THIS ARTICLE WAS PUBLISHED IN 2007 IN THE BLACKWELL JOURNAL "FISH & FISHERIES" (8 315-336). THIS IS THE FINAL VERSION THAT WAS ACCEPTED FOR OPUBLICATION (AUGUST 2007). THE DEFINITIVE VERSION IS AVAILABLE AT www.blackwell-synergy.com

Model uncertainty in the ecosystem approach to fisheries.

Alternative titles:

(1) *Dealing with structural uncertainties in models of exploited ecosystems.*

(2) *Dealing with uncertainties about model structure in ecosystem-based fisheries management*

Simeon L. Hill[1], George M. Watters[2], André E. Punt[3,4], Murdoch K. McAllister[5], Corinne Le Quéré[1,6], John Turner[1]

[1]British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge, CB3 0ET, UK; [2]Southwest Fisheries Science Center, NOAA Fisheries, Protected Resources Division, 1352 Lighthouse Avenue, Pacific Grove, California 93950: [3]USA School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195-5020, USA; [4]CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart, TAS 7001, Australia; [5]Fisheries Centre, University of B.C., AERL, 2202 Main Mall, Vancouver, B.C., V6T 1Z4, Canada; [6]School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK.

Correspondence: Simeon Hill, British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge, CB3 0ET, UK

Tel.: + 223 221233 Fax: + 223 221259 Email: sih@bas.ac.uk

Abstract

Fisheries scientists habitually consider uncertainty in parameter values, but often neglect uncertainty about model structure. The importance of this latter source of uncertainty is likely to increase with the greater emphasis on ecosystem models in the move to an ecosystem approach to fisheries (EAF). It is therefore necessary to increase awareness about pragmatic approaches with which fisheries modellers and managers can account for model uncertainty and so we review current ways of dealing with model uncertainty in fisheries and other disciplines. These all involve considering a set of alternative models representing different structural assumptions, but differ in how those models are used. The models can be used to identify bounds on possible outcomes, find management actions that will perform adequately irrespective of the true model, find management actions that best achieve one or more objectives given weights assigned to each model, or formalise hypotheses for evaluation through experimentation. Data availability is likely to limit the use of approaches that involve weighting alternative models in an ecosystem setting, and the cost of experimentation is likely to limit its use. Practical implementation of the EAF should therefore be based on management approaches that acknowledge the uncertainty inherent in model predictions and are robust to it. Model results must be presented in a way that represents the risks and trade-offs associated with alternative actions and the degree of uncertainty in predictions. This presentation should not disguise the fact that, in many cases, estimates of model uncertainty may be based on subjective criteria. The problem of model uncertainty is far from unique to fisheries,

and coordination among fisheries modellers and modellers from other communities will therefore be useful.

Keywords Bayesian methods, Ecosystem Approach to Fisheries, ecosystem models, fisheries management, model uncertainty, operational management procedures.

Abstract.....2

Introduction.....4

Model uncertainty in ecosystem dynamics models......5

Approaches to Addressing Model Uncertainty I: Adaptive management.....8

Approaches to Addressing Model Uncertainty II: Robust Management.....9

The CCAMLR approach.....11

The International Whaling Commission approach.....14

Similarities and differences.....18

Approaches to Addressing Model Uncertainty III: Model Weighting.....18

Applications of the Bayesian approach to weight alternative models.....22

Approaches to Addressing Model Uncertainty IV: Examples from related disciplines.....26

Presenting results.....31

Discussion **17**

Limitations and pragmatic approaches.....33

Improving models and establishing protocols.....37

Acknowledgements.....39

References.....39

Introduction

Implementing the ecosystem approach to fisheries (EAF: Garcia *et al.* 2003) will require, *inter alia*, stakeholder agreement about management objectives, and management systems designed to achieve these objectives. In many cases, EAF will also require models that predict the effects of human activities in terms of these objectives. Unfortunately, marine ecosystems are structurally complex, spatially and temporally variable, and difficult and costly to observe, all of which lead to considerable uncertainty in model predictions. Fisheries scientists have been at the forefront of attempts to account for uncertainty in the management of living resources (Patterson *et al.* 2001; Harwood and Stokes 2003). However, effort has largely focused on uncertainty in parameter values and the process uncertainty that arises from natural variation, whereas uncertainties about model structure have received less attention. For example, Halpern *et al.*'s (2006) review of methods to evaluate rules for spacing marine reserves given different levels of uncertainty did not consider uncertainties in model structure.

Neglecting model uncertainty can lead to under-representation of uncertainty in model predictions, with important implications for management. For example, management that aims to minimise the risk of particular outcomes or to be robust to uncertainty might require very different management actions as the perceived level of uncertainty changes. Fishery modellers and managers must therefore be well-equipped to deal with the various uncertainties that can influence model predictions. Our aim is to review approaches to handling model uncertainty and to suggest practical ways to deal with it in developing the EAF. In the next section we define model uncertainty and highlight its importance in ecosystem dynamics models. In the following sections, we review approaches for addressing model uncertainty, including

approaches adopted in other disciplines attempting to make predictions about complex systems. Finally, we discuss the presentation of results that incorporate uncertainty, pragmatic approaches to dealing with model uncertainty in the short term, and strategies for reducing model uncertainty over the longer term.

Model uncertainty in ecosystem dynamics models.

There are many sources of uncertainty in resource dynamics models, and these have been classified in various ways by Francis and Shotton (1997), Charles (1998), Reagan *et al*. (2002), Harwood and Stokes (2003) and Mangel (2006). Our focus is on "model uncertainty" which arises because any single, apparently satisfactory, model can be misleading. For example, Fig. 1 shows a linear model relating recruitment of a commercially exploited fish species to sea-surface temperature. The model predicts negative recruitment at temperatures below 2˚C (Hill *et al*. 2005) and is clearly unsuitable for predicting the response to temperatures beyond a limited range.

Although model uncertainty largely concerns model structure, it can also include uncertainties about the values of parameters that determine a model's behaviour, but are not easily estimated. For example, the parameter *M* in Virtual Population Analysis (VPA) attempts to summarise all the sources of natural mortality. In contrast, multispecies VPA estimates predation mortality due to other modelled species and therefore relaxes the assumptions inherent in the value of this parameter (see Magnússon 1995). Other models are formulated so that their behaviour can be changed by altering a key parameter (such as a shape parameter in a functional relationship; Yodzis 1994). Model uncertainty also overlaps with process uncertainty as it can also concern the shape of the error distribution (e.g. normal versus lognormal; Halley and Inchausti 2002).

Models are mathematical abstractions of non-mathematical processes and there will be uncertainty about whether a given model structure is an appropriate representation of a real system. According to Patterson *et al*. (2001), fisheries scientists must give "more attention … to examining the sensitivity to alternative assumptions and model structures." Predictive models, especially in ecology, are rarely intended to provide an all-encompassing description of how a system actually works, but they are intended to forecast how certain characteristics of the system respond to specific sets of conditions. It is therefore more important that a model captures the response of interest than that it specifies the exact mechanisms producing this response. Critically, it must do this for the range of plausible input values.

There are numerous models that aim to predict the dynamics of exploited marine ecosystems. Plagányi (2007) reviews twenty such models that, in order of decreasing complexity, are broadly categorized as 1) "whole ecosystem models" that attempt to account for all trophic levels, e.g., Ecopath with Ecosim (EwE: Christensen and Walters 2004); 2) "dynamic system models" that represent both bottom-up and top-down processes, e.g., "Bay Model 2" (Fulton *et al*. 2004); 3) "dynamic multi-species models" (also known as "minimally realistic models") that represent a limited number of species with important interactions, e.g., GADGET (http://www.hafro.is/gadget, accessed 19th March 2007); and 4) "extended single-species assessment models" that take a few interactions into account, for example by treating predators as additional fisheries (e.g. Hollowed *et al*. 2000).

Ultimately, all ecosystem dynamics models simplify the structure of the food-web, the nature of ecological interactions, and the demographic structure of populations. Predictions arising from such models are likely to be particularly uncertain. The complexity of the systems leads to complexity in models and a plethora of potential

assumptions, but there are often insufficient data to evaluate models of high complexity. However, as Plagányi (2007) notes, the treatment of uncertainty in such models "has lagged unsatisfactorily behind other aspects of model development"

A number of applications using EwE exemplify the consequences of model uncertainty in ecosystem modeling. Mackinson *et al*. (2003) showed how six alternative hypotheses about the functional response of cetaceans to variations in the availability of their main prey led to different outcomes under simulated harvest regimes for both the cetaceans and the prey. Watters *et al*. (2003) explored how a pelagic ecosystem might respond to climate forcing under two different hypotheses about how physical and ecological processes in the system are linked. Both studies identified results that were robust to the alternative hypotheses and others that were sensitive. For example, Mackinson *et al*. (2003) found that intense fishing on the prey always had a longer lasting, negative, impact on cetaceans than direct removals of cetaceans themselves, while depleted cetacean populations recovered more slowly as limitations to foraging were increased. Watters *et al*. (2003) found that climate trends caused changes in the biomass of middle and upper trophic level species. However, the direction of such changes was sensitive to assumptions about the relative strength of direct and indirect climate effects on these species. Pinnegar *et al*. (2005) used EwE to explore the implications of nine alternative models of the same food-web that differed in the number of functional groups included and which weak predator-prey links were considered in the model. They showed that Ecosim predictions about food-web responses to, and recoveries from, intense pulses of fishing mortality are very sensitive to model complexity and "taxonomic bias" (where functional groups are aggregated or disaggregated on the basis of particular interest in mammals, fishes, or invertebrates).

The developers of widely distributed software packages for ecosystem modelling have included functionality that allows some uncertainties to be addressed. For example, EwE includes the Ecoranger routine to find alternative parameter combinations that satisfy Ecopath's mass-balance assumptions, and the sensitivity of Ecosim models to initial biomass estimates can be assessed using Monte Carlo simulation (Christensen and Walters 2004). GADGET allows the analyst to select and compare different forms of key functions (Begley 2005). Despite the availability of such tools, the practice of considering model uncertainty in implemented ecosystem models appears to be relatively rare.

Approaches to Uncertainty I: Adaptive management

Adaptive or experimental management seeks to reduce uncertainty through experimentation. This approach identifies candidate management actions, which are then implemented as experimental treatments in the exploited system (Walters and Hilborn 1976). The role of alternative structural models in this approach is to identify appropriate management actions and to represent the alterative hypotheses about the system. It is not necessary to fully characterize uncertainty in model predictions, just to determine whether a candidate management action could credibly achieve the desired management objectives given plausible hypotheses about the operation of the system. The credibility of alternative hypotheses is determined through the analysis of experimental results. Examples of experimental management suggest that it will not completely eliminate model uncertainty, and Walters and Martell (2004), who discuss the use of adaptive management in EAF in detail, conclude that the risks and investment involved make it economically or socially unacceptable in many cases. Furthermore, adaptive management cannot easily consider the full range of uncertain

processes (e.g. climate change and regime shift). Nevertheless, in some cases, adaptive approaches can help to weight hypotheses, thus improving the characterization of uncertainty. We provide two examples of this (Sainsbury 1988; Sainsbury *et al.* 1997; McAllister and Kirchner 2002) in a later section.

Approaches to Uncertainty II: Robust Management

A pragmatic approach to uncertainty in general, that is particularly relevant to model uncertainty, is to identify management actions that perform adequately (robustly) across the plausible range of ecosystem structures and parameter values (e.g. Ludwig *et al*. 1993; Charles 1998). The shift in focus from identifying "optimal" management actions, as implied by concepts such as maximum sustainable yield (MSY), to "adequate" management actions acknowledges both the trade-offs and the uncertainties that are inherent in managing complex systems. It is unlikely that all possible management objectives (e.g. profitability, social benefits and the potentially conflicting requirements of different natural components of the ecosystem) can be optimised simultaneously or that all model structures would suggest the same optimum.

The concept of robust management is embodied in Operational Management Procedures (OMPs: Kell *et al*. 2006). This approach, as defined by Punt and Donovan (2007), has seven distinct steps:

1. Identify and prioritize the management objectives (usually specified in national and international laws, standards and agreements) in qualitative terms.
2. Translate these qualitative objectives into quantitative performance measures. The number of performance measures should be kept small. Miller (1956)

suggested that the human mind might be able to simultaneously compare no more than about seven pieces of information. Our own experience is that when decision makers are confronted with a large number of performance measures they tend to focus on a small number of "important" ones anyway. Therefore seven seems to be a reasonable upper limit for the number of performance measures that we can expect to be given equal consideration.

3. Develop and parameterize 'operating' models representing plausible dynamics of the resource or system. Model uncertainty is dealt with by including alternative structural assumptions in the set of operating models.
4. Identify candidate management procedures (rules for analyzing data and determining management actions) and monitoring strategies.
5. Simulate the future performance of each management procedure by applying it to the resources or systems represented in the operating models.
6. Summarize the performance of each management procedure in terms of the performance measures identified in step 2.
7. Identify the management procedure most likely to meet the management objectives.

The OMP approach is exemplified by the development of the International Whaling Commission's (IWC's) management procedures for whale stocks. These management procedures are based on single-species population dynamics models, but they were tested using operating models that considered some of the effects of ecosystem variability. More recently, the Commission for the Conservation of Antarctic Marine Living Resource (CCAMLR) has begun a programme of work to identify management procedures for aspects of the management of krill that are robust to

uncertainties about the current status of the krill stock and its interactions with predators and the fishery. This section explores these two cases in more detail.

The CCAMLR approach

The CCAMLR is responsible for managing fisheries in the Southern Ocean. Its objectives include facilitating the "rational use" of marine resources; the maintenance of harvested populations at levels that ensure stable recruitment; the restoration of depleted species to these levels; the maintenance of ecological relationships between harvested, dependent and related species; and the prevention of changes to the ecosystem that are not potentially reversible within two to three decades (Constable *et al*. 2000). The last two objectives are the basis for an EAF.

The CCAMLR's implementation of an EAF in its main fishery for Antarctic krill (*Euphausia superba*, Euphausiidae) includes setting precautionary catch limits intended to account for the needs of predators, and the development of a monitoring programme that uses krill predators to detect the effects of the fishery (Constable 2004). Precautionary catch limits are determined for, and applied to, large areas designed for reporting catch statistics (e.g., FAO statistical sub-areas). In addition, the CCAMLR recognises the need to subdivide the precautionary krill catch limit for the Scotia Sea-Antarctic Peninsula region among small-scale management units (SSMUs) to minimise possible local effects on krill predators (Hewitt *et al.* 2004a). At present the effects of krill harvesting on krill predators cannot be quantified since current catches are relatively small compared to the precautionary limit, with no observed effects. It is therefore particularly important to account for model uncertainty when evaluating ways to subdivide the full krill catch limit. The CCAMLR has adopted various principles of the OMP approach in its ongoing work to address this problem. A number of operating models have been developed, but the quantification of

management objectives is incomplete. Furthermore, the candidate management actions being considered at present are not management procedures as generally understood (i.e., regular stock assessments that are used to revise catch or effort limits according to a harvest control rule). Rather, they are options for subdividing a catch limit estimated from a separate stock assessment (e.g., Hewitt *et al*. 2004b). However, regular reassessments of the data used to provide the basis for such subdivisions are being actively considered.

At a recent CCAMLR workshop, two ecosystem dynamics models (with acronyms KPFM and SMOM) were used to evaluate six candidate options for subdividing the precautionary krill catch limit in the Scotia Sea-Antarctic Peninsula region (see Hewitt *et al*. 2004a) (SC-CAMLR 2006; Plagányi 2007). The evaluations considered potential effects on the krill stock, predator populations breeding within SSMUs, and fishery performance. Both models represented the krill population and several predator populations in each SSMU, as well as the movement of krill among SSMUs. There were, however, important differences between the KPFM and SMOM, including the number of predator populations explicitly modelled in each SSMU, the main source of stochasticity in the system, the mechanism by which krill availability affects predator dynamics, and the nature of competition among predator species. The workshop explored three key sources of uncertainty: 1) the functional relationship between krill availability and predator breeding success; 2) the relative influence, on krill dynamics, of advection (both among SSMUs and from outside them) versus local production; and 3) the juvenile and adult survival rates of predators. In all three cases, the workshop agreed on plausible bounds to these uncertainties.

The two models addressed three sources of uncertainty between them, although neither model was formulated to deal with all three simultaneously at the time of the

workshop. With SMOM, the focus was on the 1st and 3rd sources of uncertainty, and a "reference set" of parameters consisting of 3 (lowest plausible value, average value, highest plausible value) * N (number of predator taxa) * P (number of uncertain parameters) parameter combinations was used. With KPFM, the focus was on the 1st and 2nd sources of uncertainty, and four parameter combinations, representing each combination of high versus low transport rates (Hill *et al*. in press) and high versus low sensitivity to krill availability in all predator populations, were used. Multiple stochastic simulations (e.g., to account for recruitment variability) using each parameter combination were performed with both models. The output included trajectories of abundance for each modelled population in each SSMU for each model run. These were plotted either as individual trajectories (Fig. 2) or as 95% probability envelopes. These plots indicate the range of plausible outcomes, but not the probability of any particular outcome. It is possible that more extreme results would have arisen from combinations of high and low settings not considered in these simulations. The use of "reference sets" allows a comprehensive evaluation of multiple parameter combinations, but could easily generate high numbers of parameter combinations (for example, in this case the values for the these parameters for up to 47 predator populations could be allowed to vary independently). Ultimately, the workshop made compromises between a comprehensive evaluation and practicality.

Model output was also presented in terms of trade-off plots showing fisheries performance versus ecosystem performance for each SSMU (e.g., Fig. 3). These plots were generated for a range of performance measures including the mean catch, the spatial distribution of the catch relative to its historical distribution, and the species-specific probability of depletion below or recovery to above predefined thresholds.

Although no attempt has been made to date to quantify the probabilities of alternative models, the approach taken nevertheless allows decision makers to evaluate the trade-offs among potential outcomes.

CCAMLR has not yet completed this work programme and no management decisions have been made on the basis of ecosystem dynamics models. Indeed, the final set of performance measures is yet to be agreed, and there are likely to be other plausible models of the system and other important sources of uncertainty. However, this workshop demonstrated some important principles for dealing with model uncertainty, namely:

1. the formulation of models so that different assumptions about ecosystem structure can be represented through different parameterisations of the same model;
2. the parsimonious attempt to bracket major uncertainties with parameters representing the extreme cases;
3. the use of stochastic simulations to capture other sources of uncertainty (e.g., recruitment variability); and
4. the use of more than one basic operating model.

The IWC approach

The IWC is responsible for the management of commercial and subsistence whaling worldwide. Its objectives for commercial whaling are to ensure that the risks of extinction to individual stocks are not seriously increased by exploitation, and to maintain the status of stocks to allow the highest continuing yield as long as the environment permits (IWC, 1981). The IWC's current approach for commercial whaling, known as the 'Revised Management Procedure' (RMP), was developed

following the ‘moratorium’ on commercial whaling that took effect in 1986. The full management system (the ‘Revised Management System’, RMS) consists of three components: a) a generic management procedure that could be applied to any baleen whale population with known stock structure (referred to as the ‘Catch Limit Algorithm’, CLA), b) the rules used to handle situations in which stock structure is uncertain (which constitute the RMP), and c) other non-scientific issues such as enforcement. The CLA is conservative in the face of uncertainty because it prohibits the harvesting of stocks below 54% of the estimated pre-exploitation level, and because it determines the catch limit as the lower 40.2^{th} percentile of a posterior distribution. Increased uncertainty therefore leads to lower catch limits.

Both generic and case-specific simulation trials have been used by the IWC’s Scientific Committee. During the development of the RMP, simulations based on a “generic” baleen whale stock examined a range of uncertainties. The aim of these simulations was to ensure that the performance of the RMP would be ‘adequate’ (in terms of the trade-off between extinction risk and long-term catch) irrespective of the true dynamics of the population being managed. Although the CLA and RMP are generic and can, in principle, be applied to any stock of baleen whales harvested on their feeding grounds, application of the RMP requires that case-specific (implementation) simulation trials be developed to ensure that performance is adequate for the uncertainties most relevant to each case. The process developed by IWC (2005a) to interpret the results of case-specific trials assigns “plausibility weights” to each of these trials (see next section). A management action (a variant of the RMP related to how catches are allocated spatially) is then selected so that conservation performance is adequate on all simulation trials considered to have ‘high’ plausibility.

Operating models have been developed to consider a wide range of uncertainties (e.g. IWC 1992; Punt and Donovan 2007), arguably more so than for any other marine renewable resource; examples are provided in the following list.

1. Bias in the assessment model on which the CLA is based. Bias arises because the assessment model is structurally different from the operating models (the model underlying the CLA is age- and sex-aggregated, while the operating models are age-, sex- and occasionally spatially-structured). Furthermore, the estimates of abundance used by the CLA, usually from scientific surveys, are sometimes made to be biased in simulations, and this bias may change over time.
2. Changes over time in biological parameters. Simulation trials have been used to examine the implications of carrying capacity and productivity increasing and decreasing over time, in part to mimic the implications of climate change and ecosystem shifts. The CLA assumes that carrying capacity and productivity are time-invariant.
3. Episodic events. Simulation trials have been used to examine the implications of large-scale increases in natural mortality that occur, on average, once every 50 years. Half of the population is assumed to die when such an event occurs.

Perhaps the major source of model uncertainty considered during the development of the RMP relates to the case-specific 'implementation' trials constructed to ensure that performance is adequate for a given species in a region (generally a part of an ocean basin such as the western North Pacific). To date, implementation trials have been developed for southern minke whales (*Balaenoptera bonaerensis*, Balaenopteridae), northern minke whales (*Balaenoptera acutorostrata*, Balaenopteridae) in the western North Pacific and the North Atlantic, and Bryde's whales (*Balaenoptera brydei*, Balaenopteridae) in the western North Pacific. In

contrast to the generic simulation trials used when developing the RMP, most of the parameters of the operating models used in 'implementation' trials are estimated by fitting the operating models to the actual data for the species and region concerned, and the models are spatially-structured.

The approach outlined above was also applied to develop 'Strike Limit Algorithms' (SLAs) for aboriginal subsistence whaling. To date SLAs, have been developed for the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*, Balaenidae) and the Eastern North Pacific stock of grey whales (*Eschrichtius robustus*, Eschrichtiidae) (IWC 2003, 2005b). Although, technically, the way SLAs were selected for these two cases was the same as for commercial whaling, there were some notable conceptual differences. For example, only case-specific trials were used for aboriginal whaling. This is because there are only a small number of aboriginal subsistence operations identified by the IWC, and it is unlikely that more will be 'accepted' (Punt and Donovan 2007), and because the objectives for aboriginal subsistence whaling ('need satisfaction') differ from those for commercial whaling ('catch maximization').

The approach developed by the IWC Scientific Committee has been applied elsewhere. In South Africa, the approach has been adopted in totality and management procedures have been developed for the hake, anchovy, pilchard and rock lobster resources (e.g. Geromont *et al*. 1999). These management procedures prescribe how the data on which scientific management advice is to be based must be collected and analysed, and how Total Allowable Catches are to be determined from those analyses. In contrast to the situation in South Africa, the process of simulating the performance of management procedures has been used in other countries, such as Australia, to guide the selection of the assessment methods that are most likely to

provide robust, accurate and precise estimates of the quantities of interest to management (Punt 2006; Punt *et al*. 2002).

Similarities and differences

For both the IWC and CCAMLR cases, the key outputs from the analyses are based on different models related to trade-offs between conservation- and use-related objectives. The IWC used an initial set of simulations to identify a generic RMP that is robust to a range of uncertainties, followed by more specific simulations to establish the trade-offs associated with uncertainties about stock structure and therefore to identify variants of the generic RMP that perform adequately for specific regions and species. The CCAMLR sets regional krill catch limits with a precautionary single-species model and is using ecosystem dynamics models to identify robust ways of subdividing this limit. However, while the IWC has attempted to assign weights, albeit largely qualitatively, to alternative structural assumptions when selecting among RMP variants, the CCAMLR's approach has been to bracket uncertainty. The former approach allows a quasi-quantitative assessment of risk whereas the latter is less informative and perhaps better suited to eliminating management actions that are likely to perform poorly than to choosing among better performing candidates.

Approaches to Uncertainty III: Model Weighting

It is difficult, but useful and arguably necessary, to quantify the plausibility of a model relative to others, particularly if those models have markedly different implications. Butterworth *et al*. (1996) proposed the following four-level scheme to assign "plausibility ranks" to the hypotheses underlying alternative models:

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

This scheme was presented to the IWC Scientific Committee. Although the scheme provides a formal structure within which to assign weights to models, it has not been used, largely because of concerns regarding the validity of the likelihood functions that are needed to apply the first level of the scheme. Instead, the IWC Scientific Committee has used a less rigorous approach, involving a "Delphi method" (e.g., Linstone and Turoff 1975), in which the Committee assigns a plausibility ranking of "high", "medium", "low", or "no agreement" ("no agreement" being treated as "medium", IWC 2005a) to alternative structural models.

If a valid likelihood function can be constructed (i.e., level one of the above scheme), the Bayesian approach is well suited to assigning weights to alternative models that use the same data. In Bayesian statistics, unobservable quantities are treated as random variables, and Bayes' theorem or rule (see below) is used to define probability density functions for the model parameters. These probability density functions convey the plausibility or degree of support for particular parameter values. Unobservable, discrete items such as alternative scientific hypotheses can also be treated as random variables. Bayesian probabilities can therefore be calculated for alternative scientific hypotheses whereby the probability reflects the plausibility of a hypothesis relative to one or more alternatives (Patterson 1999; McAllister and Kirchner 2002; Parma 2002).

The posterior probability specified by Bayes theorem or rule is directly proportional to the product of the prior probability for the hypothesis, and the probability of obtaining the observed data presuming that the hypothesis is true, i.e.:

$$\text{Posterior Prob}(H) \propto \text{Prior Prob}(H) \times \text{Prob}(\mathbf{D} \text{ assuming } H \text{ is true}),$$

where H is a hypothesis and **D** is a set of observed data. The prior probability reflects the credibility of H relative to its alternatives before evaluating the probability of the data given the hypothesis. Prior probabilities may be formulated on the basis of expert judgment or the analysis of other relevant datasets that are not included in the term which indicates the probability of observing **D** (Punt and Hilborn 1997). The probability of observing **D** given H typically has the same mathematical form as the likelihood function. The Bayes factor (Kass and Raftery 1995):

$$\text{Bayes factor} = \frac{\text{Prob}(\mathbf{D} \text{ assuming } H_1 \text{ is true})}{\text{Prob}(\mathbf{D} \text{ assuming } H_2 \text{ is true})}$$

is one way of evaluating the credibility of a hypothesis H_1 relative to that of an alternative hypothesis H_2 (e.g. Brodziak and Legault, 2005).

The probabilistic weighting of hypotheses provided by Bayes' rule provides a framework that has been applied to decision problems in fisheries management, particularly in attempts to deal with model uncertainty (Walters and Hilborn 1978; Sainsbury 1988; McAllister *et al*. 1994; McAllister and Kirchner 2002). The steps needed to apply this framework to decision analytic problems parallel those of the OMP approach and are as follows (see McAllister *et al*. 1999 for further details).

1. Identify the management objectives.

2. Formulate a set of decision options.
3. Formulate a small number of quantitative performance measures to evaluate the extent to which the objectives are met.
4. Formulate alternative hypotheses for system behaviour (see next section for examples).
5. Use Bayes' rule to compute probabilities for each alternative hypothesis formulated in Step 4 and for their (uncertain) parameters.
6. Evaluate the potential consequences of implementing each decision option in terms of the performance measures. This evaluation is done using the alternative models and posterior probability density functions for the parameters in these alternative decision options. Thus, for each decision option evaluated, a posterior probability distribution of outcomes should be computed for each of the performance measures.
7. Summarize the results obtained in Step 6 for the purpose of conveying them to decision makers (e.g., in decision tables, Hilborn *et al*. 1993; McAllister and Kirchner 2002). Typically, this would show the potential outcomes of each decision option under each alternative hypothesis and the probability assigned to each hypothesis, thus providing an objective method to weight the results obtained under each hypothesis.

There are other approaches for weighting models, including information-theoretic criteria such as Akaike weights. Also, pattern oriented modelling (POM) has been promoted as a way of developing plausible models of complex systems (Grimm and Berger 2004; Grimm *et al*. 2005; Grimm and Railsback 2005). Most applications of POM make use of agent-based models, which are not widely used for modelling exploited marine ecosystems (but see Shin and Cury 2001; Shin *et al*. 2004).

However, the principles of POM can be applied more generally. POM seeks to replicate the characteristic patterns of observed systems at a range of scales and hierarchical levels. The plausibility of models is judged against their ability to capture the full set of "essential" patterns of a system in the most parsimonious way, and their ability to predict secondary characteristics of the system that were not explicitly considered in model formulation (Grimm and Railsback 2005). This approach is data intensive, requiring sufficient observations to establish the characteristic patterns at a range of scales. While it has not been used to weight models, the principles imply that models could be weighted on a combination of model simplicity, the degree of pattern replication, and the extent to which replicated patterns are considered characteristic of the observed system.

Weighting alternative models leads to consideration of model averaging, where the weights determine how much each alternative model contributes to the overall expectation. Burnham and Anderson (1998) recommend model averaging when the goal of an analysis is to "get the best set of parameters in common to all models" but they caution against such averaging if the competing models lead to "definite, and differing, interpretations." Such caution is particularly important if the competing models imply different strategies and tactics for achieving a set of management objectives.

Applications of the Bayesian approach to weight alternative models

There are a number of examples in which Bayesian methods have been used to assign plausibility weights to alternative models in fishery management problems. These mainly concern single-species stock dynamics. However, one of the earliest examples (Sainsbury 1988; Sainsbury *et al.* 1997) attempted to distinguish among competing models of the interactions between harvesting and fish community

structure. Following intensive pair trawl fishing on the northwest shelf of Australia since the 1960s, the abundance of the two most commercially valuable species had declined considerably by the mid 1980s while that of two low value species had increased. The Bayesian approach was applied to evaluate the probability of four alternative models which explained the changes in terms of: (a) fishing and no species interactions, (b) fishing and the less valuable species negatively impacting the more valuable species, (c) fishing and the more valuable species negatively impacting the less valuable species, or (d) fishing causing reductions in habitat for the more valuable species and increases in habitat for the less valuable species. The two best-supported models were approximately equally probable given the available data in the mid-1980s (Sainsbury 1988), and an experimental fisheries management regime was implemented in the late 1980s based on the results of a decision analysis that considered yield of information and economic value as performance measures. The experimental regime was not strictly implemented, but, by 1990, it appeared that the habitat modification hypothesis was most probable (Sainsbury *et al.* 1997; Table 1). Data collected in the 1990s provided further support for this hypothesis (Sainsbury, pers. comm.). The results suggested that trap fishing should replace trawling to achieve a recovery in fish community structure and the fishery for the more valuable species.

Michielsens and McAllister (2004) used Bayesian hierarchical modelling to jointly analyse stock-recruitment datasets for several Atlantic salmon (*Salmo salar*, Salmonidae) populations and extrapolate the results to Baltic stocks for which such data were unavailable. These analyses evaluated the central tendency and cross-stock variance in the steepness parameter of the stock-recruitment relationship, and considered Beverton-Holt and Ricker models as alternative stock-recruit functions. In

subsequent population dynamic modelling of Baltic salmon stocks, the posterior probabilities for the two alternative stock-recruit models have been used as priors. Because the posterior density for the Beverton-Holt model was much higher (i.e., 0.999) only this model has been used to provide management advice (ICES 2004; 2006). The posterior median for steepness in the Beverton-Holt model was considerably less than that in the Ricker model (i.e., 0.72 vs. 1.15) (Michielsens and McAllister 2004). The application of these model weightings supported lower exploitation rates than might have been the case if the analysis had not been performed and both models were assumed to be equally likely by default.

Brodziak and Legault (2005) considered twelve alternative stock-recruitment models for each of three overfished groundfish stocks. These models differed in terms of functional form (Beverton-Holt versus Ricker), error structure, and the nature of prior information. An approximation to the Bayes' factor (Kass and Raftery 1995) was used to compute weights for each model. Finally, weighted averages for each stock were calculated using only those models that provided identifiable parameters. This restriction excluded most of the Ricker models and all Beverton-Holt models without prior information on unfished recruitment levels. The models that were included in the average for each stock generally provided similar estimates of the spawning biomass and fishing mortality rates that produce MSY, and, therefore, the averaged estimates appeared to usefully account for model uncertainty (Table 2). Nevertheless, Brodziak and Legault (2005) acknowledged that the results, in terms of fishing mortalities, were sensitive to the prior mean assumed for the steepness parameter, which was not varied in the study. It is not clear whether model averaging would have been appropriate had the prior mean varied.

Brandon and Wade (2006) also used Bayes' factors to assign plausibility weights to: (a) a model of the Bering-Chukchi-Beaufort Seas bowhead whale stock that initiates projections from an equilibrium in 1848 and assumes that carrying capacity and the parameters of the density-dependence function have not changed over time, and that the historical catches are known without error (e.g. Punt and Butterworth, 1999), (b) a density-dependent model that started the population projection in 1978, and (c) a density-independent model that started the population projection in 1978. Although the density-independent model was assigned the highest posterior probability, none of the three models could be considered implausible (e.g., using criteria developed by Jefferys 1961 and Kass and Raftery 1995). Brandon and Wade (2006) used the results of their analyses to construct model-averaged posterior distributions for key model outputs. As expected, the posterior distributions calculated using model-averaging indicated greater uncertainty than the posterior distributions based on the model that had the highest posterior probability.

McAllister and Kirchner (2002) computed posterior probabilities for four models attributing a decline in orange roughy (*Hoplostethus atlanticus*, Trachichthyidae) catches off Namibia to (a) fishery removals (b) temporary events unrelated to the fishery, (c) temporary dispersal triggered by fishing, and (d) long-term dispersal. Rather than providing relatively precise unimodal distributions for stock biomass, the initially equally probable alternative models resulted in bimodal marginal posterior distributions indicating that, overall, the stock was either heavily depleted or lightly fished (McAllister and Kirchner 2002). The posterior probabilities for the four alternative models were used in a Bayesian decision analysis to evaluate the potential consequences of various fisheries management options (Table 3). When considered across the four hypotheses, the chance of recovery without a major reduction in

fishing effort was found to be very low, and the Namibian Minister of Fisheries consequently lowered the catch quotas on all fishing grounds considerably. A few years after one fishing ground was closed altogether, monitoring detected a major re-aggregation of fish on the closed ground only (C. Kirchner pers. comm.), indicating that the apparent decline may have been, at least partly, due to temporary dispersal of the fish stocks in response to intensive fishing.

Patterson (1999) presented a method for integrating VPA-based stock assessment, reference point estimation, and management simulation using Bayesian methods to compute posterior distributions for parameters, and posterior probabilities for alternative model structures. The model structures considered were the stock-recruitment function (Beverton-Holt versus Ricker) and the shape of the observation error distribution for surveys (normal, lognormal or gamma). For Norwegian spring-spawning herring (*Clupea harengus*, Clupeidae), this process indicated high posterior support for the normal error distribution model, but was unable to distinguish between the stock-recruitment functions. Nonetheless, all alternative models were included in the assessment and management simulation processes weighted by their posterior probabilities. Patterson (1999) found that the uncertainty resulting from the inclusion of three error models (Fig. 4) was greater than that suggested by an *ad-hoc* comparison of the results from individual parametric bootstraps using each error distribution.

Approaches to model uncertainty IV: Examples from related disciplines

In this section we consider approaches used to address model uncertainty in weather forecasting, climate prediction and ocean biogeochemistry. These disciplines also use models to make predictions about complex systems and must account for uncertainties arising from simplifications about processes and scale, as well as from

limited knowledge about the interactions between processes. Numerical weather prediction (NWP) models represent atmospheric processes and are used to forecast the weather over a period of hours to days ahead (Davies *et al*. 2005). For predictions more than 10 days ahead, including climate forecasts, it is necessary to use coupled atmosphere-ocean-sea ice global climate models (GCMs). Ocean biogeochemistry (OBGC) models are used primarily to estimate the air-sea fluxes of gases like CO_2, N_2O, or CH_4, and biological processes like primary production on time scales from one to millions of years (e.g. Maier-Reimer 1993; Suntharalingam and Sarmiento 2000; Carr *et al*. 2005). OBGC models are based on the conservation of elements in the ocean (e.g. carbon, oxygen, nutrients), which are subject to changes driven by physical, chemical, and biological processes. The elements are also affected by input from the atmosphere and rivers, and output from sedimentation.

Both NWP and OBGC models represent physical processes through the application of the laws of physics to fundamental variables such as temperature, pressure and density. Uncertainty arises because these fine scale processes are modelled at relatively coarse spatial and temporal scales. For example, atmospheric processes, such as cumulus clouds, rain showers, and flow over small hills, occur on scales well below those of state-of-the-art global NWP models. In OBGC models, assumptions about parameters such as vertical diffusivity strongly influence the modelled response of the ocean to changes in the atmospheric concentration of gases such as anthropogenic CO_2. The range of vertical transport rates in different models led to a 40% difference in estimated oceanic CO_2 uptake in the 1990s (Orr *et al*. 2001; Doney *et al*. 2004). Further uncertainty arises when the models incorporate processes that are less well understood. These include cloud structure and movement in NWP models and biological processes in OBGC models. The simplest OBGC models

represent only the biological processes that lead to the export of organic matter to the deep ocean based on the surface nutrient, light, and temperature conditions (e.g. Maier-Reimer 1993). The most complex models cover the range of size and functionality of plankton and bacteria in the ocean (e.g. Gregg *et al*. 2003; Le Quéré *et al*. 2005).

It is common practice in weather forecasting and climate prediction to represent uncertainty by considering multiple runs of single models (Bourke *et al*. 2005) or of several models that vary in formulation. When these models begin from the same initial state, this is known as the ensemble approach. NWP models differ in a number of ways. Some solve the equations of atmospheric motion on a regular grid covering the globe. Others solve them in a 'spectral' form, where the atmospheric variables are held as a number of waves around the Earth. The models also differ in terms of how they handle processes such as clouds, flow over orography, and long and short wave radiation in the atmosphere. Ensemble runs of GCMs can provide broad, regional guidance on temperature and precipitation for several months ahead (Vialard *et al*. 2005). Multi-model ensemble analysis is now a standard feature of climate research, and was a key element of the recent Intergovernmental Panel on Climate Change (IPCC) fourth Assessment Report (www.ipcc.ch, accessed 19th March 2007). All members of the ensemble are subject to the same forcing scenario, such as the set of greenhouse gas emission scenarios used within IPCC. Members of the ensemble may differ in their formulation, or random noise applied to the initial state, or both. Different formulations can include or exclude processes such as interactive vegetation (Cox *et al*. 2000). Thus, the ensemble can be used to consider different levels of model complexity.

Meteorologists and climatologists are only just starting to consider how to deal with the output of an ensemble of models. The IPCC has used twenty models to investigate past climate change and how the climate may evolve in the future. These models differ in many ways including horizontal resolution and the representation of clouds, sea ice, oceans and the inclusion of vegetation. Current estimates of future conditions are based on the averaged output of all twenty models with the standard deviation of this estimate used to represent uncertainty (IPCC 2007; Fig. 5). However, more sophisticated means are being considered to deal with the output of so many models (Mueller *et al*. 2005). For example, the models can be weighted according to how they represent various elements of the present climate, such as pressure at mean sea level or temperature. Alternatively, the models can be weighted according to how well they reproduce aspects of the climate of recent decades. Of course, the ability of a model to replicate the past may not be a good indicator of how it will deal with changes in the future as a result of different forcing factors, but methods for assessing confidence in climate predictions will be a major focus in future research.

Ocean biogeochemistry models that focus on carbon have been evaluated during four phases of the international Ocean Carbon-cycle Model Intercomparison Project (OCMIP). In the earlier phases, models were primarily compared with each other and the first representation of uncertainty was the range of model results (Orr *et al*. 2001). It was acknowledged at the time that this under-represented uncertainty because the protocols for model comparison restricted the diversity of models considered, and the models shared common elements that could all have bias in the same direction (e.g. the low resolution of their grids). In later phases, models were compared with a whole suite of observations (Dutay *et al*. 2002; Doney *et al*. 2004). In this case, much of the data collection was triggered by the publication of results from the initial phases of

the model intercomparison. Extreme models were developed to bracket the full uncertainty suggested by the observations. For example, the uptake of observed CFCs and bomb ^{14}C was clearly overestimated in a model with excessive mixing, and underestimated in one with low mixing (Matsumoto *et al*. 2004), but these estimates usefully defined bounds for the uptake of anthropogenic CO_2 (Fig. 6). On the other hand, it is suspected that uncertainty is greatly under-represented by the existing models of biological processes (Le Quéré *et al*. 2005). The development of biological models is a very active field of research, with a corresponding focus on producing validation datasets to define constraints on the biological components (e.g. Rivkin and Legendre 2001; Buitenhuis *et al*. 2006).

The OCMIP has demonstrated a community approach to model development and evaluation. Model comparison identifies where uncertainties have the greatest effects on results and coordination ensures that data collection addresses these uncertainties and their implications. However, the OCMIP has concentrated on bracketing uncertainty because there are limited data with which to assign weights to alternative hypotheses (Doney *et al*. 2004).

This summary suggests that other disciplines are at a similar stage as fisheries science in dealing with model uncertainty; it is recognised as an issue and various approaches have been developed, although none provides a truly quantitative description of uncertainty in models of complex systems. The ensemble approach in particular has been recognised as potentially useful in modelling marine ecosystems (DeYoung *et al*. 2004). This approach currently allows the comparison of predictions arising from a suite of different models which can include differing levels of complexity. Thus, if output variables are standardized to facilitate comparisons, whole ecosystem models (e.g. Shannon *et al*. 2004) could be run alongside minimally

realistic models (e.g. Punt and Butterworth 1995). However, more progress is needed to ensure that the output is not biased by the choice of ensemble members.

Presenting results

Whipple *et al*. (2000) identified the uncertainty associated with complex ecosystem models as a limitation to their usefulness in management. Yet the ecosystem approach to fisheries implies a willingness to attempt management despite this uncertainty. The communication of model results, and their associated uncertainties, to decision makers is a critical step in this process.

The OMP and Bayesian decision analytic approaches make it clear that model predictions should be presented in terms of performance measures that assess whether management objectives have been achieved. Models of complex systems can forecast the dynamics of a plethora of metrics (e.g. Fulton *et al*. 2005) and produce overwhelming amounts of information, especially when the basic outputs are accompanied by estimates of uncertainty. Therefore, the performance measures must be limited in number, to around seven. Ideally, these measures should also include quantities that can be monitored in the real system so that model results could, in principle at least, be validated. More detailed results should be made available as required, but the initial summary, which is all that might be used for decision-making, must be accessible to decision makers.

It is tempting to assume that uncertainty can always be expressed in quantitative terms (such as risk, posterior distributions, probability intervals, and critical percentiles) and approaches to model uncertainty that include explicit weighting of alternative models make this more likely. However, the probabilities suggested by these approaches are not absolute, but relative to the alternatives included in the analysis. Furthermore, different models might favour different management options

and presenting results in the form of summary statistics such as averages might be misleading.

Further complexities are added when it is not possible to weight models. An unweighted set of models gives equal prominence to extreme models, but it could be inappropriate to make decisions on the basis of ‘worst case scenarios’ as this can lead to the selection of management actions based on a highly implausible model. Butterworth *et al*. (1996) liken this to refusing to cross the road because of the small, but plausible, risk of being run over.

Given these issues, it is imperative that analysts are realistic about the limitations of their approaches and present their results as the best use of available knowledge rather than as being definitive. Ultimately, making decisions about the management of complex systems involves risks and trade-offs. The role of models is to facilitate these hard decisions by making predictions that emphasise the risks and uncertainties involved.

Discussion

Limitations and pragmatic approaches

There has been increasing interest in evaluating both parameter and model uncertainty in single species systems (IWC 1992; Patterson 1999; McAllister and Kirchner 2002; Parma 2002) and several authors have advocated evaluating both forms of uncertainty in fisheries stock assessment (Punt and Hilborn 1997; Geromont et al. 1999; McAllister and Kirchner 2002). Indeed, various studies have shown that predictions of the potential responses of fish stocks to management actions are more sensitive to model uncertainty than to uncertainty over parameter values in a single model. However, serious consideration to both sources of uncertainty is still far from common practice in fisheries stock assessment.

Model uncertainty increases with the number of modelling assumptions made, which generally increases with the complexity of the system of interest because more processes can be represented. A tendency to model more complex systems has accompanied the shift in interest from evaluating the potential effects of harvesting on single populations to evaluating effects on ecosystems. Recent studies have demonstrated how predictions from ecosystem models are sensitive to underlying structural assumptions (e.g., Mackinson *et al*. 2003; Watters *et al.* 2003; Pinnegar *et al*. 2005). However, the limited attention that has been given to addressing uncertainty has typically concerned the values for parameters in a single ecosystem model. We argue that it should become common practice to account for both model and parameter uncertainty when predicting the potential responses of marine ecosystems to alternative management actions. May (1988) suggests that model construction may tell us more about the psychology of scientists than about ecology. The ideal is, therefore, to replace implicit and opaque “psychological” choices about model

structures with an explicit and transparent evaluation of the various ways in which the system could behave.

Promising approaches to dealing with model uncertainty establish sets of alternative models and assign probabilities to each member of the set. These probabilities reflect the credibility or plausibility of one model relative to alternatives in the set. Several examples already exist where this has been accomplished using the Bayesian approach, but even in these cases the initial set of models and assumptions about prior distributions rely on subjective judgement to some extent. Furthermore, the model with the highest posterior support might still be a poor representation of the actual system dynamics, and models that were not included could potentially make better predictions. Other limitations to the Bayesian approach arise because of data constraints and model complexity, and because Bayes' rule cannot be used to assign posterior weights to models that use different data sources. However, models can be weighted by prior probabilities even when the models require different datasets for parameter estimation. A relatively simple implementation is to treat alternative models as random variables in Monte Carlo simulations with the probability of inclusion determined by some quantitative or qualitative assessment of *apriori* plausibility (Hill *et al*. 2006).

Existing quantitative approaches to model weighting depend on a model's ability to mimic historical data, which is not necessarily a good indicator of its ability to predict the behaviour of the system under potentially different future conditions. For example, if historical stock sizes have been reasonably high, a comparison of models with and without depensation is likely to place greater weight on the simpler model that ignores depensation. However, this may be inappropriate if the analysis considers management actions that lead to stock sizes lower than historically observed. Finally,

data limitations are severe for complex ecosystem models. It might never be possible to properly weight hypotheses about fishery-ecosystem interactions without data collection in an adaptive management context.

The ideal of a completely objective approach and even the use of weighted models is constrained by practical difficulties. However, the requirement to implement EAFs in the near future, as called for by the 2002 World Summit on Sustainable Development, suggests that attention must be given to using current approaches and knowledge in pragmatic ways. In our opinion, this requires the use of management approaches that can be shown to be robust to uncertainty as well as practical ways of assessing uncertainty. Approaches that identify robust management actions are well established and have been applied in single-species contexts. Implementing such approaches in ecosystem contexts will require "operating models" of ecosystem dynamics to test the robustness of proposed management actions. To the best of our knowledge, the recent CCAMLR workshop represents the only attempt to use ecosystem dynamics models to identify robust management actions. This workshop did not attempt to weight alternative models. However, this is not the same as assigning equal weight to all models.

While the current state-of-the-art is far from the ideal, we cannot ignore model uncertainty in developing EAFs. Based on our consideration of fisheries and related disciplines, we believe that the following represents current best practice in representing uncertainty in the structural form of models of ecosystem dynamics:

1. Identify the purpose of the modelling exercise in terms of management objectives, and the performance measures by which the attainment of these objectives will be assessed (including the period of time over which the performance measures are to be calculated).

2. Identify the key uncertainties about the system. This should occur during the process of assembling information and formulating models; it is necessary to identify and highlight uncertainties rather than to make assumptions that constrain the models to a single view of any important process.
3. Develop models or parameterisations that represent plausible limits to each key uncertainty. Consider more than one basic model structure.
4. Always include less extreme models, assumptions and parameterisations (those between the bounds may be more plausible), and ensure that the choice of models, parameter values, and assumptions is balanced given the purpose of the modelling. For example, McAllister and Kirchner (2002) ensured that two models implying low current stock abundance of orange roughy were balanced by two that implied high current stock abundance, but low availability.
5. Establish the full range of model behaviours by considering different combinations of models and parameters. Sensitivity analysis is useful at this stage to determine the importance of each source of uncertainty
6. Consider the interaction between models and data. Specifically, do the models capture the full range of potential conditions, or just the conditions represented by the data? If the latter is the case, it may be advisable to consider adopting plausible future scenarios that extend outside of the range of historic data, as may occur, for example, under current predictions for climate change.
7. Ensure that each model is logically consistent. For example, assumptions or fixed values for key parameters in an ecosystem model will also need to be made in models used to derive prior density functions for the ecosystem model's input parameters.

8. Weight models by plausibility if information exists to do this. Ideally this weighting would be on the basis of posterior probability, but a more subjective weighting by prior probability might be necessary.
9. Run each model multiple times to incorporate the effects of parameter uncertainty and natural variability.
10. Avoid averaging model results unless the distribution of results suggested by all models is unimodal.
11. If it is possible to weight models, present the results in terms of the risk that each management objective will not be met. If it is not possible to weight models, present the results in terms of the trade-offs associated with each management action for each alternative model.
12. Make sure the assumptions and limitations of the approach are presented along with the results.

Improving models and establishing protocols

Models are formalised statements of scientific hypotheses and part of the ongoing process of improving understanding which, in itself, should eventually reduce uncertainty. As the OCMIP demonstrates, model comparisons can help to focus data collection by identifying the areas where model uncertainty has the greatest effect on predictions. Sainsbury's (1988) analysis sparked an experimental fishing regime with spatial and temporal replication and contrasting fishing "treatments." Ultimately, this experiment yielded informative data that provided preferential support to one of the hypotheses considered in the original study, and implied a management strategy that might achieve the objectives of restoring the fish community structure and recovering the fishery for valuable species.

Observation networks on the scale of those set up to study the Earth system will be of limited use in reducing model uncertainty in resource dynamics models because the spatial and temporal scales of ecosystem dynamics are smaller than whole-Earth models. Thus, observation is best concentrated on local monitoring systems that, over time, can provide feedback with which to refine models and reduce uncertainty.

However, large-scale international coordination within the fisheries science community might be useful in defining and refining the general types of models that should be included in a set of alternative models. Specific details of such models will require expertise about the local system, but global-scale coordination can help to ensure that the set of models considered when implementing an EAF are balanced and not simply a reflection of localised conventional wisdom. As with the OCMIP, global-scale coordination could allow the testing and refinement of general models through confrontation with comparative data collected from multiple systems. This would build on recent meta-analyses of fisheries data that have produced valuable insights at large spatial scales (e.g. Liermann and Hilborn 1997; Myers *et al*. 2001) and could potentially facilitate the rapid evaluation of process models such as foraging arena theory (Walters and Martell 2004).

Finally, we note that model uncertainty is common to most modelling disciplines. There is therefore a need for continued dialogue amongst practitioners in different fields. An example of such collaboration is the Ecosystem Modelling Committee established by the North Pacific Research Board (NPRB). This Committee consists of modellers from a wide range of disciplines and has developed a set of criteria for evaluating ecosystem models (http://doc.nprb.org/web/BSIERP/BSIERP_2007RFP-Full-proposals.PDF) based on evaluation techniques from a range of modelling sub-

disciplines. Concordance with these criteria will be used as part of the basis for selecting an Integrated Bering Sea Ecosystem Project by the NPRB.

Acknowledgements

This paper is a contribution to the Ocean Ecosystems and Management project in the British Antarctic Survey's DISCOVERY 2010 programme. Funding for AEP was provided by NMFS Grant NA04NMF4550330. Some of this work was supported by NSF grant # 0443751 to W. Trivelpiece, S. Trivelpiece, and GMW. Additional support was also provided to GMW by the Lenfest Ocean Program at the Pew Charitable Trust. SLH thanks Keith Reid for support, Greg Donovan for useful information at short notice and Jamie Oliver for help with Figure 4. We thank Villy Christensen, Carl Walters and anonymous referees for stimulating comments and helpful suggestions, Eva Plagányi for providing early access to Plagányi (2007), and Jon Brodziak, Katsumi Matsumoto, Kenneth Patterson, Keith Sainsbury, the CCAMLR and the IPCC and their publishers for generous permission to reproduce figures and tables from their original publications.

References

Begley J. (2005) Gadget User Guide (http://www.hafro.is/gadget/userguide/node1.html, accessed 19th March 2007)

Bourke, W., Buizza, R. and Naughton, M. (2005) Performance of the ECMWF and the BoM Ensemble Systems in the Southern Hemisphere. *Weather Forecasting* **132**, 2338-2357

Brandon, J. and Wade, P.R. (2006) Assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using Bayesian model averaging. *Journal of Cetacean Research and Management* **8**, 225-239

Brodziak, J. and Legault, C.M. (2005) Model averaging to estimate rebuilding targets for overfished stocks. *Canadian Journal of Fisheries and Aquatic Sciences* **62**, 544-562.

Burnham, K.P., and Anderson, D.R. (1998) *Model Selection and Inference: A Practical Information-theoretic Approach*. Springer-Verlag, New York, New York, USA. 353 pp.

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.

Buitenhuis, E., Le Quéré, C., Aumont, *O. et al.* (2006) Biogeochemical fluxes through mesozooplankton. *Global Biogeochemical Cycles* **20**, GB20023.

Carr, S.D., Hench, J.L., Luettich, R.A., Forward, R.B. and Tankersley, R.A. (2005) Spatial patterns in the ovigerous *Callinectes sapidus* spawning migration: results from a coupled behavioral-physical model. *Marine Ecology-Progress Series* **294**, 213-226.

Christensen, V. and Walters, C.J. (2004) Ecopath with Ecosim: methods, capabilities and limitations. *Ecological Modelling* **172**, 109-139.

Cox, P.M., Betts, R.A., Jones, C.D., Spall, S.A. and Totterdell, I.J. (2000) Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**, 184-187

Charles, A.T. (1998) Living with uncertainty in fisheries: analytical methods, management priorities and the Canadian groundfishery experience. *Fisheries Research* **37**, 37-50.

Constable, A.J. (2004) Managing fisheries effects on marine food webs in Antarctica: Trade-offs among harvest strategies, monitoring, and assessment in achieving conservation objectives. *Bulletin of Marine Science* **74**, 583-605.

Constable, A.J., de la Mare, W.K., Agnew, D.J., Everson, I. and Miller, D. (2000) Managing fisheries to conserve the Antarctic marine ecosystem: Practical implementation of the Convention on the Conservation of Antarctic Marine Living Resources (CCAMLR) *ICES Journal of Marine Science* **57**, 778-791.

Davies, T., Cullen, M.J.P., Malcolm, A.J., Mawson, M.H., Staniforth, A., White, A.A. and Wood, N., 2005: A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society* **131**, 1759-1782.

deYoung, B., Heath, M., Werner, F., Chai, F., Megrey, B., & Monfray, P. (2004) Challenges of Modeling ocean basin ecosystems. *Science* **304**, 1463-1466.

Doney, S.C., Lindsay, K., Caldeira, K. *et al.* (2004) Evaluating global ocean carbon models: The importance of realistic physics. *Global Biogeochemical Cycles* **18**, GB3017.

Dutay, J.-C., Bullister, J.L., Doney, S.C., *et al.* (2002) Evaluation of ocean model ventilation with CFC-11: comparison of 13 global ocean models. *Ocean Modelling* **4**, 89-120.

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

Fulton, E.A., Parslow, J.S., Smith, A.D.M. and Johnson, C.R. (2004) Biogeochemical marine ecosystem models I: a model of marine bay ecosystems. *Ecological Modelling* **174**, 267-307.

Fulton, E.A., Smith, A.D.M. and Punt, A.E. (2005) Which ecological indicators can robustly detect effects of fishing? *ICES Journal of Marine Science* **62**, 540-551.

Garcia, S.M., Zerbi, A., Aliaume, C., Do Chi, T. and Lasserre, G. (2003) The ecosystem approach to fisheries. Issues, terminology, principles, institutional foundations, implementation and outlook. *FAO Fisheries Technical Paper* **443**, 71pp

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

Gregg, W.W., Ginoux, P. Schopf, P.S. and Casey, N.W. (2003) Phytoplankton and iron: validation of a global three-dimensional ocean biogeochemical model. *Deep Sea Research II* **50**, 3143-3169.

Grimm, V. and Berger, U. (2004) Seeing the forest for the trees, and vice versa: pattern-oriented ecological modelling. Pages 411-428 in L. Seuront and P. G. Strutton, eds. *Handbook of Scaling Methods in Aquatic Ecology: Measurement, Analysis, Simulation*. CRC Press, Boca Raton.

Grimm, V. and Railsback, S.F. (2005) *Individual-based Modelling and Ecology*. Princeton University Press, Princeton N.J.

Grimm, V., Revilla, E., Berger, U. *et al.* (2005) Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science* **310**, 987-991.

Halley, J. and Inchausti, P. (2002) Lognormality in ecological time series. *Oikos* **99**, 518-530.

Halpern, B.S., Regan, H.M., Possingham, H.P. and McCarthy, M.A. (2006) Accounting for uncertainty in marine reserve design. *Ecology Letters* **9**, 2-11.

Harwood, J. and Stokes, K. (2003) Coping with uncertainty in ecological advice: lessons from fisheries. *Trends in Ecology and Evolution* **18**, 617-622.

Hewitt, R.P., Watters, G., Trathan, P.N. *et al.* (2004a) Options for allocating the precautionary catch limit of krill among small-scale management units in the Scotia Sea. *CCAMLR Science* **11**, 81-97.

Hewitt, R.P., Watkins, J., Naganobu, M. *et al.* (2004b) Biomass of Antarctic krill in the Scotia Sea in January/February 2000 and its use in revising an estimate of precautionary yield. *Deep-Sea Research Part II-Topical Studies in Oceanography* **51**, 1215-1236.

Hilborn, R., Pikitch, E.K., and Francis, R.C. 1993. Current trends in including risk and uncertainty in stock assessment and harvest decisions. *Canadian Journal of Fisheries and Aquatic Sciences* **50**, 874-880.

Hill, S.L., Reid, K., and North, A.W. (2005) Recruitment of mackerel icefish (*Champsocephalus gunnari*) at South Georgia indicated by predator diets and its relationship with sea surface temperature. *Canadian Journal of Fisheries and Aquatic Sciences* **62**, 2530-2537.

Hill, S.L., Murphy, E.J., Reid, K., Trathan, P.N., and Constable, A.J. (2006) Modelling Southern Ocean ecosystems: krill, the food-web, and the impacts of harvesting. *Biological Reviews* **81**, 581-608.

Hill, S.L., Reid, K., Thorpe, S.E., Hinke, J. and Watters, G.M. (in press) A compilation of parameters for ecosystem dynamics models of the Scotia Sea – Antarctic Peninsula region. *CCAMLR Science* **14,** 00-00

Hollowed, A.B., Ianelli, J.N. and Livingston, P.A. (2000) Including predation mortality in stock assessments: a case study for Gulf of Alaska walleye pollock. *ICES Journal of Marine Science* **57**, 279-293.

ICES (2004) Report of the Baltic salmon and trout assessment working group (WGBAST) 21 April – 30 April 2004. ICES CM 2004/ACFM **23,** 212 pp.

ICES (2006) Report of the Baltic salmon and trout assessment working group (WGBAST) 28 March – 6 April 2006. ICES CM 2006/ACFM **21**, 213 pp.

IPCC (2007) Climate Change 2007: The Physical Science Basis. Summary for Policymakers. 18 pp

IWC (International Whaling Commission) (1981) Report of the Third Meeting of the Special Scientific Working Group on Management Procedures Honolulu, Hawaii, 20-26 March 1980. *Reports of the International Whaling Commission* **31**, 41-47.

IWC (International Whaling Commission) (1992) Report of the Third Comprehensive Assessment Workshop on Management Procedures, Oslo, 1-8 February 1990. *Reports of the International Whaling Commission* **42**, 271-304.

IWC (International Whaling Commission) (2003) Chair's Report of the Fifty-Fourth Annual Meeting. *Annual Report of the International Whaling Commission, 2002*: 1-53.

IWC (International Whaling Commission) (2005a) 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,** 84-92.

IWC (International Whaling Commission) (2005b) Chair's Report of the Fifty-sixth Annual Meeting. *Annual Report of the International Whaling Commission, 2004*: 1-58.

Jeffreys, H. (1961) *Theory of Probability*. 3rd edn. Oxford University Press, Oxford.

Kass, R.E. and Raftery, A.E. (1995) Bayes Factors. *Journal of the American Statistical Association* **90**, 773-795.

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 Evaluation Frameworks. In: *The Knowledge Base for Fisheries Management*. (eds L. Motos, and D. Wilson). Elsevier Limited. pp. 379-407.

Le Quéré, C., Harrison, S. P., Prentice, I.C. *et al.* (2005) Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*. **11**, 2016-2040.

Liermann, M. and Hilborn, R. (1997) Depensation in fish stocks: a hierarchic Bayesian meta-analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **54**, 1976-1984.

Linstone, H. and Turoff. M. (1975) The Delphi Method: Techniques and Applications. Addison-Wesley, Boston.

Ludwig, D., Hilborn, R. and Walters, C. (1993) Uncertainty, resource exploitation and conservation: Lessons for history. *Science* **260** (17), 17-18.

Mackinson, S.J., Blanchard, J.L., Pinnegar, J.K., Scott, R. (2003) Consequences of alternative functional response formulations in models exploring whale-fishery interactions. *Marine Mammal Science* **4**, 661-681.

Magnússon, K.G. (1995) An overview of the multispecies VPA — theory and applications. *Reviews in Fish Biology and Fisheries* **5**, 195-212.

Maier-Reimer, E. (1993) Geochemical cycles in an ocean general circulation model: preindustrial tracer distributions. *Global Biogeochemical Cycles* **7**, 645-677.

Mangel, M. (2006) Commentary: Accounting for uncertainty in marine reserve design. *Ecology Letters* **9**, 11-12.

Matsumoto, K., Sarmiento, J.L., Key, R.M. *et al.* (2004) Evaluation of ocean carbon cycle models with data-based metrics. *Geophysical Research Letters* **31,** L07303

May, R.M. 1988. How many species are there on Earth? *Science* **241**, 1441–1449.

McAllister, M. and Kirchner, C. (2002) Accounting for structural uncertainty to facilitate precautionary fishery management: Illustration with Namibian orange roughy. *Bulletin of Marine Science* **70**, 499-540.

McAllister, M.K., Pikitch, E.K., Punt, A.E. and Hilborn, R. (1994) A Bayesian-Approach to Stock Assessment and Harvest Decisions Using the Sampling Importance Resampling Algorithm. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 2673-2687.

McAllister, M.K., Starr, P.J., Restrepo, V.R. and Kirkwood, G.P. (1999) Formulating quantitative methods to evaluate fishery-management systems: what fishery processes should be modelled and what trade-offs should be made? *ICES Journal of Marine Science* **56**, 900-916.

Michielsens, C. and McAllister, M. (2004) A Bayesian hierarchical analysis of stock-recruit data: quantifying structural and parameter uncertainties. *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 1032-1047.

Miller, G.A. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review* **63**, 81-97

Mueller, W., Appenzeller, C., Doblas-Reyes, F.J. and Liniger, M. (2005) Debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. *Journal of Climate* **27**, 1513-1523.

Myers, R.A., MacKenzie, B.R., Bowen, K.G. and Barrowman, N.J. (2001) What is the carrying capacity for fish in the ocean? A meta-analysis of population

dynamics of North Atlantic cod. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 1464-1476.

Orr, J.C., Maier-Reimer, E., Mikolajewicz, U. *et al.* (2001) Estimates of anthropogenic carbon uptake from four three-dimensional global ocean models. *Global Biogeochemical Cycles* **15**, 43-60.

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**, 423-453.

Patterson, K., Cook, R., Darby, C. *et al.* (2001) Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries* **2**, 125-157.

Patterson, K.R. (1999) Evaluating uncertainty in harvest control law catches using Bayesian Markov chain Monte Carlo virtual population analysis with adaptive rejection sampling and including structural uncertainty. *Canadian Journal of Fisheries and Aquatic Sciences* **56**, 208-221.

Plagányi, E.E. (2007) Methods for assessing indirect ecosystem effects of fisheries. *FAO Fisheries Technical Paper* 477. FAO, Rome

Pinnegar, J.K., Blanchard, J.L., Mackinson, S., Scott, R.D. and Duplisea, D.E. (2005) Aggregation and removal of weak-links in food-web models: system stability and recovery from disturbance. *Ecological Modelling*, **184**, 229-248.

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 Modeling* **19**, 441-464.

Punt, A.E. and Butterworth, D.S. (1995) The effects of future consumption by the Cape fur seal on catches and catch rates of the cape hakes .4. Modelling the

biological interaction between Cape fur seals *Arctocephalus pusillus pusillus* and the cape hakes *Merluccius capensis* and *M. paradoxus*. *South African Journal of Marine Science* **16**, 255-285.

Punt, A.E. and Butterworth, D.S. (1999) On assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*) using a Bayesian approach. *Journal of Cetacean Research and Management* **1**, 53-71.

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

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 G. Cui. 2002. Evaluation of management tools for Australia's South East Fishery. 2. How well do commonly-used stock assessment methods perform? *Marine and Freshwater Research* **53**, 631-644.

Regan, H.M., Colyvan, M. and Burgman, M.A. (2002) A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* **12**, 618-628.

Rivkin, R.B. and Legendre, L. (2001) Biogenic carbon cycling in the upper ocean: Effects of microbial respiration. *Science* **291**, 2398-2400.

Sainsbury, K.J. (1988) The ecological basis of multispecies fisheries, and management of a demersal fishery in tropical Australia. In: *Fish Population Dynamics,* 2nd ed. (ed. J.A. Gulland). Wiley, New York. pp. 349-382.

Sainsbury, K. J., Campbell, R. A., Lindholm, R. and Whitelaw, W. (1997) Experimental management of an Australian multispecies fishery: examining the possibility of trawl-induced habitat modification. In: *Global Trends: Fisheries*

Management (Proceedings of American Fisheries Society Symposium 20. Bethesda, Maryland, USA.) E. L. Pikitch, D. D. Huppert, and M. P. Sissenwine, eds. pp. 107-112.

SC-CAMLR. (2006) Report of the Second Workshop on Management Procedures. In: Report of the Twenty-fifth Meeting of the Scientific Committee (SC-CAMLR-XXV), Annex 4, Appendix D. CCAMLR, Hobart, Australia: 227-258.

Shannon, L., Christensen, V. and Walters, C.J. (2004) Modelling stock dynamics in the southern Benguela ecosystem for the period 1978-2002. *African Journal of Marine Science* **26**, 179-196.

Shin, Y.J. and Cury, P. (2004) Using an individual-based model of fish assemblages to study the response of size spectra to changes in fishing. *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 414-431.

Shin, Y.J., Shannon, L.J. and Cury, P.M. (2004) Simulations of fishing effects on the southern Benguela fish community using an individual-based model: Learning from a comparison with Ecosim. *African Journal of Marine Science* **26**, 95-114.

Suntharalingam, P. and Sarmiento, J.L. (2000) Factors governing the oceanic nitrous oxide distribution: Simulations with an ocean general circulation model. *Global Biogeochemical Cycles* **14**, 429-454.

Vialard, J., Vitart, F., Balmaseda, M.A., Stockdale, T.N. and Anderson, D.L.T (2005) An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model. *Monthly Weather Review* **133**, 441-453

Walters, C.J. and Hilborn, R. (1976) Adaptive control of fishing systems. *Journal of the Fisheries Research Board of Canada* **33**, 145-159

Walters, C.J. and Hilborn, R. (1978) Ecological optimization and adaptive management. *Annual Review of Ecology and Systematics* **9**, 157-188.

Walters, C.J. and Martell, S.J.D. (2004) *Fisheries Ecology and Management* Princeton University Press, Princeton.

Watters, G.M., Olson, R.J., Francis, R.C. *et al.* (2003) Physical forcing and the dynamics of the pelagic ecosystem in the eastern tropical Pacific: simulations with ENSO-scale and global-warming climate drivers. *Canadian Journal of Fisheries and Aquatic Sciences* **60**, 1161-1175.

Whipple, S.J., Link, J.S., Garrison, L.P. and Fogarty, M.J. (2000) Models of predation and fishing mortality in aquatic ecosystems. *Fish and Fisheries* **1**, 22-40.

Yodzis, P. (1994) Predator-Prey Theory and Management of Multispecies Fisheries. *Ecological Applications* **4**, 51-58.

Table 1 The prior and posterior support for each of four alternative hypotheses to explain observed changes on the North West Shelf of Australia. The hypotheses were arbitrarily assigned equal prior probabilities, which were used to calculate posterior probabilities, first with data available in 1985 and then with data from five years of experimental management up to 1990. Modified with permission from Sainsbury *et al.* (1997, their Table 1).

		Probability	
Hypothesis	Prior	1985	1990
(a) No species interactions	0.25	0.01	0.02
(b) Less valuable species negatively affect more valuable species	0.25	0.52	0.33
(c) More valuable species negatively affect less valuable species	0.25	0.01	0.03
(d) Habitat effects	0.25	0.46	0.62

Table 2 Spawning stock biomass (S_{MSY}: thousands of metric tons) and fishing mortality rate (F_{MSY}: per year) associated with MSY for Georges Bank Atlantic cod (*Gaduus morhua*, Gadidae) based on five stock-recruitment models (RBH: informative recruitment prior with uncorrelated Beverton-Holt; RABH: informative recruitment prior with autocorrelated Beverton-Holt; RZBH informative recruitment and steepness priors with uncorrelated Beverton-Holt; RZABH informative recruitment and steepness priors with autocorrelated Beverton-Holt; SRK informative slope at origin prior with uncorrelated Ricker). Modified with permission from Brodziak and Legault (2005, their Table 4).

Model	Posterior probability [-2 log (B_j)]	S_{MSY}	F_{MSY}	Evidence against
RBH	0.34 [1.4]	193.7 (36.2)	0.21 (0.03)	None
RABH	0.15 [3.5]	176.1 (39.3)	0.23 (0.05)	Positive
RZBH	0.33 [1.4]	188.7 (33.6)	0.22 (0.02)	None
RZABH	0.16 [3.4]	172.7 (34.6)	0.23 (0.03)	Positive
SRK	0.01 [8.9]	87.5 (57.4)	0.69 (0.01)	Strong
Model average		184.7 (38.2)	0.23 (0.06)	
80% credibility interval		(135.8, 233.6)	(0.15, 0.31)	

Note B_j is the Bayes' factor evaluating the credibility of model *j* relative to all alternative models.

Table 3 An example decision table resulting from a Bayesian decision analysis in which four alternative hypotheses to explain reduced orange roughy yields were examined. For each total allowable catch (TAC) policy and fishing ground, the 10th percentile (with median value in parentheses) of the expected mature biomass in 2010 as a proportion of the unexploited biomass is shown for each of the alternative hypotheses (see text for details), and all hypotheses combined (and weighted by posterior probability). The probabilities associated with each hypothesis are also given. Modified with permission from McAllister and Kirchner (2002, their Table 5).

Ground	TAC	**Hypothesis a**	**Hypothesis b**	**Hypothesis c**	**Hypothesis d**	**Combined**
Johnies						
	Probability	<0.01	0.02	<0.01	0.98	
	500 mt	0.13 (0.35)	0.60 (0.82)	0.64 (0.86)	0.06 (0.11)	0.07 (0.12)
	1000 mt	0.01 (0.16)	0.55 (0.77)	0.61 (0.83)	0.01 (0.07)	0.02 (0.08)
	1500 mt	0.01 (0.06)	0.51 (0.74)	0.57 (0.79)	0.003 (0.04)	0.01 (05)
	2000 mt	0.01 (0.04)	0.46 (0.69)	0.54 (0.75)	0.002 (0.01)	0.01 (0.02)
Frankies						
	Probability	<0.01	0.25	0.37	0.37	
	0 mt	0.29 (0.44)	0.64 (0.84)	0.68 (0.91)	0.11 (0.15)	0.46 (0.60)
	1000 mt	0.01 (0.07)	0.53 (0.73)	0.61 (0.85)	0.003 (0.04)	0.36 (0.51)
	2000 mt	0.01 (0.03)	0.41 (0.63)	0.55 (0.78)	0.001 (0.01)	0.31 (0.45)
Rix						
	Probability	0.25	0.13	0.45	0.17	
	500 mt	0.21 (0.46)	0.40 (0.66)	0.52 (0.74)	0.14 (0.29)	0.36 (0.58)
	1000 mt	0.02 (0.26)	0.22 (0.55)	0.41 (0.65)	0.01 (0.16)	0. 22 (0.46)

	1500 mt	0.01 (0.10)	0.06 (0.43)	0.28 (0.57)	0.005 (0.06)	0.14 (0.35)
	2000 mt	0.01 (0.05)	0.02 (0.32)	0.16 (0.48)	0.004 (0.03)	0.08 (0.28)
Hotspot						
	Probability	<0.01	0.12	0.01	0.87	
	200 mt	0.03 (0.24)	0.61 (0.85)	0.44 (0.71)	0.05 (0.10)	0.12 (0.20)
	250 mt	0.01 (0.17)	0.59 (0.84)	0.39 (0.68)	0.04 (0.09)	0.11 (0.19)
	500 mt	0.006 (0.03)	0.51 (0.79)	0.18 (0.57)	0.003 (0.06)	0.07 (0.15)

Figure 1 A model with limited predictive ability. The model explains 41% of the observed variability in recruitment of mackerel icefish at South Georgia (I, data labelled by year of recruitment) as a linear function of sea surface temperature (SST) when the cohort was spawned. Reproduced with permission from Hill *et al*. (2005, their Fig. 5).

Figure 2 Population trajectories generated by a spatially resolved ecosystem dynamics model (KPFM). The panels show output for three of the four modelled predator taxa in two of the fifteen spatial units (SSMUs) considered. Grey and black lines result from two competing management options. The multiple lines within each group represent the effect of uncertainties concerning krill transport rates and predator functional responses, as well as stochastic krill recruitment variation. Reproduced with permission from SC-CAMLR (2006, their Fig. 4a).

Figure 3 An example of the presentation of uncertainty in ecosystem model (KPFM) predictions. The figure indicates the trade off between a fishery performance measure and an ecosystem performance measure (the probability that the whale stock will remain above a threshold) in one of fifteen modelled spatial units (SSMUs) for each of six candidate management options.

Figure 4 Bayes posterior probability of different model structures relating to (a) stock-recruit functions and (b) observation error distributions assessed; and (c) cumulative probability that a catch will exceed a specified harvest control law, based on the posterior probabilities from a Bayesian analysis incorporating structural and

parameter uncertainties. Modified with permission from Patterson (1999, his Figs 3 & 6).

Figure 5 An example of the presentation of model uncertainty to decision makers. The right hand side of the figure shows global surface temperature increase (relative to 1980-99) predictions under two forcing scenarios (atmospheric CO_2 maintained at year 2000 concentrations, and A2: emissions increasing to 3.8 times 1990 levels by 2100). The central line for each scenario is the average over an ensemble of 20 models, while the shading represents the standard deviation of these averages. Adapted with permission from IPCC (2007, their Fig. SPM-5).

Figure 6 A representation of model uncertainty in OGBC models. Symbols with error bars are observed values (± 2 standard deviations) while the numbered symbols are mean estimates derived from various models. The estimates are of North Pacific Deep Water (diamonds, lower observation) and North Atlantic Deep Water (triangles, upper observation) versus Circumpolar Deep Water delta-14-C respectively. Reproduced with permission from Matsumoto *et al*. (2004, their Fig. 2).

FIGURE 1

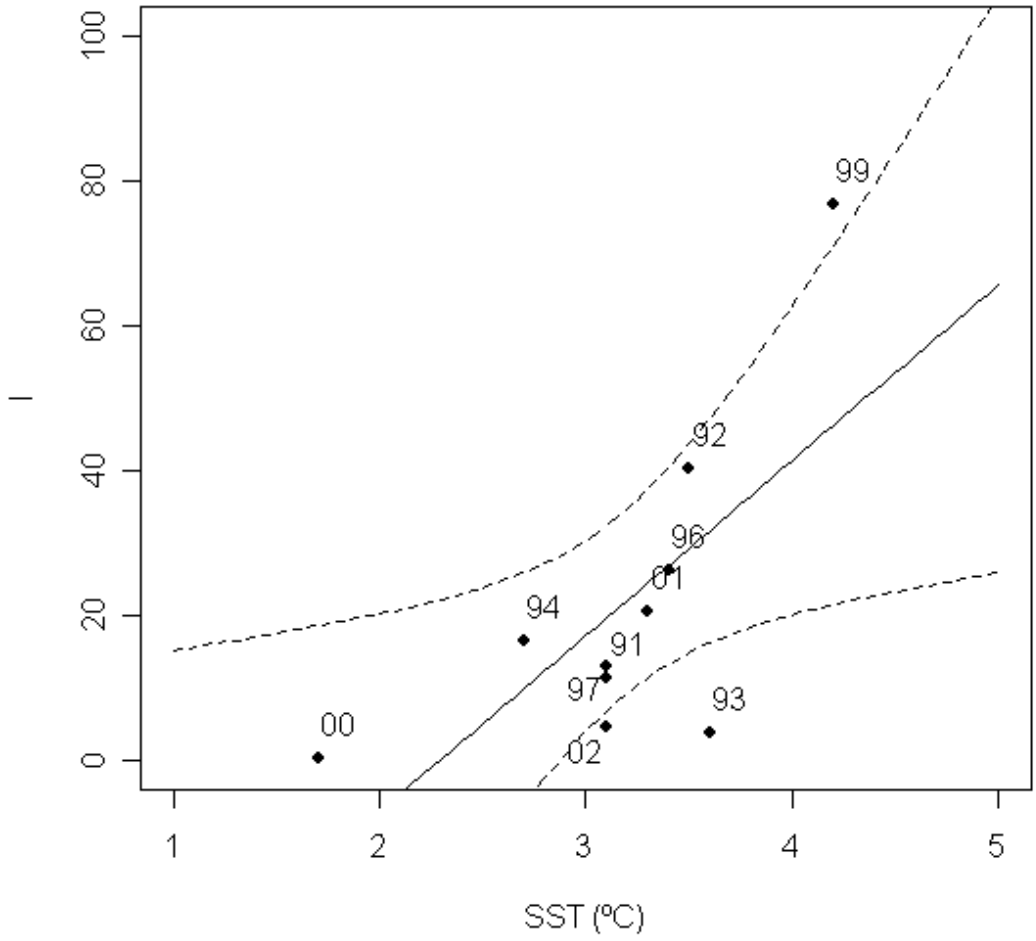


FIGURE 2

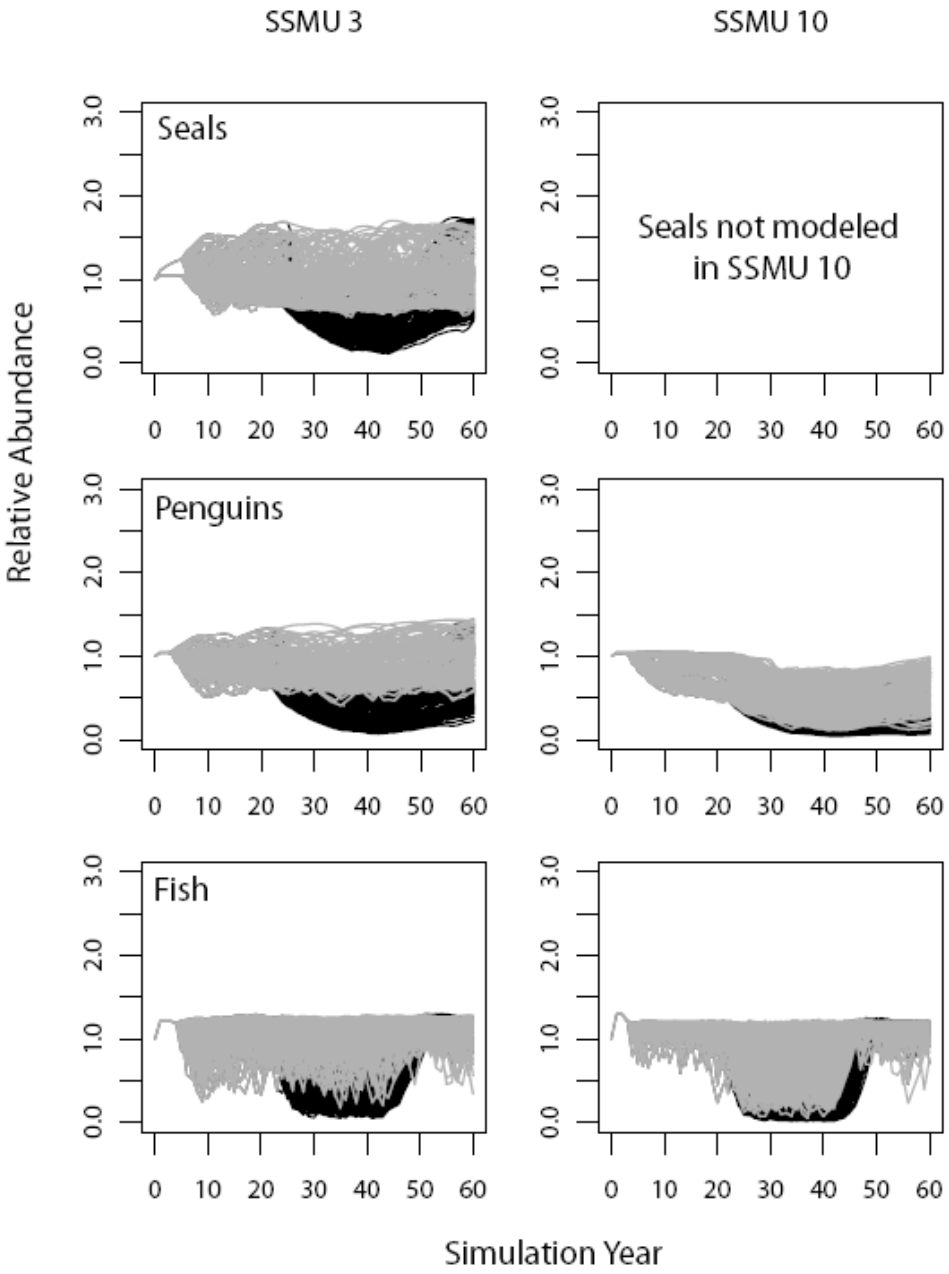

SSMU 3
SSMU 10
Seals
Seals not modeled in SSMU 10
Penguins
Fish
Relative Abundance
Simulation Year
0.0
1.0
2.0
3.0
0
10
20
30
40
50
60

FIGURE 3

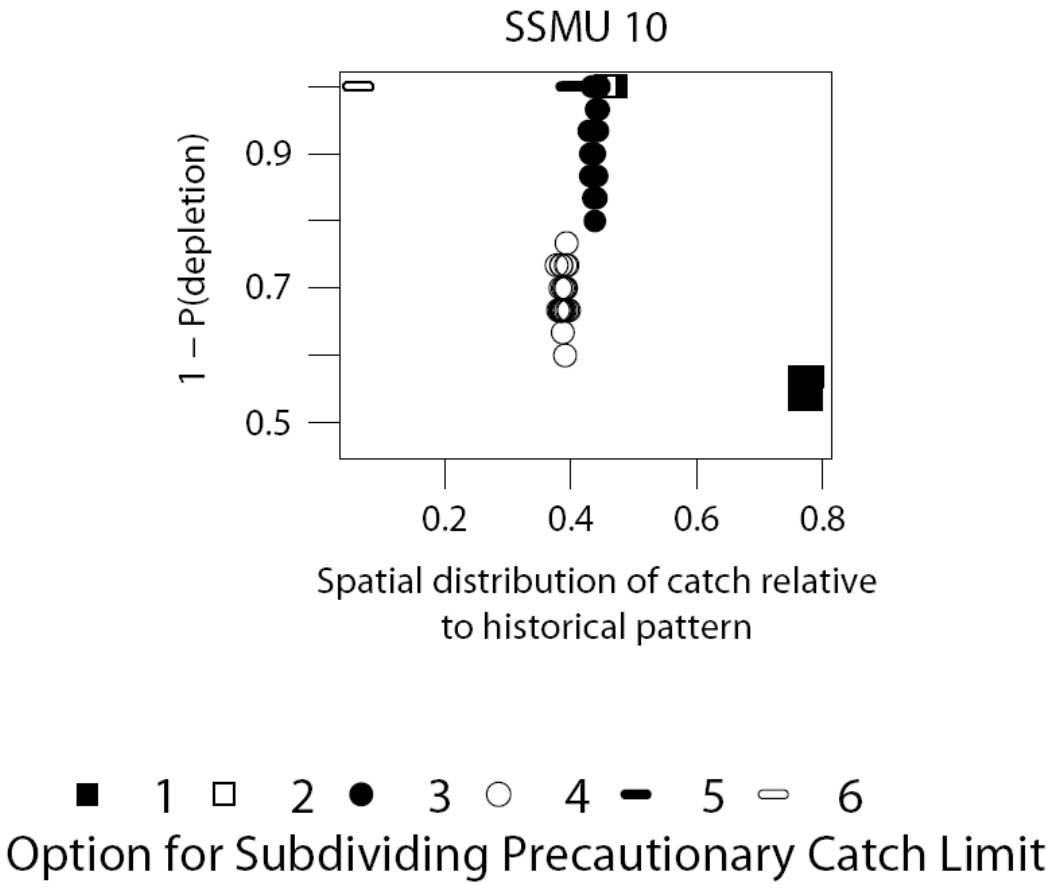

SSMU 10
1 – P(depletion)
0.9
0.7
0.5
0.2
0.4
0.6
0.8
Spatial distribution of catch relative to historical pattern
1
2
3
4
5
6
Option for Subdividing Precautionary Catch Limit

FIGURE 4

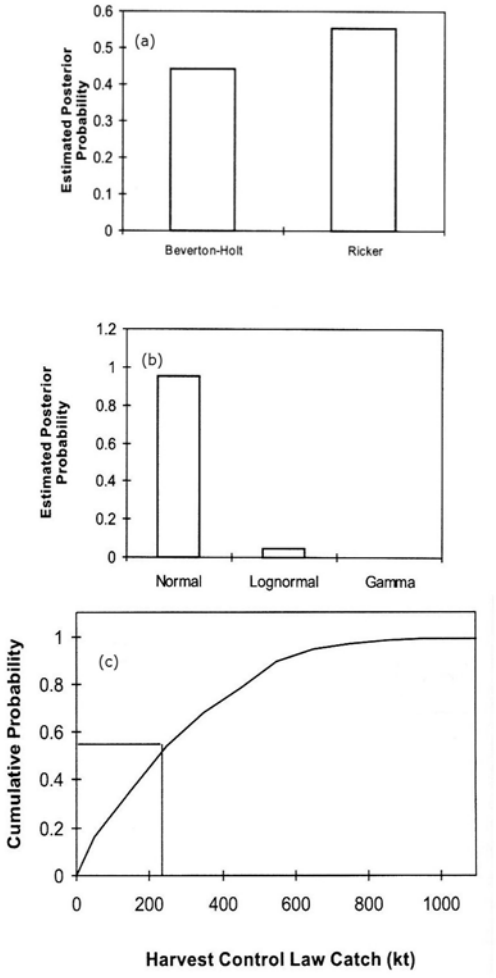

(a)
Estimated Posterior Probability
0.6
0.5
0.4
0.3
0.2
0.1
0
Beverton-Holt
Ricker
(b)
Estimated Posterior Probability
1.2
1
0.8
0.6
0.4
0.2
0
Normal
Lognormal
Gamma
(c)
Cumulative Probability
1
0.8
0.6
0.4
0.2
0
0
200
400
600
800
1000
Harvest Control Law Catch (kt)

FIGURE 5

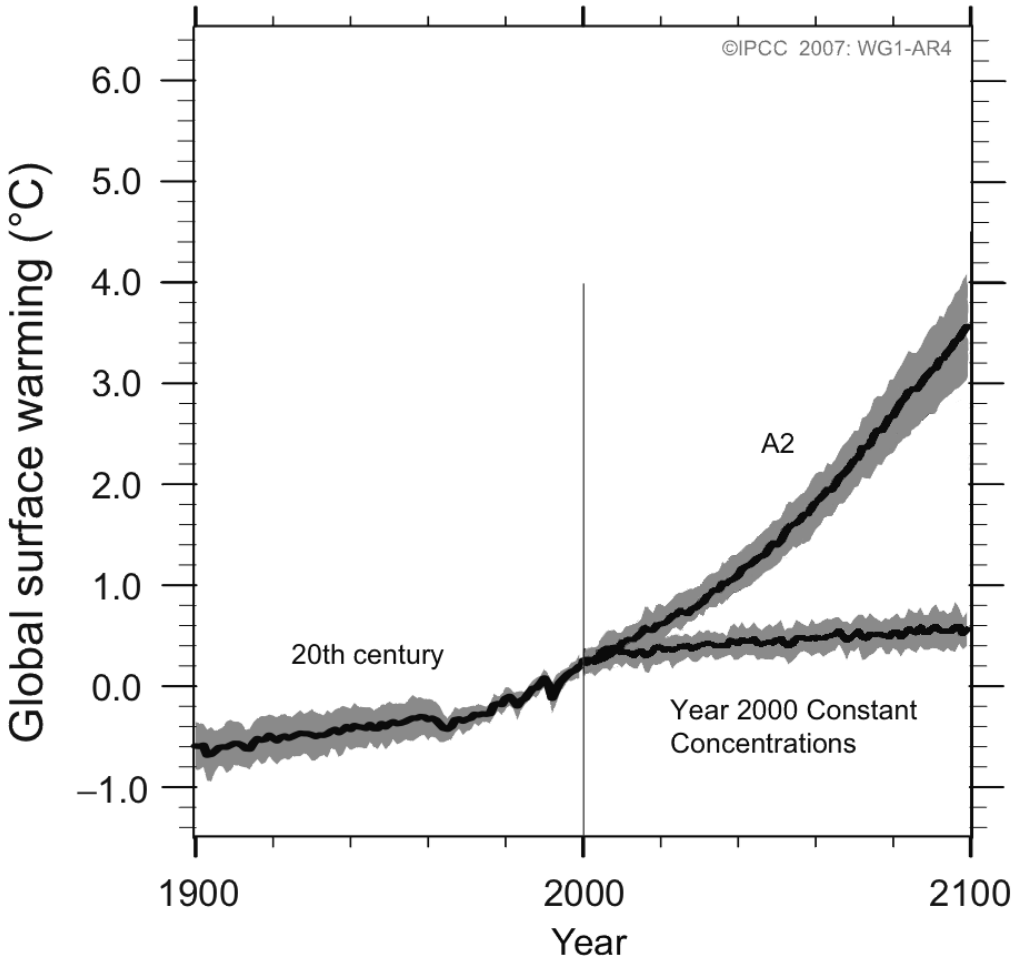
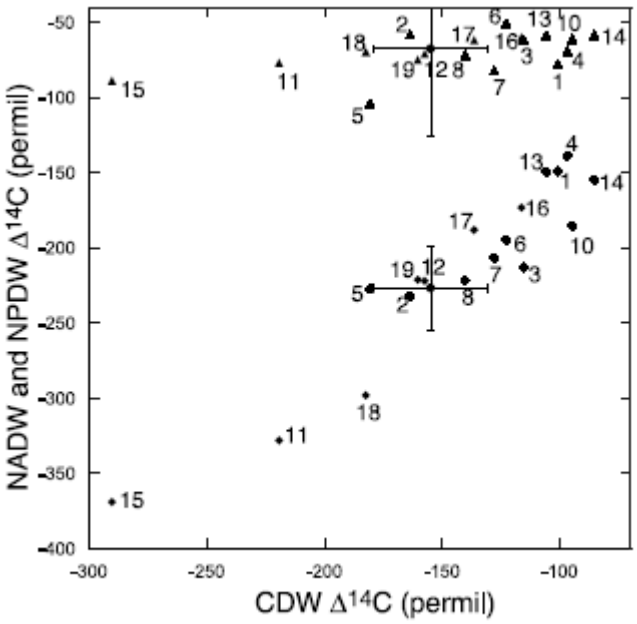

©IPCC 2007: WG1-AR4
Global surface warming (°C)
6.0
5.0
4.0
3.0
2.0
1.0
0.0
–1.0
1900
2000
2100
Year
A2
20th century
Year 2000 Constant
Concentrations

FIGURE 6


NADW and NPDW Δ14C (permil)
CDW Δ14C (permil)
-50
-100
-150
-200
-250
-300
-350
-400
-300
-250
-200
-150
-100