The approach recently developed by the IWC Scientific Committee for taking formal account of the results of robustness trials, together with their relative plausibilities, in assessing risk when selecting between alternative candidate management procedures

D.S. Butterworth

MARAM (Marine Resource Assessment and Management Group)
Department of Mathematics and Applied Mathematics
University of Cape Town
Rondebosch 7701, South Africa

A particular raison d'etre for the management procedure (MP) approach for providing scientific recommendations on management measures such as TACs is that it takes formal account of scientific uncertainties in its assessment of risks (primarily to the resource, but also to the fishery). This is achieved by considering the results of simulation tests which project the resource forward under the MP's TAC-setting algorithm, not only under a model seen to best reflect the resource's dynamics, but also under other models consistent with alternative plausible explanations of the data available.

For acceptability, a candidate MP must demonstrate reasonable performance across statistics related to management objectives (such as low risk to the resource), not only under the "best assessment" model, but also under the "robustness tests" based upon these other models. This raises two problems however:

a) with the candidate MP tuned to provide "optimum" performance for the "best" model, performance will deteriorate to some extent for the other models – how much deterioration is acceptable, and

b) the plausibility of these alternative models also needs to be factored into the evaluation of risk in this process – extreme interpretations of the data which will lead to high probabilities of heavily reduced resource abundance can always be advanced, but need they be taken into account if such scenarios are considered to have low plausibility?

Evaluation of such robustness tests is often conducted on only a "tick test" basis – on inspection, do the associated performance statistics seem not to be substantially worse than for the "best" model? But particularly in circumstances where groups with appreciably different interests are involved in such deliberations, consensus can prove difficult to achieve on this basis. Hence it is desirable to move towards a more specific framework for formal incorporation of the results from such robustness tests in selecting between alternative candidate MPs.

This paper summarises the International Whaling Commission's (IWC's) Scientific Committee procedures developed to this end, which are set out in detail in the Annex hereto (extracted from the Report of the Committee's 2004 meeting). First, however,

some details concerning the IWC's RMP (Revised Management Procedure) for commercial whaling are necessary for background.

The IWC's RMP

The RMP is a generic procedure intended for potential application to any baleen whale resource. At its heart is the CLA (Catch Limit Algorithm) which, given historic catches and one or a series of estimates of abundance from surveys of an area, will generate a catch limit for that area through a Bayes-like application of a simple population model. The algorithm has the property (consistent with the Precautionary Approach) that, other things being equal, abundance estimates with higher variances (i.e. greater uncertainty) will result in lower catch limits being output.

The CLA was evaluated for a wide range of robustness tests, and judged to perform acceptably across a certain range of "tunings". The CLA has a control parameter that can adjust the trade-off between higher catches *vs* lower risks of unintended reduction in resource abundance regarding which a decision needs to be made in the management of any fish resource. The Scientific Committee deemed that a range from 60 (reflecting higher catches, but higher risk) to 72%[1] (for which catch dropped, but risk was reduced as well) for this tuning was acceptable. The Commission made the final choice of 72% for their adopted MP.

The CLA is designed for application to the idealized situation of a single stock (population), with no uncertainty about stock structure. This situation scarcely ever pertains in reality, so that certain rules are added to the CLA to appropriately spread catches in space (and time within the year if necessary) to limit risk in situations where there is plausibly more than one stock present and the location of the boundaries separating such stocks (or the extent to which they overlap) is uncertain. For example, if most past catches have been limited to a small area, whereas abundance estimates pertain to a much larger area over which whales are distributed, setting catches on the basis of such abundance estimates alone without further restrictions could place at great risk what might be a localized stock from which most of the past catches had been taken.

These rules require the CLA to be applied at the level of "*Small Areas*" into which the overall area surveyed is divided, with catch limits set at this smaller scale. However, because estimates of abundance calculated for smaller areas have larger sampling variances, this process leads to smaller catches overall – perhaps to an unnecessary extent. Therefore there are further rules which may also be applied, e.g. "*Cascading*" under which the CLA remains applied at a larger areal scale, but the catch limit output is then allocated amongst the constituent *Small Areas* in proportion to the abundances estimated in each.

---

[1] The details as to exactly what these numbers relate are not important for the purposes of this summary, but in brief they refer to median population levels (relative to the pre-exploitation level *K*) anticipated after the application of the CLA over a 100 year period for one of the core trials against which the CLA was tested.

The combination of the CLA with these rules for spreading catches is known as the Revised Management Procedure (RMP)[2]. In a particular instance of "*Implementation*" of the RMP, trials specific to the species and region in question, which in particular incorporate alternative hypotheses for stock structure, are developed to test and thereby select which of these rules to apply.

Taking formal account of results from robustness trials

This section summarises the key steps in the process set out in sections 3 and 4 of Appendix 2 of the Annex hereto, to be found on pgs 86-89 thereof.

Plausibility

Suggested trials are each accorded one of four weights based on the plausibilities assigned to the hypotheses that underlie them (see left side of pg 87 of the Annex). These weights are high, medium and low, and "no agreement" for scenarios for which a reasonable case can be made for a high weight but there is no consensus.

Low weight trials are not considered further, and for the purpose of this summary the "no agreement" trials can be considered to be treated identically to those accorded medium weight.

Equivalent single stock trials

A difficulty that arises in multi-stock trials is identifying whether or not the level to which management might have depleted any one of the constituent populations, or allowed such populations to recover, is acceptable in terms of risk. This is not entirely straightforward, because even in the simple case of the CLA applied to a single stock, the simulated final population size distribution after the 100-year management period typically considered is not fixed, but depends on factors such as the size of the resource when application of the RMP is initiated and its productivity.

The underlying concept adopted was that application of the RMP in a multi-stock case should be such that no stock was depleted further than would have been the case in the idealized "single stock + CLA" combination: hence thresholds for acceptable extents of depletion for multi-stock trials are developed from population abundance distributions after 100 years of application of the CLA to an "equivalent single stock trial".

---

[2] For an overview, together with references for further details, see Kirkwood, G.P. 1992. Background to the development of revised management procedures. *Rep. int. Whal. Commn* 42: 236-243.

Specific statistics used for comparison

Being generic, the RMP must cater both for situations where future catches will deplete abundance from a level initially close to pre-exploitation equilibrium (*K*), and for those where recovery is sought for a population already heavily depleted.

For the former, the population risk-related statistic chosen is the lower 5%-ile of the distribution of population size as a fraction of *K* after 100 years. To cater for the latter, again the lower 5%-ile is considered – on this occasion of the distribution of the minimum over the projection period of the ratio of the population size under the RMP to that which would have eventuated in the absence of commercial catches[3]. Note that since risk is involved, the statistics specified are lower %-iles of the distributions.

Since these two statistics are each motivated by their respective associated situations described above, and would not have much pertinence in the other situation, acceptable behaviour requires only that the threshold for one of the two is met in a particular trial.

Thresholds and decisions

Thresholds are trial-specific, with two being specified for each of the statistics above, corresponding to applications of the two extreme tunings of the CLA to the equivalent single stock trial in question: 72% as for the Commission's adopted RMP, and the less conservative 60%. Results which are (see also Annex, right side of pg 87):

    i)   above the 72% threshold fall in the acceptable category;
    ii)   above the 60% but below the 72% threshold in the borderline category; and
    iii)   below the 60% in the unacceptable category.

Decisions as to the acceptability or otherwise of different "RMP variants" (different catch-spreading rules in combination with the CLA) then result from following the flowchart in Fig. 1 on pg 88 of the Annex. Key elements of this are that:

a) Failure to achieve the acceptable threshold for **any** *high* weight trial results in a candidate RMP variant being rejected.
b) If for some *medium* weight trials, performance is considered reasonably close to the acceptable threshold (while above it for the rest), the candidate RMP variant may be classed as acceptable.
c) If the "reasonably close to acceptable" criterion of b) is not met, yet the candidate shows good catch-related performance, it might remain acceptable on a "research-conditional" basis. This requires the concurrent institution of a research programme targeted at resolving the uncertainty underlying the trial causing the difficulties, together with demonstration that if it fails to do so within 10 years, acceptable thresholds can still be reached over the 100 year projection period by substitution after 10 years of a more conservative RMP variant.

---

[3] Note that these reflect subsequent refinements by the Scientific Committee of specifications given in Table 1 on pg 88 of the Annex.

Wider application of these IWC concepts?

At first sight the IWC RMP concept of a generic approach applying across a variety of stocks and species might seem too inflexible to serve even as a starter for fisheries on a wider scale. However, both US and Australian fisheries legislation now includes (or is targeted to include) generic recovery performance criteria and catch control law restrictions – in an attempt at greater inter-resource consistency, most likely as a reaction to failures to achieve recoveries under systems that admitted greater flexibility. Furthermore similar pressures are arising from the developing ecolabelling requirements of the Marine Stewardship Council. These factors suggest that time may bring a more widespread move for fisheries towards elements of the IWC's approach.

While the IWC's constructs may be somewhat more complex than necessitate replication in detail in other fisheries situations, focus on some of the core elements of the approach might nevertheless be immediately useful in taking first steps towards linking robustness test results to rules governing candidate MP acceptability, *viz.*:

> a) categorizing trial weights, in relation to plausibility, as high, medium or low; and
> b) disregarding low weight trials, while requiring candidate MPs to meet more stringent risk criteria for high weight trials than for medium weight trials.