.I., Taylor, B.L., Prince, J.D., 2000. Stock assessment of school shark, *Galeorhinus galeus*, based on a spatially explicit population dynamics model. Mar. Freshw. Res. 51, 205–220.

Punt, A.E., Smith, A.D.M., Smith, D.C., Tuck, G.N., Klaer, N.L., 2014b. Selecting relative abundance proxies for B_{MSY} and B_{MEY}. ICES J. Mar. Sci. 71, 469–483.

Punt, A.E., Smith, D.C., Smith, A.D.M.. 2011. Among-stock comparisons for improving stock assessments of data-poor stocks – the “Robin Hood” approach. ICES J. Mar. Sci. 68, 972–981.

Punt, A.E., Su, N-J., C-L. Sun, C-L. 2015b. Assessing billfish stocks: A review of current method and some future directions. *Fish. Res.* 166, 103–118.

Restrepo, V.R., Porch, C.E., 2000. Options for conducting swordfish sex-specific assessments. Col. Vol. Sci. Pap. ICCAT. 51, 1375–1386.

Richards, L.J., 1991. Use of contradictory data sources in stock assessments. Fish. Res. 11, 225–238.

Rosenberg, A.A., Bolster, W.J., Alexander, K.E., Leavenworth, W.B., Cooper, A.B., McKenzie, M.G., 2005. The history of ocean resources: modelling cod biomass using historical records. Front. Ecol. Environ. 3, 78–84.

Sampson, D.B., 1993. The assumption of constant selectivity and the stock assessment for widow rockfish, *Sebastes entomelas.* Fish. Bull. 91, 676-689.

Sampson, D.B., Yin, Y., 1998. A Monte Carlo evaluation of the Stock Synthesis assessment program. In: Funk, F. et al. (eds), Fishery Stock Assessment Models. Alaska Sea Grant College Program Rep. AK-SG-98-01, University of Alaska, Fairbanks, p.315-338.

Schirripa, M.J., Goodyear, C.P., Methot, R.M., 2009. Testing different methods of incorporating climate data into the assessment of US West Coast sablefish. ICES J. Mar. Sci. 66, 1605–1613.

Schnute, J., 1977. Improved estimates from the Schaefer production model: Theoretical considerations. J. Fish. Res. Board Can. 34, 583–603.

Shepherd, J.G., 1999. Extended survivors analysis: An improved method for the analysis of catch-at-age data and abundance indices. ICES J. Mar. Sci. 56, 584–591.

Science Working Group (SWG) 2010. Report from the assessment simulation task team workshop on using the joint jack mackerel assessment model (JJM). South Pacific Regional Fisheries Management Organisation. https://www.sprfmo.int/assets/Meetings/Meetings-before-2013/Scientific-Working-Group/SWG-09-2010/SWG-09-JM-02A-Report-Seattle-v4.pdf (last accessed 7 August 2015).

Stewart, I.J., Piner, K.R., 2007. Simulation of the estimation of ageing bias inside an integrated assessment of canary rockfish using age estimates from a bomb radiocarbon study. Mar. Freshw. Res. 58, 905–913.

Stewart, I.J., Hicks, A.C., Taylor, I.G., Thorson, J.T., Wetzel, C., Kupschus, S., 2013. A comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework. Fish. Res. 142: 37–46.

Taylor, I.G., Methot, R.F. Jr., 2013. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fish. Res. 142, 75–85.

Taylor, I., Stewart, I., Hicks, A., Garrison, T., Punt, A., Wallace, J., Wetzel, C., Thorson, J., Takeuchi, Y., Monnahan, C., 2014. Packages 'r4ss' https://cran.r-project.org/web/packages/r4ss/r4ss.pdf (last accessed 12 August 2015)

Thorson, J.T., Wetzel, C., 2015. The status of canary rockfish (*Sebastes pinniger*) in the California Current in 2015. Pacific Fishery Management Council. 7700 NE Ambassador Place, Suite 101, Portland, OR 97200-1384, USA. 692pp.

Thorson, J.T., Taylor, I.G., Stewart, I.J., Punt, A.E., 2014. Rigorous meta-analysis of life history correlations by simultaneously analyzing multiple population dynamics models. Ecol. Appl. 24, 315–326.

Turnock, B.,J., Rugolo, L.J., 2014. Stock assessment of Eastern being Sea snow crab. pp. 41-177. 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 9950.

U.S. Department of Commerce (U.S. Doc). 2007. Magnuson-Stevens Fishery Conservation and Management Act. 2007. National Oceanic and Atmospheric Administration. http://www.fisheries.noaa.gov/sfa/laws_policies/msa/documents/msa_amended_2007.pdf (Last accessed 24 August 2015)

Walter, J.F., Porch, C.E., 2012. Three strategies for modeling terminal-year fishing mortality rates in Virtual Population Analyses of western Atlantic bluefin tuna: Retrospective patterns and consequences for projections. Aquat. Liv. Resou. 25, 333–340.

Wayte, S.E., 2013. Management implications of including a climate-induced recruitment shift in the stock assessment for jackass morwong (N*emadactylus macropterus*) in south-eastern Australia. Fish. Res. 142, 47–55.

Waterhouse, L., Sampson, D.B., Maunder, M., Semmens, B.X., 2014. Using areas-as-fleets selectivity to model spatial fishing: Asymptotic curves are unlikely under equilibrium conditions. Fish. Res. 158, 15–25.

Wetzel, C.R., Punt, A.E., 2011. Performance of a fisheries catch-at-age model (stock synthesis) in data-limited situations. Mar. Freshw. Res. 62, 927–936.

Wetzel, C.R., Punt, A.E., In press. Evaluating the performance of data-moderate and catch-only assessment methods for U.S. west coast groundfish. Fish. Res. 00, 00–00.

Williams, E.H., Prager, M.H., 2002. Comparison of equilibrium and nonequilibrium estimators for the generalized production model. Can. J. Fish. Aquat. Sci. 59, 1533–1552.

Williams, E.H., Shertzer, K.W., 2015. Technical documentation of the Beaufort Assessment Model (BAM). NOAA Technical Memorandum NMFS-SEFSC-671. 43 pp.

Worm, B., Hilborn, R., Baum, J.K. Branch, T.A.Collie, J.S, Costello, C., Fogarty, M.J., Fulton, E.A., Hutchings, J.A., Jennings, S., Jensen, O.P., Lotze, H.K., Mace, P.M., McClanahan, T.R., Minto, C., Palumbi, S.R., Parma,A.M., l Ricard, D., Rosenberg, A.A.,Watson,R., Zeller, D. 2009. Rebuilding Global Fisheries. Science 325, 578-585,

Yin, Y., Sampson, D.B., 2004. Bias and precision of estimates from an age-structured stock assessment Program in relation to stock and data characteristics. No. Am. J. Fish. Manage. 24, 865–879.

Zheng, J., Siddeek, M.S.M. 2014. Bristol Bay red king crab stock assessment in fall 2014. pp. 178-323. 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 9950.

CHAPTER 2. HOW MANY OF AUSTRALIA'S STOCK ASSESSMENTS CAN BE CONDUCTED USING STOCK ASSESSMENT PACKAGES?[9]

INTRODUCTION

Fisheries in Australia are considered amongst the world's best managed. While attempts to rank nations in terms of their ability to successfully implement fisheries (and Ocean) management should be interpreted with extreme care, Australia is consistently ranked very highly. For example, Australia was ranked 7th in terms of managing marine ecosystems [1], 4th in terms of compliance with the FAO Code of Conduct for Responsible Fisheries [2], and 7th in terms of overall performance in terms of ecosystem based management [3]. Moreover, only 6.4% of the 170 assessed Australian stocks are considered to be overfished [4]. Fishery management in Australia is, however, not without major challenges. Specifically, the amount of government funding available for monitoring is lower than in the USA and western Europe, and Australia has many low value data-limited species and fisheries [5]. This can be attributed partially to the low volume (and total value) of fisheries in Australia [4], and that most jurisdictions use partial cost recovery from the industry for fisheries management, i.e. a user pays system.

Notwithstanding the limited resources for monitoring, assessments and management, most jurisdictions in Australia (six State, one Territory, and the Federal (Commonwealth) government) have adopted an approach to management that involves conducting quantitative assessments based on fitting population dynamics models for stocks considered to be of recreational and commercial importance, and applying some form of harvest control rule to determine management actions. This is most clearly formulated at the Commonwealth level where a Harvest Strategy Policy has been developed [6], which specifies the need for target and limit reference points, and provides defaults for these reference points and the maximum probability that stocks are reduced below the limit reference point. The target reference point for Australian Commonwealth managed fisheries is B_{MEY}, the biomass corresponding to Maximum Economic Yield, which is generally larger than the conventional target for fisheries management, the biomass corresponding to Maximum Sustainable Yield, B_{MSY} (e.g. [46, 47]). The Australian Commonwealth Harvest Strategy Policy allows for the use of proxies for B_{MEY} ($1.2 \times B_{\mathrm{MSY}}$), where the proxy for B_{MSY} is taken to be $0.4B_0$, or 40% of the pre-fishery spawning biomass [7] if these reference points

[9] Published in *Marine Policy* with authors: Catherine M. Dichmont, Roy A. Deng, and Andre E. Punt

cannot be estimated using a quantitative model. Management actions justified based on economic considerations and quantitative model-based stock assessments are also common at the Australian state level [e.g. 8, 9]. A National Harvest Strategy Guideline [10] has been developed and, although it has not been widely adopted as of writing, it has been endorsed by the heads of the Australian and state and territory government agencies responsible for fisheries through the Australian Fisheries Management Forum. No specific reference points are prescribed in this Guideline.

The trend in the USA, Europe and New Zealand and in Regional Fisheries Management Organizations is towards stock assessments being conducted using a small number of stock assessment packages (see, for example, the review of stock assessment packages used in the USA by Dichmont *et al*. [11]. Dichmont *et al*. [11] identify the benefits of using packages rather than user-developed software as: (a) substantially increased flexibility to explore alternative assessment configurations, (b) ease of peer-reviewing, (c) increased confidence that the assessment is correctly coded and tested, (d) the availability of tools to explore uncertainty and summarise model fits to data, (e) increased collaboration amongst assessment scientists on generic questions related to stock assessment practice given a common software platform is being used, (f) faster development time for an assessment, (g) increased ability for a new analyst to take over a stock assessment because they are familiar with the package, and (h) a large user base to facilitate further development and improvement (and to detect errors).

However, as will be shown in this Chapter, only a relatively small proportion of Australian assessments are currently being conducted using stock assessment packages, and only in one region of Australia. This Chapter provides a summary of the model-based stock assessments used in Australia (these being the most demanding in terms of data requirements as well as the need for skilled analysts and that support management of the most valuable fisheries), and examines the extent to which it would be possible to conduct assessments in Australia using the stock assessment packages used most commonly for conducting ‘integrated’ assessments, particularly those packages available in the USA and New Zealand, since fisheries management in these countries aligns most closely to the Australian policy environment. The lessons learnt from this review will help Australia plan a ‘business model’ for how to provide ongoing scientific fisheries management advice.

MATERIALS AND METHODS

SURVEY QUESTIONNAIRES

A two-step survey process was undertaken. The first step involved contacting each Australian fisheries research organisation (or individuals within consulting companies) that had undertaken stock assessments in the past decade to obtain a list of the species (and stocks within species) with assessments and the primary contact for the assessment. For the purposes of this study, stocks assessments were defined as "model-based assessments used for tactical fisheries management that includes an optimisation component." The second step involved contacting each assessment analyst and requesting information for each assessment. The questionnaire was in the form of an Excel file with default dropdown menus for several questions, but with the ability to provide written comments for all questions (see Supplementary Table 1 for the entire questionnaire). The four categories of questions were:

- an overview of the method of assessment on which the assessment was based, including whether it was designed for generic use (i.e. for multiple fisheries or stocks) and whether it has been subject to peer-review;
- information about the software used to implement the assessment, including whether it is available publically, the language on which it is based, and its ongoing maintenance;
- specifics of the technical aspects of the model on which the assessment is based; and
- information on how uncertainty is characterised.

All the responses were collated into a single file and checked by the authors of this Chapter. We were unable to obtain a response from New South Wales. Any queries were sent to the respondents for further comment. Respondents did not always provide references to the methods on which their assessments were based, so these were sourced by the authors and then sent to the respondents for checking.

DATA ANALYSIS

The species for which stock assessments are conducted in Australia were categorised into seven broad taxonomic groups: "abalone", "scallops", "crabs", "lobsters", "prawns", "sharks", and "finfish". These represent a subset of the 11 species groups for which species summaries are provided in the Australia's Status for Key Australian Fish Stocks Reports (SAFS; e.g. [4]). The other species groups included in SAFS are pipis, octopus, squids, and bugs, but no tactical model-based assessments are available for species in these groups. The stock assessment methods on which Australia stock assessments were based were categorised in terms of the structure of the population dynamic model underlying the assessment (Table 2.1). As an initial summary, the assessments were then categorised by: a) jurisdiction, b) research organisation that conducted the assessment, c) programming language used; d) input data required; e) population dynamic structure, and f) how the method and application were documented. The results for c), d) and f) were summarised by population dynamic structure as well as by jurisdiction.

Table 2.1. Possible population dynamic structures, and freely available stock assessment packages used in the USA and New Zealand that implement these structures (see Table 2 for explanation of package abbreviations and associated references).

Population dynamic structure	Definition	Assessment package with similar model dynamics
Delay-difference	Delay-difference model	AMAK, ASAP, BAM, CASAL, CSA, MULTIFAN-CL, SCALE, SS, SSS, VPA, VPA2BOX, XSSS
Production	Surplus production or biomass dynamics model	ASPIC, BSP1, BSP2, XDB-SRA*
Age	Age structured model	AMAK, ASAP, BAM, CASAL, MULTIFAN-CL, SCALE, SS, SSS, VPA, VPA2BOX, XSSS
Age-sex	Age-structured model that separates the sexes	CASAL, SS, SSS, XSSS
Age-area	Age-structured model that represents populations spatially	CASAL, MULTIFAN-CL, SS
Age-sex-area	As above, but also represents populations by sex	SS
Age, Size	Age-structured model that can fit to size data (but remains fundamentally age-structured)	CASAL, SS
Age-sex, Size	Age-sex structured model that can fit to size data	CASAL, SS
Age-area, Size	Age-area structured model that can fit to size data	SS
Age-sex-area, Size	As above, but also represents populations by sex	SS
Age-size	Model that is both age and size structured	SS
Age-size-sex	Age-size structured model that separates the sexes	SS
Age-size-area	Age-size structured model that represents populations spatially	SS
Age-size-sex-area	As above, but also represents populations by sex	SS
Size	Size-structured model	CASAL, GMACS
Size-sex	Size-structured model that separates the sexes	CASAL, GMACS
Size-area	Size-structured model that represents populations spatially	NONE
Size-sex-area	As above, but also represents populations by sex	NONE

* XDB-SRA divides the population to mature and immature components, but does not contain a standard delay-difference structure so is not considered a delay-difference model for the purposes of this Chapter.

Each assessment was classified as “unique”, “semi-generic” or “package” based on the assessment method applied (i.e. whether the assessment method (a) could not be applied to another stock, (b) could, with reasonably minor code and input model adjustments, form the basis for an assessment of another stock or species, or (c) was based on a freely available stock assessment package that can easily be adapted to another similar stock or species without code changes and without need to review the method or code).

The assessments that are currently not based on stock assessment packages (i.e. the “unique” and “semi-generic” assessments) were then linked to available stock assessment packages used in the USA and New Zealand (Table 2.2) based on the population dynamic structure on which the assessment was based to provide an initial set of packages that could be used to conduct the assessment. A final set of assessments that could be conducted using packages was then identified taking into account the data sources used for parameter estimation and other features of the assessment. Those assessments that cannot be conducted using packages will remain “unique”, and the reasons for these cases are identified and discussed.

Table 2.2. Summary of the stock assessment packages that could have been used to assess the Australian stocks.

Package abbreviatio n	Package name	Population dynamic structure that can be represented	References
AMAK	Assessment Method from Alaska	Delay-difference; Age	[24]
ASAP	Age Structured Assessment Procedure	Delay-difference; Age	[25]
ASPIC	A Stock Production Model Incorporating Covariates	Production model	[26,27,28]
BAM	Beaufort Assessment Model	Delay-difference; Age	[29]
BSP1	Bayesian Surplus Production model -1	Production model	[30]
BSP2	Bayesian Surplus Production Model -2	Production model	[31]
CASAL	C++ algorithmic stock assessment laboratory	Delay-difference; Age; Age-sex; Age-area; Age, Size; Age-area, Size; Size-sex	[32]
CSA	Collie-Sissenwine Analysis	Delay-difference	[33]
GMACS	Generalized Modeling for Alaskan Crab Stocks	Size; Size-sex	Athol Whitten (pers. Commn)
MULTIFAN-CL	MULTIFAN-CL	Delay-difference; Age; Age-area	[34]
SCALE	Statistical Catch At LEngth	Delay-difference; Age;	NOAA Fisheries Toolbox
SS	Stock Synthesis	Delay-difference; Age; Age-sex; Age-area; Age-sex-area; Age, Size; Age-area, Size’ Age-sex-area, Size; Age-size; Age-size-sex; Age-size-sex-area	[35]
SSS	Simple Stock Synthesis	Delay-difference; Age	[36]
VPA	Virtual Population Analysis	Delay-difference; Age	[37]
VPA2BOX	Dual Zone VPA Model	Delay-difference; Age	[38,39,40,41]
XDB-SRA	Extended Depletion-Based Stock Reduction Analysis	Production model	[42,43]
XSSS	Extended Simple Stock Synthesis	Delay-difference; Age	[43,43]

RESULTS

REPRESENTATIVENESS OF THE DATA SET

The total (commercial plus recreational) catch of all marine species by Australia was 152,689t in 2013, with a value of AUS$1.4 billion. Of this catch, 92% is included in the 2014 SAFS (Table 2.3). The total catch represented within the 76 model-based stock assessments[10] is 50,776t. Almost all of the catch of lobsters (75%) is included in stock assessments, while more than 50% of the catch of scallops, prawns and sharks are included in stock assessments. In contrast, no assessments are available for pipis, octopus, squids and bugs, and less than 15% of the catch of crab stocks is included in assessments. The largest catch by volume is of finfish, and of that catch 37.2% is included in stock assessments.

Table 2.3. Summary of the catch for 2013 reported in the SAFS [43] by species group and the catch by species group that is included in model-based stock assessments. The total catch taken by Australia in 2013 was 152,689t. Thus, the catch in the SAFS is 92% of the total catch (90% of the total value). Assessments for ten stocks of finfish with stock assessments are not included in the SAFS.

Species group	Catch (t) in the SAFS	Catch (t) in the stock assessments	Percentage in stock assessments
Molluscs	10,196	3,171	31.1
Abalone	2,958	744	25.2
Pipis	582	0	
Scallops	4,137	2,427	58.7
Octopus	96	0	
Squids	2,423	0	
Crustaceans	31,746	19,043	60.0
Crabs	4,191	611	14.6
Bugs	734	0	
Lobsters	9,514	9,194	96.6
Prawns	17,308	9,238	53.4
Sharks	2,986	2,153	72.1
Finfish	70,918	26,409	37.2

Table 2.4 lists the amount of catch of the seven species groups represented in stock assessments by jurisdiction. Stock assessments are available for over 80% of the catch for Commonwealth and Northern Territory fisheries. In contrast, stock assessments are only

[10] Actually 66 because 10 of the stocks are not included in the SAFS.

available for 27% of the catch by Tasmania fisheries, 7% of the catch by South Australian fisheries, and 11% of the catch by New South Wales fisheries. The low proportion for Tasmania is due to the lack of assessments for abalone and scallops while 31,981t of the catch of finfish off South Australia is of Australian sardine, which is not assessed to provide estimates of fishing mortality and biomass. The low proportion of the New South Wales catch included in stock assessments is a result of few assessments for finfish and prawns.

Table 2.4. Breakdown of the catch (t) by species and jurisdiction based on SAFS [43]. The regional fisheries management organizations are: JZIA (lobster and prawns), the Indian Ocean Tuna Commission (finfish), the Western and Central Pacific Fisheries Commission (finfish), and the Commission for the Conservation of Southern Bluefin Tuna (finfish). The final row indicates the proportion of the catch by jurisdiction included in stock assessments.

Species group	Commonwealth	New South Wales	Northern Territory	Queensland	South Australia	Tasmania	Victoria	Western Australia	Regional Fisheries Management Organizations
Abalone	0	113	0	0	617	1,248	744	166	0
							(744)		
Scallops	189	0	0	2,427	0	1,260	0	261	0
	(0)			(2,427)		(0)		(0)	
Crabs	0	296	227	2,670	597	27	10	364	0
		(0)	(227)	(384)	(0)	(0)	(0)	(0)	
Lobsters		139		181	1,552	1,064	306	5,686	586
					(1,552)	(1,064)	(306)	(5,686)	(586)
Prawns	5,839	1,300	0	5,564	1,813	0	17	2,671	103
	(3,507)	(569)		(5,059)	(0)		(0)	(0)	(103)
Sharks	1,700	16	112	329	215	3	7	603	0
	(1,697)	(3)	(111)	(328)	(12)	(1)	(1)	(0)	(0)
Finfish	9,170	4,528	4,051	6,492	33,410	381	1615	4,707	7,832
	(9,000)	(192)	(3,562)	(6,031)	(1,059)	(31)	(159)	(1,837)	(4,539)
Total (%)	84.1	11.4	89.4	71.9	6.8	27.3	44.6	50.8	61.4

Table 2.5. Summary of the 45 assessments conducted using stock assessment packages or semi-generic assessment methods.

Assessment method	No of assessments	Type of User	Training available?	Manual?	Regular maintenance?	Diagnostics produced automatically?	Methods for quantifying parameter uncertainty	Basis for projections	Basis for method review
Packages									
CASAL	1	Experienced	No	Yes	Yes	No	Asymptotic, Bayesian	Catch	[31]
Stock Synthesis	16	Experienced	Yes	Yes	Yes	Yes	Asymptotic, Bayesian, Bootstrap	Catch, Effort	[34]
Semi-generic									
QDAF monthly sex-length model.	3	Experienced	Yes	No	Yes	Yes	Asymptotic	Catch, Effort	N/A
QDAF semi-generic fish model	4	Beginner / experienced	Yes	No	Yes	Yes	Asymptotic	Effort, Catch	N/A
SA Slice-partition age-length model	3	Beginner / experienced	No	No	Yes	Yes	Asymptotic	Catch, Effort	[44]
Southern rock lobster model	3	Experienced	No	Yes	No	Yes	Asymptotic	Catch, Effort	N/A
Stochastic Stock Reduction Analysis	9	Beginner	Yes	Yes	Yes	Yes	Bayesian	Catch	N/A
WDAF age-structured Integrated model	6	Beginner / developers	No	No	Yes	Yes	Asymptotic	Catch	Independent review

SYNTHESIS OF ASSESSMENTS

The assessments were conducted between 2005 and 2015, with 46 of the assessments conducted during 2014 and 2015. Queensland and the Commonwealth commissioned the bulk of the assessments (23 and 22 respectively; Fig. 2.1a). In addition, the Commonwealth also provided scientific support for the Australian delegation to the scientific meetings of the Commission for the Conservation of Southern Bluefin Tuna, while Queensland and the Commonwealth jointly provided four assessments for species in the Protected Zone Joint Authority (PZJA) (Fig. 2.1a). There is close to a one-to-one mapping of assessment service providers and the corresponding State or Federal jurisdiction (Figs 2.1a,b). For example, the ten Northern Territory assessments were undertaken by the Northern Territory Department of Primary Industry and Fisheries (NTDPIF) in workshops where NDPIF and other analysts undertook assessments of key species. The Commonwealth Scientific and Industrial Research Organisation (CSIRO) conducted all but one of the Commonwealth assessments, with the assessment of Western Gemfish conducted by another federal agency, the Australian Bureau of Agricultural and Resource Economics and Sciences. The CSIRO also undertook the assessment for Southern Bluefin Tuna and one of the assessments for the PZJA. There are only a few exceptions to this one-to-one mapping between a jurisdiction and its direct service provider. In particular, consulting companies provided assessments to the Victorian Department of Environment and Primary Industries and the Australian Federal Government.

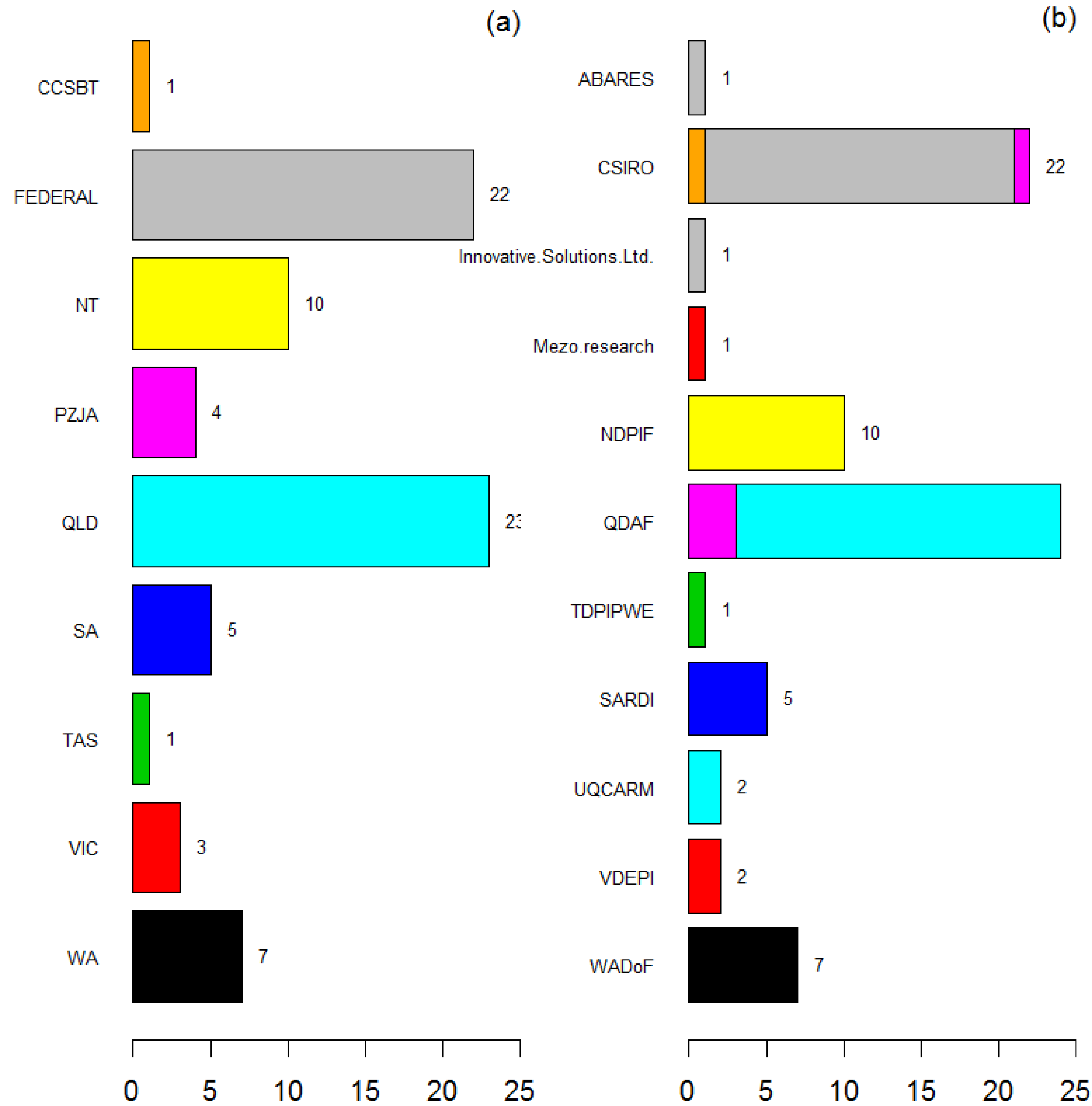

Figure 2.1. Number of assessments by jurisdiction (left) and service provider (right). Colour coding corresponds to the jurisdiction (left) and predominant jurisdiction the agency delivers to (right). The abbreviations are as follows. CCSBT: The Commission for the Conservation of Southern Bluefin Tuna; FEDERAL: Commonwealth of Australia; NT: Northern Territory; PZJA: the Protected Zone Joint Authority; QLD: Queensland; SA: South Australia; TAS: Tasmania; VIC: Victoria; WA: Western Australia; ABARES: the Australian Bureau of Agricultural and Resource Economics and Sciences; CSIRO: Commonwealth Scientific and Industrial Research Organisation; NDPIF: Northern Territory Department of Primary Industry and Fisheries; QDAF: Queensland Department of Agriculture and Fisheries; TDPIPWE: Tasmania Department of Primary Industries, Parks, Water and Environment; SARDI: South Australian Research and Development Institute; UQCARM: University of Queensland Centre for Applications in Natural Resource Mathematics; VDEPI: Victoria Department of Environment and Primary Industries; WADoF: Western Australia Department of Fisheries. NSW was not included in this review (that undertakes 1 assessment) as it was not a respondent to the survey.

The 76 assessments were conducted using 39 stock assessment methods (2 packages, 6 semi-generic methods, 31 unique assessments; Table 2.5, supplementary Table 2.2). The most common programming language used to implement the assessment methods was Auto-differentiation Model Builder (ADMB) [12], which is freely available software that takes model specifications written in a C++ pre-processor language and creates executable code that includes the function and its derivatives with respect to each estimable parameter (Fig. 2.2). Although 16 assessments were implemented using Matlab, all of these were developed by the Queensland Department of Agriculture and Fisheries (QDAF) staff. Several assessments extensively used R. Purposes of R included: a) calling programs written in another language, such as ADMB; b) fitting models coded in R; and c) automating analysis of the output data (the latter being the most common). The assessments conducted by the NTDPIF were implemented in Excel and/or Visual Basic (on its own or within Excel), whereas Excel was rarely used outside this jurisdiction. Overall, the bulk of the assessments (68) were implemented using freely available software (WinBUGS, ADMB, and R) and software that is part of Microsoft Office (VB and Excel).

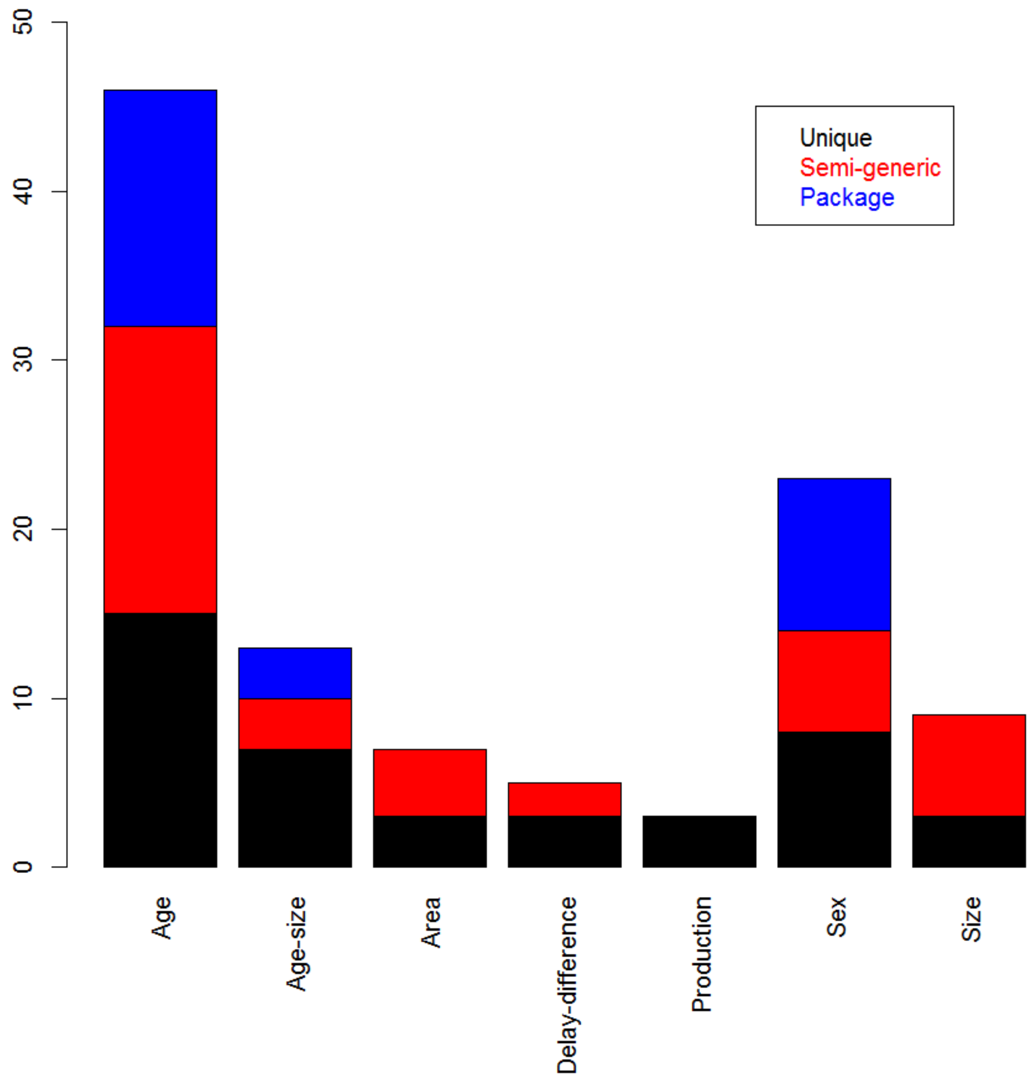

Figure 2.2. Summary of programming languages used in Australia to implement stock assessments. VB= Visual Basic; ADMB=Automatic Differentiation Model Builder.

A variety of data types are used in Australian stock assessments. These include the data types traditionally included in stock assessments, such as catch data, CPUE and survey index data, discard rate data, catch and survey age- and size-composition data, catch and survey conditional age-at-length data, and tagging data. Apart from catch data, CPUE is the most common input data type for Australian assessments (Fig. 2.3). Age and size data are also key inputs to these assessments. Tagging data are only included in ten assessments, typically those that are "unique". Not surprisingly, non-traditional data sources such as habitat area, genetic and Vessel Monitoring System (VMS) data tend to be used in a small number of "unique" assessments. River discharge is used as a covariate in the stock-recruitment relationship and to determine growth transition matrices for Barramundi in Queensland [13]. The assessment of sharks off Queensland is multi-species so requires data on species frequency.

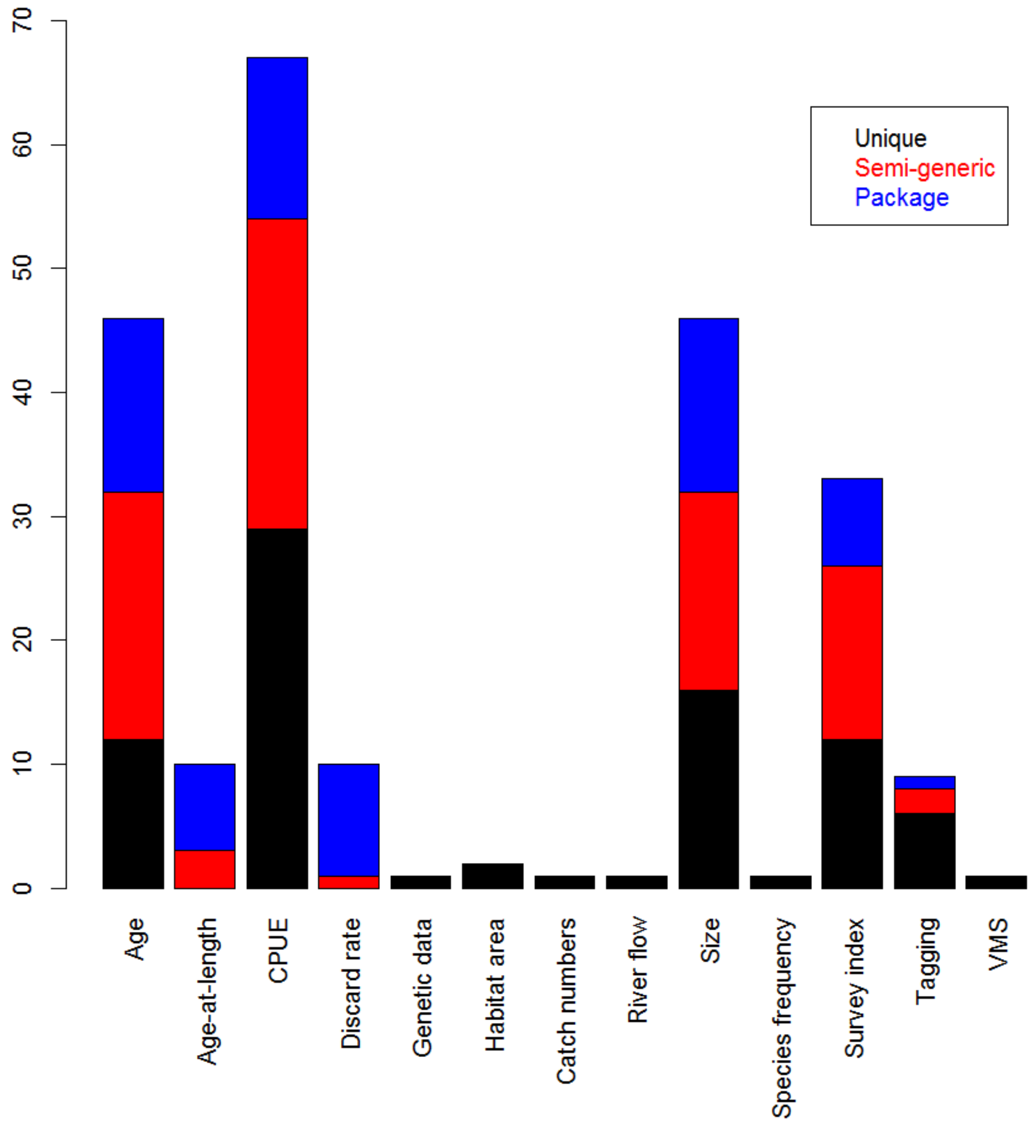

Figure 2.3. Input data used for assessments categorised as "unique", "semi-generic" or a "package" (other than catch). VMS denotes "Vessel Monitoring System".

A breakdown of the assessments by their population dynamic structure and whether the assessment was conducted using a "unique", "semi-generic" and "package" assessment method is outlined in Fig. 2.4. Assessments based on the stock assessment packages Stock Synthesis and CASAL were only conducted for fish species found in the South East Shark and Scalefish Fishery (SESSF). Eighteen of the 28 "semi-generic" assessments were conducted for finfish species (Table 2.6a). The relationship between species group and population dynamic structure is even more evident when the "unique" assessments are considered – only 12 of these assessments were for finfish (Table 2.6a). The bulk of the assessments were based on age-structured population dynamics models, irrespective of whether the assessment was "unique", "semi-generic" or "package" (Fig. 2.4). This probably reflects that a large number of the assessments were for finfish species. Several assessments were based on sex-disaggregated models. However, only "unique" and "semi-generic" assessments were spatially-disaggregated. Fewer than ten of the assessments were based on size-structured models. The use of surplus production and delay-difference models was rare.

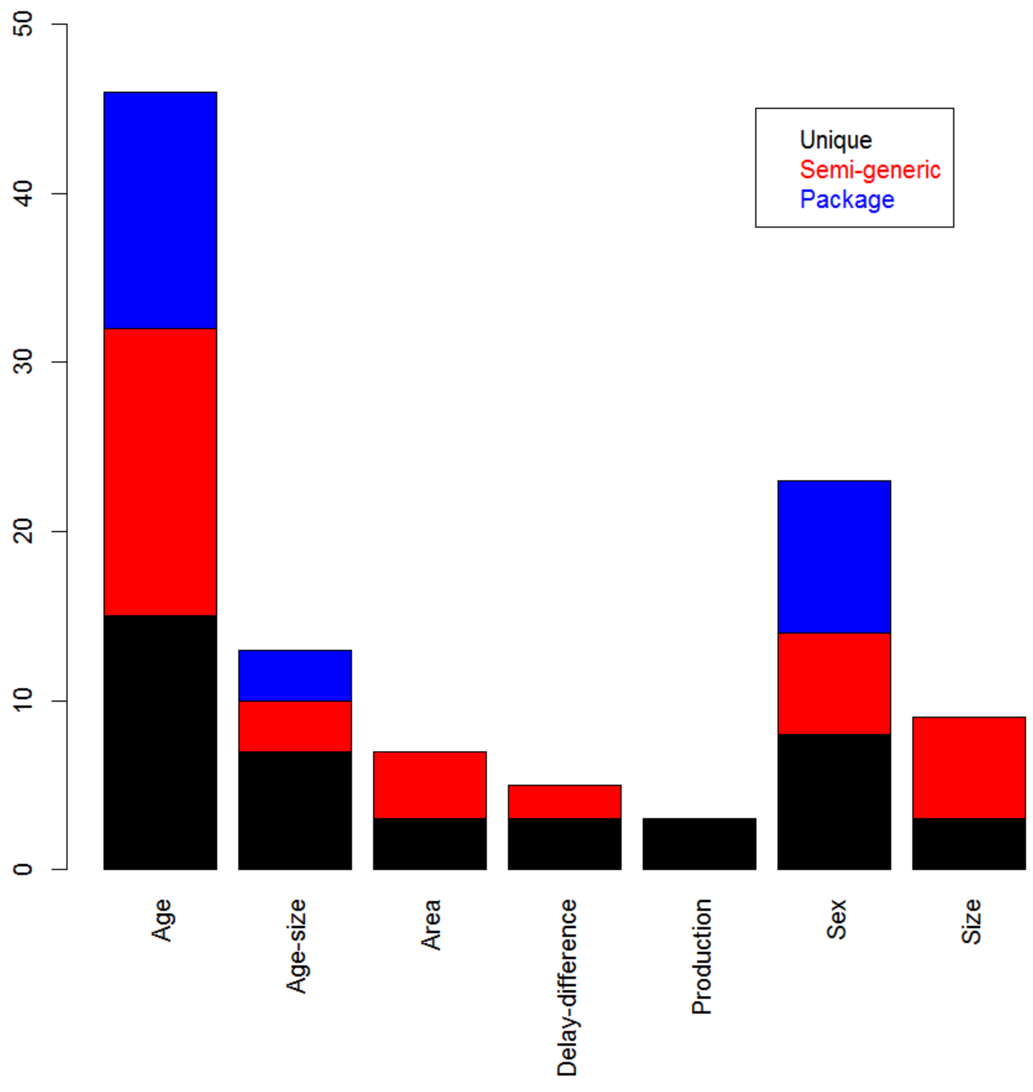

Figure 2.4. Assessments categorised by population dynamic structure and whether they are based on "unique", "semi-generic" or "package" approaches. Note that assessments classified as sex-structured ("Sex') or area-structured ("Area") may also be counted under other model type categories.

Table 2.6. Number of assessments that are classified as "Unique", "Semi-generic" or a "Package" by species group (a) and in terms of whether they can be conducted by packages (b).

(a) Breakdown of assessed species by type and species group

Type of assessment	Abalone	Crabs	Finfish	Lobsters	Prawns	Scallops	Sharks	Total
Unique	1	1	12	3	9	2	3	31
Semi-generic	0	1	18	3	4	0	2	28
Package	0	0	17	0	0	0	0	17

(b) Can be assessed using a package

	Abalone	Crabs	Finfish	Lobsters	Prawns	Scallops	Sharks	Total
No	0	1	4	4	5	1	3	18
Yes	1	1	43	2	8	1	2	58

The 76 assessments quantify estimation uncertainty using Bayesian, bootstrap, and asymptotic methods (Table 2.5; Supplementary Table 2.2), as well as by varying the assumptions of the assessment to form sensitivity tests. The "unique" and "semi-generic" assessment methods tend to quantify uncertainty using a single approach (Bayesian or asymptotic) whereas Stock Synthesis can quantify uncertainty using all three ways to quantify uncertainty, and also has the facility to generate bootstrap data sets to enable its estimation performance to be evaluated semi-automatically. Almost all of the packages include the facility to undertake projections, with projections in which catches are used to drive the future population dynamics the most common option (Table 2.5, Supplementary Table 2.2). The outputs from the projections tend to be time-trajectories of spawning or total biomass, but some of the assessments (e.g. those for Tiger Prawns in the Northern Prawn Fishery and off Queensland, or rock lobsters off southern Australia) also provide results in the form of time-trajectories of revenue and profit (e.g. [14]).

WHICH ASSESSMENTS CAN BE CONDUCTED USING STOCK ASSESSMENT PACKAGES?

Thirty-one of the 76 assessments undertaken in Australia were classified as "unique", i.e. the code implementing the assessment was developed specifically for that application (Table 2.6a; Supplementary Table 2.2). A further 28 assessments, predominantly those conducted by the QDAF, were classified "semi-generic" where the same assessment method can be applied to a group of stocks or species following minor changes to the code used to implement the assessment method.

The 76 assessments were evaluated in terms of whether their population dynamic structure matched those of at least one of 16 freely-available packages in Table 2.2 (Supplementary Table 2.3). One of the "unique" assessments (that for Western Rock Lobster) did not have a population dynamic structure that matched that of at least one of the packages and this was the case for five of the "semi-generic" assessments. Thus, there are six stocks for which no assessment package exists that could potentially be used to conduct the assessments. When other features of the assessments are taken into account, the number of stocks for which no assessment package could be used increases from 1 and 5 ("unique" and "semi-unique" respectively) to 13 and 5.

The bulk of finfish assessments can be conducted using CASAL or Stock Synthesis while GMACS or CASAL would be appropriate for several of the assessments based on size-structured population dynamics models. Table 7 summarises the reasons that packages cannot be used for 18 of the stock assessments. Of the 12 assessments for which a package exists that has the required population dynamic structure, but they cannot be assessed using a package, four use tagging data to estimate fishing mortality (Stock Synthesis can make use of tagging data, but only to estimate movement), two are multi-species, two use habitat data to make catchability comparable spatially, one explicitly models migration while another is highly spatially-structured (very many spatial areas), one uses river discharge as a covariate defining growth, and the final one is fitted to catch-in-numbers and catch-in-weight simultaneously.

Table 2.7. The stocks which cannot be assessed by a currently existing package and the reasons for this

Species / stock name	Model type	Reason no package can be applied
No package with the correct population dynamic structure exists		
FEDERAL - Tiger Prawn and Endeavour Prawn	Spatially-structured production model	
QLD - Blue Swimmer Crab	Size-sex-area-structured	
QLD - Eastern King Prawn	Size-sex-area-structured	
QLD - Western King Prawn	Size-sex-area-structured	
TAS - Southern Rock Lobster	Size-sex-area-structured	
WA – Western Rock Lobster	Size-sex-area-structured	
Other features preclude use by packages		
CCSBT - Southern Bluefin Tuna	Age, Size	Uses genetic and tagging data[1]
FEDERAL – Gummy Shark	Age-sex	Uses tagging data[1]; effort saturation
FEDERAL – School Shark	Age-sex	Uses Tagging data[1]
FEDERAL - Tiger Prawn and Endeavour Prawn	Delay-difference	Weekly time-step, uses economics data to calculate the total allowable effort, multi-species
FEDERAL - Tiger Prawn and Endeavour Prawn	Size-sex-structured	Weekly time-step, uses economics data to calculate the total allowable effort, multi-species
PJZA - Tropical Rock Lobster	Age-structured	Involves migrations
QLD – Barramundi – East Coast and Gulf of Carpentaria	Age-size	Uses tagging data[1]
QLD – Barramundi – Fitzroy	Age-size	River discharge impacts growth rate
QLD – Coral Trout	Age-area	Uses habitat data to scale catch rates
QLD – Scallop	Age-structured	Highly spatial-structured model using VMS data
QLD – Shark	Age-area	Uses habitat data and species composition data
SA - Southern Rock Lobster (Northern Zone)	Production	Uses data on catch-in-numbers and –in-weight simultaneously

[1] To estimate mortality rather than growth

DISCUSSION

Thirty-seven percent of the Australian fisheries catch (in 2013) that is included in the Australian Status for Key Australian Fish Stocks Reports 2014 is for species for which model-based stock assessments exist, although this percentage could be increased to 60% if a model-based assessment of Australian Sardine off South Australia could be conducted. Thus, the availability of stock assessments on which to base management advice is important for the ability of management to achieve its goals. However, this proportion is lower than is the case in the USA. This is attributable primarily to the fact that a high proportion of Australia's catch is of invertebrates such as Abalone that have proved to be notoriously difficult to assess. In contrast, countries such as the USA and New Zealand have a few species, the catch of which constitutes a large proportion of the total catch by the country (Walleye Pollock, *Theragra chalcogramma*, in the case of the USA and Hoki, *Macruronus novaezelandiae*, in the case of New Zealand). However, there are many stocks of finfish and prawns in Australia that are potentially amenable to assessment, were the resources (including analysts) available to conduct them, which would increase the proportion of the total catch by Australia that is included in stock assessments.

Eighteen of the 76 assessments (24%) could not be conducted using packages that are freely available and used to conduct assessments in the USA and New Zealand. This proportion is much higher than would be the case for USA federally-managed species. There are several reasons for this:

- Australia has a high proportion of invertebrate stocks which need to be assessed using size-structured (or age-size-structured) models. By contrast, most invertebrate stocks in the USA are managed by the States rather than the federal government and few State-managed invertebrate stocks have stock assessments. In addition, for those invertebrate stocks that are managed by the USA federal government, the assessments are currently conducted using case-specific (i.e. unique) stock assessment methods.
- There are no packages that can deal explicitly with species for which age-size-area models are most appropriate and for species that need to be assessed using spatially-structured production models. In principle, it might be possible to configure Stock Synthesis as a production model and consider multiple areas.
- Some of the assessments make use of non-traditional data types such as VMS, habitat information, and river flow, which cannot be used in the currently-available packages. Stock Synthesis can link parameters to an environmental covariate, but the assessment of Fitzroy Barramundi links growth and recruitment to multiple environmental covariates.

The number of stock assessments that can be conducted using packages is perhaps somewhat over-estimated because while Stock Synthesis can represent age-size structured dynamics using its platoon feature [15], platoons only represent age-size dynamics approximately. If assessments based on age-size models are considered not be implementable using packages, the number of Australian stocks for which packages cannot be used increases from 18 to 27.

The comparisons of this Chapter are based on assuming that the analysts would make the same assumptions regarding the structure of their assessments had they had to use packages. In principle, the analysts conducting assessments may have changed their approaches somewhat had they known they had to use packages, with results not appreciably different from those of current approaches. It is, however, beyond the scope of this Chapter to examine this issue quantitatively.

The evaluation of whether assessments could be conducted using packages was based on packages used in the USA and New Zealand. Packages exist in other jurisdictions. In particular, Fisheries Library in R (FLR; [16]) includes several assessment tools, and is widely used in Europe, along with SAM [17], SURBA [18], XSA [19], a4a [20] and ICA [21]. However, these packages do not contain the features that would allow the 18 (or 27) assessments to be conducted using packages. In contrast, the package GADGET [22] is based on an age-size-area-structured population dynamics model, but this package has not been widely used outside of Iceland and Norway.

Depending on whether Stock Synthesis is considered to be adequate to implement age-size structured assessments, 49 or 58 of the 76 stock assessments could be conducted using freely-available packages that have been simulated tested to assess their performance and published in the peer-reviewed literature. This would be an increase of either 32 or 41 stock assessments that are conducted using packages. Beyond the reduced costs associated with maintaining and updating packages is that the review process for packages tends to be more rigorous. Table 2.5 and Supplementary Table 2.2 shows that the methods used for “unique” and “semi-genetic” stock assessments used in Australia tend not have been published in the peer-reviewed literature. Instead, these methods of stock assessment tend to be published in conference proceedings and reports to management and funding agencies. Review of some of these approaches has, however, involved publication in peer-reviewed scientific journals.

Adoption of packages would not only increase the likelihood that the method of assessment is correctly coded and documented, but also that analysts could spend more time focusing on analysis of the basic data on which assessments rely and/or conducting more assessments. As is the case in the USA, the use of a small number of packages should increase the availability of the pool of analysts capable of reviewing stock assessments, given increased familiarity with the input files, the output files, and the assumptions underlying the assessment method. In principle, use of packages could also increase the number of analysts capable of conducting assessments because such analysts would need to be trained primarily to apply available packages and interpret the results, rather than being involved with coding of models and their associated estimation methods. The ability of analysts to conduct more assessments tends to be higher for packages because of the availability of manuals, training and example data sets (Table 5), even though they require analysts who are considered to be experienced modellers.

Other advantages with the use of packages in contrast to “unique” assessments include the automatic provision of diagnostic tables and figures to evaluate model fits and outputs used for peer-review and management decision making. These diagnostics and outputs tend to be available for the assessments conducted using packages and “semi-generic” assessment methods, but less frequently for “unique” assessments (Table 2.5; Supplementary Table 2.2). In addition, there is a much higher frequency of maintenance plans for packages than for “semi-generic”, and particularly “unique”, assessment methods.

So why are so few stocks assessed using packages in Australia compared to other jurisdictions? There is likely not to be a single answer to this question, but the reasons include that there is no package that could be used to conduct the assessment because of the structure of the population dynamics model and because the assessment relies on a data source that cannot be used by a package (or that analysts believes this to be case). This problem could potentially be overcome if analysts in Australia work with package developers in the USA, New Zealand and perhaps Europe to modify their packages. For example, the availability of a package that represents populations using an age-size-sex-structured population dynamics model would seem to be valuable for many stocks and species. Similarly, extending packages such as Stock Synthesis or CASAL to include the use of tagging data to estimate fishing mortality (as well as growth and movement) would be beneficial to assessments in Australia and elsewhere. Another reason for the relatively large number of “unique” stock assessments is that several jurisdictions in Australia require management advice in the form of economic outputs. Australia is currently the only jurisdiction that sets catch limits to achieve economic objectives so it is hardly surprising that existing stock assessment packages do not feed their results into bio-economic models.

Another reason for not using packages relates to the nature of the assessment processes used in the various Australian jurisdictions. There appears to be a desire for stock assessments to be “in-house” in Australia so that it is possible for analysts to make changes to the assessment as needed. This is certainly a valid concern [11]. However, experience in the SESSF where almost all of the assessments are based on Stock Synthesis is that it is possible to work with the model developers to make changes to ensure that the package addresses jurisdiction-specific issues. The prevalence of “in house” assessments in Australia may also be a reflection of the lack of a forum for stock assessment scientists to discuss common issues. Such a forum exists in the USA due to the availability of, for example, National Stock Assessment Workshops, and the Center for the Advancement of Population Assessment Methodology (CAPAM; www.**capam**research.org/). The latter group is developing best practices for stock assessment (e.g. [23]) that will feed into the development process for assessment packages.

In conclusion, it could be possible to increase the number of model-based stock assessments in Australia. However, this would require increased collaboration across jurisdictions. The benefits for doing so could be substantial and lead to a greater number of assessments.

REFERENCES

[1] J. Alder, S. Cullis-Suzuki, V. Karpouzi, K. Kachner, S. Mondoux, W. Swartz, P. Trujillo, R. Watson, D. Pauly, Aggregate performance in managing marine ecosystems of 53 maritime countries, Marine Policy 34 (2010) 468–476.

[2] T.J. Pitcher, D. Kalikoski, K. Short, D. Carkey, G., Pramod, An evaluation of progress in implementing ecosystem-based management of fisheries in 33 countries, Marine Policy 33 (2009) 223–232.

[3] T.J. Pitcher, D. Kalikoski, G. Pramod, K. Short, Not honouring the code, Science 457 (2008) 658–659.

[4] M.J. Flood, I. Stobutzki, J. Andrews, C. Ashby, G.A. Begg, R. Fletcher, C. Gardner, L. Georgeson, S. Hansen, K. Hartmann, P. Hone, P. Horvat, L. Maloney, B. McDonald, A. Moore, A Roelofs, K.A. Sainsbury, T. Saunders, A.D.M. Smith, C. Stewardson, J. Stewart, B.S. Wise, Status of Key Australian Stocks Reports 2014, Fisheries Research and Development Corporation, Canberra, Australia, 2014, pp. 622. http://fish.gov.au/Pages/SAFS_Report.aspx (last accessed 22 September 2016).

[5] OECD, OECD Review of Fisheries: Policies and Summary Statistics 2015, OECD Publishing, Paris, 2015.

DOI: http://dx.doi.org/10.1787/9789264240223-en

[6] Department of Agriculture and Water Resources DAWR. Commonwealth Fisheries Harvest Strategy Policy Guidelines, Australian Government Department of Agriculture, Fisheries and Forestry, Canberra, Australia, 2007, pp. 55. http://www.agriculture.gov.au/fisheries/domestic/harvest_strategy_policy (last accessed 25 July 2016).

[7] N. Rayns, The Australian government's harvest strategy policy. ICES Journal of Marine Science 64 (2007) 596–598.

[8] R. McGarvey, J.M. Matthews, J.E. Feenstra, A.E. Punt, A Linnane, Using bioeconomic modeling to improve a harvest strategy for a quota-based lobster fishery, Fisheries Research 00 (In press) 00-00.

[9] C. Gardner, K. Hartmann, A.E. Punt, S. Jennings, B. Green, Bio-economics of commercial scale translocation of southern stock lobster, Fisheries Research 162 (2015) 29–36.

[10] S.R. Sloan, A.D.M. Smith, C. Gardner, K. Crosthwaite, L. Triantafillos, B., Jeffries, N. Kimber, N. National Guidelines to Develop Fishery Harvest Strategies. FRDC Report – Project 2010/061. Primary Industries and Regions, South Australia, Adelaide, March. CC

BY 3.0, 2014, http://frdc.com.au/research/Final_Reports/2010-061-DLD.pdf (last accessed 28/7/2016).

[11] C.M. Dichmont, R. Deng, A.E. Punt, J. Brodziak, Y-J. Chang, J.M. Cope, J.N. Ianelli, C.M. Legault, R.D. Methot, C.E. Porch, M.H. Prager, K. Shertzer, A review of stock assessment packages in the United States.,Fisheries Research 183 (2016) 447-460.

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

[13] M. Tanimoto, J.B. Robins M.F. O'Neil, I.A. Halliday, A.B. Campbell, Quantifying the effects of climate change and water abstraction on a population of barramundi (*Lates calcarifer*), a diadromous estuarine finfish, Marine and Freshwater Research 63 (2012) 715–726.

[14] A.E. Punt, R.A. Deng, C.M. Dichmont, T. Kompas, W.N. Venables, S. Zhou, S., Pascoe, T. Hutton, R. Kenyon, T. van der Velde, M. Kienzle, Integrating size-structured assessment and bio-economic management advice in Australia's Northern Prawn Fishery, ICES Journal of Marine Science 67 (2010) 1785–1801

[15] I.G. Taylor, R.D. Methot, Hiding or dead? A computationally efficient model of selective fisheries mortality, Fisheries Research 142 (2013) 75–85.

[16] L.T. Kell, I. Mosqueira, P. Grosjean, J-M. Fromentin, D. Garcia, R. Hillary, E. Jardim, S. Mardle, M.A. Pastoors, J.J. Poos, F, Scott, R.D. Scott, FLR: an open-source framework for the evaluation and development of management strategies, ICES Journal of Marine Science 64 (2007) 640–646.

[17] A. Nielsen, C.W. Berg, Estimation of time-varying selectivity in stock assessments using state-space models, Fisheries Research 158 (2014) 96–101.

[18] C.L. Needle, Survey-based assessments with SURBA, Working Document to the ICES Working Group on Methods of Fish Stock Assessment, Copenhagen, February 2003, 2003.

[19] J.G. Shepherd Extended survivors analysis: An improved method for the analysis of catch-at-age data and abundance indices, ICES Journal of Marine Science 56 (1999) 584–591.

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

[21] K.R. Patterson, G.D. Melvin, Integrated catch at age analysis version 1.2. Scottish Fisheries Research Report, 56. Aberdeen: FRS, 1996, 66pp. http://www.gov.scot/Uploads/Documents/No%2058.pdf (last accessed 25 July 2016).

[22] J. Begley, Gadget User Guide., 2012, http://www.hafro.is/gadget/files/userguide.pdf (last accessed 25 July 2016).

[23] P.R. Crone, M.N. Maunder, J.L. Valero, J.D. McDaniel (Eds.), Workshop Series Report Center for the Advancement of Population Assessment Methodology (CAPAM). NOAA/IATTC/SIO. 8901 La Jolla Shores Dr., La Jolla, CA 92037, 2013, 46 pp.

[24] Anon. Assessment Model for Alaska Description of GUI and Instructions, 2015, Available from https://github.com/NMFS-toolbox/AMAK/blob/master/docs/AMAK%20Documentation.pdf (last accessed 25 July 2016).

[25] C.M. Legault, V.R. Restrepo, A flexible forward age-structured assessment program, Collective Volume of Scientific Papers ICCAT 49 (1998) 246–253.

[26] M.H. Prager, ASPIC: A Surplus-Production Model Incorporating Covariates, Collective Volume of Scientific Papers ICCAT 28 (1992) 218–229.

[27] M.H. Prager, A suite of extensions to a nonequilibrium surplus–production model. Fishery Bulletin 92 (1994) 374–389.

[28] M.H. Prager, Comparison of logistic and generalized surplus-production models applied to swordfish, *Xiphias gladius*, in the north Atlantic Ocean, Fisheries Research 58 (2002) 41–57.

[29] E.H. Williams, K.W. Shertzer, Technical documentation of the Beaufort Assessment Model (BAM), NOAA Technical Memorandum NMFS-SEFSC-671, 2015, 43 pp.

[30] J. Brodziak, A. Yau, J. O'Malley, A. Andrews, R. Humphreys, E. DeMartini, M. Pan, M., Parke, E. Fletcher, Stock assessment update for the main Hawaiian Islands Deep 7 bottomfish complex through 2013 with projected annual catch limits through 2016, U.S. Dep. Commer., NOAA Tech. Memo., NOAA-TM-NMFS-PIFSC-42, 2014, 61 pp.

[31] M.K. McAllister, A generalized Bayesian surplus production stock assessment software (BSP2), Collective Volume of Scientific Papers ICCAT 70 (2014) 1725–1757.

[32] B. Bull, R.I.C.C. Francis, A. Dunn, A., McKenzie, D.J. Gilbert, M.H. Smith, CASAL (C++ algorithmic stock assessment laboratory): CASAL user manualv2.07-2005/08/21, NIWA Technical Report 127 (2012).

[33] J.S. Collie, M.P. Sissenwine, Estimating population size from relative abundance data measured with error, Canadian Journal of Fisheries and Aquatic Sciences 40 (1983) 1871–1879.

[34] D.A. Fournier, J. Hampton, J.R. Sibert, MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*, Canadian Journal of Fisheries and Aquatic Sciences 55 (1998) 2105–2116.

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

[36] J.M. Cope, Implementing a statistical catch-at-age model (Stock Synthesis) as a tool for deriving overfishing limits in data-limited situations, Fisheries Research 142 (2013) 3–14.

[37] R. Conser, J.E. Powers, Extensions of the ADAPT VPA tuning method designed to facilitate assessment work on tuna and swordfish stocks, Collective Volume of Scientific Papers ICCAT 32 (1990) 461-467.

[38] C.E. Porch, S.C. Turner, Virtual population analyses of Atlantic bluefin tuna with alternative models of trans-Atlantic migration, Collective Volume of Scientific Papers ICCAT 49 (1999) 291–305.

[39] V.R. Restrepo, C.E. Porch, Options for conducting swordfish sex-specific assessments, Collective Volume of Scientific Papers ICCAT 51 (2000) 1375–1386.

[40] C.E. Porch, V.R. Restrepo, S.C. Turner, G.P. Scott,. Virtual population analyses of Atlantic bluefin tuna incorporating movement and tagging data, Collective Volume of Scientific Papers ICCAT 44 (1995)183–190.

[41] C.E. Porch, S.C. Turner, J.E. Powers, Virtual population analyses of Atlantic bluefin tuna with alternative models of transatlantic migration: 1970-1997, Collective Volume of Scientific Papers ICCAT 52 (2001) 1022–1045.

[42] J. Cope, E.J. Dick, A. MacCall, M. Monk, B. Soper, C. Wetzel, Data-Moderate Stock Assessments for Brown, China Copper, Sharpchin, Stripetail, and Yellowtail Rockfishes and English and Rex Soles in 2013. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220, 2015, 298 pp. http://www.pcouncil.org/wp-content/uploads/Data-Moderate_Assessments_2013_FINAL_160116.pdf (last accessed 25 July 2016).

[43] C.R. Wetzel, A.E. Punt, Evaluating the performance of data-moderate and catch-only assessment methods for U.S. west coast groundfish, Fisheries Research 171 (2015) 170–187.

[44] M. Flood, I. Stobutzki, J. Andrews, C. Ashby, G. Begg, R. Fletcher, C. Gardner, L. Georgeson, S. Hansen, K. Hartmann, P. Hone, P. Horvat, L. Maloney, B. McDonald, A. Moore, A. Roelofs, K. Sainsbury, T. Saunders, T., Smith, C. Stewardson, J. Stewart, B. Wise, (Eds.), Status of Key Australian Fish Stocks Reports 2014. Fisheries Research and Development Corporation, Canberra, 2014, http://www.fish.gov.au/pages/safs_report.aspx (last accessed 25 July 2016).

[45] R. McGarvey, J.E. Feenstra, Q. Ye, Modeling fish numbers dynamically by age and length: partitioning cohorts into 'slices', Canadian Journal of Fisheries and Aquatic Sciences 64 (2007)1157–1173.

[46] C.W. Clark. The Economics of Overexploitation, Science 181 (1973) 630-634.

[47] C.W. Clark, G.R. Munro,U.R. Sumaila, U.R. Limits to the Privatization of Fishery Resources, Land Economics 86 (2010), 209-218.

SUPPLEMENTARY TABLES

Supplementary Table 2.1. The survey questionnaire completed for each assessment. "&" indicates that multiple options are available, including "other" and * indicates that text providing more information was requested. In cases where "other" was selected, additional text was requested.

Questions	Dropdown options
A. Overview of the Assessment Techniques	
Software name	
Owner organisation(s)	
Developer(s)	
When was the last major version released	
When was it last used for an assessment?	
Has it been published?	internal report, external report, journal article, book chapter, conference proceeding[&]
Documentation of any external peer-review of the **assessment method**	
Have **applications** been subject to external peer-review?	Yes / No / Unknown*
For which fishery was the method initially developed?	
Has it been used for multiple fisheries?	Yes / No / Unknown*
Designed for generic use?	Yes - flexible, No – specialised*
Which types of fisheries can it support?	TAC, TAE[&]
B. Use of the software	
Is the source code available?	Yes / no / unknown*
Is a complied version available?	Yes / no / unknown*
Is training available?	Yes / no / unknown*

Is there a manual and test data set?	Yes / no / unknown*
What skills do users have to have to use the model?	Model developer / experienced modeller / beginner&

C. Software and development

Programming language?	ADMB,C, C++, Excel, FORTRAN, JAVA, Matlab, VB, VB-net, WinBUGS&
Implemented for which platforms?	Windows, Unix, Linux&
Does code exist to read outputs?	Yes / no / unknown*
Is there an automatic system to produce diagnostics?	Yes / no / unknown*
What is the process of modification?	
Is the model regularly maintained?	Yes / no / unknown*
Are on-going resources dedicated to support model maintenance and development?	Yes / no / unknown*

D. Characteristics of the assessment

Does the model include spatial structure?	Yes / no / unknown*
Does the model include multiple fleets?	Yes / no / unknown*
What is the species-structure of the model?	Single-species / multi-species / multiple species*
Does the model include multiple stocks of a species?	Yes / no / unknown*
What is the temporal resolution of the model?	
Is the model deterministic or stochastic?	Deterministic, stochastic, unknown (* if stochastic)
What data can it use?	Catch data, CPUE, effort data, tagging data, survey data, length-frequency data, age-frequency data, weight-frequency data, conditional age-at-length data &
What time-series of outputs does it provide	Biomass (B), numbers, fishing mortality(F), B/B_{ref}, F, F/F_{ref} &
What reference points?	B_0, B_{MSY}, F_{MSY}, BMEY, $F_{X\%}$, $B_{X\%}$, none&
Can it use data from commercial, recreational, indigenous fisheries?	commercial, recreational, indigenous fisheries, all &

What kind of projections can be conducted?	Catch, effort[&]
Under which circumstances would the use of this application be inappropriate?	
E. Characteristics of uncertainty	
Does it characterize uncertainty	Yes / no / unknown*
What fit diagnostics are available?	
Can one easily undertake sensitivity tests	Yes / no / unknown*
Has the method been simulated tested?	Yes / no / unknown*

Supplementary Table 2.2. Summary of the 31 assessments conducted using “unique” stock assessment methods.

Assessment method	Type of User	Training available?	Manual?	Regular maintenance ?	Diagnostics produced automatically?	Methods for quantifying parameter uncertainty	Basis for projections?	Basis for method review
VIC Abalone	Experienced	No	No	No	No	Bayesian	Catch	Independent review
QLD Banana prawn	Beginner, experienced	No	No	No	Yes	Asymptotic	Catch, Effort	No
QLD Barramundi - Fitzroy	Beginner, experienced	No	Yes	No	Yes	Asymptotic	Catch, Effort	Tamimoto *et al.* (2012)
QLD Barramundi - East Coast and Gulf of Carpentaria	Developer	No	No	Yes	No	Bayesian	Catch, Effort	No
QLD Coral Trout	Developer	No	No	Yes	No	Bayesian	Catch, Effort	No
PZJA Endeavour Prawn - Torres Strait	Beginner, experienced	No	No	No	Yes	Asymptotic	Catch, Effort	No
PZJA Endeavour Prawn - Torres Strait	Beginner, experienced	No	No	No	Yes	Asymptotic	Catch, Effort	No
QLD Grey Mackerel	Developer	No	No	No	Yes	Bayesian		No
FEDERAL Gummy Shark	Experienced	No	No	Yes	No	Asymptotic	Catch	No
QLD Morton Bay Tiger and Eastern King and Greasyback Prawns	Experienced	No	No	No	No	Unknown	Catch, Effort	No
NT Mud Crab	Experienced	Yes	No	Yes	No	Asymptotic	Catch, Effort	No
QLD Mullet	Beginner, experienced	No	No	No	Yes	Asymptotic	Catch	No
QLD Pink Snapper	Beginner, experienced	No	Yes	Yes	Yes	Bayesian	Catch, Effort	No
FEDERAL Red-legged Banana Prawns	Beginner, experienced	No	No	Yes	Yes	Asymptotic	Catch, Effort	No

QLD Red Throat Emperor	Developer	No	No	No	Yes	Asymptotic	Catch, Effort	No
QLD Scallop-1	Beginner, experienced	No	No	No	Yes	Asymptotic	Catch, Effort	No
QLD Scallop-2	Developer	No	No	Yes	No	Bayesian	Catch, Effort	No
FEDERAL School Shark	Experienced	No	No	Yes	No	No	Catch	No
QLD shark	Developer	No	No	Yes	No	Bayesian	Catch, Effort	Independent review
QLD Snapper	Experienced	No	No	No	Yes	Bayesian	Catch, Effort	Independent review
CCSBT Southern Bluefin Tuna	Experienced	No	Yes	Yes	Yes	Asymptotic, Bayesian	Catch	Independent review
SA Southern rock lobster - Northern Zone	Developer	No	No	Yes	No	No	Catch	No
QLD Spanish Mackerel	Developer	No	No	Yes	No	Bayesian	Catch, Effort	No
QLD Tailor	Experienced	No	No	Yes	Yes	Asymptotic	Catch, Effort	No
Qld Tiger Prawn - Morton Bay	Experienced	Yes	Yes	Yes	No	Asymptotic	Catch, Effort	Kienzle *et al*. (2014)
FEDERAL Tiger and Endeavour Prawn	Experienced	No	No	Yes	Yes	Bayesian	Catch	Zhou *et al*. (2009)
FEDERAL Tiger and Endeavour Prawn	Experienced	No	No	Yes	Yes	Asymptotic	Catch, Effort	Punt *et al*. (2010)
FEDERAL Tiger and Endeavour Prawn	Experienced	No	No	Yes	Yes	Asymptotic	Catch, Effort	No
FEDERAL Tropical Rock Lobster	Experienced	No	No	Yes	Yes	Asymptotic	Catch, Effort	No
QLD Tropical Snapper	Beginner	Yes	No	Yes	Yes	Asymptotic	Catch, Effort	No
WA Western Rock Lobster	Experienced, developer	No	No	Yes	Yes	Bayesian	Catch, Effort	Independent review

SUPPLEMENTARY TABLE 2.2 REFERENCES

Kienzle, M., Courtney, A. J., and O'Neill, M. F. (2014). Environmental and fishing effects on the dynamics of brown tiger prawn (*Penaeus esculentus*) in Moreton Bay (Australia). *Fisheries Research* **155**, 138–148.

Punt, A. E., Deng, R. A., Dichmont, C. M., Kompas, T., Venables, W. N., Zhou, S., Pascoe, S., Hutton, T., Kenyon, R., van der Velde, T., and Kienzle, M. (2010). Integrating size-structured assessment and bio-economic management advice in Australia's Northern Prawn Fishery. *ICES Journal of Marine Science* **67**, 1785–1801

Tanimoto, M., Robins J. B., O'Neil, M. F., Halliday, I. A., and Campbell, A. B. (2012). Quantifying the effects of climate change and water abstraction on a population of barramundi (*Lates calcarifer*), a diadromous estuarine finfish. *Marine and Freshwater Research* **63**, 715–726.

Zhou, S., Punt, A., Deng, R., Dichmont, C., Ye, Y., and Bishop, J. (2009). Modified Bayesian biomass dynamics model for assessment of short-lived invertebrates: a comparison for tropical tiger prawns. *Marine and Freshwater Research* **60**, 1298–1308.

Supplementary Table 2.3. Stock assessment packages and the number of Australian "Unique", "Semi-generic" or "Package" stock assessments that could be conducted using each package. The "Initial" list was based on the population dynamic structure of the model and the "Final" list considered data types and other unique characteristics of the assessment.

Package abbreviation	Australian Package					
	Unique (31)		Semi-generic (28)		Package (17)	
	Initial	Final	Initial	Final	Initial	Final
AMAK	18	6	19	6	14	0
ASAP	18	6	19	6	14	0
ASPIC	3	1	0	0	0	0
BAM	18	10	19	13	14	1
BSP1	3	0	0	0	0	0
BSP2	3	0	0	0	0	0
CASAL	27	12	24	21	17	10
CSA	3	2	2	0	0	0
GMACS	2	1	2	2	0	0
MULTIFAN-CL	18	6	19	6	14	0
SCALE	18	6	19	6	14	0
SS	25	15	22	22	17	17
SSS	18	2	19	2	14	0
VPA	18	6	19	6	14	0
VPA2BOX	18	6	19	6	14	0
XDB-SRA	3	0	0	0	0	0
XSSS	18	6	19	6	14	0
Any	30	16	24	24	17	17

CHAPTER 3. PRESENT AND FUTURE STOCK ASSESSMENT NEEDS

INTRODUCTION

This chapter is final stage of the project. Freely available packages that align with Australia's needs have been reviewed in Chapter 1 and Australian assessments have been collated and aligned in terms of their relative uniqueness (Chapter 2). In this Chapter, the key questions being asked are:

1. What are the different stock assessment research provider's needs and priorities,
2. What is their capacity and willingness to use packages as well as bespoke models, and
3. What is their investment model in stock assessment?

METHODS

Semi-structured interviews were undertaken by Drs Little and Dichmont over a two-week period covering each of eight jurisdictions that provide stock assessment expertise in fisheries – Northern Territory (NT), Queensland (Qld), New South Wales (NSW), Victoria (Vic), Tasmania (Tas), South Australia (SA), Western Australia (WA) and the Commonwealth (CSIRO) (Table 3.1). Only a single interview was undertaken per jurisdiction, but multiple staff members could attend. Most of those interviewed stated they also sought input from other staff and managers. The results from Chapters 1 and 2, together with the survey questions were generally provided to the interviewees in advance of the phone hook-up. Interviewees could provide written input beforehand and afterward. Several jurisdictions used this opportunity to add and clarify their comments. A summary of the interviews was also provided to the group, as well as this chapter for their perusal and comment.

The survey questions are provided as Supplementary material to this chapter. The survey was divided into four broad sections covering the jurisdiction's stock assessment Vision, Business case approach, Capacity and Needs. The 'Key to the Needs' section was the question asking the interviewees to identify their top three stock assessment needs.

Table 3.1: List of interviewees by jurisdiction. Note: interviewees also sought input from their own organisations before their interview.

Jurisdiction	Interviewee(s)
NT	Thor Saunders
Qld	Mick O'Neill
NSW	Geoff Liggins
Vic	Paul Hamer, Harry Gorfine, Mellissa Schubert
Tas	Klaas Haartman, Craig Mundy
SA	Stephen Mayfield, Rick McGarvey
WA	Brent Wise, Norm Hall, Ainsley Denham, Alex Hesp
CSIRO	Geoff Tuck, Éva Plagányi

The results were summarised by linking comments to keywords and then totals of how often these were mentioned. Scores in terms of how well the vision and business case was defined based on the interviewees' response were also provided by the project team. Similarly, for the top three needs (and additional needs mentioned elsewhere during the interview), a word cloud was created removing obvious words or terms such as "stock assessment" etc. (see report cover).

Interviewees were also asked as part of the Capacity section, on how well they linked with Australian universities. This was not part of the original set of questions, but became important during the discussions.

RESULTS AND DISCUSSION

A summary of key points addressed during the interview is provided in Table 3.2. More details were provided, but not included to avoid identity of the jurisdiction becoming obvious. There was quite a diversity of comments on how well stock assessment funding and capacity are linked with overarching policies. In some jurisdictions, this link was strong as well as transparent, in some strong but less transparent, and in others it was seen as extremely weak. The eastern States stood out in that they are in the process of reviewing their fisheries policies (but at various stages in that process) and this creates difficulties to predict their short- to medium-term needs in terms of stock assessment (as one would expect).

There are also some jurisdictions that are "rich" in stock assessment capability and these tend to respond differently in terms of impediments and needs. Some jurisdictions have enough capacity to undertake their own stock assessments and compete for external stock assessment projects (but mainly when there are synergies back to their own assessment needs) (SA, CSIRO, Tas, Qld), whereas in the other extreme, NT and Vic intend to mainly outsource their stock assessments. WA mainly concentrates on undertaking their own assessments. Those jurisdictions that outsource did not express a concern about access to stock assessment skills.

There were common discussion points about difficulties concerning lack of succession planning within Australia and several of the jurisdictions. Comments about a program of building assessment expertise in the 1990s were mentioned as being successful. Several jurisdictions struggle to obtain new or replacement stock assessment staff – they may have many applicants, but few have the necessary skills required for the job. This is potentially a growing issue as more of the existing stock assessment staff start changing their roles through reducing their time or retiring.

However, linking with universities to build this capacity was mostly not successful for several jurisdictions. The main reasons for this were that:

- Most Australian universities tend not to train students in the skills needed for stock assessment,
- students with the right types of skills entering university tend to move to more lucrative fields,
- training students is not core business in most of the interviewees agencies yet requires enormous inputs from staff (also related to the issue that the skills are not appropriate without further training), and
- good students cannot be kept after graduation, because of lack of flexibility or employment policies within the jurisdictions.

These results therefore highlighted that this issue still needs to be addressed despite initiatives already underway.

Table 3.2. Summary scores of interviews for the different jurisdictions. BC – Business Case; U – model user; D – model developer; B – Both; I – mainly in-house capacity; O – mainly outsourcing stock assessment; Y – Yes; N – No; G – Government; R – Recreational levy; I – industry; SLA – Service Level Agreement; MOU – Memorandum of Understanding.

Jurisdiction*	Clear policy	Clear policy to BC process	Policy and BC freely available	Funding source	BC process	Model user (U) or developer (D) or Both (B)?	In-house (I) or outsource (O) or Both (B)?	Undertake assessments in other jurisdictions?	Build capacity	Comment on ability to build capacity	Comment on links to uni's and the effectiveness thereof	Size (FTE)	Bespoke view	Database process
1	High	High	High	G	SLA	D	I	N	Y	No good applicants even if have positions		5	Both	Excellent
2	High	Medium	Medium	G	MOU	D	I	Y	N	Difficult to get good applicants	Too theoretical; too much input required; not well trained for purpose	9	Both	Good
3	Low	Low	Low	G; R; I	Depends on funding source	U	B	N	N			1	Package	Poor
4	Medium	Medium	Low	G	Direct with Minister	U	O	N	N	Rely on getting external input. Can be relatively easily expanded		1	Both	Poor
5	Low	Medium	Low	G	SLA	D	I	Y	N	Few good applicants	Hard to attract good maths/stats people and retain people trained in uni	3	Both (but own package)	Good
6	Medium	Medium	Medium	G	SLA	D	I	Y	N	Few good applicants - long time to get	Uni's don't have capacity or right skills	4	Bespoke	Good
7	Low	High	Medium	G	SLA	B	I	Y	N			2	Both	Excellent
8	Low	Low	Low	G	SLA	U	B	N	N		Uni's do not have right skills	1	Package	Poor

* Using codes to represent jurisdictions to protect their identities.

Jurisdictions were highly reliant on their various databases (commercial logbook, biological, fishery independent and recreational survey data), but particularly reliant on good quality control and easy access. This was a major issue in many jurisdictions, which varied from Tas that is aiming (and is well underway) for seamless, almost real-time online provision of data and reporting services, to Vic and NT that have an old database that is difficult to access, requires skill to extract and clean, and is time consuming to use. Qld also has online data system available to the public.

NT, NSW and Vic mentioned that accessing data from the data managers of their database group was an impediment to timely delivery. The need for seamless and regular linking to the data, almost real-time entry of data and automation of reports was a very consistent discussion point. However, this has only been achieved in part in Tas, which stands out in how far they are down this process. Interviewees were very interested in sharing the technology of how to automate reporting systems and streamlining data entry, extraction etc. There would be real advantage in setting up a system that would allow this sharing of ideas, knowledge and code.

The following skills (although overlapping to some degree) are what jurisdictions look for in stock assessment scientists: Applied statistics (mentioned by 4 jurisdictions), Mathematics (2), GIS (1), Data analysis (2), Database (2), Data-limited assessments (1), Modelling (2), CPUE standardisation (1), Population dynamics (1), Programming (2), Knowledge of assessments (1), Communication (1), Management strategies (1) and Applied outcomes (1). From these answers, a wide range of analytical skills are required.

None of the jurisdictions prescribed software their stock assessment staff are required to use. That tended to allow the assessor to choose his or her own in this regard. There was a sense that R (https://CRAN.R-project.org/) and AD Model Builder (ADMB: www.admb-project.org/) were favoured and there was some expectation that their staff know these software products. Indeed, seven of the eight jurisdictions listed R as being used and four mentioned ADMB. There seemed to be an eclectic use of software tools beyond these two including Fortran™ (3), Excel™ (3), C or C++™ (2), Stock Synthesis [Methot (1990); Methot and Wetzel (2013): 2], Matlab™ (1), Genstat™ (1), Mathematica™ (1), CASAL (Bull et al., 2005: 1), sql server™ (1), perl™ (1) and SAS™ (1). Many have moved away from the professional software products such as SAS and Matlab in favour of freeware such as R and ADMB due to the reduction in software costs (although some may argue these increase time taken to learn and use these products).

Jurisdiction's needs (if they were to move to greater use of assessment packages) were that there are no Australian-centric stock assessment packages that link to Australia's policy systems; there is no length-based model packages specifically not bio-economic length-based models (i.e. there is no widely supported and freely available package for lobsters, abalone and prawns); that any package being used in Australia is open source and is well reviewed, and trying to find the time and capacity to learn a package. Impediments to moving to packages (Table 3.3) overlapped with their needs and were mainly about in-house capacity; time to learn package; length of time to transition to a package and in-house technical problems e.g. installing non-official programs on computer.

Table 3.3. Impediments to transitioning to packages

Keyword	Frequency
In-house capacity	2
None	2
Lack of bio-economic length-based model package for hard to age animals	4
Time	2
Not interested in transitioning	2
Length of time to transition	2
In-house technical problems e.g. installing non-official programs on computer	1
Time to learn package	1
Need to work with bespoke and packages in parallel	1
Outsourced person using package not working as a team	1
No succession plans	1
Biometricians shared with all agriculture	1
Lack of Australian-centric packages that link to our harvest strategies and ecology	3

Interest in moving to packages was varied and depended on the purpose of application and need. The main reason against such movement was that models have already been tailored to suit many fisheries. Often using both bespoke and packages combined for the same species was seen as an advantage – it was seen as good to test different methodologies. Those jurisdictions that outsource tended to prefer packages as long as the assessor was qualified to use the package. Often this process of outsourcing using an external expert was seen as preferable, because the model could be run by someone internally after the development phase was completed. Two jurisdictions have developed internal "packages". For example, Qld has developed their own age-based model, covers dynamics for age/sex/time/length/growth-type group/fleets, that they use on finfish fisheries. They also intend to release it at some stage for general use i.e. it will become a freely available and supported package. SA, Vic and Tas share a length-based model developed by André Punt called LenMod which is being maintained in the GitHub version control system (www.github.com) and can be used and extended upon request to their needs. SA also had heavy investment in models such as qR (McGarvey et al., 2007) that they prefer over any package. CSIRO, on the other hand, uses both bespoke and packages as shown in Chapter 2. In terms of adapting packages to fit Australia's needs, CSIRO has worked with Dr Methot, the author of Stock Synthesis to make changes to his code to address Australian Government needs. Over the years this has been a very successful collaboration. This may be a risky strategy to rely on for all cases.

Possible steps to moving part-way or fully to using packages were:

- to share code libraries i.e. take what fits rather than move wholly to a package
- to share model components of existing bespoke models
- to share standard sets of codes rather than develop a single bespoke model
- create formalised sharing relationships through the Research Provider Network (RPN)
- create and share Australian bespoke model of models already well developed
- develop a rigorous review system of models and code
- create a generic bio-economic length-based model for Australia
- share bespoke models between jurisdictions e.g. LenMod
- use packages (but must be open source) as can then share with others and draw on global experience and review process; and
- develop capacity in existing packages, especially Stock Synthesis and CASAL.

From a stock assessment development perspective, criticisms of bespoke models were a lack of version control and a reliance on internal review systems, and a lack of developers with enough knowledge to identify and fix bugs that inevitably creep into the computer code. Sharing models through packages allows more developers to critique and contribute to the work.

There was in-depth discussion by all jurisdictions on how limited the inter-agency collaboration is in Australia with regard to detailed stock assessment modelling. Collaboration did occur within species groups e.g. rock lobster, and abalone, but less so between groups e.g. between finfish and lobster scientists. Collaboration also tended to occur around principles and broad methods, but less around technical details, coding and model development. Scientists felt isolated and that the standard of review process within and among agencies was not rigorous enough for them to learn. There was often not enough time to learn new methods and, if the time was available, there was no Australian based training on software such as Stock Synthesis, CASAL etc. They were reliant on the few times international stock assessment scientists (e.g. André Punt) visit and collaborate in Australia. Many mentioned the success of the two workshops held as part of the EU, Canada, New Zealand Knowledge based Bio-economy program where ideas were exchanged.

From a stock assessment user perspective, there was also a feeling that other people such as managers and industry do not understand stock assessment and what it takes to undertake an assessment. With continual pressure on budgets, stock assessment delivery is expected for less. A professional development course for managers and other users of stock assessment on basic stock assessment, without resorting to coding skills would address this, by providing an understanding of the constraints and demands, as well as pitfalls and weakness that assessment scientists confront.

A periodic course/workshop that delivers capacity building to fishery managers and other stock assessment users, could be coupled to a more technical course/workshop for hands-on users and developers of stock assessment. Such a forum would increase collaboration, and disseminate new techniques and allow discussion of issues. This could start with sharing methods on the new initiatives underway in several jurisdictions on cost-saving automated data streamlining and report writing processes.

The solutions listed below address the top three priorities of the jurisdictions (Figure 3.1) where the top responses included:

i. more and better quality data (“More and better data” in Figure 3.1) especially for data-limited, multi-species and recreational fisheries,
ii. a greater level of collaboration (feeling isolated and not growing in expertise) (“Collaboration”);
iii. build more qualified stock assessment expertise (few qualified applicants) (“More assessors”);
iv. reliable, open source bio-economic length-based model (“Length model”);
v. train students with assessment expertise (“Students’ expertise”); and
vi. proper assessment review process (beyond RAGs and similar bodies).

Other priorities of the jurisdictions included: clearer fisheries policies such as harvest strategies (“Clearer policies”); better understanding of data quality and clearer processes on how to fix these (“Better data quality”); improved extension/communication across Australia to understand assessment, outputs, limitations etc. (including appreciation of the field) (“Greater understanding”); streamlining data entry, extraction, analysis and reporting processes (including QA/QC links with assessors and data people) - near real time and automated (“Streamlining”); capability to support stock assessment, time constrained (“Added support”); consistency in fisheries and data management staff (“Consistency”); increased capacity whether internal or external (“Increased capacity”); improved data collection systems (“Better data systems”); sharing of data limited techniques (“Greater sharing”); difficulties assessing biomass (not just fishing mortality) given the data (“B estimation methods”); embracing new technologies for data collection (“New technologies”) and methods to avoid re-invention of the wheel (“Avoid re-invention”).

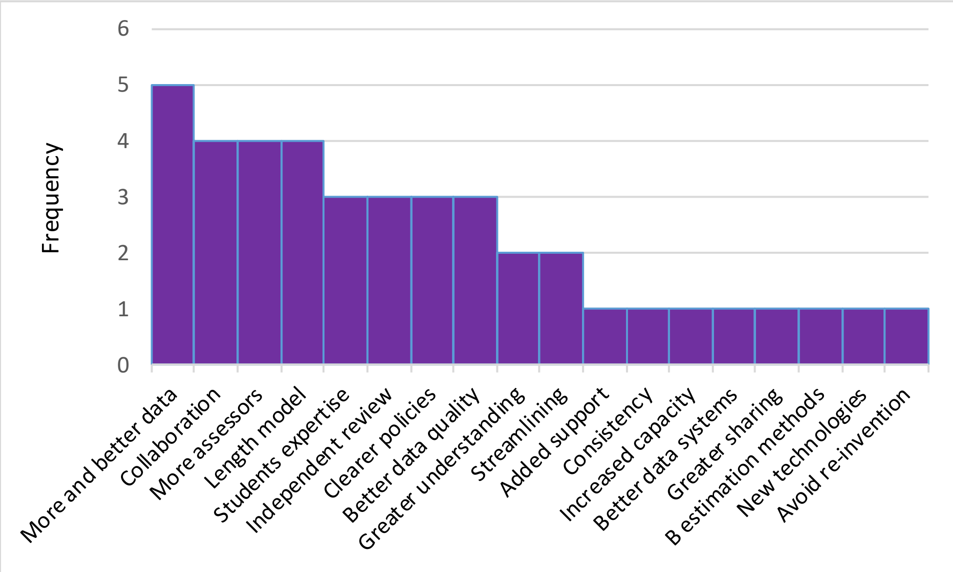

Figure 3.1. Number of times a jurisdiction mentioned a topic as a priority. Total interviews were 8

CONCLUSIONS AND RECOMMENDATIONS

1. Most jurisdictions have their own technologies that would be valuable to others. Sharing of these is difficult in the present environment. There is a need for a **national clearinghouse for stock assessment code and other streamlining tools** to foster sharing and collaboration among research providers.

2. A **national committee** through the Australian Fisheries Management Forum (AFMF) or RPN should be established with the goal to develop collaborations and share knowledge in stock assessment capability.

3. Most jurisdictions were open to using bespoke or packages depending on which was best for the task. The gradient of being predominantly against the use of packages to mostly using packages tended to be inversely proportional to the jurisdiction's own level of stock assessment expertise (and therefore awareness of desirable features often missing in packages) and investment in past assessments. **One common element was the need for a freshly developed bio-economic length-based model for species that are difficult to age. It is therefore recommended that a workshop is run bringing all the length-based assessors together, including from overseas, to determine the need for such a model and the steps required to develop and share such a model.**

4. There was a general view that the input from groups such as Resource Assessment Groups (if they existed at all) and the review process of scientific journals was not enough to fully capture the range of assessment work, and review a stock assessment. Resource Assessment Groups often lack the expertise to technically review a stock assessment (Penney et al. 2016). Although several options were discussed it was accepted that the

USA approach using the Centre for Independent Experts (https://www.st.nmfs.noaa.gov/science-quality-assurance/cie-peer-reviews/index) was too expensive, but a "lite" version of that may work, as well as approaches such as that undertaken in South Africa where a consistent review panel meet every year to discuss various assessments as part of a medium-term review program. **It is recommended that a small desktop review be undertaken on different international stock assessment peer review processes with a cost-benefit analysis.**

5. To allow greater sharing and collaboration of stock assessment scientists, **it recommended that a biennial or annual forum for assessors to share ideas and technologies should be developed similar to the USA National Stock Assessment meeting. This may be a mixture of less regular in-person meetings combined with other types of format. The priorities are:**

    a. **A workshop that shares techniques to streamline data collection, collation and reporting,**
    b. **Data-limited assessment techniques and available tools, and**
    c. **A course on CASAL or Stock Synthesis.**

6. To address the need for more understanding of stock assessment from non-assessment staff, **it is recommended that a stock assessment training course for managers and industry be run. This should link to existing courses and resources that are developed above.**

7. The most common statement made in terms of priorities was the need for **more and better quality data**. This was not a request for more and better data to develop more data-rich assessments, but rather a need for:
    a. More or even some data for data-limited fisheries so that a data-limited assessment could be undertaken,
    b. **More (frequent and species) recreational data,** the paucity of which is an impediment for assessing even quite valuable fisheries,
    c. **Better quality control processes of existing data being collected.** Enormous time and resources go into cleaning and checking data.

8. In common were concerns about the average age of stock assessment scientists in Australia. Attempts to link with universities tended not to be seen as successful due to blockages of good student supply all the way to absence of flexible employment systems to employ good graduates when they do occur. **A national discussion on succession planning is required, such as the past program of supporting the employment of stock assessment experts, a review of existing programs and how these can be improved and added investment in stock assessment development of young scientists.**

9. Jurisdictions, especially those that outsource their assessment work, stated that there was a demand by managers, scientists and industry to understand whether an assessment has followed best practice. Although Penney et al. (2016) provide a starting point by providing a standard guideline for research provision, it does not prescribe approaches to ensure or verify best practice e.g. sensitivity tests, retrospective analyses. **A guideline on stock assessment best practice for consistency between assessments should be developed.**

10. Despite their standardised approaches, many packaged stock assessments still remain expensive options for assessing the status of a stock, especially in terms of the human capital required to operate them. **Data-limited packages** are increasingly being used but, in general, lack statistical robustness, while statistical packages remain applicable only to stocks that can justify the costs (the risk-cost-catch trade-off: see Dowling et al., 2013; Dichmont et. al., 2016). There is increasing demand however to reduce the costs of statistical stock assessments, possibly through automation. Automated processes, in data acquisition through to model balancing are needed, and could reduce management costs. **A fisheries research priority should be developed by FRDC (after discussion with the Research Providers Network) for potential funding that includes data and stock assessment streamlining. This priority should as a first step articulate the steps required, where sharing of approaches would beneficial and the cost benefits thereof.** The risk of automating the stock assessment process to a greater degree than currently experienced is that an automated process would lose the human eye, or capability of checking the output. A balance between automation and fully manual would need to be addressed on a case by case basis, but tools on how to do this should be **shared** (as recommended above).

REFERENCES

Bull, B., Francis, R.I.C.C., Dunn, A., McKenzie, A., Gilbert, D.J., Smith, M.H., 2005. CASAL (C++ algorithmic stock assessment laboratory): CASAL user manualv2.07-2005/08/21. NIWA Technical Report 127.

Dichmont, C.M., Fulton, E., Gorton, R., Sporcic, M., Dowling, N., Little, L.R., 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

Dowling, N., Dichmont, C.M., Smith, A., Smith, D.S. 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. DOI http://dx.doi.org/10.1016/j.marpol.2012.12.009.

R. McGarvey, J.E. Feenstra, Q. Ye, 2007. Modeling fish numbers dynamically by age and length: partitioning cohorts into 'slices', Canadian Journal of Fisheries and Aquatic Sciences 64, 1157–1173.

Methot, R.D., 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. Bull. Int. North Pacific Fish. Comm. 50, 259–277.

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.

Penney, A.J., Bromhead, D., Begg, G., Stobutzki, I., Clarke, S., Little, R., Saunders, T., Martin, J. 2016. Research and science information guidelines for Australian Fisheries. FRDC Project 2014-009, 17 pp.

SUPPLEMENTARY MATERIAL: INTERVIEW SURVEY QUESTIONS

QUESTIONS REGARDING STOCK ASSESSMENT STRATEGY, CAPACITY AND NEEDS

ETHICS STATEMENT

This interview will be in a free-flowing format where we will only be keeping handwritten notes i.e. it will not be recorded. Any mention of specific names and figures will not be made publicly available in the minutes. Any commercial-in-confidence information (when we are made aware of this) would also not be made available or recorded in the minutes. We endeavour to keep the final reports more descriptive unless you feel that a more factual account would be better. You are welcome to ask your own questions.

BACKGROUND

An FRDC funded project entitles “Stock assessment review integration” was supported by AFMF to investigate the strategic and operational use of stock assessment within Australia and compare these practices with those from overseas. The objectives of the project are to a) review existing stock assessment methods used in Australia; b) review Australian stock assessment needs, and model developer and user capacity; c) review methods used and reviews undertaken elsewhere in the world; d) assess the relative merits of off-the-shelf versus case-specific assessments and **e) with input from the different jurisdictions, provide recommendations for a possible set of investment models**.

In this interview, we would like to address e), since we have received extensive information from you on your existing assessments.

Attached is a review of Australia’s stock assessment output, based on responses we got from a survey undertaken in 2015. We have calculated how many assessments your organization and others undertake and for which jurisdiction, and also some key characteristics of these assessments. This is attached to the questionnaire for your perusal – we hope it is of value to you.

We have also analysed whether any of these stock assessments could have been undertaken by an off-the-shelf stock assessment package or not. **It would be greatly appreciated if you could discuss the results with your stock assessment leaders whether our analyses are technically correct.** The key finding of the work is that many of the assessments undertaken in Australia are bespoke home-grown software developed in-house that is rarely independently reviewed. In contrast, the USA for example, has moved to using many more packages where these are simulation tested and independently reviewed. This means that the only outstanding task of a reviewer for a specific case study is to review whether the application of the software tool is appropriate.

Of course, using software packages versus home grown software has many trade-offs. Learning to use a package can be time consuming initially and flexibility of what can be done can also be reduced. At the same time, if a team of assessors use the same package there is much more within team collaboration and less reliance on a single assessor. Software bugs tend to be reduced with the use of a package, but there is reliance on the developer to maintain the product.

DEFINITIONS

In our review, we have defined "stock assessments" as model-based assessments used for tactical fisheries management that includes an optimisation and statistical estimation component.

We would also like you to think of you and your staff capacity in stock assessment in terms of model developers (those that can code and write new stock assessment models) and model users (those that are proficient in using other models including checking for errors and poor convergence).

STOCK ASSESSMENT STRATEGIC VISION

1. What is your strategic vision for stock assessment in your jurisdiction?
    a. Present?
    b. Future?
2. Do you have a written strategic plan for your stock assessment capacity and funding?
3. What are the key points within your strategic plan?
4. Is it freely available to others? (source)
5. Do you think the above will change over the next five years? If so, how and why?
6. Any comments?

STOCK ASSESSMENT BUSINESS CASE

1. Do you have a stock assessment business case?
2. What are the main drivers of this case?
3. If so, what are the key ingredients and over which period does it apply?
4. Do you see your organisation as being mainly about a) developing in-house stock assessment capacity, b) outsourcing capacity or c) both?
5. Are you trying to transition into a different business case i.e. build more capacity or outsource more?
6. What are the reasons for these proposed changes?

STOCK ASSESSMENT CAPACITY

1. How many $ per annum do you spend on stock assessment (excluding monitoring)?

2. How many staff in FTE do you have as a) model developers and b) model users?
3. What is the biggest source of funding for this work?
4. How secure is this funding source?
5. How reliant is this capacity on specific skill sets?
6. How reliant is this capacity on specific databases?
7. How reliant is this capacity on specific software?

STOCK ASSESSMENT NEEDS

1. Given what you have read about our attached review, have you seen any issues?
2. What are your thoughts with regard to mostly bespoke models versus using freely available packages as a main?
3. Would you be interested in transitioning to using packages more?
4. What would be your main impediments to undertaking this potential change?
5. Can you name you top three needs in terms of stock assessment?

CHAPTER 4. INTERNATIONAL STOCK ASSESSMENT REVIEW PROCESSES

INTRODUCTION

An international review of stock assessment peer review was undertaken. The reason for this is that the above analyses showed that most of the Australian stock assessments were published as external reports that had gone through some form of internal review. Fewer than half of the stock assessment methods were published in journals. There was a general view that the input from groups such as Resource Assessment Groups (if they existed at all) and the review process of scientific journals was insufficient to fully capture the range of assessment work, and review a stock assessment. Resource Assessment Groups often lack the expertise to technically review a stock assessment (Penney et al. 2016). A small desktop study on different international stock assessment peer review processes was undertaken. This included information from a special session entitled "National and International Perspectives on Improving Science and Management through peer review – symposium" which was initiated as part of the American Fisheries Society Conference in Tampa (August 18-24, 2017). Presentations on the peer review systems were provided on the USA (the Center for Independent Experts (CIE); South-east Data, Assessment, and Review (SEDAR); Northeast Regional Stock Assessment Workshop (SAW) with a Stock Assessment Review Committee (SARC)), Canada, New Zealand, Australia, EU particularly ICES, tuna RFMOs, CCAMLR and MSC (see the program at bottom of the document). South Africa has also been added as several Australian stock assessment analysts have been involved in this system.

The review has found that three broad systems can be found internationally – a) an independent in-person panel of experts; b) an independent desktop review, and c) a Working Group (WG) (or Resource Assessment Group (RAG)) system.

a) The most independent form of review is also the most expensive. It uses an independent body to appoint independent reviewers that write either a panel report and/or independent review reports. Examples of this are the Center for Independent Experts in the USA and the Marine Stewardship Council assessment review process. Another form of in-person panel system that is implemented only in South Africa, which involves appointing a panel that meets each year and reviews assessment work undertaken through the year, but with a fair amount of consistency in panel members. Since the panel is essentially the same group of reviewers, ongoing work and parts of assessments can be reviewed as the work progresses. The former approach emphasises independence, but not continuity whereas the latter approach emphasises continuity.

b) The less expensive option is to undertake a desktop review using an independent expert(s). While still maintaining independence, the costs are much reduced. However, the in-person knowledge transfer is lost, which can be important for the reviewer to understand the fishery's background and the data available for assessment purposes. At present, this system tends to not be the most common approach.

c) Several forms of Working Group approaches can be found. The first is for the various WGs in the standard assessment and management process to review each other's work. This keeps the cost down, but allows criticism and cross connections. It does not emphasise independence, but uses available resources at the expense of increased WG workload. The second is an independent (not initially part of the assessment being reviewed) reviewer(s) who actively participates in the WG and therefore becomes another assessor available to the group. This forfeits independence, but brings in fresh views. The third is a WG system, which is highly critical (often highly adversarial), either because several nations are involved or the WG is open, including independent consultants paid by other users e.g. the fishery. The final form of WG is a mostly collaborative WG where all inputs tend to be included while the model is being built (wherever possible). The latter would be the AFMA RAG process, which is the least independent form of peer review process of all those described, where some would see the collaborative nature (which has merits) as not performing an independent peer review role.

In all cases, peer reviews through journals are sporadic, partially because stock assessments are often seen as not novel enough to be published. Furthermore, some assessors are very regular publishers, but many are not.

CONCLUSIONS AND RECOMMENDATIONS

Apart from urgent ad hoc reviews, undertaking regular independent panel- based reviews is probably too costly for Australia, especially under the current cost recovery model. However, if a panel model is considered, the process undertaken in South Africa where a fairly consistent group of individuals forms the panel and annually undertakes multiple nominated assessment reviews over a week, could be the best approach in terms of costs and benefits. Another option is for the panel to review multiple species or generic issues.

If a panel approach is not supported, then the following options could be considered:

1. For all jurisdictions: FRDC facilitates an in-depth desktop review process linked to key controversial and/or valuable species SAFS chapters, where the reviewers comment on key science behind the chapter, including the stock assessment (if any), and

2. Commonwealth: AFMA to set up a cross -RAG review system for the Commonwealth fisheries where each RAG reviews the other RAGs' assessments.

To facilitate more consistency between assessments for review by the public and experts, a more standard system of assessment documentation including Appendices of model settings would be beneficial.

REGIONAL SUMMARIES

USA: SUMMARY OF PRESENTATION “NATIONAL STANDARD ON SCIENTIFIC INFORMATION FOR MARINE FISHERIES MANAGEMENT IN THE UNITED STATES” (PRESENTER: BILL MICHAELS)

The USA has several Acts and policies that lead to a high degree of peer-review. These include the Information Quality Act 2001, the Magnuson-Stevens Fishery Conservation and Management Act 2007 (MSA) and the OMB Peer review bulletin 2004. These require that the science quality of both the methods and conclusions need to be peer-reviewed if any government action will have a substantial impact on people’s livelihoods. Minimum peer review requirements are defined. There is a further requirement to also continuously improve the science behind decision-making. The MSA establishes the Regional Fisheries Management Councils and a series of committees within each Council, including the Scientific and Statistical Committee (SSC). The entire process of reviewing science and providing scientific advice to the Council is publicly transparent (e.g. all minutes and reviews available online). The MSA has 10 National Standards (NS) of which NS2 relates to scientific information where peer review is a key aspect of good science. The role of the SSC is to evaluate the science being undertaken, the reviewers’ comments and to provide advice to the Councils. SSC members can also be an observer or the independent chair of the Panel-based review. Throughout there is an importance placed on transparency, so anyone (industry, eNGO, etc) is able to attend the review. Independence is clearly defined in the various policies, but mainly imply that reviewers cannot be NMFS employee for a CIE review, have any association or been a previous reviewer of the fishery for at least three years, nor a history of advocacy involvement with the fishery (these are applied to the reviewer and their family). Usually, therefore, the reviewers are international.

Assessments require a series of reviews depending on the degree of change – internal, external desktop, ad hoc panel and an external panel. Federal management of fisheries in the USA involves seven regions with some differences in the review process among regions, but all conform to the common standards defined in the MSA NS2. The benefits of this process are seen as ensuring scientific integrity, matching the form of review with the requirements, transparency, improving trust with stakeholders and is inclusive.

USA: “CENTER FOR INDEPENDENT EXPERTS – NOAA FISHERIES’ PEER REVIEW PROGRAM” (PRESENTER: STEPHEN BROWN)

Although the National Marine Fisheries Service (NMFS) manages the review process in all regions, it is independently co-ordinated by various regional bodies. In the case of the northwest, this body is the Centre for Independent Experts (CIE). The CIE process is about 20 years old and reviews stock assessments and additional special subjects such as ecosystem modelling. There are two levels of review, a minor update and a major update (benchmark) of an assessment review.

The former is usually undertaken as a desktop review whereas the latter are in-person panel-based reviews.

Desktop reviews involve up to three reviewers, but these are undertaken semi-independently at the scientists' home base. Only individual reports are produced. The total cost is about USD53k.

Panels usually consist of three independent reviewers with a total of 14 remunerated days. The panel is not allowed to produce consensus panel reports i.e. they have to represent differences in opinion. The reviewers also have to provide individual reports. The total cost of a single review is USD91k.

The process is that the relevant Fisheries Management Council (somewhat similar to a Management Advisory Council) from NMFS produces the Terms of Reference. Based on this, NMFS staff develop a Statement of Work (SoW) that provides more detail of what is required and includes the Terms of Reference. The CIE then uses the SoW to align the skills needed with their list of available scientists. A larger list than required is provided and the Council or Science Center provide comments on this list, but CIE approves the final list. In some cases, an SSC chair is also part or the chair of the panel. This CIE process prevents the Council choosing the CIE panel members. The panel (if required) and individual reports are provided to CIE who, after initial review for compliance, provides the reports to the Council. The panel selection is based on technical expertise, independence, conflict of Interest and availability. The CIE reviewers are from Europe (245), Canada (223), Oceania (165), South America and Caribbean (8) and the USA (144).

Since inception, the CIE has undertaken 320 reviews (786 review reports) on fishery assessments (most), methods (next most), protected species, status review, salmon, socio-economics and ecosystem models.

The CIE process itself has been reviewed and is regarded as a "gold standard". It can address a wide range of disciplines, is well established, meets the National standards and is well funded. The challenges, on the other hand, are that the process is expensive, time consuming, there is a tension between timeliness, throughput, transparency and thoroughness, it lacks a less intensive alternative, there is a lack of standards regarding stock assessment methods and their review, and there is a limited pool of stock assessment reviewers.

USA – "PEER REVIEW OF HIGH PROFILE BENCHMARK STOCK ASSESSMENTS IN THE NEW ENGLAND AND MID-ATLANTIC REGIONS: SAW/SARC PROCESS" (PRESENTER: JAMES WEINBERG)

This regional Council covers the fisheries in the mid-Atlantic and New England, and related State waters that fall within their jurisdiction. Species include cod, haddock, surf clam, red clam, squids, spiny dogfish and sea bass. A large influence on recent assessments is evidence of a changing environment, through sea surface temperature increasing over the past century, which has accelerated in the most recent decade. As a result, assessments now often include some broader ecosystem aspects.

Assessments are prepared by Stock Assessment Working groups (SAW) and reviewed by an independent panel of stock assessment experts called the Stock Assessment Review Committee (SARC). The peer review process is seen as allowing new ideas to be included in future assessments, giving public acceptance as well as providing high-level reviews. However, some question the need for international reviewers as this makes the process expensive and brings in scientists not familiar with the fishery/species, rather than using local expertise.

SAW WG members are appointed using a formal process. The SARC review panel consists of an independent chair drawn from the relevant SSC and 3 CIE members. Selection of the independent reviewers needs to be independent of scientists doing the assessments. The whole process is overseen by the Northeast Regional Coordinating Council (NRCC) that consists of directors and chairs of the Fisheries Management Councils and the relevant regional organisations. The NRCC decides when and which assessment is to have a benchmark review, whereas the review process is carried out by the Northeast Fisheries Science Center. In this case, there is an attempt by the WG to arrive at a consensus, although this is not required. For transparency, the review results are provided to the Council in a public forum. The challenges are that there are many assessments (driven by three FMCs) so it is often difficult to choose between in-depth or update reviews resulting in competition for scarce staff resources. A new planning process has been undertaken that has allowed for a two-track assessment system, longer planning horizons and more flexibility.

USA: "THE IMPORTANCE OF PEER REVIEW FOR SEDAR STOCK ASSESSMENTS" (PRESENTER: JULIA BYRD)

This region has several Councils (aka. co-operators). The Southeast Data, Assessment, and Review (SEDAR) steering committee runs the review process and includes the co-operators. They also define the policy, procedures and review schedules. The Councils, therefore, have to follow these procedures. Transparency in this process is seen as key. SEDAR started in early 2000 due to fisher demand amid perceptions that the science was hidden in a "back room". All documents are provided on the internet. It is important that there is a consistent formal review process that lends validity to the science, and provides recommendations for technical and process improvements. Most of the reviews focus on specific stock assessments, but periodically they address special topics such as stock structure.

There are three types of reviews: 1) Benchmark – in-person independent review, 2) standard – review through the SCC, and 3) update undertaken by the SSC. There is a data workshop, assessment workshop and review workshop where the review itself takes place. Initially review panels were very diverse with a CIE chair, but not everyone had the requisite expertise and the reviews often strayed into management considerations. The chair also didn't have the necessary technical expertise or knowledge of the fishery. The system moved to a three-person CIE panel with an external NMFS chair. Managers and fishers were observers. Due to lack of staff, the process moved again to an SSC chair, CIE and SSC person on the panel with managers and fishers as observers. The SEDAR steering committee also tried to incorporate the CIE expertise earlier in the process i.e. one reviewer attended all three meetings. This did not work well as it was more useful if the reviewer became more directly involved with the work, and later CIE panellists reviewed earlier reviewers' work. Over time, the documentation was increased and the dissemination process was improved. The steering committee also improved region- and fishery-specific information to provide the panellists with more background. Since traditional stock assessment processes often don't work in the Caribbean, early assessments were often rejected. The assessors started to explore data-poor assessment methods and undertook a Caribbean fisheries-specific data evaluation workshop, which has meant more recent approaches are being approved.

Challenges are that it adds time and costs to the process, the review becomes an anchoring point in the assessment schedule, fear of failing a review can inhibit innovation, difficulty incorporating the CIE earlier in the process and that the review drives the tone of the report and documentation. Some believe that it does not help end users. In future, the process will remain and evolve as in the past, including finding ways to increase involvement of outside high-level expertise.

CANADA – "PEER REVIEW AND SCIENCE ADVICE IN CANADA" (PRESENTER: JAKE RICE)

The original process was based on the ICES experience, which was to include an independent scientist in the assessment development itself rather as an independent reviewer as per the CIE process. International working groups formalised the ICES rules and these were influential in the early Canadian process until 1992. For this period, there were regional sub-committees of Canadian Atlantic Fisheries Scientific Advisory Committee that shared the role of reviewing ("challenge" format) the other regions' assessments i.e. one region will prepare the work and other regions would then criticise it. Each region was expected to review every other regions' work. All chairs of the sub-committees and assessment lead authors would then report the findings and recommendations in advisory format. These committees were closed in 1992 after the cod fishery closure.

In 1997, the Canadian stock assessment secretariat was formed, which shared many features of the previous process. New features were a higher degree of inclusiveness by including academics, industry and NGOs. There was a more proactive initiation of advisory actions and more discussion of applicable Terms of References with stakeholder groups. There was a higher degree of communication through Press conferences or Notices to the press at the end of key meetings, publication of advice within 15 days of the meetings and a complete proceedings document. In 1989 and 1999 the Oceans and Species at Risk Act was developed. These initiated more zonal and national meetings, often because more travel money was available. This means that there was a greater need for framework advisory meetings to standardise on, for example, what is a consistent interpretation of precaution and allowable harm. The integrated advice was facilitated by a One Window Canadian Science Advisory Secretariat.

In the mid-2000s, there was a significant cut to government budgets that affected travel to meetings. This resulted in fewer, smaller and shorter meetings and less diversity of view in the peer review process. The advice was drafted to consensus at the meetings.

Presently, there has been a reversal of budget constraints, but no real changes in practice, although some would say scientists are less restricted in what they could say.

EU – "THE NORTHEAST ATLANTIC EXPERIENCE WITH PEER REVIEW OF FISH STOCK ASSESSMENTS" (PRESENTER: HENRIK SPERHOLT)

Fisheries related to the International Council for the Exploration of the Sea (ICES) catch about 10 million tons of fish each year. The Northeast Atlantic is now a sustainable fishery. ICES aims to produce advice based on best available science by having good quality assurance processes, and being transparent, unbiased independent and relevant to management. Benchmark reviews are conducted about every five years through an expert group. In between the assessment is run as an update. In the benchmark, methods and a data series are developed. The workshop process is that the peer review occurs in conjunction with the benchmark workshop. This integration with the benchmark assessment development means that the input can be integrated right away and one can avoid an extra link. Three external reviewers are invited. 'Independent' in this context can be a scientist from another ICES area to external from both ICES and Europe. The reviewers are ordinary participants in the workshop. It is important to note that ICES stocks are often multi-national. The assessment reports contain a two- to four-page summary that includes statements on whether the reviewers agree with the working group reports or not. Benchmark assessments look at data quality, methods including reference points, forecasts, environmental drivers and species interactions. The strengths of the ICES process are that there is better linkage to state-of-the-art science, better exchange knowledge across the ecosystem and cost as reviewers are not paid. Weaknesses are that the reviewer is no longer an independent part of the review process, it is difficult to attract the best scientists as (unpaid) reviewers and for reviewers to grasp the complete background of the biology, the fishery and the data. Reviewers are thus inclined to focus more on methods and statistics. Integration of environmental drivers and multi-species interactions are almost lacking. Fmsy and other reference points get (strangely enough) a low priority in meetings. There is considerable momentum around assessment and forecast, but the reviews run out of time for the remainder of the factors. 186 benchmark assessments were undertaken between 2007-2016. Only half of the assessments have been through this process.

AUSTRALIA – "IMPROVING FISHERIES SCIENCE AND MANAGEMENT VIA PEER REVIEW: AN AUSTRALIAN PERSPECTIVE" (PRESENTER: CATHY DICHMONT)

Australia has a Federal system that consists of States, Territories and the Commonwealth. In general, fisheries within 12 nm of the coast fall within the relevant State/Territory jurisdiction, whereas offshore, cross-jurisdictional or international (e.g. CCSBT, WCPC, IOTC) fisheries fall within the Commonwealth jurisdiction. The peer review process of assessments is very variable across jurisdictions. It is important to note that Australia (especially at the Commonwealth level) has adopted Harvest Strategies (not always requiring a data-rich assessment) that are Management Strategy Evaluation (MSE) tested. All research agencies have some form of an internal review system of reports before release. The strength of this assessment review depends on agency depth in terms of assessment skills and the degree to which the review process is monitored. These reviews generally check whether the report is technically correct and well written. Stock assessment scientists are also expected to publish their work in peer-reviewed journals to expose the technical aspects of their work to peer review. However, publication volume is very assessor- and organisation-dependent and more than half of the assessments are not published in this format. Assessment reports are mostly publicly available on the internet. All Commonwealth Harvest Strategies have been MSE tested and most of these tests have been published in the peer reviewed literature. The Commonwealth system relies on two committees that provide the review process of stock assessment in a given fishery. The first is an appointed technical group of experts, the Resource Assessment Group. These contain assessment scientists including the main assessor and an external assessment scientist, biologists, economists, managers and industry. The second group, the Management Advisory Committee (MAC) consist of eight members and an independent chair. These are appointed positions. Both committees are chaired by independent appointees that have no direct conflict of interest with the fishery or the research providers. All assessments are monitored by the RAG, who have strong input to the data, model settings and model type. Both committees have strong control on conflict of interest. This is an extremely effective system if the committee has the required resources. Assessment reports are published online, but there is often a delay when this happens and previous versions of assessments are hard to find. The RAG review process intends to verify findings, improve acceptance of an assessment, is more a product of the RAG and assessors. The RAG also plays an important role in setting research directions.

Some State jurisdictions have a similar system, although the RAG-equivalent committees are usually not independent of the MAC as is the case in the Commonwealth i.e. they are sub-committees of their "MAC". However, many jurisdictions stopped using this process and recently, are trying to re-institute variants of this system. The assessment report format varies by RAG and jurisdiction. The assessment code is usually not freely available or peer reviewed. There is a large variability in peer review rigour between assessments and jurisdictions. Only a few *ad hoc* in-depth reviews have been undertaken, usually in response to a controversy. Generally, there does not seem to be an industry demand for reviews, given that cost recovery would mean they would have to partly fund the reviews.

The Status of Australian Fisheries Stocks (SAFS) is published every two years. This is a fairly new system so new species are being added, based on value, volume and sensitivity. An assessment *per se* is not being reviewed, but the stock status of the assessment (if one has been used) is reviewed. Each species chapter is reviewed by three external, independent reviewers. Chapter authors can reject suggestions without any return to the reviewers. The process is managed by the major funding agency, FRDC. The main reason for SAFS is to enable community and retail acceptance and information transfer.

Challenges with the Australian system are that there is no unified peer review system, the RAGs are not independent as they become involved in producing the product, few assessments are published in peer reviewed journals, and code and details are not checked. Furthermore, the RAGs have less and less time to undertake in-depth reviews of each assessment – they rely in part on the tests of the model undertaken and checking the results, on the research providers' internal review process and on assessors' reputation.

In the future, there will be increasing use of RAG-like bodies in those jurisdictions presently without them – at least for valuable or controversial fisheries. Adoption of harvest strategies are on the increase, and some of these are being MSE tested. Any more rigorous and independent review process would need resourcing, a source of funds for which presently remains unclear.

NEW ZEALAND – "SCIENCE PEER REVIEW: WHAT, WHEN, WHERE, AND HOW SHOULD WE REVIEW FOR COST-EFFECTIVENESS?" (PRESENTER: MARTIN CRYER)

All fisheries are managed by the Ministry for Primary Industries (MPI) mostly through ITQs. MPI contracts most of its science to NIWA. 65% of the costs are recovered from industry. The payment for research by industry has an impact on them and has had unintended consequences, including purchasing opposing research. There is a published standard for science (Mace and Penney, 2010) that defines key principles and responsibilities based on peer review, integrity, objectivity, and reliability/repeatability (PIER). The standard formalised long-standing practice, but ensured what was expected in terms of quality. These responsibilities lie on buyers and doers of the research. Standing Working Groups are the major entity for peer review. These are mostly fishery-specific, but others include recreational fisheries, mammals, RFMOs etc. Staged technical guidance is available, especially for costlier, novel, complex or contentious research i.e. there is a process of working through various stages of developing an assessment. The Working Groups are chaired by a scientist and run by the MPI. There is a formal Conflict of Interest process. Meetings are open to anyone that agrees to the Terms of Reference. For some meetings up to 40 scientists, managers and advocates – sometimes lawyers – attend. Decision making is ideally by consensus. For transparency, all the documents are published online. There is also a fisheries assessment plenary meeting that summarises approaches and findings across all the working groups. Effects of fishing etc. is a compendium document on the impacts of fisheries. Standing working groups are difficult to chair because attendees will find weaknesses in assessments. It is a highly confronting experience for some scientists. The approach risks losing its independence and objectivity through gaming by some participants. It is expensive as MPI convenes about 100 full day meetings each year. Each stage in routine assessments can take up to eight meetings. If the assessment is contentious then the process is similar to the SAW process in the USA where an independent panel is established. This increases the time and cost. The fully independent review will have three or four invited experts, there are formal Terms of References and a meeting of 4-5 days. Stakeholders are allowed only to the presentations. The report is written by the reviewers.

For small reviews an MPI staff member can be tasked check the work (not often used). End of the process have grades – 1 High quality (most assessments are ranked 1); 2 mixed (use with caution), 3 low quality so do not use or ungraded (not peer-reviewed).

NZ intends to continue using this standard, but fisheries management is a much more diverse concept than single species stock assessments. Research is now more diverse and new Working Groups may be needed and new expertise may be included. There is also increasing workload on WGs to review work commissioned by others.

TUNA RFMOS – "PEER REVIEW IN REGIONAL FISHERIES MANAGEMENT ORGANIZATIONS, THE EXAMPLE OF ICCAT" (PRESENTER: DAVID DIE)

There are five tuna Regional Fisheries Management Organisations (RFMOs) in combination managing fisheries in all oceans of about 5 million tons (2013). About 76% of the fished species are deemed healthy, 14% in need of strong(er) management and 10% are intermediate. The ones that are healthier tend to be those species of lower value. About 130 countries are signatories in at least one of these RFMOs. Apart from SBT, votes tend to be dominated by developing countries. The scientific processes in each RFMO are slightly different. Three have Scientific Committees that fall under a Commission. The International Commission for the Conservation of Atlantic Tunas (ICCAT), Indian Ocean Tuna Commission (IOTC) and Commission for the Conservation of Southern Bluefin Tuna (CCSBT) have Secretariats that perform the data management provided by each jurisdiction, research and meeting logistics, and the Commissions are provided assessment advice through the Scientific Committees. This differs from IATTC that undertakes the science internally at the Secretariat by staff from the Secretariat. For WCPC, the scientific assessment work and research are mostly contracted to the South Pacific Commission. These structural differences have led to differences in the review process.

Reviews of the organisational processes have occurred for all RFMOs at least once. IATTC undertakes independent reviews through a contract to Moss Adams Consultants. However, the identity of the reviewers are not known. These reviews have tended to be organisational reviews and investigate all aspects from fish stock health, human resources planning, management, operating structure, financial planning, efficiency and effectiveness. A three person panel is formed to conduct ICCAT reviews, consisting of an independent chair, scientist/RFMO member, scientist, fishery manager, or a law expert. IOTC committees consist of a chair with legal fisheries backgrounds, a science expert, six IOTC representatives, two NGO representatives (e.g.

International Seafood Sustainability Foundation and Pew), and two members of other RFMOs. CCSBT forms an internal working group of the CCSBT executive secretary and five CCSBT member representatives. Performance is reviewed internally to produce a panel report and an independent expert checks the report.

The internal peer review of the science in the ICCAT, IOTC and CCSBT have fewer national scientists involved, more at the Working Groups and a large number of scientists involved in the Scientific Committee. Usually, small groups have undertaken the research and everyone else attempts to criticise it. The Scientific Committee reviews all scientific products. Scientific capacity is highest at the Working Group level, and less at the plenary and national level. For the IATTC and the WCPFC, only a small group of people review the assessment either at the secretariat or within SPC. Scientific products go to the Scientific Committee of the full Commission, where there are many scientists that are available to review the science, but not all have the requisite expertise.

ICCAT evolved its process since 2000, when its first review of white marlin occurred in 2002 and in 2013 it reviewed bluefin tuna. There is a 2015-2020 science strategic plan in which it plans to undertake two reviews each year. The selection of reviewers is done with the ICCAT secretariat and the working group chairs. The Terms of References are developed by the same group. Reviewers are remunerated. Challenges are that the selection of candidates is internal, there are not enough available expert reviewers, there is a lack of clear processes to follow up on recommendations of reviewers, and the reviews have been limited to assessments and not included other scientific products. One possible solution is to undertake the selection process through other tuna RFMOs.

ANTARCTIC – "PEER-REVIEW OF FISHERIES MANAGEMENT IN THE COMMISSION FOR THE CONSERVATION OF MARINE LIVING RESOURCES (CCAMLR)" (PRESENTER: ALISTAIR DUNN)

The Convention on the Conservation of Antarctic Marine Living Resources (CCAMLR) was established in 1982 with 25-member states (11 countries acceded). It regulates harvesting of Antarctic marine species. The CCAMLR Convention (treaty) enshrines EBFM through sustainable harvesting, but also taking into account the effects of fishing and ecosystem effects. The Scientific Committee advises on the science and comments on regulations. Membership is well defined. The Secretariat is located in Hobart. The Scientific Committee meets each year and agrees on the science advice by consensus. The science is undertaken by technical Working Groups (WGs), where the assessments process mostly involves the WG Statistics, Assessment and Model and the WG Fish Stock Assessment. The scientific advice is scrutinised by experts within the WGs, where most of the review process is undertaken. The final advice has to be reviewed by the WG and the Scientific Committee. Documents are published online.

Stock assessments are undertaken either annually or biennially. The advice is developed the Assessment Methods WG. The secretariat may provide papers that are then reviewed by member scientists. There are about 50 scientists in WGs. Submitted WG papers, WG and SC reports are publically available. Most work is published in the CCAMLR and other peer-reviewed journals. The advice from technical WGs is based on consensus, although if members do not agree, this gets recorded. There is a high standard of science, and this science is inclusive and transparent to reach consensus. The science is clear and available to all members. There is an informal adoption of the NZ PIER principle.

Fisheries can be complex and the models can also be very complex requiring high levels of skill. CCAMLR also has a large number of fisheries to assess, so meeting schedules are full.

The external review process is mainly through publication in the peer-reviewed literature. External observers also comment on the scientific advice at the SC. Some of the CCAMLR fisheries are also MSC-certified (MSC comments are provided into the CCAMLR process). Occasionally, experts are invited to WGs.

There have been recent discussions concerning whether some of the reviews should adopt something similar to the CIE approach.

MARINE STEWARDSHIP COUNCIL – "THE ROLE OF PEER REVIEW IN PROVIDING ASSURANCE IN MSC CERTIFICATION" (PRESENTER: DAVID AGNEW)

The Marine Stewardship Council (MSC) has set standards related to fishery certification. So far, there are 296 certified fisheries, 17 suspended and 67 in assessment. Once certified, a fishery has to go through another full assessment every five years. Each fishery has to go through an independent review process. About 10% of fisheries fail. There are two certification systems – one that certifies the fishery and the other the chain of custody. There is an enormous burden on MSC to retain public trust in the system. The certification system is based on the FAO guideline on eco-labelling. MSC certification consists of three principles – target, ecosystem and management – and therefore one reviewer per principle. A fishery needs a score of 80 to pass, with benchmarks at 60, 80 and 100. The guidelines are a very long document. The review process is run through a third party certification program. There is a Board of Directors, Stakeholder Council and a Technical Advisory Board. MSC developed the standard and certification standard. Conformity Advisory Bodies (CAB) and third-party assessors apply the standard. There is an independent peer review of the assessment report and also, subsequently, a public review of the assessment report. An independent adjudicator oversees any objections and appeals. Accreditation Services International checks the performance of the CAB. This is a long certification process with stakeholders engaged throughout.

Prior to 2014, the CAB was the main body that managed the process, including selecting the peer reviewers. Peer reviewers are nominated by the certification bodies. MSC's role was to review the quality of the reviews and the uptake of the peer reviewers' comments. A review of the process considered other processes including the CIE and journal peer review system. Problems identified include that many stakeholders expected a fishery to fail and then it passed; there is an uneven quality of peer reviewers, training of peer reviewers and contract conditions; and direct contracting of peer reviewers by CABs is seen as a conflict of interest. There were eleven certification bodies all competing with each other, so quality was suffering.

The solutions to strengthening quality control and removing conflict of interest issues are to: 1) use a single contractor for all peer reviews of assessment reports; 2) create an independent oversight committee drawn from MSC's Stakeholder Advisory Council and the Technical Advisory Board; 3) implement consistent training requirements for all peer reviewers; 4) conduct third-party scrutiny of qualifications and potential conflict of interest of peer reviewers; 5) implement specific timelines for undertaking a peer review to make assessment predictable; 6) implement a consistent rate of pay for peer reviewers and contract duration (2 days); 7) employ two peer reviewers per assessment (can be increased if needed at the discretion and cost of MSC); 8) provide quality assurance by MSC and Peer Review College's third party scientists; and 9) give the peer reviewers an opportunity for comment on the final report if the CAB has not adopted the original peer review comments.

The new system has a Peer Review College which reports to MSC and a Conformity Advisory Body (contracted by MSC). The Conformity Body receives a Client Action Plan from the fishery clients, which then send the draft report for peer review to the Peer Review College, which in turn sends the draft report for peer review to the peer reviewers (contracted by MSC). Third party scientists contracted by MSC approve the selection of the peer reviewers' shortlist. Once the peer reviewed report has been received by the College this is sent back to the Conformity Advisory Body (contracted by MSC). An ASI (contracted by MSC) is only required if a potential conflict of interest is highlighted.

Twenty-three fisheries were in this pilot revised program, the previous rate of 49% of comments taken up by CAB has improved to 65%; and five peer reviews achieved >80% CAB acceptance of their comments. MSC does the training and MSC contracts, but the CAB makes verification decisions. Given the success of the trial, the new system will be used for all assessments from now on. This new system is very cost efficient and able to draw on MSC resources for additional peer review if listed. Presently 45 peer reviewers are listed and are now being taught in universities.

SOUTH AFRICA (NOT PRESENTED BUT ADDED FOR INTEREST)

In South Africa, fisheries are managed (including data collection and monitoring) by the Fisheries Management Branch of the Department of Agriculture, Forestry and Fisheries (DAFF). On the other hand, most of the assessments are contracted to the Marine Resource and Assessment and Management Group in the University of Cape Town. Several species-specific WGs exist that contribute to or criticise the assessments. Each of these has a series of meetings during the year, ultimately providing advice on the Total Allowable Catch. Assessments are formally reviewed in the WG, but capacity varies. Occasionally, the WGs can be adversarial if, for example, an industry-appointed consultant tables a separate approach to a key assumption to the assessment. Each WG member and chair is appointed, but the majority of these come from DAFF. No more than about three of the 8-10 members are external. Observers are allowed who come from various stakeholder groups, such as industry and NGOs, who often contribute to discussions. Consensus is preferred.

South Africa has also used a unique approach of appointing a reasonably set panel of independent experts that meet each year to review aspects of assessments and ecosystem models. Rather than review a single assessment each year, they review components of multiple models that have been updated during the year. The international panel is used to deal with both important fisheries, and cases which are topical and/or controversial. Although panel members are usually the same from one year to the next (beyond retirements or the panel member being unavailable), with new members added if a topic is being addressed that is not covered by existing panel members. One of the panel members is the rapporteur. Documents are available by request and observers, including industry-appointed scientists, can ask to attend. Panel members are not remunerated.

The set of species whose assessments are to be reviewed are selected by Fisheries Management Branch of the DAFF and the Marine Resource and Assessment and Management Group, and a specific set of questions is identified for the Panel (although the Panel can comment on issues in addition to the identified questions). Reviews involve presentations of the assessments in a workshop format followed by requests for clarification and additional analyses. Extensive documentation is provided, which is made available via a web-site (http://www.maram.uct.ac.za/maram/workshops), except for copyrighted material such as published papers. The Panel develops its report, which is primarily a set of conclusions and recommendations (related to questions identified) by consensus. The recommendations are ranked in order of importance by the Panel and subsequent annual Panels review progress related to the recommendations. The Panel report is presented to the workshop, with requests for correction of errors of fact (but not the conclusions of the Panel). The Panel report is then presented in a less technical manner in a public forum.

CONFERENCE SPECIAL SESSION PROGRAM

1. Brown, S.K., Shivlani, M., Chandler, M. Center for Independent Experts – NOAA Fisheries' Peer Review Program

2. Michaels, W. National Standard on Scientific Information for Marine Fisheries Management in the United States

3. Weinberg, J. Peer Review of High Profile Benchmark Stock Assessments in the New England and Mid-Atlantic Regions: SAW/SARC Process

4. Byrd, J., Carmichael, J., Neer, J. The Importance of Peer Review for Sedar Stock Assessments

5. Rice, J. Peer Review and Science Advice in Canada

6. Sparholt, H. The Northeast Atlantic Experience with Peer Review of Fish Stock Assessments

7. Dichmont, C.M., Punt, A.E., Deng, A., Little, R. Improving Fisheries Science and Management Via Peer Review: An Australian Perspective

8. Cryer, M. Science Peer Review: What, When, Where, and How Should We Review for Cost-Effectiveness?

9. Die, D. Peer Review in Regional Fisheries Management Organizations, the Example of Iccat

10. Reid, K., Dunn, A. Peer-Review of Fisheries Management in the Commission for the Conservation of Marine Living Resources (CCAMLR)

11. Agnew, D., Hoggarth, D., Bruford, G., Good, S. The Role of Peer Review in Providing Assurance in MSC Certification

REFERENCES

1. Mace, P.M. and A.J. Penney. 2010. Overview of New Zealand’s fisheries science peer review processes. 24 pp. Wellington, New Zealand.

IMPLICATIONS

There are continual cost pressures on stock assessments. Assessors are feeling pressure to undertake complex work for a smaller budget, yet they find that there is not always good understanding of how much is required by managers and industry. Assessors also often feel isolated – Australia is a large country with assessors spread through eight jurisdictions with, it seems, little time or resourcing to share and communicate meaningfully.

Australia is seen as being in the forefront of the assessment world in terms of quality of assessors, yet there is a perception by assessors outside of Australia that these have not always contributed to providing tools (beyond described in journal papers) for others to use. Some of this lack of contribution to the international community is the culture of the assessors themselves as well as the culture of the Australian organisations in which they reside – there are few resources or even the incentive to move in that direction. In the USA, a very high level of scrutiny by review panels and other processes have resulted in much higher value being placed on developing in-house peer reviewed open source packages.

Many of the assessments within Australia could have been undertaken with these packages, but past investment and a perception that the models do not produce what is required by the jurisdiction does create some resistance to their use.

However, there are positive recommendations listed below that were gleaned from the project team's review and interviews by jurisdictions' stock assessment research providers. These provide invaluable ways forward to address the isolation and increase sharing within Australia and globally.

RECOMMENDATIONS

1. Several off the shelf stock assessment packages are available, especially those from the USA that correspond reasonably well to Australian policies. Assessors should be aware of software available and interact freely with the authors. Often these would be able to modify code to jurisdictions needs, where possible.

2. Joining the shareware community would be beneficial – Australia is seen as stock assessment capacity rich, but also has contributed little to the software package resources available. Do we need a culture shift within the Australian assessment community to be more sharing? Or is this about lack of time and resources? It would be worth investigating the impediments for moving this forward although many are mentioned in Chapter 3 and below.

3. The 76 surveyed model-based assessments undertaken in Australia reflect 37% of the 2013 catch recorded in Australia's Status for Key Australian Fish Stocks Reports (or 34% of the total catch in 2013). All but 18 (or 24 if full rather than approximate age-size-structured models need to be used) of the stock assessments could have been conducted using stock assessment packages used in the United States and New Zealand. Adoption and use of packages for more stocks in Australia should increase the likelihood that results are based on correctly-coded models whose estimation performance is widely understood, reduce the time needed to conduct assessments, and speed up the peer-review process. The availability of training, manuals, and example data sets for stock assessment packages should partially address their additional complexity. Additional benefits, in terms of numbers of assessed stocks, could occur if Australian stock assessment scientists develop a forum to collaborate and share methods. These results are applicable to many other jurisdictions that undertake stock assessments. Of course, this statement needs to be tempered against the counter view expressed by some that the transition would be more costly to more of a black box model, given the past investment in a bespoke model. Despite these valid opposite views, assessors should use whatever tools are available.

4. Most jurisdictions have their own technologies that would be valuable to others. Sharing of these is difficult in the present environment. There is a need for a national clearinghouse for stock assessment code and other streamlining tools to foster sharing and collaboration among research providers.

5. A national committee through the AFMF or RPN should be established with the goal to develop collaborations and share knowledge in stock assessment capability.

6. Most jurisdictions were open to using bespoke or packages depending on which was best for the task. The gradient of being predominantly against the use of packages to mostly using packages tended to be inversely proportional to the jurisdiction's own level of stock

assessment expertise and investment in past assessments. One common element was the need for a freshly developed bio-economic length-based model for species that are difficult to age. It is therefore recommended that a workshop be run bringing all the length-based assessors together, including from overseas, to determine the need for such a model and the steps required to develop and share such a model.

7. There was a general view that the input from groups such as Resource Assessment Groups (if they existed at all) and the review process of scientific journals was not enough to fully capture the range of assessment work, and review a stock assessment. Resource Assessment Groups often lack the expertise to technically review a stock assessment (Penney et al. 2016). Although several options were discussed, it was accepted that the USA approach using the Centre for Independent Experts (https://www.st.nmfs.noaa.gov/science-quality-assurance/cie-peer-reviews/index) was too expensive, but a "lite" version of that may work, as well as approaches such as that undertaken in South Africa where a consistent review panel meet every year to discuss various assessments as part of a medium-term review program. Another option is for the panel to review multiple species or generic issues. If a panel approach is not supported, then the following options could be considered: a) For all jurisdictions: FRDC facilitates an in-depth desktop review process linked to key controversial and/or valuable species SAFS chapters, where the reviewers comment on key science behind the chapter, including the stock assessment (if any), and b) Commonwealth (because it already have RAGs): AFMA to set up a cross-RAG review system for the Commonwealth fisheries where each RAG reviews the other RAGs' assessments.

8. To allow greater sharing and collaboration of stock assessment scientists, it recommended that a biennial or annual forum for assessors to share ideas and technologies should be developed. This may be a mixture of less regular in-person meetings combined with other types of format. The priorities are:
    a. A workshop that shares techniques to streamline data collection, collation and reporting,
    b. Data limited assessment techniques and available tools, and
    c. A course on CASAL or Stock Synthesis.

9. To address the need for more understanding of stock assessment from non-assessment staff, it is recommended that a stock assessment training course for managers and industry be run. This should link to existing courses and resources that have been developed.

10. The most common statement made in terms of priorities was the need for more and better quality data. This was not a request for more and better data to develop more data rich assessments, but rather a need for:

a. More or even some data for data limited fisheries so that a data limited assessment could be undertaken,

b. More (frequent and species) recreational data which is an impediment for assessing even quite valuable fisheries,

c. Better quality control processes of existing data being collected. Enormous time and resources go into cleaning and checking data.

11. In common were concerns about the average age of stock assessment scientists in Australia. Attempts to link with universities tended to not be seen as successful due to blockages of good student supply all the way to absence of flexible employment systems to employ good graduates when they do occur. A national discussion on succession planning is required, such as the past program of supporting the employment of stock assessment experts, a review of existing programs and how these can be improved and added investment in stock assessment development of young scientists.

12. Jurisdictions, especially those that outsource their assessment work, stated that there was a demand by managers, scientists and industry to understand whether an assessment has followed best practice. Although Penney et al. (2016) provide a starting point by providing a standard guideline for research provision, it does not prescribe approaches to ensure or verify best practice e.g. sensitivity tests, retrospective analyses. A guideline on stock assessment best practice for consistency between assessments should be developed.

13. Despite their standardised approaches, many packaged stock assessments still remain expensive options for assessing the status of a stock, especially in terms of the human capital required to operate them. Data-limited packages are increasingly being used but, in general, lack statistical robustness, while statistical packages remain applicable only to stocks that can justify the costs (the risk-cost-catch trade-off: see Dowling et al., 2013; Dichmont et. al., 2016). There is increasing demand however to reduce the costs of statistical stock assessments, possibly through automation. Automated processes, in data acquisition through to model balancing are needed, and could reduce management costs. A fisheries research priority should be written to include data and stock assessment streamlining, and cost reduction. The risk of automating the stock assessment process to a greater degree than currently experienced is that an automated process would lose the human eye, or capability of checking the output. A balance between automation and fully manual would need to be addressed on a case by case basis, but tools on how to do this should be shared (as recommended above).

FURTHER DEVELOPMENT

The recommendations highlight several next steps. Most of these addresses the expressed need for greater cross-jurisdictional sharing. This work did not cover the huge literature and growing number on data limited in light of a large body of work globally on this topic. However, some of our findings may in some years be largely similar to the data rich ones if the recommendations in this report does not be progressed, at least in part.

EXTENSION AND ADOPTION

This work has had huge input from:

a. the USA community of stock assessment scientists;
b. responses from stock assessment scientists on their assessments to the survey of Chapter 3;
c. interviews by stock assessment research providers on their research needs, capacity and priorities;
d. comments by these interviewees on the Australian review of Chapter 2;
e. comments by interviewees on their summaries and Chapter 3; and
f. presentations and discussion as part of the American Fisheries Conference attendees and experts in peer review processes in Chapter 4.

The comprehensive discussions, critiques and inputs have been greatly appreciated.

PROJECT COVERAGE

1. Dichmont, C.M., Deng, R.A., Punt, A.E. 2016. How many of Australia's stock assessments can be conducted using stock assessment packages? **Marine Policy**. 74: 279–287. https://doi.org/10.1016/j.marpol.2016.09.033

2. Dichmont, C.M., Deng, R.A., Punt, A.E., Brodziak, J., Chang, Y.-J., Cope, J.M., Ianelli, J.M., Legault, C.M., Methot Jr, R.D., Porch, C.E., Prager, M.H., Shertzer, K. 2016. A review of Stock Assessment Packages in the United States. Fisheries Research. 183: 447–460. https://doi.org/10.1016/j.fishres.2016.07.001

3. Presentation at the 147th Annual Meeting of the American Fisheries Society in Tampa, FL from August 20-24, 2017 entitled Improving "Fisheries Science and Management Via Peer Review: An Australian Perspective".