IOTC 2020 SKIPJACK STOCK ASSESSMENT
Invited Expert Report

Michael J. Schirripa

October 26, 2020

1. Overall opinion about the meeting.

Overall. The assessment author should be commended on the work put into this assessment. Despite the need for a video meeting and the possible necessity to abbreviate some aspects of the assessment process, the author and team covered a great amount of breadth and detail. The work and presentations were very complete and identified some of the major uncertainties in the assessment model and data. The overall process of the assessment was seemingly very transparent and comments from the attendees were welcomed and addressed. The assessment document itself was complete and extensive. While I cannot make the determination, I assume that the assessment addressed every comment or issue brought up at the data preparatory meeting.

2. The stock assessment process

One of the most positive aspects of this process was the use of the team approach. It was encouraging that several assessment scientists attended and were ready to offer ideas and help carry out the requested tasks. Not from the point of view of reviewing the work, but more for the collaborative nature of how the assessment progressed throughout the meeting. This team had a strong background in stock assessment as well as involvement in the various tropical tuna assessed for IOTC, which made them familiar with the issues common to all assessments.

The manner in which the tagging data was modeled was shown to be quite critical. So much so that the use of the various methods used were included several of the axis of uncertainty. This puts great importance on the assumptions made when considering this data. It would have been helpful to see how the model responded to removing the tagging data entirely; however, this may not have been compatible with a two-area model. If skipjack movement is associated with changes in oceanographic features, as many tropical tuna are, it may be worthwhile to explore the possibility of annual dynamic stock mixing rather than one stationary value. Since the tagging data was included in the model to inform abundance and fishing mortality, the model sensitivity to this data should be investigated with detail to ensure its influence is fully understood. This is especially true if the same tagging data is used for multiple species assessments. I think it worthwhile to consider a detailed study/workshop of this dataset as it effects several assessments, perhaps in the same manner. Collaboration with ICCAT Atlantic Ocean Tropical Tuna Tagging Program (AOTTP) might want to be considered.

3. Exploration of uncertainty

Data. The data presentation was helpful and well done. The Figures were innovative and informative. Regarding the assessment, it seems the assessment author did a complete examination of the data. A decision that perhaps could have been made prior to the meeting is whether to include the longline lengths. Indications were that they were likely biased. Thus, if there is no faith in them, then trying to fit them could mean unnecessary sacrifices and/or misspecification somewhere else in the model.

Data weighting. Appropriate data weighting is an ongoing investigation (CAPAM). Each year of each CPUE time series was assigned a standard error (SE) of 0.10. Using an accepted value of SE for each CPUE is not necessarily problematic. In fact, the annual SEs for a given CPUE from the standardization process (e.g. GLM), while often used, are not appropriate for inclusion into the stock assessment. The relative year-to-year variation in SE from the standardization process, however, can be used as a surrogate for year-to-year variation in the SE of the CPUE. One approach is to choose a minimum SE (e.g. 0.10). Any SE's from the standardization process less than 0.10 are assigned a value of 0.10, and any SE's greater than 0.10 are assigned their actual values from the standardization.

Another approach to weighting the CPUE information is to estimate the added variance of each CPUE. This request was addressed by the assessment author but the presentation of the results did not seem to draw much attention from the attendees. This, despite the fact that the exercise estimated quite a difference in additional variance between the two indices of abundance, about 0.18 for PL index, and 0.025 for PSLS index. Thus, the adjusted variance model gives much more weight to the PSLS index. This result does not support the assumption of equal standard errors for each year and index. The variance re-weighting resulted in a larger population size and the difference between the two models was at least as large as some differences seen between some of the models contained in the grid.

Attempts at adjusting the effective sample size (ESS) of the lengths via such accepted practices (such as Francis reweighting) was not presented. In fact, the Francis weighting method suggested that the length compositions should/could have been adjusted further with an estimated variance multiplier. Given the demonstrated influence of the length information, some exploration of the Francis weighting of the length compositions may be helpful.

Worthy of exploration may be the inclusion of the detailed bias adjustment options in Stock Synthesis. The bias adjustment used:

> *1960 #_last_yr_nobias_adj_in_MPD; begin of ramp*
> *1983 #_first_yr_fullbias_adj_in_MPD; begin of plateau*
> *2018 #_last_yr_fullbias_adj_in_MPD*
> *2018 #_end_yr_for_ramp_in_MPD (can be in forecast to shape ramp, but SS sets bias_adj….*
> *1 #_max_bias_adj_in_MPD (typical ~0.8; -3 sets all years to 0.0; -2 sets all non-forecast yrs….*

Was somewhat different from the estimated alternative:

> *Estimated alternative inputs to SS control file associated with blue line in figure:*
> *1972.3 #_last_early_yr_nobias_adj_in_MPD*
> *1993.3 #_first_yr_fullbias_adj_in_MPD*
> *2018.0 #_last_yr_fullbias_adj_in_MPD*
> *2029.8 #_first_recent_yr_nobias_adj_in_MPD*
> *0.9189 #_max_bias_adj_in_MPD (1.0 to mimic pre-2009 models)*

This modification may or may not be influential.

Model diagnostics. By fixing the steepness and M parameters (and assuming no meaningful error in the catch), one is essentially solving for R0 rather than fitting it. This approach will make the CV's associated with SSB for any given model much smaller than they really are. This approach also renders the profiling on R0 diagnostic deceptive. While there is evidence that a minimum was reached, this is because there is only one value of R0 that corresponds to the fixed values for steepness, sigma-r and M (and the fixed catch). The conclusion that R0 is well estimated is not truly valid in this case as it is uncertain exactly how much the data is driving the estimate versus the fixed values. Furthermore, quantities derived from R0 will reflect this same issue (e.g. MSY). Nonetheless, there is still value in the profiling as it shows the conflicts in the data and the value of the R0 parameter.

Interesting is the innovative manner that is being proposed for selection of models in the grid: the proportion of diagnostic tests that each of the models pass. The higher the percentage of tests passed the more plausible the model is deemed. For a given cut-off, say 60%, all models passing 60% of the diagnostic tests get included in the gird. But why 60%? This is difficult to justify objectively. I think that perhaps this method, while objective and reproducible, may be a bit too mechanistic. It assumes that all diagnostics have equal weighting and that the issues flagged by each are of equal concern. I think it to be a very useful tool to summarize the diagnostic results and should be included in an assessment, but to use the numerical scores as a means to choose which models will go into the grid of uncertainty I think ignores several important factors (such as, are the models being correctly specified).

Approach to account for uncertainty. The assessment process used the “model ensemble” approach to create a grid of possible alternative states of nature. The presentation of this approach was very useful in deciding which models should be included in the grid of uncertainty. The selection of the alternative states of nature that make up the grid is of course critical to the overall management advice.

When creating alternative states of nature for a grid of uncertainty, it is standard practice to use various values for important biological parameters (i.e. M, h, etc.). Also appropriate is the inclusion/exclusion of suspect data sets (e.g. tagging). It is also acceptable to create two alternative states of nature by (i) estimating/using variance re-weighting, (ii) not estimating and using variance re-weighting. However, I do not think is appropriate is to create two alternative states of nature by using two different lambda values (e.g. lambda = 1.0 and 0.1) on the observational data partition. Biological data has natural bounds, such as 0.2-1.0 for steepness, and perhaps 0.01 and 1.0 for natural mortality (likely even narrower). Variance re-weighting and/or Francis re-weighting of the length compositions is an objective calculation. Lambda values, on the other hand, are not bound nor are they derived from objective calculations. Thus, the alternative values are essentially arbitrary. The difference in the values chosen (in the case of 1.0 vs. 0.10 is 10x difference) can have a great influence on the degree of uncertainty of the results generated over the entire grid.

I think with the inclusion of “effort creep” as an axis of uncertainty was a good choice. In fact, I think that the question of effort creep over time due to technological factors (i.e. larger boats, better electronics, cell phones, improved fishing gear, etc.) is deserving of even more attention, realizing however that changes in catchability are very hard to quantify outside the assessment model itself. The paper presented that attempted to quantify this creep used a novel approach, but the assumptions were broad and originated from a different species. Nonetheless, if all parties involved agree that effort creep is an issue, then it represents at the least a good starting point for further analysis. In an effort to keep the grid of uncertainty manageable, sensitivity of the model to the assumption of effort creep could perhaps be compared to the other axes of uncertainty in the grid for consideration for replacement. An option to estimating the “effort creep” outside the model would be to fit a trend of the increasing catchability within the model and thus using all of the observational data to inform the fit. This could be accomplished by estimating an annual catchability deviation (with or without an informative prior) or random walk. This would allow for error on estimation of the deviates of q rather than dictating it via changing the data itself without error.

4. Main controversies and discussions

A great deal of debate ensued around the proposed "effort creep" in the purse seine fishery and the PSLS CPUE. With only two CPUEs being used in the model, this was an important point of contention. There was debate as to whether a creep of 0% should be part of the grid. Nevertheless, to keep the grid balanced it was decided that it would remain. I fully support this decision. It was still a discussion point when writing the report. In fact, there was discussion of it being even greater than 25 percent.

5. Ideas for future improvements

Overall, I believe the IOTC skipjack assessment and the process used to derive management advice is quite sound. The entire team should be commended on the successful completion of their task under adverse conditions. Also noteworthy is the opportunity that all meeting attendees were offered.

The "model ensemble" approach seems to be gaining favor in the stock assessment community. In that regard, this assessment was thorough in its examination of justification of the alternative states of nature used to comprise the grid of uncertainty. When using to many models for advise, it would be helpful to include histograms of each of the models on the margins of the KOBE "snail tracks" plot. The R-code for this plot is freely available. Selection of the models included in the grid needs further thought (for all stock assessments). It is difficult to attempt to include all the important axis of uncertainty without "blowing up" the grid.

I think various techniques in data weighting for the base model is worth exploring. If the effects are minimal then that result in of itself is useful.

Given the uncertainty surrounding the tagging data, as well as its wide use for several IOTC tropical tuna assessments, a singularly devoted workshop to thoroughly scrutinize the data and ensure current assumptions and uncertainties are understood is justified. Such a workshop devoted to analysis could not only benefit the stock assessments that use this data but may also provide insights to the study design of the tagging effort.

The model configuration used in this assessment is quite complex, which may well be necessary to capture many of the important aspects of the fishery. However, if not already explored, a simpler model configuration could be evaluated to determine if the current level of complexity is necessary to capture the important fishery dynamics and provide sound advice.