Southern bluefin tuna 1998 Peer Review Panel

Executive Summary

The data available on southern bluefin tuna and the techniques used by scientists to analyse these data are comparable to those utilised elsewhere in the world. Although we have some concerns about the data and models used for assessing southern bluefin tuna, our major concern is with the process and group dynamics that lead to the report provided to the Commission for the Conservation of Southern Bluefin Tuna (CSBT) and the lack of agreement on what advice should be provided.

The process we have observed in the Stock Assessment Group (SAG) and the Scientific Committee (SC) cannot be described as scientifically neutral. In other scientific processes we are familiar with, scientists participate firstly as individuals, not as national representatives. There is no formal role for heads of delegations in the technical part of the assessment meetings, and the report is prepared and adopted collectively without status or party distinction. In these fora, it would be seen as totally inappropriate for heads of delegations to agree on the scientific report before it is discussed and adopted in plenary session.

We believe that these problems are due in part to history, and in part to inappropriate responsibilities given to, or assumed by scientists. For the fishery management process to function properly and to be transparent, there needs to be a clear separation between science and management. Separating the process into a technical part, the SAG, and an advisory part, the SC, may be a first step in that direction, but it does not appear to be sufficient under present conditions. As an interim measure, the CCSBT should constitute a facilitating panel of three to five independent scientists to guide the SAG/SC process towards consensus advice. If the SAG/SC cannot reach consensus, the panel itself would provide the advice to the CCSBT. Three years should be sufficient for the SAG/SC to get acquainted with a consensus building process.

The combined catch of member and non-member countries is currently not sustainable under some model formulations considered. Effective means of monitoring and controlling all catches should be sought. In the mean time, a precautionary approach would be for member countries to set aside a portion of the TAC to account for the catches by non-member countries.

All analyses indicate a substantial reduction in spawning stock biomass over time and considerably lower recruitment in recent years than 20 years ago. Also, many analyses suggest that fishing mortality (F) on the older ages is now lower than it was 15 years ago, but such a decrease is not detected for the younger ages. It would therefore seem beneficial to reduce F on the younger ages in order to allow a greater proportion of the currently low recruitment to survive to the parental stock.

With respect to data, although the available catch at age appears relatively good to reconstruct past stock trajectories[1], there is a paucity of consistent and reliable stock size indices. Therefore, there are considerable uncertainties about current stock size, and, as a corollary, it is difficult to assess the effect of recent and/or future management measures. There is therefore an urgent need to develop means of assessing the response of the stock to management actions.

The uncertainties are considerable in fisheries science, and it is very rare that there is a single ‘correct’ interpretation of the data, with all the others being necessarily ‘incorrect’. The SAG/SC provides a curious treatment of the uncertainties, and it is probably one of the rare cases in fisheries science where some uncertainties are overstated. This is linked to a lack of interpretation of assessment results and we believe that the CCSBT would be better served by receiving a smaller, but more clearly delineated, set of possible scenarios.

[1] However, as indicated in the body of the text, the current practice of grouping individuals age 12 and older into a plus group makes this process more complicated and uncertain than is necessary.

A section on recommendations is provided below. The main body of the report follows after the recommendations.

Recommendations:

1. Advisory process

The CCSBT was implementing its restructured advisory process of a Stock Assessment Group and a Scientific Committee for the first time in 1998. The separation of the process into an observation and analysis phase and an interpretation/advisory phase works well in other fishery arenas and, based on discussions with meeting participants, seems to have provided some improvement to the CCSBT process compared with previous years. However, further guidance on the roles the two committees is needed. For example, does the Scientific Committee interpret the science from the Stock Assessment Group, do they formulate possible management scenarios for consideration, and do they provide management advice? Should they be called a Scientific Committee or are they a management advisory group? A step forward has been taken in defining these two groups, and further progress can be made by defining further their roles.

i. Chairing of meetings:

The current practice of selecting the chair for a meeting from the host country has been in effect since the beginning of the CCSBT and it appears to be functioning satisfactorily as far as the administrative preparation of meetings is concerned. However, there would be considerable benefits in choosing a chairperson for several consecutive meetings for the scientific sessions of both the SAG and the SC. The chair must be chosen based on the personal suitability of the candidate, and unrelated to the country of origin, laboratory or affiliation. This is standard practice in similar organisations where the chairs are normally elected for two to four meetings, and their role is perceived as pivotal in achieving the objectives of the commissions. The CCSBT should adopt this approach for the SAG/SC process.

The responsibilities of the elected chairs would include circulating a draft agenda, delegating rapporteurs, making other preparatory arrangements, and guiding the meeting to the production of a consensus report to the CCSBT. The election of chairs can be extended to groups reporting to the SAG/SC and if *ad hoc* groups are created, for one or more meetings on specific topics, a chair for the subgroup should similarly be chosen until their work has been successfully completed. The chair of the subgroup would have the responsibility of reporting the results either to the SAG or to the SC. It would be desirable that the chair be able to understand both official languages of the CCSBT, if not speak it.

An effective, efficient, knowledgeable independent chair is not a guarantee that a meeting will successfully reach a consensus agreement on the report to be presented to the CCSBT; it is also necessary for the participants to have the desire to reach that goal. We assume that CCSBT Commissioners would like to receive a consensus report from the SAG/SC process, and therefore, they should make it known to their scientific staff.

ii. Meeting agenda:

Setting a structured timeline for presentation and discussion of papers is necessary for a meeting to run smoothly. Having the chair set the agenda with approximate times associated for each agenda item should help further. Once this is done, priority items can be highlighted and dealt with. The chair’s suggestion, at the 1998 SAG meeting, that detailed technical questions be clarified outside the meeting, helped move things along at several critical points. However, the SAG meeting spent too much of its time on relatively minor issues, on previous miscommunications, and on technical incompatibilities. This prevented discussion on more important and pressing issues such as projections, data collection, data quality control, and establishment of joint research priorities. Discussion on what future research is needed to resolve uncertainties and diverging interpretation should take place earlier in the meeting and be allocated sufficient time because it is unlikely that analyses and discussion of past data will resolve these uncertainties.

iii. Report to the Commission:

The Commission and the Scientific Committee have taken a significant step in agreeing on a format for the report submitted by the Scientific Committee to the Commission. It is unfortunate, however, that the current content of the report is not as informative and useful as it could be. Part of this stems from group dynamics and the inability to reach consensus as discussed elsewhere in this report, but part also stems from the difficulty that exists in communicating uncertainty in a meaningful and useful way. When communicating uncertainties, scientists should include the possible consequences of various decisions based on what is understood about the system and based on what management objectives have been outlined by the Commission. Decision tables, control rules, and biological reference points would all be useful tools to employ. This process should be allowed to evolve within the SBT arena as it has, and continues to do so in other arenas.

iv. Peer review process:

A peer review process can have several objectives: obtaining independent scientific advice for management, achieving an independent quality control of the products and procedures, facilitating the formation of consensus, incorporating new ideas and methods, etc. Depending on the objective(s) to be reached, different structures and formats would be appropriate.

If the intent of the Commission is to periodically obtain independent quality control and if the review panel is NOT to be considered part of the assessment process, it would be preferable to conduct the review outside of the normal assessment cycle. The review panel would then be able to ask questions of the stock assessment group as a whole and the focus of the meeting would be on the review and not on the completion of the assessment. This type of peer review is employed by many fisheries agencies every five years or so as a periodic check on their process, procedures and products.

If the objective is to obtain independent scientific advice for fishery management decision making, the Commission itself could employ directly its own stock assessment staff as done by both the Inter-American Tropical Tuna Commission (IATTC) and the International Pacific Halibut Commission (IPHC). This option allows the science to be conducted in an independent and objective framework following the needs and priorities as directly defined by the Commission, and not mediated through other national agencies. A variation of this option would be to delegate the peer review process of the assessments to other established bodies, either involved in tuna research (ICCAT, IOTC, IATTC) or not (ICES), or even to contract the assessment to an independent set of stock assessment scientists. This might alleviate some of the polarisation problems currently acutely felt by the Commission, but there is no *a priori* guarantee that there would be wider involvement of other parties in these processes. Issues of continuity and quality control, with potentially different assessment teams conducting the assessment each year, would need to be addressed under this scenario, but several fisheries, specifically in New Zealand, are being assessed in this manner.

If the objective is to get new ideas and keep the scientific process up to date and scientifically on track, then bringing external scientists into the SAG/SC assessment process, either as independent participants hired either by the CCSBT or by industry, should be considered. There was an extended debate during the 1998 meetings on the need and criteria for external scientists. External scientists can be brought in to provide expertise on specific issues, but they can and should also have a larger role as facilitators, as scientists with independent point of views, or as chairs.

The role of the 1998 Review Panel was to observe the deliberations of the SAG and the SC. In that context, the Panel remained observers of the process, asking questions of clarification outside the meeting, in order to review both the assessment and the process used to produce it. In principle, this provides the Commission with an independent viewpoint on the process, a critique of the process, and perhaps an interjection of new ideas that will help improve the process. The Commission might consider different formats and objectives under which a review panel might operate.

In a situation where the participating parties cannot reach consensus on the assessment, the review panel could be involved directly in the assessment process. A future review panel could therefore act as arbitrator/conciliator of the analyses presented. The USA National Marine Fisheries Service (NMFS) has recently adopted this approach on the Pacific Coast of the USA. Stock Assessment Teams present their assessment (generally one assessment per stock is presented, but two per stock have recently been submitted, as is the case for southern bluefin tuna) to one of the Stock Assessment Review Panels for its consideration. The Panel may ask that further or different analyses be conducted during the meeting. When possible, the assessment team(s) produce a single stock summary report and the Panel makes an advisory report based on the joint stock summary report. However, if the differences of opinion are such that the assessment teams cannot agree on joint stock summary, then the Review Panel would make its advisory report taking into account all available reports.

If the current situation where the CCSBT does not get a consensus view from the SAG/SC process is unsatisfactory, and if the CCSBT finds it desirable to obtain a scientific consensus view, we recommend that a modification of the panel process developed on the West Coast of the USA described above be implemented in the CCSBT for a fixed three year time period. The panel should number at least three members and no more than five. The main objectives would be to help scientists reach consensus, develop a standard informative stock status report for the Commission and produce an advisory report. The panel would have decision powers on the stock status summary and advisory reports.

2. Translation and interpretation:

Japanese and English are the two official languages of the CCSBT and both languages should continue to be used in the SAG/SC process. The 'sequential' translation[2] currently used is not an efficient use of time and it is not conducive to the rapid resolution of issues. Simultaneous interpretation with appropriate technical support should therefore be implemented. At first, the costs may seem considerable, but the gain in efficiencies should clearly outweigh the increased costs.

Alternatively, or as a complementing measure, scientists could be trained in the other official language of the Commission. This would be a worthwhile investment, as most scientists will be involved in the process for several years.

3. Data exchange:

Data exchange should take place through the Secretariat or through a neutral party to avoid the frustrations that are being expressed when one side depends on the data from another and does not receive it in time. This may not improve the situation immediately in terms of data availability, but it might nevertheless improve the working climate between member countries because the complaints would be directed either at the Secretariat or by the Secretariat at the parties, rather than from one party to the other.

On a related subject, clarification of catches, landings, product forms, etc. to ensure that all southern bluefin tuna killed are accounted for, should be handled before the meeting starts, perhaps through the Secretariat of the Commission or directly between individuals involved. Alternately, these questions should be resolved by the national agencies responsible for the collection of statistics within each country.

4. Documentation:

For most scientific meetings, whether advisory in nature or otherwise, a deadline is set for the submission of titles to an upcoming meeting. For large international symposia, the deadline is often several months ahead of the conference while for advisory meetings in support of fisheries management activities, the deadline is closer to the meeting. Normal operating procedures dictate that papers that are received by the organising secretariat sufficiently ahead of time, typically one month, will be copied by the organisers while the authors of those that are not received in time

[2] The current approach is that the speaker enunciates a portion of his/her intervention, the interpreter translates, then the speaker enunciates another portion, followed by the interpreter's translation.

must bring sufficient copies for distribution. Generally, only papers whose titles have been submitted within the deadlines will be considered. Similar rules exist for the Stock Assessment Group (SAG) and the Scientific Committee (SC), but they are not strictly applied. Although this may appear of relatively minor importance, the *ad hoc* treatment of documentation results in lost time, possibilities for differing treatment of papers, and is therefore having the potential to create unnecessary tensions between the parties. Existing rules should therefore be strictly applied, although the chair of a meeting should be allowed to request that a working paper be prepared and submitted during a meeting.

Key documents should be made available in a series clearly identified with the CCSBT, either in a traditional printed publication and/or electronic/web based.

More than 40 working papers were either tabled or presented at the 1998 SAG/SC, a sign of scientific vitality. However, there is room for reducing the number of WP without compromising the information made available to the SAG/SC. In particular, it would be desirable to combine into a single national working paper the data and analyses necessary to repeat the assessment: total yearly catches by area, gear, country; weights-at-age, catch-at-age, standardised CPUE, stock size indices (including tagging), etc. At present, it is necessary to consult several documents to assemble the various pieces of the assessment.

5. Frequency of assessments:

Given the low resolution of the indices of stock size currently available to calibrate the assessment, and considering that it is unlikely that one more year of data will change the situation, the CCSBT should consider not asking for an updated assessment in 1999. There are sufficient methodological issues to be resolved and agreed to before an updated assessment is produced. Instead, a stock assessment workshop could be held to agree on a how to stabilise the main parameters of the assessment for the next 3-5 years, through an CCSBT sponsored workshop, involving experienced scientists from outside the normal CCSBT stock assessment sessions.

6. Retrospective analysis:

Two types of retrospective analyses are recognised in fisheries science, and while they are related, they typically represent different parts of the assessment process.

The first is a model-based retrospective where given the current data and model formulation, the assessment model is run under data sets which have been reduced progressively by one year's worth of data at a time. Thus the model's performance is examined in retrospect to see what biases become evident as more data are added. This type of retrospective analysis is important as a diagnostic device and should be conducted annually to monitor the behaviour of the assessment model.

The second is a process-based retrospective where the results presented in previous stock assessment documents are compared to examine the uncertainty in the process as a whole and to provide some perspective on how management measures based on previous assessments relate to those currently in practice. This type of retrospective should be updated periodically, and specifically when there are significant changes in the assessment or management procedures.

7. Technical recommendations

i. Data collection practices in non-member countries could be expanded and improved through CCSBT sponsored mutual assistance programs, and the southern bluefin tuna stock assessment group should assist in their design and implementation.

ii. Reliable indices of stock size or of exploitation are necessary to obtain reliable stock assessments. Indices of stock size for recruits, juveniles, and adults are therefore necessary. Mechanisms such as acoustic and aerial survey for monitoring stock size should be further developed and implemented.
iii. Catch at age should be calculated directly from the length frequencies for the first few ages where the modes are clearly identifiable.
iv. Direct age determination of the age of fish in the commercial catch and in the surveys should be continued and encouraged.
v. The plus group should not be tuned. There is no disadvantage to extending the age distribution in the catch at age to at least age 20. Should the SAG/SC decide not to, or be unable to extend the age range for the next assessment, they should base the estimate of the plus group in the terminal year on the relationship between the F on the last true age and the plus group in that year, and backcalculate the plus group for previous years based on that estimate.
vi. The replacement yield component of the assessment model should no longer be used.
vii. The various standardisations of the CPUE series show globally the same trends for 1969 to 1992-93 within ages or age-groupings. The approaches reduce to two with the Habitat and Variable Square, on one hand, producing virtually identical trends for all practical purposes, and the Geostatistical and Constant Squares also showing virtually identical relative trends. The SAG should therefore reduce the number of standardisation approaches considered to no more than two options and use these two approaches consistently for a few years.
viii. Standard goodness of fit criteria should be used in model selection.

Other recommendations can be found in the body of the report.

1 Introduction

The Commission for the Conservation of Southern Bluefin Tuna (CCSBT) formally began operation in 1994. The Commission's Scientific Committee (SC) can be considered as the successor of the Committee of Trilateral Scientists from Australia, Japan and New Zealand that had been conducting analyses and co-ordinating research on southern bluefin tuna (SBT) from 1982 to 1994, including yearly detailed assessments.

Recently, the CCSBT has developed a new process for the assessment and formulation of advice and this process was approved at a Scientific Process Workshop held in Sydney, May 14-16, 1997. A main feature of the new scientific process is that the technical part of the assessment is separated in time (and possibly space) from the advisory part of the process. Therefore, different people can be involved in each part of the process, making it theoretically possible to get the best use of time.

The CCSBT convened a Review Panel of independent stock assessment and scientific fishery advisors to evaluate the process and the methods. The Review Panel consisted of Dr. Syoiti Tanaka, adviser to the Institute of Cetacean Research, Dr. Patrick Sullivan, Cornell University, and Jean-Jacques Maguire, a private consultant, currently chairing the Advisory Committee on Fisheries Management of the International Council for the Exploration of the Sea. Brief biographies of the Panel members are provided in Annex 1.

Terms of Reference: The Panel was requested to "*undertake a review and provide a report to the Commission on the quality of the scientific analyses and methods being used by the Scientific Committee, to assist in seeking a greater level of consensus in the stock assessment process, which is to include a review of:*

1. *Existing data used in stock assessment;*
2. *Availability and necessity of data to be used in the stock assessment;*
3. *Hypotheses and structure of assessment model;*
4. *Quality and appropriateness of tuning indices;*
5. *Method and hypotheses to standardise indices;*
6. *Biological parameters used in assessment;*
7. *Sets of weightings assigned to uncertainties;*
8. *Hypotheses and structure of models used in projections;*
9. *Methods treating uncertainties in models;*
10. *Process to evaluate calculations and computer codes;*
11. *Process to incorporate new techniques and/or new information;*
12. *Process to review newly incorporated information and;*
13. *Quality and format of the Report to the Commission* "

These terms of reference contain elements of technical review as well as review of the process. Given the mode of operation chosen, that is, assist at the entirety of the Stock Assessment Group and Scientific Committee sessions, we understood that our mandate included evaluating and advising on the overall process used by the SC and SAG.

2 Existing data used in stock assessment

The diversity and quantity of data used in the southern bluefin tuna assessment is average compared with other stock/fisheries. The assessment uses catch at age that is available from 1969 to 1997, but some information on the size composition of the catch is available since 1951, longer than for several other stocks. However, few indices of stock size, either from commercial catch per unit effort (CPUE) or independent of the fishery are available. A number of factors, including technological innovations, changes in fishing strategies, changes in management measures, and changes in fish distribution or species composition, complicate the interpretation of the CPUE data. Monitoring and correcting for the influence of these factors is an ongoing process that must be dealt with at a technical level, but corrections are not always possible, and this should be recognised at the advisory stage. Such problems exist in all fisheries assessments using fishery-dependent stock size indices. We outline here several of these issues as they relate to SBT, and suggest some methods to deal with them.

2.1 Accounting for all SBT landings

The wide-ranging nature of the southern bluefin tuna and its fisheries means that landings are widely dispersed and made under diverse conditions. Therefore, accounting for all the landings of SBT is not a straightforward undertaking. The magnitude of the landings of southern bluefin tuna involved is not always clear, and as a result, there were relatively protracted discussions to ascertain if all the catches have been accounted for, if double counting is happening, or if landings labelled tuna are in fact southern bluefin tuna. Such technical clarifications, while important, should be conducted outside of the meeting so that a brief and accurate report can be supplied, thus leaving time available to discuss issues that are more either more substantive or involve a larger number of participants.

Catches by non-member countries, because by definition they are not controlled by member countries, pose a critical problem for assessment, management, and conservation of southern bluefin tuna. The combined catch of member and non-member countries is currently not sustainable under some model formulations considered. Effective means of monitoring and controlling all catches should be sought. Encouraging non-member countries currently catching southern bluefin tuna to become members of CCSBT may facilitate achieving this objective. Furthermore, data collection practices in these countries could be expanded and improved through CCSBT-sponsored mutual assistance programs. If such programs are considered, the SBT stock assessment group should assist in their design and implementation.

In the mean time, a precautionary approach would be for member countries to set aside a portion of the TAC to account for the catches by non-member countries.

2.2 Catch at age

Catch at age is currently calculated from length frequencies using a growth equation, based on tagging experiments conducted in earlier years. Available length frequencies suggest that calculating catch at age for the first few ages (probably 0 to at least 3 and perhaps age 5) directly from the modes in the length frequencies rather than through the growth equation would improve the reliability of the age composition of the catch for those ages. This could resolve the poor consistency of catch at age at younger ages which may be due to the cohort slicing method cutting through modes which in fact would correspond to a single year-class.

The application of cohort slicing, however, appears to result in good year-class consistency for older ages, the correlation coefficient between the catch at age 18 and that at age 19 for the same year classes is r=0.80 for the period 1969 to 1997. This is probably an overestimate of the real consistency, and an artefact of the cohort slicing method. But it nevertheless suggests that compared with the current approach, there are no disadvantages, from the point of view of year-class size estimation, in extending the modelled catch at age past the current 12+ group. Extending the catch at age matrix past age 12 removes the need to make an assumption about either the F on the plus group or the relationship between F on age 11 and 12+ in an age range where the change in F sometimes appears to be rapid between successive ages. Extending the catch at age would also account more easily for the recently observed broad age composition observed in the Indonesian catches on the spawning grounds.

In the future, the feasibility of deriving the age composition of the catch by direct ageing of a random sample of the catch rather than using age-length keys should be investigated. Co-operation with scientists from non-member countries, where much of this age sampling currently takes place, should be maintained and should be expanded immediately to countries not currently included in order to obtain some information from those fisheries.

Gathering age information requires co-operation from the fishing industry and will involve scientific training and co-operation between member countries. The cost and effort associated with gathering age information will result in greater clarification of issues regarding year-class strength, survivorship, growth, production, and spawning stock potential. Using modal analyses to separate the catch at age for younger ages, and collecting direct ageing information for older ages will likely show considerably more year-class size variability because the cohort slicing approach has a tendency of smoothing neighbouring year-classes.

The method agreed in 1994 to estimate catch at age requires that when there are less than 200 fish lengths measurements for a given catch, substitution should take place to the nearest adequate sample in time, gear, and space. This approach has been applied even when the total catch is less than 200 fish and all the fish have been measured. Substitution is not appropriate for the purpose of calculating removals at age when all the fish caught have been measured. Whether the catch at age thus calculated is the most appropriate to be used in deriving indices of stock size is a different matter. Using the actual catches, rather than making a substitution, may result in catches for some 5X5 squares being different, but it is unlikely to make large differences in the overall catch-at-age estimates. Nevertheless the appropriate method should be used.

2.3 Handling and availability of catch statistics

Discussion on catch statistics, length distribution substitution, and catch-at-age calculations should take place and be resolved in advance of the Stock Assessment Group meeting. Similarly, the details of national procedures for handling such basic data should be clearly documented in advance of the meeting.

Modern computer science allows a database to be created without the data having to physically reside in a central location. A common database would provide a single source for all of the baseline data, so there would be no misunderstanding of what is defined as the agreed-upon data. Baseline data, particularly those used as input to assessment models, should be accessible from a centralised database. Having the Secretariat serve as this data repository may have several benefits. First, a centralised database that can be updated periodically and referred to by all parties should reduce misunderstandings with regard to what data sets are available and agreed upon for use. Of course, documentation noting where the data came from and when it was last updated will need to be included with the database, but this is straightforward. Second, having such a database would also facilitate structuring timelines for update and accessibility of data in preparation for upcoming assessments. It would also make the parties providing the data responsible to the Secretariat (and therefore the Commission), rather than to each other, for supplying data and meeting timelines that may be otherwise poorly specified or poorly understood.

2.4 Interpretation and analysis of catch statistics

Even though there is considerable discussion under the basic data agenda item, these discussions relate mostly to how the data are treated, and not what they mean. For example, when examined in an historical perspective, the age composition of the catch appears to have changed over the years (Figure 1). The proportion of the catch made up of ages 0 to 5 increased steadily in the 1960s and 1970s and through the late 1980s, while the proportion of ages 6 to 10 decreased steadily over the same period. The proportion of ages 11 and older in the catch originally decreased, but has been increasing since the early 1980s.

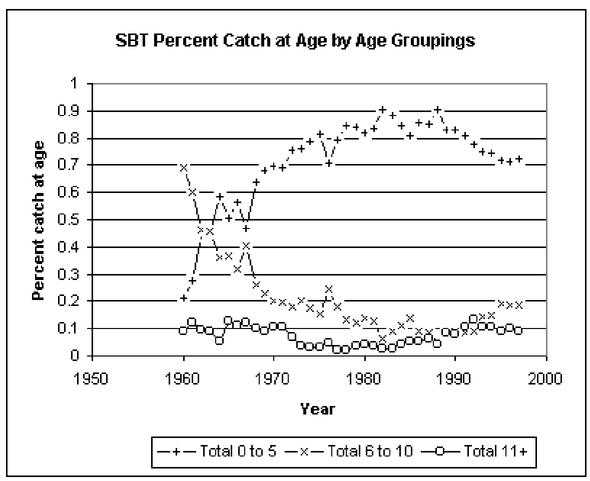


Figure 1: Southern bluefin tuna catch at age by age-groupings.

Interestingly there appears to be a trend over time in the ratio of reported landings to the product of catch at age times weights at age as shown in Figure 2.

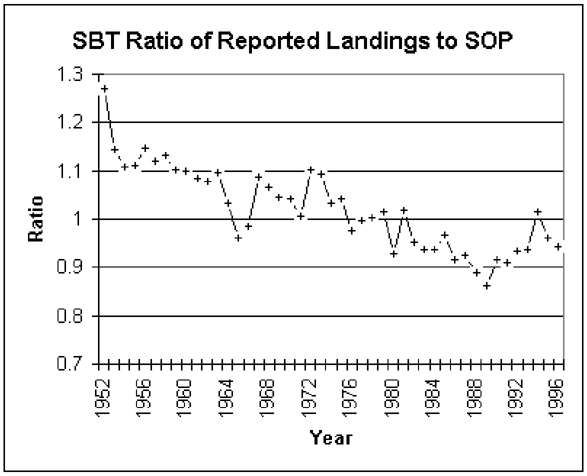


Figure 2: Southern bluefin tuna ratio of reported landings to the sum of products (SOP) of catch at age multiplied by weights at age.

Such a trend in the ratio is unexpected. A ratio greater than 1.0 implies that either the catch numbers or the weights at age are underestimated, and conversely a ratio smaller than 1.0, implies that either the numbers caught or the average weights are overestimated. Variability in the ratio is expected, but a trend is not. This is a minor point, but such inconsistencies in the basic data should be resolved.

3 Availability and necessity of data to be used in the stock assessment

Stock assessments require that some indices of changes in stock size or in fishing mortality be available. It is difficult to derive indices of stock sizes from the commercial fishery that are consistent over time. This is due to the wide-ranging nature of southern bluefin tuna, the diversity of the fisheries, and the changes in the behaviour of the fleets either in responses to management measures, changes in market conditions, technological changes in fishing, navigation and/or fish finding equipment. Therefore, in addition to fishery-dependent indices of stock size, it is useful to develop indices of stock sizes that are not affected by the problems identified. Such indices are described as being fishery independent, but this does not mean that they have to be derived solely by the research institutions independently from the commercial fishing operations. On the contrary, there are advantages in using the commercial fleets to obtain such 'fishery-independent' indices of stock size (for example reduced cost, improved relationship with the trade).

Several so-called fishery-independent indices have been introduced and, in some instances, used in the SBT stock assessments. These include aerial surveys, acoustic surveys, and information gathered through tagging experiments. While several technical and practical problems exist in implementing these approaches, their use is of primary importance in better defining the current stock size. Recruitment, adult stock abundance, and survivorship are the three defining components of any stock assessment. Continued support should be given to all three components, but the aerial survey, if satisfactorily developed, may be the most expedient method for ascertaining stock recovery through changes observed in recruiting year-classes.

Several competing hypotheses on how to interpret the indices of stock size currently exist leading to significantly different pictures of stock status. Additional modelling exercises will not provide a solution as to which of these hypotheses is the most likely. With regard to the specific issue of how stock abundance is distributed over the fishing grounds (i.e. constant squares vs. variable squares assumptions), it is only by gathering data from those areas where no commercial fishing takes place that this problem is likely to be resolved in the short term. However, as with any scientific approach, a statistical design for gathering the data must be implemented so that the data are cost-effective, unbiased, answer the specific question of concern, and are precise enough to contribute significant information to the assessment.

Tagging programs are another source of valuable information about the stock. However, the proportion of recovered tags that are returned, sometimes called the reporting rate, is one of the important sources of uncertainty in estimating fishing mortality, natural mortality, and migration rates from tagging experiments. The proportion of recovered tags that are returned is likely to vary from fishery to fishery, country to country, and also over time. Therefore, in order to decrease such uncertainty, it might be preferable that the recovery of tags be under controlled conditions, for example, by using tags that can be collected by observers and independent of the fishery (e.g. coded-wire tags, pit tags, pop-up tags). There are costs, which are not insignificant, associated with such tagging programs, however, the greater precision and reduced bias associated with them may make the cost increase trivial relative to what is spent on current tagging programs.

Good assessment models, even those using state-of-the-art techniques, cannot substitute for high quality data, and the best assessment model will not provide reliable results if the basic catch at age and stock size indices are unreliable or uninformative. Every effort should be made to collect continuous and consistent time series of high quality data from every aspect of the fishery so that informative analyses can be made to support decision making. In particular, it would be useful to make available catch and effort information for every major sector of the fishery, including the surface fishery. These data may not resolve the uncertainties in current stock size, but their interpretation will undoubtedly provide some information useful in assessing stock size.

The current differing interpretation of recent trends in the southern bluefin tuna stock can be linked directly to the paucity of informative stock size indices. In order to be effective, any fishery management system must be credible to all interested parties. Credibility depends on many factors, but it is unlikely to be established and maintained unless the means to monitor the effect of management are put into place. This starts with good sampling and data collection programs, but it also includes management monitoring the responses of the stock and the fishery to management actions.

4 Hypotheses and structure of assessment model

4.1 Statistical age-structured models

While the modelling approaches used by SBT scientists are sophisticated by global standards, alternative approaches like those proposed by Hilborn et al. (1998) indicate that some modification and update of current models could prove useful. Specifically, stock assessment scientists might consider the benefits of a forward-calculated assessment that incorporates error explicitly for the catch-age data. The backward VPA method used assumes that the catch at age is known without error. It is therefore of the utmost importance in VPA applications to use only those ages and years where there is reasonable confidence in the catch-at-age data. Other methods are not so critically dependent on the catch at age being known without error and can better account for such errors. These ‘statistical age-structured’ methods are becoming more and more widely used.

4.2 Structure of the plus group

The decision to combine catches of older ages into a 12+ category is problematic in the southern bluefin tuna assessment. The selectivity in the age range 10 to 20 years old appears to vary not necessarily smoothly over time, and the contribution of the 12+ group is far from negligible, representing more than 25% of the catch weight on average since 1990. Fitting the plus group in a VPA based assessment, although used in some situations not very different from the southern bluefin tuna, is generally not recommended. Instead, some relationship between the F on an earlier age group and the plus group is made. However, in the case of southern bluefin tuna, as mentioned above, it is likely that the relationship of F over age has changed over time and there is little information available to estimate that relationship outside of the model. In the best of cases, when reliable indices of the plus group are available, tuning the plus group would be a relatively hazardous undertaking. In the absence of reliable indices, as is the case for southern bluefin tuna because of changes in selectivity and targeting plaguing the CPUE series, it is doubtful that the tuning provides real information on the 12+ stock size trends. Therefore, the plus group should be moved to at least age 20+ as this would remove the need to have all 14 plus group options in the sensitivity analysis runs.

Extending the age-composition past the 12+ age group takes on added importance given the recent observations of the age composition of catches by Indonesian vessels on the spawning grounds. Extending the age composition, and using statistical catch-at-age models would allow the incorporation of such information in the assessment. It would be interesting to see what combination of recruitment, exploitation rates and natural mortality rate can reproduce the observed age composition of the catches on the spawning grounds by Indonesia.

We therefore recommend that tuning the plus group, as is currently done should be abandoned. Our preferred option is to extend the age range in the catch-at-age matrix at least to age 20 and perhaps older. In fact, there are no obvious disadvantages to such an extension as the correlation between the catch at age of successive ages on a year-class basis are very good until at least age 18 to 19 where the correlation is r=0.80 for the period 1969 to 1997. Therefore, the year-class signal which is currently obtained using age 11 as the last true age will not be corrupted by extending the age composition. Should the SAG/SC decide not to, or be unable to extend the age range for the next assessment, they should base the estimate of the plus group in the terminal year on the relationship between the F on the last true age and the plus group in that year, and hereto backcalculate the plus group for previous years based on that estimate.

4.3 Replacement yield

The replacement-yield component of the model influences the fit in a way that is model based and not data based. Unlike other population dynamic components of the model, such as annual survivorship, which we know must be represented in some form, the replacement-yield component represents what we *hope* is true. Biologically, we hope that recruitment is sufficient to replace the spawning stock biomass that produced it, but in reality it may not. One situation where it is not realistic, for example, is during a regime shift. Furthermore, having such a component in the model may make a potentially unrealistic or uninformative model appear realistic and informative. It may also serve to confound or mask trends in residuals that, if observed, could serve to allow other more informative models to be realised. It is recommended that this component no longer be used.

4.4 Examining model fit as a test of hypotheses

The subject of including model fit as a criterion in the analysis of VPA results was raised on several occasions in both the SAG and the SC. However, the discussion was not about using conventional goodness of fit/Maximum Likelihood criteria, and it seems that such conventional model- fit criteria are not examined to any extent. This may be difficult to do given the hundreds of VPA runs being considered. Due consideration to goodness of fit, in the conventional sense, is appropriate and should be encouraged because incorporating the results of unrealistic VPA formulations in assessment summaries will lead to an erroneous understanding of the stock and may mislead management.

The assessments conducted by the SBT scientists compare favourably with those performed elsewhere in the world, but the interpretation of the stock status information that is prepared from the assessments is lacking. This may be the result of an inappropriate definition or perception of the role of the scientists with respect to those of the fishery managers. We find that the information contained in the stock status report prepared for the Commission is lacking and insufficient.

5 Quality and appropriateness of tuning indices

Tuning indices are a necessary component of age-structured analyses such as the VPA used to assess southern bluefin tuna (see section 3 on availability and necessity of data to be used in the stock assessment). Tuning indices can come from fishery-dependent and/or fishery-independent data sources. Several different indices may exist from these sources. They often represent different types of information needed in the assessment and sometimes these indices provide contradictory measures of stock status. Assessment authors often choose different weights for each index in an attempt to represent the information present in each. Currently, in the southern bluefin tuna assessments, there is no explicit weighting of the data series, regardless of the information they provide.

The southern bluefin tuna VPA relies heavily upon a fishery-dependent CPUE index. However, the interpretation of commercial CPUE is complicated by technological innovations, changes in fishing strategies, changes in management measures, and changes in fish distribution or species composition. There are currently two modelling hypotheses in use for SBT to represent contradictory interpretations of this index (see below in section 6 on methods and hypothesis to standardise indices). This contradiction has led to significant differences in interpreting SBT stock status.

One method proposed to resolve this conflict involves expanding the fishery-dependent CPUE index through experimental fishing. The idea is that there exist areas where fish density is unknown because no fishing takes place there, and that by fishing in these areas (following an experimental design) information will become available where none currently exists. If this approach is pursued, care should be taken to ensure that an informative experimental design is implemented and that the stock will not suffer as a result of the experiment. Conceivably, there may be alternative methods for gathering this information outside the experimental fishery arena, such as through an acoustic survey for example but there may be many practical as well as methodological limitations to such alternatives. Nevertheless, such alternatives should be discussed as part of the process.

More broadly, efforts must continue to expand and develop fishery-independent tuning indices. Those provided by the aerial and acoustic surveys of juveniles will become increasingly important. Measures of stock status taken consistently over time form the basis of all good stock assessments. Periodic or one-time-only measures even though they may be precise, are often not as informative as those taken over an extended time period. Such indices, although they generally involve an investment in time and human resources, will ultimately prove cost effective as the stock changes and alternative management procedures are considered.

6 Method and hypotheses to standardise indices

Current CPUE standardisation approaches do not account for changes in efficiency or changes in targeting over time, unless these are associated unambiguously with area specific targeting.

Clearly the variable squares approach, which assumes that there are zero fish in all of the squares that were once fished, but that are not fished in a given year, is not a realistic assumption. It must be considered as an extreme case and there is a high probability that the true CPUE is somewhat higher. The same cannot be said of the constant square assumption, and although it is unlikely that the abundance in unfished squares will be exactly the same or higher than in fished squares, this approach could be realistic in some years, and not in others.

Neither hypothesis is likely to be true all the time. Southern bluefin tuna geographical distribution is likely linked to yet to be identified environmental variables, including the distribution of their prey, which are themselves strongly influenced by physical variables such as temperature, salinity and currents. It is therefore likely that the size of the habitat available to southern bluefin tuna varies from year to year. This is well recognised by the fishermen, and tuna fleets have used sea surface temperature maps to orient their activities since the late fifties. Although incorporating environmental co-variates in the analyses of CPUE data may help resolve some of the uncertainties, it would not be able to account for changes in technology and/or targeting in response to changing market conditions.

As shown in Figure 3, the various standardisations of the CPUE series globally show the same trends for 1969 to 1992-93 within ages or age groupings. It is clear that the approaches reduce to two with the Habitat and Variable Square, on one hand, producing virtually identical trends, and the Geostatistical and Constant Squares also showing virtually identical relative trends. The SAG should therefore reduce to no more than 2 options, the number of standardisation approaches considered, and use these two approaches consistently for a few years. The reason to recommend no changes for a few years is because there is unlikely to be enough resolution in past data to choose among alternative approaches.

The CPUE often show very rapid decreases from 1969 to 1973 with the decline for older ages appearing particularly steep. It is the Panel's understanding that the fishery moved away from the spawning aggregations in 1973, and it is possible that the rapid decrease in previous years may also be the result of changes in fishing practices. In theory, if these were changes in the geographical areas of fishing, they would be accounted for by the catch rate standardisation, provided there were enough observations to properly estimate the area-month effects. Fishery changes or biological reasons should be investigated to explain the steep decrease in CPUE from 1972 to 1973.

There is a marked decrease in the absolute numbers of fish caught for almost all ages from 1972 to 1973. This is suggestive of either a fundamental change in the way the fishery was being prosecuted, a change in the fishing effort, in the location of fishing, or in sampling. There is also a small decrease in tonnage caught between 1972 and 1973, but it is not as striking as the decrease in the numbers caught.

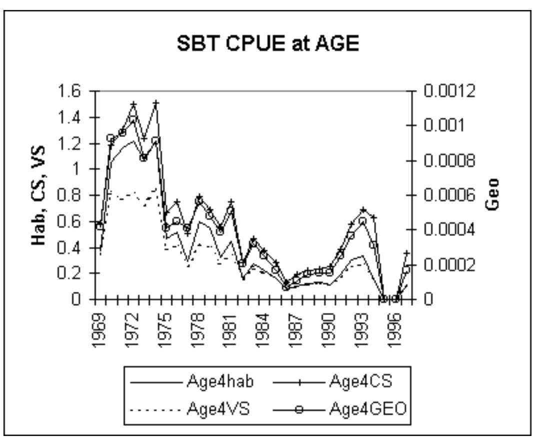
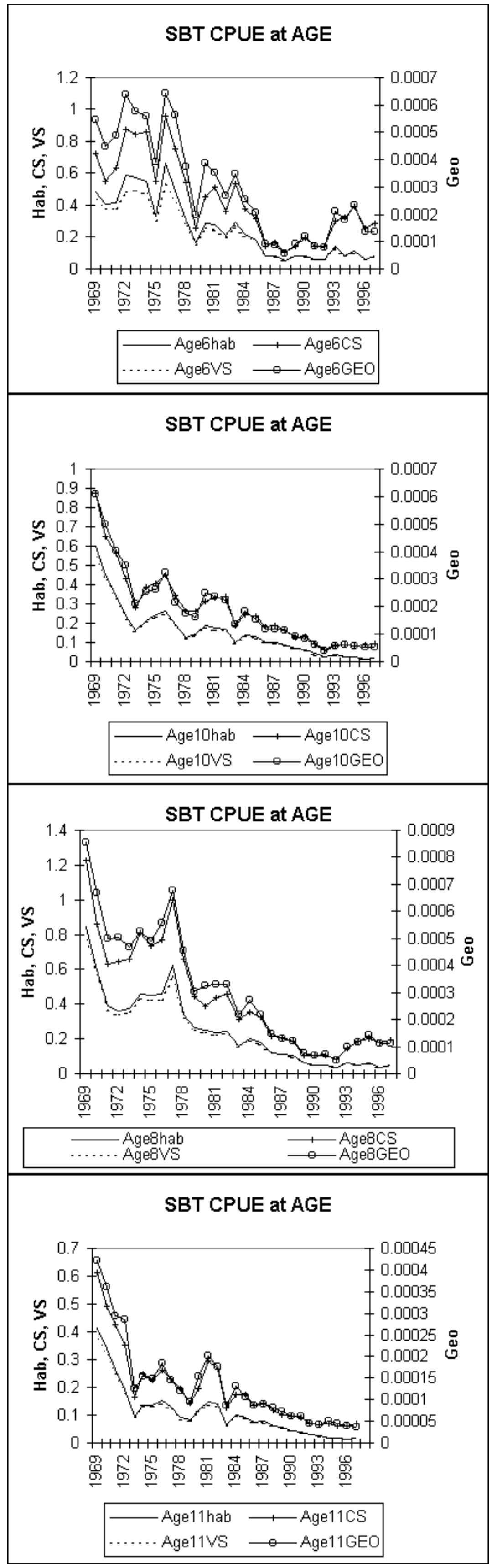


Figure 3: (above and right hand column) Southern bluefin tuna comparison of various CPUE standardisations. The legend gives the age (or age group) considered; hab=habitat, CS=constant square, VS=variable square, GEO=geostatistical.



7 Biological parameters used in assessment

Currently, 10 vectors of natural mortality at age are used in some of the assessments. This is not necessary because the differences between the vectors are small. The number of vectors certainly can be reduced to three, the minimum values at age, either the mean or the median, and the maximum. In order to encapsulate the uncertainties, it would be sufficient to use the minimum and the maximum values at age.

Min	Median	Mean	Max
0.200	0.400	0.400	0.500
0.200	0.366	0.356	0.483
0.200	0.300	0.313	0.467
0.200	0.250	0.270	0.450
0.200	0.233	0.246	0.367
0.200	0.217	0.223	0.283
0.200	0.200	0.200	0.200
0.175	0.175	0.175	0.175
0.150	0.150	0.150	0.150
0.125	0.125	0.125	0.125
0.100	0.100	0.100	0.100
0.100	0.100	0.100	0.100
0.100	0.100	0.100	0.100

A continuous tagging experiment to estimate survival curves might be useful in estimating natural mortality, and therefore remove the need to use more than one M vector.

In recent assessments, southern bluefin tuna age 8 and older had been assumed to be fully mature, while those of age 7 were assumed to be immature. There appears to be relatively little information on the maturation of southern bluefin tuna (Farley and Davis, 1998). The results of the two assessments presented to the SAG did not consistently use the same maturity schedule. In one case, the assumption was as in recent assessments, i.e. fish age 8 and older (8+) were assumed to be mature, while in the other, in addition to 8+, fish age 10+, and 12+ were assumed to be fully mature, while younger ages were assumed immature. This led to some confusion about that the differences in spawning stock biomass (SSB) trends seen in some graphs which were believed to have been due to model or coding differences rather than to different assumptions about the maturity schedule.

The discussion on the maturity schedule was complicated by the same words having different meanings for the participants. A clear distinction should be made between maturity and contribution to reproduction. Although a fish may be physiologically mature, it will not necessarily make a large contribution to the egg production either in quantity or in quantity (Trippel 1998). The issue has several implications. The first and most obvious is that it is the production of viable eggs that is important for studying the relationship between the parental stock and the subsequent recruitment, and work should be undertaken to study that relationship for southern bluefin tuna.

Of more immediate concern to the review Panel and related to the assessment is the changing of the measuring standard for SSB. Even though it is the production of viable eggs that is important for studying the relationship between the parental stock and the subsequent recruitment, that information is rarely available. Normally, it is assumed that egg production is directly proportional to SSB, but often, as is the case for southern bluefin tuna, SSB itself is poorly estimated because there are few data on maturity at age and how it changes over time. In these cases, the sum of the biomass over agreed age groups is taken as a proxy for SSB that is itself a proxy for the egg production[3].

Once agreement has been reached on the standard to be used in monitoring SSB, changing the standard should follow the rules suggested below in the section on the process to incorporate new techniques and/or new information. For credibility and transparency reasons, it is desirable to maintain consistency from one assessment to the other and that changes are incorporated gradually.

In order to be effective, any fishery management system must be credible to all interested parties. Credibility depends on many factors, but it is unlikely to be established and maintained when substantial changes are introduced in indicators such as the proxy for SSB without forewarning and clear supporting evidence. Even when the evidence for the change is strong and conclusive, common practice in other fora is to compare the assessment results under both the current assumption and the proposed change for one or more years before the change is actually made. This makes it possible to identify the reasons for changes compared with previous assessments.

[3] There is evidence (Marshall et. al. 1998) that SSB, even when the maturity schedule is reliably estimated, may be a poor proxy for the production of viable eggs.

8 Sets of weightings assigned to uncertainties

As indicated below in methods of treating uncertainties in models (Section 10), the assessment results should be further summarised, with the objective of more clearly presenting the range of uncertainties for population parameters important for fisheries management, i.e. recruitment, SSB, and fishing mortality. The country-specific prior weights assigned to various model specifications are an attempt to summarise information. However, these weightings do not explicitly incorporate the 'likelihood' of individual models. This approach is commonly used in decision tables (discussed under Quality and format of Report to Commission in Section 14).

Weightings are currently assigned by each delegation to each model considered based on that delegation's belief of how well that model represents reality. Hundreds of models are considered (with hundreds more considered for analysis of sensitivity to other considerations). As reviewers we cannot comment on the appropriateness of one delegation's weightings relative to another. However, rather than having the SAG or SC present a summary based on the average of all model results based on these weightings, it may be more appropriate for managers to consider a few key alternatives presented with the associated delegation weights.

For example, it was clear from the papers presented and from SAG and SC discussions that key alternatives to consider are those resulting from hypotheses based on the variable squares vs. constant squares CPUE indexes, and those resulting from the various plus group options. Other model alternatives either produced similar results or received equal weighting from all delegations. Thus these few key alternatives could be fully represented by two-by-two or three-by-three decision tables. Alternatives presented in these tables could show the likelihood of achieving reference points (e.g. the probability of the stock achieving the reference level by the year 2020), the associated yield over that time period under that model, and each delegation's belief in the probability of that scenario. Managers would then have in their hands all the information necessary to see the consequences of different alternatives, associated degrees of belief, and a clear specification of the trade-off in stock biomass and yield. Furthermore, managers would have information to gage the adequacy of their decisions relative to the hypotheses considered and could better prioritise analysis of future research and management options.

9 Hypotheses and structure of models used in projections

The relationship between parent stock and subsequent recruitment is the single most critical factor in calculating projections of future stock sizes and catches under various exploitation scenarios. The stock and recruitment data for southern bluefin tuna in the retrospective part of the assessment, 1969 to 1998, is used to derive possible future relationships between SSB and recruitment. Aside from the more or less steady decrease in recruitment over time, these results are peculiar for two reasons. First, recruitment appears stable for relatively long periods of time (Figure 4), that is it does not vary much from one year to the next.

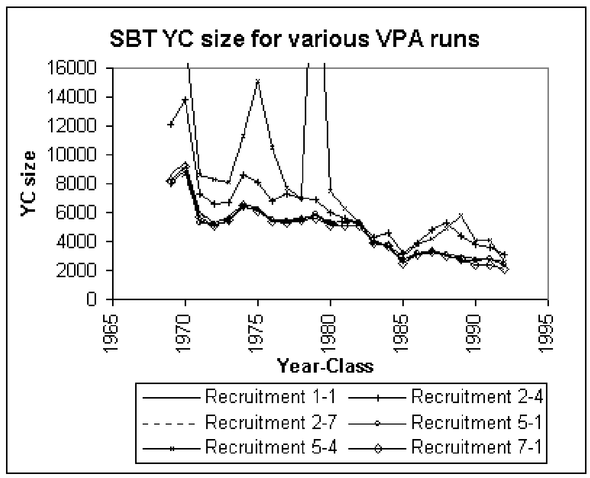


Figure 4: Southern bluefin tuna recruitment versus time from various VPA runs. The legend refers to the VPA run number identifier in working document CCSBT-SC/9807/17. The vertical scale has been truncated to better separate the lines corresponding to the various VPA runs.

Second, recruitment appears to be linearly related to parental stock size, with lower SSB apparently leading to smaller recruitment. The relationship, for some of the VPAs is especially tight (Figure 5).

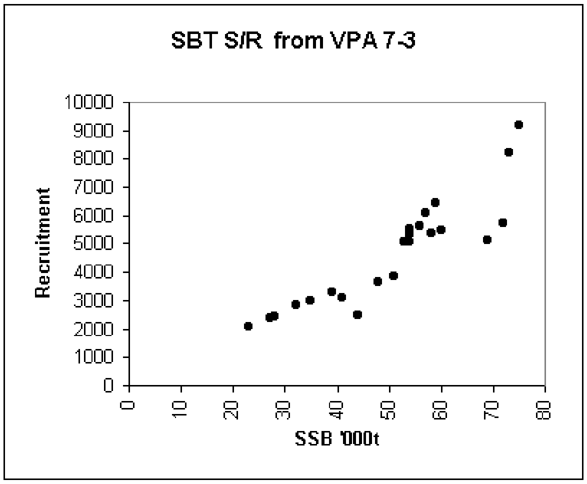


Figure 5: Southern bluefin tuna stock and recruitment relationship from VPA 7-3 (working document CCSBT-SC/9807/17).

The stability of recruitment is probably a result of smoothing related to the cohort slicing method used to convert lengths to ages. The real variability in year-class size is expected to be larger than shown in the assessment results.

Klaer et. al. (1996) have investigated the performance of stock assessments and stock projections for the 1982 – 1995 assessments. In the 1990 to 1995 assessments (their Figure 2), the terminal year, that is the most recent year with catch data, generally had the lowest SSB in the time series, and the projections suggested that it would increase, sometimes very steeply in the upcoming few years. The authors conclude (p. 10, conclusion 1) that "The bias is present in the most recent estimates of numbers at age produced by VPA, and is not influenced by any assumptions about future recruitment…".

There was insufficient time for the Panel to undertake a retrospective analysis on its own, and the subject was only briefly discussed at the SAG/SC. However, based on the study by Klaer et. al. (1996), it would appear that the problem in the assessment reviewed related to the VPA estimates of partially recruited year-classes, which explains the rapidity of the predicted recovery. If the problem had been with the assumed or fitted stock and recruitment relationship, the recovery would have taken at least 8 years to be seen. As indicated elsewhere in this report, the main problem is perhaps not so much in predicting future stock trends under various management scenarios, as with the inability to measure future changes in stock size as a result of management measures.

The majority of the results available suggest that there is a strong relationship between the SSB and subsequent recruitment. Stock and recruitment relationships for marine species are rarely as clear as indicated for southern bluefin tuna. Although the relationships observed provide strong incentive to rebuild the southern bluefin tuna SSB as quickly as possible, it is not impossible that factors other than the SSB also play an important role in determining year-class strength. If this were the case, it could be that the SSB is low because past recruitment has been low due to environmental conditions, not the reverse. The possible environmental influence on year-class size should therefore be studied.

10 Methods treating uncertainties in models

Several types of uncertainty must be recognised in an assessment. First, there are structural differences that reflect alternative hypotheses about how the population behaves, that is, is the population model adequately reflecting the way the population really behaves. Second, in addition to model uncertainty, the data used to fit the model are not sampled with complete precision, and for any given model there are different data inputs representing different aspects of the fishery each with their own sampling variability and uncertainty. The information content of these data, in turn, reflects not only the properties of statistical sampling, but also the degree to which they provide insight on process. These two sources of uncertainty are necessarily treated differently.

In terms of structural differences, the SAG participants have chosen to explicitly identify alternative model formulations of the same basic VPA approach. While, agreement may not always exist between participants as to what models are appropriate, the range of models encompasses to some extent the beliefs of the stock assessment group as a whole. Ideally, one would hope that a consensus might be reached about model structure, but baring that, there are several approaches one can take to presenting alternatives. If the number of alternatives is not too large, it may be best to present the results of the key alternative models to the fishery managers. Weights can be given to the degree of belief in each alternative, if the stock assessment participants can agree on these weightings, however it may still be important for the managers to see the independent alternatives with their weights to better judge the consequences of management actions. Here a decision table with consequences to future stock size and yield under each alternative might be the most informative approach. In addition to VPA models, it may be informative to routinely run and update simpler models, such as production models.

Relative to information in the data, the models currently used for southern bluefin tuna give equal weighting to each data component without taking into account sampling uncertainty or differences in information content. What the appropriate weightings should be, however, is a complex decision process. Nevertheless, it appears that some steps forward are being made in this area to better represent what the different sources of information contribute to the assessments. Bayesian information criteria are useful, but care should be taken in using fully automated approaches, especially when the various sources of information provide conflicting indications.

The current treatment of uncertainties may in fact make it overly complicated to properly assess the uncertainties in stock status and parameter estimates. It would be useful to reduce the number of parameter estimates used in the sensitivity runs, and vectors of parameters should be arranged to provide results in a progressive manner from pessimistic to optimistic or vice-versa on factors important for fisheries management decisions. For advisory purposes, it may not be necessary to provide the results of all runs. It may be sufficient to provide lower and upper limits and the median.

It is imperative that uncertainties, which exist in the assessments, be translated into consequences (for spawning stock biomass and for fishery yield) relative to decisions for management. While there are uncertainties in the SBT assessment, there is also information. Presenting only uncertainties to management without some suggested protocol for interpreting and dealing with it is inadequate.

11 Process to evaluate calculations and computer codes

Verification of computer code, and calculations, is best dealt with in a small technical working group. Much of this can be accomplished without having to meet in person (e.g. via e-mail and by exchanging the code and data files, and having it archived at the Secretariat). However, if the modelling and assessment process continues to change rapidly on SBT, it might be necessary to organise face-to-face meetings of the technical working group at least once several months in advance of the formal SAG meeting. Such a meeting is necessary to confirm the finer details of the coding and data input. The stock assessment group was stalled several times this year by technical incompatibilities (perceived or otherwise) that might have been resolved earlier in the process by a small working group.

Several types of process checking should take place. First, a verbal specification of what the computer code is to accomplish, a specification of how it should be implemented, and a listing of what input criteria are to be used must be agreed upon by all parties prior to coding. Doing this up front will save hours of coding and calculations wasted on poorly specified problems.

Second, the computer code itself must be checked to see that it accurately reflects these specifications. First a conceptual check is made, and then a computational check takes place. The later generally involves matching model estimates with those gathered from an independent or known source. For computer estimation procedures, like VPA, this usually involves applying the procedure to a known data set, possibly one created by a simulation whose baseline parameters are known. Another type of verification is the visual and computational comparison of code created by independent parties. If everything is coded correctly then there is a match computationally and conceptually. In other words, it adds up and it makes sense.

Of course, a verification of the code, that is a checking for conceptual and computational accuracy, does not guarantee that the underlying processes governing the stock are accurately characterised by the code. Nevertheless, it does mean that the computer code is operating as expected and can be used to explore a limited set of hypotheses about the stock.

The parties involved in southern bluefin tuna stock assessment take great effort to make their stock assessment models algorithmically consistent. Consistency between algorithms is desirable in the sense that independent stock assessment groups can cross check results, confirming the quality of the code and data inputs. However, algorithmically different models could (and perhaps should) be proposed within as well as between the different stock assessment groups. Alternative algorithms can represent different statistical approaches or even different hypotheses about how the stock behaves. If alternative models are proposed, they still must be subject to the quality control criteria specified in previous paragraphs. This would include clear specification of objectives, model documentation, shared computer code and input data, and adequate verification procedures.

12 Process to incorporate new techniques and/or new information

As in other stock assessment and fishery management fora, stock assessment models for southern bluefin tuna and their data will continue to evolve as part of the scientific process. There is not enough time during the annual assessment meeting to introduce a new scientific concept or data set and have it adequately reviewed for its scientific merit by all delegations. Small changes or adjustments to models or data should, of course, be implemented immediately, but significant conceptual changes require time to develop, implement, and be reviewed and understood by all parties. A mechanism must be developed to address this issue, since the lack of one has already lead to significant communication problems between participants and is impeding the scientific process. A possible mechanism might involve proposing a new concept or data set during an assessment meeting one year in advance of its use in the assessment. The proposal would involve full model or sampling design specifications so that all participants would be able to investigate and fully understand the implications of the proposal intersessionally. Questions, criticisms, or suggested modifications should be raised no later than 3 months prior to the meeting where it will be implemented. It would have to be explicitly stated that the concept is proposed for use in the assessment in the following year so that all participants would recognise that the proposal is aimed at implementation. If, after the proposal has been reviewed (that is after one assessment cycle) there are still disagreements between participants as to its significance or scientific credibility, then both the proposed method and its critique should be included in the assessment advice.

12.1 Process to review newly incorporated information

A protocol for reviewing a new concept or data set must exist so that the process is clear for both those who submit the new data or the new concept and those who are to review their proposal. A critical review of each proposal should be submitted by each delegation as an information paper to the authors and to the Secretariat. The review papers must, at the very least, be submitted prior to the deadline date for the next stock assessment meeting. Ideally, however, such reviews should be submitted to the authors of the proposal as early on as possible in the process so that communication and clarification of ideas can proceed more rapidly. Reviews should include comments on the scientific quality of the work, its appropriateness in addressing the problem specified, and, if necessary, it should include suggestions for improvement or expansion of the work.

13 Assessment results

Although considerable time is spent discussing the choice of parameters to be used in various VPA runs, little time was spent on interpreting the results of the various analyses and on reducing the number of possible interpretations to a more manageable number of possibilities. In this section of our report, we will try to interpret some of the assessment results and contrast them with the results from simpler approaches.

It is often informative to make a graph of the landings versus fishing effort to try to understand the evolution of a fishery. The southern bluefin tuna landings are plotted versus the Japanese longline (JLL) fishing effort for 1955 to 1997 (Figure 6). Four different periods are easily identifiable:

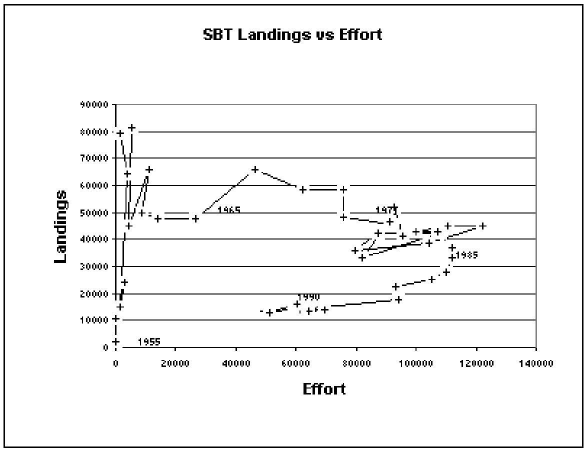


Figure 6: Southern bluefin tuna landings (tonnes) vs Japanese longline fishing effort.

1. high landings were achieved at low fishing effort when the fishery started in the 1950s;
2. from 1965 to the early 1970s, the fishing effort increased rapidly while landings only decreased slightly: this period presumably corresponds to a geographical expansion of the areas fished;
3. from 1973 to the mid-1980s, the fishery seemed to have reached some form of equilibrium at landings in the range of about 35,000 to 45,000t and fishing effort centered around 100,000 units;
4. starting in the mid 1980s both fishing effort and landings decreased rapidly and a new equilibrium seems to have been reached in the 1990s with landings around 15,000t and fishing effort centered on 60,000 units.

This simple analysis has clear shortcomings. For example, the share of the total catch caught by the JLL decreased from about 95% of the total in the 1950s to slightly less than 40% in 1996, therefore, the JLL effort in the later part of the series is not representative of the total fishing effort expanded. The graph of JLL landings vs JLL effort is essentially similar, and the objective of this exercise is to point out that there appears to exist at least 3-4 distinct periods in the fishery and that these should be taken explicitly into account when analysing the SBT fishery data.

The number of parameters and their possible values accepted for inclusion in analyses by the SAG/SC means that 216 VPA runs or more have been included in some assessment documents submitted to the SAG/SC. This number of VPA is too large to meaningfully interpret and analyse in the time available given the form of the SAG/SC meetings and its predecessors. Therefore, the number of runs **must** be reduced and extreme results identified in sensitivity runs as being highly unlikely should be eliminated from further considerations.

To illustrate how this elimination process could proceed, a sample of 6 recruitment and SSB series have been subjectively selected from VPA runs included in documents presented to the SAG/SC, based mostly on the variability they showed in SSB over time. The results for year-class size versus time were presented earlier in Figure 4 and those for SSB versus time are presented in Figure 7. As can be seen, the recruitment series are relatively consistent with one another while there is considerably more variability in the absolute magnitude of the SSB and also in the trends over time[4].
The large number of VPA results presented in the assessment documents, even if they are presented only as sensitivity tests, confuse the interpretations and the discussions. Not all of these interpretations are realistic, and unrealistic ones should be discarded, and not presented. There are no benefits in repeatedly conducting the same unrealistic analyses.

[4] The variability in SSB is not translated into corresponding variability in recruitment because the SSB from the runs selected is approximated by the age 12 and older biomass, i.e. the plus group, and the calculation of the plus group is not linked with population estimates for earlier ages in the VPA.

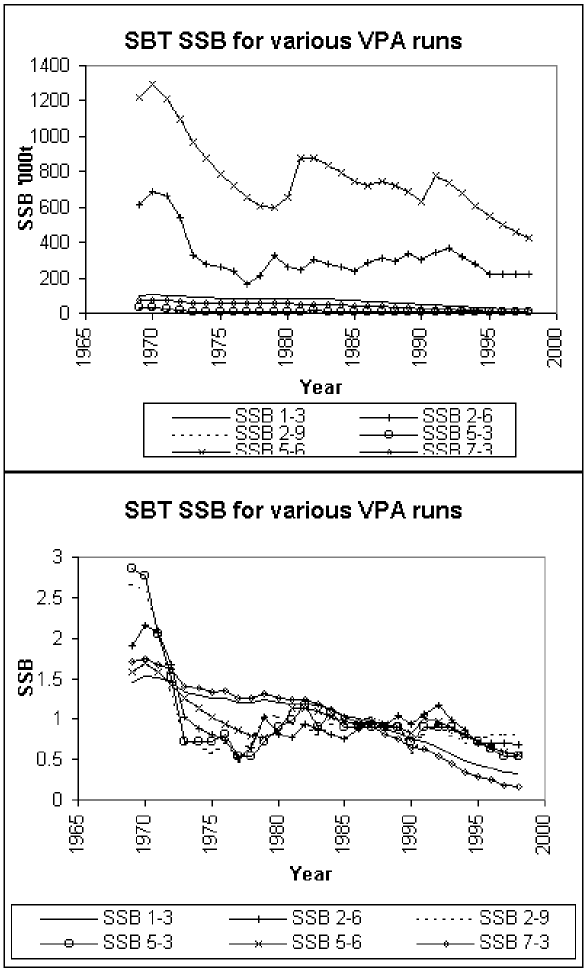


Figure 7: Southern bluefin tuna SSB vs time. The untransformed data are shown in the upper panel, while each series is scaled by its own average in the lower panel. The legend refers to the VPA run identifier in CCSBT-SC/9807/17.

If too many VPA runs are made and presented there is a risk that basic trends contained in the catch-at-age matrix will not be identified.

It is well recognised (Laurec and Shepherd 1983) that there is not sufficient information in the catch at age itself to estimate current stock size without external information on stock size or on fishing mortality. However, in cases where there are few conclusive indices of stock size and/or exploitation rates, assumptions on how recent F relates to F in previous years, on an age by age basis, can be used as external information. For exploratory purposes, using the catch at age 0 to age 19 and the natural mortality vector labelled V2 in Table 4 of Tsuji and Takeuchi (1998), we have assumed that F for each age in 1997 (the terminal year) was equal to the average for the years 1993 to 1996 for that age and that F on age 19 for every year was equal to the average for ages 12 to 16 in the same year. We have used Pope's (1972) approximation to the VPA equation and we have not included the plus group in our calculations. F at age in 1997 was iteratively replaced by the new average for 1993-96 until the maximum difference for all ages between successive runs was smaller than 0.001. The process was initiated from two different starting F (F=0.04 and F=0.20, for all ages) and converged to the similar values as indicated in Figure 8.

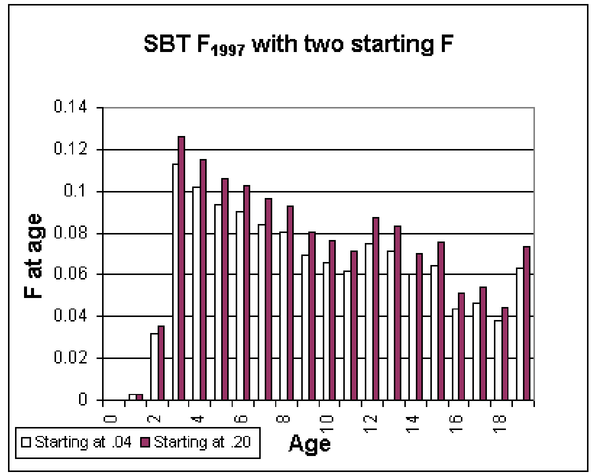


Figure 8: Southern bluefin tuna estimated F at age in 1997 from two initial F values, F=0.04 and F=0.20.

Results from such a simple analysis should not be over interpreted, but we believe that the changes in exploitation patterns (also known as partial recruitment or selectivities) (Figure 9) over time can be informative. Similarly, trends in the population numbers (Figure 10) of various age-groupings are probably relatively robust until the early 1990s.

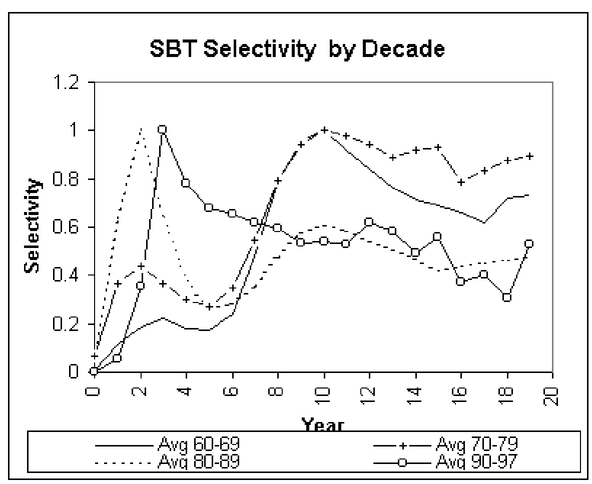


Figure 9: Southern bluefin tuna selectivity at age by

decade from a simple cohort analysis (see text for details).

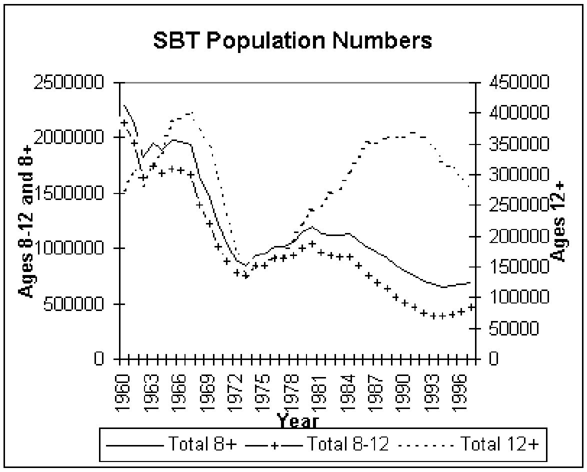


Figure 10: Southern bluefin tuna results of simple cohort analysis (see text). Note the different vertical scale for the 12+ grouping.

Considering the widespread concern about the health of the southern bluefin tuna stock, it is paradoxical that the exploitation pattern has progressively shifted from being low on young age-groups (ages 6 and less) compared to ages 12 and older in the 1960s, to those age-groups becoming fully exploited in the 1980s and 1990s as can be seen by plotting the average F for ages 10+ and for ages 3-6 versus time (Figure 10). Although the average F on ages 10+ appears to have decreased steadily since the early 1980s with the 1997 value being possibly in the order of one third of the average for the 1960 to 1997 period, the average F for ages 3-6 may have continued to increase until the late 1980s with the 1997 value possibly being almost 10% higher than the 1960 to 1997 average.

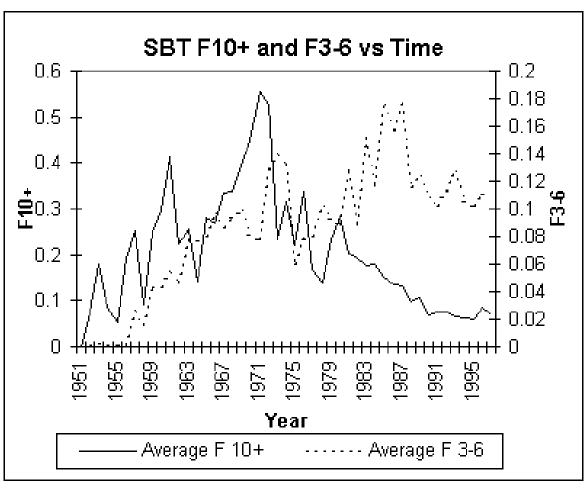


Figure 11: Southern bluefin tuna estimated F from a simple cohort analysis (see text for details).

The trend in recruitment from this simple analysis is broadly similar to those reported in the assessment documents. Recruitment appears to be surprisingly stable over time, probably an artefact of the cohort slicing procedure used to calculate the catch at age for all ages, including those young ages where modes are clearly identifiable. As mentioned elsewhere in this report, the catch at age should be derived directly from the modes in the length frequencies for the ages where such modes are identifiable.

It is very rare that fishery data will unambiguously lead the analyst to a uniquely correct solution. Generally, therefore, there is room for differing interpretation, and a scientist may get trapped in discussing the details of analyses, especially when the models are getting complex, rather than trying to identify features that are common to most analyses and interpretations. In the case of the southern bluefin tuna, all analyses indicate a substantial reduction in spawning stock biomass over time and considerably lower recruitment in recent years than 20 years ago. Also, many analyses suggest that F on the older ages is now lower than it was 15 years ago, but such a decrease is not detected for the younger ages. It would therefore seem beneficial to reduce F on the younger ages in order to allow a greater proportion of the currently low recruitment to survive to the parental stock.

14 Quality and format of the Report to the Commission

14.1 Format

The content of the Report to the Commission by the Scientific Committee as proposed in the Report of the Scientific Process Workshop is specified in Attachment N of that report and shown here for reference:

- Introduction
- Review of the Fishery
- New Scientific Information
- Current Management Measures
- Status of the Stock
- Management Implications
- Review of ERS Report
- Matters Referred by the Commission
- Advice and Recommendations

The report content, as specified, is a fairly complete representation of what is typically included in fishery management reports. However, some other report components might be considered:

- A brief review of the life history: distribution, growth, spawning time and areas, etc. as well as a brief history of the fishery.
- Biological reference points.
- Specification of management control rules as defined by Commission.
- Projections under management specified control rules.

While these components conceivably could be subsections of report sections already specified (e.g. under Current Management Measures, or Status of the Stock), we feel their importance should be stressed.

A biological reference point refers to a critical statistic on the stock that can be used for comparison with summary statistics to reflect the status of the stock. Comparing these reference points such as Bmsy, B0, Fmsy, and Fmax, (which represent theoretical points) with current stock statistics as well as with points available for other fisheries broadens the context in which fisheries management decisions can be made. While caution should be exercised in not oversimplifying the system, reference points are useful for characterising expectations about the system.

A management control rule is simply a quantitative specification by management of how the total allowable catch (TAC) should change under different stock conditions. For example, the TAC may be specified as a fixed percentage of the available biomass, as a limit on the fishing mortality rate, or as a constant catch for a certain number of years followed by a fixed increase. Such quantitative specifications of how management, the fishery, and the harvest will respond to estimated changes in stock biomass allows stock assessment scientists to explore the performance of the specified control rule under current hypotheses about stock dynamics. This way, managers can explore various management options and their consequences under a variety of scenarios. If the true behaviour of the stock is close to that specified by the stock assessment models, then projections based on these control rules should be informative to management. If the stock does not respond as expected then information is gained, that is learning occurs, and the stock assessment and management control measures should be modified based on that information.

For completeness and reference, the control rule, though specified by the Commission and not the Scientific Committee, should be included in the Scientific Committee report. The Scientific Committee might recommend different control rules for the Commission to consider given the Commission's objectives for the stock. Projections under the control rules then provide the best approach for synthesising the available information for use in understanding the potential consequences of management actions. Projections always involve uncertainty, but managers should be able to take into account this uncertainty in formulating policy. Even where alternative hypotheses exist about stock behaviour, stock assessment scientists and advisors should be able to describe the consequences of different management actions under each hypothesis.

A decision table where possible options and possible outcomes are outlined is one way of doing this. For example, if the control rule specifies that a constant proportion of the adult biomass (20%) is allowed to be harvested, and if there are currently two hypotheses about how the stock behaves (that is constant squares vs. variable squares CPUE), then the estimated stock biomass for the coming year would likely be different under the two hypotheses. As a result, the TAC (defined by the control rule as a proportion of the stock biomass) would be different for each hypothesis. Results from projections of each stock scenario could then be considered under the TAC for each hypothesis as shown in the table below:

Probability of stock reaching 1980 SSB by 2020
(a hypothetical example)

	If constant square model is true	If variable square model is true
TAC* under constant square	XX%	YY%
TAC* under variable square	ZZ%	XX%

*Computed as a constant proportion of biomass

If the TAC as applied corresponds to the correct set of scenarios, then the probability of reaching the management specified reference value by 2020 is XX% for both cases. If the real state of nature is contrary to the hypothesis assumed, the probability differs. Similar tables could provide information on the projected yield from the fishery under each scenario as well as the likelihood that the model being considered represents how the stock truly behaves. Thus the decision-makers could see the gains and loses associated with decisions available and can act accordingly.

For southern bluefin tuna, decision tables applied to key stock scenarios and management options might provide a concise outline of management options and their consequences. Such an approach might be especially useful in situations where a single consensus assessment is not achieved, and would definitely provide more information to the decision-makers than is available through the current procedure of averaging multi-model results.

14.2 Quality

The format employed by the Scientific Committee in their report to the CCSBT is consistent with that proposed in Appendix N of the Scientific Process Workshop Report. Unfortunately, the content of the report does not appear as informative as one would hope and does not contain what we would consider to be all the information necessary to make fishery management decisions.

The Scientific Committee has made some progress this year in deriving consensus opinions from their deliberations, and the rapporteurs have clearly worked hard at capturing the essence of the discussions. But in the end, only a few statistics describing stock status come through in the report and it is very difficult to discern from this report how the status of the stock is changing. Important issues, such as the concern over non-member catch, are clearly represented, but measures of stock status are somewhat obscured. In our experience, it is critical that managers see measures of stock trends over time. This could be achieved by providing standardised CPUE measures, or outputs from the stock assessment models. This could have been provided in graphs or tables showing recruitment trends as is commonly done in similar organisations, which appeared to be robust to many model formulations, or spawning stock biomass trends, which were more varying but adequately showed the decline over the history of the fishery and gave some indication of the uncertainty across models. Such presentations would provide not only a historical context for the status of the resource, but would also give managers a chance to see the scale of variation (and therefore uncertainty) involved with the assessment.

It is recommended that graphs and tables be included showing not only trends in catch and effort, but also trends in standardised CPUE, recruitment, and spawning stock biomass. Predicted trends in yield and abundance would also prove useful, although predicted relative changes in biomass, recruitment, and other stock status indicators might be preferred.

15 Achieving consensus

Stock assessment and fisheries management are complex processes subject to uncertainty. In science, uncertainty is addressed by observation, generation of ideas, formulation of hypotheses, and by critically testing hypotheses. Management addresses uncertainty by examining likely costs and benefits of decisions based on the best available information, choosing an action, and examining the results of that action. The global objective is to understand and make the best use of the resource, but the scientific objective must necessarily focus on understanding while management must necessarily focus on best use. Science must remain objective, while management must necessarily incorporate value. When the two approaches get mixed difficulty ensues, and it is of the utmost importance to clearly distinguish the scientific process from the fisheries management process and decision making.

Having separate venues for discussion of estimation and then implementation is a good first step. It may be important to go a step further and have one set of people conduct the scientific discussions, while having a second set receive the scientific advice and formulate the implications and management advice. It is extremely difficult for a single individual to operate objectively in both the science and the management venues. The scientist should not be limited by management values in developing hypotheses, nor should management advice be constrained by the time and energy that went into a particular assessment. In the current situation with southern bluefin tuna, values may be clouding objectivity and time spent on the process may be clouding value. In no small degree this is affecting the amount of information being passed forward to decision-makers. There needs to be a clearer separation of science and management tasks.

Historically for southern bluefin tuna assessments, the line separating these tasks was not clearly defined.

As a result, the process has become confrontational and this is affecting the science. Currently, there appears to be very little trust between parties that assessments, outside their own, are objective. Review of alternative assessments need not be confrontational, but given the history of the SBT assessment process, even a clearer separation of tasks (e.g. scientific from managerial) might not bring about immediate serenity between parties. Options that may help the process are: 1) allowing more time to see if the recently implemented changes in process will make it easier to achieve consensus; 2) having the stock assessment conducted by scientists hired either permanently or on short term contract by the CCSBT, 3) having the stock assessment performed by another commission, or 4) by a set of outside experts; or 5) bringing on an independent technical panel for the transitional years to facilitate and perhaps to arbitrate the process.

The scientific expertise of the parties is not in question here, but the process they are working under is problematic for all concerned. Stock assessment science should be neutral. We observed that this ideal is not always reached in the SAG/SC process, and on some occasions, participants behave more like advocates of a particular national point of view rather than neutral analysts trying to find out what the data were saying. Steps should be taken to ensure that only scientific issues are discussed either directly or indirectly in the SAG/SC process and that fishery management issues are dealt with in other fora by different individuals. This year was a first step in that direction with the creation of SAC and SC meetings, but the process needs a clearer separation of tasks. Even if this is done, it may take time to achieve equanimity. We recommend that a panel of 3-5 independent scientist be recruited by the CCSBT for a fixed three year time period. The main objectives would be to help scientists reach consensus, develop a standard informative stock status report for the Commission and produce an advisory report. The panel would have decision powers on the stock status summary and advisory reports.

References

Farley, J.H. and T.L.O. Davis. 1998. Reproductive dynamics of southern bluefin tuna, *Thunnus maccoyii*. Fish. Bull. 96: 223-236

Hilborn, R., D. S. Butterworth, J. N. Ianelli. 1998. A statistical time-series model for stock assessment of southern bluefin tuna. CCSBT-SC/9807/31.

Klaer, N., K. Sainsbury, T. Polacheck. 1996. A retrospective examination of southern bluefin tuna VPA and projection results, 1982-1995. CCSBT/SC/96/31.

Laurec, A. and Shepherd, J.G. 1983. On the analysis of catch and effort data. J. Cons. Int. Exp. Mer. 41:81-84.

Marshall, C.T., Kjesbu, O.S., Yaragina, N.A, Solemdal, P., and Ulltang, Ø. 1998. Is spawner biomass a sensitive measure of the reproductive and recruitment potential of Northeast Arctic cod (Gadus morhua)? Can. J. Fish. Aquat. Sci. 55. P. XXXX

Pope, J. G. 1972. An Investigation of the Accuracy of Virtual Population Analysis Using Cohort Analysis. ICNAF Research Bulletin No. 9. 10 pp.

Trippel, E.A. 1998. Egg size and viability and seasonal offspring production of young Atlantic cod. Trans. Am. Fish.Soc. 127: 339-359.

Tsuji, S. and Y. Takeuchi. 1998. Stock assessment and future projections of southern bluefin tuna – 1998. CCSBT-SC/9807/27

Annex 1: Brief biographies of the Panel members

Jean-Jacques Maguire

Jean-Jacques Maguire earned his M.Sc. in biology from Laval University in Sainte-Foy, Québec in 1984. He worked for the Canadian Department of Fisheries and Oceans from 1977 to 1996, in Dartmouth Nova Scotia, Québec City, Québec, and Mont-Joli, Québec. Mr. Maguire has led stock assessment teams in DFO and participated in stock assessment review processes on both coasts of North America in both Canada and in the USA, in the International Council for the Exploration of the Sea (ICES) and in the International Commission for the Conservation of Atlantic Tunas (ICCAT) whose bluefin tuna working group he chaired. He chaired both the Pelagic and the Groundfish subcommittees of the former Canadian Atlantic Fisheries Scientific Advisory Committee before chairing its Steering Committee. Mr. Maguire is now a private consultant in fisheries science and fisheries management and he is the current chair of the Advisory Committee on Fisheries Management of the ICES. His expertise is in the area of groundfish, pelagics and large pelagics fish stock assessment and fisheries management.

Patrick J. Sullivan

Patrick J. Sullivan earned a B.S. in biological science in 1979 from the University of California, Davis, and earned a M.S. in fisheries biology in 1986 and a Ph.D. in biostatistics in 1988 from the University of Washington. His experience in fisheries science includes work as a statistician modelling acoustic sampling data for the U. S. National Marine Fisheries Service. For the last ten years, however, he has been a population dynamicist with the International Pacific Halibut Commission, and is currently Assistant Professor of Population and Community Dynamics in the Department of Natural Resources at Cornell University in Ithaca, NY. He was an Invited External Scientist to the Scientific Committee on Southern Bluefin Tuna in 1992 and 1993 and was a member of the Scientific and Statistical Committee, providing advice to the Pacific Fisheries Management Council. He is currently a member of the Scientific and Statistical Committee that advises the New England Fishery Management Council.

Syoiti Tanaka

Syoiti Tanaka studied applied mathematics at the Faculty of Engineering, University of Tokyo. He started his career on studies of fish population dynamics at Tokai Regional Fisheries Research Laboratory, Fisheries Agency, and continued his work through the Ocean Research Institute, University of Tokyo and the Tokyo University of Fisheries. His work covers from statistical estimation, through stock assessment to fisheries management. He participated in the project of the Scientific Committee of the International Whaling Commission developing the Revised Management Procedure, and also proposed a method to estimate the natural mortality rate of minke whales from catch at age data and absolute abundance data obtained from Japanese research takes under special permit. He is at present an advisor to the Institute of Cetacean Research and attends the annual meetings of IWC/SC.