(DRAFT) EXTERNAL REVIEW OF ATLANTIC YELLOWFIN TUNA ASSESSMENT IN 2019

R. Methot[1]

SUMMARY

This document is an external review of the 2019 ICCAT yellowfin tuna assessment. The data preparatory meeting was held in April 2019 and the stock assessment meeting in July 2019. The author did not attend either of these meetings. This review was conducted by correspondence in September 2019 based on SCRS reports and draft documents provided by the ICCAT secretariat.

RÉSUMÉ

Ce document est un examen externe de l'évaluation de 2019 de l'ICCAT sur l'albacore. La réunion de préparation des données a eu lieu en avril 2019 et la réunion d'évaluation des stocks en juillet 2019. L'auteur n'a assisté à aucune de ces réunions. Cet examen a été réalisé par correspondance en septembre 2019 sur la base des rapports du SCRS et des projets de documents fournis par le Secrétariat de l'ICCAT.

RESUMEN

El presente documento es un examen externo de la evaluación del rabil de ICCAT de 2019. La reunión de preparación de datos se celebró en abril de 2019 y la reunión de evaluación de stocks en julio de 2019. El autor no asistió a ninguna de estas reuniones. Esta revisión se realizó por correspondencia en septiembre de 2019 sobre la base de los informes del SCRS y los proyectos de documentos proporcionados por la Secretaría de ICCAT.

KEYWORDS

Yellowfin tuna; stock assessment; external independent review

This document is an external review of the ICCAT yellowfin tuna assessment. The data preparatory meeting for this assessment was held in April 2019 and the assessment meeting in July 2019. I did not attend either of these workshops. This review was conducted by correspondence in September 2019 based on SCRS reports and draft documents provided by the ICCAT secretariat.

This assessment was called for in the tropical tuna workplan for 2019 due to concerns that the yellowfin tuna stock may currently be overfished and undergoing overfishing. This assessment conducted by the yellowfin tuna assessment group ("Group") made substantial improvements in the data and analysis, particularly with regard to a combined fleet CPUE index, and growth rates and natural mortality rates based on age data. I commend the Group for their efforts to accomplish this improved assessment. The result shows the estimated population abundance to be shifted higher, such that the mean estimate of fishing mortality rate is below overfishing and the estimated stock abundance is above the overfished threshold. I conclude that this result is the best scientific information available regarding the status of this stock.

Due to the limited time in which to accomplish this review and the fact that I was not able to attend the in-person workshops, I have not reviewed all aspects of this assessment in detail. In particular, I have not attempted a detailed review of the catch, effort, and length composition data development, and I have not attempted a review

[1] NOAA Science Advisor for Stock Assessments.

of the JABBA model. For these items, the brevity of my report should be interpreted as an indication that the Group has employed standard practices and has made reasonable evolutionary progress on those topics. My review will focus principally on various aspects of the application of the Stock Synthesis model, and on the development of the uncertainty grid. The terms of reference for this external review include ten factors considered individually below.

TOR 1) Evaluate the adequacy, appropriateness, and application of data used in the assessment

Data used in this assessment conform to standard practice for assessment of tropical tunas. This includes a time series of catch by each major fleet, a time series of length composition data by fleet where available, and a time series of standardized catch per unit effort (CPUE) for a subset of the fleets. All data were updated through 2017 and some through 2018. The lack of composition data for 2018 limits the timeliness of projections provided by this assessment.

1. Joint longline index – The CPUE index used in the assessment is significantly improved by taking a spatial, multi-fleet approach to its construction. This has been done for other tropical tuna and billfish stocks and is a good step towards more explicit spatial assessment. CPUE is inherently a measure of the density per unit area of fish in the locations being fished. CPUE becomes an index of total population abundance only by averaging the CPUE over a consistent spatial range that represents the majority of the stock. This is achieved for yellowfin tuna by taking the CPUE averaged over the tropical zone, area 2, as the primary index of the trend in abundance of the overall stock. One of the model alternatives used the indices from all three areas. I have not closely examined the GLM used to process the CPUE data, but the logic of using the three areas as independent indices of total stock abundance should be carefully compared with the assumptions for the area term in the GLM. I concur that year interaction terms should be random effects.

2. Age – Another substantial innovation in this 2019 assessment is the use of age data to revise the estimate of natural mortality and growth. These have been validated using bomb radiocarbon and reader comparisons have produced estimates of ageing imprecision. My evaluation of the use of these data for natural mortality and growth is covered in TOR #2.

TOR 2) Evaluate the adequacy, appropriateness, and application of methods used to assess the stock and if appropriate recommend alternative approaches to be accomplished in the future

1. The three models: MPB, JABBA, and SS are all potentially useful. They use varying degrees of simplifications to model the dynamics of yellowfin tuna. MPB is the simplest, so makes the greatest assumptions about how simply the yellowfin tuna population operates. JABBA relaxes some of these assumptions by allowing for process error and required priors on productivity to achieve reasonable estimates. The need for these priors and the high correlation between r and K is not surprising because the time series has rather little contrast. SS relaxes many assumptions and brings more types of data into the analysis to calibrate relationships, like natural mortality, selectivity, and growth, that replace those assumptions. In particular, SS gains information on Z from the length composition information. For yellowfin tuna, I concur with the Group in concluding that the shift over time towards fleets that target smaller fish can bias the MPB and JABBA results and the fleet-specific SS is better able to handle this shift. I add that MPB and JABBA use the joint LL index, which is a large fish index, even though much of the catch is from smaller fish.

2. Natural mortality (SS) –The maximum observed age has increased from 10 to 18. The Group has used this revised estimate of maximum age to estimate that natural mortality is 0.35, which is lower than the previous estimate of 0.55. I note that even 0.35 may be an overestimate because the stock has been fished for a few multiples of 18 years so the recent observations of maximum observed age is probably less than it would have been if observed from an unfished stock. The Group has conducted good sensitivity analyses with regard M. I recommend work in future assessments to attempt to improve the baseline M estimate and to include a reasonable range on M in the uncertainty grid. I recommend that in the SS modeling that the maximum population age be increased to 18. SS incorporates appropriate calculations to deal with growth and mortality within the current plus group of 10, but increasing the plus group age to 18 will improve the accuracy with which older fish are tracked over time.

3. Growth (SS) – The advent of improved age data has also allowed for improved calculation of growth. The Group has appropriately included the age-at-length data in the SS model to allow for internal estimation while taking into account the effect of length selectivity and ageing imprecision. In doing this, they appropriately used only the bomb-radiocarbon validated ages plus, for young fish only, the ages from daily growth increments. The resulting estimates of growth appear accurate and consistent with the length composition data. The Group found some model instability while estimating growth and did final model runs with growth parameters fixed. This instability is not surprising given the large number of data elements in the model and inevitable sensitivity to the relative weighting among data types. They should continue to investigate simultaneous growth estimation in future iterations of this assessment, especially as the volume of age data increases.

    a. The Group configured SS to model the CV of length-at-age as a function of mean length-at-age. Unfortunately, the lower CV at older ages produces an illogical decline in standard deviation of length-at-age for older ages. An improved approach available in SS would be to model the standard deviation of length-at-age to be a non-decreasing function of age.
    b. Note that Richards coefficient is estimated to be 0.11, which is close to 0.0. Richards is undefined at 0.0 as implemented in SS and a null value (e.g. equivalent to von Bertalanffy) is a value of 1.0.

4. Selectivity (SS)
    a. Areas 1 (fleet 16 - north) and 3 (fleet 18 - south) have somewhat smaller fish and the SS modeling of fisheries in those areas finds higher selectivity at smaller sizes than longline fisheries operating in area 2. This is the areas-as-fleets approach used in other assessments. In this approach, the selectivity is with respect to the total population, not the population found in that fleet's geographic area. As more is learned about movement patterns of yellowfin among these areas, a spatially structured assessment with movement between areas and area-specific selectivity could be more accurate.
    b. The same selectivity is used for fleets 3, 4, 5, 6 which are actually just the four seasons of a single fishery. Similarly for FAD fleets 7-10. Splitting these fisheries into four seasonal fleets provides flexibility for consideration of seasonal fishery selectivity, but no investigation of seasonality was made and there is no obvious seasonal pattern to the residuals. A much simpler model configuration would use one fleet for the purse seine and one FAD fleet, each with catch in each season.
    c. For the purse seine fleet, the spline selectivity has the last knot at 145 cm, but 25% of fish are larger than this. So, this configuration imparts an asymptotic selectivity pattern. It is not clear if asymptotic selectivity was intentional or not.
    d. I note that JABBA uses the joint LL index (fleet 17) as its index of stock abundance. SS shows this fleet to have selectivity shifted to larger fish. The JABBA approach uses combined catch from all fleets, so some of the expected impact of this catch may not be tracked well by the fleet 17 index.

5. Catchability (all models) – None of the modeling approaches consider the possibility that catchability has changed over time or has a non-linear relationship too stock abundance. This is rarely addressed in assessments and a defensible alternative may not be feasible. Nevertheless, the results that have been obtained are conditional on assuming that catchability is constant over time.

TOR 3) Evaluate the methods used to estimate population benchmarks and stock status (e.g., MSY, FMSY, BMSY, or their proxies)

The methods used for estimation of these quantities conform to common practices. In JABBA, they are inherent to this modeling approach and are influenced by the priors used for population productivity. In SS, they come from equilibrium calculations based on a Beverton-Holt spawner-recruitment relationship for which the steepness is fixed at 0.8 or 0.9 and incorporate information on M, growth and selectivity. The effort to create a productivity prior for JABBA from the age-structured analysis in SS is a very good idea.

TOR 4) Evaluate the adequacy, appropriateness, and application of the methods used to evaluate future population status, given the commissions objectives

Future population status is evaluated by conducting multi-year projections at various fixed catch levels. I note that fixed catch projections can fail in complex multi-area models or for fleets that select only older fish. The planned MSE for alternative harvest policies should develop robustness to such conditions.

I note that the stock is near Bmsy and F is near Fmsy, so it is inevitable that fluctuations in recruitment and actual catch will cause fluctuations in B and F that exceed the threshold Fmsy and or Bmsy thresholds. A MSE, or simply stochastic projections with SS, can evaluate the expected frequency of these occurrences and evaluate control rules with buffers to reduce this frequency to an agreed level.

TOR 5) Evaluate the adequacy, appropriateness, and application of methods used to characterize the uncertainty in estimated parameters. Comment on whether the implications of uncertainty in technical conclusions are clearly stated

The approach used in the assessment is described in SCRS-2019-121 as:
"The modeling approach also follows the structure of the 2018 BET assessment (SCRS/2019/111) to build an initial reference grid, then develop a series of sensitivities, screen these sensitivities across a suite of diagnostics and eventually build an uncertainty grid to account for both model and structural uncertainties in the eventual provision of management advice (ISSF 2018)."

Unfortunately, the draft report did not contain a description of the method for multivariate sampling that is used with SS to get the uncertainty in these projections, so I cannot comment on any potential shortcomings in the approach. No shortcomings are apparent from the results, but a better description and not just a reference should be included in future.

The uncertainty grid approach is a logical and useful way to obtain a mean result and provide information regarding uncertainty in this result. However, I have some concern that the two factors (level of spawner-recruitment steepness (h) and inclusion of the buoy index) chosen to represent axes of uncertainty has underestimated the total uncertainty in the result. The factor h has only a small effect on the fit of the model to the data with the log likelihood changing by only 1.5 units. The primary effect of h will be on MSY and on long-term forecasts with fixed catch. The buoy index has its primary effect by boosting estimates of recent recruitment to levels that are not verified by other data sources and the rationale for including this as a primary uncertainty factor is not clearly articulated.

There are two other factors, natural mortality and growth, that could be considered in the future for more complete characterization of uncertainty. Natural mortality was investigated and a good sensitivity analysis was conducted. The Group found poor model performance when they conducted a preliminary investigation into M as an uncertainty factor. In the future, they should re-initiate this effort using a smaller range of M values. Similarly, growth was investigated and then parameters were fixed for the final set of model runs. A range of growth parameters is a logical axis of uncertainty for the grid.

There will be correlations and infeasible combinations when investigating a grid across fixed values for steepness, M, and growth. Hence a future investigation that uses all these factors will need to prepare for these correlations.

TOR 6) Comment on whether the stock assessment results are clearly and accurately presented in the detailed report of the stock assessment

The results are clearly presented and accurately represent the model outputs.

TOR 7) Comment on potential improvements on the stock assessment SCRS process (CPC participation, transparency, objectivity, documentation, uncertainty characterization, etc.) as applied to the reviewed assessments

I have limited basis to comment on the process because I was not present at the meetings. The documents are clearly presented. A roadmap/index to the sequencing of these documents would be very useful for reviewers.

TOR 8) Comment on the adequacy of the workplan for the assessment and whether it was adequately addressed by the data or assessment working groups

The tropical tuna workplan for 2019 was ambitious and much was accomplished. The plan called for full access to 2018 data for this assessment, but much of the data was only available through 2017. More timely data would improve estimation of recent stock trends.

TOR 9) Consider the research recommendations provided by the working group and suggest any additional recommendations or prioritizations warranted. Clearly denote research and monitoring needs that could improve the reliability of future assessments. Recommend an appropriate interval for the next assessment considering control rules or management strategy in effect.

Four research recommendations from the report are:

- The Group recommended evaluation of approaches to improve the estimates of M, and to develop uncertainty grids that consider the correlations between key biological parameters for example, M and steepness so that biologically implausible combinations can be identified and eliminated.
- The Group recommended increasing the sampling and ageing of small (≤ 65 SFL, particularly < 30 cm SFL) yellowfin using daily ring counts and otolith weight to better understand the dynamics of growth for earlier years, and the apparent slow initial growth/two-stanza pattern.
- The Group recommended that Venezuela scientist and the Secretariat review the size data for 2006 and other years as outliers were identified for this particular year in several fleets.
- The Group recommended that the Venezuela catch and effort data from the longline fisheries should be included, if possible, in the next development of a multi-national joint LL CPUE index.

I concur with all these recommendations. The recommendation regarding small fish growth should be accompanied by research on seasonality of recruitment as growth cannot be interpreted well without birthdate information.

In addition, I recommend more research on the seasonal and spatial distribution of the stock. Where is the major spawning area(s) and where is most recruitment? Is there seasonal movement? Is there evidence that fish recruit both N and S and then move towards tropics?

More work should be done on the young fish CPUE index before it provides a reliable recruitment index. I recommend that the use of it in SS should be with a fat-tailed error structure so it will be less sensitive to extreme values.

Finally, this assessment, like all tropical tuna assessments, is highly dependent on the validity of fishery CPUE being an adequate indicator of the trend in stock abundance over multiple decades. A multi-species approach to investigation of temporal patterns in catchability seems better than efforts on each species in isolation. Alternative approaches to stock abundance estimation, such as close-kin genetics, should be considered as well.

TOR 10) Prepare a Peer Review Report which should specifically address each TOR. Complete and submit this Peer

Review Report along with a summary no later than the two weeks after completion of the assessment meeting.
This document constitutes the Peer Review Report. A separate summary has not been prepared. My overall conclusion is that this assessment is a good advancement over the previous assessment and the scientific advice contained in this assessment is a good basis for management. My review of the assessment finds some areas for potential improvement after further work, but finds no errors or shortcomings that prevent it from being a basis for management.