Estimating uncertainty in fish stock assessment and forecasting

Kenneth Patterson[1]*, *Robin Cook*[1], *Chris Darby*[2], *Stratis Gavaris*[3], *Laurence Kell*[2], *Peter Lewy*[4], *Benoît Mesnil*[5], *André Punt*[6]**, *Victor Restrepo*[7], *Dankert W. Skagen*[8] *& Gunnar Stefánsson*[9]

[1] FRS Marine Laboratory, PO Box 101, Victoria Road, Torry, Aberdeen, AB11 9DB, UK; [2] CEFAS Laboratory, Lowestoft, UK; [3] DFO, St Andrews Biological Station, Canada; [4] DIFRES, Charlottenlund, Denmark; [5] IFREMER, Laboratoire MAERHA, Nantes, France; [6] CSIRO Marine Research, Hobart, Australia; [7] ICCAT, Madrid, Spain; [8] IMR, Bergen, Norway; [9] MRI and University of Iceland, Reykjavik, Iceland.

Abstract

A variety of tools are available to quantify uncertainty in age-structured fish stock assessments and in management forecasts. These tools are based on particular choices for the underlying population dynamics model, the aspects of the assessment considered uncertain, and the approach for assessing uncertainty (Bayes, frequentist or likelihood). The current state of the art is advancing rapidly as a consequence of the availability of increased computational power, but there remains little consistency in the choices made for assessments and forecasts. This can be explained by several factors including the specifics of the species under consideration, the purpose for which the analysis is conducted and the institutional framework within which the methods are developed and used, including the availability and customary usage of software tools. Little testing of either the methods or their assumptions has yet been done. Thus, it is not possible to argue either that the methods perform well or perform poorly or that any particular conditioning choices are more appropriate in general terms than others. Despite much recent progress, fisheries science has yet to identify a means for identifying appropriate conditioning choices such that the probability distributions which are calculated for management purposes do adequately represent the probabilities of eventual real outcomes. Therefore, we conclude that increased focus should be placed on testing and carefully examining the choices made when conducting these analyses, and that more attention must be given to examining the sensitivity to alternative assumptions and model structures. Provision of advice concerning uncertainty in stock assessments should include consideration of such sensitivities, and should use model-averaging methods, decision tables or management procedure simulations in cases where advice is strongly sensitive to model assumptions.

Keywords Bayes, bootstrap, fish stock, fisheries assessment, model-averaging, Monte Carlo, uncertainty

***Correspondence and present address:**
Kenneth Patterson, European Commission, DG FISH, rue de la Loi 200, B-1049 Brussels, Belgium
E-mail: kenneth.patterson@cec.eu.int
Tel.: 32 2 2991111
Fax: 32 2 2955621
****Present address:**
School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 8195, USA

Received 21 Aug 2000
Accepted 16 Feb 2001

Introduction

The requirement for scientists to provide information to managers on uncertainty about stock assessments and forecasts arises in part from Article 7.5 of the FAO code of conduct for Responsible Fisheries (Anonymous 1995) which includes the commitment: "States should apply the precautionary approach widely to conservation, management and exploitation of living resources.... In implementing the precautionary approach, States should take into account, *inter alia*, uncertainties relating to the size and productivity of the stocks, reference points, stock condition in relation to such reference points, levels and distribution of fishing mortality and the impact of fishing mortalities ...". A similar phrasing can be found in Article 5(c) of the UN agreement on the Conservation and Management of Straddling Stocks (Anonymous 1994a; Sainsbury *et al*. 2000).

Partly on account of national commitments to these agreements, consideration of uncertainty has become an important part of the fisheries management decision process in regard to assessments of the current state of fishery resources as well as for short-, medium- and long-term forecasts. Usually there is a desire to know the probability of certain events, such as obtaining a low spawning-stock size. Furthermore, there is a growing public awareness that fish stock assessments and forecasts are not precise. Managers and stakeholders need to understand the nature and implications of uncertainty so that appropriate decisions are made. Treating assessments as exact can lead to management actions that are driven by error and natural variability rather than by new information.

Provision of fisheries advice in a form in which uncertainty is explicitly recognised and quantified is therefore becoming a standard requirement from management agencies. Scientists have developed a gallimaufry of recondite statistical techniques to identify the most important uncertainties and to evaluate their impacts for management purposes in response to this need.

These techniques have developed more or less independently over the last decade or so, and we contend that it is now appropriate to examine the present state of the art. This is particularly the case because methods considered computationally unfeasible in the past (such as Bayesian analysis and bootstrap) are increasingly now part of routine practice given the availability of computers of sufficient power.

To examine the state of the art of uncertainty estimation, we draw from our own experiences from several specific applications where attempts have been made to characterise uncertainty in fisheries stock assessment results. Based on an analysis of similarities and differences in these applications we draw out the key considerations for quantifying uncertainty. We first outline the management purposes for which uncertainty estimation is needed. This is followed by a review of the issues for which choices need to be made when estimating uncertainty in age-structured assessments and forecasting. These are classified according to structural model, error model and inference paradigm with associated estimation methods. We distinguish quantities or parameters having values of direct concern for management purposes (parameters of interest, typically spawning biomass or fishing mortality) from other parameters which describe processes not usually amenable to management (e.g. catchability or recruitment variability). For clarity, we introduce some definitions:

- By 'structural model', we refer to the set of deterministic relationships that are used to represent reality.
- By 'error models', we refer to the statistical descriptions that are used to describe the variability of observations of quantities that are governed by the structural model, to statistical descriptions of prior belief about model parameters, and to statistical descriptions of some highly variable and unpredictable processes.
- By 'inference paradigm', we refer to the set of assumptions upon which rests the translation from the data, structural model and error models to an estimate of uncertainty in a quantity of interest.

We stress the assumptions underlying the various statistical estimation methods rather than the technical detail. We reference both published case-studies and additional cases drawn from the 'grey' literature but which have in the most part served as a basis for management decisions. Here, we consider only examples of methods that model the age-structure of the population, but many of the considerations also apply to other approaches such as surplus-production modelling (e.g. Butterworth and Andrew 1984; Punt and Hilborn 1996).

Finally, and based on the foregoing review, we present our opinions about the state of the art of quantifying uncertainty in fisheries, and the comparative advantages and disadvantages of various approaches. As this is an evolving field in which little detailed work comparing alternative methods has been conducted, we acknowledge that many of our remarks are speculative and encourage other practitioners to challenge them and hence advance the field.

Management purposes

Stock assessment results can be used to support both decisions involving the establishment of strategies in the medium or long term, and the determination of the appropriate short-term tactics to conform to such strategies. Short-term tactics are usually measures that are implemented when a substantive new piece of information is obtained about a stock, e.g. a new estimate of stock size. Long-term strategies identify the desired state of the fisheries system. Medium-term strategies identify a means for altering the state of a fisheries system from its present state towards the desired state over a chosen time period. The design of a fisheries stock assessment model and, if required, a corresponding forecast is inescapably influenced by the nature of the decision which it will be used to support.

Establishing strategies

A topic closely related to estimating uncertainty is the design of management systems that are robust to the levels of uncertainty that are either measured or are believed to exist in fishery systems. Fisheries management decisions of the strategic type aim to identify policies that achieve declared objectives. Objectives tend to be broad, general statements about achieving acceptable compromises between obtaining high yields in weight or in value, maintaining employment on a sustainable and economically viable basis, and maintaining some low risk of resource depletion. Once objectives are defined, strategies designed to achieve the objectives may be sought.

Alternative strategies may be compared with respect to quantities of interest like central tendency for yield, magnitude of annual fluctuations in yield, exploitable biomass levels (reflecting catch rates and costs) and frequency that biomass may drop below some threshold. The temporal trajectories of these and other quantities of interest need to be examined over typically long time periods to contrast the consequences of candidate strategies. Such long-term forecasts typically aim to find steady-state characteristics for a range of assumptions about management (e.g. fishing mortality, selection pattern). The analysis can involve finding equilibrium points such as the fishing mortality rate that maximises yield, either by conventional means (Thompson and Bell 1934; Sissenwine and Shepherd 1987) or by introducing stochastic considerations and estimating stationary distributions for the quantities of interest (Skagen 1999). Calculations of reference points relating to long-term equilibria (such as MSY, maximum sustainable yield) and to stock collapse can be useful in addressing management questions of the strategic type and in designing long-term harvest strategies. Once a favoured strategy has been identified, it may be possible to devise a preagreed rule for managing changes in catches in response to new information about stock size.

The expected consequences of implementing such rules can then be evaluated by modelling the data-gathering, assessment procedure and implementation as part of an extensive simulation experiment (Butterworth *et al.* 1997; Kirkwood 1997; Butterworth and Punt 1999). Often, however, the decision-making process is embedded in a complex negotiating process which is not amenable either to preagreement or to simulation modelling. The important innovation introduced by the simulation experiment approach is that the consequences of uncertainty due to the assessment procedure and imperfect implementation of regulatory measures can be investigated directly. Such studies have been used to address a wide range of questions in fisheries, e.g. conditions for stability in exploited ecosystems (Collie and Spencer 1993), sensitivity to environmental variation (Basson 1999), choice of appropriate threshold levels under alternative biological assumptions (Zheng *et al.* 1993) and many other case-studies such as those given in Kruse *et al.* (1993) and Payne (1999). The studies tend to focus on the consequences for management decisions of some population dynamics parameters being subject to some type or degree of uncertainty, whether this is assumed or measured. The conclusions of studies of this type depend substantially on the estimates of uncertainty that are either assumed or derived elsewhere. Hence, we do not here review in detail such studies, which are seen principally as uses of the uncertainty estimates which are the subject of our paper.

On occasion, it is desired to devise tactics to achieve a medium-term objective. Medium-term forecasts are used more to support policy decisions over several years, and to represent the evolution of a population from an initial state towards another under different management regimes and states of nature, and hence to investigate the associated gains and costs of moving to different states. The forecast period for a medium-term forecast is usually related to the longevity of the species in question and, for many fish stocks, this tends to be in the range 5–10 years. In general this means that the population at the start of the forecast, which has typically been estimated, is completely replaced with unknown recruitment by the end of the forecast period. The results are therefore dependent on current estimated state of the stock and on an adequate representation of the population dynamics of the stock because the results are based on unmeasured future recruitment, survival, growth and maturation. Both uncertainties in the initial state of the population and uncertainties in the representation of population processes can have large influences in the outcome of such modelling exercises. Medium-term projections may be used in an *ad hoc* way to guide management decisions over a number of years (Anonymous 2000a).

Determining tactics

Short-term forecasts are generally used to support decisions about tactics or management measures (e.g. setting a quota) in relation to an established decision rule and reference points derived from the policy analysis. Such forecasts typically require estimates of the probabilities that parameters of interest would exceed their reference points, associated with the various alternative actions being considered for the immediate future fishing season. The quantities of interest in short-term projections are usually exploitation rate and biomass or spawning biomass, and changes in these quantities. The results are often presented in the form of the probability of some event occurring as a function of the quota (Fig. 1). The

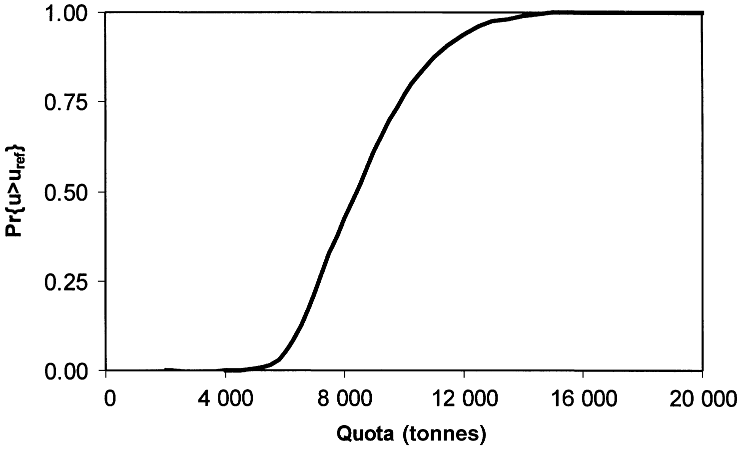

Figure 1 This type of presentation has been used to indicate to management agencies the probability that exploitation rate in the short term (u) will exceed a target exploitation rate (u_{ref}) for a range of possible quota options in the short term. Uncertainty in the target rate u_{ref} is not usually admitted. In principle, short-term management decisions can be based on a probability of exceeding the target rate which may be considered acceptable. Similar presentations have been made for biomass references.

results of short-term forecasts are strongly dependent on the abundance of recruited year-classes and that of year-classes that are about to enter the fishery. They are not influenced much by year-classes that will not be fully selected in the short term. Consequently, issues such as the form of the stock–recruitment relationship or changes in growth or predation are largely irrelevant for short-term forecasts.

Though uncommon, tactics such as quotas may be established for several future fishing seasons, generally not more than two or three years. These situations are more dependent on accurate predictive recruitment models.

Uncertainty and conditioning

It is inescapable that estimation of uncertainty is conditional on a set of assumptions or expert beliefs about some aspects of the analysis. This we refer to as 'conditioning' of the estimation. An excellent description of structural models and conditioning assumptions made in fisheries is provided by Quinn and Deriso (1999). Some examples of conditioning include:

- Predicating forecasts on a single-species population model
- Assuming lognormal distribution of errors for surveys
- Assuming a constant, known natural mortality
- Assuming recruitment will follow a Beverton–Holt function.

Assessment forecasts are predicated on a large number of conditioning assumptions, though many of these choices may not be explicit. The choices on which to condition are always subjective to some extent and no account is taken of the uncertainty surrounding how the conditioning is achieved. In other words, uncertainty due to stochastic noise about an accepted model is what is usually presented for management purposes, and this is only a part of the true uncertainty. Obviously, the skill required of a stock assessment scientist is the ability to make appropriate choices about conditioning assumptions, and the choices made will be very specific to each stock.

Some conditioning choices may be made to tailor an application for a particular management purpose. Typically, management questions relating to determination of immediate management measures depend mostly on uncertainty about the current state, while management questions relating to strategic issues are mostly affected by uncertainties in population dynamic processes. Accordingly, for example, models used in the former situation might condition on stationary somatic growth without significant loss in accuracy while models used in the latter situation may need to condition on density dependent somatic growth. Other conditioning choices can be considered 'biological' in nature in that they will be made on the basis of historic data about each stock for which a method is applied, such as the choice of stock–recruitment relationship. These conditioning choices often are the subject of informed, stock-specific discussions.

We classify conditioning according to choice of structural model, choice of error model and choice of inference paradigm. Within these classes, we consider the key conditioning choices that need to be addressed when conducting assessments and forecasts, and some of the alternative assumptions that are made for each. The issues and options identified cover the range that we have commonly encountered when providing management advice in the North-East Atlantic, Eastern Canada and Australasia (see Appendix). However, it should be recognised that the most important issue for a specific case may be one that is also unique to that case [see Table 2 of Butterworth and Punt (1999)]. This becomes increasingly likely once attempts are made to consider multispecies impacts.

Conditioning on structural models

In this section we describe some of the assumptions most frequently made about mortality, catchability, growth and maturation, dependence of recruitment on stock size, and ecosystem considerations. For convenience, we discuss here a few cases where the structural model in question is simply a parameter with a fixed value.

Mortality processes

Most single species age-structured models partition mortality into two components, fishing mortality, F, associated with the fishery harvest, and natural mortality, M, associated with all other causes of death. Denoting population abundance in numbers by N, the mortality dynamics are described by the system of differential equations:

$$\frac{dN}{dt} = -(F+M)N \tag{1}$$

$$\frac{dC}{dt} = FN \tag{2}$$

or their finite difference analogue. Solving the differential equations using year as the unit of time (y) and annual time increments yields the familiar exponential decay and catch equation for fish of age a.

$$N_{a+1,y+1} = N_{a,y}e^{-(F_{a,y}+M_{a,y})} \tag{3}$$

$$C_{a,y} = \frac{F_{a,y}N_{a,y}\left(1-e^{-(F_{a,y}+M_{a,y})}\right)}{(F_{a,y}+M_{a,y})} \tag{4}$$

It is not possible to estimate a complete array of age- and time-varying natural mortality. Often a single common value, or at most a few values common over large blocks, is prescribed. In many applications natural mortality is assumed known without uncertainty. Although it is highly desirable to consider uncertainty in natural mortality, it is often the case that no appropriate data exist for such estimation. If natural mortality can be treated as an estimable parameter of the model (e.g. Haist *et al.* 1993; Smith and Punt 1998; Patterson 1999), the data and, where applicable, the priors determine its value. In contrast, Monte Carlo methods that generate a value for natural mortality from some prior are based on the assumption that the assessment data provide essentially no information on natural mortality (Restrepo *et al.* 1992; Powers and Restrepo 1992; Mesnil 1993a, 1995). Almost invariably, natural mortality is considered a stationary process and forecast natural mortality for projections is drawn from the same estimated or assumed distribution. One additional, widely used approach is to condition a stock assessment on a fixed value for natural mortality but to admit that forecast natural mortality for projection may change by an amount for which a prior distribution may be specified (Anonymous 1993).

Though estimation of a complete array of age- and time-varying fishing mortality is technically possible, results tend to be unreliable and this approach is not common. Two frequently encountered techniques are employed to reduce the number of parameters required to be estimated. If the stock assessment model is conditioned on precise catches at age, then the structural model reduces to the conventional virtual population analysis (VPA) assumptions, i.e. the catch equation may be applied deterministically. This model is of course widely used and forms the representation of the mortality processes in the adaptive framework (ADAPT) (Gavaris 1988) and survivors analysis (Doubleday 1981) and its extensions (XSA) (Shepherd 1999). Additional conditioning constraints about the relative fishing mortality rates on some of the ages in the last year or on the older ages, or similar constraints about relative catchabilities on some older ages (see Catchability section below) will be required to obtain acceptable results. Without such constraints, VPAs tend to solutions for all $F \to 0$ and all $N \to \infty$.

Alternatively, if observation error is admitted for the catch at age, then it is necessary to formulate a structural model to represent belief about underlying mortality processes. The 'separable' models based on the approach of Fournier and Archibald (1982) condition on the assumption that fishing mortality F is dependent only on a year-effect f_y and on an age-effect S_a.

$$F_{a,y} = f_y S_a \tag{5}$$

Often, only observation errors are admitted, but a more recent approach is to admit that deviations from this underlying process do occur and can be estimated in a time-series framework (Gudmundsson 1994, 1998; Deriso *et al.* 1998; Ianelli and Fournier 1998; Millar and Meyer 2000).

In a surplus-production modelling context, Punt and Butterworth (1993) pointed out that the presence of a process error in a situation assessed by means of an observation-error estimator could lead to substantial negative bias in the estimates of variance. However, using similar approaches Pella (1993) reported that methods based on the Kalman filter could fail to estimate the precision of parameter estimates owing to singularity in the information matrix, when process error was admitted. Pella recommended the use of bootstrapping methods instead. Similar considerations probably apply in an age-structured context also.

When estimating mortality and population abundance, some analysts choose to include an assumption that either or both fishing mortality and population abundance are not very likely to have departed from the averages of recent years. Such 'shrinkage' (Copas 1983) of parameter estimates towards historic values is, in effect, including a form of prior in the analysis. As in the case when any priors are included in an analysis, there is a risk that it may introduce undesirable biases both in the parameters and in the variance estimates.

The choice of structural model may be highly dependent on the institutional background of the analyst, as there are clear geographical preferences. For most demersal stocks assessed in the North-East Atlantic area, the preferred analytic method is XSA. In the North-West Atlantic, ADAPT is widely used, whilst the 'separable' models are the methods of first choice in the North-East Pacific, Australia and for pelagic stocks in the North-East Atlantic. Biomass-dynamic models seem more widely used in South Africa and for the assessment of tuna populations than elsewhere. However, such choices also reflect regional data availability.

Catchability

Assumptions regarding the relationship between the indices of abundance and population numbers (catchability) are required by most methods of fisheries stock assessment. Given the importance of indices of abundance when conducting stock assessments, errors made when selecting a model for catchability can lead to major biases in estimates of quantities of interest. Traditionally, it has been assumed either that catchability is time-invariant (but age-specific) or that catchability changes in a simple fashion with abundance or time (Pope and Shepherd 1985). The assumption of time-invariant catchability is usually appropriate where an index of abundance is derived from a survey with standardised methodology, but where the index is derived from commercial fishing activity it is often appropriate to entertain hypotheses of time-dependent or stock-dependent catchability. Some assessment methods, and particularly XSA (Shepherd 1999), require further conditioning on assumptions of relative catchability at age. For example, it is usual in the application of this method to constrain some number of catchability parameters at older ages to be equal. More recently, approaches that allow for the possibility that catchability varies randomly over time have been developed (e.g. Ianelli and Fournier 1998; Porch 1998). These approaches treat changes in catchability between years as random variates and hence model catchability by means of an autoregressive procedure. Another option is to constrain catchability to a value which is 'known' from the survey design.

Changing conditioning on catchability assumptions can result in large changes in perceptions of stock size. For example, sensitivity of perceptions of mackerel stock size to assuming unit catchability in the egg survey biomass estimates was explored by Kolody and Patterson (1999). Residual diagnostics about the catchability relationship may be examined for goodness-of-fit but clear favouring of one catchability model may not be evident. It has been shown that for northern cod and some other stocks, the choice of catchability model may have a large influence on perceptions of both the point estimate and the attached uncertainty estimate of stock size (Walters and Ludwig 1994; Walters and Maguire 1996; Walters and Pearse 1996; Walters and Bonfil 1999). These studies show that much greater uncertainty is admitted if nonlinear hypotheses about catchability are entertained. Also, estimates of recruiting year-class abundance and hence of catches in the immediate future may be highly sensitive to catchability model choice (e.g. Faroes haddock, Anonymous 1999a).

Growth and maturity

It is important to consider uncertainty in all components of spawning stock biomass (numbers of fish, weight of fish and the proportion mature at each age). Density-dependent effects on growth and maturation are believed to be important factors that can have an impact on medium- and long-term forecasts (Patterson 1997; Helser and Brodziak 1998; Punt and Smith 1999). However, it is not common practice for these effects to be included in assessments or forecasts, more common practice being to assume that maturity- and weight-at-age are time-invariant. This is perhaps surprising as in several cases growth processes are known to have undergone long-term changes (Patterson 1997; Lilly 1998; Clark *et al.* 1999), and density dependence of growth has also been documented (e.g. Toresen 1986, 1988).

To the extent that one may expect larger fish to be more mature, one may expect that variations in maturity would be associated with changes in growth and condition. It is therefore probably important to preserve the linkage between weight- and maturity-at-age when conducting assessments and forecasts. A variety of different approaches to making allowance for uncertainty in weight- and maturity-at-age in some future year have been considered, and are described below under 'Error models'.

Stock–recruitment models

It is necessary to model the relationship between stock and recruitment to conduct medium- and long-term forecasts. This relationship is typically difficult to quantify for many reasons, including very high natural variability, large measurement errors and lack of sufficient data contrast. The parameters of the stock–recruitment relationship can be estimated in one of two ways. Either the historical estimates of stock and recruitment from a population model are treated as 'data' in an analysis subsequent to the stock assessment (e.g. Anonymous 1993; Bell and Stefánsson 1998; Skagen 1999) or the parameters of the stock–recruitment relationship are estimated along with other parameters in the population model-fitting process (e.g. Fournier and Archibald 1982; Haist *et al.* 1993; Smith and Punt 1998). In the latter instance, the estimates of other parameters (e.g. recruitment, fishing mortality) and their variances can be affected depending on the relative weight assigned to the stock–recruitment relationship in the objective function.

Many parametrisations for the stock recruitment are possible, including various forms with an autoregressive structure. The most common parametric relationships are the Ricker and Beverton–Holt relationships with lognormal errors. There are also applications of *ad hoc* (e.g. assuming a two-line model) and nonparametric (e.g. Evans and Rice 1988; Restrepo 1998) procedures. Recently, attention has begun to be focused on whether the stock–recruitment relationship may exhibit depensation (Myers *et al.* 1995; Liermann and Hilborn 1997) and Punt and Smith (1999) include a depensatory stock–recruitment relationship as one scenario in an evaluation of management procedures for Australia's eastern stock of gemfish.

Often, there is poor contrast in the historic observations of stock size, because data only exist over a time period when the stock has been relatively stable (in either a high or a low state). Under such conditions, only calculations conditioned on subjective assumptions about the form of the stock–recruitment relationship outside the observed range of stock sizes can be made. Uncertainty about the appropriate subjective choice may often largely dominate uncertainty in the forecasts. Nonparametric approaches to extrapolating outside the range of observed data have also been used in long-term forecasts (e.g. Cook 1998). A further problem is that that if there is poor contrast in the data, the estimated parameters in a stock–recruitment model will be highly correlated, which can create problems if they are treated as independent.

Some stocks have population dynamics that are dominated by infrequent but extremely large year-classes. Recruitment modelling remains problematic in such cases and, although forecasts based on mean expected recruitments can be calculated, the value of these for management purposes remains unclear. It is often argued that recruitment variability is dominated by environmental or other effects which are exogenous to the population dynamics, resulting in regime shifts or periods of high or low recruitment (Gilbert 1997). Modelling such dependencies has attracted a great deal of research effort, and many workers have shared the hope expressed by Hjort (1909) that "it will be possible by means of hydrographical researches to foretell important ocurrences both on land and in connection with the fisheries". However, in a recent review, Myers (1998) concluded that only a small proportion of reported relations between environmental variables and recruitment could be verified. Exceptions were generally stocks at the edge of the species' geographical range. Even for those cases in which robust relations describing the relationship between recruitment and environmental variables can be described, it is not possible to characterise recruitment variability unless externally specified probability distributions for such determining effects can be generated. The only means currently used to model the consequences of this is the introduction of various orders and magnitude of error autocorrelation when generating fluctuations in recruitment about a deterministic stock–recruitment relationship.

It is clear from the above discussion that the appropriate form of the stock–recruitment relationship remains a significant issue of debate for many stocks.

Thus, in addition to taking into account uncertainty about the values for the parameters of the relationship given a particular form, and variability in the annual recruitments conditioned on annual stock sizes, uncertainty about the form of the relationship needs to be considered. Though not routinely done, model uncertainty in the form of the stock–recruitment relationship can be incorporated in estimation procedures (Geiger and Koenings 1991; Adkinson and Peterman 1996; Patterson 1999).

Multispecies processes

Estimation of interspecific effects—usually predation—has been attempted through both age-structured (Multispecies VPA, Pope 1991) and biomass-structured approaches either in steady state ('ECOPATH'; Polovina and Ow 1985; Christensen and Pauly 1996; Walters *et al.* 1997) or as a biomass time series (Freeman and Kirkwood 1995). Fitting age-structured ecosystem models are very demanding of data, and so have been applied mostly in areas where extensive data resources exist, such as the North Sea (e.g. Pope and Macer 1991), Georges Bank (Collie and Tsou 1996) and the Baltic Sea (Sparholt 1994), or have been applied to a small subset of species (e.g. Punt and Butterworth 1995). Biomass-structured approaches have been used widely in data-poor situations (e.g. many papers in Christensen and Pauly 1993; Shannon and Jarre-Teichmann 1999; Bundy *et al.* 2000). The two approaches may lead to rather different perceptions of ecosystem dynamics even when applied to the same data (Christensen 1995).

Perceptions of long-term stock dynamics based on multispecies models may be very different from those based on comparable single-species models (Daan 1987; Pope and Macer 1991; Punt and Butterworth 1995), although this is not a general rule. Collie and Tsou (1996) found similar long-term trends with models of either type. However, the uncertainty introduced by the possibility of admitting multispecies effects is likely to be most important for strategic decisions, such as choice of target biomasses or exploitation rates. For this purpose, multispecies VPA and fleet interaction models arguably hold more promise than models of the 'ECOPATH' family (Larkin 1996).

Within the ICES area, use of multispecies models for management advice has generally been limited to the use of estimates of natural mortality derived from multispecies models, but still used within single-species models (Sissenwine and Daan 1991; Bogstad 1993). This is arguably because short-term forecasts are relatively insensitive to the changes in natural mortality that may occur owing to varying levels of predator abundance, and for this purpose there is little or no advantage in using the much more complex multispecies models (Bogstad *et al.* 1995; Magnusson 1995; Daníelsson *et al.* 1997; Tjelmeland and Bogstad 1998; Stefánsson *et al.* 1997, 1998, 1999). For example, Stefánsson and Baldursson (1998) compared different scenarios of model complexity involving Icelandic cod, capelin, shrimp, three species of whale and two species of seal. This study concluded that increasing model complexity had relatively little impact on catch forecasts (10% variation in mean cod catches between scenarios).

Clearly additional uncertainty may be introduced according to the possibility of conditioning advice on a multispecies rather than a single-species model. Extending the complexity of an assessment model also extends the complexity of the conditioning choices available. For example, multispecies model estimates can be highly sensitive to the assumed form of the predation relationship, about which there is little empirical information (Larsen and Gislason 1992; Magnusson 1993; Gislason and Sparre 1994; Rindorf *et al.* 1998). Furthermore, in such complex models, obscure parameter interactions may lead to inappropriate conclusions being drawn (Hilden 1988). Owing to the structural complexity of these models and the multiplicity of plausible error structures, implementation of any of the uncertainty-estimation procedures that we discuss below has only been very rarely attempted (Stefánsson and Baldursson 1998), and in such cases can prove computationally almost prohibitive.

Conditioning on error models

We describe some assumptions about the variability of various processes and the variability with which different sorts of information are gathered. These are grouped by type of information concerned.

Total catch

Most stock assessments condition on reported catch data (i.e. assume that these are known exactly), but reported catches frequently suffer from systematic bias due to a number of causes. These may include weaknesses in the catch reporting system, such as problems in gaining access to points of landing, or deliberate misreporting of catches in response to regulations. In addition, landings may only reflect part of the true catch if fish are discarded at sea. Where significant discarding occurs, but is unsampled, the use of landings data may be inappropriate. These distortions in the recorded catch, if uncorrected, lead to biases in estimates of quantities such as stock size and fishing mortality rate (Sinclair *et al.* 1996; Deriso *et al.* 1998; Patterson 1998). Furthermore, the nature of misreporting and discarding is such that even if estimates of these effects exist, there is considerable uncertainty about their magnitude and properties.

Accounting for systematic bias in catch information can sometimes be achieved in the estimation process of a stock assessment where adequate fishery-independent data exist and this may provide a means of characterising the associated uncertainties needed for forecasting (Anonymous 1999b). However, where misreporting and discarding respond to management actions in an unpredictable manner, it is extremely difficult to characterise uncertainty, yet this can be crucial in medium- or long-term forecasts. Problems of systematic bias in catches can be investigated by conducting assessments using a range of alternative catch series (e.g. Smith and Punt 1998). In most cases the difficulty in estimating even a distribution for catch misreporting has resulted in a failure to incorporate this source of uncertainty. The consequences of not doing so are likely to be highly case-specific, yet in general terms, more likely to result in bias in estimates of fishing mortality rates than in estimates of stock size and dependent quantities (Patterson 1998).

Catch-at-age

Catch-at-age data consist of information on catch and on proportion at age derived from sampling of the fishery. As noted above, the catch is often assumed known without error. When the structural model employed is the conventional 'VPA' type, it is also necessary to assume that the measurement error for the proportion at age is negligible. If it is assumed that the proportion-at-age data are observations subject to non-negligible measurement error, it is then necessary to select an appropriate observation error model for the catch at age. The most common assumptions are multinomial error (e.g. Fournier and Archibald 1982) and lognormal error (e.g. Deriso *et al.* 1985). Of these alternatives, the assumption of multinomial error would seem *a priori* to be the most appropriate as it accounts for correlations among age groups. Nevertheless, the lognormal assumption has been the more common in practice (e.g. Deriso *et al.* 1985; Kimura 1990; Smith and Punt 1998). This may be because stratified sampling schemes are often implemented, which would result in lower coefficients of variation at high and low ages than if purely random samples were taken. It would arguably be prudent to test the appropriateness of these competing models, but such evaluations are rarely undertaken (Crone and Sampson 1998).

Indices of abundance

Indices of abundance are usually obtained from research vessel surveys but occasionally also from commercial catch-rate information or tagging experiments. They comprise the most important information available for estimating the size of a fish stock, and the uncertainty about the relationship between the indices and population abundance is likely to be the major source of uncertainty for short-term forecasts. Consideration of uncertainty in indices of abundance is therefore central to estimating uncertainty in stock size using an assessment model. In general, the indices of abundance are assumed to be independent and lognormally distributed, despite problems encountered with zero observations. Some assessments (Patterson 1999) consider the alternative assumptions that the indices of abundance are normally or gamma distributed. Non-parametric methods for estimating uncertainty do not require specification of an error distribution but it is still necessary to ensure that the residuals are independent and identically distributed. When commercial catch-per-unit-effort information is used as an index of abundance, it would be reasonable to expect correlations between the catches at age and the abundance index. Account is rarely taken of such correlations.

Variances for different data sources

In addition to the error conditioning within the catch at age or within each index, it is usually necessary to estimate the relative variance between the various data arrays. The point estimates from the assessment are often very sensitive to how this is done. Alternatives are to apply maximum likelihood or iterative reweighting techniques (i.e. to condition the estimate of the data variance on the fitted stock assessment model) (e.g. Darby and Flatman 1994) or to heavily weight some parts of the data to estimate their inherent variability. Substantial doubt has been expressed about the stability of such estimators in situations where data are limited. Also, estimating data variances using Bayesian methods can be computationally prohibitive (Anonymous 1999c). This issue is related to the question of whether all data should be included in a single model-fitting exercise or alternatives explored using each data set independently and it is also related to the error-in-variables question (e.g. Polacheck *et al*. 1993; Schnute and Hilborn 1993; Stefánsson 1998). Additional external constraints may be imposed on estimates of data variances. For example, taper weighting (Cleveland 1981) may be used to make the point estimates robust to data errors earlier in the time-series.

Another approach is to estimate data variances externally to the assessment model. For example, simple multiplicative models [general linear models on log-scale—Shepherd and Nicholson (1986, 1991)] have been used for the estimation of the variances of various fisheries data sets, ranging from catch at age (Pope 1993) to mean weight and proportion mature at age (Anonymous 1994b). Such methods have been structured analogously to an analysis of variance, and depend on conditioning a variance estimate on a number of linearising assumptions, which typically include assumptions about time-invariant selection and catchability. In spite of these strict assumptions, these models are seen to explain most of the variation in the data and therefore often give plausible estimates of the error variation.

Sample estimates of survey data variances have been calculated by various means, most often by design-based estimators (Smith 1990) assuming a statistical error distribution within areas of homogeneous abundance (Leaman 1981; Gunderson 1993; Stefánsson 1996), geostatistical approaches (Rivoirard *et al*. 2000), cluster analysis (Williamson 1982) and resampling methods (Aglen 1989; Warren 1994; Smith 1997). However, there is usually concern that sampling variability is only a part of the overall survey variability (Lowe and Thompson 1993), and variance estimates so obtained are rarely used directly in assessments.

Sample estimates of catch rate variances can be obtained from model-based approaches that aim to remove confounding effects, e.g. vessel, gear, area and season, from the abundance signal. Analytical estimates are directly available (Gavaris 1980; Kimura 1981) although resampling methods have also been applied (Stanley 1992). As with survey variances, however, variance estimates so obtained are rarely used directly in assessments.

Constraints on variances, reflecting prior beliefs about the reliability of different data sources, have been the topic of much discussion at assessment working groups. Despite this, there is relatively little knowledge or study of the way in which making prior assumptions about data variances modifies any subsequent estimates of uncertainty. One cannot but suspect such effects to be large, given the often substantial effects constraints on data variances have on the point estimates.

Weight and maturity at age

If variability in growth and maturity is to be modelled, perhaps the simplest assumption is that observations of weight at age and maturity at age are made of fish which have time-invariant growth and maturation in the population, the annual differences being due to measurement error. A nonparametric representation of this variability was described by Skagen (1999), in which a year is drawn at random from the historical data series and both maturities and weights at age for this year are used for the future year to preserve the correlation between maturity at age and weight at age. A parametric approach to the same concept was provided by Patterson and Melvin (1996) amongst others. In this example, maturity at age is logit- or arcsin-transformed and the transformed variable is assumed to be normally distributed. The amount of variation in the normal distribution can be estimated from the historical data on interannual change in maturity.

Alternatively, one may make an assumption that variability occurs in the growth processes rather than in the observation. In the procedure used by Bell and Stefánsson (1998), a year is drawn at random and the changes in maturity and weight by cohort, from that year to the previous one, are used to determine the change in maturity and weight between the year of interest and its previous year (if maturity at age exceeds one, it is truncated at one).

More complex models can be contemplated, e.g. representing maturity at age using a growth or time-series model, which may even be functionally related to density (e.g. Smith and Punt 1998) and the sizes of other populations in the ecosystem. Nevertheless, none of the above approaches is able to mimic sudden regime shifts which, although rare events, are likely to have profound impacts on strategic decisions.

Recruitment

There is considerable diversity in how allowance is made for annual variability in recruitment when conducting forecasts. The annual recruitment is usually determined by using the deterministic component of the stock–recruitment relationship to find the expected recruitment for a given stock size and this is then modified by adding a residual to it. These residuals can be either resampled, based on an already fitted stock–recruitment relationship, or modelled as an autoregressive process (Anonymous 1993; Skagen 1999). Although these approaches seem intuitively appealing, they sometimes lead to time-streams of recruitment that appear biased and more variable than the historical observations of recruitment (Skagen 1999).

Inference paradigm and associated estimation methods

Two inference paradigms are in common use in fisheries: Bayes and frequentist. A third, the likelihood paradigm (Edwards 1972) is used only rarely. Wade (1999) provides an extremely helpful comparative overview of the three inference paradigms. The fundamental difference between the approaches is that frequentist methods treat parameters as unknowns that posses a 'true' value. In the likelihood and Bayes approaches, parameters are considered to be random variables. The Bayes methods are subjectivist in conception and provide a formal mechanism for the updating of belief upon acquisition of new data. Hence, to use Bayes methods it is necessary to have a formal description of the state of knowledge before information is analysed. Appropriate formulation of such 'prior' belief is often elusive. Likelihood methods are free of such prior distributional assumptions about parameters, though—as for all methods—they are dependent on belief about the appropriate representation of processes and form of error distributions.

Consider a situation in which a population with true parameters ξ is sampled and yields observations $\mathbf{D}$. The data provide an estimate of the ξ using an estimator $\hat{\xi}(\mathbf{D})$. Frequentist methods consider $P[\hat{\xi}(\mathbf{D})|\xi]$, the distribution of parameter estimates on repeated sampling, from which, e.g. confidence intervals for ξ can be calculated. Although it is formally incorrect to do so, such confidence statements are often loosely interpreted by making an inverse probability assumption and using such confidence statements as statements about $P(\xi|\mathbf{D})$, i.e. the probability of alternative values of ξ given the data.

In a Bayes approach, belief about ξ [expressed as a prior $P(\xi)$] is updated according to the probability model $P(\xi|\mathbf{D}) = P(\xi).\ P(\mathbf{D}|\xi)/\int P(\mathbf{D}|\xi)\ P(\xi)$, which can then be used to compare belief about alternative ξ. Under a likelihood paradigm, confidence in alternative ξ can be compared in proportion to $P(\xi|\mathbf{D})$ directly. The relative probability (termed likelihood) of alternative values of one element n of ξ, ξ_j say, is calculated as $P[\xi_j|\mathbf{D}\ (\xi^*_{n,n<>j})]$, in which $\xi^*_{n,n<>j}$ represents the parameter values corresponding to the maximum of the probability function P excluding parameter ξ_j, given the data $\mathbf{D}$ and fixed ξ_j. Fisheries stock assessment models are typically nonlinear in the model parameters, and uncertainty estimation using any of the methods necessitates the use of computationally intensive numerical techniques. More significantly, the nonlinearity of the models affects the sampling distributions. For frequentist methods, these effects have been studied and are associated with estimation bias. Various techniques have been developed to deal with estimation bias arising in complex nonlinear models and are discussed for the frequentist estimation methods. A Bayesian analogue to bias is not recognised and hence ways to deal with displacement of the posterior caused by nonlinearities have not been investigated.

Bayes methods

Bayesian approaches to summarising and quantifying uncertainty are widely used as the basis for decision analysis in a broad range of fields (Clemen 1996) and have recently been advocated for general use in the area of fisheries management advice (Punt and Hilborn 1997; McAllister and Kirkwood 1998). They provide a formal framework for calculating the probability of alternative values of population parameters given a particular data set, one or more structural models on which the calculation is conditioned and a number of beliefs about models or parameters from external sources (expert judgement or other, independent analyses) which are specified as prior distributions. Bayesian methods can be used to formally admit several structural models (Sainsbury 1991; Geiger and Koenings 1991; Draper 1995; Adkinson and Peterman 1996; Patterson 1999).

These methods therefore distinguish formally between the data for the stock being assessed (which contribute to the likelihood function) and inferences based on other sources (specified as prior distributions). Outputs from analyses of this type are usually presented as the probabilities associated with alternative values of the parameter of interest, integrated over the admitted distribution of the other parameters. Numerical integration methods have been developed to allow this computation to be calculated reasonably conveniently [Markov Chain Monte Carlo (MCMC) and Sample Importance Resample (SIR)]. Recent examples of fisheries applications of MCMC include Patterson (1999), Patterson and de Cárdenas (2000) and Millar and Meyer (2000) using age-structured data, Meyer and Millar (1999) using a delay-difference model, and McAllister and Ianelli (1997) using SIR. Conceptually, Bayes approaches to representing uncertainty are extremely simple. However, because analytical solutions are often not possible, the need to calculate numerical integrations of high dimensionality can place a large burden on the computational skill of the Bayesian analyst, although more recent software is making this less of a problem (e.g. 'WinBUGS', MRC Biostatistics unit and Imperial College School of Medicine, London, UK; 'AD Model Builder', Otter Research, Sidney, Canada). 'WinBUGS' is documented by Spiegelhalter *et al*. 1995.

Bayes techniques integrate prior information and data in a cohesive manner, allowing the introduction of conditioning in a structured way through the form of the likelihood function (and its underlying structural model) and through the specification of prior distributions. This strength of Bayesian methods, however, introduces complexity due to the requirement to specify prior distributions for all of the parameters. The appropriate choice of priors (including the attempts at specifying priors that represent ignorance) offers scope for extensive debates that rarely reach unambiguous conclusions. Non-parametric analogues of Bayesian techniques are not available.

Developing subjective prior distributions can be subject to considerable difficulty, as seen in the following examples.

(i) Priors are often chosen to be 'noninformative' although this need not be the case and, indeed, some authors (e.g. Punt and Hilborn 1997) have argued that the ability to include informative priors is one of the major benefits of Bayesian methods. It is usually the case that a prior that is noninformative for one quantity is highly informative for another (e.g. a uniform prior on current F corresponds to a very informative prior for current N within a VPA or ADAPT context because these parameters are linked structurally through the catch equation). Therefore, it is questionable whether any prior should be argued to be noninformative. The metric used when defining the prior (uniform, uniform on a log scale, etc.) is seldom considered when priors are specified based on bounds for parameters even though this decision can impact results substantially.

(ii) It has been observed that scientists have a tendency to under-estimate the true uncertainty when developing priors (Punt and Hilborn 1997).

(iii) It is easy, unintentionally, to base prior distributions on the outcomes from previous assessments of the same stock. This can be a major problem for Bayesian assessments but, because they do not update the priors, this is not necessarily a problem for frequentist methods.

Likelihood methods

Likelihood methods formally describe the probability of alternative model parameters given the observed data. This is arguably more attractive conceptually, if one admits that parameters are random variables, than frequentist methods—which describe the probability of alternative data sets being observed, given 'true' model parameters. Likelihood methods are also free of the need to define prior probability distributions, an advantage perceived by some in contrast to Bayes methods. However, where the results of Bayes analyses can be presented as a distribution for a parameter of interest integrated over the possible values of other parameters, in a likelihood approach a probability distribution for a parameter of interest is conditioned on the maximum-likelihood values of the other parameters. Numerical solutions to such problems are conveniently available only where the interest parameter is one of the formal model parameters, and not a transformation or projection from them. Also, calculating conditional likelihoods for some nuisance parameters such as variances entails particular difficulties (Royall (1997) covers the topic of eliminating nuisance parameters in detail). Plausibly for this reason, few instances of the application of likelihood methods are to be found in fisheries science, and those examples apply to simple models only. For example, in a surplus-production modelling context, Wade (1999) used a grid-step method to calculate likelihood profiles for the ratio of population size to carrying capacity. Values of three other model parameters were adjusted to find the maximum likelihood, subject to constraining the ratio to one value within a range of values in the step search. Punt and Hilborn (1996) also address the calculation of likelihood profiles in a surplus-production modelling context. Application of the approach to age-structured assessments with many parameters appears computationally problematic, although as Wade (1999) suggests, MCMC methods may be of use in such cases.

Frequentist methods

Frequentist approaches to statistical inference are, in the broadest sense, a type of decision making based on probability (Hoel 1971). As noted above, these methods rely on calculating a probability model describing the sampling distribution of estimators derived from the data, given particular parameter values which are assumed to be true. This model can be used to construct a confidence distribution for a parameter. A confidence level describes the probability that a calculated confidence interval will encompass the true value (e.g. 95% of confidence intervals calculated from a large number of data realisations would contain the true value). In practice, only one realisation is available, and its corresponding confidence distribution is the basis for the claim that the 'true' parameter value will be contained within an interval with a prescribed confidence, given a particular set of data and a structural model. This confidence distribution can be used to inform fisheries management decisions.

Results from these methods are therefore conditioned on there not being any relevant information left out of the observed data. Outputs from analyses are usually presented as confidence distributions for parameters of interest and are derived from the sampling distributions of their estimators, a trivial conversion for simple problems but more difficult for complex situations (Efron 1998; Schweder and Hjort 1999). For complex problems, results are obtained using the delta method approximation or bootstrap resampling techniques.

Frequentist methods do not offer a structured framework to incorporate prior information, to admit multiple model structures or to inject subjective belief. Frequentist analogues for a cohesive approach to integrating prior information and data have been proposed (Schweder and Hjort 1999) but are not in common use. 'Model weighting' as a technique to admit multiple model structures has only recently received greater attention (Buckland *et al.* 1997). Adjunct Monte Carlo methods (described below) have been employed to account for subjective beliefs about uncertainty in parameters for which the data are uninformative. Despite these limitations, frequentist methods remain attractive because they do not require specification of prior distributions for parameters, a significant complication as illustrated above. Frequentist methods using a Monte Carlo adjunct, however, require probability distributions for the model parameters that will be thus simulated. They serve the same role as priors in a Bayesian analysis although no allowance is made for the data to update the priors. A further attraction of frequentist methods is the availability of nonparametric techniques, permitting relaxation of the error distribution assumptions.

Delta method. The 'delta method' is a technique for deriving approximate variances and covariances from the Hessian matrix of mixed second derivatives of the objective function with respect to the model parameters. The confidence distributions of parameters of interest, being functions of the model parameters, can be derived from the delta estimates of the variance-covariance of the model parameters. This requires either an assumption about the distribution of the parameter of interest (such as spawning stock biomass or exploitation rate), or assumptions about the distributions of some of the estimated model parameters (such as population abundance or log-scale population abundance in the terminal year). These distributional assumptions impose conditioning on the uncertainty estimation which is additional to the conditioning assumptions made in the stock assessment model.

The covariance of model parameter estimators is calculated from the Hessian matrix assuming linearity near the solution. Estimators of both the model parameters and the parameters of interest can be biased for nonlinear models. An analytical approximation for the bias of model parameter estimators was developed by Box (1971).

Approximate estimates of variance and bias of parameters of interest can be derived analytically from the distributions of the model parameters. Ratkowsky (1983) gives a 'delta-like' approximation for the bias of functions of parameter estimators. Gavaris and Van Eeckhaute (1997) applied a naïve adjustment for the bias by shifting the confidence distribution to account for the magnitude of bias. We refer to these approaches as the 'analytic delta' and as the 'analytic shifted delta' when adjusted for bias.

The first-order approximation is not very reliable with small sample sizes and nonlinear models. A better approximation can be obtained by determining model parameters the distribution of which is approximated reasonably well by a Gaussian and then using numerical techniques to approximate the distribution of the parameters of interest.

It is particularly common to assume that the log-scale stock size in numbers at age in the first projection year comes from a multivariate Gaussian distribution (e.g. as advocated by Richards *et al.* 1998). Gavaris and Van Eeckhaute (1997) reported that log-scale population abundance parameter estimators displayed close to linear behaviour, therefore assuming a Gaussian distribution for them would be preferable to assuming a Gaussian distribution for the parameters of interest. This assumption should be tested because it may be questionable for some stocks for which the likelihood function is not adequately represented by a quadratic function about its minimum.

Under this assumption, log-scale population abundances are simulated from a multivariate Gaussian distribution with the estimated mean and covariance characteristics. The simulated values are submitted to the projection algorithm to derive the values of the interest parameter. These values are used to construct an empirical distribution function from which confidence statements may be drawn. We refer to this approach as the 'numerical delta'. When the population abundance parameters are adjusted for their bias before being submitted to the projection algorithm (Gavaris and Sinclair 1998), we refer to the approach as the 'numerical shifted delta'.

In the North-West Atlantic, the analytic shifted delta (e.g. Gavaris and Van Eeckhaute 1997) and the numerical shifted delta (e.g. Sinclair *et al*. 1996) have been used to provide advice for management purposes but results from bootstrap methods have been favoured recently. Many calculations within the framework of the International Council for the Exploration of the Sea (ICES) are provided for management purposes based on the numerical delta [e.g. sensitivity analysis (Cook 1993); use of 'WGMTERM' method (Anonymous 1993); 'ICP' (Patterson and Melvin 1996)]. The main difference is that the former uses only estimates of the variances of population abundances in the terminal year of the assessment, whereas the latter uses estimates of the covariances of estimated population abundances, fishing mortality and selection at age. Limited comparison of analytical shifted delta and numerical shifted delta, where the covariances of population abundance parameters were ignored for the numerical approach, did not reveal much difference though there was some indication that ignoring covariances could result in under-estimation of uncertainty (Sinclair and Gavaris 1996).

The analytic approach has the advantage of being computationally very efficient, but is relatively inflexible in that it is difficult to introduce or admit uncertainties other than uncertainty in those parameters that are estimated formally in a model-fitting procedure. A Gaussian approximation is not strongly supported for parameters of interest for fisheries management. Partly for these reasons, the analytic method is not currently used widely. Equivalent computations can be calculated numerically and, partly because of the greater flexibility to incorporate uncertainty for parameters external to the formal estimation and the more plausible assumptions about a Gaussian approximation for log population abundance, it has been applied widely in the ICES area (Anonymous 1993; Patterson and Melvin 1996).

Bootstrap. The fundamental idea of the bootstrap is to substitute the empirical distribution function as a simple estimate for the distribution of the sampling errors. This in turn can be used to infer confidence statements about parameters, as described above. Either the actual distribution of the observations can be used to construct a nonparametric empirical distribution function by resampling, or else a parametric model, with parameter values derived from the data, can be used to describe the empirical distribution. In either case, the data-based simulation must reproduce replicate samples with the same characteristics as the observed sample.

The simplest bootstrap technique for making inference statements and estimating uncertainty is the percentile method introduced by Efron (1979). This is implemented by simply simulating a large number of replicate samples and submitting them to the estimation procedure to obtain bootstrap replicates of the estimator. The bootstrap replicates are used directly to construct the bootstrap distribution of the estimator on which confidence statements can be based. Two important qualities of the percentile method are the transformation-respecting property and the range-preserving property. Transformation respecting means that confidence statements are not altered by any monotone parameter transformation. Range preserving means that the confidence intervals always fall within the allowable range, i.e. intrinsically positive quantities such as biomass will not have negative confidence limits, as can occur with some other methods. Not all bootstrap techniques possess these properties.

For complex estimation procedures, such as nonlinear stock assessment models, the percentile method can be improved by accounting for bias (Efron 1982) and for variance acceleration (Efron 1987). Variance acceleration is a quantity used by Efron to describe the rate of change of the standard error of a parameter estimate with respect to the true value of the parameter. This bias-corrected and accelerated bootstrap is an automatic way of ensuring a proper translation from a sampling distribution to a confidence distribution. These bootstrap variants retain the transformation-respecting and range-preserving properties. We describe further below the parametric and nonparametric variants of this approach.

Parametric bootstrap. We noted above that when using a parametric bootstrap, the empirical distribution function is estimated using a parametric model of the data. This is implemented by assuming that the observations were drawn from a distribution with a particular parametric form. The parameters describing the distribution are estimated from the sample observations. Bootstrap replicate samples are generated by drawing from the distribution characterised by the estimated parameters. This variant of the bootstrap relies on parametric conditioning of error distributions for the observed data.

The distributions used to generate the data may be derived from statistical analyses external to the model estimation (e.g. resampling from very-fine-scale sampling information; Kell *et al.* 1999) or by using Shepherd and Nicholson's (1991) method to estimate the mean and variance of age and year effects to specify the distributions, as used by Bell and Stefánsson (1998) and Stefánsson and Bell (1998).

Nonparametric bootstrap. One way of circumventing the need to make distributional assumptions about observation errors is to use the observed data as the estimate of their empirical distribution function directly. This is implemented by generating bootstrap replicate samples by resampling with replacement from the observed data. For complex problems, this procedure may be onerous or impractical. Smith and Gavaris (1993) were able to implement this approach for a stock assessment model in a special case where the sampling intensity for the stratified random survey was large enough to permit resampling. This approach, if it can be implemented, produces the ideal bootstrap as it does not condition on error distributions or model results.

A simpler and more practical nonparametric bootstrap is based on resampling model residuals and is thus model conditioned (Efron and Tibshirani 1993). This is implemented by generating bootstrap replicate samples by resampling from the model residuals about the fit and adding these to the model-predicted values for the observed data (e.g. Deriso *et al.* 1985; Mohn 1993; Anonymous 1997a). This method does not directly condition on the assumption of a particular parametric form for the residuals, but it does condition on the assumption that the residuals (often after some chosen transformation such as the lognormal) are independent and identically distributed.

Adjunct Monte Carlo. Monte Carlo is a simulation method in which the parameters of the distributions used to simulate replicate values are not based on observed data, in contrast to the bootstrap. This approach can be used as an adjunct to the delta method or the bootstrap in circumstances where it may be desirable to represent additional uncertainty that is not estimable, such as that of some input parameters the values of which are typically assumed based on expert knowledge. Monte Carlo replicate samples for parameters which are not estimated in the assessment may be generated from a distribution that is specified completely externally to the data or analysis, and used as an adjunct to the bootstrap replicate samples (Restrepo *et al.* 1991, 1992; Mesnil 1993a, 1995; Hanchet *et al.* 1998). Such applications typically involve using random-number generators to provide realisations from assumed distributions and to fit complex nonlinear models to the resulting realisations. However, where the assumed distributions are not too complex and where linear or simple nonlinear models are in question, fuzzy arithmetic tools (e.g. fuzzy regression) can be used to similar ends, with some gain in simplicity and clarity (Saila and Ferson 1998).

Although lacking a firm theoretical basis (Poole *et al.* 1999), the Monte Carlo method is very flexible and can be used to explore the consequences of uncertainty about parameters that are normally considered fixed when conducting assessments. For example, variability can be introduced in natural mortality, weights at age and future recruitments. On completion, empirical distributions are obtained for all quantities of interest for management and can be interpreted to evaluate the probability, or risk, of relevant outcomes under the simulated regimes. The key issue with this approach is how to condition the analysis on appropriate probability distributions. These can, as is the case for Bayesian approaches, be based on expert opinion as described above.

Monte Carlo methods have been criticised (Poole *et al.* 1999) because they can provide misleading results if the data included in the objective function provide information on the parameters that are assigned 'external' distributions (but are not updated by the data). An example of the latter is the use of tagging data to estimate a growth curve which is in turn used to generate catch-at-age data which are then input into an ADAPT analysis. A bias will occur if the data used in the ADAPT analysis provided information on the growth curve. In such cases, it is preferable that all observations that are informative with respect to the same parameters should be analysed in a single framework. If different parts of the data are analysed in separate stages, then the result of the analysis will be unlikely to reflect the appropriate distribution of parameter values, because conflicts or coincidences in the data will not be represented.

Although examples can be constructed where these concerns can severely bias the results, such examples are unlikely to be common for most fisheries applications because it is seldom the case that the data provide information about the quantities assigned 'external' distributions in Monte Carlo analyses (e.g. natural mortality).

Summary of inference methods

Issues related to methods used to estimate uncertainty are perhaps more general than issues related to structural and error model conditioning or model uncertainty. The philosophical debate about whether parameters are random variables for which we can discuss probability distributions or whether parameters are unknown 'true' quantities for which we can discuss confidence distributions appears to have little practical significance in cases where the data are highly informative. However, in other cases inferences drawn about population parameters can be strongly dependent on the inference paradigm used (Wade 1999). Within each inference paradigm, proper interpretation of statistical properties is possible. The ability of estimation methods associated with Bayes, likelihood or frequentist paradigms to accommodate the diverse conditioning requirements for fisheries problems is, however, a relevant concern.

Frequentist and Bayesian techniques require the practitioner to gain some expertise in order to avoid pitfalls. Delta methods rely on parametric assumptions about the distributions of parameters of interest and on linear and quadratic approximations which may not be justifiable in some circumstances. Bootstrap techniques depend on replicating samples with the same properties as the original observations, sometimes a tricky matter. Accurate bootstrap confidence statements require that measures are taken into account for estimation bias and variance acceleration when these are important. There is a lack of agreement on how to assess whether the algorithms for computing Bayesian posteriors have converged adequately. Elimination of nuisance parameters in likelihood methods is technically difficult, and has therefore limited their use in fisheries cases.

Application of Bayes methods rather than frequentist or likelihood methods can result in different estimates of the mean and dispersion of the quantities of interest, because the Bayes methods allow incorporation of prior knowledge and beliefs into the analysis. This is advantageous if the introduction of such priors has a sound basis, but uncritical introduction of prior beliefs can result in unexpected effects, especially if inferences about different quantities are drawn from different sources when these are related through the model or are correlated. Punt and Butterworth (2000) describe this problem thoroughly for an instance involving bowhead whales. A further problem, also unique to the Bayesian approach, is the Borel paradox. This arises when (unintentionally) more priors are specified than there are parameters; the result is then a nonunique prior. For example, it may appear plausible to represent ignorance about exploitation rate and abundance by specifying separate, uninformative priors but in many assessment models the estimates of these parameters will be very closely correlated, or even structurally linked. In such cases, the result can be to introduce highly informative priors unintentionally. Even experienced practitioners (e.g. Raftery *et al*. 1995) can make this type of mistake (Wolpert 1995). Overall, we stress that the specification of a noninformative prior requires substantial care.

Concluding this section, we offer our opinions about some properties of uncertainty estimation methods that we consider desirable in general terms as follows.

(i) The method should have known statistical properties. Some applications of the Monte Carlo method may involve a number of untested (but plausible) assumptions, the statistical properties of which are unknown. In general terms, we express our preference for the use of statistical distributions derived from data when attempting to quantify uncertainty, though there is a place for the use of arbitrary Monte Carlo distributions when exploring robustness to the precision of the data.

(ii) Accounting for displacement of frequentist confidence distributions or Bayesian posterior distributions caused by nonlinear models is probably a desirable property. Though estimates of bias are routinely reported in many assessments for the North-West Atlantic, bias-corrected confidence distributions are not always used and assessments in other areas often do not examine bias. Though bias-corrected and accelerated bootstrap techniques are recommended for routine application to obtain confidence distributions, bias correction of point estimates requires case-specific consideration, as the precision with which bias is estimated can be low (Efron and Tibshirani 1993).

(iii) Accuracy of probability coverage with respect to the parameters of interest is also a desirable property. Anonymous (1999d) examined probability coverage for nonparametric bootstraps based on ADAPT and found that substantial improvements in performance can be achieved by applying a bias-correction factor. In a Bayes context, bias is less easy to interpret and has, as yet, received little attention. Whatever approach is chosen, we recommend that methods should be tested to evaluate the accuracy of probability coverage before they are used for advisory purposes.

(iv) Some methods (e.g. delta methods) make more restrictive conditioning on error distributions and on variance structure than others (e.g. nonparametric bootstrap). Efron and Gong (1983) point out that the jackknife is almost a bootstrap conditioned on a linear approximation and express the opinion that the delta method is inferior. Comparative studies of the practical implications of these assertions for fisheries problems are limited. Gavaris (1999) found that inferences based on analytical delta methods diverged from results obtained with bootstrap methods, while Sinclair and Gavaris (1996) determined that results based on numerical delta and bootstrap methods were largely comparable. One should prefer methods that rely either on fewer distributional assumptions or on those that are shown to be robust to mis-specification of such assumptions.

(v) For routine use, it would be preferable to apply an approach that can be implemented quickly and understood relatively simply. However, ease of application should not be used as a reason to reject a slower, more complex but more defensible approach!

Applications in example cases

Having outlined the management purposes and the range of conditioning assumptions associated with structural models, error models and inference, we now contrast the usage of such devices in the applications we introduced. The features for each specific application are summarised briefly in Table 1, and an expanded description of these cases is provided in the Appendix. It should be noted that while we believe these examples are typical of those in use in age-structured assessments world-wide, the choices made in their development were influenced by several factors including the species being assessed and the purpose for which the results were used. These examples are drawn from a range of data-rich situations within which the estimation of parameters and the estimation of uncertainty using statistical procedures has at least the appearance of being a tractable proposition. Clearly these methods are inapplicable in data-poor situations, where the greatest uncertainties would be due to lack of knowledge rather than to any estimable uncertainty or variability.

Three applications considered only the projection phase of the decision problem and drew from independent analyses for the required input. This is a convenient approach which simplifies computation time considerably but does not address the suitability of those independent analyses. Further, care must be taken to ensure that the inputs derived from those independent analyses are not linked, possibly giving different results if the analyses were completed simultaneously in one estimation framework.

With respect to management purposes it is noteworthy that most applications were designed to address strategic decision issues. Only the Eastern Georges Bank haddock case-study focused exclusively on tactical decisions for setting of annual total allowable catch (TAC) in relation to prescribed established reference points, while three cases were aimed solely at exploring harvest strategies and policies. Three of the applications (North Sea plaice, Icelandic cod and eastern gemfish) evaluated management procedures to help managers formulate a harvest rule for setting future quotas while the remaining applications concentrated on the future implications of different levels of fishing mortality. The North Sea plaice applications that evaluated management procedures took a short cut. This involved assuming that the process of applying a stock assessment method to simulated data to obtain estimates of the age structure of the population can be modelled by simply assuming that the estimated numbers at age are randomly distributed about the actual numbers at age.

The structural models were quite diverse but there were notable commonalities. Only three applications included estimation of M in order to capture that source of uncertainty. The northern hake application attempted to capture uncertainty in M by augmenting the bootstrap estimation procedure with a Monte Carlo based on an externally prescribed uniform distribution. Other applications prescribed an assumed known constant. There was a fairly equal mix of 'simple VPA' and 'separable VPA' modelling of F, probably reflecting the analysts' perceptions regarding the quality of catch sampling. Only the Bayes random walk method departed from time-invariant index catchability models to permit a random walk process in these parameters. Modelling of growth and maturity was not incorporated in most cases, possibly reflecting the belief that these processes have a second-order effect. The impact of recruitment modelling, especially for strategic evaluations, was well recognised and included in several analyses, though diverse approaches were used to handle the problem. Most of the methods ignore the stock–recruitment relationship when fitting the assessment model. Instead, they assess uncertainty in the values for the parameters of the stock–recruitment relationship by re-estimating the parameters of the stock–recruit model for each bootstrap replicate/draw of the variance–covariance matrix of the parameters. This was probably owing to the failure of the data to support any parametric stock–recruitment relationship and the associated reluctance to have estimates of stock status influenced by such conditioning. Only the study on Norwegian spring-spawning herring explicitly considers model-structure uncertainty by placing priors on different stock–recruitment relationships. The Icelandic cod example is unique among the examples considered because account was taken in this case of multispecies impacts. Only two cases allowed for error in the total reported catches: for West Scotland herring by including additional misreporting parameters in the estimation, and for eastern gemfish by exploring two alternative time series of catch. The admission of error in the catch at age was associated with the use of 'simple VPA' or 'separable VPA' structural models and, as indicated earlier, probably reflected beliefs about the quality of catch sampling. Indices were most often assumed lognormal or, for nonparametric methods, independent and identically distributed on the log scale, about the model. The Norwegian spring-spawning application attempted to use information in the data to choose between Gaussian, lognormal and gamma distributions during optimisation. Variances for different data sources were most often assigned, though there were some attempts at estimation. Error in weight and maturity were typically not admitted, corresponding to the absence of modelling for these processes.

While many cases were typified by a dominant method of estimating uncertainty, generally associated with the method used for the uncertainty in population abundance, these methods were often augmented by different techniques for other sources of uncertainty. This was particularly the case for frequentist approaches. Only the Eastern Georges Bank haddock application considered adjustment for estimation bias.

From this comparison of case-studies, it is clear that although they are designed for generally similar purposes, to provide managers with statements about the probability of various outcomes of alternative management actions, there is a large variety in the conditioning assumptions used. We have identified a large number of issues for which conditioning assumptions must be made when estimating uncertainty in fish stock assessments and forecasts. Some of these have a direct biological significance whilst others are implicit in the detail of the statistical methodology used. In some cases, we have been able to identify cases where assessment results have been shown to be highly sensitive to conditioning assumptions. In many other cases the sensitivity of assessments and corresponding uncertainty estimates to different conditioning choices is not documented or remains unknown.

It is difficult to draw general conclusions about the extent to which the conditioning used in real applications has been appropriate because conditioning is highly specific to species and management purpose. Nevertheless, there are some cases in which the estimated uncertainty has clearly been unduly under-represented owing to inappropriate conditioning. For example, for North Sea herring (Anonymous 1998) there is clearly quite a good deal of uncertainty related to different perceptions of stock size indicated by different index series, with fishing mortality in the range 0.1–0.7 depending on ascribed variances for the various indices. More dramatically, for Norwegian spring-spawning herring, structural uncertainty about the stock–recruitment model indicated a range of stock sizes from 2.3 million to 12.08 million tonnes (Anonymous 1997b). However, the estimates provided to the managers did not convey the full range of uncertainty in these cases. One may add that in two cases commonly viewed as significant failures to forecast the development of fisheries, Peruvian anchovy (Csirke 1988) and northern cod (e.g. Myers *et al.* 1997), failure to consider the full uncertainty by relaxing inappropriate conditioning can be viewed as a contributing factor.

Discussion

Provision of uncertainty estimates for management advice in fisheries is a relatively young science, with few examples to be found before the introduction of such practices at the International Whaling Commission in the early 1980s (e.g. Beddington and Cooke 1981; Kirkwood 1981). The earliest of the age-structured fish stock assessment examples we consider in detail was due to Restrepo *et al.* (1991). Since then, many methods have been developed for more or less similar purposes, and estimates of uncertainty based on these have been used for management

Table 1 Summary of the principal characteristics of some specific applications of uncertainty estimation methods. *M*, natural mortality; *F*, fishing mortality.

	Assessment and projection								Projection only		
	Northern hake	Norwegian spring-spawn herring	West Scotland herring	North Sea plaice [1]	Bayes random walk	Georges Bank haddock	Eastern gemfish	North Sea herring	North Sea herring	North Sea cod	Icelandic cod
Management purpose	Strategic	Strategic	Tactical and strategic	Strategic	Tactical [2]	Tactical	Tactical and strategic	Strategic	Strategic	Strategic	Strategic
Structural model											
M	Prescribed uniform distribution	Estimated	Prescribed constant	Prescribed constant	Estimated	Prescribed constant	Estimated	Prescribed constant	Prescribed constant	Prescribed constant	Prescribed dynamic
F	Determined by catch; constrained at oldest age	Determined by catch; constrained at oldest age	Separable	Determined by catch	Separable (in expectation)	Determined by catch; constrained at oldest age	Separable	Separable	Determined by catch	Determined by catch	Determined by catch
Catchability	Time-invariant by index and age	Time-invariant by index and age	Time-invariant by index and age	Time-invariant by index and age; constrained at oldest age	Random walk by index and age	Time-invariant by index and age	Time-invariant by index and length	Time-invariant by index and age	N/A	Time-invariant by index and age; constrained at oldest age	N/A
Growth and maturity	Time-invariant	Not modelled	Not modelled	Not modelled	Not modelled	Not modelled	Prescribed growth parameters; estimated maturity ogive	Not modelled	Time-invariant	Not modelled	Modelled as stochastic increments
Recruitment	S–R estimated after assessment; Ricker model	Optimised over Beverton–Holt and Ricker models	Not modelled	Not modelled	Not modelled	Not modelled	S–R estimated in assessment; specified model form	S–R estimated in assessment; specified model form	S–R estimated after assessment; Beverton–Holt model	S–R estimated after assessment; specified model form	S–R estimated after assessment; specified model form
Ecosystem	Not modelled	Not modelled	Not modelled	Not modelled	Not modelled	Not modelled	Not modelled	Not modelled	Not modelled	Not modelled	Inter- and intra-species interactions

Table 1 *continued*

	Assessment and projection								Projection only		
	Northern hake	Norwegian spring-spawn herring	West Scotland herring	North Sea plaice [1]	Bayes random walk	Georges Bank haddock	Eastern gemfish	North Sea herring	North Sea herring	North Sea cod	Icelandic cod
Error model											
Total catch	No error	No error	Uniform prior for misreporting multiplier	No error	No error	No error	Two fixed catch series	No error	No error	No error	No error
Catch-at-age	No error [3]	No error	Lognormal	No error	Multivariate normal	No error	Lognormal	Lognormal	No error	No error	No error
Indices	Lognormal	Poisson (tags); optimised over Gaussian, lognormal and gamma	Lognormal	Lid, on log scale, about model	Multivariate normal	Lid, on log scale, about model	Lognormal	Lognormal	N/A	N/A	N/A
Variances	Assigned equal	Assigned equal	Assigned equal	Iterative re-weighting	Estimated	Assigned equal	Assigned	Assigned equal	N/A	N/A	N/A
Weight and maturity	No error	No error	No error	No error	No error	No error	No error (weights); uniform priors for L_{50} and L_{full} (maturity)	No error	N/A	No error	Modelled as stochastic increments
Estimation	Parametric bootstrap for population augmented by Monte Carlo for *M*, non-parametric residual bootstrap for S–R, resample observations for weight and maturity	Bayes MCMC	Bayes MCMC	Non-parametric bootstrap for population augmented by Monte Carlo for other parameters	Bayes MCMC	Analytical shifted delta; non-parametric residual bootstrap, bias corrected	Bayes SIR for population augmented by parametric residual bootstrap for S–R projections	Numerical delta for population augmented by nonparametric residual bootstrap for S–R projections	Numerical delta for population augmented by nonparametric residual bootstrap for S–R projections, resample observations by year for weight and maturity	Numerical delta for population augmented by nonparametric residual bootstrap for S–R projections	Numerical delta for population augmented by Monte Carlo for other parameters

[1] Description of the assessment model, the operating model was different. [2] Not yet implemented. [3] Estimation assumes no error but replicates are simulated with error.

© 2001 Blackwell Science Ltd, FISH and FISHERIES, **2**, 125–157

decisions. We have shown that there is considerable subjective and case-specific tailoring of population dynamics models in terms, for example, model structure, catchability constraints and variance constraints. Furthermore, a large number of choices are made about which parameters or observations are to be admitted as uncertain, and how to deal with this uncertainty. There is not any obvious consensus in the scientific community as to how these choices should be made.

This is perhaps because relatively little testing of the applications and their assumptions has been reported in the literature. One notable exception is Anonymous (1999d), in which the accuracy of probability coverage of bias-corrected ADAPT nonparametric bootstraps is examined. There is therefore little basis either for assessing the performance of the methods, or for advocating any particular standardisation.

We would wish to consider, in the face of this untested methodological diversity, whether all, some, or any of the methods provide managers with probability statements that reflect the probabilities of the eventual real outcomes of the management action they evaluate. Although we acknowledge that constraints of time and manpower mean that assessments of risk are necessarily less thorough than would be desired, we contend that ultimately uncertainty estimation techniques should meet four criteria. First, the method should be properly conditioned, i.e. it should only take as axiomatic assumptions which either have a very low probability of being wrong or for which any mis-specification has very little consequence. Second, the distributional assumptions made about parameters or data admitted as uncertain should be validated. Third, if there remain significant uncertainties due to alternative model structures or alternative conditioning choices being perceived as having high probability, then such uncertainties should be transmitted clearly. Lastly, we stress that the conditioning decisions should dictate the estimation method, rather than institutional convention or use of standard software.

We have stated a preference for 'properly conditioned' models. In practice it is often the case that a fairly wide range of structural models, often with very different consequences for management, may appear to represent the data almost equally well, and an ideal of proper conditioning is not often attainable. This may also be the case for some particular parameters—that is, a wide range of values for a parameter, such as M, represent the data almost equally well but have very different consequences. In such cases of indeterminacy, it is most important that uncertainty due to model structure should be presented in an interpretable form. One approach is to construct probability distributions using a number of structural models, weighted according to some 'prior' belief or posterior perception about the validity of each (e.g. Schnute and Hilborn 1993; Adkinson and Peterman 1996; Patterson 1999). These 'model-averaging' approaches allow a number of alternative model choices to be admitted as plausible hypotheses, yet allow provision of advice in the traditional form of expected values and associated probability distributions (Draper 1995). An alternative is the consideration of many (possibly radically different) models by evaluation of management procedures (e.g. Anonymous 1994c; Punt and Smith 1999). For example, in their evaluation of the impact of seal culls on harvest levels of Cape hake, Punt and Butterworth (1995) found that the choice concerning which species to include in an overall ecosystem model had a much greater impact on the final conclusions than the uncertainty about a single ecosystem model. When the data are not informative about a parameter or structural model, care should be taken to ensure that optimisation is not influenced by peculiarities in the data which would result in spurious estimates. The topic of model averaging is familiar to those working with Bayesian methods and is beginning to receive attention among those favouring frequentist methods (Buckland *et al*. 1997). For a full appreciation of uncertainties, however, it may also be useful to consider the consequences of adopting an extreme case when another case may be true.

Unfortunately, examining alternative models for strategic decisions is uncommon. However, such examinations do seem to be on the increase as witnessed by the fact that three of the applications in Table 1 (those to Norwegian spring-spawning herring, eastern gemfish and Icelandic cod) made some attempt to do this. There are probably several reasons for not considering model uncertainty, ranging from the computational aspects of developing and fitting different models, and lack of qualified personnel, to (unfortunately) institutional barriers. An arguably even greater problem is the common practice of ignoring the possibility of several competing models once a single (commonly used) model appears to fit the available data adequately. Much effort in stock assessment is expended in attempting to identify the 'best' assessment model and then to condition advice on that choice. In many cases, robust identification

Table 2 Example decision table (adapted from Hilborn *et al*. 1993): expected average yields for three possible states of nature and four possible actions.

Policy	Expected average yield over 5 years (in thousands of tonnes) at given virgin biomasses (and probabilities)			
	0.9 Mt (0.57)	1.5 Mt (0.4)	2.1 Mt (0.03)	Expected value
150 Kt constant quota	136	150	150	148
150 Kt quota in 1993; 39% harvest rate thereafter	137	186	202	168
200 Kt quota in 1993; 39% harvest rate thereafter	138	194	216	173
200 Kt quota in 1993; 26% harvest rate thereafter	122	154	166	141

Kt, kilotonnes; Mt, million tonnes. This shows the expected average yield over a 5-year time horizon (in thousands of tonnes) for three possible states of nature and four possible actions. The possible states of nature (corresponding to plausible alternative conditioning choices) are virgin biomasses of 0.9, 1.5 and 2.1 Mt with probabilities of 0.57, 0.40 and 0.03, respectively. The possible actions include constant quotas or specific quotas in 1993 followed by harvest rates.

of a definitive model proves difficult. It would be more productive to explore the consequences of plausible alternative model choices and to present estimates of uncertainty for management purposes that reflect this uncertainty in appropriate conditioning.

The presentation of advice to managers that takes into account uncertainty in appropriate conditioning choices is not straightforward. Possibilities include the model-averaging approaches mentioned above, but Hilborn and Walters (1992) and Hilborn *et al*. (1993) advocate making such uncertainty explicit in the form of a decision table (Table 2), which is an intuitively attractive solution. However, the task of attaching probabilities to alternative conditioning choices is not a trivial one, and indeed may be computable only in certain specific cases. Expert opinions, plausibly formulated as Bayes priors, may be of help in such instances. Nevertheless, the approach has not been adopted widely, even though it was proposed for fisheries purposes some 8 years ago. Examples include applications to orange roughy off Namibia (McAllister, personal communication), coastal sharks off USA (McAllister and Pikitch, in press) and hoki off New Zealand (Punt *et al*. 1994). This may be in part because scientific advice presented in this form is considerably harder for management agencies to interpret. Indeed, in cases where there are many users of such advice with different sectoral objectives, the possibility of conflicting interpretations of such a decision table may make its application unfeasible. Model-averaging methods may be preferable in such situations, but some would argue that the difficulties of effectively conveying the uncertainty about conditioning choices in a manner that supports practical decision making is a significant incentive for the continued prevalence of seeking the 'best' assessment. For example, attempts to reflect uncertainty in structural conditioning (Anonymous 1999e) can lead to lack of credibility in the scientific analyses (Anonymous 2000b).

The ability to conduct the evaluations needed to quantify the uncertainty associated with assessments and forecasts has increased rapidly in recent years. The challenge of computational power is close to being overcome. However, with the technical aspects less of a concern, increased attention must be given to being more careful in testing and selecting the assumptions underlying assessments and to examining whether these sophisticated methods of quantifying uncertainty are able to deliver what they promise.

Making confidence statements that adequately represent the probabilities of eventual real outcomes is the ultimate goal. One should ideally wish to know how well the methods now in use do serve to represent these eventual, real probabilities. In principle, one may test the performance of an uncertainty estimation using simulations in which no mis-specification of conditioning is incorporated. This allows evaluation of the performance of the estimators used and the quality of the associated estimates of bias, probability coverage and accuracy of confidence statements. Simulation experiments can also be extended to investigate the sensitivity of confidence statements to alternative mis-specifications of conditioning choices, which may provide information about the robustness of such statements and the robustness of alternative harvesting rules.

However, a useful test of such confidence statements would be a comparison of the confidence statements with the eventual distribution of real outcomes. Within specific stocks, evaluation of the appropriateness of conditioning choices is likely to be a lengthy process of trial and error, as forecast distributions are compared with eventual outcomes on an annual basis. It may be possible, however, to draw general conclusions from a comparison of confidence statements and eventual outcomes across many stocks.

We conclude that despite much recent progress, fisheries science has yet to identify a means for identifying appropriate conditioning choices such that the probability distributions which are calculated for management purposes do adequately represent the probabilities of eventual real outcomes. We consider that where estimates are sensitive to conditioning choices, the task of communicating uncertainty to fisheries managers is not accomplished adequately by providing a simple estimate of variance or a risk threshold based on such an estimate. In order to complete such a task a thorough exploration of the various conditioning choices and their consequences will be required. Although provision of fisheries advice with estimates of uncertainty is a clear goal, there are obvious pitfalls in providing incomplete estimates of uncertainty. Provision of incomplete uncertainty estimates is likely to result in the eventual occurrence of events outside the range of those that had been foreseen in the scientific advice, which will cause a loss of credibility in the uncertainty estimates and in the associated point estimates. In cases where important model uncertainties exist, management procedure simulations can be used to help identify decision rules that are robust to such uncertainty. Decision tables and model averaging may prove to be useful additional tools for the presentation of a more complete view of the uncertainties relevant for management purposes, but they have not yet achieved general acceptance. We recommend greater efforts to use and promote such approaches.

Acknowledgements

The authors are pleased to acknowledge financial support from the European Community's Fifth Framework Programme (FAIR-PL98-4231) and the support of the directorates of CEFAS, CSIRO, DFO, FRS, ICCAT, IFREMER, IMR and MRI. Dr Kevern Cochrane and an anonymous refereee assisted greatly by making extensive suggestions for improvements. Dr R. Fryer and Dr C.M. O'Brien also made helpful comments. The views presented here are not necessarily those of the European Commission, nor do they necessarily reflect any future policy of the Commission.

Appendix

This section provides background on the case-studies summarised in Table 1. We discuss some recent examples of uncertainty-estimation procedures that have been used in the provision of management advice for: cod (*Gadus morhua*, Gadidae) in the North Sea and around Iceland, hake (*Merluccius merluccius*, Merluccidae) in the North-East Atlantic, herring (*Clupea harengus*, Clupeidae) in the North Sea—to the West of Scotland and in the Norwegian Sea (Norwegian spring-spawning stock), plaice (*Pleuronectes platessa*, Pleuronectidae) in the North Sea, haddock (*Melanogrammus aeglefinus*, Gadidae) on the Eastern Georges Bank and gemfish (*Rexea solandri*, Gempylidae) off Eastern Australia.

North Sea cod

Up until the early 1990s, ICES advice concentrated almost entirely on short-term forecasts of stock size with 'target' fishing mortality rates indicated in relation to yield-per-recruit criteria. While stocks fluctuated with no obvious sign of serious decline, managers were inclined to choose TACs which allowed fishing mortality to continue at status quo values which often exceeded the yield-per-recruit criteria. With the decline in many stocks, tools were needed to assist in the choice of fishing mortality rates which were likely to result in stock increases above biomass thresholds within a chosen time frame. The tool described here was first used for North Sea demersal stocks in 1993 (Anonymous 1993).

WGMTERM was a simple tool designed as an add-on to the basic catch-at-age analysis in the stock assessment. Given that the majority of assessments used XSA (Darby and Flatman 1994), the program uses population estimates from that output and projects these forward for a number of years at a fixed mortality rate. Recruitment is generated from a stock recruitment function fitted to the XSA estimates of stock and recruitment. Uncertainty is included in the recruitment model and in the initial population sizes. Recruitment is modelled by one of three standard functions, Beverton–Holt, Shepherd and Ricker, and may, if desired, include a first-order autocorrelation in the residuals.

The annual reports from the ICES Advisory Committee on Fisheries Management (ACFM) contain a table of catch forecast options for a set of fishing mortality rates requested by managers. For each fishing mortality option, ACFM gives a medium-term impact statement of the type 'low probability of spawning stock biomass falling outside precautionary limit'. Many of these impact statements are taken from the analysis described above. These statements influence the manager's choice of catch limit. Thus, though the analysis is of a strategic type, it has an impact on decisions about immediate measures.

Northern hake

Another early development was a technique which augmented parametric bootstrapping of observed data with Monte Carlo simulation of 'inestimable' parameters. It was intended to guide managers away from simple point estimates into probabilistic estimates that would reflect the analyst's perception of variability in all the inputs used for the analyses. The approach was initially aimed at estimating the uncertainty of current stock status and decisions about tactics such as quotas but was extended to include evaluation of strategies based on medium-term projections. The method first proposed by Powers and Restrepo (1992) and Restrepo *et al.* (1992) was adapted to suit ICES practice with several extensions (Mesnil 1993a, 1995). Results for northern hake (Mesnil 1993b) were considered a suitable basis for advice by the ACFM.

Norwegian spring-spawning herring

The assessment of Norwegian spring-spawning herring has been perceived as highly problematic in that (i) there exist relatively few observations and those are of a rather imprecise nature, and (ii) there is also substantial perceived uncertainty as to the appropriate form of assessment model that should be used. This is largely because the stock has been at an extremely low level for some decades (during which time few observations could be gathered) and it subsequently recovered to an extremely high level after the recruitment of a few highly abundant year classes. A Bayesian form of assessment was introduced by Anonymous (1997b) in order to admit perceived uncertainty in natural mortality and model uncertainty about the most appropriate assumption concerning error distributions and the appropriate structural form of stock-recruitment model (Patterson 1999). This approach was used to explore strategic issues associated with policies of constant exploitation rate or harvest-control rules. Catch forecasts for decisions about immediate management measures can also be obtained.

West Scotland herring

A 'separable' VPA structural model was used in an attempt to quantify the uncertainty introduced by uncertainty in catch reporting levels. A multiplier on the reported catches is added as an additional parameter (with a uniform prior distribution the range of which encompasses credulity). Catch-at-age observations are assumed distributed according to the product of the true catch, the uniform misreporting error multiplier and a lognormal sampling error. Results of the West Scotland herring assessment have been presented in Anonymous (1998) for management purposes, as percentiles of the historic stock parameters, short-term catch options and medium-term projections. These results were used to decide that the assessment calculation was too uncertain to form the basis of advice and to defer the management decision until more recent survey information had become available.

North Sea herring

Two different techniques have been applied for this stock. One approach is grounded on assuming a 'separable' VPA assessment model, making an assumption of lognormal errors in the observations and assuming that the fitted population abundance parameters conform to a multivariate lognormal distribution. This method has been used in a variety of forms for North Sea herring uncertainty estimation in tactical TAC decisions and (with the inclusion of a Beverton–Holt stock–recruitment relationship) in strategic medium- and long-term management advice (Anonymous 1998). The other approach is a projection simulation designed specifically to investigate strategic issues regarding selection of an optimal harvest control rule (Skagen 1999). The contemporary state of the stock is taken from the latest assessment working group results, also assuming that the population abundance estimates conform to a multivariate lognormal distribution.

North Sea plaice

Management procedures are investigated using a simulation model. An operating model represents reality, a perception of which is gained using an assessment based on data sampled from the operating model. The model structure includes sampling, process and implementation errors, as well as the effect of feedback from management controls to the operating model. The evaluation of policies is performed in the context of the entire management procedure, that is, in the context of the particular combination of assessment technique, control rule and implementation. The results should not be viewed as providing predictions but should be regarded as a means for comparing performance of alternative management procedures. Thus, this approach is suitable for investigations of the strategic type. The approach has been applied illustratively to North Sea plaice (Kell *et al*. 1999).

Bayes random walk

Dissatisfaction with the performance of traditional index catchability models has prompted the development of approaches where catchability is assumed to be autocorrelated over time. The implementation considered here (Lewy 1999) applies a simple random walk model for each index by age group. The expected value of catchability is assumed constant while realised catchabilities are permitted to vary over time. This approach is developed within a Bayesian estimation framework.

Icelandic cod

Changes in inter- and intra-specific interactions in response to variations in stock status may have profound implications for analyses of harvest strategies and policies. In most instances there is not sufficient information to model these interactions and the processes are assumed stationary. One exception is the development from a single-species model for Icelandic cod to an eight-species model with economic concerns included (Stefánsson and Baldursson 1998). Results from these models have been used as the basis for decisions of the strategic type regarding the adoption of harvest control rules (Jakobsson and Stefánsson 1999).

Eastern Georges Bank haddock

There has been a recent evolution in Atlantic Canada away from basing annual TAC decisions on point estimates towards basing them on confidence distributions. The role of the scientific advice is to provide quantitative information on the risk that established reference points for parameters of interest, such as exploitation rate, might be exceeded for alternative TAC options. Strategic or policy decisions regarding the establishment of reference points are considered linked but more involved and independent exercises which are undertaken only periodically, while tactical decisions need to be made regularly. Reference points might be defined conceptually, requiring them to be estimated within the assessment process, but this has not occurred to date and reference points are prescribed externally. The development of tools to quantify uncertainty for tactical decisions was characterised by analytical methods initially (Gavaris and Van Eeckhaute 1997), while bootstrap techniques are dominant at present (Gavaris and Van Eeckhaute 1998). For the short-term projections required to make tactical decisions, it was considered that the most important source of uncertainty was uncertainty about the contemporary population abundance. Study to date has focused on capturing that uncertainty, with less attention given to other sources of uncertainty.

Eastern gemfish

Stock assessment for fisheries management in Australia aims towards extensive evaluation of management procedures. The aim of each Fishery Assessment Group is to conduct assessments and provide advice on the benefits of alternative management procedures. To accomplish this, an underlying operating model capable of generating future observations and representing alternative hypotheses about the dynamics of the resource is developed. The modelling is often conducted within a Bayesian framework to permit integration of 'prior' information based on either data for other species or expert opinion. Results provided to managers fall into three categories: (i) model parameters, e.g. virgin biomass, *M*, maximum sustainable yield (MSY), (ii) historical trajectories for recruitment and biomass and (iii) the probability of exceeding target or limit reference points for alternative management procedures. These management procedures range from full feedback-control approaches to sequences of fixed levels of TAC. These results are used to make decisions about strategies and to select immediate management measures such as TACs. The assessment of eastern gemfish (Smith and Punt 1998) is amongst the most advanced in Australia. It includes a wide range of alternative models and data types and has formed the basis for a formal evaluation of management procedures (Punt and Smith 1999).

References

Adkinson, M.D. and Peterman, R.M. (1996) Results of Bayesian methods depend on details of implementation:

an example of estimating salmon escapement goals. *Fisheries Research* **25**, 155–170.

Aglen, Å. (1989) Empirical results on precision–effort relationships for acoustic surveys. *ICES CM 1989/B* **30**, 28 pp.

Anonymous (1993) Report of the working group on the assessment of demersal stocks in the North Sea and Skagerrak. *ICES CM 1993/Assess* **6**, 233 pp.

Anonymous (1994a) The precautionary approach to fisheries with reference to straddling fish stocks and highly migratory fish stocks. In: *United Nations Conference on Straddling Fish Stocks and Highly Migratory Fish Stocks* (Proceedings of the UN conference New York, 14 March–31 March, 1994). United Nations, New York. **No. A/CONF 164 INF/8**, 25 pp.

Anonymous. (1994b) Report of the workshop on sampling strategies for age and maturity. *ICES CM 1994/D* **1**, 57 pp.

Anonymous (1994c) Report of the scientific committee, Annex D. Report of the sub-committee on management procedures. *Reports of the International Whaling Commission* **44**, 74–92.

Anonymous (1995) Guidelines on the precautionary approach to capture fisheries and species introduction. *FAO Fisheries Technical Paper* **No. 350.1**, 52 pp.

Anonymous (1997a) 24th Northeast regional stock assessment workshop (24th SAW): Public Review Workshop. *Northeast Fisheries Science Center Reference Document* **No. 97–11**, 67 pp.

Anonymous (1997b) Report of the northern pelagic and blue whiting fisheries working group. *ICES CM 1997/Assess* **14**, 221 pp.

Anonymous (1998) Report of the herring assessment working group for the area south of 62° N. *ICES CM 1998/ACFM* **14**, 322 pp.

Anonymous (1999a) Report of the north-western working group. *ICES CM 1999/ACFM* **17**, 329 pp.

Anonymous (1999b) Report of the working group on the assessment of northern shelf demersal stocks. *ICES CM 2000/ACFM* **1**, 402 pp.

Anonymous (1999c) Report of the herring assessment working group for the area South of 62° N. *ICES CM 1999/ACFM* **12**, 279 pp.

Anonymous (1999d) Report of the comprehensive fishery evaluation working group, Key Largo, Florida. 14–21 January, 1999. *ICES CM 1999/ACFM* **12**, 255 pp.

Anonymous (1999e) Southern Scotian Shelf and Bay of Fundy cod (Div. 4X/5Y). *DFO Science Stock Status Report* **A3–05** (1999), 25 pp.

Anonymous (2000a) Report of the CWP intersessional meeting. Working group on precautionary approach terminology and CWP sub-group on publication of integrated catch statistics for the Atlantic. 10 February–16 February. *2000 ICES CM 2000/ACFM* **17**, 67 pp.

Anonymous (2000b) Conservation requirements for groundfish stocks on the Scotian Shelf and in the Bay of Fundy (4VWX), in Sub-Areas 0.2 + 3 and redfish stocks. FRCC.2000R.1. [available at http://www.ncr.dfo.ca/frcc/index.htm on 31st. January 2001.]

Basson, M. (1999) The importance of environmental factors in the design of management procedures. *ICES Journal of Marine Science* **56**, 933–942.

Beddington, J.R. and Cooke, J.G. (1981) Development of an assessment technique for male sperm whales based on the use of length data from the catches, with special reference to the North-west Pacific stock. *Reports of the International Whaling Commission* **31**, 747–760.

Bell, E.D. and Stefánsson, G. (1998) Performance of some harvest control rules. *Working paper to the ICES working group on the application of the precautionary approach*. **MS**, 12 pp. [see reference appendix for availability (1).]

Bogstad, B. (1993) Implementation of multispecies research results in fisheries management. *Nordiske Seminar Og Arbejdsrapporter* **572**, 212–214.

Bogstad, B., Hauge, K.H., Ulltang, Ø. and Tjelmeland, S. (1995) Results of simulation studies using a multispecies model for the Barents Sea (MULTSPEC) *IWC Document* **No. SC/47/NA1**, 33 pp.

Box, M.J. (1971) Bias in nonlinear estimation. *Journal of the Royal Statistical Society of, Series B* **33**, 171–201.

Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1997) Model selection: an integral part of inference. *Biometrics* **53**, 603–618.

Bundy, A., Lilly, G.R. and Shelton, P.A. (2000) A mass balance model of the Newfoundland-Labrador shelf. *Canadian Technical Reports in Fisheries and Aquatic Science* **2310**, 171 pp.

Butterworth, D.S. and Andrew, P.A. (1984) Dynamic catch-effort models for the hake stocks in ICSEAF Divisions 1.3–2.2. *Collection of Scienific Papers of the International Commission for South-East Atlantic Fisheries* **11**, 29–58.

Butterworth, D.S., Cochrane, K.L. and De Oliveira, J.A.A. (1997) Management procedures: A better way to manage fisheries? The South African experience. In: *Global Trends: Fisheries Management* (eds E.L. Pikitch, D.D. Huppert and M.P. Sissenwine). American Fisheries Society, Bethesda, pp. 83–90.

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

Christensen, V. (1995) A model of trophic interactions in the North Sea in 1981, the year of the stomach. *Dana* **11**, 1–28.

Christensen, V. and Pauly, D. (1996) Ecological modelling for all. *Naga* **19**, 25–26.

Christensen, V. and Pauly, D. (eds) (1993) *Trophic Models in Aquatic Systems*. ICLARM, Manila, Philippines, 284 pp.

Clark, W.G., Hare, S.R., Parma, A.M., Sullivan, P.J. and Trumble, R.J. (1999) Decadal changes in growth and recruitment of Pacific halibut (*Hippoglossus stenolepis*). *Canadian Journal of Fisheries and Aquatic Science* **56**, 242–252.

Clemen, R.T. (1996) *Making Hard Decisions*, 2nd edn. PWS Kent, Boston.

Cleveland, W.S. (1981) LOWESS: a program for smoothing scatterplots by robust locally weighted regression. *American Statistician* **35**, 1–54.

Collie, J.S. and Spencer, P.D. (1993) Management strategies for fish populations subject to long-term environmental variability and depensatory predation. In: *Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations* (Alaska Sea Grant College Program Report No AK-SG-93-02) (eds G. Kruse, D.M. Eggers, R.J. Marasco, C. Pautzke and T.J. Quinn II). University of Alaska, Fairbanks, pp. 355–370.

Collie, J.S. and Tsou, T.-S. (1996) Multispecies virtual population analysis of the Georges Bank fish community. *ICES C.M. 1996/G* **28**, 16 pp.

Cook, R.M. (1993) The use of sensitivity analyses to quantify uncertainties in stock projections. *ICES C M 1993/D* **66**, 12 pp.

Cook, R.M. (1998) A sustainability criterion for the exploitation of North Sea cod. *ICES Journal of Marine Science* **55**, 1061–1070.

Copas, J.B. (1983) Regression, prediction and shrinkage (with discussion). *Journal of the Royal Statistical Society of B* **45**, 311–354.

Crone, P.R. and Sampson, D.B. (1998) Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition. In: *Fishery Stock Assessment Models* (Alaska Sea Grant College Program Report No AK-SG-98-01) (eds F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan and C.-I.Chang). University of Alaska, Fairbanks, pp. 355–370.

Csirke, J. (1988) Small shoaling pelagic fish stocks. In: *Fish Population Dynamics* (ed. J.A. Gulland). John Wiley and Sons, New York.

Daan, N. (1987) Multispecies versus single-species assessment of North Sea fish stocks. In: *International Symposium on Stock Assessment and Yield Prediction, Atikokan, Ont, Canada* (eds W.J. Christie and G.R. Spangler). *ICES Journal of Marine Science* **44**, 360–370.

Daníelsson, Á., Stefánsson, G., Baldursson, F. and Thorarinsson, K. (1997) Utilisation of the Icelandic cod stock in a multispecies context. *Marine Resource Economics* **12**, 329–333.

Darby, C.D. and Flatman, S. (1994) Virtual Population Analysis, Version 3.1 (Windows/Dos) user guide. *Information Technology Series, MAFF Directorate of Fisheries Research, Lowestoft* **1**, 85 pp.

Deriso, R.B., Quinn, T., Collie, J., Hilborn, R., Jones, C., Lindsay, B., Parma, A., Saila, S., Shapiro, L., Smith, S.J. and Walters, C.J. (1998) *Improving Fish Stock Assessments*. National Academy Press, Washington D.C.

Deriso, R.B., Quinn II, T.J. and Neil, P.R. (1985) Catch-age analysis with auxiliary information. *Canadian Journal of Fisheries and Aquatic Science* **42**, 815–824.

Doubleday, W.G. (1981) A method of estimating the abundance of survivors of an exploited fish population using commercial fishing catch at age and research vessel abundance indices. In: *Bottom Trawl Surveys* (eds W.G. Doubleday and D. Rivard). *Canadian Special Publications of Fisheries and Aquatic Sciences* **58**, 164–178.

Draper, D. (1995) Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society of, Series B (Methodology)* **57**, 45–97.

Edwards, A.W.F. (1972) *Likelihood*. Cambridge University Press, Cambridge.

Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**, 1–26.

Efron, B. (1982) *The Jacknife, the Bootstrap and Other Resampling Plans* (CBMS Monograph no. 38). Society for Industrial and Applied Mathematics, Philadelphia.

Efron, B. (1987) Better bootstrap confidence intervals, with discussion. *Journal of the American Statististical Association* **82**, 171–200.

Efron, B. (1998) R.A. Fisher in the 21st. century. *Statistical Science* **13**, 95–122.

Efron, B. and G.Gong (1983) A leisurely look at the bootstrap, the jackknife and cross-validation. *American Statistician* **37**, 36–48.

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

Evans, G.T. and Rice, J.C. (1988) Predicting recruitment from stock size without the mediation of a functional relation. *Journal of Du Conseil International Pour l'Exploration de la Mer* **44**, 111–122.

Fournier, D. and Archibald, C.P. (1982) A general theory for analyzing catch at age data. *Canadian Jounal of Fisheries and Aquatic Science* **39**, 1195–1207.

Freeman, S.N. and Kirkwood, G.P. (1995) On a structural time series method for estimating stock biomass and recruitment from catch and effort data. *Fisheries Research* **22**, 77–98.

Gavaris, S. (1980) Use of a multiplicative model to estimate catch rate and effort from commercial data. *Canadian Journal of Fisheries and Aquatic Sciences* **37**, 2272–2275.

Gavaris, S. (1988) An adaptive framework for the estimation of population size. *Canadian Atlantic Fisheries Scientific Advisory Committee Research Document* **No. 88/29**, 12 pp.

Gavaris, S. (1999) Dealing with bias in estimating uncertainty and risk. In: *Providing Scientific Advice to Implement the Precautionary Approach Under the Magnuson-Stevens Fishery Conservation and Management Act* (ed. V. Restrepo). *NOAA Technical Memorandum* **NMFS-F/SPO-40**, 17–23.

Gavaris, S. and Sinclair, A. (1998) From fisheries assessment uncertainty to risk analysis for immediate management actions. In: *Fishery Stock Assessment Models* (Alaska Sea Grant College Program Report No AK-SG-98-01) (eds F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan and C.-I. Chang). University of Alaska, Fairbanks, pp. 903–916.

Gavaris, S. and Van Eeckhaute, L. (1997) Assessment of haddock on eastern Georges Bank. *Department of Fisheries and Oceans, Canadian Stock Assessment Secretariat Research Document* **97/54**, 72 pp.

Gavaris, S. and Van Eeckhaute, L. (1998) Assessment of haddock on eastern Georges Bank. *Department of Fisheries and Oceans, Canadian Stock Assessment Secretariat Research Document* **98/66**, 75 pp.

Geiger, H.J. and Koenings, J.P. (1991) Escapement goals for sockeye salmon with informative prior probabilities based on habitat considerations. *Fisheries Research* **11**, 239–256.

Gilbert, D.J. (1997) Towards a new recruitment paradigm for fish stocks. *Canadian Journal of Fisheries and Aquatic Science* **54**, 969–977.

Gislason, H. and Sparre, P. (1994) Some thoughts on the incorporation of areas and migrations in MSVPA. *ICES CM 1994/Mini* **15**, 20 pp.

Gudmundsson, G. (1994) Time series analysis of catch-at-age observations. *Applied Statististics* **43**, 117–126.

Gudmundsson, G. (1998) Joint time series analysis of catch-at-age and CPUE data. In: *Fishery Stock Assessment Models* (Alaska Sea Grant College Program Report No AK-SG-98-01) (eds F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan and C.-I. Chang). University of Alaska, Fairbanks, pp. 199–217.

Gunderson, D.R. (1993) Surveys of fisheries resources. John Wiley and Sons, New York, 248 pp.

Haist, V., Fournier, D.A. and Schweigert, J.F. (1993) Estimation of density-dependent natural mortality in British Columbia herring stocks through SSPA and its impact on sustainable harvesting strategies. In: *Risk Evaluation and Biological Reference Points for Fisheries Management* (eds S.J. Smith, J.J. Hunt and D. Rivard). *Canadian Special Publications in Fisheries and Aquatic Science* **120**, pp. 269–282.

Hanchet, S., Haist, V. and Fournier, D. (1998) An integrated assessment of Southern Blue Whiting (*Micromesistius australis*) from New Zealand using separable sequential population analysis. In: *Fishery Stock Assessment Models* (Alaska Sea Grant College Program Report No AK-SG-98-01) (eds F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan and C.-I. Chang). University of Alaska, Fairbanks, pp. 155–170.

Helser, T.E. and Brodziak, J.K.T. (1998) Impacts of density-dependent growth and maturation on assessment advice to rebuild depleted U.S. silver hake (*Merluccius bilinearis*) stocks. *Canadian Journal of Fisheries and Aquatic Science* **55**, 882–892.

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 Science* **50**, 874–880.

Hilborn, R. and Walters, C.J. (1992) *Quantitative Fisheries Stock Assessment*. Chapman & Hall, New York and London, 570 pp.

Hilden, M. (1988) Significance of the functional response of predators to changes in prey abundance in multispecies virtual population analysis. *Canadian Journal of Fisheries and Aquatic Science* **45**, 89–96.

Hjort, J. (1909) Report regarding the herring. *Rapport et Procès-verbaux des Réunions du Conseil International pour l'Exploration De la Mer, XIIC* **5**, 17 pp.

Hoel, P.G. (1971) *Introduction to Mathematical Statistics*. John Wiley and Sons, New York.

Ianelli, J.N. and Fournier, D.A. (1998) Alternative age-structured analyses of the NRC simulated stock assessment data. In: *Analyses of Simulated Data Sets in Support of the NRC Study on Stock Assessment Methods NOAA Technical Memorandum NMFS-F/SpO-30* (ed. V.R. Restrepo). US Department of Commerce, Washington, pp. 81–96.

Jakobsson, J. and Stefánsson, G. (1999) The management of the summer spawning herring at Iceland, a case study. *ICES Journal of Marine Science* **56**, 827–833.

Kell, L.T., O'Brien, C.M., Smith, M.T., Stokes, T.K. and Rackham, B.D. (1999) An evaluation of management procedures for implementing a precautionary approach in the ICES context for North Sea plaice (*Pleuronectes platessa* L.). In: *A Review of Some Methods for Estimating Uncertainty in Fisheries* (eds K.R. Patterson, R.M. Cook, C.D. Darby, S. Gavaris, L.T. Kell, P. Lewy, B. Mesnil, A.E. Punt, V.R. Restrepo, D.W. Skagen and G. Stefánsson). *Fisheries Research Services Report* **No. 7/99**. [see reference appendix for availability (2).]

Kimura, D.K. (1981) Standardized measures of relative abundance based on modelling log (c.p.u.e.) and their application to Pacific ocean perch (*Sebastes alutus*). *Journal of Du Conseil International Pour Exploration de la Mer* **39**, 211–218.

Kimura, D.K. (1990) Approaches to age-structured separable sequential population analysis. *Canadian Journal of Fisheries and Aquatic Science* **47**, 2364–2374.

Kirkwood, G.P. (1981) Estimation of stock size using relative abundance data—a simulation study. *Reports of the International Whaling Commission* **31**, 729–745.

Kirkwood, G.P. (1997) The Revised Management Procedure of the International Whaling Commission. In: *Global Trends: Fisheries Management* (eds E.L. Pikitch, D.D. Huppert and M.P. Sissenwine). American Fisheries Society, Bethesda, pp. 91–99.

Kolody, D. and Patterson, K. (1999) Evaluation of NE Atlantic mackerel stock assessment models on the basis of simulated long-term management performance. *ICES CM 1999/S* **1**, 9 pp.

Kruse, G., Eggers, D.M., Marasco, R.J., Pautzke, C. and Quinn II, T.J. (eds) (1993) *Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations* (Alaska Sea Grant College Program Report No AK-SG-93-02). University of Alaska, Fairbanks, 825 pp.

Larkin, P.A. (1996) Concepts and issues in marine ecosystem management. *Reviews in Fish Biology and Fisheries* **6**, 139–164.

Larsen, J.R. and Gislason, H. (1992) MSVPA and prey/predator switching. *ICES CM 1992/G* **42**, 11 pp.

Leaman, B.M. (1981) A brief review of survey methodology with regard to groundfish stock assessment. In: *Bottom Trawl Surveys* (eds W.G. Doubleday and D. Rivard). *Canadian Special Publications of Fisheries and Aquatic Sciences* **58**, 113–123.

Lewy, P. (1999) Bayesian approach to fish stock assessment using MCMC and graphical models. In: *A Review of Some Methods for Estimating Uncertainty in Fisheries* (eds K.R. Patterson, R.M. Cook, C.D. Darby, S. Gavaris, L.T. Kell, P. Lewy, B. Mesnil, A.E. Punt, V.R. Restrepo, D.W. Skagen and G. Stefánsson). *Fisheries Research Services Report* **No. 7/99**. [see Appendix for availability of unpublished reports (2).]

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

Lilly, G.R. (1998) Size-at-age and condition of cod in Divisions 2J + 3KL during 1978–97. *Department of Fisheries and Oceans, Canadian Stock Assessment Secretariat Research Document* **98/76**, 66 pp.

Lowe, S.A. and Thompson, G.G. (1993) Accounting for uncertainty in the development of exploitation strategies for the Atka mackerel resource of the Aleutian Islands. In: *Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations* (Alaska Sea Grant College Program Report No AK-SG-93-02) (eds G. Kruse, D.M. Eggers, R.J. Marasco, C. Pautzke and T.J. Quinn II). University of Alaska, Fairbanks, pp. 203–231.

Magnusson, K.G. (1993) A model of predation. *Nordiske Seminar Og Arbeidsrapporter* **572**, 176–196.

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

McAllister, M.K. and Ianelli, J.N. (1997) Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Science* **54**, 284–300.

McAllister, M.K. and Kirkwood, G.P. (1998) Bayesian stock assessment: a review and example application using the logistic model. *ICES Journal of Marine Science* **55**, 1031–1060.

McAllister, M.K. and Pikitch, E.K. (in press) Why are Bayesian methods useful for stock assessment of sharks? In: *Sharks of the Open Ocean* (eds E.K. Pikitch and M. Camhi). Blackwell, New York.

Mesnil, B. (1993a) Using Monte Carlo simulations to account for uncertainties in stock assessments and biological advice for fisheries management. Application to the northern stock of European hake. *ICES CM 1993/D* **9**, 25 pp.

Mesnil, B. (1993b) Medium-term advice for the Northern Stock of European hake. *Working Document, ACFM November* **1993**, 12 pp. [see reference appendix for availability (3).]

Mesnil, B. (1995) Analyse de risque en évaluation de stocks et gestion de pêcheries. Utilisation des méthodes de Monte Carlo. In: *Les Recherches Françaises En Évaluation Quantitative et Modélisation Des Ressources et Des Systèmes Halieutiques* (Actes Du 1er Forum Halieumétrique, ENSA, Rennes 29 Juin–1er Juillet 1993) (eds D. Gascuel, J.L. Durand and A. Fonteneau). ORSTOM, Colloques et séminaires, Rennes. 105–125. [In French.]

Meyer, R. and Millar, R.B. (1999) Bayesian stock assessment using a state-space implementation of the delay-difference model. *Canadian Journal of Fisheries and Aquatic Science* **56**, 37–52.

Millar, R.B. and Meyer, R. (2000) Bayesian state-space modelling of age-structured data: fitting a model is just the beginning. *Canadian Journal of Fisheries and Aquatic Science* **57**, 43–50.

Mohn, R.K. (1993) Bootstrap estimates of ADAPT parameters, their projection in risk analysis and their retrospective patterns. In: *Risk Evaluation and Biological Reference Points for Fisheries Management* (eds S.J. Smith, J.J. Hunt and D. Rivard). *Canadian Special Publications in Fisheries and Aquatic Science* **120**, pp. 173–187.

Myers, R.A. (1998) When do environment-recruitment correlations work? *Reviews in Fish Biology and Fisheries* **8**, 285–305.

Myers, R.A., Barrowman, N.J., Hutchings, J.A. and Rosenberg, A.A. (1995) Population dynamics of exploited fish stocks at low population levels. *Science* **269**, 1106–1108.

Myers, R.A., Hutchings, J.A. and Barrowman, N.J. (1997) Why do fish stocks collapse? The example of the northern cod in Atlantic Canada. *Ecological Applications* **7**, 91–106.

Patterson, K.R. (1997) Density-dependence of growth and natural mortality in the Norwegian Spring-Spawning herring and implications for yield calculations. *ICES CM 1997/DD* **7**, 18 pp.

Patterson, K.R. (1998) Assessing fish stocks when catches are misreported: model, simulation tests, and application to cod, haddock and whiting in the ICES area. *ICES Journal of Marine Science* **55**, 878–891.

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 Jounal of Fisheries and Aquatic Science* **56**, 208–221.

Patterson, K.R. and de Cárdenas, E. (2000) Evaluation of Greenland halibut dynamics in 2J3KLMNO using separable models in Bayes and maximum-likelihood approaches. *NAFO SCR Document* **00/17**, 31 pp.

Patterson, K.R. and Melvin, G.D. (1996) Integrated catch at age analysis, Version 1.2. *Scottish Fisheries Research Report* 58, 60 pp.

Payne, A.I.L. (ed.) (1999) Confronting uncertainty in the evaluation and implementation of fisheries-management systems. *ICES Journal of Marine Science* **56**, pp. 795–1072.

Pella, J.J. (1993) Utility of structural time-series models and the Kalman filter for predicting consequences of fishery actions. In: *Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations* (Alaska Sea Grant College Program Report No AK-SG-93-02) (eds G. Kruse, D.M. Eggers, R.J. Marasco, C. Pautzke and T.J. Quinn II). University of Alaska, Fairbanks, pp. 571–593.

Polacheck, T., Hilborn, R. and Punt, A.E. (1993) Fitting Surplus Production Models: Comparing Methods and Measuring Uncertainty. *Canadian Journal of Fisheries and Aquatic Science* **50**, 2597–2607.

Polovina, J.J. and Ow, M.D. (1985) An approach to estimating an ecosystem box model. *Fishery Bulletin* **83**, 457–460.

Poole, D., Givens, G.H. and Raftery, A.E. (1999) A proposed stock assessment methods and its application to bowhead whales, *Balaena mysticetus*. *Fishery Bulletin, US* **97**, 144–152.

Pope, J.G. (1991) The ICES multispecies assessment working group: evolution, insights and future problems. In: *Multispecies Models Relevant to Management of Living Resources* (eds N. Daan and M.P. Sissenwine). *ICES Marine Science Symposia* **193**, 22–33.

Pope, J.G. (1993) A Reappraisal of the 1992 Assessment of the Icelandic Cod Stock. Directorate of Fisheries Research, Fisheries Laboratory, Lowestoft. MS, 78 pp.

Pope, J.G. and Macer, C.T. (1991) Multispecies virtual population analysis of North Sea predators for the years 1963–87. In: *Multispecies Models Relevant to Management of Living Resources* (eds N. Daan and M.P. Sissenwine). *ICES Marine Science Symposia* **193**, 46–49.

Pope, J.G. and Shepherd, J.G. (1985) A comparison of the performance of various methods for tuning VPAs using effort data. *Journal of Du Conseil International Pour L′ Exploration de la Mer* **42**, 129–151.

Porch, C. (1998) A Bayesian VPA with randomly walking parameters. *International Commission for the Conservation of Atlantic Tunas SCRS* **98/64**, 21 pp.

Powers, J.E. and Restrepo, V.R. (1992) Evaluation of Gulf of Mexico king mackerel stock assessment research: benefits and costs to management. *Miami Laboratory Contribution* **ML-90/91-71**, 38 pp. [see reference appendix for availability (4).]

Punt, A.E. and Butterworth, D.S. (1993) Variance estimates for fisheries assessment: their importance and how best to evaluate them. In: *Risk Evaluation and Biological Reference Points for Fisheries Management* (eds S.J. Smith, J.J. Hunt and D. Rivard). *Canadian Special Publications in Fisheries and Aquatic Science* **120**, 145–162.

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

Punt, A.E. and Butterworth. D.S. (2000) Why do Bayesian and maximum-likelihood assessments of the Bering-Chukchi-Beaufort Seas stock of bowhead whales differ? *Journal of Cetacean Research and Management* **2**, 125–133.

Punt, A.E. and Hilborn, R. (1996) Biomass dynamic models. User's manual. *FAO Computerized Information Series (Fisheries)* **10**, 1–62.

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

Punt, A.E., McAllister, M.K., Pikitch, E.K. and Hilborn, R. (1994) Stock assessment and decision analysis for hoki (*Macruronus novaezelandiae*) for 1994. *New Zealand Fisheries Assessment Research Document* **94/13**, 34 pp.

Punt, A.E. and Smith, A.D.M. (1999) Harvest strategy evaluation for the eastern stock of gemfish (*Rexea solandri*). *ICES Journal of Marine Science* **56**, 860–875.

Quinn II, T.J. and Deriso, R.B. (1999) *Quantitative Fish Dynamics*. Oxford University Press, New York and Oxford, 542 pp.

Raftery, A.E., Givens, G.H. and Zeh, J.E. (1995) Inference from a deterministic population dynamics model for bowhead whales (with discussion). *Journal of the American Statistical Association* **90**, 402–230.

Ratkowsky, D.A. (1983) *Nonlinear Regression Modelling*. Marcel Dekker, New York.

Restrepo, V.R. (1998) Possible use of nonparametric recruitment prediction in bluefin tuna projections. *International Commission for the Conservation of Atlantic Tunas SCRS* **98/86**.

Restrepo, V.R., Hoenig, J.M., Powers, J.E., Baird, J.W. and Turner, S.C. (1992) A simple simulation approach to risk and cost analysis, with applications to swordfish and cod fisheries. *Fishery Bulletin, US* **90**, 736–748.

Restrepo, V.R., Powers, J.E. and Turner, S.C. (1991) Incorporating uncertainty in VPA results via simulation. *Collected Volumes of Scientific Papers, International Commission for the Conservation of Atlantic Tunas* **35**, 355–361.

Richards, L.J.J.T. and Schnute and Olsen, N. (1998) A statistical framework for analysis of limit reference points. In: *Fishery Stock Assessment Models* (Alaska Sea Grant College Program Report No AK-SG-98-01) (eds F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan and C.-I. Chang). University of Alaska, Fairbanks, pp. 185–197.

Rindorf, A., Gislason, H. and Lewy, P. (1998) Does the diet of cod and whiting reflect the species composition estimated from trawl surveys? *ICES CM 1998/CC* **5**, 44 pp.

Rivoirard, J., Simmonds, J., Foote, K.G., Fernandes, P. and Bez, N. (2000) *Geostatistics for Estimating Fish Abundance*. Blackwell Science. Oxford.

Royall, R. (1997) Statistical evidence. *A Likelihood Paradigm*. London. Chapman & Hall. 320 pp.

Saila, S.B. and Ferson, S. (1998) Fuzzy regression in fisheries science: Some methods and applications. In: *Fishery Stock Assessment Models* (Alaska Sea Grant College Program Report No AK-SG-98-01) (eds F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan and C.-I. Chang). University of Alaska, Fairbanks, pp. 339–354.

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

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

Schnute, J.T. and Hilborn, R. (1993) Analysis of contradictory data sources in fish stock assessment. *Canadian Journal of Fisheries and Aquatic Science* **50**, 1916–1923.

Schweder, T. and Hjort, N.L. (1999) *Frequentist analogues of priors and posteriors*. (Statistical Research Report no. 8). University of Oslo, Oslo.

Shannon, L.J. and Jarre-Teichmann, A. (1999) A model of trophic flows in the northern Benguela upwelling system during the 1980s. *South African Journal of Marine Science* **21**, 349–366.

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

Shepherd, J.G. and Nicholson, M.D. (1986) Use and abuse of multiplicative models in the analysis of fish catch-at-age data. *Statistician* **35**, 221–227.

Shepherd, J.G. and Nicholson, M.D. (1991) Multiplicative modelling of catch-at-age data, and its application to catch forecasts. *Journal of Du Conseil International Pour l'Exploration de la Mer* **47**, 284–294.

Sinclair, A., Chouinard, G. and Currie, L. (1996) Assessment of the southern Gulf of St. Lawrence cod stock, March, 1996. *DFO Atlantic Fisheries Research Documents* **96/52**, 75 pp.

Sinclair, A. and Gavaris, S. (1996) Some examples of probabilistic catch projections using ADAPT output. *DFO Atlantic Fisheries Research Documents* **96/51**, 12 pp.

Sissenwine, M.P. and Daan, N. (1991) An overview of multispecies models relevant to management of living resources. In: *Multispecies Models Relevant to Management of Living Resources* (eds N. Daan and M.P. Sissenwine). *ICES Marine Science Symposia* **193**, 6–11.

Sissenwine, M.P. and Shepherd, J.G. (1987) An alternative perspective on recruitment overfishing and biological reference points. *Canadian Journal of Fisheries and Aquatic Science* **44**, 913–918.

Skagen, D.W. (1999) Programs for stochastic prediction and management simulation (STPR). In: *A Review of Some Methods for Estimating Uncertainty in Fisheries* (eds K.R. Patterson, R.M. Cook, C.D. Darby, S. Gavaris, L.T. Kell, P. Lewy, B. Mesnil, A.E. Punt, V.R. Restrepo, D.W. Skagen and Stefánsson G.). *Fisheries Research Services Report* **No. 7/99**. [see Appendix for availability of unpublished reports (2).]

Smith, S.J. (1990) Use of statistical models for the estimation of abundance from groundfish trawl surveys. *Canadian Journal of Fisheries and Aquatic Sciences* **47**, 894–903.

Smith, S.J. (1997) Bootstrap confidence limits for groundfish trawl survey estimates of mean abundance. *Canadian Journal of Fisheries and Aquatic Sciences* **54**, 616–630.

Smith, S.J. and Gavaris, S. (1993) Evaluating the accuracy of projected catch estimates from sequential population analysis and trawl survey abundance estimates. In: *Risk Evaluation and Biological Reference Points for Fisheries Management*) (eds S.J. Smith, J.J. Hunt and D. Rivard). *Canadian Special Publications in Fisheries and Aquatic Science* **120**, 163–172.

Smith, A.D.M. and Punt, A.E. (1998) Stock assessment of gemfish (*Rexea solandri*) in eastern Australia using maximum likelihood and Bayesian methods. In: *Fishery Stock Assessment Models* (Alaska Sea Grant College Program Report No AK-SG-98-01) (eds F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan and C.-I. Chang). University of Alaska, Fairbanks, pp. 245–286.

Sparholt, H. (1994) Fish species interactions in the Baltic. *Dana* **10**, 131–162.

Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. (1995) *BUGS 0.5: Bayesian inference Using Gibbs Sampling. Manual (Version ii)*. MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.

Stanley, R.D. (1992) Bootstrap calculation of catch-per-unit-effort variance from trawl logbooks: Do fisheries generate enough observations for stock assessments? *North American Journal of Fisheries Management* **12**, 19–27.

Stefánsson, G. (1996) Analysis of groundfish survey abundance data: combining the GLM and delta approaches. *ICES Journal of Marine Sciences* **53**, 577–588.

Stefánsson, G. (1998) Comparing different information sources in a multispecies context. Fishery Stock Assessment Models. In: *Fishery Stock Assessment Models* (Alaska Sea Grant College Program Report No AK-SG-98-01) (eds F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan and C.-I. Chang). University of Alaska, Fairbanks, pp. 120–135.

Stefánsson, G. and Baldursson, F.M. (1998) Marine mammals: a pest or a resource? *Working Paper to NAMMCO Document* **SC/6/EC/6**, 12 pp. [see Appendix for availability of unpublished reports (1).]

Stefánsson, G. and Bell, E.D. (1998) In search of stability. How to halt the decline of fish stocks. *Working paper to the ICES study group on the application of the precautionary approach*. **MS**, 14 pp. [see Appendix for availability of unpublished reports (1).]

Stefánsson, G., Hauksson, E., Bogasons, V., Sigurjónsson, J. and Víkingsson, G. (1997) Multispecies interactions in the Central Atlantic. *NAMMCO Document SC/5/ME* **13**, 14 pp. [see Appendix for availability of unpublished reports (1).]

Stefánsson, G., Sigurjónsson, J. and Víkingsson, G. (1998) On dynamic interactions between some fish resources and cetaceans off Iceland based on a simulation model. *Journal of Northwest Atlantic Fisheries Science* **22**, 357–370.

Stefánsson, G., Skúladóttir, U. and Steinarsson, B.Æ. (1999) Aspects of the ecology of a Boreal system. *ICES Journal of Marine Science* **55**, 859–862.

Thompson, W.F. and Bell, F.H. (1934) Biological statistics of the Pacific halibut fishery. 2. Effect of changes in fishing intensity upon total yield and yield per unit of gear. *Reports of the International Pacific Halibut Commission* **8**, 49 pp.

Tjelmeland, S. and Bogstad, B. (1998) MULTSPEC: a review of a multispecies modelling project for the Barents Sea. *Fisheries Research* **37**, 127–142.

Toresen, R. (1986) Length and age at maturity of Norwegian spring-spawning herring for the year-classes 1959–61 and 1973–78. *ICES CM 1986/H* **42**, 11 pp.

Toresen, R. (1988) Growth in the Norwegian spring-spawning herring for the year-classes 1950–60 and 1973–83. *ICES CM 1988/H* **35**, 8 pp.

Wade, P.R. (1999) A comparison of statistical methods for fitting population models to data. In: *Marine Mammal Survey and Assessment Methods* (eds G.W. Garner, S.C. Amstrup, J.L. Laake, B.F.J. Manly, L.L. McDonald and D.G. Robertson). Balkema, Rotterdam, pp. 249–270.

Walters, C. and Bonfil, R. (1999) Multispecies spatial assessment models for the British Columbia groundfish trawl fishery. *Canadian Journal of Fisheries and Aquatic Science* **56**, 601–628.

Walters, C., Christensen, V. and Pauly, D. (1997) Structuring dynamic models of exploited ecosystems from trophic mass-balance assessments. *Reviews in Fish Biology and Fisheries* **7**, 139–172.

Walters, C. and Ludwig, D. (1994) Calculation of bayes posterior probability distributions for key population parameters. *Canadian Journal of Fisheries and Aquatic Science* **51**, 713–722.

Walters, C. and Maguire, J.-J. (1996) Lessons for stock assessment from the northern cod collapse. *Reviews in Fish Biology and Fisheries* **6**, 125–137.

Walters, C. and Pearse, P.H. (1996) Stock information requirements for quota management systems in commercial fisheries. *Reviews in Fish Biology and Fisheries* **6**, 21–42.

Warren, W.G. (1994) The potential of sampling with partial replacement for fisheries surveys. *ICES Journal of Marine Science* **51**, 315–324.

Williamson, N.J. (1982) Cluster sampling estimation of the variance of abundance estimates derived from quantitative echo sounder surveys. *Canadian Journal of Fisheries and Aquatic Science* **39**, 229–231.

Wolpert, R.L. (1995) Comment: Inference from a deterministic population dynamics model for bowhead whales. *Journal of The American Statistical Association* **90**, 426–427.

Zheng, J., Funk, F.C., Kruse, G.H. and Fagen, R. (1993) Evaluation of threshold management strategies for Pacific herring in Alaska. In: *Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations* (Alaska Sea Grant College Program Report No AK-SG-93-02) (eds G. Kruse, D.M. Eggers, R.J., Marasco, C., Pautzke and T.J. Quinn II). University of Alaska, Fairbanks, pp. 141–167.

Appendix to references: availability of unpublished reports

(1) G. Stefánsson, Marine Research Institute, Skulagata 4, PO Box 1390, 121 Reykjavik, Iceland.

(2) K.R. Patterson, FRS Marine Laboratory, PO Box 101, Victoria Road, Aberdeen, AB11 8DB, United Kingdom.

(3) B. Mesnil, IFREMER, BP 21105, 44311 Nantes Cedex 03, France.

(4) V. Restrepo, ICCAT, Calle Corazón de María, 8, Sixth Floor, 28002 Madrid, Spain.