Review Stock Assessment of Yellowfin Tuna in the Western and Central Pacific Ocean

Malcolm Haddon
CSIRO Marine and Atmospheric Research
Castray Esplanade
Hobart 7001
Tasmania
Australia

Malcolm.Haddon@csiro.au

Executive Summary

The 2009 stock assessment document for yellowfin tuna in the Western and Central Pacific Ocean was reviewed. To do this appropriately it was necessary to also read through the 2007 assessment and a few other documents.

The authors of the 2009 assessment have made a real effort to pre-empt critical review by including diagnostics and their own critical review of the strengths and weaknesses of the assessment. They identified where the data were weakest, where the model fits were poorest, and which assumptions and structural decisions were most influential. With this list in mind they were also able to include a list of the most valuable future research and extra data gathering that could be conducted to improve the assessment. This is an excellent assessment that provides a fine example of how to present a complex assessment to a wide audience.

The data sources for the assessment were appropriate and, although there can always be more data at a better resolution and with more detail, it proved adequate to provide an assessment that can be used to assess the status of the yellowfin stock in the western and central Pacific Ocean.

Multifan-CL is the only assessment framework available that could address the complexity of the different data streams and the multitude of fisheries and gear types that make up this fishery. Its design lends itself to generating large amounts of diagnostic output and this lends itself to producing self-critical assessments. The stock assessment methods have been applied properly and are adequate and appropriate for this species and the types of data available. The assessment should be reliable but this will only really be determined through time and repeated assessments.

Many changes were made to the assessment model in 2009 relative to 2007 but these were all adequately defended and the range of sensitivities to the assumptions and structural decisions made were adequate to characterize the models performance and how these decisions and changes affected that performance. The sensitivity analyses conducted appear thorough and sufficiently complete to provide the necessary confidence in the model outputs. Further work is needed to characterize the affect of differences in the growth of younger ages expressed in different regions, in addition, the affect of the new approach to calculating the reproductive potential on the spawning biomass performance measures should be expanded.

A wide range of stock performance measures such as $B_{current}/B_{MSY}$ or $B_{current}/B_{current,F=0}$, were provided and the performance of the fishery summarized in phase plots of some of these ratios. This is a positive direction to go in, but if a decision has to be made about which measures to adopt or to move to then it is recommended that the management decisions that might derive from using the alternatives be considered retrospectively for a number of years so that an informed decision can be made that can be agreed to by all members of the WCPFC.

No projections were made in this current assessment and if such projections are required in future, then these should be specified in some detail so that the implications of specific

scenarios can be examined, rather than generalized outputs that may not be particularly informative.

The authors provide a detailed list of research recommendations but a shorter list would include a consideration of the integrity and accuracy of the varied catch and effort time series. In addition, the methods used to standardize the longline catch rate data and the relationship between longline catch rates and yellowfin tuna abundance would benefit greatly from closer examination. Finally, more work is needed to characterize the growth of the younger age classes across the regions and the means for including that in the assessment examined.

Background

Statement and History of the Problem

Background

Yellowfin tuna (*Thunnus albacares*) comprise a significant component of the tuna fishery undertaken in the Western and Central Pacific Ocean (WCPO). The fishery is characterized by spatial heterogeneity in the catches (most is taken in Region 3; see Figure 1), as well as numerous nations and fishing methods participating. Each of these factors adds complications to the data collection. This is a relatively large fishery with catches varying from 370,000 – 440,000 metric tonnes since 2000.

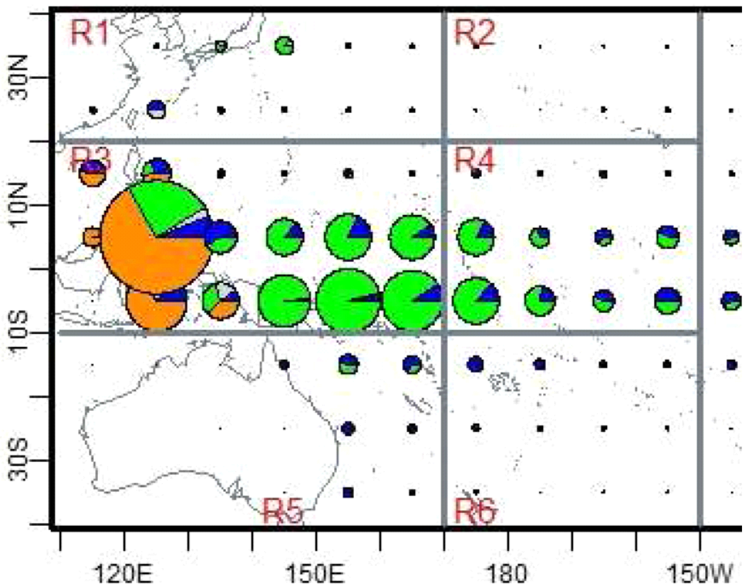

Figure 1. The spatial structure used in the WCPO Yellowfin stock assessment with an indication of the relative catches (copied for the convenience of those reading this review, directly from Langley et al., 2009).

The first formal stock assessment was conducted in 1999 with annual assessments following up until 2007. The 2009 assessment, the subject of this review, is an expansion/extension of the 2007 assessment with an update on available data but also with some important structural and parameter changes. The assessment framework is in MultifanCL (Fournier et al, 1998; described formally for yellowfin by Hampton and Fournier, 2001), which for yellowfin tuna uses a spatially disaggregated, length-based, age-structured population model to describe the stock dynamics (six regions, quarterly time periods from 1952 – 2008, 28 quarterly age classes, 95 2-cm size classes, and 24 fisheries – defined by region, gear and nationality). The data available consist of catch, effort, length-frequency and weight frequency data from the various fisheries identified as contributing to the yellowfin fishery. There are also tag-release-recapture data used to inform aspects of the assessment. The data are stratified by fishery, region, and quarterly time period. Data are available for the various fisheries in different amounts and with different levels of detail.

The 2009 assessment uses the same spatial and fishery structure as used in the 2007 assessment. The catch, effort, and size composition data were all updated to include more recent data but there were also significant changes to the model inputs. These changes were varied and included changes to the catch history for the purse seine fisheries and the Philippines fisheries, refinements to CPUE indices for the longline fisheries, an alternative natural mortality at age schedule, and an alternative estimate of spawning fraction. In addition, some structural changes were also considered, such as alternative relative weights attributed to the likelihood contributions from longline CPUE and longline size-frequency data, the presence or absence of effort creep in the long-line fisheries and alternative steepness parameters defining the spawning-stock – recruitment relationship. The selection of the different possible combinations deemed most plausible was set as the base case and the various alterations to the 2007 model were introduced as alternative scenarios so that comparisons could be made.

An important difference between the 2007 and 2009 assessments is that the 2007 assessment attempted to estimate steepness of the spawner-recruit relationship (usually a very difficult thing to do with any accuracy) whereas the 2009 assessment fixed steepness at different given values, with a base case value of 0.75. The importance of this is that the current assessment is markedly more optimistic than the 2007 assessment and this is largely due to the steepness value increasing from the estimated 0.62 to the fixed 0.75. In the 2007 assessment there was a 47% probability that there was overfishing ($F_{current}$ was larger than F_{MSY}) was occurring (Langley et al., 2007). However, the 2009 assessment concludes that the stock is neither overfished nor that overfishing is occurring.

Review Activities

The documents listed in the first part of Appendix 1 were received on 11th August 2010. The other documents listed in Appendix 1 were obtained from a variety of sources but especially from the WCPOFC web site. While the review documents were limited to the 2009 assessment and the structural sensitivity analysis of the 2007 assessment, for a more informed review the other documents in Appendix 1 were also of value in understanding the changes that had occurred between the two assessments and the context in which these assessments were conducted.

DISCLAIMER

The information in this review has been provided by way of review only. The author makes no representation, express or implied, as to the accuracy of the original information and accepts no liability whatsoever for either its use or any reliance placed on it.

Summary of Findings

The authors of the 2009 assessment have made an excellent effort to pre-empt any critical review. They have provided a detailed breakdown of the weaknesses and strengths of the current assessment, have identified where further work would be beneficial, and compared the various choices made during the assessment. While constraints of time mean that all assessments can be improved, this one provides a good example of how to present a stock assessment.

1. Comment on the adequacy and appropriateness of data sources for stock assessment.

This is a complex assessment that combines information from six regions and 24 fleets or fisheries, which are defined by region, gear used and nationality. Obtaining representative data from such a diverse and dispersed fishery must present extremely difficult operational problems.

The data available consist of catch, effort, length-frequency and weight frequency data from the various fisheries identified as contributing to the yellowfin fishery. In addition there is tag release - recapture data that informs the movement and growth estimates within the assessment. All data are stratified by fishery, region, and quarterly time periods. Data are available for the various fisheries in different amounts and with different levels of detail. In some cases operational data appear to be available on a set by set basis whereas in other cases only aggregate data at a 5º x 5º block are available. It is striking that the total catches included in the model can be lower than the summation of total reported catches (Langley *et al.,* 2009, Figure. 4). It is reported that this is because of difficulties in separating some of the aggregated catch estimates. Clearly there would be benefits in collecting even aggregate data at a finer spatial resolution (though it is understood that there might be difficulties in getting domestic policies to match the needs or wishes of this multi-national fishery; Langley et al., 2008).

Catch per unit effort data can be influential in any stock assessment but with such a complex fishery with so many components and contributors, there is a strong need to standardize this data in an attempt to remove the effects of such factors as who is fishing with what particular gear. This is done for the principal longline fisheries but it would undoubtedly be helpful to the assessment to be able to do this for some of the other significant fisheries. No doubt the WCPFC is attempting to improve the capture of operational data from other fisheries, and these attempts should be encouraged and should continue.

Some of the variations between the 2007 and 2009 assessments related to the input data. Most importantly, an alternative data set for the purse seine catches has been developed based on work that demonstrated that previous catch estimates, especially in Region 3 (the most important for catch), were biased low. Even though there remained considerable uncertainty with respect to this new catch series the corrected catch series were considered to be more reliable than the original catch series as used in the 2007 assessment and so the amended series were used in the base case for the 2009 assessment. The work to revise and correct such data series should be applauded and continued but the implications also need to be examined carefully. Contrasting the residuals from the purse seine fishery in region 3 between the 2007 and 2009 assessment (Figure 2) demonstrates some differences but considering the other changes in the assessment there is a close resemblance with an increasing variance in more recent years (although there are some wider residuals around 1990 that do not appear in the 2009 assessment).

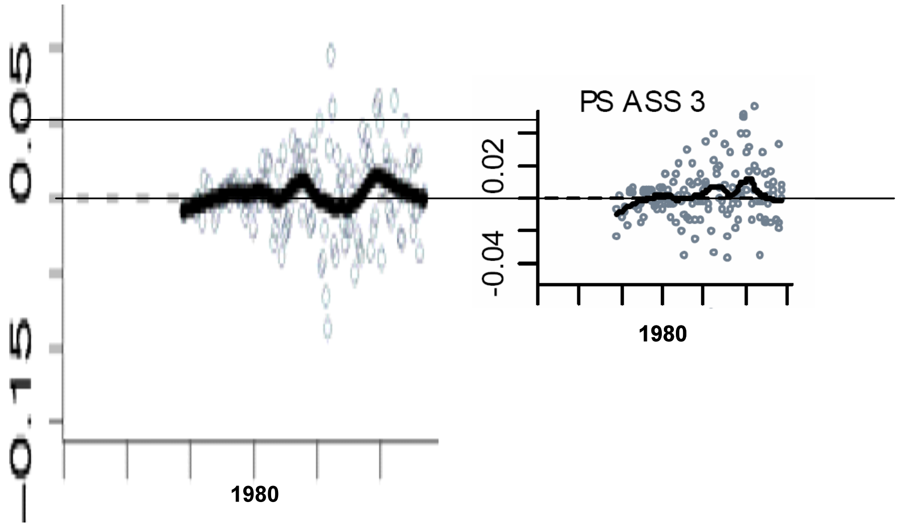

Figure 2. A comparison of the total catch residuals from the 2007 assessment (left panel) with the 2009 assessment (right panel). The 2007 graph has been stretched along the y-axis to give it approximately the same scale as the 2009 assessment graph.

While all regions are important, the scale of the fishery in region 3 suggests greater attention should be paid to the quality of the assessment fits to the data from region 3. Thus the presence of minor trends in the residuals of the total catch in the early longline ALL data and in the early ASS purse seine data in region 3 suggests that those data series may benefit from re-examination. It may also be beneficial to elucidate the reason behind the expanding variance in the purse seine catch residuals.

In summary, the data sources used in the assessment are definitely appropriate and from the perspective of being able to generate an internally consistent assessment (mostly) they are also adequate. However, the work on the purse seine catch time series demonstrates that there are likely to be real benefits from continued examination and review of the available data. At the same time, if operational data can be collected for more fisheries this would become advantageous for the fishery into the future once time series of such improved data are developed. This is clearly a long term process that needs to be institutionalized.

2. Review the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data.

The underlying assessment framework is the Multifan-CL software. Given the complexity of the internal structure of this fishery there are no other frameworks that could permit the inclusion of so many of the different fisheries with so many different gears, and include the spatial structure implicit in the operation of the fishery. It might be possible to produce an assessment using simpler models, for example, by using an age-structured production model

(Nishida, 2008). However, it would not be a simple matter trying to include the full diversity of the Western and Central Pacific Ocean fishery in such a framework. The Multifan-CL assessment framework appears to be the optimum approach at present in that it can include the various fisheries and their varied data and even include tagging data results that contribute to estimates of movement and growth within the assessment framework. By including all of these estimates within the single estimation framework, interactions between uncertain values can be combined and the whole should provide better estimates of the intrinsic uncertainty in the assessment. The real value of this is that there should be less chance of being mistakenly over-confident in any conclusions.

The current strategy of selecting a base case and presenting numerous sensitivities to changing the various options should be continued. In addition, the structural sensitivity analysis (Harley et al., 2009) provides a valuable exploratory tool for examining the implications of different choices and their potential interactions. This may involve a large commitment in computing time but it is recommended that this approach be continued as it is an efficient way to discover the most influential components in the assessment and acts as a guide to further work on improving data sets or subordinate analyses.

There are many model configurations possible within the Multifan-CL framework and the selection of the base case scenario is obviously a challenging task. The eight main model options investigated in the 2009 assessment (plus the two options trying to mimic the 2007 assessment), represent a reasonable attempt to characterize the effect of the selection made across the major options in the model. All share the steepness = 0.75 option, and the alternative correct purse seine catch series, although one option explored the effect of using the old time series of purse seine catches. There is also explicit consideration of the effect of varying the weightings applied to longline catch rate series and size-distributions, and in addition there was the option of including effort creep or not (implemented by increasing the catchability coefficients). In addition, all of the important influences and changed factors were explored for their relative effects on the assessment. By examining a wide range of different combinations of the available options for input data, model assumptions, and structural variations, the behaviour of the model in the face of the data and the uncertainty over details of model implementation are well characterized.

The authors of the assessment report demonstrate an excellent working knowledge of how to implement a Multifan-CL assessment so it is simple to conclude that the assessment was properly applied and appropriate for the species. The robustness of the outputs to an array of changes also indicates that the assessment is reliable but this can only really be determined by conducting a Management Strategy Evaluation. However, to conduct a MSE simulation for this fishery would be a huge undertaking and there are many alternative analyses that could be pursued rather than this.

Finally, it is asked whether the assessment methods are adequate for this species. This is difficult to answer except to say that no other method is available that can provide such a comprehensive assessment as Multifan-CL in the face of such complexity and such varied data inputs, so in that fashion it is clearly adequate. But of course, assessments can always be improved. The surprising outcome with the current assessment is the major change in the model conclusions relative to the 2007 assessment. With all the changes there remains little evidence that the stock is either overfished or that overfishing is occurring. Figure 57 in the assessment is a valuable graphic (the phase plot of F/F_{MSY} against SB/SB_{MSY} for the numerous model sensitivities) in that it enables one to grasp the effects of different combinations of the many options considered in the current assessment. In none of the combinations was there an indication that the Spawning Biomass to SB_{MSY} ratio was below one (in other words there was no evidence that the stock was overfished). The only combinations that suggested there was a chance that overfishing might be occurring were those sensitivities that included a steepness of only 0.55 (less than the 0.62 estimated in the 2007 assessment), but even with that circumstance the ratio of $F_{current}$ to F_{MSY} was only just greater than 1.0.

The justifications for altering the inputs and structure of the 2007 assessment are dispersed through the 2009 document but nevertheless, the effect of the changes are tested appropriately and the impacts of the changes are discussed and explained in sufficient detail to see that the authors have the necessary evidential support or adequate arguments for their choices. An example is where the effect of the selection of the steepness value is explored (e.g. Figure 58). It is clear that setting the steepness to a particular value leads to relatively tight likelihood profiles over stock status indices such as $B_{current}/B_{MSY}$. By generating composite likelihood profiles the approximate effect of uncertainty in the steepness parameter can be indicated. The authors appear to have tried hard to pre-empt criticisms that could have been leveled if they had not explored these effects. It is noteworthy that the results of 2009 assessment are markedly different to those from the 2007 assessment. This does not mean that the assessment is not reliable because the authors have made the effort to explain the sources of the differences and justify acceptance of the changes made to the modelling and data.

In short, the stock assessment methods have been applied properly and are adequate and appropriate for this species and the types of data available. Whether the assessment is reliable will really only be determined through time and repeated assessments.

3. Evaluate the assessment model configuration, assumptions, and input data and parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for.

By examining structural and input sources of uncertainty the authors of the 2009 assessment appear to have tried hard to pre-empt criticism about there implementation of this assessment framework. This is a very well executed stock assessment.

Because the steepness parameter is so influential on the outcome more discussion could have been provided over the choice of the array of fixed steepness value chosen. However, the range of values covers the plausible range and their attempts to generate combined likelihood profiles of the main management reference points across the given values of steepness is a reasonable compromise and is preferable to attempting to estimate steepness from the available data. The estimated annual recruitment time series (Langley et al., 2009, Fig. 37) indicates highly variable recruitment across regions with some unexplained extreme events in some regions early in the fishery. It is not surprising that the 2007 estimates of steepness were so poorly determined (Langley et al., 2007, Fig 50; the fact that the outcome then was so influenced by the prior used is also of concern when trying to characterize the full uncertainty involved in this estimate). The decision to move to an array of fixed steepness values rather than attempt to estimate steepness is a reasonable step in the face of the uncertainty associated with trying to estimate its value. However, there is no attempt to explain why 0.75 was considered a “moderate value of steepness” when the mode of the estimates was 0.62 (or lower depending on which prior was used). Fortunately, it is only the lowest value of steepness used (0.55) that suggested that overfishing might be occurring. When the combined likelihood profiles are examined none indicate fishery problems.

The fixed array of natural mortality with age may appear unusual in that it starts high, drops as might be expected but then rises again before dropping again. However, this is consistent with the observed change in sex ratio as the fish age and is explained by increased mortality of females with the onset of maturity. This appears like a sensible explanation but it would be more convincing if evidence (perhaps from tag return rates by gender) could be gathered that related to this issue. The authors have included a consideration of the potential impact of higher natural mortality on the youngest ages, which was identified as one of the key uncertainties in the assessment. The analyses conducted and comparisons made between the alternative natural mortality schedules again cover plausible values and identify their effects.

Once again the main models considered along with the larger array of combinations of the various inputs and data alternatives demonstrate the relative effects and the interactions that occur among the array of model variants. Given the past experience and stock assessments of this fishery, the selected input parameters appear reasonable.

The model configuration chosen leads generally to internally consistent outcomes that mostly match expectations from what is known of the biology. However, that is not to say there are no problems with the fit of the model to all data streams. There are some significant deviations between some of the predicted time series and some of the observed data. The authors, however, provide a detailed description of such deviations/failures in Section 5.4 “Fit diagnostics …. for the base case 2009” as well as in the Discussion and Conclusion section, and they appear to have identified all the important sources of internal inconsistency.

A possibly important issue for which no complete solution is presented is the failure of the model to adequately fit the smallest size modes in some of the longline fisheries and also in other fisheries (Langley et al., 2009, Fig 19). This appears to be due to the estimates of the growth of the younger age classes being biased high, especially in region 3, the most important in the fishery. However, as it is currently structured it is not possible to include heterogeneous growth between regions in the model. The authors suggest conducting a separate assessment for each region as one potential solution although presumably this would ignore movement between areas and other spatially important components of the fishery. It is recommended that this be explored further attempting to account for movement among regions. Possibly some regions can be considered together, as it appears that it is primarily region 3 that exhibits different (slower) growth, however this may be because the growth expressed in other regions was not examined or plotted.

There is discussion of the tag data and details are given of various biases, but with such spatial structuring the fits to data appear reasonable. What is missing, however, is any commentary on what can be done about how four of the estimates of tag recovery rates bump up against the upper limit allowed (Langley et al., 2009, Figure 36). This may be due, as stated, to the small number of returns from those fisheries but the affect of this misfit should be explored or at least commented on in more detail.

Another place where an obvious mismatch between the model and the data occurs is with the estimate of median fish length in the Japanese purse seine fishery (Langley et al., 2009, Figure 20). As identified in the assessment document, this is clearly due to a major change in fishing behaviour rather than a change in the stock. As a data stream that would benefit from further exploration this is an obvious candidate. Another possible option would be to treat the Japanese purse seine fishery as if it were two fleets one from 1970 to about 1985 and the other from about 1985 to the present. Imposing time blocking on time series of data contains the danger of making ad hoc changes to improve model fits; however, in this case, the change is so marked and consistent that there was clearly some large change that has not been captured by the current description of the fishery. Bubble plots of residuals for the fit to the length frequency data (Figure 21) are useful for following trends in residuals but are clearly unnecessary for the Japanese purse seine fishery. Another advantage of using Multifan-CL is that there is a plethora of diagnostic output and related plots.

In general, the model assumptions and parameterization (as described in Section 4: "Model Description – structural assumptions, parameterization, and priors", provides a clear description of what the modellers did and how they did it. The assumptions about the stock-recruitment relationship are clearly important and the decision to impose only a weak penalty on deviations from the Beverton – Holt relationship seems to be a reasonable solution to the problem of implementing such a relationship in a spatially disaggregated model. The inclusion of movement can only be done in a limited manner (assuming that movement patterns are invariant with age – which intuitively seems like a poor assumption); however, without more and extensive tagging data there is little else that can be done about that.

The change adopted in the method of calculating the spawning biomass is an attempt to move to a statistic that should more closely indicate the true spawning biomass. However, it is not clear how the estimates of spawning biomass are influenced by the "substantial reduction in the reproductive potential for older age classes relative to the values used in the 2007 assessment" (p13). The previous method of calculating the reproductive potential used the relative biomass of both sexes above the age of female maturity. With the sex ratio of the larger fish changing towards a male dominated ratio the new method of calculating the reproductive potential is a sensible change, but it would be useful to have more discussion as to the implications of this change. This may be an area where further research may be warranted.

4. Evaluate the adequacy of the sensitivity analyses in regard to completeness and incorporation of results.

The envelopes surrounding the predicted management reference points appear as reasonably smooth distributions of outcomes. The patterns appear clear and do not exhibit major gaps or unexpected shifts or changes. Again Figure 57 (the phase plot of F/F_{MSY} against SB/SB_{MSY} for the numerous model sensitivities) provides a clear demonstration that the sensitivities cover the space of variation within the model and also provides some insight into how the various combinations interacted on the outcomes.

The attempt to include an approximate mimic of the 2007 assessment was an excellent idea. Given the numerous changes made to the input data and various constants it would not be possible to step change by change from the 2007 assessment to the 2009 assessment. Nevertheless, the strategy adopted gives the reader insight into how the changes altered the outcomes from the assessment.

In some of the sensitivity analyses the weights attributed to the different sources of likelihood were modified. As correctly identified by the authors the final likelihoods of these variants are not directly comparable to those models which were not changes structurally. It would be helpful to identify those models which were directly comparable and those which are not in the graphical and tabular representations (for example in Table 7). But this is a minor point.

In brief, the sensitivity analyses conducted appear thorough and sufficiently complete to provide the necessary confidence in the model outputs.

5. Comment on the proposed population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status.

The authors present a wide range of stock performance measures and their ratios, which is good practice as long as their individual implications are understood by those making decisions about the consequent management. As is often the case, the ratios are less variable between model variants than the absolute values and, as the authors point out, those ratios based on one current value and one equilibrium estimated value tend to be more variable than those produced from two equilibrium estimated performance measures. Nevertheless, ratios such as $B_{current}/B_0$ or $F_{current}/F_{MSY}$ are reasonably well understood and their implications appear meaningful to a wide audience. Such more traditional measures should continue to be used but not to the exclusion of other, potential more informative measures.

Given the short lived nature of yellowfin tuna and the variable recruitment that appears to happen in different regions the ratio $B_{current}$ with $B_{current,F=0}$ is also of real interest as potentially being a more representative measure of the current state of the stock. As a measure of stock depletion it may be more meaningful than the more usual $B_{current}/B_0$ or $SB_{current}/SB_{MSY}$. As this fishery seems less likely ever to achieve anything other than a long term equilibrium then the notion of B_0 or even B_{MSY} may give an artificially optimistic view of the potential production, especially if the trends predicted in the recruitment are real and continue downwards. It would be useful to continue to present the $B_{current,F=0}$ measure and examine its use in more detail. As with all of these ratios and performance measures, once the question of how different management would be under one or the other ratio or measure can be determined, an informed decision can be made as to which to adopt as a guide to management.

The authors use phase plots (or Kobe plots in the Tuna world) of F/F_{MSY} against SB/SB_{MSY} to illustrate the stock status to good effect. This is both with the optimum model (as in Figure 55) and with the array of sensitivity model runs (as in Figure 57). There may be some value in adding the 95% confidence intervals derived from the likelihood profiles across the steepness parameters to highlight the degree of uncertainty in the analysis, but otherwise these are useful tools for illustrating current status in a simple, summary manner. There may be some value in plotting up in a similar way some of the other combinations of ratios possible.

It is simple to suggest a diverse array of possibilities but in the end it is best to limit any exposition to those summaries that capture how the stock has been changing through time. There is value in consistency through time in what measures and ratios of measures are used. These should continue to be used while exploring how the relatively new non-equilibrium measures and ratios would perform is they were used instead.

6. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status.

In any spatially disaggregated stock assessment model, future projections for conducting risk assessments of alternative management options are made much more difficult by the need to include solutions to the spatial distribution of recruitment and of effort. In this case this is even more difficult because of the multitude of different fishing gear used among the different fisheries.

In this assessment no specific projections were made of the future implications of alternative management options. This may reflect a difficulty that the assessment scientists faced in the 2007 assessment. The authors (Langley et al, 2007) appear to have been further hampered by a need to conform to WCPFC conservation and management measures (specifically that effort should not increase above current levels and that purse seine effort should not exceed either 2004 levels or the average of levels from 2001 – 2004. The authors made a series of assumptions that match the constraints but also account for the recent increases in Philippine and Indonesian domestic fisheries as well as accounting for those fisheries where catchability was variable. To solve the recruitment problem the authors used the estimated stock recruitment relationship and the recruitment produced was spread among regions based on the long term average distribution. This is a very difficult thing to predict and electing to use the average distribution seems reasonable. However, later statements (e.g. "if recruitment remained distributed among regions in accordance to the recent pattern of recruitment then the probability of the stock size falling below the $\tilde{B}_{MSY}$ level would be greatly increased." (p30) suggest that the average distribution of recruitment might fail to capture the usual variation in recruitment that occurs.

All of these assumptions appear reasonable, however, as the authors state, the particular outcome of the stock projections would be very sensitive to variations in the assumptions listed. The description of the projection makes it appear as if the projection was conducted in a deterministic fashion. I assume that because the projections are very sensitive to the assumptions used, that had an array of options been trialed and the results combined in an effort to capture the real uncertainty, then the result seems likely to have been uninformative. In essence, the projections were saying that if recruitment returns to the long term average and effort remains about the same then the stock biomass is likely to move towards B_{MSY}. What is not being said is that there is no knowing whether that is what recruitment will do, especially as the long term trend in recruitment appears to be downward.

If forward projections are so uncertain and yet they are still wanted, it may be more useful to those managing this resource if those managers specify the scenarios whose implications they would like to know about. For example, if the trend in recruitment continues downwards at the same rate as recently, what would be the implications?

7. Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices.

As with all complex fisheries this is a difficult question to answer; however, once again the authors have made an excellent effort throughout the document to identify areas of weakness that could benefit from further research effort. Their 13th conclusion on page 35 provides a detailed list of where the assessment would benefit from further research and what further data might be collected (e.g. age frequency data from the commercial catch) to improve the assessment.

Selecting from among the possibilities remains difficult but one guide is in the statement that "The biomass trends in the model are strongly driven by the time-series of catch and GLM standardised effort from the principal longline fisheries." p2.

Firstly, there have already been important changes made to the stock assessment through reviews of important time series of catch data. For each important time series of data it would be useful to know whether there are further reviews possible or whether the catch and effort history available can become the agreed history. Given probable difficulties in funding this would not entail actually conducting the further reviews, but a first step would be determining if such reviews might even be possible.

Secondly, it is disappointing that even though the longline catches only make up less than 20% of the catches (and a much lower percentage of the numbers caught) it is their catch rates that inform the assessment. This is disappointing because in one of the main longline fisheries (LL ALL 3) there are inconsistent signals from the size distribution data and the catch rate data. I agree with the authors that it would be useful to further investigate the methods used to standardize the longline CPUE data and also the relationship between standardized longline CPUE and yellowfin tuna abundance.

While the model diagnostics are generally remarkably good (for such a diverse and complex fishery) there remain some standout failures of fit (to some of the size frequency data) that indicates that a single growth curve for the geographical extent of the WCPO yellowfin fishery is a poor assumption. That is, the assumption that the length at age and the weight at length relationships are spatially invariant are known to be false. Even though currently it is not yet possible to include heterogeneity of growth into the assessment model it would be useful to conduct geographical specific growth studies to determine the real extent of this issue. The primary issue appears to be with juvenile growth, so movement may not be such a large problem. Once the scale of this problem is known, then options for including this into the full assessment might be suggested.

As with all suggestions for further research, there needs to be a consideration of the relative value of such further work to the assessment and the likelihood of obtaining sufficient resources to conduct the work and complete it successfully. All of the suggestions made in the 13th conclusion might be valuable but some will be simpler to implement than others.

Conclusions/Recommendation

1. Comment on the adequacy and appropriateness of data sources for stock assessment.

The data are complex and varied but there are clearly some problems that still need to be addressed. There would be value in reviewing the various data series, as has been done for the purse seine fishery catches, to increase confidence that all that can be done has been done. The work to revise and correct the various time series of catches is applauded and should be continued and expanded. In addition, initiatives aimed at improving the detail and resolution of data as it is currently collected should be pursued. Similar agencies such as CCAMLR have solutions to issues of confidentiality; if that is one of the problems, then such solutions may be useful to the WCPFC. In summary, the data sources used in the assessment are definitely appropriate and from the perspective of being able to generate a mostly internally consistent assessment they are also adequate (though there are places where they can be improved).

2. Review the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data.

The use of Multifan-CL as the assessment vehicle is the only option currently available that can handle the diverse array of different fisheries and data streams. It is able to generate the wide range of diagnostic statistics, which lends itself to the production of multiple graphical diagnostic plots allowing the assessment scientists to generate the self critical detailed assessment that the authors of the 2009 assessment have produced. The inclusion of the varied sensitivities testing the implications of the multiple choices that could be made over the structural form and emphasis placed on different data streams is also best practice.

In short, the stock assessment methods have been applied properly and are adequate and appropriate for this species and the types of data available. While it should be reliable this will only really be determined through time and repeated assessments.

3. Evaluate the assessment model configuration, assumptions, and input data and parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for.

By examining structural and input sources of uncertainty the authors of the 2009 assessment appear to have tried hard to pre-empt criticism about there implementation of this assessment framework.

The decision to include fixed values of steepness was a reasonable solution to the difficulty presented in attempts to estimate steepness within the modelling framework. This decision is emphasized because of the influence steepness has directly on the productivity. In addition, the growth characteristics are very influential on potential productivity. The assessment exhibited weaknesses with respect to how growth is estimated, with clear deviations from expected growth rates in important parts of the fishery. This is a significant problem that needs attention. Also, the new method of calculating reproductive potential appears to be a marked improvement over the previous approach used; however, this is an area that also needs further exploration and its implications for the model outcomes, particularly in performance measures involving spawning biomass, which need to be examined in more detail.

4. Evaluate the adequacy of the sensitivity analyses in regard to completeness and incorporation of results.

The treatment given to the sensitivities was through and adequate to generate the confidence that the model outputs were internally consistent. Including an attempt to mimic the 2007 assessment structure was also a useful tool.

In brief, the sensitivity analyses conducted appear thorough and sufficiently complete to provide the necessary confidence in the model outputs.

5. Comment on the proposed population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status.

It appears that decisions still need to be made over what performance measures to use as a summary of stock status and to provide management advice. A mixture of equilibrium based and non-equilibrium based measures were provided. Intuitively the non-equilibrium based measures, such as $B_{current,F=0}$ are attractive with this short lived, highly variable and relatively unstable species. However, the more traditional equilibrium based measures such as B_{MSY} or B_0 and their related ratios still have value in being easily understood. If a decision has to be made about which measures to adopt or to move to, then it is recommended that the management decisions that might derive from using the alternatives be considered retrospectively for a number of years so that an informed decision can be made that can be agreed to by all members of the WCPFC.

6. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status.

No projections were made in this present assessment. However, because of the great complexity of this fishery and the uncertainty that this introduces to such forward projections, if projections are still required it is recommended that these be specified in some detail by managers so that specific implications can be examined.

7. Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices.

The authors provide a detailed list of potentially beneficial research in the 13th conclusion. However, if a shorter list is required, then of immediate value and concern is the consideration of the integrity and accuracy of the various catch series. In addition, the methods used to standardize the longline catch rate data (which is so influential in the assessment) and the relationship between longline catch rates and yellowfin tuna abundance would benefit greatly from close examination. Finally, growth appears to be heterogeneous among regions in this fishery but this is not included in the assessment model. More work is needed to characterize the growth of the younger age classes across the regions and the means for including that in the assessment examined.

Appendix 1: Bibliography of Materials Provided

Langley, A., Harley, S., Hoyle, S., Davies, N., Hampton, J., and P. Kleiber (2009) Stock assessment of yellowfin tuna in the western and central Pacific Ocean. WCPFC-SC5-2005/SA-WP-03. Western and Central Pacific Fisheries Commission. Scientific Committee Fifth Regular Session. 10-21 August 2009 Port Vila, Vanuatu. (Secretariat of the Pacific Community: Noumea, New Caledonia.) 121 p.

Harley, S.J., Hoyle, S.D., and F. Bouyé (2009) General Structural Sensitivity Analysis for the Yellowfin Tuna Stock Assessment. WCPFC-SC5-2009/SA-IP-03. Western and Central Pacific Fisheries Commission. Scientific Committee Fifth Regular Session. 10-21 August 2009 Port Vila, Vanuatu. (Secretariat of the Pacific Community: Noumea, New Caledonia.) 30 p.

Other Material Examined

Fournier, D.A., Hampton, J., and J.R. Sibert (1998) MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. *Canadian Journal of Fisheries and Aquatic Sciences* **55:** 2105-2116

Hampton, J. and D.A. Fournier (2001) A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. *Marine and Freshwater Research* **52:** 937-963.

Hampton, J., Kleiber, P., Langley, A., and K. Hiramatsu (2004) Stock assessment of yellowfin tuna in the western and central Pacific Ocean. SCTB17 Working Paper SA-1. 17th Meeting of the Standing Committee on Tuna and Billfish SCTB 17 Majuro, Marshall Islands, 9-18 August 2004. 74 p.

Langley, A., Hampton, J., Kleiber, P. and S. Hoyle (2007) Stock assessment of yellowfin tuna in the western and central Pacific Ocean, including an analysis of management options. WCPFC-SC3-SA-SWP/WP-01. Western and Central Pacific Fisheries Commission. Scientific Committee Third Regular Session. 13-24 August 2007, Honolulu, United States of America. (Secretariat of the Pacific Community: Noumea, New Calidonia.) 116 p.

Langley, A., Wright, A., Hurry, G., Hampton, J., Aqorau, T., and L. Rodwell (2008) Slow steps towards management of the World's largest tuna fishery. *Marine Policy* **33:** 271-279.

Nishida, T. (2008) Preliminary stock assessment of yellowfin tuna (*Thunnus albacares*) in the Indian Ocean by the ADMB based ASPM. IOTC-WPTT-2008-28 rev. 66p.

Appendix 2: Statement of Work

Attachment A: Statement of Work for Dr. Malcolm Haddon

External Independent Peer Review by the Center for Independent Experts

Stock assessment of yellowfin tuna in the Central and Western Pacific Ocean

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org

Project Description: The Pacific Islands Fisheries Science Center (PIFSC) requests an independent review of the stock assessment of yellowfin tuna in the Western and Central Pacific Ocean (WCPO). The Oceanic Fisheries Programme (OFP) of the Secretariat of the Pacific Community, with collaboration from scientists participating in the Scientific Committee of the Western Central Pacific Fisheries Commission, is responsible for conducting the assessment. Results of the 2009 assessment indicate that overfishing of yellowfin tuna is not occurring in the WCPO, which is contrary to the previous assessment (2007). The most influential change in the current assessment was due to was due to assumptions regarding the steepness of the spawner-recruit relationship. Previous assessments used relatively low values of steepness while the current assessment assumes a moderate value (0.75), resulting in a more optimistic assessment of the stock status. The assessment provides the basis for scientific advice on the status of the stock that is provided regularly at both national and regional levels, and directly influences U.S. policy on resource utilization. The most recent stock assessment of yellowfin tuna in the WCPO was completed by the OFP in 2009, with collaboration from U.S. scientists, and three reviewers are requested to review the assessment. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have expertise, working knowledge, and recent experience in various subject areas involved in the review: tuna biology; analytical stock assessment, including population dynamics theory, integrated stock assessment models, and estimation of biological reference points; and

MULTIFAN-CL and AD Model Builder. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review as a "desk" review of the necessary documentation of the current assessment of yellowfin tuna in the WCPO, therefore no travel is required.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact will provide the CIE reviewers with the background documents and reports of the current assessment and sensitivity analyses to be peer reviewed. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Desk Review: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report addressing each ToRs in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Conduct an independent peer review in accordance with the ToRs (**Annex 2**).

3) No later than (TBD), each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, via email to David Die *ddie@rsmas.miami.edu*. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

9 August 2010	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
11 August 2010	NMFS Project Contact sends the CIE Reviewers the report and background documents
11-30 August 2010	Each reviewer conducts an independent peer review as a desk review
30 August 2010	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
10 September 2010	CIE submits the CIE independent peer review reports to the COTR
15 September 2010	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) Each CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) Each CIE report shall address each ToR as specified in **Annex 2**,

(3) The CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Contracting Officer's Technical Representative (COTR)
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel - NMFS Project Contact:

Gerard DiNardo
Pacific Islands Fisheries Science Center
2570 Dole Street, Honolulu, Hawaii
Gerard.DiNardo@noaa.gov Phone: 808-983-5397

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background and Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work

Annex 2: Terms of Reference for the Peer Review

Stock assessment of yellowfin tuna in the Central and Western Pacific Ocean

1.Comment on the adequacy and appropriateness of data sources for stock assessment.

2.Review the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data.

3.Evaluate the assessment model configuration, assumptions, and input data and parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for.

4.Evaluate the adequacy of the sensitivity analyses in regard to completeness and incorporation of results.

5.Comment on the proposed population benchmarks and management parameters (*e.g., MSY, Fmsy, Bmsy, MSST, MFMT*); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status.

6.Evaluate the adequacy, appropriateness, and application of the methods used to project future population status.

7.Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices.