DISCUSSION OF SCIENTIFIC ISSUES RELATED TO METARULES AND THE IMPLEMENTATION OF A MANAGEMENT PROCEDURE FOR SOUTHERN BLUEFIN TUNA

M Basson
T Polacheck
D Kolody

Prepared for the CCSBT 3rd Management Procedure Workshop Meeting
19-24 April 2004, Pusan, South Korea

Table of Contents

ABSTRACT

This paper provides a scientific discussion of metarules and implementation issues under Agenda items 5 and 6 ('Metarules, Assessments and Special Circumstances', and 'Implementation issues and other considerations') of the CCSBT Workshop on Management Procedures (April 2004). It considers that metarules are intended for exceptional circumstances. It addresses those rules and issues which are of direct scientific relevance. Circumstances under which a metarule may be required are discussed under different categories: improved knowledge about the stock, assessment results outside the range tested, clear signs of exceptional circumstances, and no data or circumstances leading to an inability to run the code of the decision rule. Metarules need to define the circumstances in which they would be invoked, a procedure to determine agreement on whether such circumstances apply and the nature of the response if such circumstances are found to have occurred. The type of metarule that might be required in each category, is considered. In some cases the response component of a metarule may simply be to recondition the operating model and/or retune the decision rule (or MP). In other cases a more complex response may be required.

Implementation issues primarily pertain to the data and other inputs required to run the decision rule, and safeguards to avoid or minimise the chances of failure of the decision rule code. A critical component of any decision rule is the reliability of the data used by the decision rule to adjust future TACs. The evaluation of an MP's performance is based on assumptions about the reliability and consistency of the data provided to it. Thus, a critical issue with respect to the implementation of an MP is the integrity and consistency of the data inputs. Mechanisms and types of verification are largely a management issue, but the scientific advice about an MP's performance is under the assumption that this issue is appropriately dealt with.

1. INTRODUCTION

This paper provides a scientific discussion of metarules and implementation issues under Agenda items 5 and 6 ('Metarules, Assessments and Special Circumstances', and 'Implementation issues and other considerations') of the CCSBT Workshop on Management Procedures (April 2004). It considers that metarules are intended for exceptional circumstances. It addresses those rules and issues which are of direct scientific relevance.

2. METARULES OF SCIENTIFIC RELEVANCE

Metarules in the context of the management procedure can be thought of as 'rules' which pre-specify what should happen in unexpected, exceptional circumstances. Metarules also have a role to play in cases where circumstances have changed substantially. Examples of this might be where the range of uncertainty used in simulation testing of the decision rule no longer overlaps with that implied by the new circumstances, or where the decision rule appears not to be performing as expected on the basis of the simulation evaluations. We therefore consider it sensible to build in a process whereby the CCSBT can periodically review whether there have been such substantial changes to warrant, for example, reconditioning and/or retuning of the chosen decision rule (the term decision rule, DR, is used here for the rule which generates a TAC from the data).

It is critical to have metarules that predefine what constitutes “exceptional” circumstances and an agreed response by the Commission so that there is both transparency and clarity in the decision making process. Metarules are not a mechanism to avoid painful decisions, but are meant to deal with truly “exceptional” circumstances. It is also important that there is a well defined process for determining when metarules should be applied. A metarule should be sufficiently well described that conditions under which it applies and the actions/process to follow are clear. However, they should also not be so rigid that we are forced to either continue doing something that doesn't make sense, or unnecessarily make changes such as recondition or retune. There is therefore a fine balance to be struck when constructing metarules. Metarules should specify either an action or a well structured process for arriving at an action within the overall MP framework.

It should, however, be noted that frequent changes to a DR, are not desirable because it can severely undermine the advantages of the MP process over annual ad-hoc decision-making. Frequent changes to a DR can also lead to actual performance being very different from that assessed during evaluation. It is therefore desirable to keep the DR as the default 'action' and only make changes when absolutely necessary. The relatively slow dynamics of the stock and the information content of one additional year's data, as seen in the historical data, suggest that changes are likely only to make sense in a multi-annual framework (i.e. not annual, rather every several years).

To summarise, a metarule should have three components:

a) definition of the exceptional circumstances
b) process for deciding whether the exceptional circumstances apply or not (deciding applicability)
c) response/action or process for arriving at an action, and a time frame for this

Examples to clarify this will be given below. We find it useful to consider four categories of issues or events which could invoke a metarule:

1. substantial improvements in knowledge, or new knowledge, about the stock biology and its dynamics, or substantially improved measures of abundance and/or fishing mortality rates;
2. stock assessment results are 'narrower' than or outside the simulated range of dynamics considered in the evaluation of the performance of the DR;
3. clear signs of exceptional circumstances (e.g. recruitment failure);
4. no data or circumstances which lead to an inability to calculate a TAC from the decision rule.

A brief discussion of each of these categories is given below. Note that the categories are not in any particular order.

Category 1: Substantial improvements or new knowledge

Examples of this category are new estimates of natural mortality, maturity-at-age or growth, particularly where they are very different from current estimates used in the conditioning and tuning of the DR. A sensible response would be to review whether the new knowledge is likely to imply stock dynamics that are outside the 'envelope' generated by the existing/underlying conditioning and simulation testing, or whether the new data would lead to a marked change in the range of uncertainty that was originally considered. If so, then reconditioning and retuning of the existing decision rule to the Commission's chosen tuning level would be a sensible approach. This should not, and need not, imply any disruption to the implementation of the DR.

In terms of the three components of a metarule, an example based on the above, could be:

a) circumstances: new or substantially improved estimates of natural mortality-at-age
b) deciding applicability: agreement in the SC that estimates are substantially improved and that results are likely to be sufficiently different that action is required
c) response: recondition and retune[1] DR to the Commission's chosen tuning level, within an agreed timeframe.

A second example could relate to maturity-at-age:

a) circumstances: new or substantially improved estimates of maturity-at-age
b) deciding applicability: agreement in the SC that estimates are substantially improved and that results are likely to be sufficiently different that action is required
c) response: recondition and retune DR to the Commission's chosen tuning level, within an agreed timeframe.

Note that parts (b) and (c) of this second example, are the same as for the first example, only the definition differs - one relates to natural mortality, the other to maturity-at-age. This suggests that one could group these together in a metarule which has a more general definition, along the following lines for example:

a) circumstances: new or substantially improved knowledge about mortality-at-age, maturity-at-age or growth

[1] Exactly what is meant by recondition and retune will need careful consideration because there are implicit technicalities which are not obvious. One simple example is that if tuning was done with respect to B2022/B2002 and in real time we're beyond 2022, how should 'retuning' be done? 'Reconditioning' is likely to imply a different extent of changes to the operating model depending on what the new knowledge is, and the process could be slow and labour intensive.

(parts (b) and (c) stay as in the above examples)

This illustrates the potential for metarules to be grouped together, or to cover several similar circumstances, thus avoiding the need to set up large numbers of separate metarules. We think it best to define fewer general metarules, because the alternative of defining very detailed metarules, has a high possibility of missing something out, ending up with an unmanageably large number of metarules and being very time consuming.

Category 2: Stock Assessment Results differs from simulated dynamics

There are two reasons why one may want to run periodic stock assessments: (i) to assess whether stock abundance and dynamics are within or similar to the 'envelope' generated by the existing/underlying conditioning and simulation testing and (ii) to measure the performance of the implemented DR. This category (stock assessment) is obviously linked with the one above (new knowledge), in that new knowledge is likely to feed into a stock assessment at some stage. However, here we primarily have in mind a stock assessment updated with more years' catch and CPUE data.

Recall that many of the decision rules considered so far do not provide an estimate of abundance, and even if it does, it is usually not based on a 'full' assessment. If stock assessment results (not just in terms of biomass, but also in terms of other estimated quantities such as selectivity patterns, for example) are far outside the range implied by the conditioning and simulation evaluations, then reconditioning and retuning of the existing rule, to the previously agreed tuning level, may be needed. Again, there should be no need to disrupt the implementation of the DR. This is particularly true if the TAC is not changed/recalculated annually, because a periodic assessment review can be scheduled to occur in between the years when the TAC is calculated.

An example of a metarule in this category may be:

a) circumstances: the stock assessment is substantially outside the range of simulated stock dynamics considered in evaluations
b) deciding applicability: agreement in the SC that this is the case
c) response: recondition and retune DR to the original tuning level, within an agreed timeframe.

As more data accumulate, a stock assessment could in future also imply a range of dynamics that is narrower than (though still falling inside) the range used in the simulation evaluations. This could, for example, come about if better estimates of steepness can be obtained. In such a case, reconditioning and retuning could be done with the appropriate changes to the range of uncertainty that needs to be considered.

There may be cases when a stock assessment, with or without other changes to inputs (e.g. new knowledge), could imply stock dynamics that are so different that the tuning level needs to be revisited, for example, because it can no longer be achieved. This can become part of the response, for example, "recondition and retune to the original tuning level UNLESS there is agreement that the tuning level should be revisited." (See footnote 1, however). There would then be a further description of the course of action when the tuning level needs to be revisited.

A second reason for doing an assessment is to measure the performance of the implemented DR. In the same way that the simulation trials were used to assess the performance of different DRs, performance measures can be set up to assess whether the implemented DR is likely to achieve its built in goals. This would again only be required periodically, and a metarule would only be invoked if there is strong evidence that the DR is not performing as anticipated.

Category 3: Unexpected exceptional circumstances

Metarules in this category are meant to handle a situation where there are clear signs of exceptional circumstances, particularly problems that are potentially serious, e.g. recruitment failure. This is particularly relevant where the 'signs' are unlikely to be detected by a DR because that information is not used by the DR. As noted above, care needs to be taken when constructing metarules, because sufficient flexibility is required to be able to respond when absolutely necessary but, at the same time, circumstances need to be sufficiently well described to avoid frequent interference with the DR. Frequent interference or adjustments to outcomes from a DR could severely change the realised performance of that DR compared to its performance as assessed via simulation evaluations.

One possibility is to use something like the approach that has been taken while developing the MP. For example, one could set up a framework to do an assessment every 'n'th year (where 'n' would have to be decided - see section on Timeframe below), and in years when an assessment is not being done, to consider a set of indicators to monitor. The metarule could then specify a course of action if the SC assesses that there are 'exceptional circumstances' on the basis of the indicators. The response to this category of events is more likely to be immediate management action or intervention, rather than reconditioning and/or retuning which could be done over a slightly longer timeframe.

An example of a metarule in this category could be something like this:

a) circumstances: clear signs of recruitment failure
b) deciding applicability: agreement in the SC that there are clear signs of recruitment failure
c) response: Commission to decide on the immediate management action to take

Category 4: No data or incomplete data

Metarules in this category are meant to cope with situations where the data required by the DR are not available, or where the data are incomplete (e.g. only partial catch data are available).

An example of a metarule for such a situation may be:

a) circumstances: no CPUE data or no CPUE can be provided to the decision rule in the form required (we assume the DR uses CPUE to calculate a TAC)
b) applicability: agreement in the SC that these data or inputs are not available
c) response: In a TAC-setting year (i.e. the DR needs to be applied but cannot) - Commission to decide on the immediate management action to take. If it is NOT a TAC-setting year (i.e. the DR does not have to be run) - describe a course of action to try to obtain those data for the next application of the DR, or action to ensure data in the following year(s)

In some fora, a suggested response to missing data (including, for example, surveys used as inputs to DRs) is to reduce the existing TAC by some predefined percentage at each TAC-setting occasion, until the DR can again be run with all required inputs. Presumably this holds unless another metarule which directly affects the TAC-setting is invoked.

Some aspects of this category may in fact become part of the implementation details and could in principle be programmed into the DR rather than being an 'external' metarule (see below under ‘Implementing the DR code’).

3. IMPLEMENTATION ISSUES OF SCIENTIFIC RELEVANCE

Implementation issues primarily pertain to the data and other inputs required to run the decision rule, and safeguards to avoid or minimise the chances of failure of the decision rule code. A critical component of any decision rule is the reliability of the data used by the decision rule to adjust future TACs. The evaluation of a DR's performance is based on assumptions about the reliability and consistency of the data provided to it. Thus, a critical issue with respect to the implementation of an MP is the integrity and consistency of the data inputs. Mechanisms and types of verification are largely a management issue, but the scientific advice about an DR's performance is under the assumption that this issue is appropriately dealt with.

Data related issues

Recall that a decision rule has two components: (1) the data and other input parameters, and (2) the method for using those data and parameters to calculate a quota or TAC. The current set of evaluations were all based on a consistent set of simulated data. When implementing a DR, it is important that the inputs to the rule are well defined to ensure consistency. Implementation requires:

- specification of the data required
- whom to submit to (e.g. Secretariat), in what form, by when
- if inputs are derived from raw data, a description of how the inputs should be derived (e.g. CPUE, or catch at age)
- who should do data preparation for input to DR and by when
- who should run the DR (e.g. Secretariat)

Implementing the DR code

The last point above also relates to issues of transferring and validating the code of the chosen DR. For example, if the Secretariat is to run the DR, then the 'developer' would have to transfer the code to the Secretariat, and there may be a need to verify that the code is running properly (i.e. same results for simulation trials run on Secretariat set-up as those presented to SC). One could also ask whether there is a need to verify that the code does what the description of the method i.e. the documentation, claims. (If not, a reasonable approach may be to take the code as given, since that is what forms the basis of the evaluation and subsequent choice, and to modify the documentation accordingly.)

There may be advantages to making the code robust to missing data, and this could potentially incorporate aspects of a metarule associated with missing data. For example, if the DR is run every 3 years, and the data are available for the most recent year, but not for the year before that, one option would be to make the decision rule capable of handling that, rather than

producing an error message! Such an approach would have the advantage that there will not be a need to continue having to invoke a metarule as soon as there was one incident of 'missing data'.

Mismatch between TAC and actual Catches

The current set of evaluations are all based on the assumption that the catches taken are equal to the TAC set by the decision rule. A substantial mismatch between the TAC and actual catches would require the consideration of two issues. First, whether the DR needs to and/or does properly distinguish between the two. For example, the limits on changes in TACs, built into the decision rule, are presumably meant to apply to the TAC rather than to the actual catch. Also, where the DR uses catch data, is it coded to use TACs or actual catches (e.g. when fitting a stock production model)?

Second, a large mismatch between TACs and actual catches would affect the actual performance of the DR compared to its expected performance evaluated via simulations. The Commission would therefore need to consider which measures to take to avoid large mismatches between TACs and actual catches (if catches are much greater than TACs), in order to maintain the performance of the DR. Consideration needs to be given to metarules which may be required if measures are not effective in avoiding mismatches. When the mismatch is such that the catches are much less than the TACs, there may still be a need for a metarule which at least aims to establish why this is occurring (e.g. lack of fish or external socio-economic factors).

Making performance measures operational

This was hinted at under 'Stock assessments' above. It is likely that the CCSBT would want to be able to assess periodically, whether they are likely to achieve the agreed management objectives, i.e. whether the DR is performing as indicated by the simulation results. There is likely to be a need to define sensible operational performance measure which would only lead to a metarule being invoked if the performance is deemed to be well outside the expected performance. In principle, the performance measures used in the simulation trials could also be used in reality. In practice, the details such as the frequency of TAC-setting and the first year of implementation, and the requirements of the Commission may imply a need to devise slightly different performance measures. For example, B2012/B2002 may be useful for comparing the relative performance of two rules, but may be less informative about the performance of the implemented rule with respect to meeting management objectives.

Timeframe issues

Evaluation of the performance of decision rules were made under the assumption that the rule was implemented for the full period (until 2032) in an unchanged and uninterrupted way. If frequent changes are made to any aspect of the decision rule, its control parameters or its TAC-output (e.g. via reconditioning, retuning or ‘overriding’ of the TAC by a metarule) then the actual performance of the DR could be very different from what was assessed during simulation evaluations. One point of view with regard to the implementation of an Management Procedure (MP) is that one could simply set the MP running without intervention for at least the time-period for which it was tested. The reasoning behind this is that a key point of the process of management procedure evaluation is the quantification of risk and the choice of an MP is based on choices of acceptable levels of risk. It is, however, recognised that this is dependent on a range of assumptions, for example, the assumption that future actual dynamics (of stock, fishery etc.) would be within the range of simulated

dynamics. If this is not the case, then it could be highly risky and misleading to continue applying the decision rule without any adjustment to take into account the changed dynamics.

The key question is then: how often should we “check” whether adjustments are required (e.g. do an assessment)? How often should “new knowledge” be incorporated? In practice, there will be advantages if these activities (e.g. stock assessment) can be scheduled to occur in between years when TACs are set, if they are not set annually. It was noted above that frequent changes to a DR (or MP) can severely undermine the advantages of the MP process and affect the realised performance of a DR compared to its performance in simulation trials. There is a fine balance to strike between being too reactive or being too rigid with regard to making changes to a DR. Also, care needs to be taken that one is not simply adjusting a DR to "noise" in the data. Given the relatively slow dynamics of SBT and relatively low information content of one year's data, multi-year timeframes for reviews should be considered.

4. CONCLUDING REMARKS

Given the above discussion, we consider that fewer general metarules are preferable to a large number of very detailed metarules. The general metarules should still contain the key components (a,b,c defined above), but the focus is more likely to be on a well-defined process for dealing with the exceptional circumstances and arriving at appropriate action.

We also consider that small changes that make very little difference to outcomes and/or frequent changes to the DR should be avoided. The default should be that the DR applies unless "exceptional" circumstances invoke a metarule. There should, of course, be some metarules to deal with 'emergency' situations and which would require an immediate response. Other metarules which deal with less urgent and slower timeframe changes (e.g. improved knowledge about some aspect of stock dynamics), can be set up with a response that need not be immediate. In some cases the amount of work involved may mean that the response cannot be immediate.

Although it makes sense to consider a regular 'review' of the performance of the DR and stock status, we consider that the relatively slow dynamics of SBT and relatively low information content of one year's data imply that a multi-year timeframe should be considered.