CCSBT-ESC/0709/17

The management procedure: options for ways forward.

Marinelle Basson
Tom Polacheck
Campbell Davies

Prepared for the CCSBT 8th Meeting of the Stock Assessment Group and the 12th Meeting of the Extended Scientific Committee
4-8 September, and 10-14 September 2007, Hobart, Australia

Table of Contents

Abstract

A large amount of time and resources have been put into developing and testing management procedures (MP) for the management of Southern Bluefin Tuna. The CCSBT Scientific Committee (SC) had recommended, and the Commission adopted, an MP to implement, when the large unreported longline catches were revealed. The current large uncertainties about historic catches, length frequencies and longline CPUE mean that little confidence can be associated with the results of the previous testing of decision rules. This testing was done using a conditioned operating model and projection procedures that were highly dependent on the catch, length frequencies and CPUE estimates, but which did not take into account the large uncertainties associated with these as a result of the overcatches. Also, all the MPs tested, depended on there being a single informative CPUE series derived from Japanese vessel reported logbook data. An outstanding concern is whether a CPUE series based on unverified logbook data is an appropriate quantity on which to base a decision rule (or MP).

We consider that a management procedure is the most appropriate approach for providing scientific advice on catches. However, any management procedure needs to be linked with processes that ensure that the data used in the decision rule are reliable and verified. This paper explores the options available to the SC for: (1) addressing the historic catch, length frequency and CPUE uncertainties in the existing simulation models, both the conditioned historic part and the projection part of the model and (2) using indicators or alternative approaches as a basis for a decision rule in the near future. Without additional information that can be used to reduce some of the large uncertainties in the historic data, it will be extremely difficult, if not impossible, to adequately quantify the uncertainties (e.g. bias and precision) in order to obtain reliable probabilistic predictions of the performance of different decision rules. In such a case, it may only be possible to identify “best” and “worst” case scenarios. This paper is intended as a starting point for discussion in the SAG and SC in order to provide guidance to the Commission on options for future work.

1. Introduction

In 2005 the CCSBT Scientific Committee (SC) recommended a Management Procedure (MP) to the Commission. This was after five years of work by member scientists, the independent software developer and the independent panel, to develop and test MPs for Southern Bluefin Tuna. The Commission had adopted an MP to implement when, in 2005, “independent reviews of the Japanese SBT market data and of the Australian SBT farming operations to determine whether or not over-catching is occurring relative to the total allowable catch” were instigated (Paragraph 23, Report of the Twelfth Annual Meeting of the Commission). Results indicated that there had been large unreported longline catches of SBT over a period of at least 15 years (Report of the Special Meeting of the Extended Commission, 2006) The magnitude of the unreported catches, and the potential implications for longline CPUE, have substantial implications for the whole Management Procedure process and the management of SBT in general. Results also indicated that there appeared to be little scope for over-catch via misreporting of tuna numbers in the surface fishery, but results suggested that the age frequency of the catch may have been biased (see para 54 of the Report of the Seventh SAG meeting, 2006).

Although the importance of testing MPs for robustness to uncertainty (and bias) in historic catches had been recognised by the MP Working Group, the SAG and SC (see past reports), an approach to incorporate this uncertainty could not be agreed. Therefore, the MP testing was conducted under the implicit assumption that the historic catches were reported without error or bias. Similarly, the CPUE series used in the conditioned operating model (OM), assumed no bias, other than an assumed level of effort creep which formed part of some scenarios. It was also assumed that the estimated length frequency of the longline catches were unbiased and known for all catches. The conditioned OM is highly dependent on, and results are dominated by, the CPUE and total catch series. The size- and age-frequency data affects the estimated underlying stock structure. The extent of the unreported historic catches, and their potential implications for historic CPUE, means that very little confidence can be associated with the results of the previous testing of decision rules. The operating model itself, therefore, needs to be reconsidered in the light of the large uncertainties in historic catch, CPUE and size frequency distributions. Unfortunately, it is not a simple matter of re-running the model with 'the new/correct data' and obtaining new estimates of parameters, because there is (currently) no single set of 'new' or correct data. It also seems unlikely that there will be sufficient information to ever be able to reconstruct the historic data with a great deal of certainty. If so, we consider it scientifically inappropriate to attempt to develop a single 'agreed' dataset. Instead, the aim should be to characterise the uncertainty and biases in the historic data – or at least the plausible range for these. As a result, we consider there will continue to be large uncertainties associated with the historic catch and effort data for SBT.

It is, therefore, clear that the operating model needs to be reconsidered. Its structure, formulation and how it deals with uncertainty, as well as, the estimation of parameters and quantities used in the projections for testing MPs need to be reviewed. In the first section of this paper we consider the options for ways forward in this regard.

Once the operating model has been reconstructed and/or reconditioned, in our view it is not just a matter of re-testing the original "short list" of candidate MPs. This is because all the MPs tested depended on there being a single informative CPUE series derived from Japanese vessel reported logbook data. Some of the MPs also used the total catches (e.g. MPs based on fitting a Fox stock-production model), and longline size frequency information. In the light of the now large uncertainties about historic catches, CPUE and associated size frequencies, there is increased concern whether a CPUE series based on unverified logbook data is an appropriate quantity on which to base a decision rule (or MP). In the second part of this paper we consider the potential issues with CPUE in more detail and list potential alternatives with their advantages and disadvantages.

In spite of the implications of the overcatches for the current MP, we still consider that a simulation tested management procedure is the most appropriate approach for providing scientific advice on catches. We are optimistic that there are ways forward to achieve this. However, this is only worth doing if there are processes in place to ensure that the data used in the decision rule (or MP) are reliable and verified. If this cannot be achieved, then no MP can be expected to perform according to its evaluation in tests, and the scientific advice that can be provided would be very limited. This paper is intended as a starting point for discussion in the SAG and SC in order to provide guidance to the Commission on options for future work to achieve an agreed management procedure for SBT.

2. Addressing uncertainty in historic data

In this section, we try to constrain the discussion to the non-technical detail of 'ways forward', focusing on advantages, disadvantages and practicalities of alternative broad approaches. We consider the issues of conditioning of the operating model (OM) and then implications for generating projections to enable us to test (or re-test) MPs. The conditioning of the OM is in fact very closely linked to doing a stock assessment and the discussion here is therefore also relevant to any future stock assessment, though the focus is on MP development and testing.

2.1 The conditioning phase of the operating model

Recall that the conditioning phase of the operating model is the process whereby the model parameters are estimated from historic data to ensure that the model's performance is compatible with the historic, observed, data. In this document we use the term operating model (OM) to refer to both the structure of the model, i.e. the population and fishery dynamics equations, as well as the conditioning assumptions and equations which allow fitting of the model to data.

In our view there are three main problems facing the process of re-conditioning the existing operating model, or even of developing a new operating model. The first problem, already noted above, is that major components of the historic data on which the current OM was conditioned and the MPs tested, have now been shown to be highly biased (catches, possibly CPUE and size frequency) or unknown (e.g. size frequency of unreported catches).

The second problem is that the two independent reviews do not provide alternative input data series for direct use in the OM or a statistical framework for turning the results from the review into quantities needed for the OM. In order to turn the outcomes of the reviews into inputs for the OM, several additional technical assumptions need to be made. For example, there is a lag between catches being physically taken from the population, and those fish appearing in Japanese market statistics. When interpreting estimates of SBT in the Japanese Market Review it is therefore necessary to make some assumption about this lag (see e.g. paragraph 26 of the Report of SC11, 2006). More importantly perhaps is that the review does not provide information to determine if, and how, both CPUE and the size distribution may have been affected.

The third problem is that the scenarios in the two reviews do not (yet) constitute scientifically justifiable set of scenarios for direct use in MP testing. There are two difficulties here: (1) the reviews do not provide a statistical framework for estimating the biases and variances associated with the estimated catch series and (2) there was only limited consideration given to sensitivity to underlying assumptions. Thus, the outcomes in the reviews are themselves based on a number of assumptions for which only limited direct data were available. It seems likely that there are alternative assumptions which have not been considered, or provided (for whatever reasons), but which would also be plausible and which could potentially affect the estimated magnitude, and temporal pattern, of the overcatches. This is not to criticise the reviews, but to highlight that, in the context of scientifically testing an MP to ensure robustness to uncertainties and biases, it is particularly important to adequately represent the full range of significant uncertainties and biases. At this stage the SAG/SC have not been able to review or evaluate whether the scenarios are likely to adequately cover the underlying uncertainty or not.

Of particular concern is the possibility that there may be quite different plausible assumptions which could affect the patterns of quantities over time. For example, calculations in the market review were based on the assumption that the proportion of SBT that goes through the market was constant over time with no variability. This begs the question whether a time-varying proportion may also be plausible, and possibly more realistic. It might be inconvenient if there are not enough data to estimate any such trends, but this does not mean that the possibility of trends or patterns should be ignored. The pattern over time could be important for the uncertainty in the total catches, and hence for the stock dynamics.

From a scientific point of view, identifying individual scenarios is not the only way (or even the best way) of characterising uncertainty. An alternative approach is to derive probability distributions for quantities at each step in a process, and then explicitly propagate the associated uncertainty through the entire process. This is quite a different way of approaching the problem but if this could be done, it would provide a much better basis for constructing the models and processes whereby MPs could be tested for robustness to uncertainty in the historic data.

At this point it is useful to identify two broad approaches for taking the additional uncertainty in input data into account in a conditioned OM:
1. ‘discrete’ scenario modelling using the existing framework (existing OM framework)
2. creating a new framework which explicitly models the uncertainty in input data and possibly considers alternative model structure for the population and fishery dynamics (new OM framework)

There is a third option, which is really a continuum and consists of a combination of the two approaches. It may, for example, turn out to be simpler to treat some aspects of the uncertainty in input data as scenarios, and other aspects using a different approach. This would require some changes to the OM, but not as substantial as for option 2 above.

The ‘scenario’ approach is essentially similar to the approach used in 2006 to explore the implications of the findings of the two reviews. A scenario can, for example, be constructed by taking an assumed historic catch series, assumed CPUE series and assumptions about the size frequencies of the catches. This approach can therefore use the same OM as used before (possibly with some minor modifications), and it explicitly continues to be based on total catches as an input[1].

[1] For the purpose of this discussion we ignore the additional complication that the current OM requires catches by fleet, rather than total catches, as inputs, but this is a non-trivial aspect of the characterisation of the uncertainty.

Under the new framework approach, we are not, at this stage, proposing any particular model structure, or formulation, because the detail will need to be carefully considered depending on past experience and the availability of the necessary pieces of information. In general though, the notion is to reconsider something closer to a full Bayesian approach. Under the 'new framework' approach we envisage a very wide range of possible model structures for taking the additional uncertainty into account. This could mean changes to the underlying population and fishery dynamics equations and/or changes to the way in which the model is conditioned to historic data.

The idea is to reconsider which quantities are assumed as given , which quantities are estimated, which inputs are treated as "known" and measured without error, and which inputs are used in the objective function to fit the population/fishery model to the data. For example, the total catches should arguably no longer be treated as "known" inputs, but maybe rather as random variables with expected values and variances. It may be possible to use the reported catches as lower bounds for catch, and model the additional catch as a random variable. (See Hammond and Trenkel (2005) for an example of such an approach, noting though that they assume the availability of a survey catch rate index.) At this stage it is impossible to judge whether a full Bayesian approach is feasible, or whether it would be necessary to develop a hybrid model (e.g. where only some aspects are dealt with in a Bayesian way) in order to deal with constraints in our ability to specify priors for some quantities.

A new framework may also reconsider how the model should be structured depending on how the data can sensibly be input. One might for example reconsider whether the fleet definitions are the most appropriate or useful in the light of how the overcatch can be characterised. For the surface fishery, for example, the findings of the review suggest that the catches should be input in terms of numbers rather than weight. It may even be worth considering whether a much simpler population model, such as a stock-production model, should be used

There are distinct advantages and disadvantages of the two approaches. First we primarily focus on advantages and disadvantages for conditioning; implications for projections and MP testing are considered further below.

Advantages of the 'scenario' approach for conditioning:

- This approach is obviously much faster to program and to adapt the existing software to be able to do conditioning runs with a given scenario of input data. Some such runs were done last year at the SAG, in the scenario modelling (see Report of SAG7). There will, however, be a need for additional pre- and post-processing software to automate the process of evaluation and to make it less error prone.
- The assumed input dataset for each scenario will be obvious rather than implied by a distribution or estimated sets of parameters.

Disadvantages of the 'scenario' approach for conditioning:

- How should the set of scenarios be constructed to include the plausible range of uncertainty, and to constitute a scientifically justifiable set? There is also little data for determining what would constitute the range of uncertainty that needs to be considered.
- Without additional information on the relative plausibility of each scenario, how should results from different scenarios be combined (both for 'Assessment' and MP evaluation purposes)? This point arguably relates more to the projection/testing phase of the MP where, in the past, each scenario has been given a relative weight and the MP has then been evaluated by combining results from all scenarios according to their relative weights.
- There is likely to be a very large number of different scenarios because there are several data types that are uncertain: a) longline fishery total catch, b) how the catch is allocated between the longline fleets in the model, c) the longline size frequencies associated with unreported catches, d) Japanese longline CPUE, e) surface fishery total catch weight and f) associated size frequency. This is not a simple set of 'independent' scenarios, and this will make the process of determining weights even more difficult.
- The number of scenarios will become even more unwieldy when we recall that there were already a large number of 'grid points' (what we previously called scenarios) associated with several parameters in the current MP (e.g. value of steepness, assumptions about the Indonesian fishery selectivity, etc.). Recall just how difficult it was to look at diagnostics and implied 'plausibility'

(goodness of fit) for all conditioning runs. With a much expanded set of scenarios this task will become close to impossible.

With regard to the choice of relative weights for scenarios, we note that the SAG/SC have used a combination of approaches (relative values of the objective function component and/or pre-specified weights) to obtain weights for the scenarios used during the recent round of MP testing. The 'scenarios' in that context did not refer to different catch scenarios, but to different assumptions about steepness, Indonesian fishery selectivity etc. Some aspects of this approach are quite ad-hoc, with a particular problem being the fact that the objective function is not a proper likelihood. This problem will be substantially greater in the current context of an expanded set of scenarios which are not independent.. Determining weights in this situation is likely to be problematic without additional data to constrain the range of plausible hypotheses for how the overcatch occurred. For example, on what basis would one assign weights to the degree to which the overcatch was comprised of small or large fish?

There is an alternative way in which scenarios can be used, which avoids the need for specifying relative weights, and avoids integrating over all scenarios. This is not so much an issue for conditioning, but it is relevant in the context of projections and MP testing. This is discussed further under Projections and MP testing.

Advantages of the 'new framework' approach for conditioning:

- This approach offers a way to more succinctly quantify uncertainty (e.g. through random effects, probability distributions with means and variances) and to do so in a 'continuous' rather than discrete/scenario way.
- The approach is potentially more statistically rigorous than the 'scenario' approach, though we recognise that there are always likely to be some aspects that simply cannot be dealt with rigorously.
- The approach also offers the potential for a framework which might get away from having to come up with explicit weights for a given scenario, though priors would then be required for the inputs (see below under disadvantages). This approach may allow a more rigorous way of combining priors for different components which are not independent, though this is by no means easy
- It provides a chance to improve aspects of the previous conditioning (e.g. how the 1990s tagging data were fitted) and to incorporate some additional relevant data ,(e.g. aerial survey and more recent tagging data, the latter which are also affected by the assumptions about historic catches.

Disadvantages of the 'new framework' approach for conditioning:

- This will be a time-consuming approach, because it will involve at least some reformulating of the existing operating model and rewriting the software accordingly
- It will not be an easy exercise because there are many options for, and different ways of, developing a new framework. Developers (scientists) are likely to have different philosophies about ways of tackling the problem. This is not a problem in itself, but it does add to the time it will take to develop a new agreed framework.
- There will be a need to define priors for inputs, and it may be very difficult to find the necessary information and/or to agree on priors, just as with weights for scenarios.
- Ideally there should be some form of testing of new methods, models and software, even if just for the reason of 'good practice'. However, if a Bayesian approach is taken there may be a need for an MCMC approach to integrate over uncertainty, and in this case simulation testing is very difficult.

It is worth briefly commenting on the second dot point above, under disadvantages of a new framework, and the process that might be followed if a new framework is to be developed. An approach where the development (particularly the early stages of such development) is done 'by committee' is unlikely to be productive or successful. In our view, it would be important to establish a process whereby developers are encouraged to bring ideas and preliminary model frameworks to the relevant meetings (SAG or special workshops), then to allow further independent development, before commencing a more collective approach, as was done in the case of the previous MP development.

For both approaches, scenario modelling and new frameworks, there will be a need to carefully consider how the uncertainties in the various historic data sets are quantified. This is probably the most important first task facing the SAG/SC. If the SAG and SC only has the contents of the two reviews on which to base subsequent OM development, conditioning and MP testing, and if there is no further factual information on which to base relative weights for scenarios, or priors for a Bayesian framework, then we (the SAG/SC) are very limited in what can be achieved. If so, we would have to evaluate, and explicitly state, the limitations of (a) the extent of uncertainty that can be built into the conditioned OM and (b) the level of robustness that this will imply for an MP tested under these conditions. This obviously also has important implications for the kind of advice that can be provided in a 'stock assessment' sense.

The question has been asked whether another option would be to ignore all historic data. We do not, however, consider this to be a sensible or helpful option. It seems highly unlikely that the reported catches were not taken, so they can at least be interpreted as lower bounds on total catch. What is currently more of a concern though is the reliability of the historic Japanese longline CPUE series which has played a major role in past stock assessments and the conditioned OM. We note that a start has been made by the CPUE WG to explore this issue further.

2.2 The projection and testing phase of the operating model

The projection or testing phase of the model is the process whereby the operating model is run forward in time, with a feedback management procedure (MP) used to determine catches in each future year. The MP takes simulated data, and in some cases also historic, 'observed' data, from the operating model, establishes a total allowable catch (TAC), which is then applied to the population. This cycle is repeated for, say 30 years, and many replicates are run to reflect uncertainties in underlying quantities (e.g. steepness in the stock-recruit relationship), as well as random variability in quantities such as recruitment. Results are then summarised in order to evaluate the performance of the MP.

This discussion is primarily in the context of the MPs considered so far, and the kinds of input data they have used: historic CPUE, size frequency data and, in some cases, also historic catches. We consider other potential indicators in Section 3 below. The testing of MPs which are based on entirely different input data, has different issues which we do not consider here. Note though, that projections from any conditioned OM, used to test MP performance, will face many of the issues discussed in this section.

The first question that arises, is whether it is still appropriate to consider any management procedures (MPs) that are based on Japanese longline CPUE. It is easy to continue to simulate 'future' CPUE data, but the question is whether, in reality, the procedures required to ensure that the CPUE data used in the MP are reliable and verifiable are likely to be achieved, either in the short or medium term. In addition, changes in the domestic management of the Japanese fleet, and hence potential changes in its operation that might affect the CPUE time series, have been noted (para. 172, SAG7). This should also be considered, when deciding how to proceed.

This then leads to the question whether other data, which provide measures of abundance and/or fishing mortality, should be considered for use in the decision rule (e.g. the aerial survey, tagging). In reality, one may have more confidence in some of these other measures of abundance. Of course how to simulate such data in the future would be a considerable challenge. Other measures of abundance are discussed in more detail in Section 3.

The third question is whether it is still appropriate to test the behaviour of MPs without any implementation error associated with future catches. This would only be appropriate if there is high confidence that future total catches (including non-member catches) would be very close to (say within 1 or 2% of) the total recommended catches generated by the MP. Even then, it may be prudent to include in the software, the ability to model some form of implementation error (*sensu* Francis and Shotton 1997).

Even if it is considered that CPUE-based MPs are still worth considering (note, for example the notion of 'survey' or 'sentinel' catch-effort data discussed in Section 3, below), the re-testing of the MPs that have already been tested during the CCSBT MP development process, will not be straightforward. In addition, there are interesting implications for implementation. In the previous round of development and testing, there was no distinction between the historic data used in the OM for conditioning and the historic data used in the MP. Now there needs to be a distinction and an explicit discussion about the interpretation of 'historic data' used in an MP. We discuss this further below, but it is first worth recalling some of the detail of the candidate MPs that were tested in the recent CCSBT MP work.

The final set of four candidate MPs (see Report of the Fourth Meeting of the MP Workshop, May 2005) use different aspects of the historic data, and are therefore likely to be affected in different ways. Essentially all the MPs are structured in such a way that they use a set of inputs to obtain a single quantity which then determines how the TAC in the previous year (t, say) should be adjusted to obtain a TAC for the following year (t+1, say). In this regard, all the MPs use the "current" catch, which, in testing, was assumed to be equivalent to the "current" TAC. The sets of inputs used by the four MPs differ. Two are based on fitting a Fox stock-production model to catch and CPUE data (CMP_1 and CMP_2). CMP_2 is in fact the one that has been recommended by the SC and adopted by the commission (Report of the Tenth meeting of the SC, September 2005). The other two candidate MPs are based on CPUE, particularly the level and trend in recent CPUE (CMP_3 and CMP_4). All four MPs also use information on the proportion of young fish in the catch or the CPUE.

The two MPs which fit the Fox model, use the full time series of catches and CPUE, and the time series will continue to increase in length as the projection data are added. The other two MPs use changes in recent CPUE (3 years in one case, 10 years in the other case), and the time-series used in the MP will "shift" along, but remain the same length. In other words, the historic data will be "remembered" in the case of the FOX model, but "forgotten" in the other two cases.

The Scenario approach

To illustrate the notion of distinguishing between the historic data in the OM and in an MP, consider the following hypothetical situation. Assume that the 'scenario' approach is being used for conditioning and testing, and that there are five different scenarios (numbered 1 to 5, but with no particular ordering implied nor is this to imply what a likely number of sufficient scenarios would be) for the total catch and CPUE. Also assume that we are considering an MP which requires "total catches" and "CPUE" time series as inputs (e.g. to fit a Fox stock-production model). The first question is which historic data scenario should be used to construct the MP? One could use the Scenario 1 data in the MP and then test its performance with the OM model conditioned to the data from scenario 1, 2, 3, 4 and 5 in turn. It would definitely be inappropriate to only test this MP against the OM conditioned to scenario 1 data, because this would not test robustness. On the other hand, one could also construct an MP based on data from any of the other scenarios, or based on, say, the average catch from some or all scenarios. In fact, the "historic catch series" used in the MP can be seen as MP-specific parameters which are up to the MP developer to determine. The most appropriate way to deal with this, consistent with the MP procedure, would be to provide each MP with the full range of historical catches and alternate CPUE series and it would be contingent on the MP to decide how to use these in formulating its decision rule. Note for example, that in the previous round of testing, the OM contained scenarios based on each of the 5 CPUE series, but the MPs were based on the median CPUE series. In each case it is obviously important to test each MP (with whichever catch or CPUE series) against the full set of scenarios.

Results for such and exercise may form a grid similar to this hypothetical example for a single structure of an MP, but different versions of the MP using different inputs for the 'historic catches':

		OM Conditioned to:					
		Sc1	Sc2	Sc3	Sc4	Sc5	Summary
MP version:	V1. uses catches from Sc1						MP_V1 over all OMs
	V2. uses average catch over all scenarios						MP_V2 over all OMs

	V3. uses catches from Sc3, multiplied by 1.2						MP_V3 over all OMs

Each 'cell' is meant to represent some result, for example, the probability that biomass in 2014 is lower than the biomass in 2004, for each version of the MP, tested under a specific scenario. In this example, the MP structure is the same in each case, but each version uses a different time-series for its historic catch inputs (e.g. MP_V1, MP_V2, etc., down the left hand side). Each version of the MP is tested under the OM conditioned to a specific scenario (e.g. OM_Sc1, OM_Sc2 etc., across the top). These specific results are obviously relevant. The summaries along the right hand side represent the summary outcomes of a particular MP version over all the OM scenarios. (At this point we deliberately avoid the detail of how the summaries are derived from the individual results!).

The reason for going into this level of detail here, is to help illustrate the implications of how we view the uncertainty in the data in MP development and testing, AND how that relates to the subsequent implementation of an MP. In reality, an MP should provide a single outcome, a single 'TAC'. In short, the question is, what do we use as the (so-called) "historic data" in an MP when it gets implemented?

This is where it becomes useful to make a distinction between the "historic data" scenarios in the OM and the "inputs" in the MP. It is important to note that IF the "catch" and "CPUE" quantities in the MP are no longer seen as being 'historic data'[2] (since there is no longer a single set of "historic data"), they can almost be thought of as MP "parameters". In some sense, almost any set of MP inputs that lead to good performance of the MP in future, could be used (in practice, there will obviously be limited sets of 'sensible' or well-performing inputs). This means that (a) there can be more columns than rows (not all OM scenarios have to be turned into MPs), and (b) there can be more rows than columns (some MPs can have 'inputs' which do not really constitute a plausible scenario for the OM) in the grid above. This may sound strange, but the issue is that the inputs to the MP may no longer be "data" as we saw it in the past.

Having said this, there are obvious pitfalls with such a change in view. There will need to be some continuity and compatibility between the assumed 'historic' inputs and the simulated (or expected real) future inputs to avoid potentially strange behaviour in the early years of actual implementation. This is particularly true when an MP works by updating its previous TAC to obtain the next TAC. There should be no doubt that the future inputs to the implemented MP need to be based on reliable, verifiable real data which cannot be treated as "MP parameters".

This discussion suggests that there may now be a need to test several "versions" of an MP, each of which is based on a different set of inputs which may, or may not, coincide with a scenario used to condition the OM. It would be very important to pre-determine the criteria by which MP performance would be evaluated, and to emphasise the limited interpretation of the 'inputs' in the MP. (For example, the 'inputs' may NOT constitute anything like the 'most plausible' catch scenario, or if a particular MP with a particular set of catch inputs performs best, this does not imply that series should, for example, be used in a stock assessment).

It is worth noting that there are other ways of dealing with the uncertainty in historic data as inputs to an MP. One could construct an MP which explicitly uses all the data scenarios as inputs but then combines its outcomes in some way to obtain a single outcome. Different ways of combining outcomes can be evaluated in the testing phase. It is therefore important that such MPs are tested in exactly the same way that they are intended to be used when implemented.

Although this discussion mostly mentioned catches, similar arguments apply to CPUE, but with the additional comment that, for MPs which only use subsets of the CPUE time-series, the issue of which inputs to use for the historic part of a time-series, is likely to only be a problem in the short term, if it is assumed that the future data are obtained from a process that ensures reliable, verifiable data.

[2] Technically the "historic data" in this case used in conditioning are not actual data (i.e. observed quantities) but are outcomes from models of observed quantities – e.g. reported catch, market statistics, etc.).

Two options for treating scenarios in testing MP performance

In the discussion above we deliberately avoided a description of how summary results might be obtained. One option for treating scenarios is to try to integrate over projection results by specifying a weight for each scenario, and combining results according to those weights. Stated another way, the projections are done by sampling from the scenarios according to the specified weights. This approach, however, relies on having relative weights for scenarios, and this is never easy to specify or agree on. Experience from the most recent round of MP testing suggests that the setting of weights for a much expanded set of scenarios will be problematic. In the absence of information on, or a basis for, setting weights, equal weights are often used, but this in itself implies the assumption that all those scenarios are equally likely or plausible. It becomes particularly problematic when results are sensitive to the assumed weights.

The other option for dealing with scenarios is not to integrate results, but to keep them separate and to focus on robustness of an MP under each scenario. This was discussed in Polacheck and Kolody (2003), and at the second meeting of the MP workshop (2003). Under this approach, “acceptable limits” would be agreed for critical performance indicators in relation to management objectives. MPs would then be evaluated to see if they met these levels, and only MPs that satisfied the criteria across all scenarios would be considered robust.

At that stage (MPWS 2 in 2003), the MP working group decided not to adopt the robustness approach; this was most likely due to difficulties in agreeing on robustness criteria, or what was referred to as “acceptable limits” above.

A new framework

In the case of a new framework, the issue of setting weights for scenarios would be replaced by agreeing on priors for inputs and being able to define an appropriate likelihood function. The inputs will not just be the uncertain components of historic data, but will very likely include quantities such as steepness in the stock-recruit relationship. Although there would not be explicit scenarios for historic data, there may be a more elegant way of handling the testing.. However, the construction of inputs to an MP, will still likely be complicated. For example, if the conditioning model uses priors and estimates posteriors for, say total catches, then testing could be done by sampling from the posteriors. In practice it will not, of course, be as simple as this given the need for a coherent time-series and compatibility between different data components. Nonetheless, a full modelling framework may greatly facilitate this exercise. We should add that, at this point, we are raising issues to start discussions; closer consideration of detail may well reveal many technical difficulties[3]. It is not at all clear whether it would be possible to construct proper priors. For example, it may be very difficult to specify priors for the way in which any assumed overcatch should be split between the longline fleets currently defined in the OM. Although the way in which to test MPs for robustness against uncertainty in the historic data may follow more obviously from a new modelling approach, it is unlikely to be much simpler or less time-consuming to run.

Under this approach, it becomes potentially more difficult to see how to deal with the issue of which ‘data series’ to use in an MP, in the abstract. This is particularly true if there is no input data series of catches or CPUE as such, but rather priors or a mixture of, say lower bounds for catches, with priors for additional catches. Once a framework has been developed, this should become clearer. There are, however, general notions which can be considered for developing common-sense approaches. For example, if the conditioned OM implies (or estimates) something like posterior distributions, these could be used to construct an input series for an MP. However, if we again disconnect the MP inputs from the OM inputs (or assumptions about historic data), then there are many other potential ways of constructing such inputs, or simply using MPs which do not rely on those inputs so directly. This does, of course, become more of a problem with regard to CPUE, unless only a short subset of the CPUE time series is used.

Further General comments

[3] It is (sadly) quite likely that we would end up having to be satisfied with an inelegant, hybrid approach because of practical constraints.

In theory one might consider that, if future input data to the MP can be treated as reliable and therefore without (or with only a very moderate level of uncertainty), the historic uncertainty should not affect its performance substantially in the long term. We have already noted that some MPs do not even use much historic data, but our assessment of their performance depends critically on the historic data used in conditioning. In the short term, however, different assumptions about what the recent historic catch or CPUE had been, could have a substantial effect on the performance of some MPs. Also recall that all MPs tested so far, were based on adjusting the previous year's TAC to obtain the subsequent year's TAC (or the next TAC, if set in 3-year blocks), implying another possible short-term effect of uncertainty in historic catch data. Also, given the longevity of SBT such effects are likely to enter the spawning biomass in the medium term. Even if all MPs perform similarly relative to each other under different scenarios for historic data, the actual tuning and 'absolute' performance (for example, in terms of the percentage time that SSB falls below some level) is likely to be very sensitive to assumptions about the historic catches input to the operating model and to the MP. The degree of sensitivity may depend on the structure of the MP.

Before embarking on another testing phase for SBT, it may be prudent to explore whether scientific testing of an MP is in fact currently feasible for SBT, given the uncertainties and biases in the historic data. This could be done by developing a checklist of data, inputs and conditions required in order to conduct scientific testing of MPs. We note that scientists working on management procedures in the context of the IWC (International Whaling Commission) have developed such an approach, after an experience of a case in which the background data. and updates to the background data during the test phase, made it difficult to constrain the universe of scenarios to a manageable fixed set.

3. Indicators for decision rules

In this section we ask what is available as a potential indicator for use in a management procedure (or, to use a more direct term, a 'decision rule') for SBT. In all the previous MPs that have been tested, Japanese longline CPUE was the key indicator, whether used with or without other quantities such as catch and/or size frequency data. Two important aspects that need to be considered when looking for alternative indicators are: (a) what historic data are available to characterise the indicator for testing in projections, and (b) what is the likelihood of getting reliable and verifiable future data for that indicator. In the next set of sections, several potential candidates are considered in turn.

3.1 Longline CPUE

Even before the discovery of the unreported catches, which may have affected the historic CPUE, there were recognised concerns about longline CPUE being used as an indicator. The standardisation of the data has been the subject of a large amount of research, particularly because of concerns about the substantial changes in spatial coverage of the fleet over the past 4 decades. Also, at the time the MP testing started, there was still no agreement on a single standardised series to use, and the median of five series was used for testing purposes. It is worth recalling that, although the series all look quite similar visually, the MP testing revealed sufficient differences in OM behaviour, and hence MP performance, between series, that the five series were retained as an 'uncertainty axis' in the conditioning phase of the model (see MP Working Group reports for more information). The projections generated the 'median' CPUE, a single series, for use by the decision rules, though quite how this was to work in practice (i.e. when the MP was implemented) had not been thoroughly discussed or agreed.

The unreported longline catches that were discussed at length at the SAG/SC meetings in 2006, may have affected the CPUE, but it is still unclear whether this is the case, and if so, to what extent. The CPUE working group has started to consider the implications. At best, it may be possible to develop a scientifically justifiable single series based on data, if the uncertainties about the effects of unreported catch on the historic data can be resolved, though this seems unlikely. If not, and there are many plausible CPUE series, then this uncertainty has to be retained in the testing phase of MPs (as discussed above). We note that even though it would be much more convenient to have a single

CPUE series, it would be unscientific to ‘force’ the choice of a single series. This is particularly true for CPUE as an input to the OM for conditioning. With regard to which historic CPUE series is input to the MP, it may be less crucial in the medium term, though it may prove to be important in the short term; refer to the discussion in section 2.2 above.

We consider that there is very little point in doing further testing of MPs based on Japanese (or any other) longline CPUE, unless there is a process in place (or being developed, with a high level of confidence that it will be put in place) to ensure that logbook catch and effort data are collected in such a way that those data are reliable and verifiable. Even if this can be done, the concerns always associated with CPUE as indicators of abundance will persist. For example changes in the spatial and/or temporal distribution of the fleet can make standardisation impossible; effort creep may be hard to standardise for and impossible to detect. Further, the change from an olympic to a per boat quota system is likely to make it more difficult to interpret CPUE data.

In general, the dangers of relying on CPUE are well known from experiences in other fisheries and we have many concerns about the use of CPUE from commercial fishing as the sole indicator in a decision rule, or stock assessment. More confidence could possibly be associated with a CPUE indicator based on protocols akin to a ‘survey’, for example the ‘industry-based, scientifically-designed CPUE sampling’ discussed at SAG7 (see paragraph 170 of SAG7 Report, 2006). This is again considered in section 3.4 below.

3.2 Tagging data

Although the historic tagging data (from the 1990s) were included in the conditioning of the OM, the more recent tagging data have not yet been incorporated into the OM, and none of the original set of MPs tested so far used the tagging data. In principle, it would be possible to use tagging data, or quantities estimated from tag returns, as inputs to a decision rule or MP. The problem is that reporting rates are required to reliably interpreted tag returns as indicators of fishing mortality or stock size, and ideally, this should be done in conjunction with catch at age data. Also, spatial heterogeneity (e.g. incomplete mixing) combined with a lack of adequate spatial distribution in the release of tags can bias any estimates from tagging experiments. Without additional pieces of information and a more comprehensive experimental design for the release of tags, the uncertainty about the interpretation of the data (i.e. in estimates of quantities such as current harvest rate) is greatly magnified.

In terms of testing, there may be some difficulties. Analyses of recent tag return data (from the 2000’s) suggest that fish movement patterns may have changed between the 1980s and the 2000’s, and that these patterns differ by age class (see e.g. CCSBT-ESC/0609/15 and CCSBT-ESC/0709/19). If this is the case, it may be difficult to characterise the nature of the tagging data in order to simulate future release and recovery data unless a more comprehensive design were used for the release of tags so that such spatial effects could be accounted for. (Note that if such changes have been occurring they would also increase the uncertainty and interpretation of CPUE abundance indices in which there is large areas and time periods without any effort).

As in the case of CPUE data as indicators in an MP, there would only be an advantage in testing MPs which use tagging data, IF there are processes in place which are likely to lead to reliable tag returns estimates of reporting rates, and ideally of estimates of catch at age. It is also insufficient to only obtain reporting rates for the surface fishery, because this severely limits the information content of the tagging data. Many of the problems of associated with reporting rates potentially could be overcome with the use of automated methods for tag detection, or by using genetic “tagging”. In addition, a regular tagging program would have to be established, and the tag releases would need to follow a design which allows for much better spatial coverage than was obtained in the most recent SRP tagging programs.

3.3 Aerial survey

There is now a reasonable time-series of a ‘standard’ (i.e. comparable between years) aerial survey index (1993-2000 and 2005-2007). In theory, the index can quite easily be incorporated into the current OM framework. This was illustrated in the assessment work conducted in 2004 (Basson et al. 2004). There are, however, three important points to note. First, the index only represents the relative abundance of juveniles (mostly ages 2-4). Second, the index only represents the relative abundance of juveniles that are in the Great Australian Bight in each year. If this proportion is close to constant over time, then this should not be a problem. If this proportion varies greatly or, more importantly changes with autocorrelated trends, then the use of the index, particularly on its own, in an MP could be risky. The apparent changes in spatial dynamics of juveniles based on conventional and archival tag results is a concern in this regard. Third, there are still concerns about the differences between different spotters, which cannot be fully resolved retrospectively.

It is the issue of the partial coverage of the juvenile stock that implies that the aerial survey index may be risky if used as an overall index of juvenile abundance. In theory, this statement could be tested in a simulation context, but such testing is best done on the basis of some knowledge of the broad scale movement dynamics of juveniles. Research on juvenile spatial dynamics is underway (see e.g. CCSBT-ESC/0709/20), but although most of the field work have been completed, the data analyses have only recently started. Results are not yet available to be incorporated into a simulation framework.

Having noted this, we do however also emphasise that this is about the only fishery independent index of Southern Bluefin Tuna that is currently available to us, and in this regard, we consider it to still be an important candidate. It may, for example, be worth thinking about other ways of using the index, rather than just as a proxy for overall juvenile abundance. For example, can it be used with concepts of, say, ‘escapement’ as a management tool?

Although we are discussing the different data sources as if they would be the sole component (or indicator) in an MP, this is obviously not necessary, and may well not be the best approach. For example, many of the MPs that were tested used the CPUE data together with some size frequency data and, in some cases, also the total catches. It would be possible to construct an MP that used the aerial survey index together with other data. In this regard, however, we note that combining indicators within a model framework (such as a simple ‘assessment’) has many advantages over combining indicators in ad-hoc ways.

3.4 Other options

There are two other options for obtaining estimated quantities that may have merit in future in spite of the fact that there are currently no historical data of this kind.

The first is the direct estimation of spawning stock abundance based on the genetic ‘close kin’ approach (i.e. sampling adults and juveniles and using genetics to identify and count parent-offspring pairs). The approach is discussed in detail in CCSBT-ESC/0709/18, and we note that data collection of adults through the Indonesian size and age-frequency sampling program, and of juveniles through sampling at processing plants in Port Lincoln, have already started. Although this work is in a very early stage of development, we consider this to be potentially a very powerful tool for obtaining another fishery independent estimate of abundance, in this case of the adult stock. In theory, such an estimate can be incorporated into an MP, though it is admittedly still much too early to even consider any kind of testing of such an MP. However, it is also of interest to note that the cost of providing “routine estimates” is likely to be quite economical, relative to alternatives, once the initial development work has been completed.

The second is the notion of ‘survey’ or ‘sentinel’ catch-effort data. This was briefly mentioned above under ‘longline CPUE’. The notion is simple. Commercial longline vessels, possibly even a subset of the total fleet, would be required to do a certain number of very specific longline sets each year. For broader spatial coverage, this could involve all longline fleets. The actual setting and gear would be predefined so that it is comparable and compatible between all vessels and over time. The time (e.g. week or month) and location (e.g. within a specified grid square) would also be specified. This obviously needs to be done in a practical manner, but it would allow for persistent coverage over certain areas and months. The data from such ‘research sets’ should be observed and recorded by an independent observer, just as if it were part of a ‘scientific’ survey. The catch and effort data should then be much more amenable to standardisation and interpretation as an index of abundance (though it is still, of course, likely to suffer from some problems – e.g. spatial and temporal coverage may never be extensive). The main advantage is that this can potentially be instigated very soon and it has been demonstrated to be feasible and provide useful data in other international fisheries (e.g. Exploratory longline fisheries in the Antarctic). Work would need to be done to design such a 'survey', but the scientific aspect of this should be feasible within a relative short time.

It is, however, important to note that such an approach is still likely to suffer from some of the problems associated with commercial CPUE. In particular, it is unlikely that it would be practical to obtain, and maintain, the level of spatial coverage and the level of sampling intensity that would provide an index of abundance with acceptably low variance and with a low probability of bias because of the practical and logistic constraints.

4. Conclusions

The increased uncertainties in the historic catches, CPUE and size frequency data have important implications for the implementation of an MP in the short term. The agreed MP uses historic catch and CPUE data and cannot be used without revision because, (a) we can no longer rely on results from testing based on the reported data, and (b) we can no longer rely on the MP tuned on the basis of these data. There are also implications for the medium term, because the revision of the agreed MP, or the development of any new interim, or long term MP, is unlikely to be a simple task.

At this point we emphasise our strong view, that the goal to aim for, should still be an agreed management procedure, with all its components. We consider the MP approach will have the highest likelihood of achieving the goals set out by the SAG7 (paragraph 169) and noted by the SC11, that "data collection and MP development in the next 5-10 years should be prioritized to focus on rebuilding the stock to a point where the biological and economic risk associated with the current high depletion and high fishing mortality is greatly reduced". While we are also optimistic that progress can be made in this regard, the following strong caveats apply.

Progress can only be made, and further development and testing of MPs are only worth doing, if there are processes in place to ensure that the data used in the implemented MP are reliable and verified. If this cannot be achieved, then there would be no basis for determining what the likely performance of an MP would be.

Progress towards further scientific testing of MPs, to ensure robustness to historic uncertainties and biases in data, can also only be made if the uncertainties can be characterised. At the very least, the plausible ranges need to be specified. There is therefore an urgent need to discuss how, and whether, the uncertainties can be appropriately characterised. This is irrespective of decisions about what to use as indicators in an MP in the short (or long) term, and irrespective of the approach we choose for dealing with uncertain historic data in the conditioned operating (or assessment) model. If a scientifically justifiable set of scenarios, or description of uncertainty, cannot be obtained, the chances of developing a robust MP and of providing robust advice will be severely limited. It may only be possible to give indications of the implications of different scenarios, without having any confidence about whether they represent 'best' or 'worst' case scenarios.

If a scientifically justifiable set of scenarios, or description of uncertainty, can be obtained, it is much more likely that a robust MP could be developed, and robust advice be provided. We are, however, concerned that the substantial uncertainties in the historic data would now make it even more difficult to provide reliable 'absolute' advice on the performance of MPs. Most importantly though, is the need for two processes to run in parallel: i) further MP work and ii) progress on actual data collection procedures. It goes without saying that there is little point in testing an MP based on, say, tagging data, if there is little likelihood of obtaining the necessary data, at the necessary level of reliability, once the MP has been implemented.

As noted above, the operating model itself will need to be reviewed in the light of the uncertainty in historic data, irrespective of whether the original MPs are re-tested or whether new MPs are developed. There are two potential approaches: a 'scenario modelling' approach, which will be relatively fast to implement, or a new framework that would involve the development of a new conditioned OM. Such a new OM need not be more complicated than the existing OM; it could in fact be less complicated. There are advantages and disadvantages to both approaches, and these should be considered when discussing how to proceed.

With regard to the design of MPs, it is fair to question whether commercial CPUE is an appropriate or reliable indicator to use in an MP. Although there are some advantages in developing a 'survey' or 'sentinel' catch-effort data scheme, there are still likely to be concerns about the spatial coverage and the variances achievable from such a series. Even if such a program can be started soon, it would still take at least three years before it could even begin to be used as a time series. Currently, the only fisheries independent index that is available, is the Line Transect Aerial survey index, but since it is unlikely to represent all juveniles in the population, it needs to be used with caution. The most recent tag return data suffer from issues associated with reporting rates and concerns about the limited spatial coverage of tag releases.

We should, however, also note that, even if commercial CPUE are not used in an MP, reliable fisheries data, such as catch, effort, size/age frequency, are still required for the characterisation, and hence understanding of, the fishery dynamics, as well as for stock assessment purposes.

On the whole, we consider that an approach which uses several indicators together may be the most promising, particularly in the short term. Candidate components could be a 'survey' catch-effort series, the aerial survey and some form of tagging, but with special attention paid to the problems encountered in the most recent CCSBT conventional tagging program. In all cases where fishery-based information are required it is, as mentioned before, essential to put appropriate processes in place to obtain reliable and independently verifiable data. In the case of fisheries independent information, it is assumed that the scientific program would be designed and conducted in a way that would obtain reliable data.

In the medium to longer term, we consider that the proposed 'close-kin' genetics approach has great potential for becoming a suitable fishery independent index of spawning biomass which could be used in an MP, as well as in stock assessments. In addition, the genetic fingerprinting used in that approach could serve a dual purpose in a genetic tagging program.

The ideas and suggestions presented here are intended as starting points for discussion in the SAG and SC about ways forward with regard to a Management Procedure for SBT. Although there is potentially much work ahead to adequately characterise the uncertainty in the historic data, and then to develop and test a new MP, and although we foresee many technical difficulties, we consider that there are ways forward. Even if it proves impossible to resolve the uncertainties in the historic data, there is now the opportunity and a stronger need than ever before to make progress towards setting up processes and systems which would deliver reliable data on which an MP, stock assessment, and scientific advice can be based.

References

Basson, M., D. Kolody, T. Polacheck, A. Preece, J. Hartog. 2004. Assessment and projections of SBT stock assessment an summary of agreed stock status reference points. CCSBT-ESC/0409/23.

Francis, R. I. C. C. & Shotton, R. (1997). "Risk" in fisheries management: A review. Canadian Journal of Fisheries and Aquatic Sciences, 54, 1699-715.

Hammond, T.R. and V.M. Trenkel. 2005. Censored catch data in fisheries stock assessment. ICES Journal of Marine Science, 62: 1118-1130.

Polacheck, T. and D.Kolody. 2003. Synthesising Performance of Candidate Management Procedures across different operating model scenarios. CCSBT-MP/0304/08.