Purported flaws in management strategy evaluation: basic problems or misinterpretations?

Douglas S. Butterworth, Nokome Bentley, José A. A. De Oliveira, Gregory P. Donovan, Laurence T. Kell, Ana M. Parma, André E. Punt, Keith J. Sainsbury, Anthony D. M. Smith, and T. Kevin Stokes

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

Rochet and Rice, while recognizing management strategy evaluation (MSE) as an important step forward in fisheries management, level a number of criticisms at its implementation. Some of their points are sound, such as the need for care in representing uncertainties and for thorough documentation of the process. However, others evidence important misunderstandings. Although the difficulties in estimating tail probabilities and risks, as discussed by Rochet and Rice, are well known, their arguments that Efron's non-parametric bootstrap re-sampling method underestimates the probabilities of low values are flawed. In any case, though, the focus of MSEs is primarily on comparing performance and robustness across alternative management procedures (MPs), rather than on estimating absolute levels of risk. Qualitative methods can augment MSE, but their limitations also need to be recognized. Intelligence certainly needs to play a role in fisheries management, but not at the level of tinkering in the provision of annual advice, which Rochet and Rice apparently advocate, *inter alia* because this runs the risk of advice following noise rather than signal. Instead, intelligence should come into play in the exercise of oversight through the process of multiannual reviews of MSE and associated MPs. A number of examples are given of the process of interaction with stakeholders which should characterize MSE.

Keywords: management procedure, management strategy evaluation, Monte Carlo simulation, risk estimation, uncertainty.

Received 26 August 2009; accepted 19 January 2010; advance access publication 5 March 2010.

D. S. Butterworth: Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7701, South Africa. N. Bentley: Trophia Ltd, PO Box 60, Kaikoura 7300, New Zealand. J. A. A. De Oliveira: Centre for Environment, Fisheries and Aquaculture Science, Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK. G. P. Donovan: International Whaling Commission, 135 Station Road, Impington, Cambridge CB24 9NP, UK. L. T. Kell: ICCAT Secretariat, Corazon de Maria 8 280002 Madrid, Spain. A. M. Parma: Centro Nacional Patagonico, Puerto Madryn, Chubut, Argentina. A. E. Punt: School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA, and CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart, Tasmania 7001, Australia. K. J. Sainsbury: University of Tasmania, Private Bag 46, Hobart 7001, Australia. A. D. M. Smith: CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart, Tasmania 7001, Australia. T. K. Stokes: 59 Jubilee Rd, Khandallah, Wellington 6035, New Zealand. Correspondence to D. S. Butterworth: tel: +27 21 6502343; fax +27 21 6502334; e-mail: doug.butterworth@uct.ac.za

Introduction

Management strategy evaluation [MSE, also known as the management procedure (MP) approach] was developed to address problems identified with more traditional management approaches (e.g. Kirkwood, 1992, 1996), initially with respect to the management of commercial whaling by the International Whaling Commission (IWC). Critical appraisals of MSE are opportune, particularly given the slow but steady moves towards increased application of the approach at both national and international (e.g. Regional Fisheries Management Organization, RFMO) levels (Punt, 2006).

As developers of and hence advocates for MSE, because of the great improvements the approach can bring to fisheries management (e.g. Butterworth and Punt, 1999; Smith *et al.*, 1999; Parma, 2002; Kell *et al.*, 2006; Punt and Donovan, 2007; De Oliveira *et al.*, 2008; Bentley and Stokes, 2009), we welcome pertinent warnings offered by Rochet and Rice (2009; hereinafter referred to as R&R) concerning some aspects of the approach that are potentially subject to poor implementation. Their criticisms centre on some of the ways in which uncertainty is represented and risk levels are estimated in MSE, although they do note that "these tools improve practice compared with ignoring uncertainty and applying *ad hoc* decision-making".

However, in welcoming R&R's critique, we also consider that it is important to draw attention to some important misunderstandings evident in their paper that bring some of their conclusions under question. Further, we are concerned that some of the issues raised by R&R may be misconstrued as basic flaws of MSE, rather than concerns about some technical aspects of its implementation.

The most important misunderstanding seems to be what precisely is meant by the term "management strategy evaluation". It is essential to be clear about the concept and implementation of MSE, which has come to have a specific meaning in a fisheries context (see the Glossary in Rademeyer *et al.*, 2007). MSE is a process for the evaluation of decision rules for providing scientific recommendations for management measures (such as total allowable catch, TAC, or total allowable effort) for a marine renewable resource, or a set of linked species. This evaluation is made against user- and resource-orientated objectives set by managers that should take into account the various stakeholders' views and needs. Importantly, the data inputs to such rules are prespecified, and in line with the precautionary approach (FAO, 1996; Richards and Maguire, 1998), the rules are checked using simulation to ensure that they are likely to provide reasonably robust performance given the plausible range of scientific and operational uncertainties considered to apply to the resource and associated fishery, group of fisheries, or even ecosystem (Sainsbury *et al.*, 2000; Smith *et al.*, 2007). It is especially this simulation testing, which takes proper account of feedback effects, which differentiates MSE from the more traditional "best assessment" approaches to providing fishery management advice (Butterworth, 2007).

MSE can be applied at three levels:

1. development of the specific and operational MP for a particular fishery, e.g. those for the major South African fisheries summarized in Plagányi *et al.* (2007), or the strike limit algorithms developed for subsistence fisheries by the IWC (IWC, 2002, 2004);
2. the evaluation of generic MPs, such as the IWC's revised MP (RMP) for baleen whales on their feeding grounds (IWC, 1999), which can be applied to a number of stocks [note that specific implementations of the generic RMP are also evaluated using the simulation approach, e.g. IWC (2008)];
3. more broadly the evaluation of management strategies in the form, for example, of general formulations for harvest control rules where, as R&R state, the process can *inter alia* be used to identify some formulations that will not work and can hence be eliminated.

For brevity of expression in what follows, we will use the acronym "MP" to refer to the outcome from MSE at all three of these levels. Nevertheless we note that MSE does not extend to broader concepts still, such as those underlying, for example, the European Common Fisheries Policy, which R&R also term a "management strategy", or national "harvest strategy standards" such as the US National Standard Guidelines that expand upon the Magnusson–Stevens Act (NMFS, 1998). However, MSE could, and we would suggest should, be used to evaluate how well a specific MP, with its embedded monitoring, assessment method, and decision rule, could be expected to achieve these broader goals (Kell *et al.*, 2005a).

In what follows, we comment first on some positive and then on some questionable viewpoints advanced by R&R and argue that most of the problems raised concerning MSE are addressed through its proper interpretation and implementation.

Features of a poorly implemented MSE

Some of the criticisms made by R&R do not apply to MSE *per se*. Rather they reflect instances of poor implementation of MSE, as elaborated below.

If MSE is to address the requirements of the precautionary approach so as to identify MPs that are robust to scientific uncertainties, which was the key factor behind the development of that approach, those uncertainties have to be realistically reflected in the simulation "trials" on which the testing of proposed MPs is based. Among other things, care needs to be taken not to input probability distributions, e.g. Bayesian priors, that are unrealistically narrow, or misleadingly low estimates of risk would result. Equally, problems can arise if questionably high levels of uncertainty are input to the process—see the debate in Kolody *et al.* (2008) and Butterworth (2008a), as well as in Butterworth *et al.* (1996). The proponents of MSE have never suggested that it can compensate for poor data inputs or inappropriate assumptions, nor has it been suggested that it can calculate very precisely the differences in outcomes expected under marginally different strategy selections, any more than can other quantitative analyses. In essence, these, and a number of other reservations raised in the Discussion section of R&R, are drawing attention to the GIGO (garbage in garbage out) principle, with which we naturally concur.

Similarly, we agree with R&R that the complete MSE process must be thoroughly documented for each application, particularly the details of the simulation trials considered and their motivations, so that the reliability of eventual outputs from the process is susceptible to thorough independent review. The documentation that accompanies IWC reports of the implementation reviews of their MPs for both commercial and aboriginal whaling (e.g. IWC, 2009—see Appendix 6 for North Atlantic fin whales, *Balaenoptera physalus*; IWC, 2004—see Appendix 3 for eastern North Pacific grey whales, *Eschrichtius robustus*) provides useful benchmarks for best practice for such documentation.

The whole exercise involves an assessment of risk in some way, and this necessarily requires inferences about the lower portions of probability distributions (typically 5 or 10%). It would be a poor analyst who showed no cognizance of the difficulties of estimating tail probabilities, and the sensitivity of such estimates to the probability distribution shapes assumed, particularly for low percentiles.

The above alone should be sufficient to confirm that MSE, provided properly applied to avoid poor practices such as those discussed above, does not remove "intelligence" from the fishery management process, a concern raised by R&R in their Discussion section. However, there are different places where "intelligence" can be applied to modify intermediate outcomes in the multistep process of MSE, and here we have some reservations concerning R&R's suggestions about when such interventions should occur, for reasons on which we elaborate in the penultimate section of this manuscript.

Paradoxes or misinterpretations?

R&R assert that stochastic modelling of uncertainty, which applies to Bayesian assessments of stock status as well as to MSE, is paradoxical because "it implies knowing more than deterministic approaches: to know the distribution of a quantity requires more information than only estimating its expected value". However, use only of a deterministic value, such as an expectation, ignores the facts that that value had to be estimated in some way and that estimation would not have been exact; there may be associated bias and certainly lack of precision. Surely, analyses that assume uncertainty does not exist cannot be argued to be preferable in a management context to others that attempt to take some account of such uncertainty, though imperfectly? Certainly, the former would be totally at variance with the precautionary approach, and ignoring uncertainty has been a major cause of the failure of some natural resource management attempts, e.g. not accounting for uncertainty about the value of natural mortality *M* for orange roughy (*Hoplostethus atlanticus*) in New Zealand (Mace *et al*., 1990), and about the productivity of groundfish species off the west coast of North America (Ralston, 1998, 2002). The argued paradox seems to us a consequence rather of a mis-focus away from the basic question: which of a range of possible MPs or strategies best meets management objectives? Or conversely, what is the basis for considering that a particular MP is likely to achieve its intended outcome in the context of the information and uncertainties in the observation, assessment, and management processes?

R&R also claim that non-parametric bootstrap re-sampling, a method which provides probability distributions without the need to assume parametric forms and which has been widely applied in statistics since being introduced by Efron (1981, 1987), "underestimates the probabilities of low values", and hence also of the risks of undesirable outcomes. However, their justification for this conclusion is flawed. (Furthermore, the results shown in Figure 3 of R&R demonstrate an overestimation of risk, not an underestimation, as they state.) The sparser distribution of observations farther from the centre of their distribution will certainly lead to a lower proportion of resamples drawn from those outer compared with central regions, but that is exactly in line with the relatively lower level of the distribution in such outer regions. Therefore, this in no way implies a necessary bias, as argued by R&R. It simply means that the bootstrap process estimates probabilities near the centre of distributions more precisely than near the tails, a problem that may be severe when the sample size is small, as in the examples provided by R&R. Bootstrap percentiles can be biased because the bootstrap distribution is centred on the estimated rather than the true value of the parameter in question, and methods exist to correct for this (Efron and Tibshirani, 1993), but the issue raised by R&R relates to variance, not bias.

Although R&R's critique that the tails of distributions are often poorly estimated may be true in absolute terms, MSE is almost always concerned with relative performance among alternative MPs (and the trade-offs involved across multiple objectives), rather than in absolute levels of performance for a particular MP (Smith *et al*., 1999).

MSE does not purport to know more about issues that are uncertain, but instead addresses uncertainty by concentrating on evaluating the robustness of MP performance to that uncertainty. R&R state that "Robustness may be tested by combining the full suite of related simulations probabilistically into a combined probability density function of possible outcomes, and evaluating the risk of an undesirable outcome from each management option". Of course this could be done, but it is certainly not, in our experience, a standard practice in MSE implementations. Most MSE studies integrate across uncertainty in some parameters/variables, such as variation in future recruitment and future observation error, and some go an extra step to integrate over a plausibility-weighted reference set of operating models (OMs) describing a range of highly plausible scenarios (e.g. CCSBT, 2005a). However, when it comes to evaluating robustness, one wants to demonstrate this robustness in anticipated performance under simulation across different OMs, not to consider only some tail-probability risk criterion being met for a single probability distribution generated from a weighted set of all OMs. The relative plausibility of different scenarios (OMs) has to be taken into account to some extent, because clearly no MP can be robust to every conceivable situation (Butterworth *et al*., 1996). However, although there are standard methods for calculating relative probabilities associated with parameter estimation uncertainty within a single model structure, assigning plausibility weightings across different model structures, e.g. how many stocks of the species are present in the area under management, is recognized as a far more difficult and certainly as yet not fully solved problem.

Probably the most advanced approach to date is that developed in the IWC, which accords scenarios corresponding in particular to different model structures a high, medium, or low plausibility on qualitative grounds ("intelligence"). It then disregards the low ones and requires the performance of candidate MPs to satisfy less stringent conservation-related criteria for trials with medium plausibility weighting compared with the criteria that apply for those given high weighting (IWC, 2005). The focus is therefore on consistency across stocks and over time in evaluating risk, rather than claiming a single integrative measure of risk as a probability calculated in absolute terms.

If indeed MSEs are not "powerful at discriminating the best approach" (R&R), is that of great consequence? As indicated above, the priority in selecting management approaches is robustness to unavoidable uncertainty rather than seeking an optimality that is in any case difficult to define given the inevitable trade-offs in performance between catch on average, catch variability, and the risk of depleting the resource to too low a level. We are therefore somewhat bemused by R&R's concern that MSE is being used "as if small differences in analytical results are meaningful biologically and in management". We equally would be concerned if this was common practice in MSE, but we are not aware of any instances where the outputs from an MSE process have been used in that manner. If such instances do exist, they are the exception rather than the rule. Indeed, in our experience, the far greater problem in current stock assessments is for small differences in estimated values, e.g. the estimated current fishing mortality *F* compared with a target *F* each year, to be treated as meaningful despite the stock assessment model not including important sources of uncertainty. The resulting scientific overconfidence and lack of robustness have arguably contributed to overexploitation of several fisheries (see, e.g., Mace *et al*., 1990; Ralston, 1998, 2002).

Finally, R&R's argument that "implementation uncertainty is very likely to greatly exceed any nuanced quantitative differences amongst simulation results" is hardly an argument against MSE *per se*. If implementation issues are among the major of the set of uncertainties that apply in a particular situation, the trials used for testing associated candidate MPs should certainly be attempting the not always straightforward task of making appropriate allowance for implementation error. For example, Dichmont *et al*. (2006) characterized implementation uncertainty in Australia's northern prawn fishery (NPF) based on historical decision-making, and incorporated that source of uncertainty in the scenarios considered in the simulation trials used to compare MPs for the two commercially important tiger prawn species (*Penaeus esculentus* and *P. semisulcatus*) in the NPF. The implementation of MSE for the management of a tropical trawl fishery (Sainsbury, 1991) explicitly included a high level of implementation uncertainty. The evaluation of alternative composite fishery management scenarios by Fulton *et al*. (2008) goes to considerable lengths to represent both implementation uncertainty and the response of fishers to changes in management arrangements. Open source software available to apply MSE to fisheries explicitly includes implementation uncertainty in several of the management processes (Kell *et al*., 2007). In any case though, the feedback nature of MPs will, at least partially, adjust for implementation error such as failure to take account of unreported catches (e.g. IWC, 1992).

Alternatives?

R&R admit that they can offer no cure to a number of problems they raise in relation to MSE. They suggest that MSE be augmented by alternative methods to evaluate risks, including qualitative methods and *post hoc* analyses. However, much of this is already accepted and implemented best practice where MSE is used. The Multi-Criteria-Decision-Making context of the interpretation of outputs from the simulation-testing process, which incorporates more than one measure of risk and emphasizes the key role of trade-offs among conflicting objectives, is widely recognized. Regular reviews are frequently mandated, e.g. five-yearly in the IWC (IWC, 1999) and four-yearly in South Africa (Appendix 2 of Rademeyer *et al*., 2008), and these allow further knowledge and experience gained in the interim to be taken into account.

We agree that qualitative methods (e.g. Smith *et al*., 2007) have a useful supplementary role to play alongside more quantitative approaches, from several perspectives, including facilitating stakeholder engagement, and the speed and cost with which analyses can be completed. However, qualitative methods tend to be less adept at incorporating uncertainty and can be poor at capturing the often unexpected effects of feedback, including future learning (Smith *et al*., 2009). *Post hoc* analyses can be even more problematic, because the limited number of real world "experiments" to hand, combined with severe problems of attribution of cause, make learning from such analyses difficult. Ongoing contention over explanations for the failure of NW Atlantic cod (*Gadus morhua*) to recover (Shelton *et al*., 2006), and for the decline of Steller sea lions (*Eumetopias jubatus*) in the North Pacific (National Research Council, 2003), are cases in point. Moreover, analyses that provide only qualitative or directional management insight are necessarily incomplete; the first question that will be asked by a manager when advised that the TAC needs to be reduced is going to be: "By how much?".

More to the point, we believe, are the advantages that MSE brings over previous approaches, an aspect that R&R do acknowledge partially. These are elaborated in more detail in Punt (2006), Butterworth (2007), and Punt and Donovan (2007), but it is worth re-emphasizing some of the more important ones, such as a structured approach to taking uncertainty into account, being able to make allowance for the feedback (learning) contribution forthcoming from future data as the MP is put into practice, and providing a basis to consider trade-offs between interannual catch variability and resource risk, or between short-term pain and achieving resource rebuilding targets, so that the aim of socio-economic stability can be factored in more objectively.

R&R quote the Sparholt *et al*. (2007) observation that despite the implementation of the precautionary approach by ICES for providing scientific advice, demersal stocks have still declined. However, Cadrin and Pastoors (2008) showed that of 137 management units for which advice is provided by ICES, only 17% actually have the necessary estimates to implement a precautionary control rule, and 61% have no estimates of reference points at all. The failure has been related to not implementing scientific advice in an effective management context, and to not evaluating the extent to which management strategies are able to meet the objectives set out for them, given uncertainties. MSE has helped address these failings by developing management advice for stocks for which detailed assessments have yet to be developed (Kell *et al*., 2005a, b; STECF, 2008).

The appropriate role for "intelligence"

We certainly agree with R&R that MSEs provide a mechanism for a desirable strengthening of "the use of intelligence in developing approaches to fisheries management". As discussed above, the properly conducted process of setting up the simulation trials required for MSE and the interpretation of the results already requires full exercise of that attribute.

R&R subsequently comment, however, that few OMs "can simulate the acts of scientists, managers, and fishers making choices in each year of an ongoing cyclic process of fishery-assessment-advice-management plan-fishery" and that "this is the wrong time to remove intelligence from the process. If all their quantitative details are taken as reliable, then simulation-based MSEs can only evaluate impoverished procedures, those without the need for human intervention. This is not the type of management procedure we want".

What R&R apparently advocate is, we believe, unwise and indeed quite the reverse of the philosophy underlying levels 1 and 2 of MSE (see above), which insists on the MPs adopted being exactly of the form that has been simulation-tested, and to be implemented in the "auto-pilot" mode, i.e. the management measures they output to be applied unchanged unless there is compelling evidence for the need for some amendment (what are termed exceptional circumstances). Jurisdictions such as Australia that use the MSE at level 3 to inform both policy and choice of fishery-specific harvest strategies without fully adopting an MP do provide more room for human intervention in annual decision-making, but even this is formulated as the application of meta-rules that constrain the freedom of decision-makers to simply override application of the strategy without sound justification (Smith *et al*., 2008).

R&R motivate their position on Punt (1997) having used Laurec–Shepherd tuning rather than ADAPT to simulation-test VPA-based management "because the latter requires making educated choices that cannot be simulated in an operating model". If a procedure cannot be simulated, how can one have confidence that it will perform better than one which has been? A counter-example is provided by Punt's (1993) simulation-based comparison of the age-aggregated production model and the VPA-based management. The educated choice would surely be for VPA, because this can make use of additional information (catch-at-age data). Yet as the case investigated showed, compared with the production model, VPA added nothing in terms of greater catches on average, or less risk to the resource, but led to far greater interannual variability in the catches, i.e. the addition of "intelligence" led to a deterioration in performance in meeting one management objective without any compensating gains for others. It is surely far better to develop an MP that can be simulated, and thus that can be shown to be likely to meet the user- and resource-based objectives identified.

The philosophy apparently espoused by R&R of, even after conducting an MSE, returning to the annual cycle of scientifically debating the management action to follow an assessment process which itself may be contested, reintroduces many of the problems that levels 1 and 2 of MSE are intended to avoid, by providing a default decision rule which has been tested and proved to be robust. On what bases could one claim that "intelligence" routinely entered at the stage of developing management advice leads to improved performance in the longer term? Or does it instead only "tinker", by providing management recommendations that follow noise rather than signal in the data, and reopen the door to *ad hoc* decisions that lack a clear and consistent basis? A particular problem with this last approach is that it admits a role for strengthening advocacy positions over time as resource indices deteriorate, resulting in the necessary management response of adequate reductions in catch and/or effort being enacted too little too late. Avoiding such problems underlay the IWC's (and subsequently a number of other organizations') decision to opt instead for the MP auto-pilot route of levels 1 and 2 of MSE.

Even when advocacy is not a problem, frequent changes in models or data are often introduced in an effort to provide the best stock assessment every year. Such changes may lead to unnecessary disruptions in the management advice, when a simpler, consistent decision rule applied every year might perform better in the long term (e.g. Parma, 2002). Relatively simple estimation models are often both sufficient and more robust than their "more realistic" counterparts as the basis for MPs, because of the high noise-to-signal ratio typical of fisheries data. For example, in the development of the IWC's RMP, the relatively simple Cooke population estimation model that was eventually adopted performed at least as well in simulations as other MPs based on more complex models that attempted to match features of the underlying OMs of whale population dynamics, such as time-lags, more closely (IWC, 1992). This is not to deny any role for such more realistic models, but this is generally at the OM level in providing necessary realism for the possible underlying dynamics of the system in the simulation trials used for testing purposes, rather than in calculating management measures to be applied in practice. Therefore, for example, a recent FAO workshop on modelling ecosystem interactions for informing an ecosystem approach to fisheries concluded that although such ecosystem models were not yet at the stage where they might be used directly to provide management recommendations, they did have a role to play as OMs in evaluating the performance of simpler models used as the basis for MPs (FAO, 2008).

There is a role for "intelligence" at the stage of implementing an MP for either of levels 1 or 2, but it is rather in the oversight role of a pilot regularly checking that his autopilot continues to achieve the results planned. This includes checking that exceptional circumstances have not arisen in the form of compelling scientific evidence that the resource has moved outside the range of circumstances (including implementation error) spanned by the simulation testing, which necessitates that the advice provided by the MP be overruled (for level 3 of MSE as applied in Australia, the constraints on human intervention provided by the meta-rules play a similar role). This should be a rare occurrence—it is not a mechanism to allow annual "tinkering" with the scientific recommendations output by an MP. Best practice for MSE implementation includes documentation of an agreed process for such an eventuality, as well as a regular MP review process during which "intelligence" can be brought to bear in considering whether amendments to an MP are required [Butterworth, 2008b; and see specific examples in CCSBT, 2005b, and Appendix 2 of Rademeyer *et al.*, 2008; the 5-year regular IWC MP review process (IWC, 1999) includes the option to call for an "exceptional" review under certain circumstances].

Stakeholder participation

R&R admonish a frequent absence in MSE publications of details of how stakeholders were involved in the MP development process, given that the opportunity that MSE provides for such involvement, and hence for more likely acceptance of the product, is often advanced as one of its major benefits.

Indeed, an interesting feature of the growing application of MSE and MPs is increasing acceptance by industry of the process, because it provides them with greater planning security by removing the uncertainty associated with the outcome of the annual assessment/TAC scientific debate. Failures to report details of these interactive processes may reflect the fact that in many fora, these processes settled down some time ago and are now fairly routine.

At an international level, in the IWC, there has been a concerted effort to ensure stakeholder participation. This is normally carried out via national delegations, which often include representatives of environment and fisheries ministries, industry, and NGOs. With respect to the development of the MPs for subsistence whaling, the chair of the scientific group working on development talked directly with representatives of the subsistence communities throughout the development process. The primary interests of stakeholders have been: determination of the user- and resource-orientated objectives and the balance between these; performance measures of proposed procedures against those objectives; and data and analysis requirements and availability. Technical details of approaches have rarely been the focus of concern of the stakeholders. The Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) has so far adopted the approach of a broad interchange between scientists and managers (Commissioners) at its annual meetings in its work towards development of an MP to govern spatial allocation of the krill (*Euphausia superba*) catches allowed in the Scotia Sea (CCAMLR, 2008). Moreover, stakeholders such as the fishing industry and NGOs have been party to relevant discussions by virtue of the fact that they are either involved as members of national delegations or serve as observers at the Commission or Scientific Committee level. In the Commission for the Conservation of Southern Bluefin Tuna (CCSBT), where adoption of an MP is nearing finality (though after some hiccoughs—see Kolody *et al.*, 2008), there have been several special meetings held to allow for interaction between scientists, managers, and industry representatives on the details of the MP as it is developed (e.g. CCSBT, 2005c), and an initiation of this process at NAFO for Greenland halibut (*Reinhardtius hippoglossoides*) took a similar route (NAFO, 2008).

In Europe, there is increased emphasis on stakeholder participation and the ecosystem-based approach to fisheries management in the reform of the CFP (Gray and Hatchard, 2008). The development of MSE is therefore an important tool because it allows a broad array of European, national, and local stakeholders to be involved. For example, MSE is being used in a participatory modelling process in the JAKFISH project (http://www.imares.wur.nl/UK/research/marinefisheries/projects/jakfish/), so that scientists and stakeholders can jointly develop flexible and transparent models to evaluate alternative management strategies before implementation. This will allow stakeholders to decide upon the best assessment, monitoring, and MPs, based upon incomplete knowledge of stock structure and subpopulation catches for the management of the herring (*Clupea harengus*) meta-population (Kell *et al.*, 2009), and to help perform bioeconomic evaluation of different management measures (Tserpes *et al.*, 2009).

Nationally, in South Africa, the primary discussions take place in species-specific scientific working groups on which industry and managers are present as observers, but in practice can and often do participate fully in discussions (Butterworth, 2008b). These discussions have focused particularly on the trade-off arising from restrictions on interannual TAC variability that industries seek for reasons of stability, and the associated lesser average fishing mortality levels and hence lower catches on average over time that such restrictions require.

Involvement of stakeholders, including environmental NGOs as well as the fishing industry, is a basic and widespread feature of the Australian system of fisheries governance. This stakeholder participation has been an important aspect of developing and applying the MSE approach in Australia (Smith *et al.*, 1999).

Cox and Kronlund (2008) used MSE to develop practical stakeholder-driven harvest policies for groundfish fisheries in British Columbia, Canada. They found that MSE helped in developing co-management, because in traditional stock assessment there is only a scientific choice of what constitutes the best assessment method, and the long-term policy consequences of particular assessment model choices are rarely evaluated. Therefore, MSE offered a potential vehicle for addressing both policy and process conflicts in fishery co-management.

In New Zealand, stakeholder (commercial, recreational, customary Maori, and environmental) involvement in scientific and management meetings is the norm for all fisheries. In addition, all decision-making, including the setting of TACs, is subject to statutory public consultation. Although most stocks are managed using the traditional assessment approach, MPs have been developed to guide TAC-setting for some rock lobster stocks since the mid-1990s (e.g. Starr *et al.*, 1997). MP development and recommendations and consequent TAC advice have been provided by a national multistakeholder group with responsibility for providing advice directly to the Minister of Fisheries in place of the usual government-led processes. The MP development processes have been highly collaborative involving stakeholders, scientists, and government officials. Based on advice from the multistakeholder group (http://www.nzrocklobster.co.nz/rl-mandocs) reflecting changing stakeholder goals, successive ministers have decided upon a sequence of 5-year MPs since 1996 and have also implemented TAC cuts and increases as indicated by the adopted MPs and advised on annually by the multistakeholder group.

Acknowledgements

We thank three anonymous reviewers for their comments on an earlier version of this paper.

References

Bentley, N., and Stokes, T. K. 2009. 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.

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, pp. 381–397. Ed. by K. Tsukamoto, T. Kawamura, T. Takeuchi, T. D. Beard, and M. J. Kaiser. TERRAPUB, Tokyo.

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., 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.

Cadrin, S. X., and Pastoors, M. A. 2008. Precautionary harvest policies and the uncertainty paradox. Fisheries Research, 94: 367–372.

CCAMLR. 2008. Report of the Twenty-Seventh Meeting of the Scientific Committee, Hobart, Australia, 27–31 October 2008. Part 1. CCAMLR, Australia. 635 pp.

CCSBT. 2005a. Report of the Special Management Procedure Technical Meeting, Seattle, USA, 15–18 February 2005. 47 pp. www.ccsbt.org/docs/pdf/meeting_reports/ccsbt_12/report_of_MPTM.pdf (last accessed August 2009).

CCSBT. 2005b. Report of the Tenth Meeting of the Scientific Committee, Narita, Japan, 9 September 2005. 90 pp. www.ccsbt.org/docs/pdf/meeting_reports/ccsbt_12/report_of_SC10 (last accessed August 2009).

CCSBT. 2005c. Report of the Management Procedure Special Consultation, Canberra, Australia, 23 May 2005. 15 pp. www.ccsbt.org/docs/pdf/meeting_reports/ccsbt_12/report_of_MP_SpecialConsultation (last accessed August 2009).

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., Kell, L. T., Punt, A. E., Roel, B. A., and Butterworth, D. S. 2008. Managing without best predictions: the management strategy evaluation framework. *In* Advances in Fisheries Science. 50 Years on from Beverton and Holt, pp. 104–134. Ed. by A. Payne, J. Cotter, and T. Potter. Blackwell Publishing, Oxford. xxi + 546 pp.

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

Efron, B. 1981. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika, 68: 589–599.

Efron, B. 1987. Better bootstrap confidence intervals. Journal of the American Statistical Association, 82: 171–185.

Efron, B., and Tibshirani, R. J. 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.

FAO. 1996. Precautionary approach to fisheries. Guidelines on the precautionary approach to capture fisheries and species introductions. FAO Fisheries Technical Paper, 350/1. 57 pp.

FAO. 2008. Fisheries management. 2. The ecosystem approach to fisheries. 2.1 Best practices in ecosystem modelling for informing an ecosystem approach to fisheries. FAO Fisheries Technical Guidelines for Responsible Fisheries, 4(Suppl. 2, Add. 1). 78 pp.

Fulton, E. A., Smith, A. D. M., and Smith, D. C. 2008. Alternative management strategies for Southeastern Australian Commonwealth Fisheries. Stage 2. Quantitative management strategy evaluation. Report to the Australian Fisheries Management Authority, CSIRO, Hobart, Australia.

Gray, T., and Hatchard, J. 2008. A complicated relationship: stakeholder participation and the ecosystem-based approach to fisheries management. Marine Policy, 32: 158–168.

IWC. 1992. Report of the scientific committee, Annex D. Report of the sub-committee on management procedures. Reports of the International Whaling Commission, 42: 87–136.

IWC. 1999. The revised management procedure (RMP) for baleen whales. Journal of Cetacean Research and Management, 1(Suppl.): 251–258.

IWC. 2002. 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, 4(Suppl.): 148–177.

IWC. 2004. 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, 6(Suppl.): 185–223.

IWC. 2005. Report of the scientific committee, Annex D. Report of the sub-committee on the revised management procedure. Appendix 2. Requirements and guidelines for implementation. Journal of Cetacean Research and Management, 7(Suppl.): 84–92.

IWC. 2008. Report of the second Intersessional Workshop on the Western North Pacific Bryde's Whale Implementation, Yokohama, 10–14 December 2006. Journal of Cetacean Research and Management, 410(Suppl.): 449–510.

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.

Kell, L. T., De Oliveira, J. A. A., Punt, A. E., McAllister, M. K., and Kuikka, S. 2006. Operational management procedures: an introduction to the use of management strategy evaluation frameworks. *In* The Knowledge Base for Fisheries Management, pp. 379–407. Ed. by L. Motos, and D. C. Wilson. Developments in Aquaculture and Fisheries Science, 36. Elsevier, Amsterdam.

Kell, L. T., Dickey-Collas, M., Hintzen, N. T., Nash, R. D. M., Pilling, G. M., and Roel, B. A. 2009. Lumpers or splitters? Evaluating recovery and management plans for metapopulations of herring. ICES Journal of Marine Science, 66: 1776–1783.

Kell, L. T., Mosqueira, I., Grosjean, P., Fromentin, J-M., Garcia, D., Hillary, R., Jardim, E., *et al.* 2007. FLR: an open-source framework for the evaluation and development of management strategies. ICES Journal of Marine Science, 64: 640–646.

Kell, L. T., Pastoors, M. A., Scott, R. D., Smith, M. T., Van Beek, F. A., O'Brien, C. M., and Pilling, G. M. 2005b. 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., Pastoors, M. A., Mesnil, B., Korsbrekke, K., Abaunza, P., *et al.* 2005a. An evaluation of the implicit management procedure used for some ICES roundfish stocks. ICES Journal of Marine Science, 62: 750–759.

Kirkwood, G. P. 1992. Report of the Scientific Committee. Annex I. Background to the development of revised management procedures. Reports of the International Whaling Commission, 42: 236–243.

Kirkwood, G. P. 1996. Assessing the precautionary nature of fishery management strategies. FAO Fisheries Technical Memorandum, 350/2: 141–158.

Kolody, D., Polacheck, T., Basson, M., and Davies, C. 2008. Salvaged pearls: lessons learnt from a floundering attempt to develop a management procedure for southern bluefin tuna. Fisheries Research, 94: 339–350.

Mace, P. M., Fenaughty, J. M., Coburn, R. P., and Doonan, I. J. 1990. Growth and productivity of orange roughy (*Hoplostethus atlanticus*) on the north Chatham Rise. New Zealand Journal of Marine and Freshwater Research, 24: 105–119.

NAFO. 2008. Report of the NAFO Scientific Council Study Group on Rebuilding Strategies for Greenland halibut, Instituto Espanol de Oceanografia/Centro Oceanográfico de Vigo, 21–23 February 2008, Vigo, Spain. NAFO SCS Document, 08/13. 31 pp. www.nafo.int/publications/frames/science.html (last accessed August 2009).

National Research Council. 2003. Decline of the Steller sea lion in Alaskan waters: untangling food webs and fishing nets. National Academies Press, Washington, DC.

NMFS. 1998. Magnuson–Stevens Act Provisions; National Standard Guidelines; Final Rule. Federal Register, 63: 24212–24237.

Parma, A. M. 2002. In search of robust harvest rules for Pacific halibut in the face of uncertain assessments and decadal changes in productivity. Bulletin of Marine Science, 70: 455–472.

Plagányi, É. E., Rademeyer, R. A., Butterworth, D. S., Cunningham, C. L., and Johnston, S. J. 2007. Making management procedures operational—innovations implemented in South Africa. ICES Journal of Marine Science, 64: 626–632.

Punt, A. E. 1993. The comparative performance of production-model and ad hoc tuned VPA based feedback-control management procedures for the stock of Cape hake off the west coast of South Africa. *In* Risk Evaluation and Biological Reference Points for Fisheries Management, pp. 283–299. Ed. by S. J. Smith, J. J. Hunt, and D. Rivard. Canadian Special Publication of Fisheries and Aquatic Sciences, 120.

Punt, A. E. 1997. The performance of VPA-based management. Fisheries Research, 29: 217–243.

Punt, A. E. 2006. The FAO precautionary approach after almost 10 years: have we progressed towards implementing simulation-tested feedback-control management systems for fisheries management? Natural Resource Modelling, 19: 441–464.

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

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. African Journal of Marine Science, 30: 291–310.

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.

Ralston, S. 1998. The status of federally managed rockfish on the US West Coast. *In* Marine Harvest Refugia for West Coast Rockfish: a Workshop, pp. 6–16. Ed. by M. Yoklavich. NOAA Technical Memorandum, NMFS NOAA-TM-NMFS-SWFSC-255.

Ralston, S. 2002. West coast groundfish harvest policy. North American Journal of Fisheries Management, 22: 249–250.

Richards, L. J., and Maguire, J. J. 1998. Recent international agreements and the precautionary approach: new directions for fisheries management science. Canadian Journal of Fisheries and Aquatic Sciences, 55: 1545–1552.

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.

Sainsbury, K. J. 1991. Application of an experimental approach to management of a tropical multispecies fishery with highly uncertain dynamics. ICES Marine Science Symposia, 193: 301–320.

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.

Shelton, P. A., Sinclair, A. F., Chouinard, G. A., Mohn, R., and Duplisea, D. E. 2006. Fishing under low productivity conditions is further delaying recovery of Northwest Atlantic cod (*Gadus morhua*). Canadian Journal of Fisheries and Aquatic Sciences, 63: 235–238.

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.

Smith, A. D. M., Sachse, M. L., Fulton, E. A., Smith, D. C., Prince, J. D., Knuckey, I., Walker, T. J., *et al.* 2009. Evaluation of alternative strategies for management of Commonwealth fisheries in south eastern Australia. Fisheries Research and Development Corporation, Final Report 2003/061, Canberra.

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., Smith, D. C., Tuck, G. N., Klaer, N., Punt, A. E., Knuckey, I., Prince, J., *et al.* 2008. Experience in implementing harvest strategies in Australia's south-eastern fisheries. Fisheries Research, 94: 373–379.

Sparholt, H., Bertlesen, M., and Lassen, H. 2007. A meta-analysis of the status of ICES fish stocks during the past half century. ICES Journal of Marine Science, 64: 707–713.

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. New Zealand Journal of Marine and Freshwater Research, 48: 1093–1101.

STECF. 2008. Scientific, Technical and Economic Committee for Fisheries (STECF). Report of the Working Group on Harvest Control Rules (SGRST 08-02), 9–13 June 2008, Lowestoft, UK. Ed. by M. Haddon, and F. Hoelker. JRC 49098, doi:10.2788/38645.

Tserpes, G., Tzanatos, E., Peristeraki, P., Placenti, V., and Kell, L. 2009. A bio-economic evaluation of different management measures for the Mediterranean swordfish. Fisheries Research, 96: 160–166.

doi:10.1093/icesjms/fsq009