FISH and FISHERIES

Management strategy evaluation: best practices

André E Punt[1,2], *Doug S Butterworth*[3], *Carryn L de Moor*[3], *José A A De Oliveira*[4] *& Malcolm Haddon*[2]

[1]School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, 98195, USA; [2]CSIRO Oceans and Atmosphere, GPO Box 1538, Hobart, TAS, 7001, Australia; [3]Marine Resource Assessment and Management Group (MARAM), Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, 7701, South Africa; [4]CEFAS Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk, NR33 0HT, UK

Abstract

Management strategy evaluation (MSE) involves using simulation to compare the relative effectiveness for achieving management objectives of different combinations of data collection schemes, methods of analysis and subsequent processes leading to management actions. MSE can be used to identify a 'best' management strategy among a set of candidate strategies, or to determine how well an existing strategy performs. The ability of MSE to facilitate fisheries management achieving its aims depends on how well uncertainty is represented, and how effectively the results of simulations are summarized and presented to the decision-makers. Key challenges for effective use of MSE therefore include characterizing objectives and uncertainty, assigning plausibility ranks to the trials considered, and working with decision-makers to interpret and implement the results of the MSE. This paper explores how MSEs are conducted and characterizes current 'best practice' guidelines, while also indicating whether and how these best practices were applied to two case-studies: the Bering–Chukchi–Beaufort Seas bowhead whales (*Balaena mysticetus*; Balaenidae) and the northern subpopulation of Pacific sardine (*Sardinops sagax caerulea;* Clupeidae).

Correspondence:
André E Punt, School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195, USA
Tel.: +1 (206) 221-6319
Fax: +1 (206) 685-7471
E-mail: aepunt@uw.edu

Received 26 Jun 2014
Accepted 22 Oct 2014

Keywords Fisheries management, management procedure, management strategy evaluation, simulation, stakeholders, uncertainty

 DOI: 10.1111/faf.12104

Introduction

Management strategies (also referred to as management procedures; Butterworth 2007, 2008a,b) are combinations of data collection schemes, the specific analyses applied to those data and the harvest control rules used to determine management actions based on the results of those analyses. Management strategy evaluation (MSE),[1] the evaluation of management strategies using simulation, is widely considered to be the most appropriate way to evaluate the trade-offs achieved by alternative management strategies and to assess the consequences of uncertainty for achieving management goals. Butterworth *et al.* (2010a) list three primary uses for MSE:

1. development of the management strategy for a particular fishery;
2. evaluation of generic management strategies; and
3. identification of management strategies that will not work and should therefore be eliminated from further consideration.

One specific use for MSE, particularly in the United States where the forms of the harvest control rules for federal fisheries management are constrained by the Magnusson–Stevens Act (MSA), is to quantify the impacts of uncertainty associated with management strategies adopted at present, and to identify the 'realizable' performance which can be achieved given the quality of the data available and the types of uncertainties which are inherent in the system being managed.

Butterworth (2007) contrasts MSE with the traditional approach to providing management advice, which involves conducting a 'best assessment' of the resource, evaluating uncertainty using confidence intervals and sensitivity tests, and providing a recommendation for a management action based on applying some harvest control rule or by conducting constant catch or constant fishing mortality projections. That paper explains how MSE overcomes many of the concerns with the traditional approach, including that the full range of uncertainty can be taken into account and that decision-makers consider longer term trade-offs among the management objectives, instead of focusing on short-term considerations only.

For the purposes of this paper, a MSE must address the fact that the data and models on which management strategies are based are subject to uncertainty. Consequently, analyses in which fishing mortality can be set and implemented exactly (e.g. Punt and Butterworth 1991) are not considered to be MSEs, even though such analyses may be useful in terms of understanding the dynamical properties of exploited ecosystems.

Management strategy evaluation has been used extensively to understand the expected behaviour of management strategies, but is increasingly being implemented to select management strategies for implementation in actual fisheries. The earliest use of MSE for such selection occurred in South Africa, where the control rules used to set total allowable catches (TACs) for the anchovy *Engraulis encrasicolus*, Engraulidae, and later the sardine *Sardinops sagax*, Clupeidae, fisheries were selected using what has since become known as MSE (Bergh and Butterworth 1987; Geromont *et al.* 1999; De Oliveira and Butterworth 2004). MSE has also been used in South Africa to select management strategies for the Cape hake *Merluccius paradoxus* and *M. capensis*, Merlucciidae (Rademeyer *et al.* 2008), rock lobster *Jasus lalandii* and *Palinurus gilchristi*, Palinuridae (Johnston and Butterworth 2005; Johnston *et al.* 2008) and most recently horse mackerel, *Trachurus trachurus capensis*, Carangidae (Furman and Butterworth 2012) fisheries. The use of management strategies that have been tested using simulation has been routine for the major fisheries in South Africa for some 20 years.

Management strategy evaluation has been used extensively by the International Whaling Commission (IWC) since the late 1980s to select management strategies to calculate potential catch limits for commercial whaling and determine actual strike limits for aboriginal subsistence whaling

[1]A term introduced into the fisheries lexicon by Smith (1994). To the extent possible, the nomenclature for MSE outlined by Rademeyer *et al.* (2007) is followed throughout.

(Punt and Donovan 2007). The use of MSE accelerated internationally following a 1998 ICES Symposium on Confronting Uncertainty in the Evaluation and Implementation of Fisheries-Management Systems, which included several papers illustrating the methods underlying MSE and then current applications (Butterworth and Punt 1999; Cooke 1999; Smith *et al.* 1999).

Management strategy evaluation has been used by the Commission for the Conservation of Southern Bluefin Tuna (CCSBT) to select a management strategy for southern bluefin tuna *Thunnus maccoyii*, Scombridae (Kurota *et al.* 2010; Anonymous 2011). The Potential Biological Removals method, used to determine upper limits on anthropogenic removals of marine mammals in the USA, was also developed using MSE (Wade 1998). Similarly, MSE was used to evaluate a by-catch management control rule for seabirds (Tuck 2011). Outside of South Africa and the IWC, MSE has been applied most extensively in Australia where it has been used to compare and select management strategies for the Southern and Eastern Scalefish and Shark Fishery, SESSF (Punt *et al.* 2002; Wayte and Klaer 2010; Little *et al.* 2011), the Queensland spanner crab *Ranina ranina*, Raninidae fishery (Dichmont and Brown 2010), the Northern Prawn Fishery (Dichmont *et al.* 2008, 2013), the fishery for southern rock lobster *Jasus edwardsii*, Palinuridae off South Australia (Punt *et al.* 2012a) and the Tasmanian abalone fishery (Haddon and Helidoniotis 2013). The management strategies used to recommend catch limits for southern rock lobster off New Zealand have also been selected using MSE (Starr *et al.* 1997; Breen and Kim 2006).

Management strategy evaluation has been applied extensively to European fisheries to explore the performance of management strategies theoretically (Kell *et al.* 2005a,b, 2006), but few applications have resulted in strategies being formally implemented. The International Council for the Exploration of the Seas (ICES) provides a list of 18 management plans for North East Atlantic stocks that have been evaluated using MSE approaches since 2008 (ICES 2013). As an advisory body to the governments of ICES member countries and the European Commission, ICES bases its advice on these management plans if advice recipients have agreed that they can be used as a basis for that advice and provided the MSEs have shown them to fulfil ICES' precautionary criteria (ICES 2012). If this does not apply, ICES reverts to its own MSY framework, and if there is no basis for giving MSY-related advice, takes account of precautionary considerations (ICES 2012). The European Commission has its own advisory body, the Scientific, Technical and Economic Committee for Fisheries (STECF, established to advise on matters pertaining to the conservation and management of living aquatic resources) that performs impact assessments of proposed management plans, and may make use of MSEs for this purpose (STECF 2011a).

In North America, MSE has been applied to evaluate management strategies for the fishery for the northern subpopulation of Pacific sardine *Sardinops sagax caerulea*, Clupeidae, and the control rule used for this fishery from 1998 until 2012 was based on a MSE (PFMC 1998), as was the 2014 revision to this control rule (Hurtado-Ferro and Punt 2014). A revision to the management strategy adopted became necessary when the estimated relationship between recruitment success and environmental factors changed given new information. MSE has also been used to establish a management strategy for sablefish *Anoplopoma fimbria*, Anoplopomatidae off British Columbia (Cox and Kronlund 2008), for West Greenland halibut *Reinhardtius hippoglossoides*, Pleuronectidae (Butterworth and Rademeyer 2010; NAFO 2010) and for pollock *Pollachius virens*, Gadidae off eastern Canada (Rademeyer and Butterworth 2011).

Management strategy evaluation has recently been used to evaluate alternate management strategies for Tristan rock lobster (*Jasus tristani*) for three of the islands that form the Tristan da Cunha group of islands (Johnston and Butterworth 2013, 2014; Butterworth and Johnston 2014).

The focus for most previous MSEs has been single-species systems. However, MSE has also been used to evaluate management strategies to achieve multispecies or ecosystem objectives (Sainsbury *et al.* 2000; Fulton *et al.* 2007; Dichmont *et al.* 2008, 2013; Plagányi *et al.* 2013).

Management strategy evaluation is at the interface between science and policy. While it would be desirable to keep science and policy separate, there is a link. Decision-makers need to identify the desirable outcomes that any management strategy adopted should aim to achieve, while scientific analyses (the MSE) can inform the decision-makers on the feasible ranges of trade-offs. A well-structured MSE will utilize the links between policy and science, but ensure that a 'wall of science' remains whereby decision-makers do not decide scientific issues and scientists do not make policy decisions (Field *et al.* 2006).

While MSE is widely acknowledged to be the most appropriate way to compare management strategies, and the basic approach has been summarized in many publications, actual uses can differ markedly with regard to the likelihood that the resulting management strategy actually provides the best trade-off amongst the management objectives and is robust to uncertainty. Furthermore, it is well recognized that poorly conducted MSEs are likely to be less useful for management purposes than using the traditional best assessment approach coupled to essentially *ad hoc* advice (Rochet and Rice 2009; Butterworth *et al.* 2010b; Kraak *et al.* 2011). This paper therefore outlines the process of conducting MSEs and identifies a set of 'best practice' guidelines (Table 1). These proposed best practices for MSEs should assist in facilitating that MSEs are conducted in the most appropriate manner so that the resulting management strategies are best able to achieve their goals. The focus of the paper is on single-species applications of MSE, although applications that consider multispecies and ecosystem aspects are also considered. The extent to which these guidelines have been followed in practice is illustrated for two case-studies: the management strategy for the northern subpopulation of Pacific sardine and that for bowhead whales, *Balaena mysticetus* Balaenidae, in the Bering Sea, Chukchi Sea and Beaufort Sea.

MSE – the basics

The basic steps that need to be followed when conducting a MSE (Fig. 1) are as follows:

1. identification of the management objectives in concept and representation of these using quantitative performance statistics;
2. identification of a broad range of uncertainties (related to biology, the environment, the fishery and the management system) to which the management strategy should be robust;
3. development of a set of models (often referred to as 'operating models') which provide a mathematical representation of the system to be managed; an operating model must represent the biological components of the system to be managed, the fishery which operates on the modelled population, how data are collected from the managed system and how they relate to the modelled population (including the effect of measurement 'noise'); in addition, an implementation model is required that reflects how management regulations are applied in practice; note that more than a single operating model is nearly always required because of the need to cover the range of the ever-present uncertainties, which include the imprecision of the values of parameters estimated from fits to data, as well as structural uncertainties such as how many reproductively separate stocks of a species are present in the region considered;
4. selection of the parameters of the operating model(s) and quantifying parameter uncertainty (ideally by fitting or 'conditioning' the operating model(s) to data from the actual system under consideration);
5. identification of candidate management strategies which could realistically be implemented for the system;
6. simulation of the application of each management strategy for each operating model; and
7. summary and interpretation of the performance statistics; this may lead to refinement of the relative weighting of the management objectives as the simulation process develops and continues to provide more refined results to inform the quantitative trade-offs among competing goals.

The feedback loop between the management strategy and the operating model(s) is a fundamental aspect of MSE and is the particular feature, which distinguishes it from simple risk assessment where the implications of unchanging management regulations (e.g. constant TAC) are evaluated by use of projections. Simple risk assessment overestimates risk through failing to take account of management reactions to the information provided by future data.

Management strategy evaluation is not the same as conducting projections from a stock assessment, although a stock assessment may form the basis for the operating model(s) which are core to a MSE. Specifically, MSE takes feedback control into account, that is it takes account of the collection and use of future data on the status of the managed system. In addition, stock assessments usually involve selecting a single model structure and estimating the parameters of the model.

Table 1 Summary of the best practice guidelines.

Selection of objectives and performance metrics

- Involve decision-makers and stakeholders (e.g. using workshops) throughout the process to ensure the performance statistics capture the management objectives and are understandable.
- At a minimum, report statistics related to average catches, variation in catches and the impact on stock size.

Selection of uncertainties

- Consider a range of uncertainties, which is sufficiently broad that new information collected after the management strategy is implemented should generally reduce rather than increase this range.
- Include trials for each potential source of uncertainty (unless there is clear evidence that the source does not apply) and for the factors considered in Table 3.
- Consider the need for spatial structure, multiple stocks, predator-prey interactions and environmental drivers on system dynamics; modelling the last by imposing trends on the parameters of the operating model is often sufficient to understand its implications.
- Include predation effects using minimum realistic models and examine the potential for technical interactions amongst major fished species, especially in multispecies fisheries.
- Divide the trials into ‘reference’ and ‘robustness’ sets.
- Use Bayesian posterior distributions to capture the parameter uncertainty for each trial if possible.

Identification of candidate management strategies

- This should be the primary responsibility of the stakeholders/decision–makers, but with guidance from the analysts given the limitations of the management strategy evaluation (MSE). Care needs to be taken that the management strategy can be implemented in practice.
- Evaluate the entire management strategy. In cases in which the management strategy is complex, this may be impossible computationally, in which case a simplification of the assessment method is needed – the nature of the simplification should be based on simulation analyses.

Simulation of the application of the management strategy

- Check that operating model and management strategy are consistent with reality; projections into the future should generate quantities, such as past assessment errors and levels of variability in biomass and recruitment, on the same scales as those estimated to have occurred in the past.
- Conduct tests of the software, for example using ‘perfect’ data before conducting actual analyses.
- Base recommendations for management actions in management strategies only on data which would (with near certainty) actually be available.
- Document any assumptions regarding parameters assumed known when applying the management strategy.

Presentation of results and selection of a management strategy

- Develop a process, so that the decision-makers understand the results of the MSE and the range of trade-offs which are available to them.
- Use effective graphical summaries which are developed collaboratively with the stakeholders.
- Identify whether there are ‘performance standards’ which must be satisfied to eliminate some possible management strategies immediately and hence simplify the final decision process.
- Select a method for assigning a plausibility rank to each trial and take these ranks into account when making a final selection among candidate management strategies.

Other

- Include ‘Exceptional Circumstances’ provisions which specify the situations under which a management strategy’s recommendations may be over-ridden.
- Include a schedule for when formal reviews of the implemented management strategy will take place.

Although an aim of a stock assessment is to quantify uncertainty, it is rarely possible to capture all the key sources of uncertainty within the confines of a stock assessment, in particular ‘outcome uncertainty’ (see below), and to carry uncertainty forward fully into the provision of management advice. MSE can also be used when it is not possible to apply standard methods of stock assessment, as is common in data-poor situations.

Although not the focus of the present paper, Marasco *et al.* (2007) observe that the results from a MSE may be used not only to choose amongst the candidate management strategies, but also to identify future research and monitoring goals. In addition, the results of a MSE can be used to evaluate how well existing monitoring and data analysis methods are able to reflect the true status of the system with reasonable accuracy (see e.g. Ful-

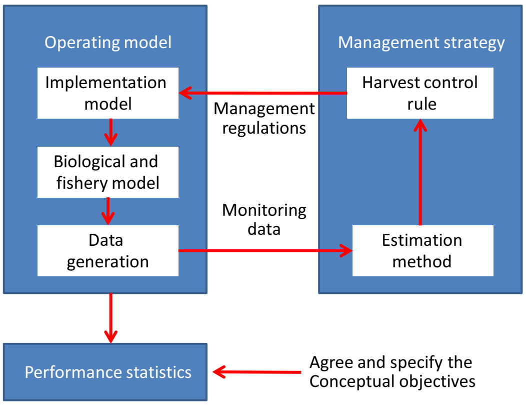

Figure 1 Conceptual overview of the management strategy evaluation modelling process.

ton *et al.* 2004; for an evaluation of ecosystem indicators). Marasco *et al.* (2007) also emphasize the need to continue to monitor the system following the implementation of a management strategy. Consistent with practice in, for example, the IWC and South Africa (Butterworth 2007; Punt and Donovan 2007), they stress the need to review and revise the MSE periodically, as consolidated outcomes from future monitoring and research become available.

Overview of the case-studies

Bering–Chukchi–Beaufort Seas bowhead whales

Bowhead whales in the Bering, Chukchi and Beaufort Seas are considered to be a single stock, separate from the stocks in the Okhotsk Sea, the Davis Strait and Hudson Bay, and off Spitsbergen. This stock, often referred to as the Bering–Chukchi–Beaufort (or BCB) Seas stock of bowhead whales, has been subject to hunting by aboriginal peoples off Alaska (USA) and Russia for centuries. In common with other stocks of bowhead whales, it was severely depleted by commercial whaling, which occurred between 1848 and 1914 in the case of the BCB stock. Commercial whaling on the BCB bowhead stock ceased once whaling there became economically non-viable, but aboriginal whaling continues at low levels.

Management of bowhead whales is challenging because individuals can live beyond 100 years (George *et al.* 1999). In addition, the location of the population and the fishery makes monitoring difficult (it involves ice platform sighting surveys of bowhead whales as they migrate through leads which open as the ice thaws). The aboriginal hunt of bowhead whales off Alaska and Russia is managed under the IWC. Management for aboriginal whaling is based on strike limits, which are the number of strikes of whales permitted within a season. Management advice is based on the number of strikes rather than numbers of whales landed because of the need to account for mortality when animals are struck but subsequently not landed ('lost').

Each country wishing to take whales for aboriginal subsistence purposes must provide the IWC with a 'Need Statement' which documents the number of annual strikes needed to satisfy the requirements of aboriginal peoples in terms of nutrition and culture. Management advice in the context of the BCB bowhead whales relates to whether the need requested can be satisfied without impacting the ability to achieve conservation-related management goals; this contrasts with commercial whaling, where the aim is to maximize the catch subject to the same constraint. The development of a management strategy for aboriginal subsistence whaling, and in particular for the BCB bowhead whales, commenced in 1995 after a management strategy for commercial whaling was adopted in 1994 (IWC 1994). A 'Strike Limit Algorithm' (SLA) was later adopted as the management strategy for the BCB bowhead whales in 2003 (IWC 2003). Prior to the use of the SLA, evaluation of whether the need requested was consistent with the IWC's conservation-related objective involved comparing the proposed need in terms of strikes with an estimate of a lower percentile (usually the lower 5th percentile) of a distribution for the replacement yield (the number of animals removed from the population each year which will keep the population at its current level; Givens *et al.* 1995).

The development of the SLA involved the IWC identifying management objectives for aboriginal subsistence whaling, obtaining a 'need envelope' from hunters and their scientific representatives (the range of possible maximum need levels by year over the next 100 years), developing operating models tailored to the dynamics of the BCB bowhead whale population, and simulating the application of candidate SLAs (equivalent to management strategies). The operating models for the BCB bowheads were case-specific, rather than generic as was the case for commercial whaling, because this was considered likely to lead to an improved ability to satisfy the management goals and because there are only a few aboriginal whaling fisheries. The development process was competitive, with several sets of 'developers' 'competing' to best satisfy the management goals. However, as it happened, the final selected SLA was none of these individual SLAs, but rather an average of the best two.

Northern subpopulation of Pacific sardine

The northern subpopulation of Pacific sardine is harvested off Mexico, the USA (including Alaska) and Canada. The population dynamics of Pacific sardine, in common with those of many small pelagic fish species, are characterized by large changes in abundance, driven primarily by environmental conditions. The long-term nature of these fluctuations has been confirmed for Pacific sardine using samples of fish scales from sediment cores in the Santa Barbara basin (Soutar and Isaacs 1969, 1974; Baumgartner *et al.* 1992). Sardine populations in the Santa Barbara basin are estimated to have peaked at intervals of approximately sixty years. The biomass and catch of Pacific sardine increased rapidly during the 1930s until the mid-1940s, and declined thereafter. The decline was likely due to a combination of environmental conditions leading to poor recruitments and very high fishing mortality rates.

The biomass of Pacific sardine began to rebuild during the 1980s, and by 1991 a directed fishery was re-established. The Pacific sardine fishery was managed by the State of California until 2000 when management authority was transferred to the Pacific Fishery Management Council (PFMC; Hill *et al.* 2011). Harvest Guidelines for Pacific sardine between 1998 and 2012 were set using a harvest control rule of the form (PFMC 1998):

$$\mathrm{HG} = (\mathrm{BIOMASS} - \mathrm{CUT\text{-}OFF}) * \mathrm{FRACTION} * \mathrm{DISTRIBUTION}$$

where: HG (Harvest Guideline) is the allowable catch for each management year; BIOMASS is the estimate of the biomass of Pacific sardine aged 1 and older obtained from an age-structured stock assessment model; CUT-OFF is 150 000 mt and is the escapement threshold below which fishing is prohibited; FRACTION is a temperature-dependent exploitation fraction which ranges from 5 to 15%; and DISTRIBUTION is the average proportion of the coastwide biomass in USA waters, estimated at 0.87. In addition, there is a maximum allowable catch regardless of biomass such that HG ≤ MAXCAT, where MAXCAT is 200 000 mt. The purpose of CUT-OFF is to protect the stock when biomass is low. The purpose of FRACTION is to specify how much of the stock is available to the fishery when BIOMASS exceeds CUT-OFF. The DISTRIBUTION term recognizes that the stock ranges beyond USA waters and is therefore subject to foreign fisheries. In PFMC (1998), FRACTION was determined on the basis of a 3-year running average of the temperature at Scripps Pier, La Jolla, USA.

The overarching management plan for all coastal pelagic species (CPS) managed by the PFMC was modified in 2011 to be consistent with the 2006 reauthorization of the MSA. This involved formally introducing how the overfishing limit (OFL, the annual catch amount consistent with an estimate of the annual fishing mortality that corresponds to maximum sustainable yield) is calculated, as well as the acceptable biological catch (ABC, a harvest limit set below the OFL that incorporates a buffer against overfishing to take account of scientific uncertainty).

The specifications of the harvest control rule adopted in 1998 were determined using simulations in which the population dynamics were represented by a production model where productivity was related to an environmental variable (PFMC 1998). Results of assessments conducted after 1998 were analysed during a workshop in February 2013 (PFMC 2013) which suggested that the temperature at Scripps Pier no longer exhibited the same trends as most other measures of temperature for the offshore waters to the west of North America (McClatchie *et al.* 2010). Rather, the relationship between recruitment, spawning biomass and temperature was strongest when temperature was based on sea surface temperature obtained from CalCOFI samples (PFMC 2013).

The results from the February 2013 workshop formed the basis for developing a set of operating models for the northern subpopulation of Pacific sardine, as well as an initial set of candidate management strategies (PFMC 2013). The process of selecting the operating models and the candidate management strategies was iterative, involving presentations by the analysts to the PFMC as well as its Scientific and Statistical Committee, Coastal

Pelagic Species Advisory Panel and Coastal Pelagic Species Management Team. The PFMC took advice from these advisory bodies as well as from members of the public, including industry and environmental non-governmental organizations (ENGOs), and then directed the analysts. Hurtado-Ferro and Punt (2014) summarize the most recent MSE results, along with the specifications for the operating models and candidate management strategies.

Best practices for MSE

Establishing objectives and performance statistics

One of the main strengths of MSE is that the decision-makers clarify their objectives. Objectives for fisheries management can be categorized as either 'conceptual' ('strategic') or 'operational' ('tactical'). Conceptual objectives are generic, high-level policy goals. For example, the conceptual objectives for CPS off the USA west coast (i) promote efficiency and profitability in the fishery, including stability of the catch; (ii) achieve 'Optimum Yield' (OY); (iii) encourage cooperative international and interstate management of CPS; (iv) accommodate existing fishery sectors; (v) avoid discards; (vi) provide adequate forage for dependent species; (vii) prevent overfishing; (viii) acquire biological information and develop a long-term research programme; (ix) foster effective monitoring and enforcement; (x) use resources spent on management of CPS efficiently; and (xi) minimize gear conflicts (PFMC 2011). These goals are largely self-consistent, but this need not always be the case. For example, the conceptual objectives for aboriginal subsistence whaling (i) ensure that risks of extinction are not seriously increased by whaling; (ii) enable native people to hunt whales at levels appropriate to their cultural and nutritional requirements (i.e. satisfy their 'need'); and (iii) move populations towards and then maintain them at healthy levels. Objective (ii) may be in conflict with objectives (i) and (iii) for some populations.

To be included in a MSE, the conceptual objectives need to be converted into operational objectives (expressed in terms of the values of performance measures or performance statistics). This usually involves translating each conceptual objective into one or more operational objective(s) and performance statistic(s). For example, the conceptual objective of 'avoid overfishing' could be represented operationally as 'the annual probability that the stock drops below 20% of the unfished level should not exceed 5%'. However, some conceptual objectives may link to multiple operational objectives. For example, the conceptual objective 'achieve OY' could be quantified by the operational objectives 'maximize catch in biomass', 'minimize the interannual variation in catch' and 'maximize the economic rent to the fishing industry', amongst others.

The operating model(s) should be developed so that performance statistics can be calculated. For example, when there are explicit ecosystem and economic objectives, the operating model(s) may need to include a fleet dynamics model (Ulrich *et al.* 2007) or models of how fishing impacts ecosystem components other than the target species (Schweder *et al.* 1998), as well as related performance statistics.

It is inevitable that some of the objectives will be in conflict to some extent, and one aim of MSE is to highlight trade-offs among the objectives as quantified using performance statistics. For example, increased monitoring efforts may allow higher catches for the same level of risk (see Fig. 2), but the increased monitoring will come at a financial cost. The more common trade-offs are between risk to the resource and benefits to the fishery, and between average catch and variation in catch (the less variability in catch permitted, the lower the average catch needs to be able to accommodate catch reductions needed at times the resource might be at a low abundance). It is critical to ensure that the decision-makers are aware of these trade-offs. One way to achieve this is to use a utility function which balances the various factors in providing a single number. However, efforts to base MSEs on utility functions have generally been unsuccessful because decision-makers (and stakeholder groups) wish to see how well each candidate management strategy achieves each objective and how they trade off.

The difficulties associated with conflicting objectives become more challenging when management strategies are developed for fisheries which target multiple species, or when there are multiple stakeholder groups which fish using different gears or may have markedly different objectives (e.g. commercial and recreational sectors within a fishery). This is because what is seen as the 'optimal' state of the system will differ among stakeholders. Few management strategies that have been

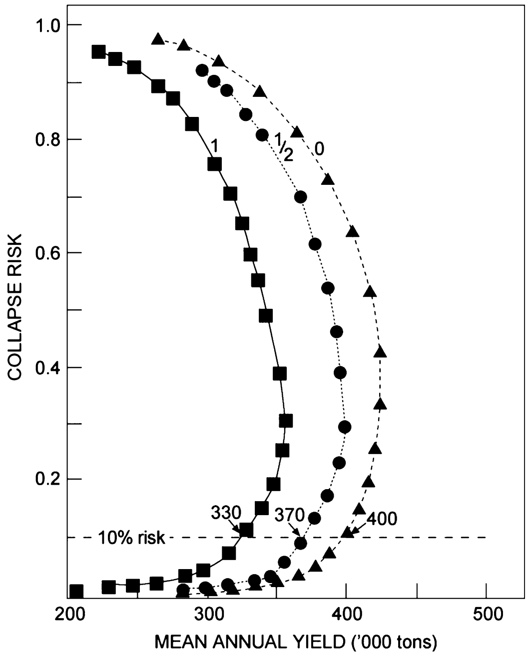

Figure 2 Relationship between risk and reward for South African anchovy ('collapse' is defined here as the spawning biomass falling below 10% of its average unexploited level, and risk reports the probability of that happening at least once during a 20-year period). Each line indicates a different level of survey precision (1: current precision; ½: double the survey effort; 0: perfect information; reproduced from Bergh and Butterworth 1987).

implemented have addressed the issue of between-species trade-offs. One notable exception is the South African fishery for anchovy and sardine. Here, there is a trade-off between anchovy and sardine catches because of an unavoidable by-catch of juvenile sardine with anchovy, which decreases the TAC possible for the more valuable adult sardine. This multiple species allocation problem was addressed over one period by allowing each company with rights to each of the two species to choose its preferred trade-off. First, the TACs that would follow under each company's desired trade-off were calculated; next quotas were allocated to that company which were computed by multiplying the TACs related to its trade-off by the proportional right to the combined (sardine and anchovy) fishery as a whole that it had been awarded; finally, the TACs themselves were calculated by summing the quotas awarded to each company for each species (De Oliveira 2003; Butterworth *et al.* 2012).

Best practice in terms of specifying objectives, particularly operational objectives, is through the use of inclusive workshops (Cox and Kronlund 2008; Mapstone *et al.* 2008; PFMC 2013). Workshop participants need to be representative of the decision-makers and other stakeholders, and efforts should be made to ensure that the decision-makers are fully aware of which decisions are theirs (weighting objectives, and selecting management strategies to be tested) and which decisions are primarily technical. Progress in such workshops may be facilitated by providing draft specifications that can be criticized, expanded upon or rejected outright.

The statistics used to evaluate the performance of alternative candidate management strategies should be chosen, so that they are easy for decision-makers and stakeholders to interpret (Francis and Shotton 1997; Peterman 2004). Butterworth and Punt (1999) comment that standard deviations or coefficients of variation of catch limits are difficult for many stakeholders to understand. Experience suggests that stakeholders find it much easier to relate to statistics such as the fraction of years in which the catch is less than some desirable level. Care should be taken to avoid having too many performance statistics. While it may seem desirable to have, for example, a number of performance statistics to quantify catch variation (IWC 1992), the final decision process is made considerably easier if the number of performance statistics is small, so that they can easily be summarized graphically. In any case, experience suggests that such catch variation statistics are often highly correlated with each other.

It is common to include performance statistics such as the probability of dropping below some threshold level [such as the minimum stock size threshold (MSST) defined in the USA MSA, or 20% of the estimated unfished biomass, B_0]. However, while dropping below MSST has implications in the USA (leading to the requirement for a rebuilding plan), the use of a metric such as the probability of dropping below a fixed fraction of B_0 can be criticized both because any such level is somewhat arbitrary, and because there is seldom evidence for threshold or depensatory effects. Nevertheless, in relation to answering questions of direct interest to decision-makers, such policy-related performance statistics may need to be included in the set reported. ICES (2013) notes that there are three ways to define the probability of dropping below a threshold: (i) the average probability (over simulations and years) of being below the threshold; (ii) the probability (over simulations) of dropping below the threshold at least once during each projection; and (iii) the maximum annual probability (over simulations) of being below the threshold over the projection period. Other ways to summarize these probabilities exist, including, for example, the probability in a given year. de Moor *et al.* (2011) comment that the probability of dropping below a threshold depends on the extent of process error, and define a performance statistic that evaluates risk in terms of the extent to which this probability increases with fishing, relative to its value in the absence of fishing.

A complicating factor with performance statistics that pertain to population size relative to B_0 or the MSY level is how these are to be defined in a changing environment and when there is time-varying predation (A'mar *et al.* 2009a,b, 2010). Usually, this problem has been resolved by replacing carrying capacity in these statistics by the population size which would have occurred had there been no catches (IWC 2003). However, for multispecies operating models, this can lead to counter-intuitive results where the unfished level is actually lower than the fished population size (Mori and Butterworth 2006).

Although most operational objectives relate to the fishery and the conservation of the species on which it depends, increasingly these objectives include ones pertaining to ecosystem impacts of fishing (Dichmont *et al.* 2008) and economic objectives (Dichmont *et al.* 2008; Anderson *et al.* 2010). Performance statistics can also relate to the management system itself. For example, in a MSE to evaluate management strategies for overfished USA west coast groundfish stocks, Punt and Ralston (2007) considered performance statistics such as when rebuilding was estimated to have occurred compared to how long this had been anticipated to take when the rebuilding plan was developed, and how often a rebuilding plan failed. This was because these were issues of interest to the decision-makers, which also related to the confidence stakeholders and the public have in the management system.

In the context of BCB bowheads, the conceptual objectives were selected by the IWC (i.e. the Commissioners). The operational objectives (and related performance statistics) were selected by the Scientific Committee of the IWC to reflect the intent of the conceptual objectives. These included statistics related to (i) the proportion of the nutritional and cultural need requested by aboriginal communities which could be satisfied, (ii) the delay in rebuilding to the population size corresponding to MSY caused by the mortality permitted and (iii) measures of the variation in the number of strikes permitted. No performance statistics specifically related to extinction risk were considered because none of the management strategies explored led to an appreciable risk of extinction – indeed the probability of extinction was zero for all management strategies and (plausible) simulations. The performance statistics related to the delay in rebuilding were hard to interpret, so that the final conservation-related statistics were based on simpler concepts such as the lowest ratio of population size to carrying capacity and the ratio of population size to carrying capacity after 100 years of simulated management.

The performance statistics for the Pacific sardine MSE were initially proposed during a workshop with stakeholders (PFMC 2013); these statistics were then refined based on input from the PFMC and its advisory bodies. The final set of performance statistics included conventional statistics related, for example, to average catches and variation in catches. However, the performance statistics also included quantities such as the proportion of times that the fishery was closed or its catch was <50 000 t, the average number of consecutive years the fishery was predicted to be closed, and the proportion of years that the biomass of animals aged 1 and older was <400 000 t. The last statistic was a proxy for indications of whether the biomass is sufficiently low that predators may be impacted.

Selection of uncertainties to consider and selection of operating model parameter values

Ideally, the range of uncertainties considered in a MSE should be sufficiently broad that new information collected after the management strategy is implemented should reduce rather than increase the range (Punt and Donovan 2007; IWC 2012a). However, in practice, it is seldom the case that it is possible to come close to incorporating all the pertinent uncertainties fully for any given situation, and choices are needed as to which uncertainties are the most consequential and reflect more plausible alternative hypotheses. Several attempts (Francis and Shotton 1997; Haddon 2011a) have been made to characterize sources of uncertainty. For the purposes of this paper, five sources of uncertainty are distinguished.

1. Process uncertainty: variation (usually assumed to be random, though sometimes incorporating autocorrelation) in parameters often considered fixed in stock assessments such as natural mortality, future recruitment about a stock–recruitment relationship and selectivity.
2. Parameter uncertainty: many operating models are fit to the data available, but the values estimated for the parameters of those operating models (e.g. fishery selectivity-at-age, the parameters of the stock–recruitment relationship and historical deviations in recruitment about the stock–recruitment relationship) are subject to error.
3. Model uncertainty: the form of relationships within an operating model will always be subject to uncertainty. The simplest type of model uncertainty involves, for example, whether the stock–recruitment relationship is Beverton–Holt or Ricker, whether a fixed value for a model parameter is correct, or whether fishery selectivity is asymptotic or dome-shaped. However, there are other more complicated types of model structure uncertainty such as how many stocks are present in the area modelled, the error structure of the data used for assessment purposes, the impact of future climate change on biological relationships such as the stock–recruitment function, and ecosystem impacts on biological and fishery processes.
4. Errors when conducting assessments, which inform the catch control rule that is being evaluated using the MSE: management advice for any system is based on uncertain data. Consequently, the data that inform catch control rules need to be generated in a manner which is as realistic as possible. Uncertainty arises when the model used for conducting assessments and providing management advice differs from the operating model, or the data are too noisy to estimate all key parameters reliably.
5. Outcome (or 'implementation') uncertainty: the impact of fishers and other players in the management system on the performance of management strategies has long been recognized (Rosenberg and Brault 1993; Rosenberg and Restrepo 1994). The most obvious form of this type of uncertainty is when catches are not the same as the TACs – typically more is taken or the decision-makers do not implement the TACs suggested by the management strategy. However, there are many other sources of outcome uncertainty, such as that associated with catch limits set for recreational fisheries and regulating discards. In some cases, this source of uncertainty has been found to dominate all the others (Dichmont *et al.* 2008; Fulton *et al.* 2011a).

In general, the evaluation of management strategies proceeds by first identifying the set of factors which are perceived to contribute the most to the uncertainty for the case in question. There will usually be factors for each of the five sources of uncertainty listed above. For example, factors could be 'the extent to which carrying capacity changes into the future', or 'the variation in realized catches about those intended'. Each factor will have a number of levels: for example, different rates of change in carrying capacity or variations in realized catch about the intended catch. Trials would then be constructed by selecting a level for each factor and thereby represent the range of uncertainty about a hypothesis to be considered in the evaluation. Best practice for a specific case involves explicitly addressing each of these uncertainties, or at least indicating how the uncertainties reflected were selected. Minimally, a MSE should consider (i) process uncertainty, in particular, variation in recruitment about the stock–recruitment relationship; (ii) parameter uncertainty relating to (a) productivity and (b) the overall size of the resource; and (iii) observation error in the data used when applying the management strategy. Which uncertainty is most important will be case-specific. For example, process uncertainty is unlikely to be very important for the management of large whale populations, whereas this uncertainty could be very consequential for a short-lived species such as Pacific sardine; the uncertainty factors considered in the MSEs for the two case-studies unsurprisingly differed markedly (Table 2).

Best practice is to divide MSE trials into a 'reference' (or 'base case') set of trials and a 'robustness' set of trials (Rademeyer *et al.* 2007). The reference trials are considered to reflect the most plausible hypotheses (see below for further comments on assigning plausibility to trials) and hence form the primary basis for identifying the 'best' management strategy, while the robustness trials are used to determine whether the management strategy behaves as intended in scenarios that are fairly unlikely, even though they are still plausible. While it is clearly desirable to conduct trials for all combinations for the levels for each factor (Kurota *et al.* 2010), this is often computationally impossible except when the management strategies being evaluated are fairly simple (Carruthers *et al.* 2014), and even then, conducting a MSE could be very computer-intensive depending on how many trials are run. Although partial factorial designs can be used to address this difficulty (Schweder *et al.* 1998), it is more common to select 'base' levels for each factor (in some cases multiple 'base levels'), and then to develop trials which involve varying each 'base' level in turn, perhaps also adding a few trials in which multiple factors are changed from their 'base' levels.

Table 2 Factors related to uncertainties considered in the simulation trials developed to test the management strategies for the Bering–Chukchi–Beaufort (BCB) Seas bowhead whales and the northern subpopulation of Pacific sardine.

BCB bowhead whales	Pacific sardine
Population dynamics	
• Inherent productivity	Extent of variation in recruitment
• Shape of the production function	Time-varying natural mortality
• Process error in calving rate	Time-varying productivity[1]
• Time trends in carrying capacity	Changes in selectivity spatially
• Time trends in productivity	Time-varying selectivity
• Occasional catastrophic mortality or recruitment events	Time-varying weight-at-age
• Time lags in the density dependence function	
• Alternative stock structure hypotheses[2]	
Data related	
• Survey frequency	Extent of auto-correlation in biomass estimates
• Average bias of survey estimates	Extent of variation in biomass estimates
• Trends in bias of survey estimates	Biomass estimates non-linearly related to true abundance
• Survey CV	
• Bias in reported catches	
Implementation related	
• Survey conducted to maximize strike limits	Only the USA follows the control rules

1All trials allowed for some variation in productivity due to environmental effects, but the manner in which productivity was related to the environment was varied in these trials.
2Conducted during the 2007 *Implementation Review* (International Whaling Commission 2008a,b, 2009, 2014).

Kraak *et al.* (2011) assert that the choice of sources of uncertainty included in MSE simulations often is quite arbitrary, and the uncertainties chosen do not necessarily reflect the key sources of uncertainty. They note that some MSEs conducted in Europe ignore spatial structure and whether egg production rather than spawning biomass drives recruitment. If these were indeed key uncertainties for the resources concerned, the scientists conducting those MSEs would clearly have been in error in ignoring them.

Best practice would involve trials based on at least a standard set of factors (Cooke 1999), so that the simulations extend over the set of uncertainties found to have had a large impact on the performance of management strategies for other systems (a list is given in Table 3). Most early operating models considered a single stock, ignored climate drivers of recruitment, growth and natural mortality, and treated the area being managed as a single homogeneous region. Each of these limitations can be overcome, particularly given the availability of sufficient computing resources. For example, although Butterworth and Punt (1999) commented that there were very few operating models which accounted for spatial structure when they conducted their review in 1998, subsequently Punt *et al.* (2005), IWC (2008a,b, 2009, 2014), Punt and Hobday (2009) and Carruthers *et al.* (2011) have all developed operating models which can, to some extent, account for spatial structure.

Climate and environmental variation is increasingly recognized as factors which often need to be included when evaluating management strategies. Two basic approaches have been adopted. The first is to include these factors in end-to-end models which represent entire ecosystems from physical processes to high trophic levels and fisheries, such as Atlantis (Fulton *et al.* 2011b) and Ecopath-with-Ecosim (Gaichas *et al.* 2012). The second is to relate environmental change to values of parameters empirically (Punt *et al.* 2014). Under the latter approach, environmental change can be

Table 3 List of factors, whose uncertainty commonly has a large impact on management strategy performance, which should be considered for inclusion in any management strategy evaluation.

Productivity	*Data-related issues*
• Form and parameters of the stock–recruitment relationship.	• CVs and effective samples sizes of data.
• Presence of depensation.	• Changes in the relationship between catchability and abundance.
• Extent of variation and correlations in recruitment about the stock–recruitment relationship.	• Changes in survey bias (fishery-independent data).
• Occasional catastrophic mortality or recruitment events.	• Survey and sampling frequency.
	• Ageing error.
	• Historical catch inaccuracy (bias).
Non-stationarity	*Outcome (Implementation) uncertainty*
• Changes in the stock–recruitment relationship.	• Decision-makers adjust or ignore management advice.
• Time-varying natural mortality (potentially a multispecies operating model).	• Realized catches differ from total allowable catches due to mis-reporting, black market catches, discards, etc.
• Time-varying carrying capacity (regime shift; linked to environmental variables or multispecies effects).	
• Time-varying growth and selectivity.	
Other factors	
• Spatial and stock structure.	
• Technical interactions.	
• Time-varying selectivity, movement and growth.	
• Initial stock size (unless it is estimated reliably when conditioning the operating model).	

modelled by linking environmental variables to the parameters that determine the dynamics of the population represented in the operating model (A'mar *et al.* 2009a; Ianelli *et al.* 2011; Punt 2011), or regime shift changes in parameters can be modelled (A'mar *et al.* 2009b; Wayte 2011; Szuwalski and Punt 2013). Most studies in which biological parameters are driven by environmental effects are conducted in circumstances where the relationships between the environment and the population dynamics are largely unknown (Hurtado-Ferro and Punt 2014). Most previous MSEs have allowed only one parameter of the operating model to exhibit time trends. However, it is possible to force a number of operating model parameters to do so. For example, Kell and Fromentin (2010) explored the performance of a VPA-based management strategy where both recruitment and migration varied as a function of the environment, while Punt *et al.* (2013) investigated the robustness of a management strategy for rock lobsters off Victoria, Australia, to time trends in natural mortality, catchability and growth.

Ecosystem effects, in particular biological and technological interactions, can be addressed within the context of end-to-end models such as Atlantis and Ecosim. However, most current investigations of the impacts of ecosystem effects on the performance of management strategies have been based on models of intermediate complexity for ecosystems (MICE; Punt and Butterworth 1995; Schweder *et al.* 1998; Plagányi 2007; A'mar *et al.* 2010; Howell *et al.* 2013), primarily because it is possible to parameterize these types of models by fitting them to monitoring data, although this renders the conclusions case-specific (Plagányi *et al.* 2014).

Technical interactions are probably easier to include in operating models given that there are usually direct data on catches and by-catches by fleet. Such interactions have been included in the MSEs conducted by De Oliveira and Butterworth (2004) for South African sardine and anchovy, by Dichmont *et al.* (2006a) for two prawn species off northern Australia, by Punt *et al.* (2005) for two shark species of southern Australia and by Kraak *et al.* (2008) for the flatfish complex in the North Sea. Dichmont *et al.* (2006a) and Kraak *et al.* (2008) model effort allocation based on economic incentives that lead to technical interactions among species.

How realistically the data are generated will directly impact the performance of any assessment method, and therefore also of any management strategies which depend on the results of the assessment. For example, most simulation studies generate age/length composition data from the survey or fishery catch in a way that matches the distributions assumed when fitting the assessment model (Bence *et al.* 1993; Sampson and Yin 1998; Radomski *et al.* 2005). However, this means that even very small sample sizes can appear to be extremely informative. In contrast, the residual patterns for actual stock assessments are often suggestive of both overdispersion and model misspecification. It is important to ensure that a number of plausible relationships between indices and true abundance are considered when assessments rely on fishery-dependent index data.

Best practice for parameter uncertainty for a given model structure is to sample parameters values from a Bayesian posterior distribution, or less ideally to use bootstrap samples or to sample parameter vectors from the asymptotic variance–covariance matrix for the parameters. Constructing Bayesian posterior distributions or developing bootstrap distributions for parameters can, however, be very intensive computationally.

Although the ideal is to evaluate management strategies using a trial structure which has been developed for a given stock or system, this may be impossible to achieve for data-poor situations. Nevertheless, it remains important to evaluate management strategies for data-poor situations, especially when the management strategies use proxies for measures of biomass; consequently, extensive testing of management strategies for data-poor situations has been undertaken, particularly in Australia (Haddon 2011b; Little *et al.* 2011; Plagányi *et al.* 2013, in press) and New Zealand (Bentley and Stokes 2009a,b). In these cases, there is a value in developing management strategies which can be applied generically. Naturally, generic management strategies would not be expected to perform as well as a management strategy that has been developed for a specific case (Butterworth and Punt 1999). When an evaluation of generic management strategies is to be undertaken, it is necessary to ensure that a broad range of species life histories are explored, along with a broad range of hypotheses regarding the quality of past and future data, and the state of the stock when the management strategy is first applied (Wiedenmann *et al.* 2013; Carruthers *et al.* 2014; Geromont and Butterworth in press-a). The values for the operating model parameters in this case would be selected based on generic considerations, and values for species which are characteristic of those to which the management strategy is to be applied.

Finally, it would be naive to believe that it is possible to identify all key uncertainties correctly, and it should not come as a surprise that some potential uncertainties not taken into account during the development of a management strategy turn out to be consequential. Kolody *et al.* (2008) drew attention to a key uncertainty (underestimation of historical catches) that was not initially considered during the development of a management strategy for southern bluefin tuna (*T. maccoyii*). They also questioned whether analyses of historical data, for example, as part of the process of conditioning the operating model(s) to data will capture the full extent of uncertainties. This problem should not imply that it is not worthwhile to conduct a MSE, but rather emphasizes that the earlier view that management strategies can be developed to run on 'autopilot' for a large number of years is likely flawed. Thus, the value of management strategies including 'Exceptional Circumstances' provisions and conducting regular *Implementation Reviews* (see final section) is high and justified, even if it entails additional work. Butterworth (2008a) emphasizes that the operating models considered in MSE analyses should remain 'broadly comparable' with the data. In practice, this means that use of, for example, strict model selection criteria to weight trials should be considered very carefully; in particular, use of, for example, AIC-weighting or the analytic hierarchy process (Merritt and Quinn 2000) should only be considered when there is confidence that the likelihood function is reliable (which is often not the case because the data inputs are not completely independent, as is usually assumed). Best practice in cases when the data used to parameterize the operating model are in conflict, for example when the various indices of abundance exhibit different trends, is to develop alternative operating models which represent each data set (Butterworth and Geromont 2001).

Identification of candidate management strategies which could realistically be considered for implementation

Ultimately, the management strategy chosen should reflect the policies developed by the decision-makers. Management strategies can be divided roughly into those that are model-based and those that are empirical, although some management strategies could be considered to be a mixture of the two types of strategies (Starr *et al.* 1997). Broadly, model-based management strategies usually involve two stages (see below), although some management strategies such as the IWC's Revised Management Procedure (IWC 1994) integrate the two stages to the point that it is impossible to distinguish them. For southern bluefin tuna, the model-based part of the management strategy is in effect a biologically plausible smoother of the two abundance indices used, with the actual harvest control rule having more in common with empirical harvest control rules than the more traditional model-based versions (Anonymous 2011).

The first stage in a model-based management strategy involves applying a stock assessment method (which may be much simpler than the methods used to develop the operating models that provide the basis for the MSE simulation testing process), and the second involves taking the results from that stock assessment model as the input for a harvest control rule. Several jurisdictions, including the USA and Australia, apply complex model-based management strategies, at least for their 'data rich' stocks. Despite the process being very intensive computationally, these types of management strategies have been evaluated using simulation (Dichmont *et al.* 2006b; A'mar *et al.* 2008, 2009a,b, 2010; Anonymous 2011; Fay *et al.* 2011; Punt *et al.* 2013). Model-based management strategies tend to lead to lower variation in terms of, for example, TACs than empirical approaches that do not constrain the estimated dynamics using population models (Butterworth and Punt 1999; Anonymous 2011), although this effect may be alleviated by imposing constraints on the extent of interannual change permitted in catch limits (see below).

In contrast to model-based management strategies, empirical management strategies do not utilize a population model to estimate biomass, fishing mortality or related quantities for use in harvest control rules. Rather, they set regulations such as TACs directly from monitoring data, usually after some data summary methods have been applied (e.g. CPUE standardization for catch and effort data). For example, the empirical harvest control rule used to recommend annual catch limits for the South African sardine involves setting catch limits as a constant proportion of the resource abundance estimated from the most recent hydro-acoustic survey. This rule is then subject to a number of constraints, or meta-rules, such as a maximum TAC and a maximum amount by which the TAC can decrease interannually. By removing this latter constraint during years of high TACs, the rule was designed to be flexible enough to allow the industry to take advantage of the occasional 'booms' that are a feature of this highly variable resource, without increasing the risk of the resource dropping to an undesirably low level (de Moor *et al.* 2011).

Rademeyer *et al.* (2007) remark that empirical management strategies are easier to test and are often easier to explain to decision-makers, but have the disadvantage that there might not be a clear basis for determining the target at which the resource will eventually equilibrate (Little *et al.* 2011). Examples of management strategies implemented which are empirical are those for hake, rock lobster, horse mackerel, anchovy and sardine off South Africa, for rock lobsters off South Australia and Tristan da Cunha, for West Greenland halibut and for pollock off eastern Canada. Most empirical management strategies base management decisions on trends in an index of abundance. However, there is a move towards 'target'-based rules, where TAC changes depend on the difference between the most recent level and the target for some abundance-related index (Little *et al.* 2011; Rademeyer and Butterworth 2011; Geromont and Butterworth in press-b), because the resultant catch limits tend to show less variability without impacting performance on other statistics such as average catch and risk to the resource. An example of an empirical 'target'-based rule is that used to recommend annual catch limits for the South African south coast rock lobster: the annual TAC is adjusted up or down from that recommended for the previous year according to whether the most recent measure of standardized CPUE is above or below a target value, with the extent of TAC adjustment proportional to the magnitude of the difference between the recent CPUE and the target value (Johnston *et al.* 2014). Management strategies can also be based on changes in metrics defined from age and size compositions (Butterworth *et al.* 2010b; Wayte and Klaer 2010; Fay *et al.* 2011).

Many management strategies impose constraints on how much catch limits can vary from 1 year to the next. For example, the management strategies for Australia's SESSF include 10 and 50% rules, which state that no change in TAC up or down will be larger than 50% of the current TAC; similarly, if a predicted change is <10% of the current TAC, then no change is made. In South Africa, both the hake and rock lobster management strategies include maximum TAC changes of either 5 or 10%, although these are over-ridden if appreciable declines in abundance become evident from the indices monitoring resource abundance. These minimum change rules have the advantage of smoothing out what might be noise from the management strategy output arising from noise in its data inputs.

Most of the management strategies considered in MSEs have been based on the conventional data used for stock assessments (e.g. catches, indices of abundance, age/length composition information). However, it is possible to develop management strategies, particularly for data-poor situations, using non-conventional data. For example, McGilliard *et al.* (2010) and Babcock and MacCall (2011) developed management strategies that use the ratio of the density inside and outside of marine protected areas to adjust limits on catch and effort in fished areas. Wilson *et al.* (2010) extended these approaches to use data on the proportion of old fish in the population. Christensen (1997) defined (and evaluated) a management strategy in which limits on effort are a function of the economic rent from the fishery, while Pomarede *et al.* (2010) evaluated one based on estimates of total mortality. The management control in most management strategies changes based on the data collected (feedback strategies), although some management strategies for data-poor situations are effectively non-feedback, setting management controls based, for example, on historical catch only. The performance of non-feedback strategies is, however, generally poor (Carruthers *et al.* 2014).

While the candidate management strategies which could be adopted should be identified by the decision-makers (or their advisers), best practice for MSE is also to evaluate additional management strategies to better understand the behaviour of the strategies identified by the stakeholders and decision-makers. In particular, it is a valuable exercise to apply variants of a management strategy in which the state of the stock is known exactly by the management strategy because this provides an upper limit to the 'value of information'. In addition, having results for 'reference' strategies, such as the strategy which maximizes average catch, can be useful for determining whether or not differences in performance statistics among management strategies are meaningful.

Most management strategies involve changes in the values of traditional management instruments such as catch limits, the total amount of effort or the length of the fishing season. However, MSE can also be used to evaluate novel management strategies such as that of Kai and Shirakihara (2005) that involves changing the size of a closed area based on the results of monitoring data.

It is essential that the management strategies being tested or compared are fully specified and can be implemented both for the operating models and in reality. Best practice is to simulate the management strategy exactly as it would be applied in reality, and this is commonly done when the management strategy is empirical (De Oliveira and Butterworth 2004; Little *et al.* 2011; Punt *et al.* 2012a; Carruthers *et al.* 2014), or the assessment method is not very demanding computationally (Kell and Fromentin 2009). It is becoming easier to evaluate complex management strategies given the increased availability of, for example, distributed computing including cloud computing. However, simulating very complicated management strategies such as those that involve fitting a statistical catch-at-age model can still require considerable computation (e.g. a single set of 100 simulations of 45 years to evaluate the actual management strategy for Gulf of Alaska walleye pollock took over 3 weeks on a fast desktop computer) and run the risk that fully automated fitting procedures may not find the global minimum that would be detected in the comprehensive searches typical of 'best assessment' approaches. Consequently, it is common to approximate application of a management strategy, for example by assuming that the estimates of biomass are log-normally distributed about the true biomass, perhaps with autocorrelated errors (DiNardo and Wetherall 1999; Hilborn *et al.* 2002; Anderson *et al.* 2010; Punt *et al.* 2012b).

However, ICES (2013) comments that it is generally not sufficient to simply add random noise to quantities derived from the operating model, and express concern that only 4 of the 18 MSEs which they reviewed had simulation tested the actual assessment. Failing to simulate application of the actual assessment method allows a broader set of hypotheses to be explored quickly, but the risk is that the actual error distribution associated with assessments does not match that assumed, and hence the values of the performance statistics are incorrect. In the extreme, the resultant relative ranking of management strategies may become incorrect. The justification for using an approximation to a management strategy may be examined by running a few simulations for the actual management strategy and the approximation, and comparing the results to ascertain whether the approximation is adequate. For example, ICES (2008) compared a 'full' and 'shortcut' MSE and found that the ranking of the performance of the harvest control rules evaluated changed when conducting a shortcut MSE compared to a full MSE (i.e. the best performing harvest control rule was different for the two evaluations).

The management strategy adopted for the BCB bowhead whales is based on averaging the strike limits from two SLAs (IWC 2003): (i) an empirical relationship between the strike limit and estimates of carrying capacity, the replacement yield predicted for the year for which a strike limit is needed, and the current stock size (Johnston and Butterworth 2000; Givens 2003); and (ii) a control rule based on the concept of adaptive Kalman filtering (a combination of Kalman filtering and Bayesian methodology; Dereksdóttir and Magnússon 2003). Both SLAs included ways to restrict interannual variation in strike limits, a factor considered very important during the selection process for a SLA. In particular, the component SLAs included a 'snap to need' feature which sets the strike limit equal to the need if the strike limit indicated by the algorithm is very close to the need.

The management strategy used for Pacific sardine is based on a set of control rules (Fig. 3) that are applied to an estimate of age 1+ biomass from a stock assessment model. The value for the FRACTION parameter may depend on the value of an environmental variable. The MSE for Pacific sardine (Hurtado-Ferro and Punt 2014) did not simulate application of the actual stock assessment process, but instead generated estimates of biomass directly from the operating model. Nevertheless, the extent of the errors

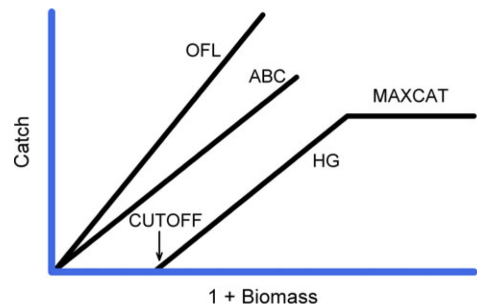

Figure 3 Harvest control rules applied to the northern subpopulation of Pacific sardine. The OFL is the overfishing level, which is based on the fishing mortality corresponding to maximum sustainable yield. The ABC is the acceptable biological catch, computed as the overfishing limit (OFL) reduced to account for scientific uncertainty. CUT-OFF determines the 1+ biomass at which the harvest guideline (HG) is zero, and MAXCAT is the maximum catch possible under the control rule.

associated with the biomass estimates for Pacific sardine was selected using a simulation evaluation of the actual stock assessment method (Hurtado-Ferro *et al.* 2014) in an attempt to ensure realism.

Simulation of application of each management strategy for each operating model

The actual process of linking the data generation phase of the operating model with the management strategy is generally straightforward, even if the process of conducting the simulations and summarizing the results can be very time-consuming. The difficult issues with MSE at this stage are primarily related to software development. There are several ways to minimize the chances of errors due to software coding, and use of these methods is best practice for MSE.

1. Base the operating model(s) and the management strategy on software that has been developed for broad application and has been tested extensively, such as Stock Synthesis (Methot and Wetzel 2013; Anderson *et al.* 2014; Maunder 2014), or use tools specifically developed to evaluate management strategies (Kell *et al.* 2007; Hillary 2009). However, in many instances, it is necessary to develop software for a specific case given the nature of the management strategy being evaluated and the hypotheses considered plausible – capturing the full range of uncertainty and of potentially

appropriate candidate management strategies should take priority over using available software.

2. Conduct simulations in which the system dynamics are deterministic, the operating model matches the estimation component of the management strategy, and the data are generated without error. In this situation, it should be possible for an analyst to heuristically predict the state of the system in the future fairly accurately (e.g. the stock should equilibrate at B_{MSY} if a strategy is based on a target fishing mortality of F_{MSY}, while if a strategy has a target level based on CPUE the stock should equilibrate at this level unless there are response-delay factors that induce oscillations) and compare this with where the MSE predicts the system will be. This provides a basic test to ensure that the coding of the operating model and of the management strategy is correct.
3. Conduct simulations in which the system dynamics are deterministic, the assessment model underlying the management strategy (if required) matches the operating model, and the data are generated with random error. Again, this provides a case where it is relatively straightforward to predict the results of the analyses.

The second and third of these steps also provide a way to eliminate poor management strategies from further consideration; it is virtually certain that a management strategy will not perform adequately in complex trials if it performs poorly when the data are not subject to error or there is no process error in the system.

The number of simulations for each trial (10 000 in the case of Pacific sardine and 100 in the case of the BCB bowheads) should be selected to ensure that the percentiles of the distributions on which performance statistics are based can be calculated with the precision required for the decisions to follow. The number to achieve a particular precision for probability-based statistics can be calculated taking into account that the simulations are independent (ICES 2013), and probability measures based on counts are therefore binomially distributed. Note that a very (perhaps prohibitively) large number of simulations may be needed if the decision-makers wish to draw conclusions based on very precise estimates of the lower fifth or first percentile of the distribution for some output from the operating model(s).

The number of years for which the operating model is projected will depend on the life history of the species under consideration. The number should be chosen, so that it is possible that the management strategy can impact the dynamics of the system and should cover 1–2 generations at minimum to allow for transients arising from response delays linked, for example, to the age at maturity. For example, the number of years for short-lived species such as sardine can be quite low while this number will be much higher for species such as bowhead whales.

It is essential, and hence best practice, that the management strategy bases recommendations for management actions only on data which would actually be available, and any assumptions regarding parameters assumed known when applying the management strategy need to be clearly documented (e.g. that natural mortality is assumed to be known exactly). One way to achieve this goal is to have separate segments of software for the operating model and for the management strategy, and to pass information (and management recommendations) between the operating model and management strategy using input and output files or their software equivalent.

The same set of random numbers should be used for all simulations for each trial, so that differences between candidate management strategies reflect the differences between the strategies themselves and not the consequences of different sets of observation and process errors.

Most management strategies assume that the data needed to apply them are always available (e.g. surveys are conducted at the expected frequency). However, this assumption might not be met in practice (e.g. a survey may not take place because of mechanical problems), and Butterworth (2008b) highlights that a management strategy should ideally also include specifications for how to provide management advice in circumstances in which anticipated data are not available. A related aspect is that the management strategy should ideally reward the provision of extra data and penalize the reverse situation. For example, the IWC's Revised Management Procedure reduces whale fishery catch limits to zero if new survey estimates do not become available within a specified time period (IWC 2012b).

Presentation of results and selection of a management strategy

Ultimately, the selection of a management strategy is not a scientific enterprise, but involves addressing trade-offs. This task lies primarily within the purview of decision-makers and policy. In principle, the selection of a management strategy could be automatic if a utility function was selected, which reflects the desired trade-offs amongst the objectives, and probabilities could be assigned to each alternative operating model configuration. However, this is rarely possible, and the authors know of no examples where a management strategy which has actually been implemented was selected this way.

There are almost always trade-offs among the management objectives. Consequently, it is desirable to provide results for a number of candidate strategies. Evaluation by the decision-makers of the trade-offs amongst the management objectives achieved by each candidate strategy may lead to a better understanding of what is possible, and even to changes to the relative ranking of management objectives. However, the results of management strategy simulations can be extensive and complicated, and the entire MSE process may be difficult for non-experts to comprehend. In South Africa, the details of the assumptions and sources of uncertainty were communicated, but statistics such as probability distributions were found hard to interpret (Cochrane *et al.* 1998). A better understanding of some of the trade-offs, particularly that between catch and catch variation, can be achieved by 'real-time gaming' of the MSE, which involves the decision-makers managing simulated populations where they are provided with the data which would actually be available on an annual basis. Walters (1994) provides an overview of the use of gaming to compare management options, including some best practices. Gaming has been used successfully in the South African fisheries (Butterworth *et al.* 1993). However, many MSE analyses are very computationally intensive, making real-time gaming impractical.

Stakeholders need to be involved in the decision process. However, more than that, they also need to be integrated within the entire MSE development process, including problem formulation, and even perhaps selecting the assumptions on which projections are based. This is, however, seldom easy and can be very time-consuming. Pastoors *et al.* (2007) describe an instance where stakeholders evaluated a MSE based on the extent to which hindcasts of the operating model could reproduce the observed dynamics of how TACs were set and whether the trends in stocks and catches proceeded 'logically'. Their advice was to present results relative to reference levels rather than in absolute terms, so as to reduce some of the concerns which stakeholders expressed.

As emphasized by Rademeyer *et al.* (2007), the basis for selecting a management strategy has to be clear to all stakeholders and should be made as simple as can be justified. Although much of the literature has focused on trade-offs among the objectives, some systems have fixed constraints. For example, the USA MSA effectively prohibits fishing mortality exceeding F_{MSY} for long periods, while adoption of a management strategy that would lead to high probabilities of decline of BCB bowhead whales would be considered unacceptable. Miller and Shelton (2010) identify an approach to selecting a management strategy based on 'satisficing', in which there are certain minimum standards for any candidate strategy, and only those candidates who satisfy these standards can be considered for possible adoption. Care should be taken not to require management strategies to meet performance statistic targets defined in terms of extreme tail probabilities, for example implementing a standard such as 'the probability of overfishing on an annual basis should not exceed 0.1%', because such probabilities are likely to be very poorly determined (Rochet and Rice 2009; Kraak *et al.* 2011). In cases in which the decision-makers require high certainty about a particular outcome, it is imperative that the analysts convey the likely level of precision possible from a MSE and that the major strength of a MSE lies in comparing the relative performance of alternative management strategies.

The first step in the process of selecting a management strategy should be explaining all of the options to the decision-makers, and placing the management strategies evaluated in the context of current management arrangements (Dowling *et al.* 2008). The value of effective graphical summaries cannot be over-emphasized. Some simple rules for constructing graphical summaries of results (see Figs 4 and 5 for examples) are to define 'best' performance for all operational objectives to be a high value for the associated performance statistics, and not to display too many performance statistics or management strategies on a single plot (contrast Figs 4 and 5 in this regard).

Perhaps most importantly, graphical approaches to summarizing performance statistics should be selected in collaboration with the decision-makers who need to understand and use them. For example, the axes in Fig. 5 were defined to report on the major areas of concern for stakeholders. 34 performance measures were identified by fishers, processors and local community, as well as given legislated fisheries and conservation objectives to across social, economic and ecological aspects (Fulton *et al.* 2014). For transparency, all of these measures were reported on, but it was not until the outcomes were aggregated and summarized around the major topic areas (using Fig 5 and other similar plots) that the relative performance and trade-offs between the objectives were clear. The axes represent natural groupings of the performance measures, but also highlight key concerns of the various stakeholders. Note that the industry and management efficiency axes used inverted performance scores, so that a larger score reflected better performance for all axes.

A key step in selecting a management strategy is dealing with the fact that not all of the trials reflect equally plausible hypotheses. This is partially addressed by assigning some trials to a reference set and the remaining trials to a robustness set (see above). However, other approaches are possible. For example, the IWC has adopted a set of guidelines for interpreting the results of trials to evaluate management strategies for commercial whaling. Specifically, trials are assigned to one of three categories ('high plausibility', 'medium plausibility' or 'low plausibility') by the Scientific Committee of the IWC (2012a). The required conservation performance of acceptable management strategies, expressed in terms of the values for performance statistics, is pre-specified for each

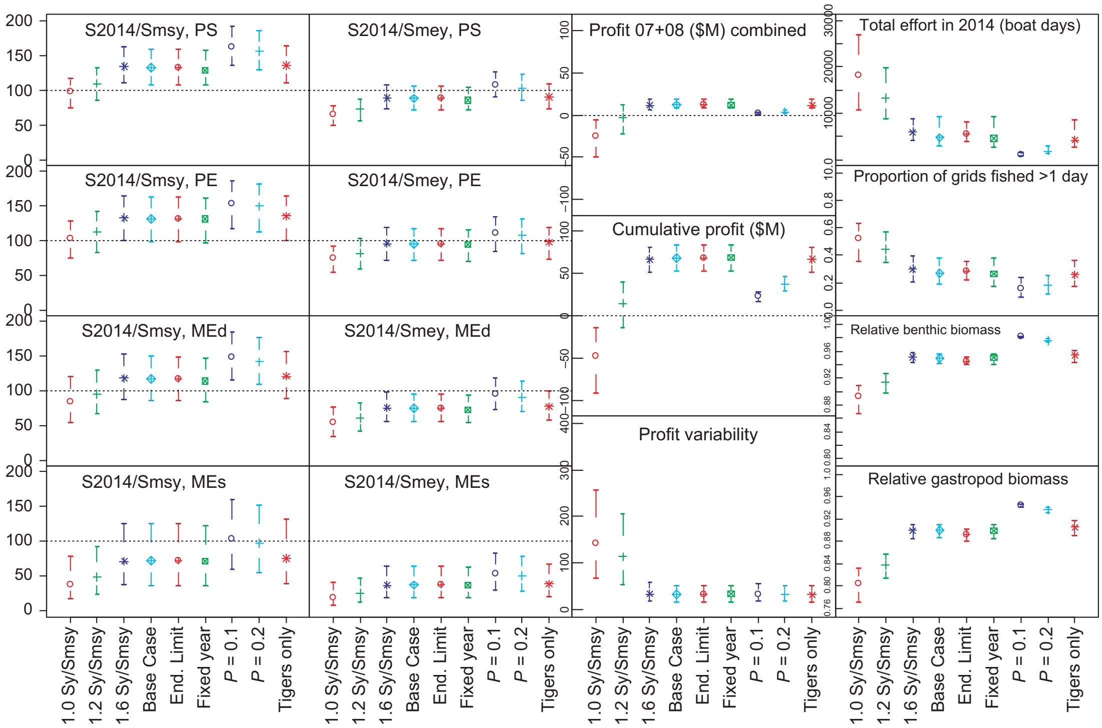

Figure 4 Biological, economic and ecosystem performance measures for a variety of management strategies for Australia's Northern Prawn Fishery (reproduced from Dichmont *et al.* 2008). The symbols indicate distribution medians, and the bars cover 95% of the simulation distributions. The performance statistics relate to spawning biomass relative to that at which MSY and maximum economic yield are achieved for four species (first two columns) and profit and its variability (third column). The right-most column shows the total effort in 2014, the proportion of grids fished for more than 1 day in 2014, the total benthic biomass relative to unfished levels, and the biomass of gastropods in 2014 relative to unfished levels. The management strategies differ in terms of the target biomass, the extent of precaution, and whether assessments for only two of the species form the basis for changes to effort limits.

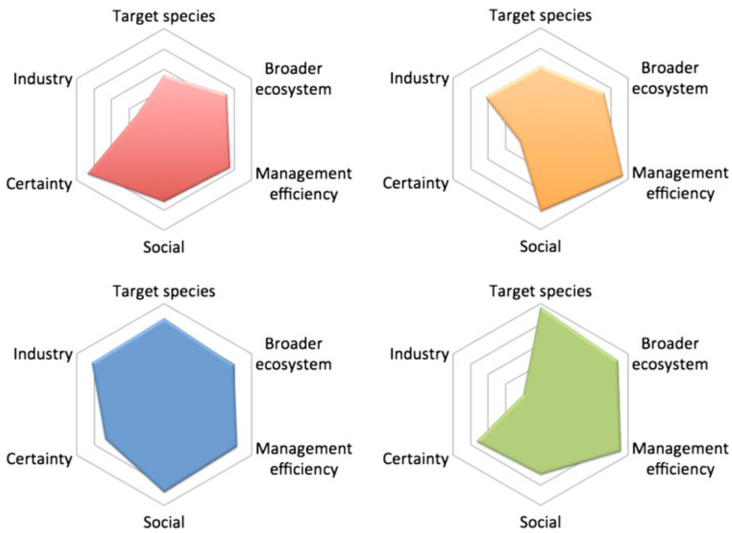

Figure 5 Example of plots which qualitatively compare four management strategies across six general areas of mean performance for a large multisector, multispecies fishery in southeastern Australia (E. Fulton, CSIRO, personal communication). A better result for a performance statistic is indicated by a vertex which is further from the centre of each hexagon.

category, which essentially (though not entirely – see IWC 2012a, for details) automates the process of selecting a 'best' management strategy. The assignment of plausibility for a trial is based on assigning a plausibility ranking to the level for each factor on which the trial is based ('high', 'medium', 'low' or 'no agreement'), with levels for which there is no agreement being treated as 'medium'. The ranking of a trial reflects the lowest rank assigned to each level of the factors on which it is based (thus to be categorized as a 'high' plausibility trial, the levels of all the factors included in the trial need to be considered to be of 'high' plausibility). Any trials considered to be 'low' plausibility are assigned a 'low' rank and ignored. This approach has been applied to select management strategies for the western North Pacific minke whales (IWC 2014), the western North Pacific Bryde's whales (IWC 2010) and the North Atlantic fin whales (IWC 2009).

In an effort to provide an improvement to simply selecting plausibility ranks based on expert judgement, Butterworth *et al.* (1996) proposed four sets of criteria with which plausibility ranks might be assessed:

1. how strong is the basis for the hypothesis in the data for the species or region under consideration?
2. how strong is the basis for the hypothesis in the data for a similar species or another region?
3. how strong is the basis for the hypothesis for any species? and
4. how strong or appropriate is the theoretical basis for the hypothesis?

Although this approach was presented to the Scientific Committee of the IWC, it was never adopted, and in general weights are almost always assigned based on expert judgement.

An alternative approach to addressing plausibility in selecting a management strategy is to assign weights to each trial and to compute integrated values for the performance statistics. However, this involves selection of quantitative weights upon which it is likely to be even more difficult to reach agreement than on assigning trials to categories of plausibility. Moreover, integrated performance statistics may obscure low plausibility trials for which performance is very poor (Rademeyer *et al.* 2007). Those authors also comment that stakeholders may benefit from being shown results of individual catch and population trajectories, as these tend to give a better impression of variation than statistics such as CVs and variances, which may be difficult for some stakeholders to understand.

Assignment of quantitative weights for plausibility becomes necessary if decision-makers wish to draw conclusions based on some percentile of the distribution of a performance statistic and the MSE is being conducted over a reference set of operating models. This was the case in the CCSBT, where the use of this set, rather than working only with a single reference case operation model, rendered consensus much more easily achieved in the Scientific Committee. Subsequently, the Commission requested its Scientific Committee to report results for reaching a target recovery level of 20% of the estimate unfished abundance by 2035 with 70% probability [see final agreed management strategy specifications reflected in CCSBT (2011)]. To provide such results, integration across the reference set became necessary.

While providing percentile results for a single operating model is a relatively objective process, as the statistical basis to take account of the associated stochastic effects is well established, extending to a reference set creates some difficulties. This is because the results will depend on the choice of which models are included in the set and how they are weighted, which is much less straightforward. Given that balance (between more optimistic and more pessimistic scenarios) is usually seen as a desired feature of a reference set of operating models, estimates of the medians of performance statistics would be expected to remain relatively robust and reliable. However, care should be taken in the interpretation of high and low percentiles of distributions for a reference set, as these will not be as firmly established as in the case of a single reference case operating model.

In the BCB bowhead case, the Chair of the group tasked with developing and testing alternative SLAs briefed the IWC as well as representatives of the hunting communities. In particular, as a key objective of the SLA was to satisfy the nutritional and cultural needs of aboriginal communities rather than to maximize catch, an important input to the analyses was the ‘Need Envelope’. This function was obtained through discussion with the hunters and their scientific representatives, and formed the basis for specifying performance statistics such as the fraction of total need over 100 years which could be satisfied.

In contrast to the bowhead case-study, the MSE for Pacific sardine was developed in the context of a USA Regional Fishery Management Council process. This allows for input by stakeholders, state and federal analysts, and the public during the development of management decisions. The structure of the MSE was initially developed during a workshop (PFMC 2013) which included biologists familiar with Pacific sardine and its relationship with the environment, modellers (including assessment biologists and ecosystem modellers), representatives of the advisory bodies of the PFMC, and stakeholders (conservation and industry). The MSE structure was then subjected to peer review through the PFMC’s Scientific and Statistical Committee on several occasions. Input from stakeholder groups included interpretation of the results of the simulations in the context of the objectives which each such group considered most important (Parrish 2014).

Did the case-studies follow ‘best practice’?

The two case-studies highlighted in this paper followed best practice to different extents. Both case-studies involved stakeholders and decision-makers at various points in the development and selection process, and included default performance statistics. The range of uncertainties was wider in the bowhead case-study, and there are some uncertainties which are likely important for Pacific sardine which were not explored (such as that the USA fishery operates at some times on the southern as well as the northern subpopulation). Such omissions were due to limited time being made available to conduct the MSE. In actual development and implementation, limited time frames are common and constitute a constraint on achieving best practice.

Neither of the case-studies explicitly considered predator-prey interactions as these were not seen as likely to have large impacts; the sardine case-study did however explore environmental impacts on recruitment, and both case-studies accounted for spatial structure to some extent. The bowhead case-study represented parameter uncertainty by sampling parameter vectors from a posterior distribution, whereas the sardine case-study explored this uncertainty through sensitivity testing.

The candidate management strategies for Pacific sardine were selected by the stakeholders and the decision-makers, whereas these were identified by the competing teams of ‘developers’ in the bowhead case. In contrast to the bowhead SLA, the actual management strategy for sardine was not simulated exactly because it was not the assessment itself (which is based on a statistical catch-at-age analysis) that was simulated. Rather, this assessment was approximated by true biomass from the operating model plus autocorrelated log-normal error. However, an attempt was made to assess the likely level of assessment error.

Both case-studies applied standard programming techniques to attempt to ensure that the code implementing the operating model(s) and management strategies was correct, but only in the sardine case were deterministic analyses undertaken. The code implementing the operating models for the bowhead case was developed by a member of the staff of the IWC and independently checked by one of us (AEP). Neither case-study conducted a thorough comparison of whether the operating model and management strategy produced results of projections consistent with reality through, for example, comparing variability in assessment outcomes with historical results, although some checks were carried out for sardine. Neither of the management strategies adopted included 'Exceptional Circumstances' provisions, although *Implementation Reviews* are mandated and have been conducted for the bowheads. The SLA for the BCB bowheads was subject to an *Implementation Review* in 2007 (IWC 2008a,b, 2009, 2014) and 2013 (IWC 2013). The 2007 *Implementation Review* focused on the possibility that the BCB stock may actually consist of two stocks as well as that different age and sex classes migrate differently. However, it did not lead to a change to the SLA developed for the BCB stock because the performance of this SLA was not markedly impacted by the multi-stock scenarios examined.

Both case-studies relied on graphical and tabular summaries, and both involved trying to educate the decision-makers on how to interpret the results from the MSE. Performance standards were adopted for interpreting the results of the trials for bowheads (IWC 2003), but the comparison of alternatives for Pacific sardine was based primarily on finding an acceptable trade-off among the performance statistics. The trials for the bowhead case were divided into 'reference' and 'robustness' trials, with most focus during selection given to the 'reference' set.

In summary, the application of MSE for bowheads followed the proposed best practice guidelines to the largest extent possible, while that for sardine took several short cuts, owing primarily to the need to complete the analyses in time for management decision-making.

Final comments

Management strategy evaluation arose from the desires to deal more systematically with the issue of uncertainties and to identify management strategies that are adaptive given the collection of new data. Although the benefits of active adaptive management strategies, that is management strategies which select management actions to increase 'contrast' and hence improve the information content of the available data, have long been known (Walters 1986), few jurisdictions have been able or willing to implement such strategies (Sainsbury *et al.* 1997 being a noteworthy exception, although in that case the 'experimental unit' was primarily a foreign fishery off Australia's northwest shelf). Consequently, MSE has in practice generally involved evaluation of passive adaptive management options, that is learning about the system dynamics through ongoing monitoring but without attempting to deliberately manipulate the system to learn more about it, although the strategy developed for the mid-water fishery for horse mackerel in South Africa is an exception to this (Furman and Butterworth 2012).

Management strategy evaluation has been applied most widely in relation to fisheries and cetacean conservation and management. However, it has also been applied to explore the performance of ballast-water management options (Dunstan and Bax 2008), and recently there have been calls for MSE to be applied to terrestrial systems, including in the development of conservation plans for threatened species (Milner-Gulland *et al.* 2010; Bunnefeld *et al.* 2011; Moore *et al.* 2013). Most fisheries applications have focused on single-species cases. However, MSE can be applied to identify management strategies to achieve ecosystem and multiuse objectives. The applications in this area remain few, in particular because of the computational requirements associated with fitting and projecting models such as Atlantis. However, one would expect that the number of these applications will increase rapidly as computational constraints become less of an issue.

Management strategy evaluation has generally been used to evaluate management strategies in terms of their ability to satisfy management goals, either generically or for a specific situation, with a view to possible formal adoption and implementation. However, an additional key reason for conducting a MSE is to identify when management strategies are likely to fail, and either to identify new data collection schemes to detect when failure might occur or to revise an existing management strategy appropriately. Finally, evaluation of the management strategies on which a fishery is based is part of several eco-certification systems, including that of the Marine Stewardship Council (MSC). In the case of Tristan da Cunha rock lobster, the MSE was conducted specifically to satisfy one of the performance indicators for MSC certification.

Smith *et al.* (1999) outline the roles for the various participants in the MSE development process, including those of decision-makers, industry, conservation agencies and groups, fishery scientists and MSE analysts. As noted above, the involvement of as many of these groups as possible enhances the likelihood that the results of the MSE will be considered credible and hence the strategy actually implemented throughout the period for which it is intended to apply. Although inclusion of stakeholders in the development of management strategies is emphasized by Smith *et al.* (1999) and in many other publications, the actual number of MSEs for which there is direct evidence that stakeholders were involved throughout the entire process is rare. ICES (2013) outlines the roles of stakeholders (and decision-makers) in the MSE process as it is typically applied in Europe. The MSE developed for Australia's SESSF was guided by a steering committee of stakeholders from all sectors of the fishing industry, an ENGO, decision-makers and representatives of two key funding agencies (Smith *et al.* 2007). In South Africa, the process is taken forward in the species-specific scientific working groups of the Fisheries Branch of the Government Department responsible; these groups include observers from both industry and ENGOs who participate actively.

The establishment of a management strategy is a critical part of effective management. However, it is only one part. There still needs to be a formal process for reviewing the appropriateness of a management strategy given information collected following adoption. In Europe, apart from performing impact assessments of proposed management plans, the European Commission's advisory body, STECF, also evaluates the performance of existing management plans in relation to their original objectives (STECF 2011b; Kraak *et al.* 2013). In South Africa, reviews of management strategies are planned for every 4 years, while reviews of the CCSBT management strategy are planned for every 9 years, with the latter commonly adjusting TACs only every 3 years (Butterworth 2008b). The IWC has established a formal process for the regular (usually 5-year) review of the basis for specific management strategies, termed *Implementation Reviews* (IWC 2012a, 2013; Punt and Donovan 2007).

A management strategy is tested for the set of hypotheses considered plausible when it was first developed. However, subsequent research could indicate that those hypotheses did not include the entire plausible range. Consequently, rules have been developed (IWC 2013) for when it is necessary to temporarily stop applying the management strategy and rely on *ad hoc* adjustments to management regulations or to initiate an *Implementation Review* before one is due. The management strategies for South African fish stocks include some formal 'Exceptional Circumstances' provisions (Butterworth 2008b), as do those for southern bluefin tuna, west Greenland halibut and east Canadian pollock, but most other management strategies do not. 'Exceptional Circumstances' are generally defined to apply when the future data fall outside of the range indicated for the projections considered in the MSE. The inclusion of such provisions should be considered a standard component of best practice.

We have identified 'best practices' for conducting MSE (Table 1). The 'best practices' should be followed as closely as possible, particularly when the intent is to use the MSE to develop a management strategy for a particular fishery. However, as we illustrate for the two case-studies, a MSE can be useful even if not all of the best practices are followed strictly. This is particularly the case when the aim of the MSE is to evaluate generic management strategies rather than to propose a management strategy for implementation to a specific stock. Most critical perhaps is that the primary aim of a MSE is to identify which uncertainties are most important in terms of achieving management objectives. What is the minimum that can be done for the process still to be considered as a MSE? We would propose this to be that a MSE *considers* all sources of influential uncertainty, even if they are not all represented in the operating models, *considers* all the management objectives, even if they cannot all be reflected in the operating models, and minimally allows for uncertainty in the information on which management advice is based.

Finally, the practice of MSE continues to develop, and so, just as management strategies should be adapted under changing circumstances, MSE best practice is expected to continue to become further articulated as more experience is gained.

Acknowledgements

This paper benefited from comments by Tony Smith, Eva Plaganyi and Geoff Tuck (CSIRO Wealth from Oceans Flagship), Kelli Johnson (UW), two anonymous reviewers and the editor. This publication is [partially] funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement No. NA08OAR4320899 Contribution No. 2370.

References

A'mar, Z.T., Punt, A.E. and Dorn, M.W. (2008) The management strategy evaluation approach and the fishery for walleye pollock in the Gulf of Alaska. In: *Resiliency of Gadid Stocks to Fishing and Climate Change* (eds G.H. Kruse, K. Drinkwater, J.N. Ianelli, J.S. Link, D.L. Stram, V. Wespestad and D. Woodby). Alaska Sea Grant College Program, University of Alaska, Fairbanks, pp. 317–346.

A'mar, Z.T., Punt, A.E. and Dorn, M.W. (2009a) The evaluation of management strategies for the Gulf of Alaska walleye pollock under climate change. *ICES Journal of Marine Science* **66**, 1614–1632.

A'mar, Z.T., Punt, A.E. and Dorn, M.W. (2009b) The impact of regime shifts on the performance of management strategies for the Gulf of Alaska walleye pollock fishery. *Canadian Journal of Fisheries and Aquatic Sciences* **66**, 2222–2242.

A'mar, Z.T., Punt, A.E. and Dorn, M.W. (2010) Incorporating ecosystem forcing through predation into a management strategy evaluation for the Gulf of Alaska walleye Pollock (*Theragra chalcogramma*) fishery of ecosystem changes on the management of the fishery for walleye pollock in the Gulf of Alaska. *Fisheries Research* **102**, 98–114.

Anderson, B.S., Vermard, Y., Ulrich, C., Hutton, T. and Poos, J.J. (2010) Challenges in integrating short-term behaviour in a mixed-fishery Management Strategies Evaluation frame: a case study of the North Sea flatfish fishery. *Fisheries Research* **102**, 26–40.

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

Anonymous. (2011) Report of the 16th meeting of the Extended Scientific Committee for the Commission for the Conservation of Southern Bluefin Tuna. CCSBT.

Babcock, E.A. and MacCall, A.D. (2011) How useful is the ratio of fish density outside versus inside no-take marine reserves as a metric for fishery management control rules? *Canadian Journal of Fisheries and Aquatic Sciences* **68**, 343–359.

Baumgartner, T., Soutar, A. and Ferriera-Bartrina, V. (1992) Reconstruction of the history of pacific sardine and northern anchovy populations over the past two millennia from sediments of the Santa Barbara Basin, California. *CalCOFI Reports* **33**, 24–40.

Bence, J.R., Gordoa, A. and Hightower, J.E. (1993) Influence of age-selective surveys on the reliability of stock synthesis assessments. *Canadian Journal of Fisheries and Aquatic Sciences* **50**, 827–840.

Bentley, N. and Stokes, K. (2009a) Moving fisheries from data-poor to data-sufficient: evaluating the costs of management versus the benefits of management. *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science* **1**, 378–390.

Bentley, N. and Stokes, K. (2009b) Contrasting paradigms for fisheries management decision making: how well do they serve data-poor fisheries? *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science* **1**, 391–401.

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

Breen, P.A. and Kim, S.W. (2006) Development of an operational management procedure (decision rule) for CRA 4. New Zealand Fisheries Assessment Report 2006/53, 46 pp.

Bunnefeld, N., Hoshino, E. and Milner-Gulland, E.J. (2011) Management strategy evaluation: a powerful tool for conservation? *Trends in Ecology and Evolution* **26**, 441–447.

Butterworth, D.S. (2007) Why a management procedure approach? Some positives and negatives. *ICES Journal of Marine Science* **64**, 613–617.

Butterworth, D.S. (2008a) A commentary on: salvaged pearls: lessons learned from a floundering attempt to develop a management procedure for Southern Bluefin Tuna. *Fisheries Research* **94**, 351–354.

Butterworth, D.S. (2008b) Some lessons from implementing management procedures. In: *Fisheries for Global Welfare and Environment, 5th World Fisheries Congress 2008* (eds K. Tsukamoto, T. Kawamura, T. Takeuchi, T.D. Beard Jr and M.J. Kaiser). TERRAPUB, Tokyo, pp. 381–397.

Butterworth, D.S. and Geromont, H.F. (2001) Evaluation of a class of possible simple interim management procedures for the Namibian hake fishery. *South African Journal of Marine Science* **23**, 357–374.

Butterworth, D.S. and Johnston, S.J. (2014) Recommendations on rock lobster TACs for the Tristan outer group of islands for the 2014/15 season. MARAM Document: MARAM/2014/JUL/10, 6 pp. Available at: http://www.mth.uct.ac.za/maram/pub/2014/MARAM_TRISTAN_2014_FEB_03.pdf (accessed 10 November 2014).

Butterworth, D.S. and Punt, A.E. (1999) Experiences in the evaluation and implementation of management procedures. *ICES Journal of Marine Science* **56**, 985–998.

Butterworth, D.S. and Rademeyer, R.A. (2010) Greenland halibut MSE results for updated SCAA reference case and robustness test operating models. NAFO Document: FCWGMSE WP 10/13, 18 pp. Available at: http://www.mth.uct.ac.za/maram/pub/2010/wgmsewp10-13_MSE_update.pdf (accessed 26 June 2014).

Butterworth, D.S., De Oliveira, J.A.A. and Cochrane, K.L. (1993) Current initiatives in refining the management procedure for the South African anchovy resource. In: *Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations (Anchorage, Alaska, October 1992)* (eds G. Kruse, D.M. Eggers, R.J. Marasco, C. Pautzke and T.J. Quinn II). Alaska Sea Grant College Program AK-SG-93-02, Fairbanks, AK, pp. 439–473.

Butterworth, D.S., Punt, A.E. and Smith, A.D.M. (1996) On plausible hypotheses and their weighting, with implications for selection between variants of the Revised Management Procedure. *Reports of the International Whaling Commission* **46**, 637–640.

Butterworth, D.S., Johnston, S.J. and Brandáo, A. (2010a) Pretesting the likely efficacy of suggested approach to data-poor fisheries. *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science* **2**, 131–145.

Butterworth, D.S., Bentley, N., De Oliveira, J.A.A. *et al.* (2010b) Purported flaws in management strategy evaluation: basic problems or misinterpretations? *ICES Journal of Marine Science* **67**, 567–574.

Butterworth, D.S., De Oliveira, J.A.A. and de Moor, C.L. (2012) Are South African administrative law procedures adequate for the evaluation of issues resting on scientific analyses. *South African Law Journal* **129**, 461–477.

Carruthers, T.R., McAllister, M.K. and Taylor, N. (2011) Spatial surplus production modeling of Atlantic tunas and billfish. *Ecological Applications* **21**, 2734–2755.

Carruthers, T.R., Punt, A.E., Walters, C.J. *et al.* (2014) Evaluating methods for setting catch limits in data-limited fisheries. *Fisheries Research* **153**, 48–68.

Christensen, S. (1997) Evaluation of management strategies – a bioeconomic approach applied to the Greenland Shrimp Fishery. *ICES Journal of Marine Science* **54**, 412–426.

Cochrane, K.L., Butterworth, D.S., De Oliveira, J.A.A. and Roel, B.A. (1998) Management procedures in a Fishery based on highly variable stocks and with conflicting objectives: experiences in the South African pelagic fishery. *Reviews in Fish Biology and Fisheries* **8**, 177–214.

Commission for the Conservation of Southern Bluefin Tuna (CCSBT) (2011) Report of the Eighteenth Annual Meeting of the Commission. Available at: http://www.ccsbt.org/userfiles/file/docs_english/meetings/meeting_reports/ccsbt_20/report_of_SC18.pdf (accessed 11 November 2014).

Cooke, J.G. (1999) Improvement of fishery-management advice through simulation testing of harvest algorithms. *ICES Journal of Marine Science* **56**, 797–810.

Cox, S.P. and Kronlund, A.R. (2008) Practical stakeholder-driven harvest policies for groundfish fisheries in British Columbia, Canada. *Fisheries Research* **94**, 224–237.

De Oliveira, J.A.A. (2003) *The development and implementation of a joint management procedure for the South African pilchard and anchovy resources*. PhD thesis, University of Cape Town, [iv] + 319 pp.

De Oliveira, J.A.A. and Butterworth, D.S. (2004) Developing and refining a joint management procedure for the multispecies South African pelagic fishery. *ICES Journal of Marine Science* **61**, 1432–1442.

Dereksdóttir, E.H. and Magnússon, K.G. (2003) A Strike Limit Algorithm based on Adaptive Kalman Filtering with an application to aboriginal whaling of bowhead whales. *Journal of Cetacean Research and Management* **5**, 29–37.

Dichmont, C.M. and Brown, I.W. (2010) A case study in successful management of a data-poor fishery using simple decision rules: the Queensland spanner crab fishery. *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science* **2**, 1–13.

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

Dichmont, C.M., Deng, R., Punt, A.E., Venables, W. and Haddon, M. (2006b) Management strategies for short lived species: the case of Australia's Northern Prawn Fishery. 3. Management strategy scale and complexity. *Fisheries Research* **82**, 235–245.

Dichmont, C.M., Deng, A., Punt, A.E. *et al.* (2008) Beyond biological performance measures in Management Strategy Evaluation: bringing in economics and the effects of trawling on the benthos. *Fisheries Research* **94**, 238–250.

Dichmont, C.M., Ellis, N., Bustamante, R.H. *et al.* (2013) Evaluating marine spatial closures with conflicting fisheries and conservation objectives. *Journal of Applied Ecology* **50**, 1060–1070.

DiNardo, G.T. and Wetherall, J.A. (1999) Accounting for uncertainty in the development of harvest strategies for the Northwestern Hawaiian Islands lobster trap fishery. *ICES Journal of Marine Science* **56**, 943–951.

Dowling, N.A., Smith, D.C., Knuckey, I. *et al.* (2008) Developing harvest strategies for low-value and data-poor fisheries: case studies from three Australian fisheries. *Fisheries Research* **94**, 380–390.

Dunstan, P.K. and Bax, N.J. (2008) Management of an invasive marine species: defining and testing the effectiveness of ballast-water management options using management strategy evaluation. *ICES Journal of Marine Science* **66**, 841–850.

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

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

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

Fulton, E.A., Smith, A.D.M. and Punt, A.E. (2004) Which ecological indicators can robustly detect effects of fishing? *ICES Journal of Marine Science* **62**, 540–551.

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

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

Fulton, E.A., Link, J.S., Kaplan, I.C. *et al.* (2011b) Lessons in modelling and management of marine ecosystems: the Atlantis experience. *Fish and Fisheries* **12**, 171–188.

Fulton, E.A., Smith, A.D.M., Smith, D.C. and Johnson, P. (2014) An integrated approach is needed for Ecosystem Based Fisheries Management: insights from ecosystem-level management strategy evaluation. *PLoS One* **9**, e84242.

Furman, L.B. and Butterworth, D.S. (2012) Proposed Catch Control Rules for Horse Mackerel Bycatch in the Small Pelagics fishery and Directed Catch in the Midwater Trawl Fishery. Department of Agriculture, Forestry and Fisheries Report No. FISHERIES/2012/OCT/SWG-DEM/23, 33 pp. Available at: http://www.mth.uct.ac.za/maram/pub/2012/FISHERIES_2012_OCT_SWG-DEM_23_Horse_mackerel_juvenile_PUCL.pdf (accessed 11 November 2014).

Gaichas, S., Gamble, R., Fogarty, M. *et al.* (2012) Assembly rules for aggregate-species production models: simulations in support of management strategy evaluation. *Marine Ecology Progress Series* **459**, 275–292.

George, J.C., Bada, J., Zeh, J. *et al.* (1999) Age and growth estimates of bowhead whales (*Balaena mysticetus*) via aspartic acid racemisation. *Canadian Journal of Zoology* **77**, 571–580.

Geromont, H.F. and Butterworth, D.S. (In press-a) Generic management procedures for data-poor fisheries: forecasting with few data. *ICES Journal of Marine Science*.

Geromont, H.F. and Butterworth, D.S. (In press-b) Complex assessments or simple management procedures for efficient fisheries management: a comparative study. *ICES Journal of Marine Science*.

Geromont, H.F., De Oliveira, J.A.A., Johnston, S.J. and Cunningham, C.L. (1999) Development and application of management procedures for fisheries in southern Africa. *ICES Journal of Marine Science* **56**, 952–966.

Givens, G.H. (2003) Empirical estimation of safe aboriginal whaling limits for bowhead whales. *Journal of Cetacean Research and Management* **5**, 39–43.

Givens, G.H., Zeh, J.E. and Raftery, A.E. (1995) Assessment of the Bering–Chukchi–Beaufort Seas stock of bowhead whales using the Baleen II Model in a Bayesian synthesis framework. *Reports of the International Whaling Commission* **45**, 345–364.

Haddon, M. (2011a) *Modelling and Quantitative Methods in Fisheries*, 2nd edn. Chapman & Hall, CRC Press, 449 pp.

Haddon, M. (2011b) Management Strategy Evaluation testing of the Management Strategies used with South-Eastern Scallop Fisheries. Report prepared for the Department of Agriculture, Fisheries and Forestry (DAFF) as part of the Reducing Uncertainty in Stock Status Project, 100 pp.

Haddon, M. and Helidoniotis, F. (2013) Legal minimum lengths and the management of abalone fisheries. *Journal of Shellfish Research* **32**, 197–208.

Hilborn, R., Parma, A. and Maunder, M. (2002) Exploitation rate reference points for West Coast rockfish: are they robust and are there better alternatives? *North American Journal of Fisheries Management* **22**, 365–375.

Hill, K.T., Crone, P.R., Lo, N.C.H. *et al.* (2011) Assessment of the Pacific sardine resource in 2011 for U.S. management in 2012. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-SWFSC-487, 260 pp.

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

Howell, D., Filin, A.A., Bogstad, B. and Stiansen, J.E. (2013) Unquantifiable uncertainty in projecting stock response to climate change: example from North East Arctic cod. *Marine Biology Research* **9**, 920–931.

Hurtado-Ferro, F. and Punt, A.E. (2014) *Revised Analyses Related to Pacific Sardine Harvest Parameters*. Pacific Fishery Management Council, Portland, OR.

Hurtado-Ferro, F., Punt, A.E. and Hill, K.T. (2014) Use of multiple selectivity patterns as a proxy for spatial structure. *Fisheries Research* **158**, 102–115.

Ianelli, J.N., Hollowed, A.B., Haynie, A.C., Mueter, F.J. and Bond, N.A. (2011) Evaluating management strategies for eastern Bering Sea walleye pollock (*Theragra chalcogramma*) in a changing environment. *ICES Journal of Marine Science* **68**, 1297–1304.

ICES. (2008) Report of the Working Group on Methods of Fish Stock Assessments (WGMG), 7–16 October 2008, Woods Hole, USA. ICES CM 2008/RMC:03, 147 pp.

ICES. (2012) Report of the ICES Advisory Committee 2012. ICES Advice, 2012. Book 1, 156 pp.

ICES. (2013) Report of the Workshop on Guidelines for Management Strategy Evaluations (WKGMSE), 21–23 January 2013, ICES HQ, Copenhagen, Denmark.

International Whaling Commission (IWC) (1992) Report of the Third Comprehensive Assessment on Management Procedures. *Reports of the International Whaling Commission* **42**, 271–304.

International Whaling Commission (IWC) (1994) Report of the Scientific Committee, Annex H. The Revised Management Procedure (RMP) for Baleen Whales. *Reports of the International Whaling Commission* **44**, 145–152.

International Whaling Commission (IWC) (2003) Report of the Scientific Committee, Annex E. Report of the Standing Working Group (SWG) on the Development of an Aboriginal Subsistence Whaling Management Procedure (AWMP). *Journal of Cetacean Research and Management* **5**(Suppl), 154–225.

International Whaling Commission (IWC) (2008a) Report of the Scientific Committee, Annex E. Report of the Standing Working Group (SWG) on the Aboriginal Subsistence Whaling Management Procedure (AWMP). *Journal of Cetacean Research and Management* **10**(Suppl), 121–149.

International Whaling Commission (IWC) (2008b) Report of the Scientific Committee, Annex D. Report of the Sub-committee on the Revised Management Procedure. *Journal of Cetacean Research and Management* **10** (Suppl), 84–92.

International Whaling Commission (IWC) (2009) Report of the Scientific Committee, Annex D. Report of the Sub-committee on the Revised Management Procedure. *Journal of Cetacean Research and Management* **11**(Suppl), 91–144.

International Whaling Commission (IWC) (2010) Report of the Scientific Committee, Annex D. Report of the Sub-committee on the Revised Management Procedure. *Journal of Cetacean Research and Management* **12**(Suppl), 89–116.

International Whaling Commission (IWC) (2012a) Requirements and Guidelines for Implementations under the Revised Management Procedure (RMP). *Journal of Cetacean Research and Management* **13**(Supplement), 497–505.

International Whaling Commission (IWC) (2012b) The Revised Management Procedure (RMP) for Baleen Whales. *Journal of Cetacean Research and Management* **13**(Supplement), 485–494.

International Whaling Commission (IWC) (2013) Report of the Scientific Committee, Annex E. Report of the Standing Working Group (SWG) on the Aboriginal Subsistence Whaling Management Procedure (AWMP). *Journal of Cetacean Research and Management* **14**(Suppl), 137–171.

International Whaling Commission (IWC) (2014) Report of the Scientific Committee, Annex D1. Report of the Working Group on the *Implementation Review* for Western North Pacific Common Minke Whales. *Journal of Cetacean Research and Management* **15**(Suppl), 112–188.

Johnston, S.J. and Butterworth, D.S. (2000) Examination of a potential Strike Limit Algorithm for the Bering–Chukchi–Beaufort Seas stock of bowhead whales. Paper SC/52/AWMP6 presented to the IWC Scientific Committee, June 2000, in Adelaide, Australia (unpublished).

Johnston, S.J. and Butterworth, D.S. (2005) Evolution of operational management procedures for the South African West Coast rock lobster *(Jasus lalandii)* fishery. *New Zealand Journal of Marine and Freshwater Research* **39**, 687–702.

Johnston, S.J. and Butterworth, D.S. (2013). An operational management procedure for the Tristan da Cunha rock lobster fishery. Document MARAM/TRISTAN/2013/OCT/14, 11 pp. Available at: http://www.mth.uct.ac.za/maram/pub/2013/MARAM_Tristan_2013_OCT_14.pdf (accessed 10 November 2014).

Johnston, S.J. and Butterworth, D.S. (2014). Initial OMP candidates for the Inaccessible and Gough island fisheries. Document MARAM/TRISTAN/2014/FEB/03, 33 pp. Available at: http://www.mth.uct.ac.za/maram/pub/2014/FISHERIES_2014_SEP_SWG-SCRL_07.pdf (accessed 10 November 2014).

Johnston, S.J., Butterworth, D.S. and Glazer, J.P. (2008) OMP 2008 for the South Coast Rock Lobster Resource. Marine and Coastal Management Report No. MCM/2008/AUG/SWG-SCRL/30, 8 pp. Available at: http://www.mth.uct.ac.za/maram/pub/2008/MCM_2008_AUG_SWG_SCRL_30.pdf (accessed 10 November 2014).

Johnston, S.J., Butterworth, D.S. and Glazer, J.P. (2014) South Coast Rock Lobster OMP 2014: Initial Specifications. Department of Agriculture, Forestry and Fisheries Report No. FISHERIES/2014/SEP/SWG-SCRL/07, 14 pp.

Kai, M. and Shirakihara, K. (2005) A feedback management procedure based on controlling the size of marine protected areas. *Fisheries Science* **71**, 56–62.

Kell, L.T. and Fromentin, J.-M. (2009) A management strategy evaluation framework for Mediterranean Atlantic bluefin tuna. *Collective Volume of the Scientific Papers – ICCAT* **64**, 633–646.

Kell, L.T. and Fromentin, J.-M. (2010) Evaluation of the robustness of maximum sustainable yield based management strategies to variation in carrying capacity or migration pattern of Atlantic bluefin tuna (*Thunnus thynnus*). *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 837–847.

Kell, L.T., Pastoors, M.A., Scott, R.D. *et al.* (2005a) Evaluation of multiple management objectives for Northeast Atlantic flatfish stocks: sustainability vs. stability of yield. *ICES Journal of Marine Science* **62**, 1104–1117.

Kell, L.T., Pilling, G.M., Kirkwood, G.P. *et al.* (2005b) An evaluation of the implicit management procedure used for some ICES roundfish stocks. *ICES Journal of Marine Science* **62**, 750–759.

Kell, L.T., Pilling, G.M., Kirkwood, G.P. *et al.* (2006) An evaluation of multi-annual management strategies for ICES roundfish stocks. *ICES Journal of Marine Science* **63**, 12–24.

Kell, L.T., Mosqueira, I., Grosjean, P. *et al.* (2007) FLR: an open-source framework for the evaluation and development of management strategies. *ICES Journal of Marine Science* **64**, 640–646.

Kolody, D., Polacheck, T., Basson, M. and Davies, C. (2008) Salvaged pearls: lessons learned from a floundering attempt to develop a management procedure for Southern Bluefin Tuna. *Fisheries Research* **94**, 339–350.

Kraak, S.B.M., Buisman, F.C., Dickey-Collas, M. *et al.* (2008) The effect of management choices on the sustainability and economic performance of a mixed fishery: a simulation study. *ICES Journal of Marine Science* **65**, 697–712.

Kraak, S.B.M., Kelly, C.J., Codling, E.A. and Rogan, E. (2011) On scientists' discomfort in fisheries advisory science: the example of simulation-based fisheries management-strategy evaluations. *Fish and Fisheries* **11**, 119–132.

Kraak, S.B.M., Bailey, N., Cardinale, M. *et al.* (2013) Lessons for fisheries management from the EU cod recovery plan. *Marine Policy* **37**, 200–213.

Kurota, H., Hiramatsu, K., Takahashi, N., Shono, H., Itoh, T. and Tsuji, S. (2010) Developing a management procedure robust to uncertainty for southern bluefin tuna: a somewhat frustrating struggle to bridge the gap between ideals and reality. *Population Ecology* **52**, 359–372.

Little, L.R., Wayte, S.E., Tuck, G.N. *et al.* (2011) Development and evaluation of a cpue-based harvest control rule for the southern and eastern scalefish and shark fishery of Australia. *ICES Journal of Marine Science* **68**, 1699–1705.

Mapstone, B.D., Little, L.R., Punt, A.E. *et al.* (2008) Management strategy evaluation for line fishing in the Great Barrier Reef: balancing conservation and multi-sector fishery objectives. *Fisheries Research* **94**, 315–329.

Marasco, R.J., Goodman, D., Grimes, C.B., Lawson, P.W., Punt, A.E. and Quinn, T.J. II (2007) Ecosystem Based Fisheries Management: some practical suggestions. *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 928–939.

Maunder, M.N. (2014) Management Strategy Evaluation (MSE) implementation in Stock Synthesis: Application to Pacific bluefin tuna. IATTC Document SAC-05-10b. Available at: http://www.iattc.org/Meetings/Meetings2014/MAYSAC/PDFs/SAC-05-10b-Management-Strategy-Evaluation.pdf (accessed 11 November 2014).

McClatchie, S., Goericke, R., Auad, G. and Hill, K. (2010) Re-assessment of the stock-recruit and temperature-recruit relationships for Pacific sardine (*Sardinops sagax*). *Canadian Journal of Fisheries and Aquatic Sciences* **67**, 1782–1790.

McGilliard, C.R., Hilborn, R., MacCall, A., Punt, A.E. and Field, J. (2010) Can information from marine Protected Area be used to inform control rule-based management of small-scale, data-poor stocks? *ICES Journal of Marine Science* **68**, 201–211.

Merritt, M.F. and Quinn, T.J. II (2000) Using perceptions of data accuracy and empirical weighting of information: assessment of a recreational fish population. *Canadian Journal of Fisheries and Aquatic Sciences* **57**, 1459–1469.

Methot, R.D. and Wetzel, C.R. (2013) Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research* **142**, 86–89.

Miller, D.C. and Shelton, P.A. (2010) "Satisficing" and trade-offs: evaluating rebuilding strategies for Greenland halibut off the east coast of Canada. *ICES Journal of Marine Science* **67**, 1896–1902.

Milner-Gulland, E.J., Arroyo, B., Bellard, C. *et al.* (2010) New directions in management strategy evaluation through cross-fertilization between fisheries science and terrestrial conservation. *Biology Letters* **6**, 719–722.

de Moor, C.L., Butterworth, D.S. and De Oliveira, J.A.A. (2011) Is the management procedure approach equipped to handle short-lived pelagic species with their boom and bust dynamics? The case of the South African fishery for sardine and anchovy. *ICES Journal of Marine Science* **68**, 2075–2085.

Moore, J.E., Curtis, K.A., Lewison, R.L. *et al.* (2013) Evaluating sustainability of fisheries bycatch mortality for marine megafauna: a review of conservation reference points for data-limited populations. *Environmental Conservation* **40**, 329–344.

Mori, M. and Butterworth, D.S. (2006) A first step towards modelling the krill-predator dynamics of the Antarctic ecosystem. *CCAMLR Science* **13**, 217–277.

NAFO (2010) Report of the FC Working Group on Greenland Halibut Management Strategy Evaluation (WGMSE), 16–17 September 2010, Halifax, Nova Scotia, Canada. FC Doc 10/30, 60 pp. Available at: http://www.nafo.int/publications/frames/publications.html (accessed 26 June 2014).

Parrish, R. (2014) *Comments on the Sardine Re-Analysis*. Pacific Fishery Management Council, Portland, OR.

Pastoors, M.A., Poos, J.J., Kraak, S.B.M. and Machiels, M.A.M. (2007) Validating management simulation models and implications for communicating results to stakeholders. *ICES Journal of Marine Science* **64**, 818–824.

Peterman, R.M. (2004) Possible solutions to some challenges facing fisheries scientists and managers. *ICES Journal of Marine Science* **61**, 1331–1343.

PFMC (1998) *Amendment 8 (to the Northern Anchovy Fishery Management Plan) Incorporating a Name Change To: The Coastal Pelagic Species Fishery Management Plan*. Pacific Fishery Management Council, Portland, OR. Appendix B. Available at: http://www.pcouncil.org/wp-content/uploads/a8apdxb.pdf (accessed 26 June 2014).

PFMC (2011) *Coastal Pelagics Species Fishery Management Plan as Amended Through Amendment 13*. Pacific Fishery Management Council, Portland, OR. Available at: http://www.pcouncil.org/wp-content/uploads/CPS_FMP_as_Amended_thru_A13_current.pdf (accessed 26 June 2014).

PFMC (2013) *Report of the Pacific Sardine Harvest Parameters Workshop*. Pacific Fishery Management Council, Portland, OR. Available at: http://www.pcouncil.org/wp-content/uploads/I1b_ATT1_SARDINE_WKSHP_RPT_APR2013BB.pdf (accessed 26 June 2014).

Plagányi, É.E. (2007) Models for an ecosystem approach to fisheries. *FAO Fisheries Technical Paper* **477**, 108.

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

Plagányi, É.E., Punt, A.E., Hillary, R. *et al.* (2014) Models of intermediate complexity for ecosystem assessment to support tactical management decisions in fisheries and conservation. *Fish and Fisheries* **15**, 1–22.

Plagányi, É.E., van Putten, I., Hutton, T. *et al.* (In press) Integrating indigenous livelihood and lifestyle objectives in managing a natural resource. *Proceedings of the National Academy of Sciences*.

Pomarede, M., Hillary, R., Ibaibarriaga, L., Bogaards, J. and Apostolaki, P. (2010) Evaluating the performance of survey-based operational management procedures. *Aquatic Living Resources* **23**, 77–94.

Punt, A.E. (2011) The impact of climate change on the performance of rebuilding strategies for overfished groundfish species of the U.S. west coast. *Fisheries Research* **109**, 320–329.

Punt, A.E. and Butterworth, D.S. (1991) On an approach for comparing the implications of alternative fish stock assessments, with application to the stock of Cape Hake, Merluccius spp. off Northern Namibia. *South African Journal of Marine Science* **10**, 219–240.

Punt, A.E. and Butterworth, D.S. (1995) The effects of future consumption by the Cape fur seal on catches and catch rates of the Cape hakes. 4. Modelling the biological interaction between Cape fur seals *Arctocephalus pusillus pusillus* and Cape hakes *Merluccius capensis* and *M. paradoxus*. *South African Journal of Marine Science* **16**, 255–285.

Punt, A.E. and Donovan, G. (2007) Developing management procedures that are robust to uncertainty: lessons from the International Whaling Commission. *ICES Journal of Marine Science* **64**, 603–612.

Punt, A.E. and Hobday, D. (2009) Management strategy evaluation for rock lobster, *Jasus edwardsii*, off Victoria, Australia: accounting for uncertainty in stock structure. *New Zealand Journal of Marine and Freshwater Research* **43**, 485–509.

Punt, A.E. and Ralston, S.V. (2007) A Management Strategy Evaluation of rebuilding revision rules for overfished rockfish species. In: *Biology, Assessment and Management of North Pacific Rockfishes* (eds J. Heifetz, J. DiCosimo, A.J. Gharrett, M.S. Love, V.M. O'Connell and R.D. Stankey). Alaska Sea Grant College Program, University of Alaska Fairbanks, Anchorage, pp. 329–351.

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

Punt, A.E., Pribac, F., Taylor, B.L. and Walker, T.I. (2005) Harvest strategy evaluation for school and gummy shark. *Journal of Northwest Atlantic Fisheries Science* **35**, 387–406.

Punt, A.E., McGarvey, R., Linnane, A., Phillips, J., Triantafillos, L. and Feenstra, J. (2012a) Evaluating empirical decision rules for southern rock lobster fisheries: a South Australian example. *Fisheries Research* **115–116**, 60–71.

Punt, A.E., Siddeek, M.S.M., Garber-Yonts, B. *et al.* (2012b) Evaluating the impact of buffers to account for scientific uncertainty when setting TACs: application to red king crab in Bristol Bay, Alaska. *ICES Journal of Marine Science* **69**, 624–634.

Punt, A.E., Trinnie, F., Walker, T.I. *et al.* (2013) The performance of a management procedure for rock lobsters, *Jasus edwardsii*, off western Victoria, Australia in the face on non-stationary dynamics. *Fisheries Research* **137**, 116–128.

Punt, A.E., A'mar, T., Bond, N.A. *et al.* (2014) Fisheries management under climate and environmental uncertainty: control rules and performance simulation. *ICES Journal of Marine Science* **71**, 2208–2220.

Rademeyer, R.A. and Butterworth, D.S. (2011) Technical details underlying the Management Strategy Evaluation process leading to selection of a management procedure for Western Component (4Xopqrs5) pollock. DFO Canadian Scientific Advisory Secretariat Research Document 2011/nnn.

Rademeyer, R.A., Plagányi, É.E. and Butterworth, D.S. (2007) Tips and tricks in designing management procedures. *ICES Journal of Marine Science* **64**, 618–625.

Rademeyer, R.A., Butterworth, D.S. and Plagányi, É.E. (2008) A history of recent bases for management and the development of a species-combined Operational Management Procedure for the South African hake resource. *African Journal of Marine Science* **30**, 291–310.

Radomski, P., Bence, J.R. and Quinn, T.J. II (2005) Comparison of virtual population analysis and statistical kill-at-age analysis for a recreational, kill-dominated fishery. *Canadian Journal of Fisheries and Aquatic Sciences* **62**, 436–452.

Rochet, M.-J. and Rice, J.C. (2009) Simulation-based management strategy evaluation: ignorance disguised as mathematics? *ICES Journal of Marine Science* **66**, 754–762.

Rosenberg, A.A. and Brault, S. (1993) Choosing a management strategy for stock rebuilding when control is uncertain. In: *Risk Evaluation and Biological Reference Points for Fisheries Management* (eds S.J. Smith, J.J. Hunt and D. Rivard). Canadian Special Publication in Fisheries and Aquatic Sciences **120**, 243–249.

Rosenberg, A.A. and Restrepo, V.R. (1994) Uncertainty and risk evaluation in stock assessment advice for U.S. fisheries. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 2715–2720.

Sainsbury, K.J., Campbell, R.A.Lindholm.R. and Whitelaw, A.W. (1997) Experimental management of an Australian multispecies trawl fishery: examining the possibility of trawl induced habitat modification. In: *Global Trends: Fisheries Management* (eds E.L. Pikitch, D.D. Huppert and M.P. Sissenwine). American Fisheries Society Symposium 20, Bethesda, Maryland, pp. 107–112.

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

Sampson, D.B. and Yin, Y. (1998) A Monte Carlo evaluation of the Stock Synthesis assessment program. In: *Fishery Stock Assessment Models* (eds F. Funk, T.J. Quinn II, J. Hiefetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan and C.-I. Zhang). Alaska Sea Grant College Program Report No. AK-SG-98-01, pp. 315–318.

Schweder, T., Hagen, G.S. and Hatlebakk, E. (1998) On the effect on cod and herring fisheries of retuning the Revised Management Procedure for minke whaling in the greater Barents Sea. *Fisheries Research* **37**, 77–95.

Smith, A.D.M. (1994) Management strategy evaluation – The light on the hill. In: *Population Dynamics for Fisheries Management, Australian Society for Fish Biology Workshop Proceedings, Perth 24–25 August 1993* (ed. D.A. Hancock). Australian Society for Fish Biology, Perth, pp. 249–253.

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

Smith, A.D.M., Fulton, E.J., Hobday, A.J., Smith, D.C. and Shoulder, P. (2007) Scientific tools to support the practical implementation of ecosystem-based fisheries management. *ICES Journal of Marine Science* **64**, 633–639.

Soutar, A. and Isaacs, J.D. (1969) History of fish populations inferred from fish scales in anaerobic sediments off California. *CalCOFI Reports* **13**, 63–70.

Soutar, A. and Isaacs, J.D. (1974) Abundance of pelagic fish during the 19th and 20th Centuries as recorded in anaerobic sediment off the Californias. *Fishery Bulletin* **72**, 257–273.

Starr, P.J., Breen, P.A., Hilborn, R.H. and Kendrick, T.H. (1997) Evaluation of a management decision rule for a New Zealand rock lobster substock. *Marine and Freshwater Research* **48**, 1093–1101.

STECF. (2011a) Impact Assessment of Baltic cod multi-annual plans (STECF 11-05) (ed. J. Simmonds). EUR 24899 EN–2011, 231 pp.

STECF. (2011b) Evaluation of multi-annual plans for cod in Irish Sea, Kattegat, North Sea, and West of Scotland (STECF-11-07) (eds J. Simmonds and S. Kraak). EUR 24901 EN–2011, 358 pp.

Szuwalski, C. and Punt, A.E. (2013) Fisheries management for regime-based ecosystems: a management strategy evaluation for the snow crab fishery in the eastern Bering Sea. *ICES Journal of Marine Science* **70**, 955–967.

Tuck, G.N. (2011) Are bycatch rates sufficient as the principal fishery performance measure and method of assessment for seabirds? *Aquatic Conservation: Marine and Freshwater Ecosystems* **21**, 412–422.

Ulrich, C., Andersen, B.S., Sparre, P.J. and Nielsen, J.R. (2007) TEMAS: fleet-based bioeconomic simulation software to evaluate management strategies accounting for fleet behaviour. *ICES Journal of Marine Science* **64**, 647–651.

Wade, P.R. (1998) Calculating limits to the allowable human-caused mortality of cetaceans and pinnipeds. *Marine Mammal Science* **14**, 1–37.

Walters, C.J. (1986) *Adaptive Management of Renewable Resources*. Macmillan, New York.

Walters, C.J. (1994) Use of gaming procedures in evaluation of management experiments. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 2705–2714.

Wayte, S.E. (2011) Management implications of including a climate-induced recruitment shift in the stock assessment for jackass morwong (*Nemadactylus* *macropterus*) in south-eastern Australia. *Fisheries Research* **142**, 47–55.

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

Wiedenmann, J., Wilberg, M.J. and Miller, T.J. (2013) An evaluation of harvest control rules for data-poor fisheries. *North American Journal of Fisheries Management* **33**, 845–860.

Wilson, J.R., Prince, J.D. and Lenihan, H.S. (2010) A management strategy for sedentary nearshore species that uses Marine Protected Areas as a reference. *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science* **2**, 14–27.

