INTER-AMERICAN TROPICAL TUNA COMMISSION

2nd REVIEW OF THE STOCK ASSESSMENT OF YELLOWFIN TUNA IN THE EASTERN PACIFIC OCEAN

La Jolla, California (USA)
2-6 December 2019

REPORT OF THE MEETING

Shannon Cass-Calay (Chair), Alistair Dunn, Adam Langley, Steve Teo, Laura Tremblay-Boyer

CONTENTS

EXECUTIVE SUMMARY

The 2nd review of the stock assessment of yellowfin tuna in the eastern Pacific Ocean (EPO YFT) was conducted from 2-6 December 2019 at the Marriott Hotel in La Jolla, California. In accordance with the Terms of Reference (TORs), the aim of the review was not to conduct a review of a specific assessment in relation to the provision of management advice, but rather to identify the research to improve the assessment of EPO YFT and to provide advice that would assist the approach that should be taken in order to develop a stock assessment for the 11th Meeting of the Scientific Advisory Committee in May 2020. At the time that the review panel (RP) met, the IATTC staff were still in the process of developing a base case model for 2020. As such, the RP were unable to proceed with a formal review of all of the aspects of the assessment, but aimed instead to provide advice and recommendations on research, analyses, and potential options for developing the assessment.

IATTC staff provided background documentation, numerous presentations and responses to requests by the RP that addressed topics related to data inputs, life history, and aspects of the modelling. The review focused on the research needed to improve the current assessment. Areas of focus for the meeting included:

- What causes the mismatch in the longline and purse seine CPUE based indices of relative abundance?
- What is the most appropriate stock structure for the yellowfin tuna stock assessment?
- What is the most appropriate fishery structure for the yellowfin tuna stock assessment?
- What approach should be used to deal with the uncertainty in the length of older individuals and the impact it has on the stock assessment results?
- What is the appropriate stock-recruitment relationship?
- How should the CPUE indices of abundance be used in the stock assessment?
- Should logistic selectivity be used and for which fishery/survey?
- How should the recent increase in the size of longline caught fish be dealt with?
- Age- and sex-specific natural mortality

The Panel notes that the following topics were neither discussed nor evaluated during the review:

- Discards monitoring and estimation
- Parameterization of catchability

1. BACKGROUND

The review of the stock assessment of yellowfin tuna in the eastern Pacific Ocean (EPO YFT) was conducted from 2-6 December 2019 at the Marriott Hotel in La Jolla, CA. The meeting was chaired by Shannon Cass-Calay (NOAA SEFSC) and the other RP members were Alistair Dunn (Consultant), Adam Langley (Consultant), Steve Teo (NOAA SWFSC), Laura Tremblay-Boyer (Consultant). The review focused on the research needed to develop an assessment for the Scientific Advisory Committee of IATTC in 2020. IATTC staff had already recognized that the previous EPO YFT was not sufficiently robust for the purposes of providing reliable assessment advice. At the time of the review, the IATTC team was in the process of revising the model inputs and model structure for the 2020 assessment.

On that basis, the RP noted that the purpose of the review of the IATTC staff's assessment of the EPO YFT was not to determine whether the current or proposed assessment was adequate for providing management advice; the intention was to provide information to the assessment team likely to improve the new assessment. The goals and objectives of the review are to:

- identify the best available science for use in the assessment;
- provide an independent review of the assessment; and
- provide advice on future research and data collection that will improve the assessment and the provision of management advice.

The TORs noted that the main responsibility of the RP was to perform an adequate technical review of the assessment.

IATTC staff provided background documents prior to the meeting, documents prepared specifically for the review, and presentations on the data inputs, stock structure, biology, and assessment modelling. The RP made several requests to the IATTC staff to enable it to better understand and consider the issues. These requests, and the responses by IATTC staff are listed in **Appendix E**.

As noted above, the current stock assessment for EPO YFT had been rejected by IATTC staff and the revised model was not fully developed at the time of the review. Hence, the RP focused its review on the development of the EPO YFT data inputs, stock structure, biology and potential assessment models. These deliberations are described in **Sections 3-6** below.

The RP noted the professionalism and expertise of the IATTC staff, including Carolina Minte-Vera, Haikun Xu, Cleridy Lennert-Cody, Dan Fuller, Kurt Schaefer, Mark Maunder, and Alexandre Aires Da Silva, and expressed their gratitude for the hard work and willingness to respond to RP requests, and also for their support, provisions, and hospitality during the review.

2. GENERAL COMMENTS BY THE REVIEW PANEL

The RP thanked the IATTC staff for the available documents and presentations, but noted that it would help external reviewers if IATTC developed standardized documentation of the input data and subsequent analyses conducted on the input data to develop life history information, model parameters, length frequencies, and abundance indices. This documentation should include a description of data sources, data selection, assumptions, methods, relevant model diagnostics and a discussion.

The RP noted that more than six years had passed since the previous EPO YFT review (Martell *et al*. 2013), and that subsequent iterations and analyses for the input data and stock assessment methods were documented across a number of primary publications and report updates. In addition, a number of data input analyses and diagnostic plots were not available for public or scientific review outside of the IATTC staff. The RP recommended that IATTC staff provide benchmark documents of the analyses used to develop input data and the stock assessment model for the SAC in May 2020 as part of their development of the assessment for consideration for management advice.

The RP also noted that important questions and concerns arose regarding the abundance indices used to inform the stock assessment model, notably the Japanese longline (LL) CPUE index and the dolphin-associated purse seine (PS) CPUE index. Comprehensive documentation to facilitate the evaluation and interpretation of these indices is strongly recommended.

Finally, the RP noted that several lines of inquiries pertaining to data issues had been followed for bigeye tuna (for which the latest assessment is also under investigation by IATTC staff) but not repeated or prioritized for yellowfin tuna at the time of the review. The RP would advise that these analyses also be extended to yellowfin tuna.

3. REVIEW OF DATA INPUTS

The following inputs to the assessment were evaluated by the RP: catch, species composition, length composition and catch-per-unit-effort (CPUE). The RP did not review or evaluate the estimation of discards.

3.1. Catches

3.1.1. Purse seine catch

The catches from the purse seine (PS) fishery on floating objects (OBJs) are dominated by skipjack tuna (SKJ), followed by bigeye tuna (BET) with yellowfin tuna (YFT) representing a relatively small proportion of the total catch. The species composition of the catch is estimated from port-sampling data because of the need for accurate speciation of smaller fish (*i.e*., less than about 70 cm), mainly due to similarities between small BET and YFT. IATTC staff informed the RP that species-based port sampling has been used to estimate the species composition of the catch from the PS-OBJ fishery since 2000.

The higher catches of BET from this fishery since the mid-1990s have coincided with the estimation of higher recruitment in the BET assessment model. The review of the EPO BET assessment model identified the potential for uncertainty in the catch estimates to be influencing the trend in BET recruitment. The EPO YFT assessment models have estimated lower overall levels of recruitment from about 2000 onwards. There is potential for the under-estimation of yellowfin tuna catches from the PS-OBJ fishery to be influencing the estimates of yellowfin tuna recruitment, although there is no strong link with a corresponding over-estimation of the bigeye tuna catches.

Nonetheless, given that there are potential issues associated with the reliability of sampling catches of smaller tunas from purse seine catches, the RP recommended that:

- the sensitivity to the current species proportion be evaluated via alternative levels of catch of yellowfin from the OBJ fishery

For example, the analyst could determine the magnitude of the change in YFT PS OBJ catch that would be required to remove the step change in yellowfin recruitment occurring around 2000.

Alternatively, the change in the estimated level of yellowfin recruitment since 2000 may be associated with environmental change (*i.e.* regime shift). Therefore, the RP recommended that:

- An evaluation of the environmental mechanisms that may influence yellowfin recruitment should be undertaken once the new stock assessment has been completed.

3.1.2. Longline Catch

Longline (LL) catch-and-effort data are reported to the IATTC secretariat at an aggregated scale of 5 x 5 degrees, though some countries provide data at finer resolutions. No specific concerns were raised by the RP on the reliability of aggregated LL catches, given these catches tend to be well determined in tuna RFMOs, and individuals are caught at a size where species identification tends to be accurate.

Operational catch-and-effort data, including vessel identifier, for key fleets are not always accessible to IATTC staff, or are available only for limited periods of time or under specific constraints (*e.g.*, collaboration with visiting staff). Given the influence of standardized LL CPUE indices on the EPO YFT stock assessment, the RP recommended that:

- IATTC staff have continuous access to an updated dataset of operational catch-and-effort for all fleets operating in the convention area to ensure that robust and unbiased analyses of this dataset can be developed. This dataset should include vessel identifier and other operational variables like hooks-between-floats to inform the standardization, ideally for the full time-series.

3.2. Species composition of catch

The RP noted that species composition from the purse seine fishery was more uncertain before 2000. The RP acknowledged that this could have introduced unquantified biases into the stock assessment, but made no specific recommendation.

3.3. Length composition

The longline length composition data came primarily from the Japanese fleet. Data from the Japanese LL training vessels were not included in the assessment dataset because they were not considered to be representative of commercial fishing operations as the training vessels operate in a restricted range, mostly north of 10°N and in the Hawaiian region whereas the commercial fishing operations are more widespread.

Size composition data from the commercial LL fleet have been collected from a number of sources, primarily from commercial fishermen and, in recent years, from observers. The source of size data collected prior to the early 1980s is unknown. Most of the size data from the early period are fish weight data (kg). These data have been previously rejected from the assessment model as it was not possible to reconcile differences between individual fish weights and lengths.

Most of the LL length composition data from the early 1980s to the 2010s were collected by commercial fishermen. A comparison of fish size from multiple data sources (*i.e.*, average fish weights converted to length from commercial fishermen, average fish weights from logbook and from the weight sampling) suggest that there may have been potential positive bias in the length composition data collected by commercial fishermen.

The IATTC staff noted a recent increase in the average length of longline caught fish. Such an increase can provide an influential signal of decreasing fishing effort (impact) to the assessment model. Therefore the RP recommended that:

- Verifying that it does not result from an issue in data collection or treatment should be a priority.

The RP also considered whether the increase in average length could also correspond, at least in part, to a transition in 2010 from the collection of length measurements primarily by vessel crew to observers. However, this hypothesis was rejected because an analysis suggested that vessel crew sampled larger fish than observers, at least for the years 2011-2013, when data were available from both observers and vessel crew. Other hypotheses included changes in the area of operation or gear configuration changes, but no conclusive evidence was presented.

The LL length composition data were available by sex from about 1995, with a sex ratio skewed towards male fish at large sizes. The reliability of the sex determination is unknown as it was noted that the gonads are typically not removed from the gut cavity during on-board processing.

While the current assessment model is structured by sex, it was not fitted to sex-specific length composition data. There were no further discussions on the topic of sex ratio but the RP recommended that:

- the skew in the sex ratio should be investigated further if IATTC staff wanted to move towards additional disaggregation by sex for specific components within the assessment model.

3.3.1. VAST approach to standardize composition

To estimate the annual population-scale length composition for dolphin-associated (DEL) PS fishing IATTC staff applied the VAST software package. The intention was to incorporate the resulting population-scale length compositions in the new assessment model. The RP noted that there seemed to be limited data available prior to the early 1980s, with the proportion of dolphin sets lower than the 75% threshold used to identify vessels over the entire time period, and during this period the fishery was mainly comprised of smaller fish. Therefore, the RP recommended that:

- Additional analyses on the length composition data using the VAST approach are required before these indices can be considered robust, and the assessment team should consider the choice of vessels to include in the analysis, including the proportion of DEL sets by each vessel in an appropriate time period (*e.g.*, monthly or quarterly) as a potential covariate, and evaluate the potential for spatial confounding in the resulting analyses.

Overall, the VAST approach estimated a vulnerable (DEL PS) population with a higher proportion of larger fish relative to the nominal (catch scaled) length composition. The RP recommended:

- Understanding the source of this effect before these estimates could be considered robust for use in an assessment model.

For instance, it could stem from interpolating across areas where the CPUE of larger fish was higher (and the level of catch was low), but also from potential edge effects or the assumptions of even spatial autocorrelation across the domain.

The RP noted that there is also persistent bimodal structure in the standardized length compositions derived from the PS fishery data. This appears to be related to the combination of data from fisheries that separately operate on smaller and larger fish, and lower overall catches of fish in the intermediate length range. This may also be related to the spatial structure of the composite fisheries. If this is the case, the RP suggested that:

- It may be necessary to fit to the population length composition using a bimodal selectivity function.

More pronounced issues stemmed from a similar VAST application for the LL length composition data. The analysis was intended to generate year-quarter estimates of the population length composition over the entire EPO spatial surface. The spatial distribution of these data were limited and variable between time intervals. The nominal and VAST length compositions were very similar, suggesting VAST interpolated a relatively homogeneous length composition across the spatial range of the analysis (in each time step). This was quite surprising given that there was some evidence of spatial variation in the size of LL caught YFT across the EPO, with larger fish towards the southeastern area of the fishery. The RP expected that the VAST model would interpolate these patterns to account for differences in the spatial distribution of the length sampling in each time step and, therefore, generate length compositions that differed from the nominal distribution. It was unclear that the current implementation of the VAST model accounted sufficiently for the spatial differences in the length composition, to the extent that the analysis had removed the differences in fishery selectivity that might be associated with changes in the spatial distribution of the fishery (and/or sample collection).

- In general, the RP recommended the development of additional summaries for the input data as well as the inclusion of standard diagnostic plots, disaggregated by space and time, to inform on model performance,

3.4. Indices of abundance (CPUE)

The previous EPO YFT assessment incorporated two main sets of CPUE indices: southern LL CPUE indices derived using a GLM approach, and the nominal (catch per day) CPUE from the DEL and NOA PS fishery.

3.4.1. Longline

In preparation for the 2020 assessment, VAST was also used to standardize the CPUE of the LL fishery. The VAST model of the LL fishery includes the hooks between floats (HBF) effect directly, however the time series is partitioned into two time blocks, split in the early 1990s to account for changes from shallow to deep LL sets (as indicated from HBF). The HBF effect is still influential in the earlier part of the time series.

Given the significant change in standardization methodology, the RP recommended that:

- It would be informative to have a direct comparison between the VAST indices and more traditional GLM indices to determine whether or not the new approach to standardization is influencing the final set of abundance indices. This should be one of the sensitivity runs reported for the 2020 assessment.

The CPUE of the LL fishery appeared to be related with ENSO, and the RP noted that the indices may be reflecting changes in either catchability or abundance. While the RP noted that a geostatistical approach such as VAST should be preferred in contrast to a GLM with a fixed cell effect to account for changes in habitat, geostatistical models do not solve the issue of differentiating between abundance and catchability effects when there are no covariates representative of the one or the other. For instance, yellowfin catchability is likely to vary with changes in the depth of the thermocline or other depth-related oceanographic conditions (though not to the same extent as bigeye), but recruitment and movement are likely to be affected by other oceanography variables (*e.g.*, SST) that change during ENSO events.

No summary of input data was presented to adequately review the distribution of LL effort over the spatial extent of the VAST analysis. It was evident that the fishery has contracted considerably and this has accelerated over the last decade. This is partly reflected in the higher CVs for the indices from the recent years, although the RP noted that the CVs were still quite low (< 10-15%) over most of the time period, with the exception of the terminal years.

The RP questioned the assumption that the use of VAST is a way to control for preferential sampling, and noted that it may not be valid. VAST, like most geostatistical models, assumes that the response variable is independent from sampling intensity. This is a key issue in the use of fishery-dependent CPUE as a proxy for abundance, given that fishing effort tends to be concentrated in areas of higher catch rates for the species of interest. In the EPO, this issue, coupled with the contraction of the LL effort, increases the risk that the VAST approach applied in its current form may result in biased estimators of the quantities of interest being modelled. Recent research on this topic in a fisheries context shows that density in poorly-sampled areas tends to be overestimated, resulting in hyperstable CPUE indices. Preferential sampling is an issue that plagues analyses of fishery-dependent data in most fisheries and accounting for preferential sampling still remains an active area of research in geostatistics. To the RP’s knowledge, no analytical approach applied to fisheries systems has been able to successfully account for it. Therefore:

- The RP urged caution in the use and interpretation of VAST-derived indices as unbiased population estimators of abundance and size composition. When possible, staff should explore whether the model is sensitive to the patterns of preferential sampling occurring in the EPO.
- One option would be to develop a simulation model combining the Japanese LL effort dynamics with scenarios of spatial distribution of yellowfin tuna abundance over time.

The RP also noted that the uncertainty in the LL CPUE indices in the terminal year of the model was one of the reasons for rejecting the previous assessment as it changed the estimated stock status. However, the RP considered that the deficiencies in the recent data set are unlikely to be adequately alleviated by the application of the VAST analytical approach. LL catch and effort data had limited spatial coverage over the last decade, particularly in the eastern area of the fishery. This means that predictions for large areas of the spatial domain are unlikely to be reliable due to the lack of observations in the last decade. To better elucidate the influence of the analytical approach on the final indices, the RP recommended that IATTC staff:

- Consider the development of separate indices for areas of the fishery that have been covered by the fishery throughout the data period (northern, southwestern, southeastern). These areas could be used to generate three sets of CPUE indices using traditional approaches to standardization. The resulting indices could then be more directly compared with the results from the VAST model.

3.4.2. Purse seine

For the 2020 assessment, IATTC staff proposed to apply a CPUE index derived from the DEL PS fishery to monitor the abundance of YFT in the northern region of the fishery. The proposed CPUE index was derived using a simple implementation in VAST. The RP had several concerns with this approach as outlined during the review.

The RP had general concerns regarding the appropriateness of using catch and effort data from the DEL PS fishery, especially given that the searching component of the fishery operation was not adequately quantified, and the relative efficiency of the fishing operation had increased considerably over the period of the fishery with the introduction of new technology. The RP also discussed issues related to the evolution of the PS fishery associated with the development of fishery on floating objects, and noted that these may influence the operation of the DEL PS fishery. Insufficient information was presented to evaluate the continuity of the operation of the fishery (*e.g.* with respect to longevity of individual vessels within the fishery, introduction of new technology, and changes in the distribution of effort by set type [DEL vs OBJ]). Therefore, the RP recommended:

- A thorough characterization of the CPUE data set be conducted to enable an evaluation of the utility of the CPUE indices.

The RP also noted that the DEL 'fishery' had been defined as those vessels having reported at least 75% of their sets targeting dolphin schools when records were aggregated over the full time-series. The RP noted their concern that actual targeting practices by individual vessels may change over time, and that the definition based on the long-term average could miss gradual shifts in targeting, and could bias standardized CPUE trends if unaccounted for. The RP requested a diagnostic plot during the meeting that illustrated the amount of effort by each vessel, and the proportion of that effort identified as DEL sets. This plot suggested a strong correlation between the period of peak yellowfin CPUE and a high proportion of DEL sets by most vessels. Therefore:

- The RP recommended that the classification of the DEL fishery be reconsidered by the assessment team, and that alternative definitions be explored for the DEL fishery from quarterly or annual DEL set ratios. In addition, some members of the RP suggested that the analysts investigate the use of the proportion of DEL sets as an explanatory variable in the analysis.

The CPUE VAST model for the DEL PS data that was presented to the RP also showed some strong spatial patterns in residuals, with knots covering larger areas at the edges of the spatial domain appearing to have a higher prevalence of negative residuals. The RP recommended that IATTC staff explore the reason behind the strong residual patterns from the VAST analysis. If the pattern in residuals is caused by the assumption that the spatial covariance function is the same over space and time, the RP recommended that:

- The analysts investigate splitting the spatial domain into two or three components, fitting the model separately (as has already been done by IATTC during exploratory analysis), and verifying that the pattern in spatial residuals was reduced.
- Including oceanographic covariates relevant to yellowfin tuna biology as a density covariate might also decrease the reliance of the VAST model on the patio-temporal effects estimated for the knots.

The RP considered that oceanographic covariates could be included as either density or catchability effects, but concluded that given the lack of catchability covariates assumed in the current model, model predictions would be improved by the addition of density covariates unless there was strong evidence that an oceanographic covariate affected catchability. Therefore, to avoid the confounding that occurs if a covariate affects both density and catchability, the RP agreed that oceanographic variables should generally be included as density covariates.

The RP noted that patterns in spatial residuals were generally less severe for areas of the surface where the knots were concentrated. Therefore, the RP recommended:

- Redefining the mesh to distribute the knots evenly over the spatial domain, together with an increase in the resolution of the mesh both for the core area and the outer edges of the spatial domain, might be useful.
- A plot of the predicted value and variance for each knot should be examined to verify that there are no edge effects in the estimated surface.

The RP noted that the spatial domain of the DEL PS fishery was limited prior to 1980 and that the spatial coverage was quite poor during 1980-1985. There was also a southward and westward extension of the DEL PS fishery since 2000. The RP considered whether the spatial extension was related to the interaction with the expanding OBJ fishery, and whether the expansion of that fishery could be influencing the operations of the DEL PS fishery. The RP noted that:

- A better understanding of the interaction between the DEL and OBJ components of the PS fishery would assist the implementation and interpretation of the model.

To the RP's understanding, all spatial cells having at least one record of fishing effort were included within the spatial domain for the analysis. To ensure that poorly observed cells do not overly influence the final predictions, the RP recommended that:

- It is important to re-run the model with different threshold for the inclusions of cells within the spatial domain (*e.g.*, minimum effort in sets over a given number of quarters).
- Ideally, the spatial domain should be refined to exclude edge cells that were only anecdotally part of the fishery, but the exact threshold of inclusion would need to be determined through exploration of the inputs and of the sensitivity of model predictions on the definition of the spatial domain.

It was noted that the indices from the earlier years (pre 1980) had a higher associated CV and that those indices would be less influential in the likelihood. The RP noted that in the case of the Northern model, no other CPUE indices were available for that time period, so the model will fit to those indices anyway. However, the RP recommended the following explorations to facilitate evaluation and interpretation of the PS indices:

- Exclude early indices from the model to assess the influence of these data.
- If there are significant differences in the operation of the fishery at various intervals (*e.g.*, pre and post expansion of the OBJ fishery), it would be useful to partition the DEL PS indices into time periods to account for possible changes in catchability.

The RP noted that a depletion analysis indicated that the catchability of the PS fishery may be variable, for example the high peak in the PS CPUE might have been caused by increased catchability during a period of higher abundance. This could have been related to the availability of YFT on the surface and/or the proximity of the biomass to the main area of operation of the fleet. Similarly, the RP noted that during the period prior to 1985, there was a considerably higher proportion of non-DEL sets. This reinforces the need to exclude the earlier data from the analysis. There was also a time period during the mid-late 2000s when there was a lower proportion of DEL sets across most of the vessels. On that basis, the RP concluded that:

- It would be preferable to define the vessel selection criteria on the basis of the proportion of the DEL sets for each vessel within finer time scales (*e.g.*, year or quarter), as discussed above.

The RP also made suggestions for general procedures and diagnostic plots relevant to both the LL and the DEL PS CPUE standardization.

- Stepwise plots (Bentley *et al*. 2011) showing the effect of consecutive layers of filtering and of candidate filtering thresholds on the standardized index would be useful tools to identify whether the index of abundance is sensitive to decisions about data grooming.
- Noting that residuals over the entire model domain tend to be well behaved, plots showing disaggregated residuals by space and time should be considered whenever possible. For instance, annual QQ plots of residuals should be examined to verify whether there is a temporal trend in model fit.
- To verify whether seasonal effects are being accounted for properly, the model could be fitted separately for each quarter. If the diagnostics for the quarter-specific models are improved, some sort of additional flexibility would be required in the all-quarters model. One option may be to include oceanographic covariates that vary seasonally but unevenly over the spatial domain (*e.g.*, SST).
- Residuals disaggregated in space and time should be compared between the VAST and the traditional GLM approach as support for the proposed methodology.

- Finally, both the PS and LL CPUE indices are comprised of the two model components (probability and magnitude of catch). It would be important to understand which component of the model is most influential in determining the composite year/quarter indices. This would be best presented as the annual trends from each component (normalized) and the composite index.

4. MOVEMENT AND STOCK STRUCTURE

4.1. Movement

Some evidence was presented by the IATTC of regional residency by individuals tagged in the North and South regions. More specifically, most linear displacements from conventional tags appeared to be less than 1000 nm. This limited displacement was supported by data from archival tags that also highlighted the potential for regional residency, especially around specific locations like the Revillagigedo Islands. However, the exact amount of north-south exchange from conventional tags was hard to assess based on the visualizations presented to the RP. There also appeared to have been limited releases from offshore areas.

There was also some evidence of limited exchange between the southern and northern 'populations', from patterns in the CPUE and length-composition data. However, it was unclear that patterns in the CPUE and length-composition data were due to a real biological process, such as movement.

To improve understanding of YFT movement and stock structure, the RP considered that:

- A tagging program for YFT, especially in offshore areas and in conjunction with oxytetracycline marking of otoliths, should be designed and conducted as a priority. Noting the logistical constraints of tagging yellowfin in the offshore areas, the RP noted that collaborations with the Western and Central Pacific Fisheries Commission might be a productive path to share the logistical and financial burdens of organizing offshore tagging cruises.

4.2. Stock structure hypothesis

Previous assessments have assumed that YFT in the EPO consisted of a single, well-mixed stock, and any spatial heterogeneity was accounted for using the 'fleets-as-areas' approach. However, this approach assumes that individuals removed in any area affect the population in all areas at the same time. Exploratory work on the temporal mismatch between the DEL and LL indices (*cf.* Xu - Hypotheses for the difference between PS and LL indices) suggested that the strong cohorts indicated by the DEL and LL indices in 2001 and 2000 respectively were due to different recruitment events in 1999 and 1998 respectively. This difference in recruitment events between the DEL and LL was further suggested as evidence for YFT spatial structure in the EPO, structuring that is unaccounted for in the current stock assessment. Furthermore, an analysis of the size composition data of the various fisheries and their relationship with the DEL and LL indices (*cf*. Minte-Vera *et al*. - Development of a new benchmark model for yellowfin tuna in the EPO), suggested that the size composition data of some fisheries (some DEL, some OBJ, and some NOA fisheries) were more consistent with the DEL index while other fisheries (LL, some OBJ, and some NOA fisheries) were more consistent with the LL index.

The IATTC staff also presented evidence of YFT stock structure for the EPO. There was a single genetic study with multiple sites in the EPO which showed no differentiation of populations on a north-south gradient (Ward *et al*. 1997). However the RP notes that this study used allozyme and mitochondrial markers which are known to be very sensitive to small amount of exchange between populations[1]. Therefore

[1] Waples, R. S., & Gaggiotti, O. (2006). INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular ecology*, *15*(6), 1419-1439.

the RP recommended that:

- An updated genetic analysis using modern molecular markers like SNPs from multiple locations along a north-south gradient in the EPO might provide further information with regards to stock structure.

The IATTC staff presented additional information regarding the ensemble characteristics of the environmental variables across the EPO (*cf.* Minte-Vera *et al* - Stock and spatial structure of yellowfin tuna in the Pacific Ocean; and a response to a RP re: SST patterns in different quarters of the year). IATTC staff reported that YFT undergo continuous spawning in locations with SSTs of ≥24°C, and that spawning was optimal when SST was 28°C.

Given that the DEL index was primarily based on data in the north and the LL index was primarily based on data in the south, it was proposed that the YFT in the EPO may be modeled as two separate stocks: a 'North' stock and a 'South' stock. Therefore, two reference models were developed and presented. A 'North' model was based on the DEL index, and included the fisheries that were identified as being consistent with the DEL index. A 'South' model was based on the LL index and the fisheries that were identified as being consistent with the LL index. Both models were relatively independent of each other with no shared data. Biological parameters (*e.g.*, growth, natural mortality, stock-recruitment relationship) were identical for both models. Although the models were nominally spatial, the area definitions varied by fishery type and season. For example, the DEL fisheries were split roughly along the 5°N parallel into the 'North' and 'South' models. In contrast, all of the LL fisheries in the entire EPO irrespective of latitude were in the 'South' model, and only one OBJ fishery in Quarters two and three were placed into the 'North' model.

Based on the information that was presented, the RP felt that the YFT in the EPO might consist of two stocks and that a two-stock hypothesis could be further explored. However, the RP concluded that:

- a single-stock model should also continue to be developed as an alternative hypothesis because the evidence supporting a two-stock hypothesis was thought to be suggestive, rather than conclusive.

In addition, the single, well-mixed stock hypothesis and the two independent stocks hypothesis can be thought of as extreme cases of a continuum, while the real stock structure is likely an intermediate somewhere between those extremes. In short, there did not appear to be enough evidence to support a two-stock hypothesis as the only plausible hypothesis.

There was further evidence suggesting that YFT in the EPO was somewhere between a single, well-mixed stock and the two independent stocks. A comparison of the estimated recruitments for the 'North' and 'South' models indicated that the long-term (multi-year) scale and trend in recruitment for both independent models were largely similar. However, the recruitment for a specific quarter could differ substantially in magnitude between the models. This suggested that the large-scale source and/or environmental influences on recruitment for both the 'North' and 'South' were similar but the apportionment to the 'North' and 'South' could be highly variable. This does not preclude the hypothesis of two independent stocks because the similar long-term recruitment trends could be due to large-scale environmental conditions influencing both stocks simultaneously. However, there is a non-negligible risk that there is a common pool of recruits that gets apportioned between the two areas. Although management consequences were outside the purview of the RP, the RP thought a risk-analysis might be useful to inform managers, specifically:

- The risk of switching to a two-stock hypothesis if the one-stock hypothesis was actually true, and conversely, staying with the one-stock hypothesis while the two-stock hypothesis was true, should be better explored before moving unilaterally towards a two-stock hypothesis.

In addition to the one-stock and two-stock hypotheses presented by the IATTC staff, the RP also suggested various avenues be explored to better account for stock structure in the stock assessment model. These include, but are not limited to:

- A one-stock hypothesis but assuming only the LL indices are consistent;
- A one-stock hypothesis but assuming only the DEL indices are consistent;
- A two-stock hypothesis based on north and south spatial domains that may vary by season;
- A two-stock hypothesis with a stock based on dolphin-associated YFT and another stock of non-dolphin-associated YFT; and
- A hypothesis of a common pool of recruits that gets apportioned between two areas.

Although the 'North' and 'South' reference models were nominally spatial models, the models were more consistent with a model for dolphin-associated YFT ('North' model) and a model for non-dolphin-associated YFT ('South' model). This was because the 'North' model was based on fisheries that had data that were more consistent with the DEL index, while the 'South' model was based on fisheries that had data that were more consistent with the LL index. The DEL fisheries were predominantly in the northern hemisphere and all the fisheries in the 'North' model were in the north. However, the fisheries and LL index in the 'South' model included data from the entire EPO. There was no biological evidence to support the hypothesis that dolphin-associated YFT were different from the non-dolphin-associated YFT. As such, the RP recommended that:

- Data and models be consistent with explicit and plausible hypotheses or conceptual models. For example, if a model is nominally a spatial model, the spatial domain should be defined, and the data preparation and models should be consistent for the spatial domain.

5. EPO YELLOWFIN TUNA BIOLOGY

5.1. Growth

Age and growth of YFT in the EPO were based on OTC-validated daily otolith increments up to age-4. The RP noted the shape of the growth curve was essentially linear up to age-4, with little age data to inform the growth relationship after age-4, either from annual or daily increments. This is likely to result in a biased and highly uncertain growth model, especially the L_{inf} parameter. Unlike the BET growth model, there is currently not enough tagging data from large YFT to reduce the bias and uncertainty of the L_{inf} parameter.

IATTC staff have concluded that growth of YFT in the EPO is best estimated using daily otolith increments up to age-4, and have rejected the conclusions of other studies conducted on bigeye and yellowfin in the Pacific as well as other oceans that suggest annual otolith increments could be a plausible alternative measure of growth for tropical tuna, particularly at larger sizes. Several studies have concluded that for some longer-lived species, daily increments become very narrow once the size of the fish nears L_{inf}, making them difficult to discern and quantify, and that there can be areas within an otolith where daily increments are either lacking or difficult to interpret (*e.g.* Sardenne *et al*. 2015). This may cause an underestimate of the true age for older individuals.

Conclusive validation of annual increment deposition is not available for YFT. However, a recent preliminary study indicated that YFT in the Atlantic deposit a single opaque and translucent zone each year (Ailloud *et al*., 2020). While these results were limited in scope, the authors concluded that using daily increments underestimated the age for Atlantic YFT larger than 55 cm FL either because increments were not systematically deposited on a daily basis and/or were difficult to interpret. A study conducted in the WCPO by J. Farley (IATTC 2019) also found that clear daily increments were observed in otoliths close to the primordium but described an interruption at ~150-180 increments which suggested that counts of daily increments were not useful for ageing large/old yellowfin in the WCPO. Given this, the RP strongly recommended:

- Continued research efforts into the age and growth of YFT, especially for YFT >150 cm.
- Given the uncertainty in growth, the RP supports the estimation of the growth parameters within the integrated model. The RP also suggests that additional data sources be considered to inform the fit of the growth curve (*e.g.*, tagging data):
- The RP suggested more collaboration with WCPO scientists working on tropical tuna growth;
- The RP suggested the exploration of hybrid methods combining daily and annual increments into final age estimates to balance the strengths and weaknesses of both methods;
- Over the longer term, design a tagging program for YFT over larger areas beyond those currently sampled, in conjunction with OTC marking; and
- The RP suggested evaluating the potential effects of density-dependence in growth.

The available information suggests that alternative growth relationships are still preliminary and that this should be identified as a potential source of uncertainty in the assessment. However, the RP noted that given the uncertainty in the model structure and in the spatio-temporal interpretation of CPUE and length frequency observations, that model sensitivities with alternative growth relationships should be low priority for the 2020 assessment.

5.2. Natural mortality (*M*)

The current EPO yellowfin tuna assessment model used a mortality function derived from tagging data (Hampton, 2000) and assumed that *M* was age and sex-specific (Maunder and Aires-da-Silva, 2012). Female *M* increased after maturity, while male *M* does not. The change in female *M* was assumed to occur at 2 years (8 quarters) in EPO yellowfin based on an examination of the sex ratio.

Given the paucity of information available to estimate natural mortality, the RP did not make specific recommendations with regard to the functional forms used in the proposed population models. However, the RP did note that the assumed *M* function will be sensitive to changes in the assumed growth function and sex ratio at age, and that any changes in the parameterization of these functions will require re-estimation of *M*. The RP also recommended sensitivity runs and diagnostics as described below.

- Explore the potential to estimate *M* within the stock assessment model by pre-specifying *M* at Age-0 and estimating the difference (offset) in *M* between mature females and males (*e.g.* as informed by the sex-ratio).

The RP also noted that there are a range of values of *M* used in assessments of yellowfin tuna from other regions. The preference would be to select *M* values that have been derived from models that incorporate a substantial amount of tagging data that may inform the model about *M*, at least for the range of age classes represented in the tagged portion of the population. Given these considerations, the RP recommended that:

- IATTC staff summarize the available information on estimates of *M* and growth for yellowfin tuna, and document the estimates used in the assessment.
- IATTC surveys plausible alternative relationships, in particular taking into consideration any correlation between *M* and growth.
- Model sensitivities should include (at a minimum) a plausible alternative level of *M* that is lower than the level of overall *M* currently assumed.

5.3. Recruitment and spawner-recruit relationship

The RP did not specifically discuss the choice of the spawner-recruit relationship, but did examine the sensitivity of the model to sigma-R. Currently, the model uses a Beverton-Holt relationship with steepness fixed at 1.0 and sigma-R fixed at either 0.6 or 1.0. Due to the exploratory nature of the work presented, it was not possible to determine the most appropriate parameterizations of the spawner-recruit relationship. However, the RP recommends that:

- Standard Stock Synthesis bias-correction procedures (*e.g.*, Methot and Taylor 2011) be considered before final models are selected, including an attempt to estimate sigma-R.
- In exploratory models reviewed by the RP, it appeared there was some support for values of sigma-R larger than 0.6. If sigma-R is ultimately fixed, a likelihood profile on sigma-R is recommended to inform the selection of the most appropriate value for that parameter.

The RP also noted that the steepness value of 1.0 used for the assessment of tropical tuna, including YFT, is the highest amongst all RFMOs, but did not further consider this issue as it had already been well covered by the recent IATTC external review of bigeye tuna (IATTC 2019).

The RP also noted that overall recruitment patterns are comparable from the two sets of abundance indices although there are differences at certain periods and these differences are pronounced in some years (possibly/probably related to El Nino conditions). There is a potential to develop a spatially stratified model to accommodate differences in recruitment between the two fishery regions, but the RP also acknowledged the increased complexity of that model related to movement assumptions and relative biomass scaling between the two regions.

6. MODEL STRUCTURE

6.1. Fishery structure

The IATTC staff proposed that the YFT in the EPO be considered as two separate stocks for the purposes of assessment (see also Section 4 on Movement and Stock Structure). While there was some evidence of two potential recruitment pools (one in the north, and another in the south), the RP requested:

- Further analyses be conducted to assess the impacts of managing with two stocks rather than one, given that some exchange of individuals between stocks remained highly likely.
- In addition to the one-stock and two-stock hypotheses presented by the IATTC staff, the RP also suggested the IATTC staff explore other examples of plausible models and hypotheses with alternative stock structures (see Section 4.2).

The RP noted that the approach of using binary tree splits to develop hypotheses for the purse seine fisheries based on the homogeneity of the proportions at length data by season and spatial cell (5x10 degree cells) was appropriate, but noted that the analyses may not adequately account for seasonal growth, variable recruitment, or locations where there was incomplete length frequency data. Therefore, the RP recommended:

- Further analyses on the DEL fisheries be developed, and suggested that IATTC consider converting the length frequency to an approximate age frequency via the mean growth curve to test the current fishery definitions.

The RP noted the presence of bimodality in some of the aggregated length frequencies in the tree-regression, and suggested that this bimodality could potentially be a result of aggregating across too large a spatial scale or represent recruitment pulses. Therefore, the RP recommended:

- Further analyses of the length frequency data to help determine the cause of the bimodality.

- Further, the RP noted that additional fine-scale data are now available from the Japanese LL fleet, and suggested that the definition of the LL fisheries be re-investigated using these new data.

6.2. Selectivity

The RP noted that there was good evidence for domed-shaped selectivities in some fisheries based on the length composition of the samples. However, for fisheries catching larger fish, it was not clear whether high natural mortality or misspecification of the growth curve may have led to the choice, possibly inappropriate, of either domed-shaped or logistic selectivities. Therefore, the RP noted that:

- It may be necessary to use a more complex functional or non-parametric form of selectivity (*e.g.* cubic spline) to adequately represent the selectivity of the PS vulnerable population due to the bimodal distribution of the population length compositions.

Model trials with a cubic spline selectivity estimated a more complex selectivity (non-asymptotic) and fitted the left shoulder of the distribution. The more complex selectivity may be accounting for deficiencies in the population length composition generated by VAST which is smoothing across length data from fisheries with smaller/larger fish depending on space.

- The RP also noted there is evidence of a recent increase in the size of yellowfin tuna caught by the Japanese LL fishery (see section 3). This suggests there could be a need to explore time-varying selectivity in that fishery, nothing it could also be a feature of a stock responding to a reduction in fishing pressure.

6.3. Catchability

The RP did not evaluate information pertaining to an increase in catchability, nor make any specific recommendations regarding catchability.

6.4. Fishery Start Year

The proposed models start in 1975, requiring estimation of initial conditions using only the length composition data (as the initial equilibrium catch penalty is turned off in the likelihood). The RP noted that the initial biomass and depletion estimates may be sensitive to the specification of the initial conditions. In some instances, the preliminary models also estimated an initial equilibrium catch that was much greater than the observed historical catches from the fishery prior to 1975. This behavior can be indicative of model mis-specification and should not be ignored in the final models.

The original rationale for the model to start in 1975 was the availability of hooks per basket starting in 1975 for the Japanese LL fishery and port sampling for purse seine length composition starting in 1975. Japanese LL and PS catch data are available since the 1950s and could therefore be used to justify starting the model further back in time. Allowing the model to condition on the longer time-series of removals could improve the initial depletion estimates, particularly if the time-series can be extended back to a time period with negligible removals. Given these considerations:

- The RP recommended that the sensitivity of the model to initial conditions be explored by estimating removals prior to 1975 (*e.g.* 1950s) and/or by implementing an equilibrium catch penalty.

7. MODEL SENSITIVITY AND DIAGNOSTICS

7.1. Data Weighting

The RP noted that data weighting can have a large influence on model results, particularly if the data components are in conflict with one another, or are inconsistent with model assumptions. Data weighting in the YFT assessment model was determined by the lognormal standard deviation for the CPUE indices and the multinomial sample size for the length composition data. The CV estimates produced by VAST plus a constant (to average 0.2) were assumed for the LL CPUE index. The initial sample sizes for the PS length-composition data were based on the number of wells sampled, whereas the initial sample sizes for the LL length-composition data were set to values representing comparable weights to the main purse seine fishery while retaining the relative weighting of individual samples based on the number of fish sampled. Data-reweighting was explored using the integrated Francis (2014) reweighting procedure in Stock Synthesis. The RP did not specifically discuss or recommend any particular data weighting scheme, but noted that:

- The constant CV of 0.2 applied to the CPUE indices appeared unwarranted. Annual estimates of CV are available, and should be used in the final models. If a higher observed CV on indices is desired, the RP recommends that the CVs be scaled in such a way to preserve the interannual variation.

7.2. Diagnostics

Given that only preliminary/exploratory models were available, few diagnostics were presented to the review. The RP did not request a more complete set of diagnostics during the meeting because the models presented suggested that basic changes in data and model parameterization were still needed. For example, the lack of fit to the VAST length composition indicated a conflict between this data and other sources of information in the model (*e.g.* index, fleet composition data). Also, the lack of fit to the PS length composition data suggested the need to consider a more flexible selectivity function (*i.e.* cubic spline) for those fleets.

To examine the quality and stability of the final models, the RP recommends a comprehensive evaluation of a suite of diagnostics including:

- Improved diagnostics for the model fits to length composition data, including Pearson boxplot residuals by length class, year of observations, and potentially cohort (but less likely to be possible).
- An examination of the covariance matrix to identify evidence of improper model formulation. In particular, these tables should be scrutinized for evidence of high coefficients of variation and strong correlations between selectivity patterns and growth parameters.
- Trace plots to explore the parameter estimates relative to the phase of estimation. In a well-behaved model, the parameters should not change a great deal after the final phase of estimation. Large changes after the final phase of estimation can be alleviated by reconfiguring the phases of estimation.
- A "Jitter" analysis to examine the effect of varying initial parameter values on model results. A well-behaved model should converge on a global solution across a reasonable range of initial values.
- Likelihood profiling to evaluate model performance across a range of values for key input parameter (*e.g.* steepness, R0, sigma-R). Ideally the profile should be a smooth functional shape. Abnormal model behavior can be the result of model instability, and is indicated by numerous spikes and saw-toothed profiles.
- Bootstrapping: The performance of stock synthesis can also be evaluated by plotting the distribution of parameter estimates and derived quantities across bootstrapped replicates. In a well-behaved model, the maximum likelihood estimate should be similar to the mean of the bootstraps.
- Retrospective analysis to identify changes in estimates of population size, fishing mortality, recruitment, etc. that occur as years of data are added to, or removed from the model (*e.g.* Hurtado-Ferro *et al*, 2015). The severity of the retrospective pattern should be quantified using Mohn's ρ.
- Hindcasting (Kell *et al*. 2016) can also be a useful model diagnostic. In a hindcast the model is fit to the first part of a time series and then projected over the period omitted in the original fit.

Prediction skill can then be evaluated by comparing the predictions from the projection with the observations.

7.3. Other recommendations

The RP noted that a bridging (or 'stepwise') analysis should be included when documenting the assessment to show the impacts of successive changes in model decisions and input treatment between the previous and current assessment model.

8. SUMMARY OF PUBLIC COMMENTS

The RP received comments from the public after the conclusion of the review. These are briefly summarized here:

- Thanks were given to the IATTC scientific staff and the RP from the representative of Ecuador. Noting the lack of strong conclusions about the spatial split for the assessment model, there was concern from Ecuador about the management advice coming for the next year's assessment.
- A member of the public wanted to acknowledge the pressure for IATTC staff of having to provide management recommendations to the IATTC SAC in 2020, noting that staff are expected to present results that will inform the next conservation measure. A suggestion was given of an alternative option whereby if problems with the new YFT model were not resolved within an appropriate timeline, the SAC might recommend prolongation of the current resolution for an additional year to avoid pressure to the IATTC staff. The member of the public noted it was better to have confidence in the results than to not fully support the conclusions presented to the SAC. Thanks were also given to the RP for their considerable effort in providing a fresh point of view.
- Thanks were given to the IATTC scientific staff and to the RP for their hard work from the representative of Japan, noting also the ongoing collaborations with IATTC with regards to data.
- Thanks were given to the RP from the Korean delegation, with a mention of their motivation to contribute to the CPUE analysis for the EPO area and other work, as well as their ongoing collaboration with IATTC staff.

9. SUMMARY OF CONCLUSIONS AND FUTURE WORK

By the conclusion of the meeting, the RP did not consider any model that emerged as a strong candidate for the 2020 assessment. Instead, the RP concluded that substantial model development was still required before candidate models would emerge.

One of the key aspects of this required model development would be the production of the CPUE indices. These indices would likely include a DEL PS CPUE index (or indices) in the area north of the Equator and an LL CPUE index south of the Equator. The candidate model structure will depend to some degree on the reliability and representativeness of each index, and coherence in the trends of both indices given the relative selectivity of the corresponding fisheries.

However, it should be noted that conflict between these indices have already been identified by the IATTC staff. Hence, if the identified conflict is not resolved, it will likely be necessary to select a primary index for the entire stock (likely the LL index as for the current assessment model) and/or accommodate the differences through spatial partitioning of the model. If this spatial partitioning approach is taken, the RP recommends that the IATTC staff evaluate a range of spatial structure options that were not presented to the RP.

9.1. Spatial structure

The RP saw some evidence that there is limited exchange between the southern and northern 'populations', both from patterns in the CPUE and length-composition data, and from the movement patterns from tagging. However, it is unclear that patterns in CPUE and length-composition data are due to a real biological process. At this stage, the RP does not feel there is enough evidence to justify the two-stock hypothesis as the only plausible hypothesis.

Data and models should be consistent with explicit and plausible hypotheses or conceptual models. For example, if a model is nominally a spatial model, the spatial domain should be defined, and the data preparation and models should be consistent for the spatial domain. Given the information supporting the two-stock hypothesis was suggestive rather than conclusive, alternative models should continue to be developed that are consistent with both one-stock and two-stock hypotheses, as well as models somewhere between those extremes. If models with alternative stock structures are developed, management consequences of using a two-stock hypothesis if a one-stock hypothesis was actually true, and vice versa, should be evaluated (*e.g.* risk analysis).

Future model development should investigate the potential to incorporate more complex spatial structure within a single EPO wide model. In addition to the one-stock and two-stock hypotheses presented by the IATTC staff, the RP also suggested several examples of plausible models and hypotheses with alternative structures. These include, but are not limited to: 1) a one-stock hypothesis but assuming only the LL indices are consistent; 2) a one-stock hypothesis but assuming only the DEL indices are consistent; 3) a two-stock hypothesis based on north and south spatial domains that may vary by season; 4) a two-stock hypothesis with a stock based dolphin-associated YFT and another stock of non-dolphin-associated YFT; and 5) a hypothesis of a common pool of recruits that gets apportioned between two areas.

9.2. VAST approach

VAST in its current formulation, like most spatial models, makes the assumption that the response variable is independent from sampling intensity. There appears to be an assumption by IATTC staff that the use of VAST is a way to control for preferential sampling. This is a key issue for all fisheries relying on the use of fishery-dependent CPUE as a proxy for abundance.

In the EPO, this issue coupled with the contraction of the longline effort increases the risk that the VAST approach applied in its current form results in biased estimators of the modelled quantities of interest, even when diagnostics are good. While VAST may help with issues of uneven spatio-temporal effort coverage, it still does not resolve the issue of preferential sampling.

The RP recommends developing sensitivities (CPUE and COMP) using traditional and VAST approaches, and that the quality of the resulting model fits be evaluated using appropriate diagnostics. VAST model results should be selected over traditional approaches only if an improvement in diagnostic performance can be demonstrated. Staff should also explore whether the model is sensitive to the patterns of preferential sampling occurring in the EPO. Finally, if possible, staff should seek to explore the impacts of preferential sampling on the reliability of standardized CPUE as an index of abundance as well as investigate potential solutions, or collaborate with fisheries scientists already leading active research on this topic.

9.2.1. VAST population length composition:

The RP has reservations regarding the reliability of population-based length compositions derived from the VAST spatial modelling approach. However, insufficient information has been provided to understand the mechanism for the generation of the composite length data sets (*e.g.* spatial correlations). Useful, easily interpretable diagnostics disaggregated in space and time could be useful in this regard.

9.2.2. Index of population abundance

There is a need to evaluate the influence of the two components of the CPUE model (catch probability and abundance), including a detailed examination (in space and time) of the model residuals especially when integrated within a complex spatial modelling software such as VAST.

9.2.3. VAST environmental covariates

The RP acknowledges that there is an issue with the inclusion of oceanography covariates as density or catchability effects. Given the lack of catchability covariates assumed in the current model (beyond hooks-between-floats), the model predictions would be improved by the addition of density covariates like oceanography. Inclusion of oceanography covariates could:

- give additional information to the model for imputing CPUE across large areas without observations
- improve the ability of the model to account for seasonal effects on CPUE (beyond the use of quarter alone)
- Partly resolve the potential issue of varying spatial covariance relationship over the spatial domain, if some of this covariance structure is caused by the spatial distribution of oceanography data.

9.3. Indices

9.3.1. Purse seine indices

The PS CPUE from the dolphin fishery is likely becoming a more significant input for the stock assessment given the decline in the spatial extent of the longline fishery. There are recognized limitations in the utilization of PS CPUE indices in tuna stock assessments. Increased scrutiny is required to refine the dolphin PS catch and effort data set to ensure the data set is representative with respect to vessel operation and fleet structure, and can be standardized to yield a reliable index of abundance. It is recognized that there are likely to be long term trends in increasing fishing efficiency that have not been accounted for in the current standardization of PS catch and effort data (related to improvements in fishing technology), especially given the history of activity of the vessels selected as part of the core fleet. IATTC staff should prioritize investigating approaches to account for effort creep in the dolphin-associated purse seine indices.

More specifically, the RP recommends that IATTC staff consider:

- Working with stakeholders (*e.g.*, member countries, fishing companies) to define a timeline of technological changes that could inform on time-varying Q.
- Refining the covariates used to account for effort creep beyond the vessel identifier.
- Improving the quantification of effort (*e.g*. remove set processing time and other non-search time activities).
- Examining model sensitivity to (arbitrary) increases in fishing power over time.

9.3.2. Longline indices

Longline CPUE indices have likely become less reliable over the last decade as the fishery has contracted. However, for comparative purposes and while reliable PS CPUE indices are being developed, it is necessary to maintain an alternative model option that incorporates a LL CPUE index (derived using more established standardization approaches) as the principal index of abundance.

9.4. Life history information

9.4.1. Growth

IATTC staff concluded that growth of YFT in the EPO is best estimated using daily otolith increments. The RP noted the shape of the growth curve was essentially linear up to age 4, with little age data to inform the growth relationship of older/larger fish. Given the uncertainty in growth, the RP supports the estimation of the growth parameters within the integrated model.

- Hybrid methods combining daily and annual increments into final age estimation should be explored to balance the strengths and weaknesses of both methods.
- The RP also suggests that additional data sources be considered in that fitting (*e.g.* tagging data).

9.4.2. Natural mortality

Given the paucity of information available to estimate natural mortality, the RP did not make specific recommendations with regard to the functional forms used in the proposed population models. However, the RP notes that:

- Age-specific M is sensitive to assumptions regarding growth and sex ratio at age. Therefore, any changes in the parameterization of these functions will require re-estimation of M.
- Model sensitivities should include a plausible alternative level of natural mortality that is lower than the level of overall M currently assumed.

9.4.3. Recruitment and the spawner-recruit relationship

Due to the exploratory nature of the work presented, it was not possible to determine the most appropriate parameterizations of the S-R parameters. However, the RP recommends that:

- Standard Stock Synthesis bias-correction procedures (*e.g.* Methot and Taylor 2011) be considered before finals models are selected, including an attempt to estimate sigma-R.
- A likelihood profile on sigma-R is performed to inform the selection of the most appropriate value of that parameter if sigma-R is ultimately fixed, given it appeared there was some support for values of sigma-R larger the 0.6.
- Consider other plausible values for the S-R Relationship and account for model sensitivity to those options when discussing reference points, as these are highly sensitive to these choices, and the IATTC staff uses a default parameterization for the S-R relationship that differs markedly from that used in other yellowfin tuna assessments (steepness of 1.0).

9.5. Model diagnostics

To examine the quality and stability of the final models, the RP recommends a comprehensive evaluation of a suite of diagnostics including:

- Improved diagnostics for model fit to length composition (*e.g*. Pearson boxplot residuals by length class, year of observations, and potentially cohort).
- An examination of the covariance matrix to identify evidence of improper model formulation.
- A “Jitter” analysis to examine the effect of varying initial parameter values on model results.
- Likelihood profiling to evaluate model performance across a range of values of a model parameter (*e.g*. steepness, R_0, sigma-R).
- Bootstrapping: In a well-behaved model, the maximum likelihood estimate should be similar to the mean of the bootstraps.
- Retrospective analysis: The severity of retrospective patterns should be quantified using Mohn’s ρ.
- Hindcasting (Kell *et al*. 2016)

9.6. Other recommendations

The RP noted that the methodology used to groom, process and analyze the inputs used in the assessment and resulting diagnostics were seldom included in reports presented to the SAC. Describing the exact methodology used in producing the inputs to stock assessments presented for any given year as well as key accompanying diagnostics are key to ensuring the transparency, reproducibility and quality of the science used to inform fisheries management.

The RP feels that this is best-practice for stock assessments, even when stakeholders do not review technical documents. It is especially important when the methodology is evolving between years, as is currently the case for the CPUE standardization and the processing of the length composition data. Other RFMOs (*e.g.* ICCAT, IOTC and WCPFC) all include companion data input reports to their stock assessments.

APPENDIX A: Areas of focus identified by IATTC staff

Main areas

1. What causes the mismatch in the longline and purse seine CPUE based indices of relative abundance?
2. What is the most appropriate stock structure for the yellowfin tuna stock assessment?
3. What is the most appropriate fishery structure for the yellowfin tuna stock assessment?
4. What approach should be used to deal with the uncertainty in the length of old individuals and the impact it has on the stock assessment results?
5. Can you estimate absolute abundance?
6. What is the appropriate stock-recruitment relationship?
7. How should the CPUE indices of abundance be used in the stock assessment?
8. Should logistic selectivity be used and for which fishery/survey?
9. How should the recent increase in the size of longline caught fish be dealt with?
10. Age- and sex-specific natural mortality

Other topics of interest

1. Time varying selectivity
2. Reference points
3. Using tagging data
4. Methods used to create the length-composition data
5. Inclusion of age conditioned on length data
6. Definition of spawning biomass
7. Diagnostics
8. Research recommendations
9. Data collection
10. Data analysis

APPENDIX B: Terms of reference

1. Goals and objectives

The purpose of the review of the IATTC staff's assessment of the yellowfin stock is not to determine whether the current or proposed assessment is adequate for providing management advice; the intention is to provide information to the assessment team to improve the assessment. The goals and objectives of the review are to:

a. identify the best available science for use in the assessment;
b. provide an independent review of the assessment; and
c. provide advice on future research and data collection that will improve the assessment and the provision of management advice.

2. Review panel responsibilities

The main responsibility of the RP is to perform an adequate technical review of the assessment. The members of the Panel should disclose any conflicts of interest that could significantly impair their objectivity. Conflicts of interest include, but are not limited to, personal financial interests and investments, employer affiliations, and consulting arrangements, grants, or contracts.

The specific responsibilities of the Panel are to:

a. be familiar with the Terms of Reference;
b. review background documents, data inputs, and analytical models, along with other pertinent information (*e.g.*, previous assessments and Review Panel reports);
c. discuss the technical merits and deficiencies of the input data and analytical methods, work with the IATTC staff to correct deficiencies, and, when possible, suggest new tools, analyses, and data collection methods to improve future assessments; and
d. draft a report of the meeting, to document the discussions and recommendations.

It is the Panel chair's responsibility to coordinate the discussions so that the review is completed in the time available.

3. Public comment

Time will be allocated during the meeting for public comment. The Panel will take these comments into consideration when developing its report, as appropriate.

4. Requests for additional analyses

The meeting is intended as a technical review of the assessment methodology, and additional analyses during the meeting may be beneficial. In the course of the meeting, the Panel may request a reasonable number of sensitivity runs, additional details on the models presented, or further analyses of alternative runs. However, any such requests must be clear, explicit, and be presented in writing, and be practical in terms of the time available. They should be listed individually in the Panel's report, along with their rationale and the response. To the extent possible, analyses requested by the Panel should be completed during the meeting by the assessment team.

5. Panel report

The Panel's report should be drafted and approved shortly after the meeting. The report writing process will follow these steps:

a. Panel outlines report at meeting;
b. Panel writes and agrees draft report;
c. Panel provides draft report to IATTC staff for comment on technical accuracy; and
d. Panel reviews staff comments, and modifies report as necessary. The report will include:
a. Names and affiliations of Panel members;
b. Brief overview of the meeting (location, agenda, main recommendations by Panel, etc.);
c. Brief summary of current assessment model, data used, analyses presented, and proposed assessment model;
d. List of analyses requested by the Panel, rationale for each request, and brief summary of the response;
e. Comments on technical merits and/or deficiencies in the assessment, and recommendations for remedies;
f. Unresolved problems and major uncertainties, *e.g.*, any special issues that complicate the assessment and/or interpretation of results;
g. Data, fishery or assessment related issues raised by the public; and
h. Prioritized recommendations for research and data collection for the subsequent assessment.

The Panel and the IATTC staff will strive to resolve any differences of opinion that may arise regarding the contents of the report. Any unresolved differences of opinion must be documented and reflected in the report, which will be published as an IATTC Special Report.

References

Nokome Bentley, Terese H. Kendrick, Paul J. Starr, Paul A. Breen, Influence plots and metrics: tools for better understanding fisheries catch-per-unit-effort standardizations, *ICES Journal of Marine Science*, 69 (1): 84–88.

Martell, S.J.D., de Bruyn, P., Davies, Nick *M*., Ernst, B. (2013). Recommendation of the review panel on the IATTC assessment of yellowfin tuna. Special report 20 submitted to the IATTC, La Jolla, California, USA.

APPENDIX C: List of participants

Name	Affiliation	Email
Cass-Calay, Shannon	USA	shannon.calay@noaa.gov
Dreyfus, Michel	Mexico	dreyfus@cicese.mx
Dunn, Alistair	New Zealand	alistair.dunn@oceanenvironmental.co.nz
Fleischer, Luis	Mexico	lfleischer21@hotmail.com
Langley, Adam	New Zealand	adam_langley@xtra.co.nz
Kim, Doo Nam	Korea	doonam@korea.kr
Lee, Sung Il	Korea	k.sungillee@gmail.com
Morán, Guillermo	Ecuador	gamv6731@gmail.com
Piner Kevin	USA	kevin.piner@noaa.gov
Quiroz, Juan	Ecuador	jc.quirozespinosa@gmail.com
Satoh, Keisuke	Japan	kstu21@fra.affrc.go.jp
Schiff, Tom	USA	tschiffsd@aol.com
Teo, Steven	USA	steve.teo@noaa.gov
Tremblay, Laura	New Zealand	laura@dragonfly.co.nz
Urtizberea, Agurtzane	EU	aurtizberea@azti.es
IATTC staff		
Compeán, Guillermo		gcompean@iattc.org
Aires-da-Silva, Alexandre		alexdasilva@iattc.org
Fuller, Dan		dfuller@iattc.org
Griffiths, Shane		sgriffiths@iattc.org
Hall, Martin		mhall@iattc.org
Lennert, Cleridy		clennert@iattc.org
López, Jon		jlopez@iattc.org
Maunder, Mark		mmaunder@iattc.org
Minte-Vera, Carolina		cminte@iattc.org
Morgan, Jeff		jmorgan@iattc.org
Sarazen, Robert		rsarazen@iattc.org
Schaefer, Kurt		kschaefer@iattc.org
Valero, Juan		jvalero@iattc.org
Xu, Haikun		hxu@iattc.org

APPENDIX D: Documents available to the review panel

Technical reports:

Ailloud L., Beare D., Farley J.H., and Krusic-Golub K. (2020) Preliminary results on AOTTP validation of otolith Increment deposition rates in yellowfin tuna in the Atlantic. COLLECTIVE VOLUME OF SCIENTIFIC PAPER (76): 156-163.

Hall, M., & Román, M. H. (2019). THE FISHERY ON FISH-AGGREGATING DEVICES (FADs) IN THE EASTERN PACIFIC OCEAN. IATTC - Scientific Advisory Committee (May), 13–17.

IATTC (2012). EXTERNAL REVIEW OF IATTC YELLOWFIN TUNA DOCUMENT YFT-01-01 INTRODUCTION TO THE REVIEW OF THE ASSESSMENT OF YELLOWFIN TUNA IN THE EASTERN PACIFIC OCEAN , 2012. (October), 15–19.

IATTC (2016). STATUS OF THE TUNA AND BILLFISH STOCKS IN 2015. Stock Assessment Report 17. https://doi.org/10.7551/mitpress/3298.003.0011

IATTC (2018). REPORT ON THE TUNA FISHERY , STOCKS , AND ECOSYSTEM IN THE EASTERN PACIFIC OCEAN IN 2018 Republic of Korea United States of America.

IATTC (2019). External review of IATTC staff's stock assessment of bigeye tuna in the eastern Pacific Ocean - 11-15 March, 2019 - La Jolla, California, USA. https://www.iattc.org/Meetings/Meetings2019/BET-02/Docs/_English/BET-02-AGN_External%20review%20of%20IATTC%20staff%E2%80%99s%20stock%20assessment%20of%20bigeye%20tuna%20in%20the%20eastern%20Pacific%20Ocean.pdf

IATTC (2019). WORKSHOP ON AGE AND GROWTH OF BIGEYE AND YELLOWFIN TUNAS IN THE PACIFIC OCEAN. La Jolla, California (USA) 23-25 January 2019. https://www.iattc.org/Meetings/Meetings2019/OTM-28/PDFs/_English/OTM-28-RPT_%20BET%20YFT%20Age%20and%20growth.pdf

Martell, S. J. D., de Bruyn, P., Davies, N. M., & Ernst, B. (2013). RECOMMENDATIONS OF THE REVIEW PANEL ON THE IATTC ASSESSMENT OF YELLOWFIN TUNA.

Maunder, M.N. and Aires-da-Silva, A., (2012). A review and evaluation of natural mortality for the assessment and management of yellowfin tuna in the eastern Pacific Ocean. Inter-Amer. Trop. Tuna Comm. Document YFT-01-07.

Minte-Vera, C. V., Maunder, M. N., & Aires-da-Silva, A. (2018). Status of yellowfin tuna in the Eastern Pacific Ocean in 2017 and outlook for the future. IATTC - Scientific Advisory Committee, SAC-07-05b).

Minte-Vera, C., Maunder, M. N., Aires-da-Silva, A., & Xu, H. (2019). EVALUATING INCONSISTENCIES IN THE YELLOWFIN ABUNDANCE INDICES. IATTC - Scientific Advisory Committee, (SAC-10 INF-F), 13–17.

Peer-reviewed literature:

Aires-da-Silva, A. M., Maunder, M. N., Schaefer, K. M., & Fuller, D. W. (2015). Improved growth estimates from integrated analysis of direct aging and tag – recapture data : An illustration with bigeye tuna (*Thunnus obesus*) of the eastern Pacific Ocean with implications for management. Fisheries Research, 163, 119–126. https://doi.org/10.1016/j.fishres.2014.04.001

Francis, R. I. C. C., Aires-da-Silva, A. M., Maunder, M. N., Schaefer, K. M., & Fuller, D. W. (2016). Estimating fish growth for stock assessments using both age – length and tagging-increment data. Fisheries Research, 180, 113–118. https://doi.org/10.1016/j.fishres.2015.06.011

Hampton, J., 2000. Natural mortality rates in tropical tunas: size really does matter. Canadian Journal of

Fisheries and Aquatic Sciences, 57(5), pp.1002-1010. https://doi.org/10.1139/f99-287

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L. and Ono, K., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science, 72(1), pp.99-110. https://doi.org/10.1093/icesjms/fsu198

Kell, L.T., Kimoto, A. and Kitakado, T., 2016. Evaluation of the prediction skill of stock assessment using hindcasting. Fisheries Research, 183, pp.119-127. https://doi.org/10.1016/j.fishres.2016.05.017

Lennert-Cody, C. E., Minami, M., Tomlinson, P. K., & Maunder, M. N. (2010). Exploratory analysis of spatial–temporal patterns in length–frequency data: An example of distributional regression trees. Fisheries Research, 102(3), 323–326. https://doi.org/10.1016/J.FISHRES.2009.11.014

Lennert-cody, C. E., Moreno, G., Restrepo, V., Roma, M. H., & Maunder, M. N. (2018). Recent purse-seine FAD fishing strategies in the eastern Pacific Ocean : what is the appropriate number of FADs at sea ? ICES Journal of Marine Science, 75, 1748–1757. https://doi.org/10.1093/icesjms/fsy046

Maunder, M. N., & Deriso, R. B. (2013). A stock – recruitment model for highly fecund species based on temporal and spatial extent of spawning. Fisheries Research, 146, 96–101. https://doi.org/10.1016/j.fishres.2013.03.021

Maunder, M. N., Deriso, R. B., Schaefer, K. M., Fuller, D. W., Aires, A. M., Minte, C. V, & Steven, V. (2018). The growth cessation model : a growth model for species showing a near cessation in growth with application to bigeye tuna (Thunnus obesus). Marine Biology, 165(4), 1–9. https://doi.org/10.1007/s00227-018-3336-9

Methot Jr, R.D. and Taylor, I.G., (2011). Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 68(10), pp.1744-1760. https://doi.org/10.1139/f2011-092

Minte-Vera, C. V, Maunder, M. N., Aires-da-Silva, A. M., Satoh, K., & Uosaki, K. (2017). Get the biology right , or use size-composition data at your own risk. Fisheries Research, 192, 114–125. https://doi.org/10.1016/j.fishres.2017.01.014

Minte-Vera, C. V, Maunder, M. N., Schaefer, K. M., & Aires-da-Silva, A. M. (2019). The influence of metrics for spawning output on stock assessment results and evaluation of reference points : An illustration with yellowfin tuna in the eastern Pacific Ocean. Fisheries Research, 217, 35–45. https://doi.org/10.1016/j.fishres.2018.09.022

Sardenne, F., Dortel, E., Le Croizier, G., Million, J., Labonne, M., Leroy, B., Bodin, N. and Chassot, E., 2015. Determining the age of tropical tunas in the Indian Ocean from otolith microstructures. *Fisheries Research*, *163*, pp.44-57. https://doi.org/10.1016/j.fishres.2014.03.008

Wang, S., Maunder, M. N., Piner, K. R., Aires-da-Silva, A., & Lee, H. (2014). Evaluation of virgin recruitment profiling as a diagnostic for selectivity curve structure in integrated stock assessment models. Fisheries Research, 158, 158–164. https://doi.org/10.1016/j.fishres.2013.12.009

Xu, H., Lennert-cody, C. E., Maunder, M. N., & Minte-Vera, C. V. (2019). Spatiotemporal dynamics of the dolphin-associated purse-seine fishery for yellowfin tuna (Thunnus albacares) in the eastern Pacific Ocean. Fisheries Research, 213 (September 2018), 121–131. https://doi.org/10.1016/j.fishres.2019.01.013

Zhu, J., Maunder, M. N., Aires-da-Silva, A. M., & Chen, Y. (2016). Estimation of growth within Stock Synthesis models : Management implications when using length-composition data. Fisheries Research, 180, 87–91. https://doi.org/10.1016/j.fishres.2015.09.019

APPENDIX E: Requests and responses

The following requests were made by the RP to the IATTC staff (see responses 1 and 2)

1. Time-series of heatmaps for Japanese longline effort, longline yellowfin CPUE and bigeye CPUE for additional context and also in order to see whether there is a spatial pattern in the effort contraction in the Japanese longline fleet that might match one of the target species' CPUE. [LTB]
   The heatmaps were provided by the IATTC staff and were used as a reference during the review.
2. Can we please see a plot of the Longline CPUE over time, and compare this with an equivalent index from the aggregated length frequency index from the VAST models [AD]
3. Can we please plot the %dolphin (by effort and by catch) per vessel per month as a time series box-plot for vessels that meet the threshold of at least 75% dolphin sets [AD]
4. A time-series showing the years of activity for each vessel in the purse seine DEL fleet from 1975 onwards, with points colored or scaled according to the proportion of dolphin sets. This plot was requested for additional context on the vessel history within the fleet and to check the relevance of using a vessel factor in the CPUE standardization as a proxy for operational changes.
   The plot was provided by the IATTC staff and showed that many vessels had been active for a time period spanning many operational changes, and that there was a gradual change for some in the proportion of dolphin sets.
5. Regarding my request for the spatial comparisons of the distribution of archival tag data and purse seine fishing effort to evaluate the extent of fish movement within the domain of the area of operation of the purse seine fishery. [AL]

   I suggest using the contour maps of archival release data sets – slide 18 of Kurt's presentation.

   These could then be used to overlay the distribution of the PS DEL fishing effort for individual time period blocks (aggregated for 5? year intervals). Not necessary to do the entire time period of the PS fishery – maybe just the last 20 years.
6. Attached is a draft of the table requested on the **model structure**. We still need to fill in the information for how the purse seine catch and length comp data are calculated and the indices of abundance, but I thought you might want to look at it for the other information before tomorrow morning. You can see a map of the fisheries on the google drive. [SCC]
7. Please can you ask the assessment team to conduct an additional set of model runs (both north and south) with the all of the following changes to the current north and south models made simultaneously
    - Tighten the cv at length for the age-length relationship about 0.05
    - Relax the sigmaR value to something large, like 1.0
    - Fix the survey selectivity to be flat topped and asymptote at 1.0
    - Remove, or substantially reduce the F penalty from both models (currently set at 0.2 for year 176 – either remove it or set it to be something out of the way) [AD?]