ICES Journal of Marine Science, 56: 797–810. 1999
Article No. jmsc.1999.0552, available online at http://www.idealibrary.com on IDEAL®

Improvement of fishery-management advice through simulation testing of harvest algorithms

J. G. Cooke

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

The risks and benefits associated with a given fishery-management measure, such as choice of a total allowable catch (TAC), depend on how future measures are to be chosen. It is therefore more appropriate to assess the risks and benefits of procedures or algorithms for determining management actions rather than those of single actions. The properties of a management algorithm can be explored by simulating its behaviour in hypothetical test scenarios. In each scenario, annual data with random error are generated from a simulated fishery and used by the algorithm to determine the TAC or other management measure for the next year. The performance of a management algorithm is measured against a variety of competing criteria that reflect the extent to which it achieves the conservation of fish resources, allows a reasonable level of utilization, and provides an acceptable degree of stability in catch or effort limits. Through an iterative process of testing and development, a management algorithm can be found that is robust towards uncertainties and that can be tuned to provide the desired trade-off between competing performance goals. The approach is illustrated using the Catch Limit Algorithm developed for the management of baleen whale harvests.

Key words: fishery management, harvest algorithm, scenarios, simulation testing.

Received 26 April 1999; accepted 6 August 1999.

J. G. Cooke: Centre for Ecosystem Management Studies, Mooshof, 79297 Winden, Germany. Tel: +49 7681 6018; fax: +49 7681 6019; e-mail: jgc@cems.de

Introduction

In ICES, and in many other parts of the world, fishery scientists are required to provide regular advice to fishery managers based on biological assessments of the state of exploited fish stocks. The assessments are typically subject to a high degree of uncertainty; therefore, the advice based on them cannot be expected to be infallible. In the last 10–20 years, this uncertainty has increasingly been incorporated into management advice. Classical approaches to handling uncertainty involve the determination of confidence limits for various quantities of management interest, such as the catch level that would achieve a given target fishing mortality. Modern approaches to the handling of uncertainty include various forms of risk analysis (Hilborn *et al.*, 1993) and Bayesian methods (Punt and Hilborn, 1997). These methods enable assessments of stocks and management advice to be framed in probabilistic terms, so the manager has a quantitative indication of the range of possible scenarios consistent with the data.

As an example of probabilistic assessments, Figure 1 shows a probability distribution of estimates and forward projections of spawning stock biomass for *Thunnus thynnus* in the western North Atlantic based on recent assessments (ICCAT, 1999). ICCAT's management objectives include the rebuilding of depleted stocks towards levels providing the maximum sustainable yield (MSY). Its scientific advisers are required to advise on the level of total allowable catch (TAC) that would achieve this. The assessment of the stock is based on estimates of past catch and its age composition and indices of relative abundance derived from selected fisheries. Because these indices are subject to considerable variance, assessment of the current state of the stock is subject to considerable uncertainty. Random variability in recruitment and uncertainty in the stock–recruitment relationship contribute further to uncertainty in projections for the stock under various harvest levels. The figure shows the median and 10th and 90th percentage points of the estimated probability distribution of past levels and future projections of spawning

1054–3139/99/060797+14 $30.00/0

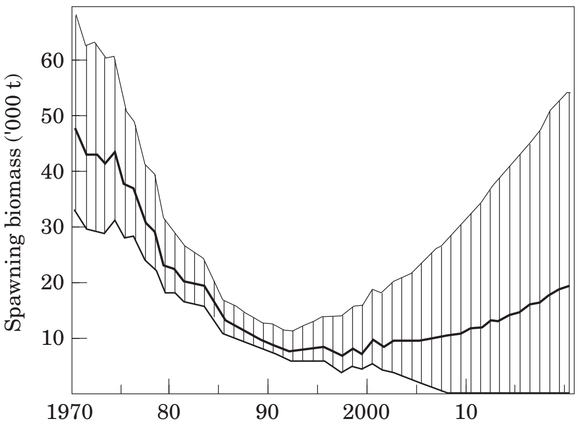

Figure 1. Probabilistic projections of spawning stock biomass for *Thunnus thynnus* in the West Atlantic: 10th percentile, median, and 90th percentile of the distribution, assuming current TAC continues 2000–2020 (based on assessments by ICCAT, 1999).

stock biomass levels under the current TAC over the next 20 years.

Taken at face value, the projections would suggest that, if the current TAC level is maintained, there is a substantial probability that the stock will decline further, so jeopardizing the management objectives, even though the median trajectory indicates that a reasonable rate of recovery would be expected. A management agency would want to know what level of risk they are running if they maintain the current TAC and put their faith in the median prediction. Unfortunately, the probabilistic projection shown does not in itself provide the information needed to assess this. The projections indicate the expected range of outcomes that would arise from setting a fixed annual TAC level for the full 20-year period from now until 2020. However, the assessments will be updated regularly with new data, in this case every 2 years, and new projections provided. What would be the risk, for example, associated with regularly setting the TAC at a level that corresponds to an acceptable recovery in the median projection?

This discussion illustrates the limitations of attempting to evaluate the risk associated with a single management action, because the consequences of a management action depend on subsequent actions anticipated in response to new data. The relevant issue is the risk associated with following a given management policy or procedure over time, rather than a single measure. A management procedure in this context means a rule for choosing a management action. In this example, the rule could be to set the TAC at a level that, according to the latest assessment, has a given probability, say 50%, of allowing the stock to recover to the target level within, say, 20 years. The question of interest is what would be the consequences of following this rule regularly over time. Relevant "consequences" here could include the average realized catch, the variability of the TAC, and the risks of failing to achieve management targets by various margins.

An approach to this type of problem is to simulate the operation of the prospective management rule under various sets of assumptions. For such a simulation approach to provide a meaningful reflection of the nature of the uncertainty faced, it is essential that the management rule be applied in the simulations, as in reality, on the basis of the data collected up to each point in time. A simulation of the management rule in which the state of the stock at each time was assumed known would miss the essence of the problem.

In most practical situations, management actions are not decided on the basis of a fixed rule, but through a less well-defined process in which a variety of relevant factors of a biological, economic, operational, and political nature are taken into account, and which may involve a process of brokerage and negotiation. However, it remains a relevant question for fisheries scientists who provide management advice to determine what would be the expected consequences, over time, of their advice being followed.

The related topic of determining an optimal management or control policy for a dynamic system, known as stochastic control theory, has also been applied to fishery-management questions (Ludwig and Walters, 1982). Stochastic control theory has its roots in the study of linear stochastic systems. In the special case of linear systems, which satisfy certain additional conditions, a simple analytic algorithm exists to find the optimal policy (Bertsekas, 1976). For such systems, the optimal solution has the elegant property that the optimal policy is equivalent to that which would pertain from ignoring the uncertainty and treating the observations as exact. This property of "certainty equivalence" unfortunately does not apply to systems with non-linear dynamics, including harvested fish stocks (Ludwig and Walters, 1982). The solution of the optimal control problem for these systems in the presence of observation error and parameter uncertainty is much more difficult. The simpler problem of optimal control for a fishery system where the parameters are treated as known is considerably more tractable, and some useful results have been obtained (Marchal and Horwood, 1998). Although techniques for solving the problem of finding an optimal control in the face of parameter uncertainty have been explored in the fisheries context, they have as yet not been widely applied to practical fishery-management problems. Technical difficulty is undoubtedly one of the reasons for this, but a further issue is that the approach focuses on finding an optimal solution with respect to a single criterion for a given model and set of assumptions about the dynamics of the resource. The problem of more practical interest is to find a robust policy that provides adequate, albeit potentially suboptimal, management performance with respect to multiple criteria over a wide range of models and assumptions about the dynamics of the resource.

The approach described here is less ambitious: no attempt is made to find formally optimal solutions. The evaluation, and hence the potential improvement, of management advice, is achieved by simulating its implementation under a range of alternative assumptions. The approach has its roots mainly in work by De la Mare (1986a), who simulated the behaviour of the process then used for the assessment and management of exploited whale stocks. The main procedure used in the 1970s and early 1980s for the assessment of whale stocks was to fit a population model, with parameters determined from biological considerations, to time series of relative or absolute abundance data. The assessment results were used to obtain estimates of the quantities required for implementation of the management procedure of the International Whaling Commission (IWC), known at the time as the new management procedure (NMP). These quantities were estimates of the MSY and of the current stock level relative to the level expected to provide the MSY. The NMP control law was then applied to determine the catch limit (TAC) as a function of the estimated MSY and the estimated stock level relative to the MSY. Simulations of the process showed that, in hypothetical repeated applications, it would not be expected to perform well over time, neither in terms of conservation of stocks nor in terms of providing reasonably stable and high levels of allowed catch (De la Mare, 1986a).

Poor performance of the existing procedures prompted a search for a more satisfactory management procedure. A number of alternative candidate algorithms for determining TACs were developed and subjected to a series of simulation tests of increasing severity. The development of algorithms proceeded mainly on the basis of trial and error: at each stage of testing, algorithms were modified to improve their performance in the tests, before moving on to the next level of tests (Kirkwood, 1992; Cooke, 1995). The algorithm that eventually emerged as giving the best balance of performance over a range of tests was selected as the catch limit algorithm (CLA), which formed the core of the IWC's revised management procedure (IWC, 1994). Some features of this particular algorithm are described below.

A simulation approach to the development and evaluation of management algorithms has also been applied to the development of management algorithms for fisheries for hake (*Merluccius* spp.) (Punt, 1992), anchovy (*Engraulis capensis*) (Butterworth and Bergh, 1993), and other species (Geromont *et al.*, 1999) off southern Africa. It has also been widely applied in the management of Australian fisheries (Smith *et al.*, 1999).

Scientific evaluation of management algorithms inevitably focuses on just one part of the management process, a part that should not be confused with the entire process. The part of the management system of most concern to fishery scientists is the process by which a recommended harvest level (such as a TAC or effort limit) is arrived at on the basis of available data. In order to distinguish it from the wider management process, this component of a management system is termed in this paper the *harvest algorithm*. Other authors use the term "management procedure" or "operational management procedure" (e.g. Geromont *et al.*, 1999; Smith *et al.*, 1999). These terms are avoided here because they are suggestive of a complete management system. The term "algorithm" presupposes a sufficiently well-defined process to be amenable to simulation. The term "harvest algorithm" covers algorithms that yield catch limits, effort limits, season or area limits, or other means controlling or influencing the level of harvest. There are at least three levels at which the approach of simulating the operation of harvest algorithms can be applied:

(i) The properties of alternative harvest algorithms can be explored at a purely scientific level. Even in the absence of well-defined management objectives, the results of such analyses may be sufficient to rule out algorithms whose performance is erratic or otherwise unsatisfactory in some obvious respect.

(ii) The simulated performance of harvest algorithms can be summarized relative to performance criteria that relate, even if only approximately, to stated or implicit objectives of the management authority. The results can be used to identify an algorithm for providing advice that serves the policy of the management authority better.

(iii) Scientists can work jointly with decision-makers and stakeholders to discuss management objectives and the extent to which they are met by alternative harvest algorithms. These are then subjected to simulation tests, the results of which are fed back into a further round of discussions. The simulation tests reveal the different trade-offs between the competing management objectives achieved by different tunings of the algorithms and thereby assist the interested parties in their choice of a harvest algorithm for their fishery.

The development of the CLA for the IWC operated mainly at levels (i) and (ii). The more advanced level (iii) approach has been advocated by Butterworth *et al.* (1992), and subsequently implemented, at least in part, for several fisheries off southern Africa (Geromont *et al.*, 1999) and Australia (Smith *et al.*, 1999). The purpose of harvest algorithm evaluation is not to determine how a fishery should be managed, but to help find an algorithm that better supports an agreed management policy. It must be stressed that none of the points made in this

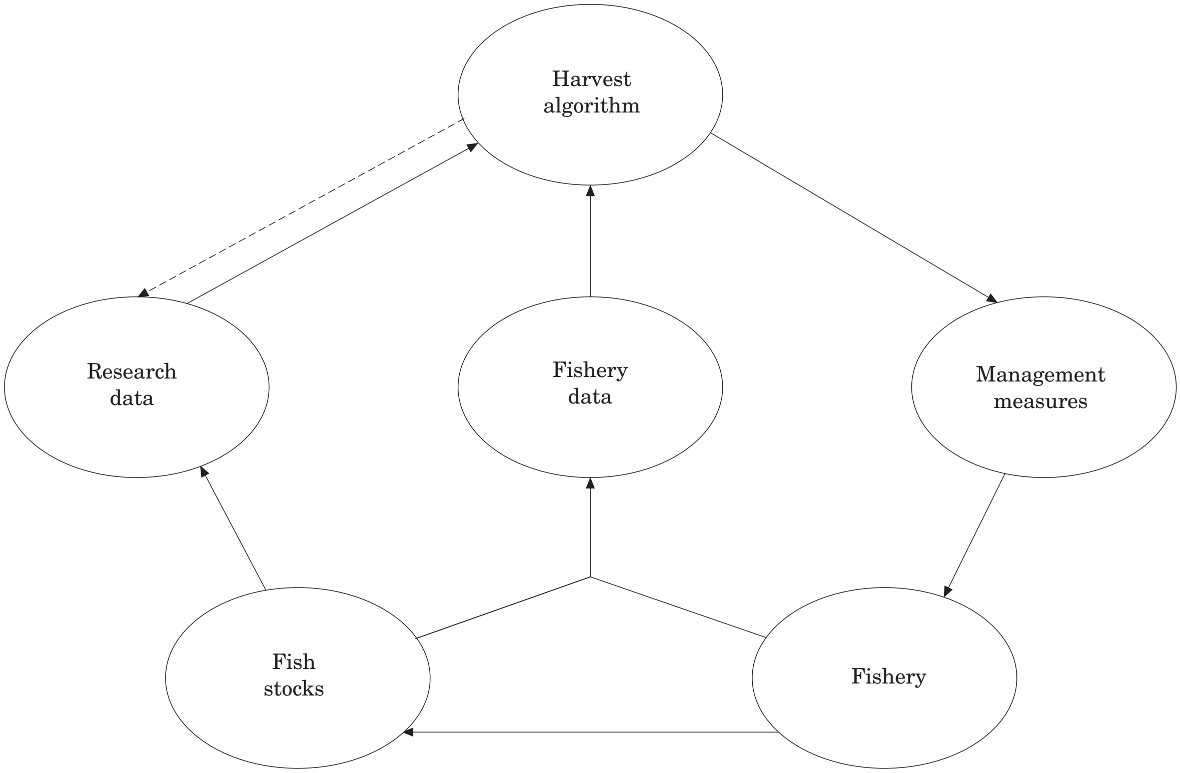

Figure 2. Structure of the harvest algorithm simulation process.

paper regarding the evaluation of harvest algorithms is intended to be prescriptive as to how fisheries should be managed.

Methods

Preliminaries

The elements required to simulate the effects of a harvest algorithm are shown schematically in Figure 2. Data relating to the fish stock are collected from the fishery, and possibly also from fishery-independent research surveys. These data are used as input into the harvest algorithm to compute a recommended harvest level (usually in terms of a catch or effort limit), a recommendation that influences the actual level of catch or effort in the fishery. The catches taken, and the fishing effort expended to obtain them, generate fishery-related data, such as the age or size distribution of the catch, and an index of catch per unit effort (c.p.u.e.). The catch also affects the fish stock, both directly and in terms of subsequent recruitment through a stock–recruitment relationship. Surveys of the fish stock may also be conducted to gather data for input into the harvest algorithm. The cycle is then repeated. A key feature is that there is no link directly from the fish stock to the harvest algorithm. The algorithm operates only on the data collected: it has no access to the true state of the fish stock.

The harvest algorithm could in principle also incorporate rules that govern some aspects of the data collection process, such as when to conduct a new survey; hence, the optional dashed arrow from the harvest algorithm to the research data.

In order to simulate the system, a model is required for each of the links in Figure 2. In the first instance, some of the models can be trivial, as noted below.

(i) Dynamics of the fish stock: The minimally realistic model will usually incorporate age structure, growth (weight-at-age), natural mortality, a stock–recruitment relationship, and variability in recruitment. More advanced models, used perhaps for sensitivity testing, can include serial correlation in recruitment variability, variability in growth and natural mortality, medium-term changes in stock productivity, spatially structured and multi-species models.

(ii) Response of the fishery to management measures: The minimal model would specify merely that the fishery makes full use of the catch or effort limit set by the algorithm. To avoid unrealistic behaviour in extreme cases, an upper bound can be placed on the level of fishing effort. For example, even when the

TAC exceeds the stock size, there will be some escapement because the catches will be limited by the power of the fishing fleet.

(iii) Impact of the fishery on the stock: A typical minimal model would be that the fishery takes the specified catch, or exerts the specified level of effort, subject to a constant, specified pattern of age-specific catchability.

(iv) Generation of data: The nature, quantity, and statistical properties of the data used for management need to be specified. Typical examples are an index of c.p.u.e. from each fishery with a specified coefficient of variation and age samples taken randomly from a specified fraction of the catch. Where fishery-independent surveys are conducted, their frequency and the statistical properties of the resulting indices of abundance (which may be age-specific) need to be specified. The coefficient of variation of c.p.u.e. can be taken to be a function of the level of catch or effort; this captures the effect that reducing the level of catch or effort can reduce the effective amount of information.

(v) The harvest algorithm: For the purpose of the simulation tests, the harvest algorithm can be treated as a black box, without reference to its internal design and functioning. The data are input and the recommended harvest level is output. In practice, the internal design of the algorithm is also of some interest, as discussed below in relation to the CLA.

The combined set of assumptions used to generate the component models (i) through (iv) can be termed a scenario. The harvest algorithm is evaluated by simulating its performance over a range of scenarios. Each scenario will involve at least some random processes. The most important of these are the sampling variability inherent in the data-generation process (observation errors) and the random variability in recruitment to the fish stock (process errors). To obtain a representative picture of how the harvest algorithm performs in a scenario, one can simulate multiple independent replicates of the scenario, typically 100 or more.

A time horizon for the simulations is required. Experience suggests that time-scales normally considered most practical, such as 10–20 years, are too short to provide a good indication of the dynamic properties of the managed system (De la Mare, 1986b). Furthermore, if only a short period is simulated, the results are dominated by the starting conditions; extra care must be taken to ensure that the scenarios cover a sufficiently representative range of starting points. For the purpose of gaining a scientific understanding of the properties of harvesting algorithms, it is advisable to run the simulations over a longer time horizon than would normally be considered of immediate practical relevance. Performance assessments for presentation to managers can be based on a more worldly time horizon.

Performance measures

In principle, evaluation of the expected performance of a harvest algorithm through simulation trials should be with respect to management objectives specified by the management authority. In practice, such objectives may not be formulated precisely, so a formal quantitative evaluation of harvest algorithm performance with respect to them is not possible. If precisely formulated objectives are provided, they should be used. Otherwise, a common-sense-based approach can be followed. This is justifiable if the conclusions are not too sensitive to the details of performance criteria.

Regardless of the details of management objectives, most modern fishery-management policies appear to be aimed, implicitly or explicitly, at steering an intermediate course between over- and forced underexploitation of fish resources, although there is considerable scope for variation in the specific criteria for what constitutes each. This goal can alternatively be expressed as achieving a compromise that allows for a reasonable level of utilization of fish resources on the one hand, while achieving an acceptably high probability of conserving the stock on the other. An additional implicit objective that is often present is that the management system should allow reasonably stable fisheries without unnecessary fluctuations in the level of permitted catch or effort.

Even these vague criteria may be sufficient for the elimination of unlikely candidates for a harvest algorithm. For example, an algorithm that yields erratically varying catch or effort limits regardless of the state of the stock would be clearly unsatisfactory with respect to any plausible management objective. Furthermore, an algorithm that yields severe overexploitation in some replicates of a scenario and severe underexploitation in others would be judged to be unreliable, regardless of the relative importance attached by the management authority to the avoidance of over- or underexploitation.

Summary performance measures need to be specified which relate, even if only approximately, to the explicit or implicit management objectives. In the first instance there are four main axes of performance to consider:

- Conservation performance: This could be measured in terms of the average proportion of time that the spawning stock biomass is above some threshold, such as the minimum biologically acceptable level (MBAL) (ICES, 1992). For evaluation of the CLA, the IWC scientific committee used the lower 5%-ile of the lowest mature stock size in the simulation period (Kirkwood, 1992). This measure has the advantage that alternative management procedures can be compared for their conservation performance without the need to specify a particular threshold stock size.

- Utilization performance: A natural choice is the average level of catch obtained over the simulation period, but possibly time-discounted to reflect the fact that the shorter term yields will usually be considered more important by those with an interest in the utilization.
- Stability: An obvious choice is the average annual change in the harvest limit. Alternatively, since reductions in catch or effort generally pose more problems for fisheries than increases, the average annual reduction in the harvest limit could be used as the performance measure.
- Reliability: Whereas stability relates to stability over time within a replicate, reliability relates to inter-replicate variability. An obvious choice is the inter-replicate standard deviation of utilization performance. The evaluation of the CLA used the lower 5%-ile of the average catch as a reliability measure; this has the advantage that it is interpretable as a level of catch that can be guaranteed with 95% confidence.

In each simulation test, all aspects of management performance that relate to the state of the fish stock or stocks are measured in terms of the state of the simulated stocks as defined in the scenario models. The harvest algorithm may itself involve a stock assessment as part of its operation, but the state of the stock as judged by the harvest algorithm's internal stock assessment model is not used to define performance criteria.

The primary purpose of computing these performance measures is not to judge the performance of a harvest algorithm relative to an absolute standard, but to provide a means by which different algorithms, or different tunings of the same algorithm, can be compared. For example, of two algorithms that perform equally well on the conservation, stability, and reliability criteria, that yielding the higher utilization performance would be preferred. Likewise, if two algorithms allow a similar level of utilization, that with the better conservation performance would tend to be preferred. Given two algorithms with similar performance on the conservation and utilization criteria, the algorithm that yielded less inter-replicate variation in utilization would be considered more reliable.

For longer lived species, unconstrained maximization of a time-discounted utilization measure tends to favour an unsustainable pattern of exploitation over time, because benefits are maximized by reducing the stock to the minimum fishable level as rapidly as possible (Clark, 1990). However, under the current treatment, such a strategy would tend to attract a low score on the measure of conservation performance and so would not necessarily be the preferred choice overall. Here, conservation and utilization objectives are treated as separate dimensions. Although they could, in principle, be converted into a common economic currency (Conrad and Clark, 1987), this would require substantial additional data and assumptions regarding fishery economics that are beyond the present scope. The assumption that the conservation objectives of most fishery-management authorities can be expressed in purely utilitarian terms is also questionable.

To address concerns that the results could be overly sensitive to a particular choice of performance measures, more than one performance measure can be computed for each of the main axes. If the different performance measures for a given axis turn out to be highly correlated, they can be reduced to one measure without appreciable loss of information. For example, the IWC Scientific Committee had difficulty determining the most appropriate measure of catch stability and computed three alternative measures. After the three measures were found to be highly correlated with each other, all but one was discarded (IWC, 1991).

Selection of test scenarios

The initial testing and development of a harvest algorithm is aimed at finding one that works satisfactorily under ideal conditions. For this purpose a limited set of basic scenarios may be sufficient. Once a sufficiently promising candidate for an algorithm has been identified, it can be tested against more demanding scenarios. The initial set of scenarios would normally include:

- A range of values for stock productivity. The most uncertain parameter tends to be the slope of the stock–recruitment relationship relative to the rate of natural mortality. The scenarios should therefore cover the full range of plausible values for this parameter combination.
- A range of starting conditions; for example, scenarios with an initially lightly fished stock and others with an initially heavily fished stock.

Once a promising harvest algorithm has been identified, its performance in more demanding scenarios can be tested. These could include:

- scenarios involving misleading data, such as spurious trends in abundance indices and misreported catches
- medium- and long-term changes in stock productivity
- spatial structure and multiple stocks
- interactions between species.

It is difficult to give specific guidance on the range of scenarios that need be considered before one can be reasonably confident that good performance of the harvest algorithm in the scenarios considered can be

Table 1. Examples of test scenarios addressing identified potential factors affecting management performance.

Nature of potential problem	Test scenarios required
Uncertainty over stock boundaries	Scenarios in which the stock boundaries for the simulated stocks differ from those used for management
Uncertainty over age-specific rates of natural mortality	Scenarios in which true rates of natural mortality differ from those used in the assessment model (if any)
Doubts over reliability of estimated catch-at-age owing to low sampling rates	All scenarios should include the sampling process by which the data used for assessment are actually obtained
Possible long-term changes in stock productivity (regime shifts)	Scenarios in which carrying capacity changes without the knowledge of the manager
Concern over shape of stock-recruitment functions	Scenarios in which true stock-recruitment function differs from that assumed (if any) in the assessment model
Concern that variance of abundance indices is over/underestimated	Scenarios in which true variance of abundance indices differs from values used for assessment

interpreted as evidence of its good performance generally. A pragmatic approach is to scan the recent reports of the scientific advisory body for the management of the species and stocks in question, and to identify all the issues that have been raised as placing potentially serious doubt on the reliability of the assessment and management advice provided. Scenarios can be constructed in which each of these potential problems pertain. The performance of the harvest algorithm under simulations of the scenarios will reveal whether the problems could potentially seriously degrade the performance of the harvest algorithm. Many of the potential problems will relate to factors that may be operating but are not known to be. In these cases, the harvest algorithm should not be "told" which factors actually pertain in each set of simulated scenarios against which it is tested. For example, Table 1 lists some issues that are raised in the latest assessment report for the *Thunnus thynnus* stock discussed in the introduction (ICCAT, 1999). Alongside each issue, the table lists the test scenarios that could be constructed to assess whether a proposed harvest algorithm is robust towards the potential problem.

The scenarios implemented for testing the CLA are listed in Table 2. The experience with developing the CLA was that, once the algorithm had been developed to a point at which it performed satisfactorily in ordinary scenarios, it was already fairly robust to the problems encountered in many of the more demanding scenarios (Cooke, 1994), but it is unclear whether this finding applies generally.

Construction of a harvest algorithm: the CLA example

The function of a harvest algorithm is to generate a catch or effort level from the input data. As noted above, it can be treated as a black box algorithm, one developed and evaluated entirely on the basis of its performance. Strictly speaking, its internal workings are irrelevant. However, experience shows that the search for a well-performing algorithm can be facilitated by following certain guidelines in its construction. These are illustrated here using the IWC's catch limit algorithm (IWC, 1994) as an example.

Table 2. Scenarios used for simulation tests of the catch limit algorithm.

Initial stocks levels ranging from 0.05 K to 0.99 K.
Different values of MSY: 0.01–0.07 of MSY stock level.
Different ages at first fishing and first reproduction.
Constant multiplicative positive and negative biases in abundance indices.
Positive and negative linear trends in biases of abundance indices.
Continuous and non-continuous (25-year gap) pre-management fishing.
Catches correctly reported and catches underreported by 50%.
Different stock-recruitment functions (left-skewed, right-skewed, various shapes).
Linear trends in parameters (r and/or K) over time.
Cyclic changes in parameters (14- and 33-year cycles).
Randomly occurring mass mortalities (50% stock die-off).
Abundance data (from research surveys) at frequencies of 1, 2, 5, 10, and 20 years.
CV of abundance data: range from 0.1 to 1.0.
Catch per unit effort (various tests conducted, resulting in a decision to abandon use of this index).
Two stocks managed as one.
One stock managed as two.
Stocks with overlapping ranges.
Stocks with shifting ranges.
Stocks with overlapping and shifting ranges.

Source: Reports of the International Whaling Commission, 1989–1994.

In brief, the CLA works by fitting a logistic surplus production model to a time series of abundance data. The variance of the parameter estimates is reduced by using Bayesian methods involving prior distributions. The recommended harvest level is a given percentile of the posterior distribution of a specified function of the estimated stock level and the model parameters. For a more in-depth discussion of the use of Bayesian methods in fitting the logistic surplus production model, the reader is referred to McAllister and Kirkwood (1998). The emphasis in the following description of the CLA is on those aspects that may be of more general relevance to the construction of harvest algorithms.

The main components of the approach followed in the construction of the CLA are detailed and discussed below.

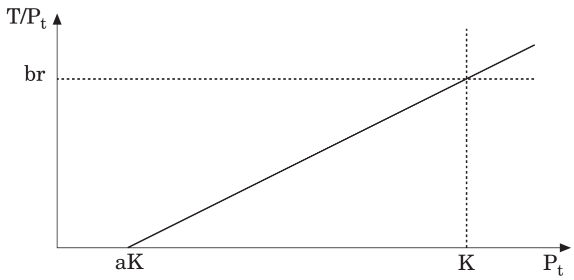

Figure 3. Control law for the CLA: TAC (T) as a function of current stock size (P_t). K (notional unexploited equilibrium stock level) and r (stock productivity) are parameters of the stock assessment model to be estimated, and a and b are parameters, and of the control law, whose values are chosen to yield the desired management performance.

Stock assessment model

The stock assessment model should capture the essential features of the dynamics of the resource that are relevant to management performance. The choice of assessment model determines the nature of the data reduction required. In the case of baleen whales, the relatively long lifespan, coupled with the physiological limits to reproductive rate, means that no single age class will contribute more than a small part of the total stock size. Hence the essential features of the dynamics can be captured by a bulk surplus production model (Schaefer, 1953). The stock assessment model used in the CLA is:

$$P_t = P_{t-1} + r(1 - P_{t-1}/K)^2 - C_{t-1} \qquad (1)$$

where P_t denotes the population in year t, C_t is the catch in year t, and r is a parameter reflecting productivity. $K = P_0$ represents the notional unfished stock level that is a feature of surplus production models. If harvests C_t are known, then the model is fully determined by the parameters K and r. Incorporation of a time-lag corresponding to the age at first fishing or reproduction does not improve performance. The exponent 2 in the second term reflects a convention among marine mammal biologists to assume a right-skewed production curve, but the effect of this on the behaviour of the model is slight.

The motivation for choosing a given stock assessment model is not to achieve a high degree of realism, but to choose a model that, when combined with the estimation procedure and the control law (see below), yields a harvest algorithm which performs well for management purposes. Performance is evaluated using the simulation tests described below. It is important that the robustness of the stock assessment model be tested by including test scenarios in which the true dynamics differ substantially from the behaviour implied by the stock assessment model.

The bulk surplus production model is not necessarily appropriate for all species. For species where individual year classes are important, an age-structured model that allows incorporation of year-class variability may be required.

Control law

The control law specifies the harvest level as a function of the model parameters. The control law for the CLA is illustrated in Figure 3. The formula is:

$$T = b\ r\ P_t(P_t/K - a) \qquad (2)$$

where T is the TAC, a and b are intercept and slope parameters, respectively, and P_t is the estimated current stock size from the fitted model. The intercept a is the level below which the harvest is set to zero. The IWC selected the relatively high value $a = 0.54$ for the intercept, to reflect a management objective that depleted stocks be allowed to recover as fast as possible. In other applications, a more gradualist policy might be preferred: a zero intercept, for example, would imply that the fishery is never closed completely. The defining characteristic of harvest control laws is that harvests are set below replacement rates at low stock levels, and above replacement rates at high stock levels.

The control law alone does not yield a unique value for the harvest limit, because it depends on the unknown parameters r, K and P_t, whose values are determined using the parameter estimation protocol.

Data reduction

The data-reduction step need not be written into a harvest algorithm, but any harvest algorithm presupposes some form of data reduction to supply the required input. The CLA requires a time series of abundance estimates, along with past catches that are assumed known. There may exist more than one time series of abundance estimates, some of which may be indices of absolute abundance and some only of relative abundance. In accordance with the general statistical principle of sufficiency (Cox and Hinkley, 1974), the full dataset should be reduced to what is relevant to the estimation of the model parameters. Provided they are non-zero it is usually acceptable to assume that abundance estimates are log-normally distributed, in which case the relevant information in a given abundance index is contained in the time series of log abundance values and its information matrix. If it is non-singular, the information matrix is equal to the inverse of the variance–covariance matrix of the index values. An extended discussion of data reduction is beyond the scope of this paper, but the following examples indicate some of the most useful techniques, all of which can be derived using standard matrix algebra (Horn and Johnson, 1985).

If there are m independent log-abundance indices, y_i, with respective information matrices H_i, then the information matrix of the combined index is given by $H=\Sigma_i H_i$ and the combined index itself is given by $y=H^- \Sigma_i H_i y_i$ (i=1, ..., m), where H^- denotes a generalized inverse of H.

Nuisance parameters associated with a given index can usually be eliminated by making appropriate adjustments to its information matrix. For example, a relative abundance index, Y_t, is related to absolute abundance by $Y_t \approx Q\ P_t$, where Q is an unknown parameter. Using the convention that lower case letters refer to the logarithms of the upper-case quantities, we have, on a log scale, $E(y_t)=q+p_t$. The nuisance parameter q can be eliminated by leaving the index values y_t unchanged and applying the following transformation to its information matrix:

$$H \rightarrow H - H\ 1\ (1^T\ H\ 1)^-\ 1^T\ H$$

where 1 denotes a vector of ones, and the superscripts ^{T} and $^-$ denote transpose and generalized inverse, respectively. If the parameter q is not totally unknown, but an estimate is available that can be assumed to be distributed as $N(\mu,\sigma^2)$ so that the corresponding estimate of Q has a lognormal distribution, then q is eliminated by the transformations:

$$y \rightarrow y - \mu 1$$

$$H \rightarrow H - H\ 1\ (1^T H\ 1+\sigma^{-2})^-\ 1^T H$$

These techniques allow the reduction of multiple abundance indices, including mixtures of relative and absolute indices, to a single index and associated information matrix. Their utility is to reduce the number of parameters remaining to be handled in the next phase of the process.

Parameter estimation

Following the data-reduction step, the parameters of the model are to be estimated. In the case of the CLA, the likelihood of the reduced data is proportional to $\exp(-\frac{1}{2}(y-p)^T\ H(y-p))$ where y is the log abundance index and p is the vector whose components are $\log(P_t)$.

Traditionally in fisheries science, parameters are estimated by maximum likelihood or least squares. However, such estimates typically have too high a variance for direct use as the basis for management. Recommended harvest levels based on such estimates can change radically each time a new data point is added. For shorter data series, there is a high probability of obtaining extreme estimates. A solution is to reduce the variance, at the expense of introducing some bias, by anchoring the estimates in some way. The method used in the CLA is to assign prior probability distributions to the unknown parameters. These are combined with the likelihood function to yield the posterior distribution using Bayes law. The Bayesian approach is becoming increasingly popular in fish stock assessments (Punt and Hilborn, 1997; Virtala *et al.*, 1998).

There is a range of views among both theoretical and practising statisticians as to the interpretation of Bayesian prior and posterior distributions. According to the pure Bayesian approach, the prior distributions should reflect beliefs about the true values of the unknown parameters, and the posterior distributions can be used to make literal probability statements about parameter values. The philosophy behind the CLA is the pragmatic one of choosing prior distributions to yield good management performance and to satisfy certain design criteria. The stock assessment model of the CLA was not chosen for its realism, and it is unclear to what extent there can be said to exist true values for the parameters of a model which is itself not true. The important issue is the performance of the algorithm across the range of uncertainties contained in the scenarios included in the evaluation.

For reasons discussed by Punt and Hilborn (1997), it is desirable that the prior for K be scale-invariant. An advantage of keeping the entire assessment process scale-invariant is that, when conducting simulation trials, it is not necessary to consider scenarios with different values of K, because everything scales in proportion to it. It is easily shown that the only scale-invariant prior distributions for K that are functions of K alone are of the form K^λ for some fixed exponent λ. If only relative abundance data are available, the requirement that the posterior distribution of K be normalizable imposes the constraint $\lambda < -1$. If absolute abundance data are available, then any value of λ, even a positive one, is admissible in principle. Negative values of λ favour lower stock sizes when data are sparse and can be termed conservative priors. Positive values for λ favour higher stock sizes.

If the prior for K is allowed to be a function of other parameters, the scale-invariance condition can alternatively be satisfied by assigning a prior distribution to any dimensionless parameter combination involving K. The

CLA assigns a uniform prior over the range (0, 1) to the relative current stock size P_t/K. Despite appearances, this is a conservative prior. For a given catch history, it implies a prior for K that is asymptotically proportional to K^{-2} for large K. The CLA uses a uniform prior for r on the range (0, 0.07): this was based mainly on the range of r values used in the simulation tests, that in turn were based on information on observed rates of increase of baleen whale populations (Best, 1993). For the CLA to be applicable to a wider range of species, it would probably be necessary to scale the prior for r to some measure of lifespan, such as the median age of the spawning stock.

The CLA also contains a prior distribution for an additional parameter that represents the bias in abundance estimates: the effect of this is to improve slightly its robustness to bias in the abundance index, but it may not be an essential feature.

The posterior distribution of the unknown parameters is obtained by multiplying the likelihood by the prior distributions, which is equivalent to adding them on a log scale: log(posterior) = log(prior) + w . log(likelihood), where w is a weighting factor. For a strictly Bayesian analysis, w should have the value 1, but for the CLA a value $w<1$ was chosen. Since the priors are fixed and the likelihood is data-dependent, lower values of w yield TACs with lower variance. With $w=1$, the harvest algorithm in simulation tests yielded unacceptably high variability, especially in scenarios where the variance of the abundance estimates was negatively biased.

It remains to be seen whether downweighting of data is a generally desirable feature of harvest algorithms. The actual variance structure of fisheries abundance indices can be quite complex, with a hierarchy of variance components at various levels (Cooke, 1997). The data are typically insufficient to permit empirical estimation of the variance with much precision (McAllister and Kirkwood, 1998). The model error arising from the mismatch between the assessment model and the true stock dynamics may also contribute to the effective variance. These factors combine to yield a considerable degree of uncertainty in the estimation of variance, such that cautious harvest algorithms may have to err on the side of underweighting the input data.

A consequence of using prior distributions or other means to reduce the variability of parameter estimates is that the parameters of the fitted assessment model will be shrunk towards central values implied by the priors. Thus, stocks with below-average productivity will tend to be overexploited to some extent, whereas stocks with above-average productivity will tend to be underexploited to some extent. As more data are accumulated that provide evidence of true productivity, the degree of under- or overexploitation will tend to diminish, but the effect will not disappear quickly under most realistic scenarios. Balancing the disadvantages of shrinkage against the advantages in terms of variance reduction is an implicit part of the process of tuning a harvest algorithm (see below).

Determination of the harvest limit

In classical likelihood-based parameter estimation, the point estimate of the unknown parameters is usually taken as the set that maximizes the likelihood, this being invariant with respect to monotonic model re-parameterization. In the case of Bayesian analysis, it is the percentiles of the posterior distribution, such as the median, that are parameterization-invariant. Other features of the posterior distribution, such as the mode or the mean, are not invariant and hence effectively undefined. The control law provides the nominal harvest limit, T, as a function of the model parameters r, K, and P_t. The actual TAC set by the CLA is a fixed percentile of the posterior distribution of T. Originally, the median (50%-ile) had been selected, but this was later modified to 42% when the algorithm was tuned to satisfy a specific conservation target set by the IWC in a reference scenario (IWC, 1994).

Note that the selected percentile of the posterior distribution of T cannot be obtained by using the corresponding percentiles of the individual parameters r, K, and P_t, as arguments to the harvest control law. The result obtained from the latter approach would depend upon the two parameter combinations chosen as the base parameter set, and would therefore not be uniquely defined.

Tuning of the algorithm

The CLA contains a number of free parameters, each of which can be adjusted to tune the performance of the algorithm to achieve the desired balance of performance over a range of simulated scenarios. The tuneable parameters include: the prior distributions for each of the parameters of the assessment model; the weighting factor used for combining the priors with the likelihood; the parameters of the harvest control law; and the choice of percentile of the posterior distribution of the harvest limit.

Changing any one of these parameters while holding the others fixed will in general improve performance relative to some performance measures, and degrade performance relative to others. Figure 4 shows how the performance of the CLA with respect to conservation, utilization, stability, and reliability criteria can vary as the algorithm is tuned by varying the percentile parameter. The results relate to a particular reference scenario, designated the D1 scenario in IWC documentation (IWC, 1994), but the details are not important for the current context. The key point is that, by varying the tuning parameter, different trade-offs between the competing objectives of conservation, utilization, and stability can be achieved. Different choices of tuning

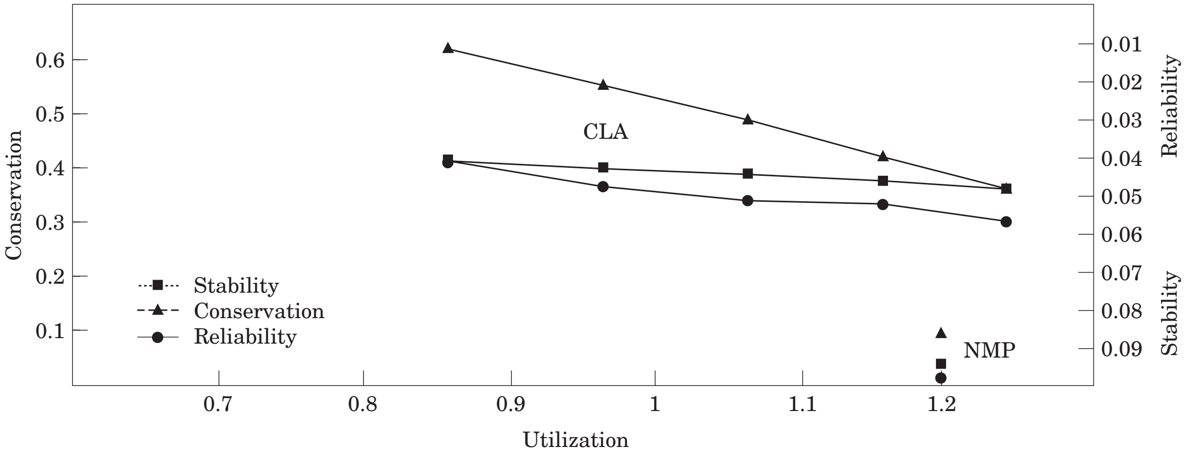

Figure 4. Performance of different tunings of the CLA, in terms of conservation (lower 5%-ile of lowest stock size), stability (mean annual relative change in TAC, with lower values uppermost), and reliability (inter-replicate standard deviation of total catch; × 0.5), plotted against utilization (mean total catch). The performance of the former management procedure (NMP) is also shown for comparison. Stock size and catch scaled to K.

parameter would in general yield slightly different trade-offs between the different axes of performance. Also shown in the Figure is the performance of the IWC's previous management procedure, the NMP. The NMP is clearly dominated by the CLA, in the sense that the latter outperforms the NMP on the conservation, stability, and reliability criteria when it is tuned to provide the same level of utilization to the NMP. If a harvest algorithm dominates an alternative in this manner across the full range of test scenarios, then it can be judged preferable to the alternative, regardless of the relative weight attached by the management authority to the different performance criteria.

Discussion

Harvest algorithms

The main components of the harvest algorithm, the stock assessment model, the control law, and the parameter estimation protocol, are evaluated in terms of their performance as a team. A single component, such as a control law, that works well in conjunction with a given stock assessment model and parameter estimation protocol, cannot be assumed to work well with different partners. The new combination would need to be subjected to a simulation testing process. There is no reason why a harvest algorithm should necessarily contain components that are separately identifiable as an assessment model, an estimation protocol and a control law, but the example of the CLA shows that this can be a practical approach to the construction of harvest algorithms.

The CLA was robust to all the single-stock scenarios listed in Table 2. The results of scenarios involving spatial structure and multiple biological stocks revealed the need for additional rules that prevent fishing effort being concentrated into a small part of the area, if conservation of local stocks is to be ensured (IWC, 1994).

Merits and limitations of simulation tests

There are a number of reasons why simulation testing of management procedures provide an important supplement to what can be learned in the field:

- Time: Field tests can only be done in real time. In the case of fisheries, it can take years or decades before the success or otherwise of a management approach, especially with regard to its effectiveness at conserving stocks, becomes apparent.
- Sample size: A consequence of the large inherent variability on fishery-management processes is that several replicates are needed to assess the performance of a harvest algorithm. A limited number of field trials could be insufficient to distinguish between good management and good luck. There are few examples in the world where multiple replicate fish stocks with similar characteristics exist.
- Diagnosis: In simulation tests, the cause of a management failure, such as a stock collapse, can be diagnosed. It is easy to determine what changes to the algorithm are required to avoid the failure. In a real management case, one might register that the stock had collapsed but would not necessarily learn what would have been required to avoid the collapse.
- Costs and learning: Failures of management in real tests incur real costs. Experience with the development of harvest algorithms, such as the CLA, is that

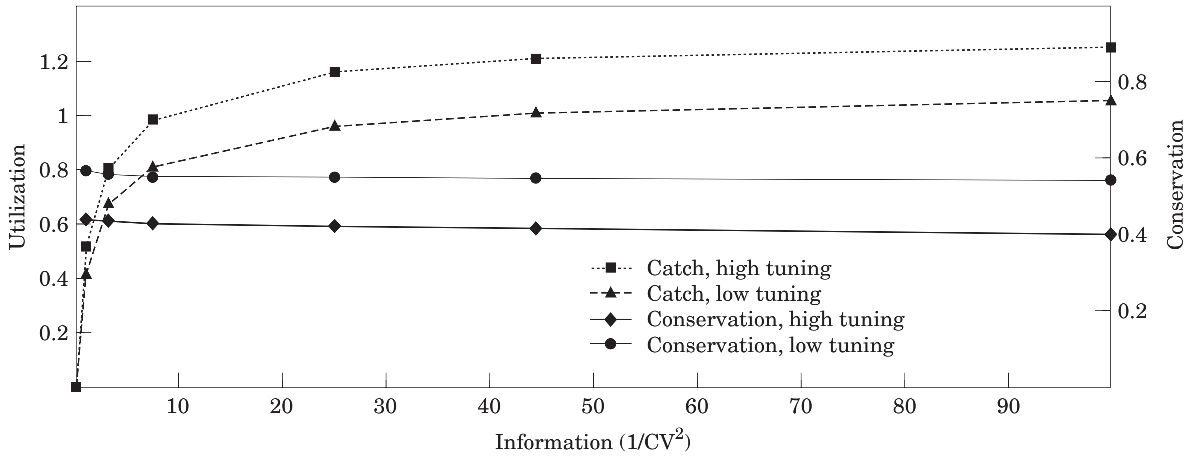

Figure 5. Relationship between conservation (lower 5%-ile of lowest stock size), utilization (mean total catch), and information level (1/(CV)^2 of abundance estimates) for two tunings of the CLA in the reference scenario. Stock size and catch scaled to K.

early versions can perform poorly relative to the final design (Cooke, 1989). Use of incompletely tested designs in the field would be wasteful.

Perhaps most compelling are the arguments against not performing simulation tests. Who would now advocate, for example, designing and producing a passenger aircraft without extensive testing of the components in wind tunnels? Not all potential defects in a harvest algorithm are of a nature that will reveal themselves in simulation tests, but it is prudent to identify and correct those that do, before incurring real risks in real fishery-management situations.

Precautionary management

Recently, there has been increasing interest in a precautionary approach to fisheries-management (FAO, 1995). Definitions and emphases vary as to what constitutes a precautionary approach. An aspect of precautionary management that is of interest in the present context is the relationship between the risk of stock depletion and the quantity of information available. The precautionary principle prescribes that lack of information should not be a reason for failing to take adequate measures to conserve stocks. This can be interpreted quantitatively by requiring that the performance of a harvest algorithm relative to conservation criteria should not deteriorate as the level of information decreases.

In the case of the CLA, the level of information is related to the frequency and precision of the surveys conducted to estimate abundance. The information gained from a survey is approximately proportional to the amount of survey effort and hence to the cost of the survey. The coefficient of variation of the abundance estimate from the survey is inversely proportional to the square root of the information. The CLA has been tuned so that the risk of depletion (as measured by the lower 5%-ile of the lowest mature stock size), for a given scenario, remains roughly constant as the level of information varies (Fig. 5). When the available information is low, there is great uncertainty over the stock size, and TACs are set low, and vice versa. The actual level of risk can be varied by tuning the algorithm. Each tuning level yields a different curve relating utilization to information, but the curves for different tuning levels have the same general shape. The positive relationship between information and utilization means that the value of information to the fishery is positive. In non-precautionary management, harvest levels are not limited until there is sufficient information to demonstrate the necessity for limits: under such regimes, information has a negative value to the fishery in the short term.

The shape of the curves of utilization versus information implies that, for constant risk, the marginal returns diminish as the quantity of information increases. A rate of utilization only slightly below the maximum can be achieved for a relatively small quantity of information without engendering an increased risk of stock depletion. The slopes of the curves approach the vertical near the origin, so, regardless of the cost per unit of information, there will be some level of information at which the benefits of utilization exceed the costs of obtaining the required information. Interpreted in this way, precautionary management need never imply a zero harvest.

Use of biological and ecological information in the design of harvest algorithms

The philosophy behind the approach advocated here to developing and testing harvest algorithms is to distinguish between: (i) the assessment model integrated into the algorithm and (ii) the various biological, ecological, and other models used to generate scenarios for the purpose of testing the harvest algorithm. The last-mentioned can conveniently be called scenario models. The assessment models are tuned for performance, whereas the scenario models are constructed to cover a broad range of plausible scenarios. The usual need to compromise between performance and realism in the choice of models is avoided. Parameter estimation is not a pressing requirement for the scenario models: if information on a parameter is poor or absent, a range of scenarios that span the plausible range of values for the parameter can be constructed. The available data are also divided into two classes: those which enter the harvest algorithm explicitly, and those which are used only in the generation of scenarios for testing harvest algorithms.

The main requirements for the scenario models are: (i) they contain sufficient realism that the performance of a harvest algorithm in each scenario can be expressed in terms that relate to real-life management objectives; (ii) they cover a sufficiently broad range of plausible scenarios such that satisfactory performance of a harvest algorithm across the scenarios considered can be taken as evidence that the algorithm itself is satisfactory.

The approach may lend itself to handling more complex management problems in situations involving ecological and technical interactions between species. The potential number of dimensions in multi-species problems can be quite large, to the extent that it would not be feasible to develop and apply a harvest algorithm in which all the important interspecies interactions were explicitly incorporated. However, there are few obstacles to incorporating what is known about interspecific interactions into the repertoire of scenarios used for testing the performance of harvest algorithms.

Closing remark

To date, development and testing of harvest algorithms, and in particular the construction of test scenarios, has required a substantial input of labour by the fishery scientists involved. This has been a limiting factor in the application of the approach to fishery management. However, as experience is gained as to which kinds of harvest algorithms work well, and as a "library" of standard test scenarios accumulates, future implementations of the approach to new fisheries may involve little more than fine-tuning of existing algorithms and test scenarios, so that more general use of the approach becomes practicable.

References

Bertsekas, D. 1976. Dynamic programming and stochastic control. Academic Press, New York.

Best, P. B. 1993. Increase rates in severely depleted stocks of baleen whales. ICES Journal of Marine Science, 50: 169–186.

Butterworth, D. S., and Bergh, M. O. 1993. The development of a management procedure for the South African anchovy resource. *In* Risk evaluation and biological reference points for fisheries management, pp. 83–99. Ed. by S. J. Smith, J. J. Hunt, and D. Rivard. Canadian Special Publications in Fisheries and Aquatic Sciences, 120.

Butterworth, D. S., Punt, A. E., Bergh, M. O., and Borchers, D. L. 1992. Assessment and management of South African marine resources during the period of the Benguela Ecology Programme: key lessons and future directions. *In* Benguela trophic functioning, pp. 989–1004. Ed. by A. I. L. Payne, K. H. Brink, K. H. Mann, and R. Hilborn. South African Journal of Marine Science, 12.

Clark, C. W. 1990. Mathematical bioeconomics: the optimal management of renewable resources, 2nd ed. John Wiley. 386 pp.

Conrad, J. M., and Clark, C. W. 1987. Natural resource economics: notes and problems. Cambridge University Press, Cambridge. 231 pp.

Cooke, J. G. 1989. Simulation studies of two whale stock management procedures. Report of the International Whaling Commission, Special Issue 11: 147–156.

Cooke, J. G. 1994. The management of whaling. Aquatic Mammals, 20: 129–135.

Cooke, J. G. 1995. The International Whaling Commission's Revised Management Procedure as an example of a new approach to fishery management. *In* Whales, seals, fish and man, pp. 647–657. Ed. by A. S. Blix, L. Walløe, and Ø. Ulltang. Developments in Marine Biology, 4. 720 pp.

Cooke, J. G. 1997. A procedure for using catch-effort indices in bluefin tuna assessments. ICCAT Collective Volume of Scientific Papers, 46(2): 228–232.

Cox, D. R., and Hinkley, D. V. 1974. Theoretical statistics. Chapman & Hall, London. 511 pp.

De la Mare, W. K. 1986a. Simulation studies on management procedures. Report of the International Whaling Commission, 36: 429–450.

De la Mare, W. K. 1986b. On the management of exploited whale populations. PhD Dissertation. University of York, York, England. 217 pp.

FAO. 1995. Guidelines on the precautionary approach to capture fisheries and species introductions, Part 1. Report of the Technical Consultation. FAO Fisheries Technical Papers, 350(1).

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

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

Horn, R. A., and Johnson, C. R. 1985. Matrix analysis. Cambridge University Press, Cambridge. 561 pp.

ICCAT. 1999. Report of the ICCAT SCRS Bluefin Tuna Stock Assessment Session. ICCAT Collective Volume of Scientific Papers, 49(2): 1–205.

ICES. 1992. Reports of the ICES Advisory Committee on Fishery Management, 1991. ICES Cooperative Research Report, No. 179.

IWC. 1991. Report of the Subcommittee on Management Procedures. Report of the International Whaling Commission, 41: 90–112.

IWC. 1994. The Revised Management Procedure (RMP) for baleen whales. Report of the International Whaling Commission, 44: 142–152.

Kirkwood, G. P. 1992. Background to the development of revised management procedures. Report of the International Whaling Commission, 42: 236–243.
Ludwig, D. A., and Walters, C. J. 1982. Optimal harvesting with imprecise parameter estimates. Journal of Ecological Modelling, 14: 273–292.
Marchal, P., and Horwood, J. 1998. Increasing fisheries management options with a flexible cost function. ICES Journal of Marine Science, 55: 213–227.
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.
Punt, A. E. 1992. Management procedures for Cape hake and baleen whale resources. PhD thesis. University of Cape Town. 875 pp.
Punt, A. E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Reviews in Fish Biology and Fisheries, 7: 35–63.
Schaefer, M. B. 1953. Fisheries dynamics and the concept of the maximum equilibrium catch. Journal of the Gulf and Caribbean Fisheries Institute, 1: 1–11.
Smith, A. D. M., Sainsbury, K. J., and Stevens, R. A. 1999. Implementing effective fisheries-management systems – management strategy evaluation and the Australian partnership approach. ICES Journal of Marine Science, 56: 967–979.
Virtala, M., Kuikka, S., and Arjas, E. 1998. Stochastic virtual population analysis. ICES Journal of Marine Science, 55: 892–904.