Simulation-based management strategy evaluation: ignorance disguised as mathematics?

Marie-Joëlle Rochet and Jake C. Rice

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.

Simulation-based management strategy evaluations are increasingly developed and used for science advice in support of fisheries management, along with risk evaluation and decision analysis. These methods tackle the problem of uncertainty in fisheries systems and data by modelling uncertainty in two ways. For quantities that are difficult to measure accurately or are inherently variable, variables are replaced by probability distributions, and system dynamics are simulated by Monte Carlo simulations, drawing numbers from these distributions. For processes that are not fully understood, arrays of model formulations that might underlie the observed patterns are developed, each is assumed successively, and the results of the corresponding arrays of model results are then combined. We argue that these approaches have several paradoxical features. Stochastic modelling of uncertainty 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. To combine the results of Monte Carlo simulations with different model formulations may be paradoxical if outcomes of concern are unlikely under some formulations but very likely under others, whereas the reported uncertainty from combined results may produce a risk level that does not occur under any plausible assumed formulation. Moreover, risk estimates of the probability of undesirable outcomes are often statements about likelihood of events that were seldom observed and lie in the tails of the simulated distributions, where the results of Monte Carlo simulation are the least reliable. These potential paradoxes lead us to suggest that greater attention be given to alternative methods to evaluate risks or management strategies, such as qualitative methods and empirical *post hoc* analyses.

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

Received 17 July 2008; accepted 18 January 2009; advance access publication 19 February 2009.

M-J. Rochet: IFREMER, Département Ecologie et Modèles pour l'Halieutique, BP 21105, Nantes Cedex 03, France. J. C. Rice: National Advisor – Ecosystem Sciences, Department of Fisheries and Oceans, 200 Kent Street, Ottawa, Ontario, Canada K1A 0E6. Correspondence to M-J. Rochet: tel: +33 2 40 37 41 21; fax: +33 2 40 37 40 75; e-mail: mjrochet@ifremer.fr.

Introduction

The precautionary approach has become a cornerstone of an improved approach to fisheries management (FAO, 1996; Richards and Maguire, 1998). The application of precaution is founded on two principles:

(i) that decision-making should be more risk-averse (i.e. precautionary) when uncertainty is higher than usual (where "usual" is relative, and may apply to assessments of a single stock over time, or across many stocks for a jurisdiction); and
(ii) that decision-making should be more risk-averse when plausibly bad outcomes involve serious or irreversible harm.

Both principles sound reasonable, and they require fisheries science to develop new or to adapt old analytical tools to make them operational.

One class of new or adapted tools involves the use of resampling methods that are designed to capture uncertainty more fully in the assessment process, so that science advisors and decision-makers actually know how uncertain is the information base for a decision (e.g. Patterson *et al.*, 2001; Berkson *et al.*, 2002), and what options really are risk-averse. The other class of tools for implementation of the precautionary approach uses simulation methods that allow the application of pre-agreed decision-rules (FAO, 1996; Garcia, 2005), and whose performance characteristics have been studied through simulation and use in fisheries management (Stokes *et al.*, 1999; Punt *et al.*, 2001; Butterworth and Punt, 2003). The combination of improved quantification of uncertainty and simulation-based management strategy evaluations (MSEs) can feed into processes for risk evaluation and decision analysis (Peterman, 2004) that are more formally structured than historical practices. However, the reliability of the more formal risk evaluation and decision analysis depends critically on the reliability of both types of tool. Their reliability depends in turn on the quantitative results of the resampling and simulations.

Without question, these tools improve practice compared with ignoring uncertainty and applying *ad hoc* decision-making. Unfortunately, they also create a new and subtle class of challenges. Moreover, in an attempt to increase transparency through making more assumptions in the analytical work of fisheries scientists explicit, opportunities are presented to substitute complex mathematical formulations for an admission that the experts really do not know what is going on. The methods need to be looked at critically with regard to both the opportunities they present for advances in practice and the new pitfalls against which practitioners must be vigilant.

The simulations consist in first building an operating model, assuming mathematical equations and parameter values for all processes from the ecology and dynamics of the resource to the dynamics of the exploiting fleets, their catch, and production. In addition, an observation model represents the scientific observation of the system, from data collection to stock assessment. The management procedure model includes this observation model and management decision rules. Then, simulations are used to explore the consequences of various management strategies. Commonly, these are repeated for several variants of the operating model, reflecting different assumptions about key processes in the stock or fishery. The foundations of this approach are not novel, because models have long been used to represent knowledge, and mathematics and computers to explore and understand the consistency and hidden content of that knowledge. What is new is the way these methods handle uncertainty. This issue is addressed by replacing deterministic parameters with stochastic (or otherwise probabilistic) ones, and a single functional relationship of interest with suites of alternative formulations of the relationships, often in various combinations. To address parameter uncertainty, variables and parameters that are inherently variable or difficult to measure or estimate with accuracy and precision are replaced by specified probability distributions. These are supposed to include the full variability and sampling distribution of the property which the parameter represents, not a single most likely or otherwise "expected" value. To address model uncertainty, processes that are not well understood are represented by arrays of alternative assumptions and corresponding functional relationships. These are intended to cover the range of processes that might underlie observed patterns, and/or the range of possible future states of nature. Uncertainty about multiple processes requires multiple arrays of assumptions and relationships, so complexity increases multiplicatively. These approaches then "capture" the full uncertainty by Monte Carlo or other probabilistic simulations, drawing numbers from the probability distributions of parameters, and running multiple scenarios using all combinations of functional relationships. Decision rules are then tested for robustness of performance in the face of the results produced by the simulations. Robustness may be tested by combining the results of a 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. Commonly the simulations are used to estimate the probability of violating a management benchmark such as B_{pa} or B_{lim}, or a proxy such as 20% virgin biomass (Butterworth and Punt, 1999; ICES, 2005a), or the probability of stock collapse (Hilborn, 1997). Alternatively, if managers have adopted a particular level of risk aversion to an event such as decline in spawning-stock biomass (SSB), then the simulations can estimate the maximum harvest that would comply with the risk tolerance (e.g. Figure 9 in CSAS, 2006). In both contexts, the decision support from such stochastic simulations must be assumed to have substantial quantitative detail behind it, particularly with regard to the likelihood of uncommon events. The multiple formulation–stochastic sampling approach also addresses the issue of estimating the severity of "plausibly serious harm", by combining the most extreme credible assumptions about processes with the most extreme credible values for parameters. All this looks very quantitative and rigorous, but it is in the appearance of rigour that new challenges and pitfalls are found, because of how models may be used to represent things that are not known.

Using alternative model formulations to "cover" the range of possible realities seems a good way of acknowledging uncertainty. However, if undesirable outcomes have not been experienced enough times to know the conditions that cause them, and particularly if undesirable events in the past appear to have had more than one cause (Hodgson *et al*., 2006; Mueter *et al*., 2007), these formulations should be viewed with scepticism, and very rarely should be considered to bracket the range of possible functional relationships. Predictions based on phenomenological models with a similar fit to data can be qualitatively completely different; this is a general result for models that are not entirely mechanistic (Wood and Thomas, 1999). Unfortunately, there is no agreed method to select the formulations to be included in the simulated array, nor to assign them weights or probabilities; yet these weights may be highly influential on the conclusions regarding plausible worse-case outcomes. Reviews such as Butterworth and Punt (1999) and Peterman (2004) have not solved any of the problems, but at least a critical assessment of their implications and associated risks is available to practitioners and users.

A similar critical assessment of the challenges and risks posed by using probability distributions rather than poorly specified parameters has not been undertaken. We argue that stochastic modelling of uncertainty is a paradoxical and at least potentially misleading approach, also producing potentially unreliable foundations for risk-averse decision-making. Two considerations mean that the detail of the quantitative support for risk-averse decisions could be illusionary. First, the information supporting a risk-averse decision is primarily the pattern in the tails of the simulated distributions, whose shapes are largely driven by assumptions regarding the variance and skewness of the sampling distribution of the parameters. When these have not been estimated carefully, the quantitative detail in the simulations lacks exactly the rigorous scientific backing that the detail invites decision-makers to infer is present in the decision support. Second, the support depends on the relative plausibility assumed for extreme events, often arising from the extreme model formulations. When MSEs are used to evaluate the risk of serious or irreversible harm, except for stocks that already collapsed often enough for the causes to be understood, the estimates rely on occurrences of events that were seldom or never observed. These events often correspond to the less well-studied hypotheses and hence are least well known and least likely to be formulated accurately and precisely.

This matters because when precautionary considerations apply, risk-averse decision-making is needed. For example, in International Council for the Exploration of the Seas (ICES) fisheries advice, B_{lim} is the biologically defined conservation boundary below which there is unacceptable risk of diminished stock productivity (ICES, 2005b). B_{pa} is used as the quantitative benchmark for harvest advice, however, serving as a risk control tool intended to ensure that the probability that true SSB is below B_{lim} is small, given the uncertainties in the assessment. This approach to positioning B_{pa} relative to B_{lim} makes advice and decisions sensitive to the shape of the tails of the parameter distribution(s) for B (and F), and less dependent on the much better specified centre of the distribution.

The use of complex mathematics and statistical tools risks giving users a false sense of rigour, whatever their relationships with the underlying knowledge. Users should not expect probabilistic advice on unlikely or extreme events to have the same accuracy and precision as probabilistic advice about events closer to the centre of the simulated distributions, yet precautionary decision support does exactly that. As with the concerns about multiple model formulations, we can offer no "cure" for these problems. However, we can illustrate the scale of the problems that are presented. We also suggest that MSEs be augmented by a greater attention to alternative methods to evaluate risks or management strategies, including qualitative methods and empirical *post hoc* analyses.

Implications of simulating distributions

When a probability distribution is used because there is insufficient information (data and/or understanding) to specify the expected value for a parameter, the substitution of a probability distribution paradoxically implies knowing more about the characteristics of the parameter than using a point estimate. To use a deterministic parameter implies knowledge of the value of the parameter. To use the distribution of a parameter acknowledges that this value is a mean and requires additional knowledge of either its variance and skewness or of a theoretical distribution and its parameters. Alternatively, non-parametric data-based methods such as bootstrapping or density estimation can be used to avoid specifying a theoretical sampling error distribution for the simulation, or the specification of a parameterized functional relationship (Evans and Rice, 1988; Rice, 1993). These approaches are limited to data-rich situations, and place the problem of dealing with uncertainty from infrequent events or inherently variable measurements even more centrally in the analyses.

If there is insufficient information to estimate the mean of a distribution reliably, there is little reason to expect to be able to estimate its variance. This can be illustrated with any biological dataset, e.g. from research survey data. When the coefficients of variation for average length and variance in length for fish taken in a series of different research surveys are plotted (Figure 1), the greater uncertainty in variance estimates is obvious. Although the surveys differed somewhat in the precision of their estimates of the mean length, the estimates of variance in length always had much higher *CV*s. Attempts to avoid the estimation of the variance or skewness of a set of observations by fitting data to a theoretical distribution does not help. Survey abundance data have extremely low statistical power to differentiate among common theoretical distributions such as Gamma, normal, and lognormal (Myers and Pepin, 1990; Trenkel and Rochet, 2003), and the same is true for estimates of SSB (Figure 2). For all stocks examined, depending on which assumption was made about the theoretical distribution from which the SSB estimates were drawn, even advice based on the modal range of estimates could be substantially different (Figure 2). The differences in likelihood are very small, illustrating the low power to choose among the distributions, yet the probabilities of low abundances calculated from these theoretical distributions are quite different

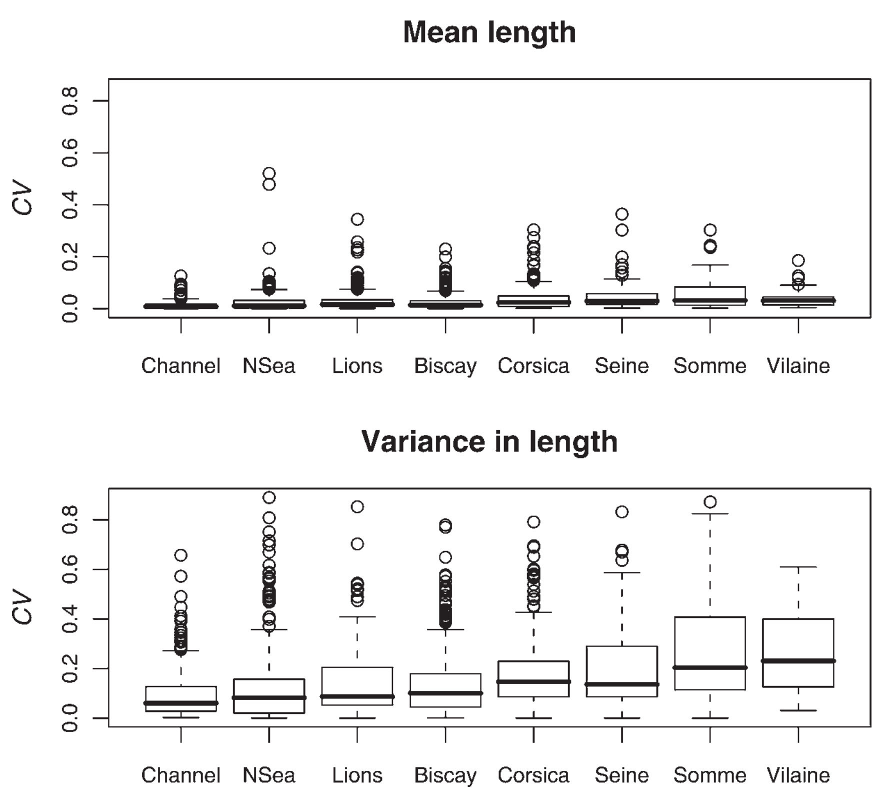

Figure 1. Comparison of the coefficients of variation (*CV*s, standard error divided by mean) of average length (top) and variance in length (bottom) for species caught in various French bottom-trawl surveys. Each data point = *CV* of a species × year estimate of the metric (from left to right: English Channel 1988–2005, $n = 20$ species; southern North Sea 1983–2005, $n = 16$; Gulf of Lions 1994–2004, $n = 27$; Bay of Biscay 1992–2004, $n = 54$; East Corsica 1994–2004, $n = 24$; Bay of Seine 1995–2002, $n = 14$; Bay of Somme 1995–2004, $n = 18$; Bay of Vilaine 2000–2004, $n = 13$). For details of the surveys, see Rochet *et al.* (2005). The heavy line represents the median, the boxes the interquartiles, and individual points are outliers.

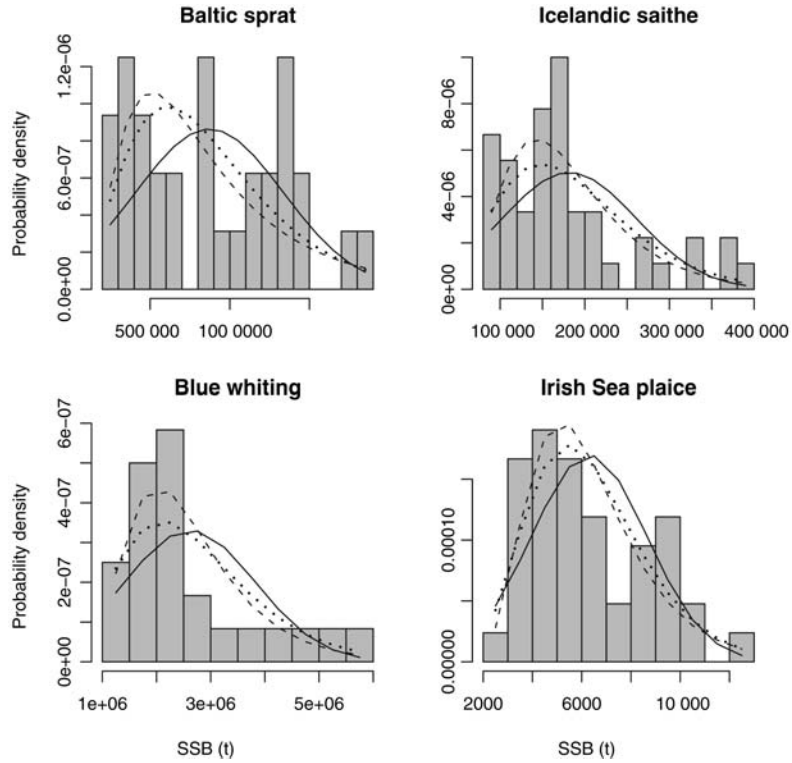

Figure 2. Example of empirical distributions of SSB for Baltic sprat (1974–2005), Icelandic saithe (1962–2005), blue whiting (1981–2004), and Irish Sea plaice (1964–2005). Fitted distributions: normal (solid), Gamma (dotted), lognormal (dashed).

(Table 1). A typical MSE work would of course not use a specified distribution for SSB but rather use "true" population numbers and then go through some deterministic model (e.g. virtual population analysis, VPA) to estimate SSB and draw inferences from the percentiles of this Monte Carlo SSB series. This may amplify the problem we outline because (i) true population numbers cannot be observed at all, so their distribution is still less well known, and (ii) errors propagate through the calculations in ways that are not easy to quantify (Pelletier and Gros, 1991).

Research survey data are often a "best case" scenario for fisheries modelling, because they can include several hundred observations to help tie down the shape of the tails of the distributions. When few data are available, common practice is that because a lognormal distribution cannot be rejected, it can be used in simulations. In such cases, even the estimates of standard deviation for the distribution can be based on a median or central value from a literature review (e.g. Punt, 1997), rather than stock-specific data. Hence, many simulations rely on parameter distributions of which neither the shape nor the dispersion are really known.

Use of data-based bootstrap methods is not a real alternative to providing reliable quantitative support for risk-averse decision-making when there is limited contrast in the data available, when few data come from the area of interest for the management decision (e.g. for stocks that have not yet been in the neighbourhood of B_{lim}), or when there might have been different causes of past stock declines. When using resampling procedures with replacement, and weighting each sampled observation equally, the uncertainty in the region where observations are most clustered dominates the bootstrapped uncertainty, because several samples will be drawn from the region where observations are most dense (often close to the centre of the distribution) for each occasion a sample is drawn from the rarer observations in the tails. As a consequence, the frequency of resampled units is increasingly biased to lie below the true percentile in the observed sample for lower percentiles (Figure 3). That is, bootstrapping underestimates the probabilities of low values, so also underestimates the risks of undesirable outcomes. If the bootstrap selection of observations is designed to sample preferentially in the region of interest (for precautionary management, in the location where

Table 1. The probability of SSB falling below B_{lim} (for plaice, B_{pa}) under three assumed distributions fitted to the same datasets (Figure 2), and goodness-of-fit (log-likelihood) of these distributions.

	$p(\mathrm{SSB} < B_{\mathrm{lim}})$			Log-likelihood		
Species	**Normal**	**Lognormal**	**Gamma**	**Normal**	**Lognormal**	**Gamma**
Sprat	0.073	0.017	0.024	−462.8	−462.1	−461.3
Saithe	0.123	0.060	0.095	−571.4	−563.6	−565.6
Blue whiting	0.177	0.128	0.166	−370.1	−365.2	−366.6
Plaice	0.087	0.040	0.059	−385.6	−382.0	−382.7

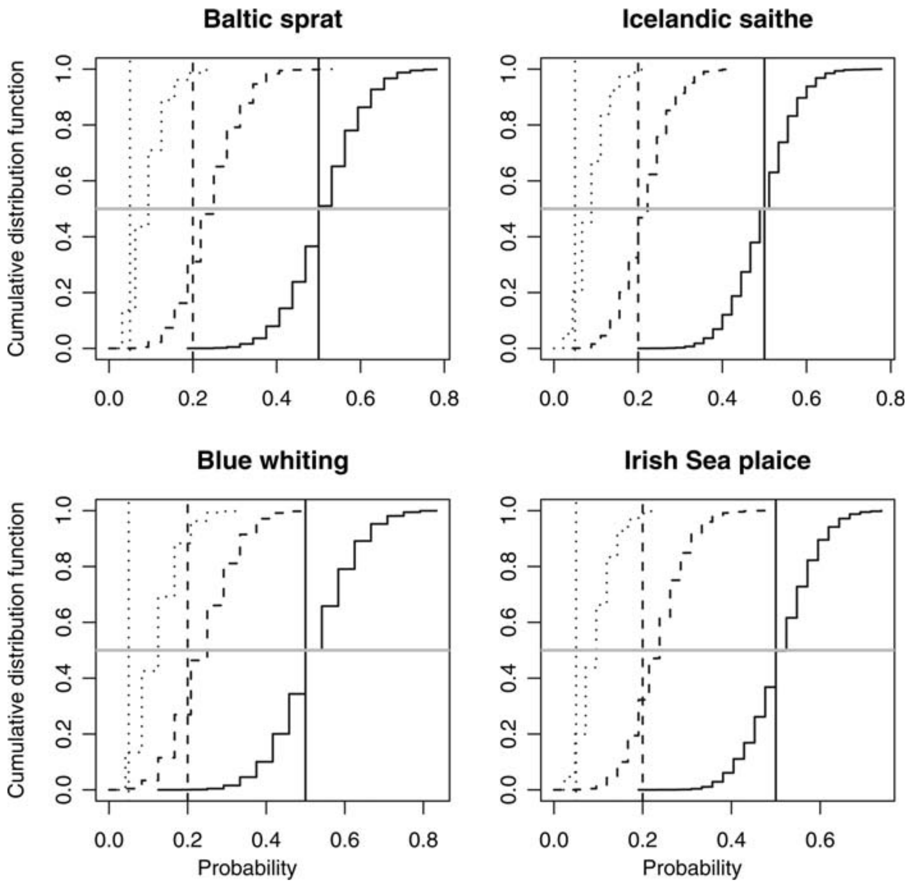

Figure 3. Cumulatived distributions of the probability of being under the 0.05, 0.2, and 0.5 percentiles of the original data. The SSB time-series used in Figure 2 were resampled 5000 times (non-parametric ordinary bootstrap in R, http://www.r-project.org/). The proportion of numbers in each sample lower than the median of the original sample was retained; the cumulative probability distribution function (CDF) of this proportion is shown (continuous line). If the bootstrapping procedure was unbiased, this should be on average 0.5, i.e. it should cross the vertical 0.5 (continuous) line approximately in the middle (near the horizontal grey line). The same reasoning applies to the 0.2 (dashed lines) and 0.05 percentiles (dotted lines). Clearly, the bootstrapping procedure is increasingly biased when one moves from central to extreme percentiles: the distance between the middle of the CDF and the target percentile increases.

$B_{\rm lim}$ is thought to lie), for stocks that have not collapsed, the probabilities are based on much less information than using the bootstrap methods to estimate probabilities of more commonly observed stock sizes.

The same types of problem arise when non-parametric density estimation methods are used to avoid specifying a functional form for poorly known relationships. When locally weighted smoothing methods are applied near the extremes of the range of observations, the few observations of the extreme events dominate the estimated probability distribution. Gaps between observations near the extremes of the range that are attributable to low sampling in the tails are treated as genuine gaps in the observations that are possible, unless very wide smoothing windows are used, then the probabilities of the events in the tails become largely determined by the more common observations near the centre of the distribution of observations. Once decision support requires estimating the probability of events beyond the range of historical observations, the situation becomes even more subjective. The estimates of probabilities are increasingly determined by the assumption of what value the property converges to at zero or infinity on the independent variable. That assumption could be convergence to the overall mean, to zero, or a linear trend from the last observations—and not any of the observations (Silverman, 1992). Hence, data-based methods for estimating both expected values and their uncertainty can also provide a spurious sense of quantitative rigour when supporting decisions about the likelihood of extreme or otherwise rare events—a form of support commonly requested by managers.

Other ways of dealing with uncertainty

Several other approaches have been or could be used to evaluate management strategies and usefully complement simulation-based methods. Here, we scan some of them with a view to illustrating various avenues rather than offering an exhaustive review or a preferred choice.

First, one can learn from experience and examine performance of management strategies, or of some of their pieces, based on how they succeeded or not, and the reasons for their success or failure (e.g. Hilborn, 2006). Because estimates by some stock assessment methods are more reliable at the beginning of the data time-series than in the most recent years, the performance of these methods at forecasting catch can be retrospectively examined (Jónsson and Hjörleifsson, 2000; Reeves and Pastoors, 2007). Retrospective analysis of the performance of the subsequent advice in terms of the relevance of the advised action against true status of the stock is also informative (Piet and Rice, 2004). Analysis of time-trends in indicators can also reveal performance of a management strategy. For example, Sparholt *et al*. (2007) showed that the demersal stocks evaluated by ICES have declined continuously since the 1950s and that neither the Common Fisheries Policy introduced in 1983 nor the precautionary approach implemented by the late 1990s had a detectable effect on trends in fishing mortality or stock biomass.

Retrospective examination of management strategy performance will be much more powerful if done in a comparative manner, i.e. comparing different management strategies across regions or fisheries or stocks (Patterson and Résimont, 2007; Dankel *et al*., 2008) or across periods with different biological or management regimes for a given stock (Simmonds, 2007). When the outcome of a particular strategy for the management of a stock is examined, the reasons for success or failure might be diverse and confounded. Bringing together outcomes of different strategies implemented in different settings, and applying explicit criteria to define successes and failures, the method will point to the characteristics of those management strategies which worked, against those of the strategies which failed. Much can be learned about the selection of cases and traits and appropriate comparison methods from the comparative method that has been developed and widely used in evolutionary biology (Harvey and Pagel, 1991).

As the experience with formalized management procedures increases, these retrospective empirical methods will be increasingly informative. However, even before any implementation, the consistency of a proposed method can be checked. One type of consistency check is simply the internal consistency of what is said and what is done. If a management strategy is not internally consistent, it may be more prone to failure than another consistent one. For example, Hauge *et al*. (2007) compared the definitions of limit and precautionary reference points in the ICES precautionary approach, and the way they are used in practice, and found that their apparent transparency is compromised by the difficulty of estimating them in a standardized way across a diversity of stocks. Another consistency check consists in using formal mathematical analysis to test if a rule really provides the expected results under ideal conditions. For example, De Lara *et al*. (2007) showed by analysing a dynamic stock model that in the ICES precautionary approach, for most stocks the objective of keeping SSB biomass above $B_{\lim}$ cannot be guaranteed in the long term just by keeping it above a limit from year to year: additional indicators of stock structure are required.

Qualitative modelling (e.g. Eisenack and Kropp, 2001) provides an attractive way of modelling vague knowledge to provide decision support. Rather than attempting to provide arrays of formulations and parameter distributions for processes that are not well known, this approach allows a formal analysis of the consequences of qualitative assumptions, such as that "catch is a dome-shaped function of effort" or "effort increased during the period 1960–1990". A combination of formal derivation and "qualitative simulation" yields system trajectories that can be categorized in groups such as "stabilization at a low level" or "sustainable use" depending on these qualitative assumptions. Similarly, qualitative analysis of dynamic models, e.g. loop analysis and shifts in equilibrium caused by press perturbations, allows one to predict the direction (not the amount) of change in variables such as stock abundance or catch induced by changes in parameters or other variables such as effort or price (Dambacher *et al*., 2002, 2003). For example, in a multifleet multispecies fishery targeting two trophic levels, decreasing catchability of piscivores will lead to an increase in their abundance, whether or not there is omnivory or technical interaction, implying that a number of complex science questions about those interactions do not have to be resolved before the right type of management action can be selected (Dambacher *et al*., in press). This can obviously be extended to examine which directions of change will be robust irrespective of model structure; when decreasing catch or increasing taxes is likely to have the expected consequences, before undertaking a costly model parameterization and simulation work.

Discussion

Without question, the MSE approach is a significant step forward in fisheries science. It has all the benefits of any modelling exercise and provides a tool to help make the precautionary approach operational. However, we are seeing it being used as if small differences in analytical results are meaningful biologically and in management. For the reasons we illustrate above, it is inherently impossible for MSE (at least with current knowledge of marine ecological processes, quality of data, and conflicts over implementation) to provide the level of quantitative accuracy and precision needed to support making finely differentiated management decisions. For example, if the goal is to rebuild a depleted stock to a pre-selected biomass level in 10 years, when knowledge of stock productivity at low biomass is highly uncertain, MSE cannot be used reliably to estimate that the goal can be reached with a particular F, but not with an F 20% higher, for example. Simulations can provide results which present themselves as having that level of precision, but it is a deception. The implied precision may not deceive those performing the evaluations, but it is often a deception that users of the results at least may exploit, whether they are believed or not.

Our alternative is first of all to obtain a general admission that it is unrealistic for managers and policy experts to expect reliable, finely nuanced discriminations from simulations. Fisheries scientists have spent 15 years admitting first to themselves and then trying to convince managers and policy-makers that they cannot provide science support for annually adjusting Fs by 10% and TACs by a few thousand tonnes. We do not want to start building a culture where either producers or users of MSEs think our tools can support discriminations of comparable precision when the analyses have to include even more processes and datasets—many more poorly specified than in traditional single-species assessments estimating only the most likely B and F each year. In other fields where risk analysis is well developed, such as nuclear power safety or cost analysis of regulatory intervention, risk estimates are sometimes reported with an excessive precision that is not supported by the available data, and they can lead to rankings of options that are not robust. Unfortunately, overly precise figures are picked up by users because they give confidence in the analyst's knowledge (Hassenzahl, 2006).

A useful practice might be to report with any MSE result the number of parameters that were actually estimated from data relevant to the situation, the number that were borrowed from the literature, and the number guessed or tuned. This could be done separately for the parameters that determine the system state and dynamics, and for those that characterize the uncertainty. Similarly, practice could include reporting comparable information about the sources and strength of support for the functional relationships included, relative to alternatives that were not. This would provide users with some appraisal of the reliability of the risk estimates.

There are valuable uses of simulation-based MSEs, such as eliminating "bad" management strategies from consideration. A management strategy that performs poorly in the full range of simulations is not likely to perform well in the real world, because the real world is always more complex and unpredictable than the simulated world. However, this does not mean that MSEs are necessarily powerful at discriminating the best approach among ones that are not bad. Simple, adaptive management strategies may not be optimal in the face of the particular forms of uncertainties/changes that were incorporated in the operating model; but in the real world, they might address better the "unknown unknowns": unforeseen changes in ecosystem factors and/or in the economy, reactions of fishers to new management plans, etc.

Overall, MSEs should be used to strengthen the use of intelligence in developing approaches to fisheries management, not as an excuse to remove it from the process. Few operating models 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. For example, Punt (1997) in his evaluation of VPA-based management, used the Laurec–Shepherd estimation procedure rather than ADAPT, because the latter requires making educated choices that cannot be simulated in an operating model. Moreover, often, choices of one player can be influenced by conjectures of what choices other players may make (Rice and Richards, 1996), and this might be difficult to simulate. Particularly because another key change going on in fisheries management is making the governance process more inclusive (FAO, 2003; DFO, 2004; EU, 2004), this is the wrong time to remove intelligence from the procedure used to evaluate the process. If all their quantitative details are to be taken as reliable, then simulation-based MSEs can only evaluate impoverished procedures, those without the need of human intervention. This is not the type of management procedure we want.

Bayesian approaches (Punt and Hilborn, 1997) amplify all our concerns about treating quantitative details as more meaningful than deserved by the quality and quantity of either the data or the theory on which the details depend. Bayesian approaches are becoming increasingly popular in many fisheries applications, in part because they allow greater use of the information that is available, differentially emphasizing datasets based on quantitative measures of their information content. However, they are still vulnerable to the concerns that we raise here: model misspecification and underestimated errors are pervasive in Bayesian analysis and lead to false precision estimates (Small and Fishbeck, 1999). More data and sophisticated techniques can improve this problem but not solve it. Instead, MSEs should not oversell their ability to differentiate among strategies that perform well enough to be considered relevant to management decisions.

It has long been argued that a major benefit of MSEs is not technical development and quantitative output, but greater opportunity for stakeholder participation and acceptance (Cochrane *et al.*, 1998; Smith *et al.*, 1999). The process of developing an MSE forces participants to agree about management objectives, and to think about the various components of the system, unknowns that might influence the success of a management strategy, appropriate criteria to evaluate a management strategy, and how to weight the various considerations. Agreeing completely with the value of such careful thinking about the management problem, and wide engagement of parties in the process of developing and implementing management strategies, we notice that few MSE publications are reporting how users were involved in the process (but see Cochrane *et al.*, 1998; Pastoors *et al.*, 2007). On the other side of the coin, we contend that progress on those fronts does not require that everything be translated into equations and distributions for this progress to be made. For example, qualitative risk evaluation (Fletcher, 2005) basically does the same job in a transparent and shared way.

Simulations have important roles in supporting management decision-making. These include finding out what really does not work under any plausible circumstances and avoiding it, and providing a process to define the management problem comprehensively and consolidate knowledge as much as possible. The quantitative results certainly serve as one source of information to support decision-making. However, they are not a substitute for applying our full intelligence in the decision support. The simulations might really tell little more than the general magnitude of the changes to be expected from a particular set of actions and the qualitative scale of the management intervention that is needed. Then it is necessary to bring in all the other sources of knowledge to first get confirmation that the scale of intervention really is needed, and second to obtain some insight into which actions are most likely to succeed in making the intervention large enough to achieve the management goals. We know from experience that implementation uncertainty is very likely to greatly exceed any nuanced quantitative differences among simulation results. Therefore, processes that are simple and adaptive, but have high compliance, are likely to be appropriate most of the time, however well structured processes may perform in MSEs.

Acknowledgements

We thank our colleagues who attended the International Symposium on Fisheries Management Strategies in Galway, Ireland, in June 2006, for inspiring discussions about MSEs. Verena Trenkel and anonymous reviewers made useful comments on earlier drafts.

References

Berkson, J. M., Kline, L., and Orth, D. J. (Eds). 2002. Incorporating Uncertainty into Fishery Models. American Fisheries Society Symposium, 27. 196 pp.

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 Punt, A. E. 2003. The role of harvest control laws, risk and uncertainty and the precautionary approach in ecosystem-based management. *In* Responsible Fisheries in the Marine Ecosystem, pp. 311–319. Ed. by M. J. Sinclair, and H. Valdimarson. FAO, Rome.

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.

CSAS. 2006. Assessment of the cod stock in northern Gulf of St Lawrence (3Pn4RS) in 2005. http://www.dfo-mpo.gc.ca/csas/Csas/status/2006/SAR-AS2006_010_E.pdf.

Dambacher, J. M., Gaughan, D. J., Rochet, M-J., Rossignol, P. A., and Trenkel, V. M. Qualitative modelling and indicators of exploited ecosystems. Fish and Fisheries, in press.

Dambacher, J. M., Li, H. W., and Rossignol, P. A. 2002. Relevance of community structure in assessing indeterminacy of ecological predictions. Ecology, 83: 1372–1385.

Dambacher, J. M., Li, H. W., and Rossignol, P. A. 2003. Qualitative predictions in model ecosystems. Ecological Modelling, 161: 79–93.

Dankel, D. J., Skagen, D. W., and Ulltang, Ø. 2008. Fisheries management in practice: review of 13 commercially important fish stocks. Reviews in Fish Biology and Fisheries, 18: 201–233.

De Lara, M., Doyen, L., Guilbaud, T., and Rochet, M-J. 2007. Is a management framework based on spawning stock biomass indicator sustainable? A viability approach. ICES Journal of Marine Science, 64: 761–767.

DFO. 2004. Atlantic Fisheries Policy Review: a Policy Framework for Management of Fisheries on Canada's Atlantic Coast. http://www.dfo-mpo.gc.ca/afpr-rppa/Doc_Doc/policy_framework/policy_framework_e.htm.

Eisenack, K., and Kropp, J. 2001. Assessment of management options in marine fisheries by qualitative modelling techniques. Marine Pollution Bulletin, 43: 215–224.

European Union. 2004. Council Decision of 19 July 2004 establishing Regional Advisory Councils under the Common Fisheries Policy (2004/585/EC). Official Journal of the European Union, L 256: 17–22.

Evans, G. R., and Rice, J. C. 1988. Predicting recruitment from stock size without the mediation of a functional relation. Journal du Conseil international pour l'Exploration de la Mer, 44: 111–122.

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

FAO. 2003. The ecosystem approach to fisheries. FAO Technical Guidelines for Responsible Fisheries, 4(Suppl. 2). FAO, Rome. 112 pp.

Fletcher, W. J. 2005. The application of qualitative risk assessment methodology to prioritize issues for fisheries management. ICES Journal of Marine Science, 62: 1576–1587.

Garcia, S. M. 2005. Fishery science and decision-making: dire straights to sustainability. Bulletin of Marine Science, 76: 171–196.

Harvey, P. H., and Pagel, M. D. 1991. The Comparative Method in Evolutionary Biology. Oxford Series in Ecology and Evolution. Oxford University Press, Oxford. 239 pp.

Hassenzahl, D. M. 2006. Implications of excessive precision for risk comparisons: lessons from the past four decades. Risk Analysis, 26: 265–276.

Hauge, K. H., Nielsen, K. N., and Korsbrekke, K. 2007. Limits to transparency—exploring conceptual and operational aspects of the ICES framework for providing precautionary management advice. ICES Journal of Marine Science, 64: 738–743.

Hilborn, R. 1997. Uncertainty, risk, and the precautionary principle. *In* Global Trends: Fisheries Management, pp. 100–106. Ed. by E. K. Pikitch, D. D. Huppert, and M. P. Sissenwine. American Fisheries Society Symposium, 20.

Hilborn, R. 2006. Fisheries success and failure: the case of the Bristol Bay salmon fishery. Bulletin of Marine Science, 78: 487–498.

Hodgson, S., Quinn, T. D., Hilborn, R., Francis, R. C., and Rogers, D. E. 2006. Marine and freshwater climatic factors affecting interannual variation in timing of return migration to freshwater of sockeye salmon (*Oncorhynchus nerka*). Fisheries Oceanography, 15: 1–15.

ICES. 2005a. Report of the Study Group on Management Strategies. ICES Document CM 2005/ACFM: 09. 66 pp.

ICES. 2005b. Report of the ICES Advisory Committee on Fishery Management and Advisory Committee on Ecosystems, 2005. ICES Advice, 5. 242 pp.

Jónsson, S. T., and Hjörleifsson, E. 2000. Stock assessment bias and variation analyzed retrospectively and introducing the PA-residual. ICES Document CM 2000/X: 9. 13 pp.

Mueter, F. J., Boldt, J. L., Megrey, B. A., and Peterman, R. M. 2007. Recruitment and survival of northeast Pacific Ocean fish stocks: temperature trend, covariance, and regime shifts. Canadian Journal of Fisheries and Aquatic Sciences, 64: 911–927.

Myers, R. A., and Pepin, P. 1990. The robustness of lognormal-based estimators of abundance. Biometrics, 46: 1185–1192.

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.

Patterson, K., and Résimont, M. 2007. Change and stability in landings: the responses of fisheries to scientific advice and TACs. ICES Journal of Marine Science, 64: 714–717.

Patterson, K. R., Cook, R. M., Darby, C. D., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., *et al.* 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish and Fisheries, 2: 125–157.

Pelletier, D., and Gros, P. 1991. Assessing the impact of sampling error on model-based management advice: comparison of equilibrium yield per recruit variance estimators. Canadian Journal of Fisheries and Aquatic Sciences, 48: 2129–2139.

Peterman, R. M. 2004. Possible solutions to some challenges facing fisheries scientists. ICES Journal of Marine Science, 61: 1331–1343.

Piet, G. J., and Rice, J. C. 2004. Performance of precautionary reference points in providing management advice on North Sea fish stocks. ICES Journal of Marine Science, 61: 1305–1312.

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

Punt, A. E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Reviews in Fish Biology and Fisheries, 7: 35–63.

Punt, A. E., Smith, A. D. M., and Cui, G. 2001. Review of progress in the introduction of management strategy evaluation (MSE) approaches in Australia's South East Fishery. Marine and Freshwater Research, 52: 719–726.

Reeves, S. A., and Pastoors, M. A. 2007. Evaluating the science behind the management advice for North Sea cod. ICES Journal of Marine Science, 64: 671–678.

Rice, J. C. 1993. Forecasting abundance from habitat measures using nonparametric density estimation methods. Canadian Journal of Fisheries and Aquatic Sciences, 50: 1690–1698.

Rice, J. C., and Richards, L. J. 1996. A framework for reducing implementation uncertainty in fisheries management. North American Journal of Fisheries Management, 16: 488–494.

Richards, L. J., and Maguire, 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., Trenkel, V. M., Bellail, R., Coppin, F., Le Pape, O., Mahé, J-C., Morin, J., *et al.* 2005. Combining indicator trends to assess ongoing changes in exploited fish communities: diagnostic of communities off the coasts of France. ICES Journal of Marine Science, 62: 1647–1664.

Silverman, B. W. 1992. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London. 176 pp.

Simmonds, E. J. 2007. Comparison of two periods of North Sea herring stock management; success, failure, and monetary value. ICES Journal of Marine Science, 64: 686–692.

Small, M. J., and Fischbeck, P. S. 1999. False precision in Bayesian updating with incomplete models. Human and Ecological Risk Assessment, 5: 291–304.

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.

Sparholt, H., Bertelsen, 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.

Stokes, K., Butterworth, D. S., Stephenson, R. L., and Payne, A. I. L. 1999. Confronting uncertainty in the evaluation and implementation of fisheries-management systems. Introduction. ICES Journal of Marine Science, 56: 795–796.

Trenkel, V. M., and Rochet, J. 2003. Performance of indicators derived from abundance estimates for detecting the impact of fishing on a fish community. Canadian Journal of Fisheries and Aquatic Sciences, 60: 67–85.

Wood, S. N., and Thomas, M. B. 1999. Super-sensitivity to structure in biological models. Proceedings of the Royal Society of London, Series B: Biological Sciences, 266: 565–570.

doi:10.1093/icesjms/fsp023