Independent review of recent IOTC yellowfin tuna assessment

Content

Executive Summary

The Independent Review Panel conducted a review of the 2021 assessment of yellowfin tuna in the Indian Ocean from 6 and 10 February 2023 at the FAO Headquarters, Rome, Italy. The assessment authors presented a summary of the main issues, an overview of the assessments of the last 10 years, the catches and length frequencies, the inputs to the stock assessment model, and the results. The Panel identified several requests for additional model runs and data analyses that the analysts addressed between meeting sessions. During the subsequent days, the Panel evaluated the responses to its requests and reviewed the background documents. The conclusions and recommendations from the draft report were presented to the analysts on 10 February 2023, and this meeting report was finalized after the review meeting.

Several areas of priority research were identified. These include: 1) develop a conceptual model specifically to improve the understanding of spatial structure to use in the assessment; 2) conduct further extensive work to improve the longline indices of abundance and form a working group involving all the tuna RFMOs; 3) evaluate the tagging data outside the stock assessment using a fine scale spatial temporal model to address concerns with mixing; 4) conduct more growth validation work and collect more aging data with a wider spatial and temporal range; 5) conduct further work on natural mortality including the use of a sex structured model.

Abbreviations

<u>Regions</u>

R1 – North west

(R1a - Arabian Sea region)

(R1b - western equatorial region)

R2 – South west

R3 – North east

R4 – South east

<u>Fisheries</u>

PS FS – purse seine free school

PS LS – purse on log sets (schools associated with FADs or natural objects)

GN – gill net

LL – longline

<u>Other</u>

LF – length frequency

ToR – Terms of Reference

YFT – yellowfin tuna

SS3 – Stock Synthesis version 3 (the stock assessment modelling program)

IO – Indian Ocean

Introduction

The Independent Review Panel (see Appendix A for panel biographies) conducted a review of the 2021 assessment of yellowfin tuna (YFT) in the Indian Ocean based on the Terms of Reference (ToR, Appendix B), including the data inputs, the settings for the diagnostic model and those for the uncertainty grid. Prior to the meeting, an online meeting was held where the ToR was reviewed. The Panel was provided with a set of background documents (Appendix C) prior to the meeting of the Panel. Email communication followed were the Panel provided some initial requests.

The review meeting took place between 6 and 10 February 2023 at the FAO Headquarters, Rome, Italy, and was chaired by Dr Mark Maunder (see Appendix D for the list of participants). The meeting was started by Dr. Gorka Merino, chair of the IOTC tropical tunas working group, who presented a summary of the main issues with the assessment and an overview of the assessments of the last 10 years (see Appendix C for the list of presentations). Dan Fu, the principal stock assessment author, followed with a presentation of the catches and length frequencies, and another presentation specifically about the inputs to the stock assessment model. The Panel identified several requests for additional model runs and data analyses that the analysts addressed between meeting sessions. During the subsequent days, the Panel evaluated the responses to its requests (Appendix E) and reviewed the background documents. The conclusions and recommendations from the draft report were presented to the analysts on 10 February 2023, and the meeting report was finalized after the review meeting.

Modeling platform

SS3 since 2015 (or 2012 according to Rishi Sharma, personal communication)

Recent reviews

Sharma 2018 (IOTC-2018-SC21-INF02)

Methot 2019

Johnson and at 2022 (IOTC-2022-WPTT24-17)

General recommendations

Extensive research is needed to improve the assessment and address the remaining issues. Enough resources should be made available to conduct the needed research. Recommendations from this review which are outlined below focus on the process and areas of research rather than specific model configurations. Similar issues occur across all stock assessments conducted by the various tuna RFMOs and we recommend a collaborative approach to address them (e.g., given the importance of the longline CPUE, a joint tRFMO working group on longline CPUE analysis should be created). A more rigorous approach for evaluation of the various data inputs through a data preparatory meeting that allows enough time to thoroughly evaluate the data before using in the assessment should be considered.

A single area model (e.g., expanded area 1) should initially be used for exploratory analysis to facilitate the evaluation of various issues. The model should be extended to two sexes to allow more comprehensive evaluation of biological and fishery processes (e.g., allow the evaluation of sex specific growth, natural mortality, and/or selectivity). Consideration should also be given to conducting exploratory analysis treating each of the 4 regions as separate stocks.

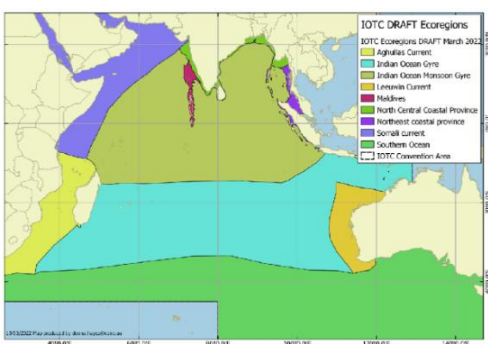
The recommendations should be considered in combination with other information such as any recommendations from the spatial modelling workshop (https://github.com/aaronmberger-nwfsc/Spatial-Assessment-Modeling-Workshop), the CAPAM Stock Assessment Good Practices workshop (http://www.capamresearch.org/GPG-Workshop) and Tuna Stock Assessment Good Practices workshop (https://www.capamresearch.org/Tuna_Stock_Assessment_Good_Practices_Workshop) as well as other CAPAM or relevant workshops, and independent reviews of other tuna assessments.

1. Stock structure

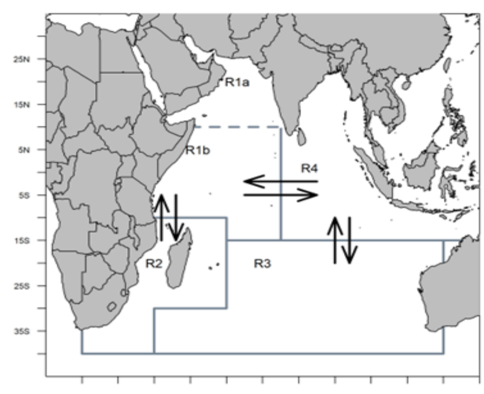
The stock assessment for yellowfin tuna in the Indian Ocean assumes that there is only one stock that has heterogeneous distribution modelled using spatial models with movement. The species is distributed in tropical and sub-tropical waters. Juveniles form mixed schools with skipjack and juvenile bigeye tuna in surface waters, and adults are found in surface and sub-surface waters. Spawning occurs mainly in the equatorial areas, with potential spawning centers in the Mozambique Channel, Seychelles archipelago, Arabian Sea, off Sri Lanka, and in the Bay of Bengal. The spawning varies seasonally and shows a peak in December to March. Both the purse-seine and the longline fisheries also show seasonal patterns, suggestive of seasonal migration, recruitment, catchability, and/or seasonal variation in favorable habitat availability. Tagging data has confirmed large scale movement within the western equatorial region. There is lack of recoveries in the eastern region, which could be due to either low effort or absence of movement. There is genomic evidence for spatial heterogeneity, with support for at least two groups, split north and south of the equator (IOTC-2020-WPTT22(AS)-12).

A map of the IO eco-regions (from IOTC-2022-WPEB18-22) was shown and compared to the catch map (Figure 1). The eco-regions are based on large-scale oceanographic features (Figure 1 bottom), Longhurst provinces and expert opinion. It is noteworthy that the ecoregions indicate epipelagic structure. The catches are mostly taken in the areas of the IO Monsoon Gyre, the Somali current and the Agulhas current. The areas of the Indian Ocean Gyre and the Southern Ocean have very little catches.

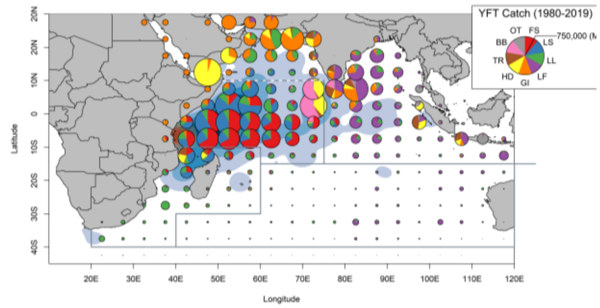
IO Ecoregions



YFT stock assessment areas



YFT catch



IO currents

Q2 Q3

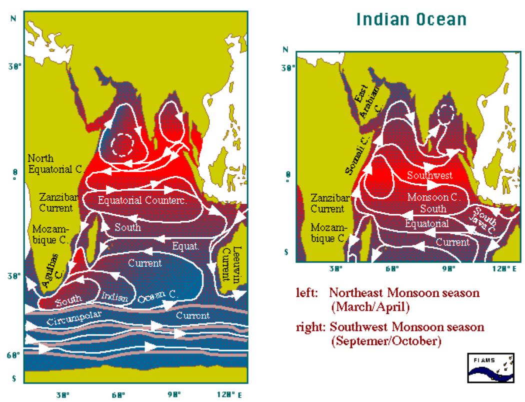


Figure 1. Top left: Indian Ocean eco-regions (IOTC-2022-WPEB18-22), top right: stock assessment areas, middle: catch distribution by gear of yellowfin tuna in the. Bottom: Indian Ocean circulation system in the northeast monsoon season and the southwest monsoon season (Le Blanc 2002).

There are some striking differences in the ecoregions and the assumed areas for the spatial model. The Somali current area and the IO Monsoon Gyre are not split in the stock assessment, which rather splits the IO Monsoon Gyre into and east and west area.

Four areas in the current assessment model were chosen based on the distribution of the fleets, the split in the Mozambique channel is based on oceanography, and the tagging data. Clustering of longline CPUE was also used to split the areas. The size data was not taken into consideration, but recent length frequency analysis seems to provide support. In previous assessments, there were 5 regions, with R1a modelled using the areas a fleet approach. Adam Langley clarified that he used to do the same as a sensitivity when he assessed this stock. The tropical regions (regions 1 and 4) are where most of the catches are taken. Some of the catches in region 3 were allocated to catches in other regions for simplicity, which decreases the number of fisheries in the model. Region 2 seems discrepant and off synchrony with the other regions. The model estimates western recruitment (Figure 2) and eastward dispersal of the recruits.

There are two regions that were discussed in detail during the review: the Arabian Sea region and the south African region. The Arabian Sea is currently modelled as a component of the western equatorial region (fisheries as areas). Thisregion lacks length composition data and indices of abundance, but may be an important area for fish of intermediate size that seem absent from the purse-seine catches. The length composition of the purse-seine fleet on free school shows a persistent bimodal pattern indicating that small fish and large fish school in the area where the PS fleet operates. The intermediate sizes either do not school, are not caught for another reason, or may not be in the area where the PS fleet operates. In this later case, it is hypothesized that they may migrate to the Arabian Sea area. Thus, obtaining better data for that area would not only allow for better estimates of the selectivity/availability of the fisheries that operate in that area, but also would contribute to a better understanding of the dynamics of the IO yellowfin tuna stock.

The southwest region, in the Agulhas current, off the coast of South Africa, is also another area of uncertainty. It is not clear if the yellowfin from that area maybe more associated with the Atlantic Ocean. The catches are not large, but the length composition data may provide signals that are not consistent with the dynamics of the core of the catches.

Request:

1. Compare the recruitments estimated for each quarter to see if there are differences, plot the recruitments using a line for each quarter

The recruitment is estimated to be higher in region 1 during quarters 3 and 4. The recruitment in region 4 shows a declining trend over time, concurrent with the increase in catches.

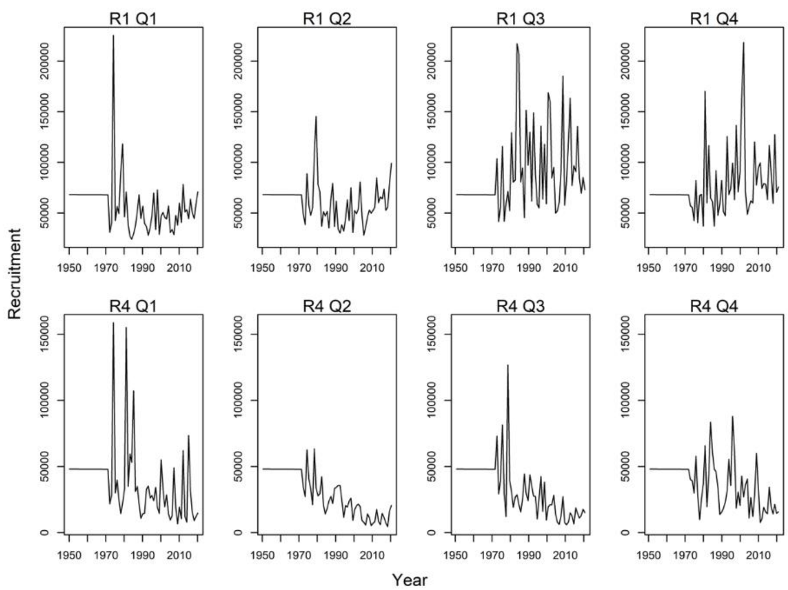
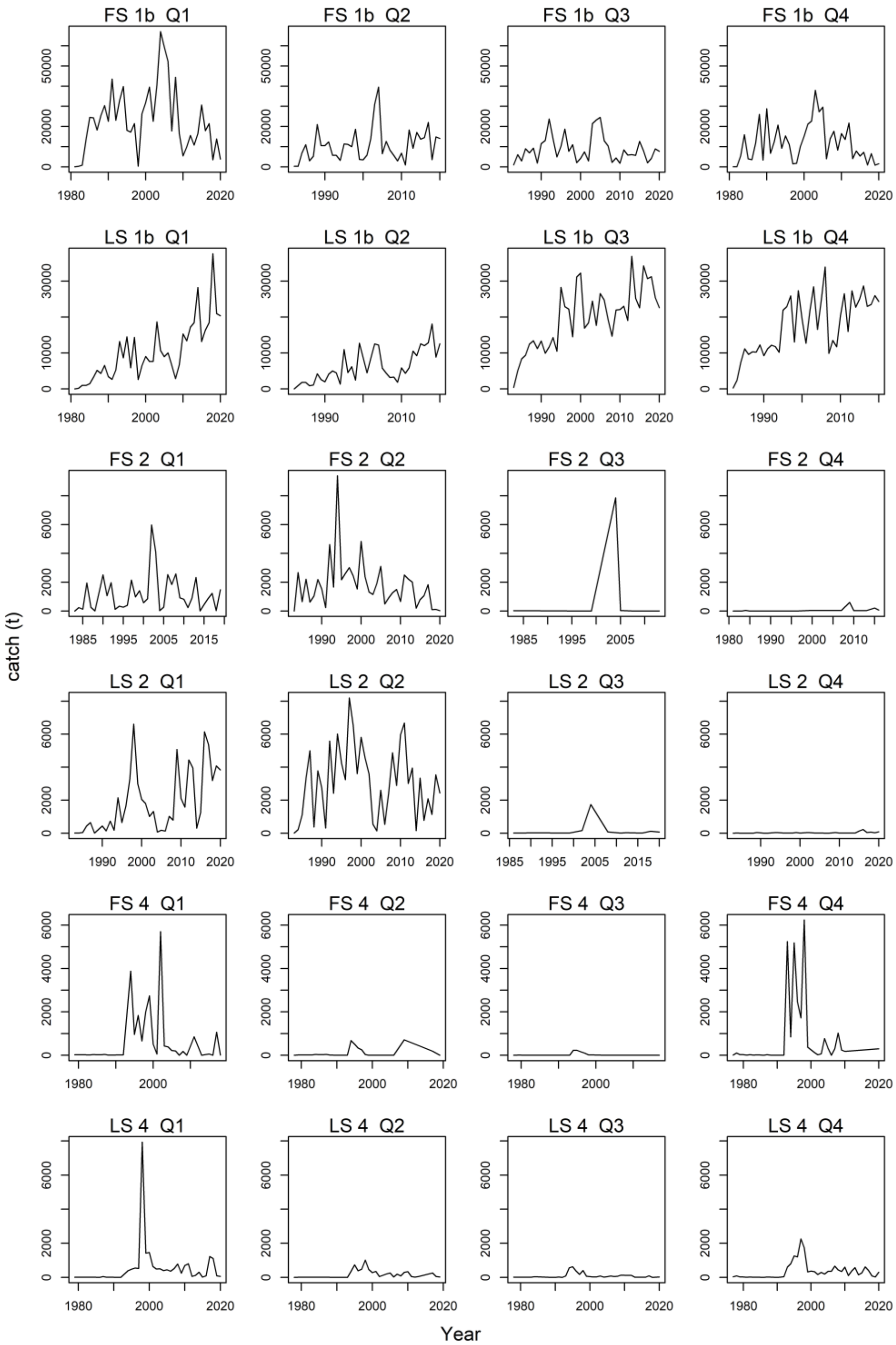
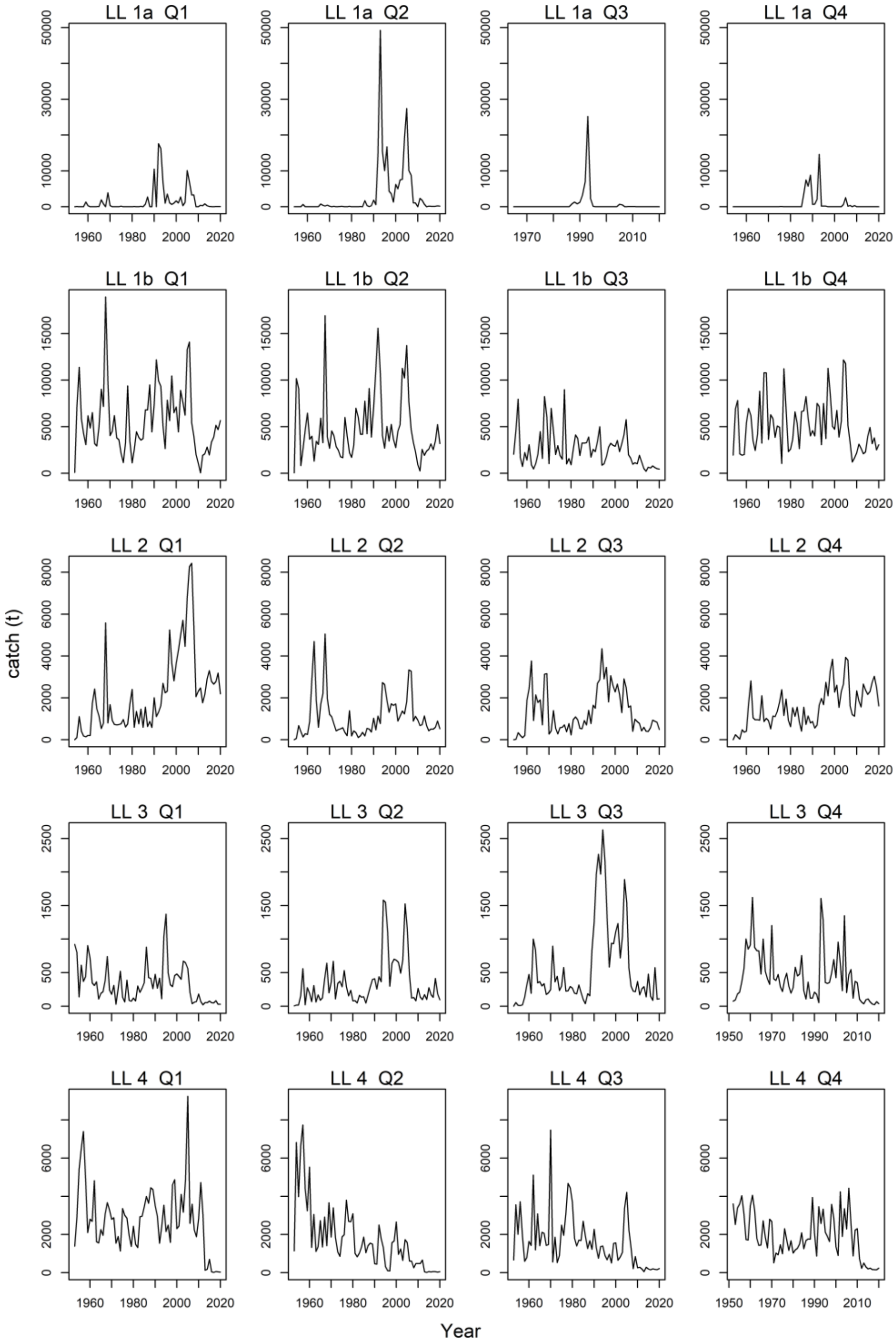


Figure 2. Estimates of recruitment by quarter for the Indian Ocean stock assessment model.

2. Plot the catches by gear and season.

The catches are very seasonal (Figure 3). In area 1, the purse-seine on free school catches are higher in quarter 1, followed by quarter 4. The purse-seine on log school catches are higher in quarters 3, 4 and 1. In area 2, the purse-seine catches are seasonal and occur almost invariably in quarters 1 and 2. Longline catches are higher in quarters 1 and 2 in areas 1b and 2, and quarter 3 in area 3. To a lesser extent, seasonality is also observed on the catches for other gears.


FS 1b Q1
FS 1b Q2
FS 1b Q3
FS 1b Q4
LS 1b Q1
LS 1b Q2
LS 1b Q3
LS 1b Q4
FS 2 Q1
FS 2 Q2
FS 2 Q3
FS 2 Q4
LS 2 Q1
LS 2 Q2
LS 2 Q3
LS 2 Q4
FS 4 Q1
FS 4 Q2
FS 4 Q3
FS 4 Q4
LS 4 Q1
LS 4 Q2
LS 4 Q3
LS 4 Q4
catch (t)
Year


LL 1a Q1
LL 1a Q2
LL 1a Q3
LL 1a Q4
LL 1b Q1
LL 1b Q2
LL 1b Q3
LL 1b Q4
LL 2 Q1
LL 2 Q2
LL 2 Q3
LL 2 Q4
LL 3 Q1
LL 3 Q2
LL 3 Q3
LL 3 Q4
LL 4 Q1
LL 4 Q2
LL 4 Q3
LL 4 Q4
catch (t)
Year

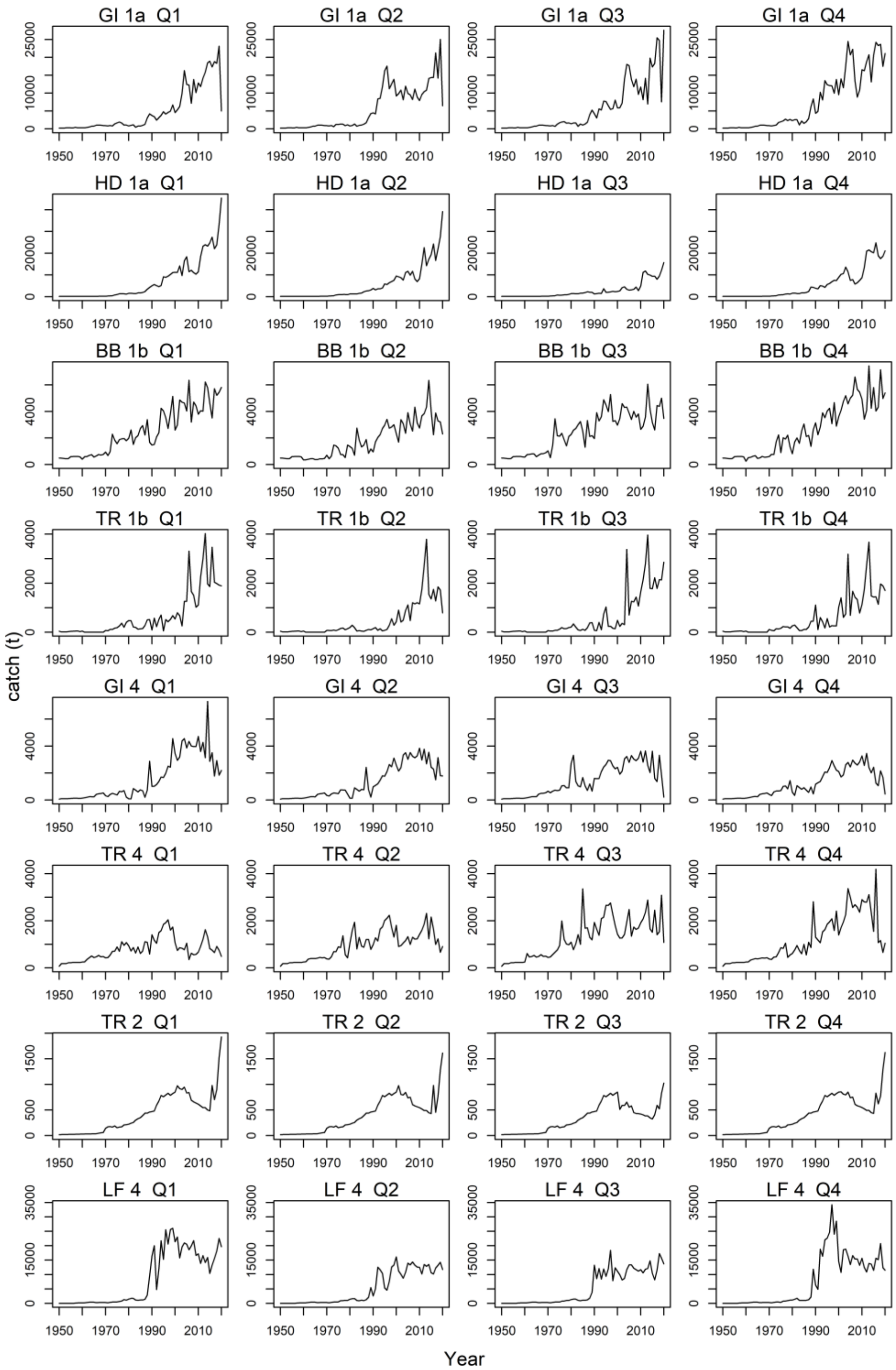


Figure 3. Catches by gear, areas and quarter for the stock assessment of yellowfin tuna in the Indian Ocean.

Recommendations:

1. <u>Comprehensive review:</u> Review all available information related to stock structure (e.g. genetics, oceanography including mesopelagic structure, biological information, length comps, etc.)

2. Conceptual model: Build a conceptual model about stock structure, making the model areas consistent with the oceanographic areas.

3. Area definitions: Use regression trees or similar analysis to explore the available length frequency data and evaluate whether it is consistent with the conceptual model. Use these analyses to guide the definitions of areas and/or fisheries.

4. Connectivity: Explore the connectivity with the Arabian sea (comparing model with and without catches and length frequencies from that area). Reconsider the stock structure, area definitions, and connectivity and how much the stock is linked to both the Atlantic Ocean stock (south African side) and the Pacific Ocean stock (in the Indonesian and Australian coasts) based on available evidence.

9. Indices of abundance: Develop an index of abundance for the Arabian sea area

6. Model complexity: For modelling, implement a one area model including the core of the catches to improve selectivity, growth, and natural mortality assumptions. Start by expanding Area 1 to the South and possibly to the East. Add other areas and movement dynamics as a second step, consistent with the conceptual model. The current spatial model does not estimate much movement between east and west. Thus, it is already effectively two independent models (east-west). If there is not significant mixing estimated in the model between east and west, then give advice separately for the two (this can be done with or without single area models)

Recommendation for future research:

1. Stock structure: Investigate stock structure and connectivity using appropriated techniques (such as archival tags and close kin analysis to have confirmation on stock structure assumptions), integrating all available information.

2. Fishery definitions

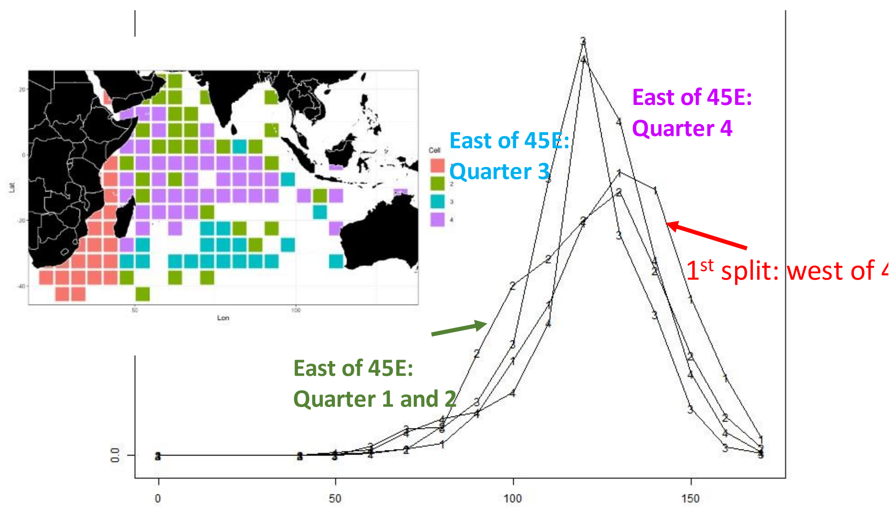
The current assessment model structures the fisheries based on region and gear type, representing a total of 21 fisheries in the model. Smaller juvenile yellowfin are caught by PS associated (FAD) sets and a number of artisanal fisheries (e.g., troll). The longline (LL) and PS free school (FS) fisheries predominantly catch large, adult fish. Although the latter fishery also catches a variable proportion of smaller yellowfin tuna. Intermediate sized fish are predominantly caught by the gillnet (GN) fishery operating in the Arabian Sea.

The current fishery definitions are based on gear type, the distribution of the fleets, the tagging data distribution, and some oceanographic patterns (e.g., the split of the Mozambique channel). Adam Langley clarified that clustering of longline CPUE was also used to split the areas. In general, the current fishery structure is appropriate based on gear type, although some revision of the fishery definitions may be required if the spatial boundaries are revised. Nonetheless, there are some persistent issues in the fishery length composition data that require further consideration

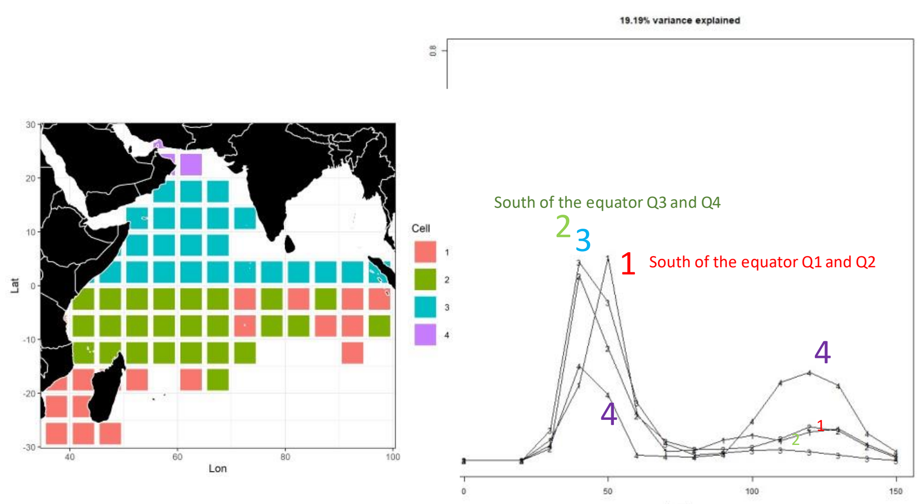
Currently, the PS FS fishery is also partitioned by fish size (length) to accommodate the variable proportion of fish in the two length modes (smaller 40-60 cm and large 100-150 cm fish). The variation in the size of fish caught indicates a higher degree of complexity in the structure of the fishery (possibly seasonal and/or spatial) that is not used in the current fishery definition and cannot be adequately modelled via a standard selectivity function. There is a comprehensive database of length sampling data from the PS fishery. Further evaluation of those data may improve the understanding of the population structure of yellowfin tuna in the western equatorial region and, thereby, improve the model definition of the PS FS fishery and the parametrization of the associated selectivity function(s) (to account for any persistent spatial and/or seasonal patterns in the length composition data). Carolina Minte-Vera presented an example on using the regression tree analysis to analyze the LF data for the fisheries. Preliminary application of the fish frequency tree analysis (Lennert-Cody et al 2010, Xu and Lennert-Cody, 2023) on the length frequencies of the longline and the purse-seine free school (FS) fisheries were done during the review (Figure 4). For the longline fisheries, the tree analysis split the southwest (area 2, east of 45°E) from the other areas as the first split, which explains the largest variability in the length frequencies. The southwest area has the largest fish. The areas east of 40°E had larger fish as well. The rest of the areas were maintained together in the second and third split, which was on quarters (second split separated quarters 1 and 2 from quarters 3 and 4, the third split separated quarters 3 from 4). For the FS data, the tree analyses show splits at the equator for both the French and the Spanish fleet, followed by seasonal splits.

Length frequency samples from individual FS sets should be evaluated to see if they include a mixture of small and large fish or whether they can be separated into schools of small fish and schools of large fish.

Longline length frequencies (10.7% variance explained)



Purse-seine (Spain, 19.19% variance explained)



Purse-seine (France, 20.9% variance explained)

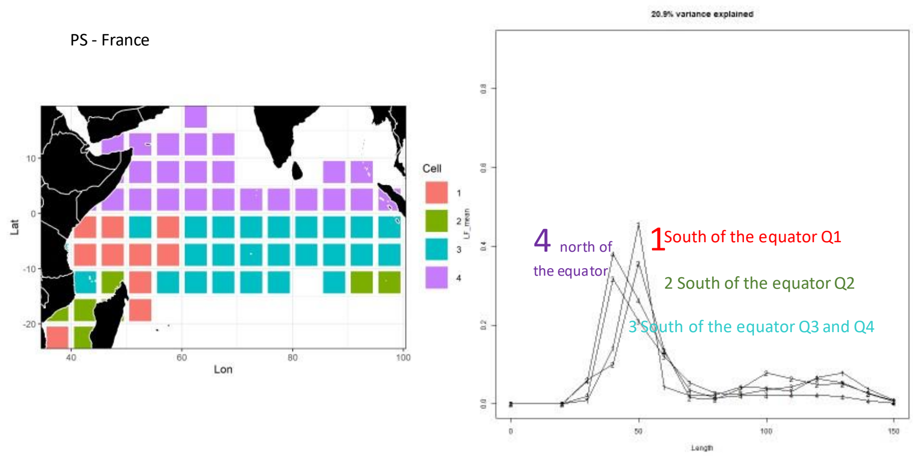


Figure 4. Preliminary application of the tree analysis on length composition for longline fisheries (Japan, China and Chinese Taipei) and purse-seine fisheries on free schools (Spain and France).

Recommendations:

Purse-seine fisheries definitions: Refine seasonal and spatial definitions of purse-seine fisheries following comprehensive review of size composition data, using, for example, regression tree analysis (Lennert-Cody et al 2010).

3. Catches

The catch was dominated by the longline fleet prior to 1980, but since then the purse seine fleet has expanded, and currently approximately half the catch is taken by the artisanal fleet (gillnets and hand line) (Figure 3). Most of the catch comes from the tropical region (regions 1 and 4).

The artisanal fisheries (especially gillnet and handline) have accounted for a substantial proportion of the total IO yellowfin tuna catch (about 50%) over the last decade and there was a marked increase in the catch from the Arabian Sea handline fishery over the last 5 years. There is considerable uncertainty associated with the estimates of the catch from those fisheries and IOTC Secretariat is currently evaluating the accuracy of those catch estimates. Until the magnitude of the catch can be verified, an alternative catch history should be evaluated as a sensitivity test of the current model. The alternative catch history should be formulated by secretariat staff in conjunction with the country experts. There is also a need to improve the length sampling from the handline fishery, particularly given the magnitude of the catch taken by the fishery in recent years.

Recommendations:

1. Catches: Include information about the reliability of the catch from each fleet in the assessment report and in the models. The uncertainty in the catch could be included in the assessment model using the CV for the catch likelihood. If the uncertainty is more related to potential bias, then alternative catch scenarios, formulated based on guidance from country experts, could be created. Alternative scenarios could be used to determine what order of magnitude underestimation in the artisanal fishery catch would be needed to substantially impact the results and management advice.

2. Sampling: Improve sampling coverage of handline fishery (catch and size).

4. Longline CPUE standardization

The main abundance indices included in the assessment model are the region specific longline CPUE indices. Consequently, the assessment results are highly dependent on the derivation of these indices and the associated model assumptions. The current assessment shares the catchability coefficient (q) between the regional longline CPUE indices, having first corrected the CPUE indices by the relative magnitude of biomass between regions (regional scaling). This is a very strong assumption that informs both the relative biomass and movement rates. However, selectivity was estimated separately for each longline fishery (logistic functions). Given the interaction between catchability and selectivity, there is somewhat of a contradiction in sharing catchability among regions but not selectivity. Sharing catchability also requires that the region scaling is done appropriately. The spatial modeling workshop, which used the IOTC YFT assessment as a simulation example, found major bias when the spatial coverage of the longline fishery did not match the spatial distribution of the stock. Therefore, further investigation into the spatial analysis of CPUE data, area weighting, and selectivity is needed. Spatial variation in growth may support separate selectivity among regions, including the possibility of dome shape selectivity, depending on how spatial structured is modeled.

For the 2018 assessment, the longline CPUE indices were derived from operational level (set) catch and effort data. However, operational level data were not available in 2021 and the CPUE analysis was conducted using catch and effort data aggregated by 1x1 degree latitude and longitude cell. The change in data structure resulted in changes in the CPUE indices, especially for the eastern equatorial region (R4); the overall decline in the 2021 R4 CPUE indices was considerably greater than the decline in the 2018 CPUE indices. There were also considerable differences in the CPUE indices from the other regions. The panel recommends further analyses to resolve the differences in the CPUE indices from different levels of data aggregation. There is also concern about whether the joint longline index is really declining so rapidly or whether it is due to inclusion of new fleets. This should also be investigated.

There is also concern that it was not possible to access the preferred (operational) data set for the 2021 assessment, resulting in a lack of consistency in the assessments between 2018 and 2021. There appear to be ongoing issues relating to the sharing of longline operational data, reducing the availability of the data and the transparency of the data analysis. The issues of data confidentiality appear to have been progressed by other tuna RFMOs. For example, SPC has access to operational data from the western Pacific longline fishery, except for data from the last 3 years. The Panel recommends that the Secretariat initiates discussions with distant-water longline nations to improve the provision and access to operational level longline data.

The Panel was unable to evaluate the technical details of the longline CPUE analysis due to complexity of the data processing and modelling approach and a lack of detailed diagnostics (e.g., trends in residuals by cluster) from the 2021 analysis. It was considered that the species-based clustering analysis represented the best approach for dealing with large changes in targeting behavior by the longline fleets; however, there are limitations in this approach. Trends by cluster should be evaluated and consideration should be given to removing clusters that do not target yellowfin. There is some evidence of change in targeting in R3 and this should be further investigated. Given the importance of the longline CPUE indices and the complexity of the analysis, the panel considered that the CPUE analysis required a separate review process, including participants from other tuna RFMOs.

The joint distant-water longline CPUE analysis resolves some of the issues regarding spatial coverage of LL fishery data, particularly the contraction of the operation of the Japanese longline fleet. However, there may be potential interactions between Nation and Area that are not adequately accounted for in the analysis. Further, there remains a complete lack of longline coverage in the Arabian Sea and trends in abundance in this area were assumed to be equivalent to the equatorial region of R1. The Arabian Sea area accounts for a large proportion of the total yellowfin tuna catch, although it appears that there are insufficient data available to develop a separate index of abundance for this area of the fishery. Progress using the joint data is hampered by access issues and these should be solved to facilitate the research on this topic. It may be useful to include average size for each longline fleet in the assessment (not fit) to see how consistent it is with the model. This may help evaluate what size data would be consistent with the decline seen in the CPUE by region.

The stock assessment model currently assumes that longline selectivity in each region is constant over time. This assumption may not be valid if there have been large changes in the species targeting and/or area fished over the history of the fishery. There are also seasonal variations in the LL CPUE indices and length compositions that are not adequately represented by the assessment model. Further modeling is proposed to investigate seasonal dynamics (movement, longline catchability and selectivity) to account for the observed seasonal patterns.

The index is fit using a constant CV of 20%, but the availability of data and other factors suggest that the CV should differ by year. Quarter specific CVs derived from the approach used to standardize the CPUE should be used in the model with an additive constant, possibly estimated, to represent model misspecification and unmodelled process error (e.g. random temporal variability in catchability). One approach to estimate the additive CV is to use an age-structure production model fit to the index of abundance and recruitment deviates estimated, which will provide a minimum additive CV.

There are marked differences in the trends in the longline CPUE indices from R2 (SW) relative to the other model regions. This may indicate that the region supports a discrete population, although the current assessment assumes those dynamics are driven by recruitment in and movement from the adjacent western equatorial region (R1). The Panel recommends further comparative analysis of the CPUE and length composition data to investigate stock structure in this region.

There is potential to improve the understanding and application of longline CPUE data through the development of spatial temporal modelling approaches (e.g. using VAST) at the regional and oceanic scale. Such an analysis may be informative regarding the definition of appropriate regional and stock boundaries and may adequately account for spatial/temporal gaps in the coverage of the longline catch and effort data. The CPUE should be spatially weighted. It is also important to spatially weight the length composition associated with the index by the CPUE to ensure that it represents the population rather than the catch. This can also be done within VAST to help fill in spatio-temporal cells with no or little composition data. If done for the whole stock (in a single stock model or for each sub-stock in a multi-stock model), it may allow the use of asymptotic time invariant selectivity.

The Panel recommends investigating the potential for the development of CPUE indices from the LF4 fishery which dominates the catch from the eastern equatorial (R4) region. This may provide an alternative set of CPUE indices for the region and reconcile the very low R4 LL CPUE indices from the distant-water fleet since the late 2000s. However, it is likely that insufficient reliable catch and effort data are available from the fishery.

Recommendations:

1. Standardization analyses: Further investigate the spatial analysis of CPUE data, area weighting, selectivity by fleet, temporal changes in selectivity, and the development of spatial temporal modelling approaches (e.g. using VAST) at the regional and oceanic scale.

2. Precision of estimates: Estimate a CV for each time step of the index and use it as indication of the relative precision of each point estimate, which can be added to the overall precision assumed for the index.

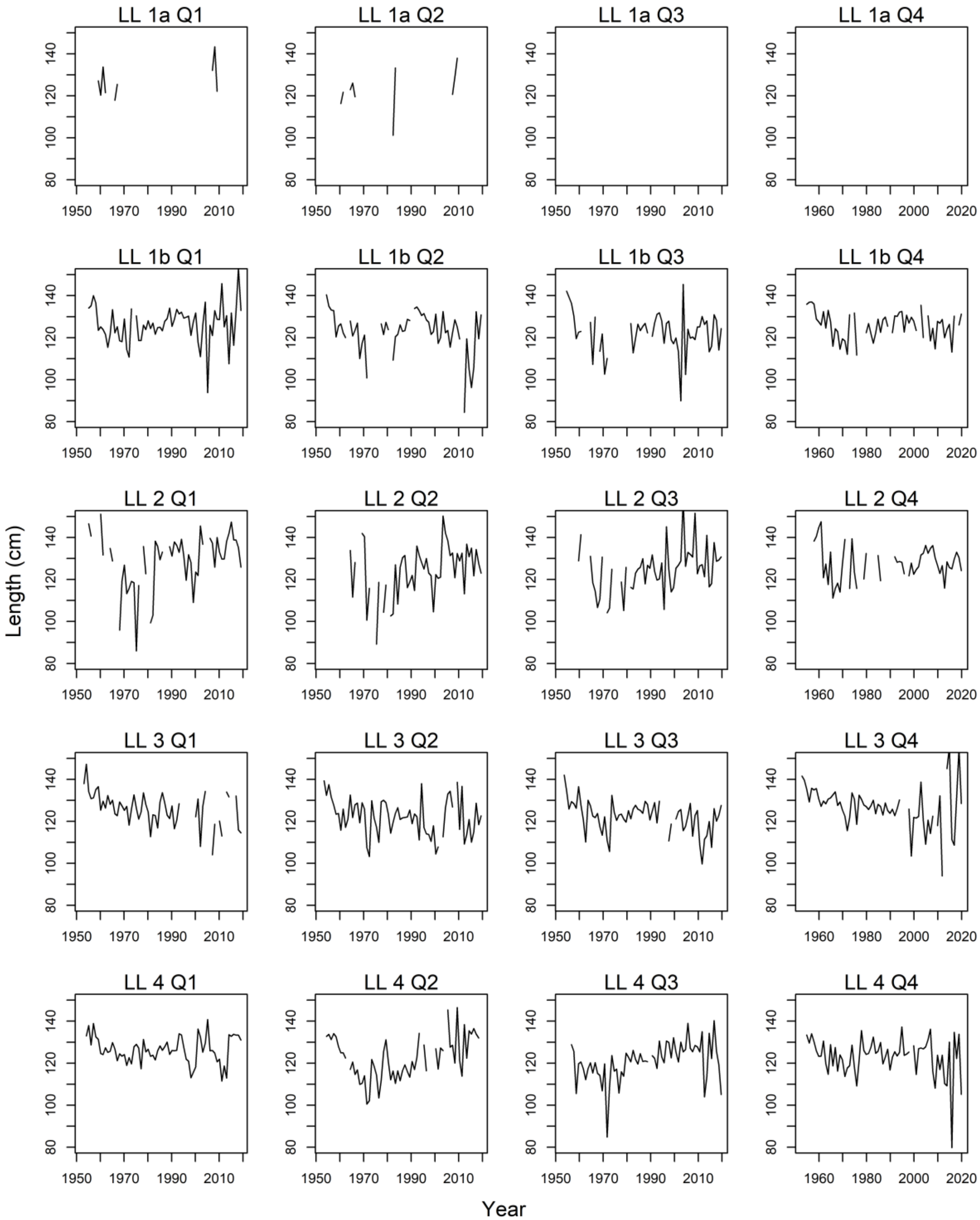
3. Length composition associated with the index: Spatially weight the length composition associated with the index by the estimated density (CPUE) to ensure that it represents the population rather than the catch, estimate a separate selectivity for the standardized index (e.g., assume it is a "survey" or a fleet with no catches)

4. Regional differences: investigating the potential for the development of CPUE indices from the LF4 fishery which dominates the catch from the eastern equatorial (R4) region.

5. Review process: Evaluate the CPUE analysis in a separate review process, including participants from other tuna RFMOs.

Request:

Plot the CPUE using lines by area and quarter in separate panels, plot the average size by using lines by area and quarter in separate panels (Figure 5).



ass

Figure 5. Average length for longline fisheries by area and quarter for yellowfin in the Indian Ocean.

5. Purse seine CPUE

The purse seine (PS) free swimming schools (FS) CPUE indices from 1991-2020 were included in the 2021 stock assessment as a model sensitivity (and included in the final uncertainty grid). It is recognized that the efficiency of the PS fleet has increased considerably over the history of the fishery. To correct for the increase in PS efficiency, the trends in the CPUE indices were effectively rescaled relative to the trend in the R1 LL CPUE indices from the corresponding period. Consequently, the corrected PS CPUE indices have the same general trend as the LL CPUE indices and, hence, the inclusion of the indices in the assessment model do not result in an appreciably different trend in stock abundance (relative to the base model). This may be a useful approach to estimate changes in efficiency of this fleet and could provide important information for management decisions, but is not likely to improve the assessment results.

There have been many other changes in the operation of the PS fishery that are unlikely to be accounted for in the CPUE standardization, in particular the changes in fleet behavior related to the expansion of the FAD fishery and the more recent introduction of catch limits for yellowfin. These factors are likely to have resulted in a decline in the effective effort that is directed towards the PS FS fishery. The Panel does not consider that changes in the relative performance of the PS fishery can be reliably determined and does not recommend the inclusion of PS CPUE indices as a primary abundance index in the assessment model.

The panel encourages continued development of methodological approaches to utilize the data from acoustic buoys to develop indices of abundance for YFT available to FADs.

Recommendations:

1. Modelling: Do not include the PS CPUE indices as a primary abundance index in the assessment model

2. Future research: Continue the development of methodological approaches to utilize the data from acoustic buoys to develop indices of abundance for YFT available to FADs

6. Length frequencies and Selectivity

The model uses multiple fleets, each with associated length composition data. Mostly these are separate physical fleets, but in some cases multiple model fleets are used to split different fishing practices of a given fleet (e.g. Free School (FS) or Log Sets (LS) in the purse seine). The purse seine (especially the FS) catches small and large fish, but very few fish in the 60-80cm range. The model structure accounts for this by splitting the fleet into two fleets in the model that represent the small and large components.

The model uses age-based selectivity, except for purse seine which has been converted to length-based in the most recent version of the model. Age-based selectivity runs faster in the model, and the assumption is that because of the fine (quarterly) time step in the model this is likely to closely approximate the actual length-based selectivity. This also allows for direct comparison with the age-based tagging data. However if there is length selectivity which is not modelled well by age then this may be suboptimal.

The quality of the data underlying the length distributions varies considerably between the fleets. The longline fleets are constrained to have logistic selectivity, while the FS purse seine for the large fish is constrained to have high selectivity for the largest fish, the others have a double normal asymmetric dome. The selectivity is considered constant over time. It should be stressed that these assumptions can have large impacts on the modelled population. This is especially true for the longline fleets which are also providing the CPUE signal to the model. The selectivity is assumed to be constant over time for all fleets, which seems unlikely given the long time span, the multispecies nature of the fishery, the changing national composition of the fishery, and the large spatial structure. It is unclear if there is enough data to relax this assumption, but it would be worth considering for the largest fisheries. It is also unclear if the PS FS fishery should also be considered to have full selectivity for the oldest fish – and this should be carefully compared with right hand limb of longline length composition data. These issues can be explored using the empirical selectivity diagnostic.

One potential data quality issue identified is that the mix of countries fishing using longlines has changed significantly in recent decades, with the fishery moving from being dominated by the Japanese to having the Japanese as a minor component of a multinational fishery. This could potentially impact on the quality of both the length distribution and the CPUE for this fleet.

The time trend in the length composition data for both the PS and LL fleets is concerning and could be a consequence of a population process (depletion or changes in growth) rather than changes in selectivity.

In one fleet, modal progression in the length distribution has been identified and used to sanity check the growth curves. This could also be investigated for the other fleets to identify if modal progression can be identified between different fleets with different sizes of capture. However, it should be noted that the gillnet fisheries showed modal progression in the early periods, but not in the latest, and it is not clear if this was caused because the fishery expanded to other areas and data is being mixed from multiple sites or some other factor.

The Review Panel evaluated the time-averaged fit to the data for all fleets, and the annual fit for several (but not all) fleets. When comparing the model to the time-averaged length distribution data, the modelled catch distributions show a good agreement with the data. However, when comparing the model fit to data on an annual basis, the match is much poorer. There is a great deal of structure in the data which is not tracked by the model. Some of this is likely sampling noise which the model cannot (and should not) track, but there are indications of structure which may be reflecting biological reality that is not captured by the model. There are multimode patterns in some years which may be worth investigating. More seriously, there are time trends in the mis-match for some fleets (i.e. the observations differ from the predictions in the same way for each fleet), with both the long line and the large fish section of the FS purse seine fleets showing similar trends.

The length composition data should be spatial raise to the catches for the fisheries and spatially weighted by the CPUE for the indices. The fisheries and the indices should have different selectivity curves, with the fisheries more likely to have temporal variation in selectivity.

Conclusions:

The overall fleet structure and choice of selectivity functions seems reasonable, although there is likely a sensitivity in the model results to the selectivity choices, especially for the longline. The length data is sufficient to estimate the time-averaged fleet selectivity adequately for the fleets. However, the model is not able to replicate the details of the length distributions on an annual basis and this should be investigated in more detail.

Recommendations:

1. Auxiliary analyses: Investigate spatio-temporal trends in the length compositions, including looking at quarter variation, using the regression tree analysis.

2. Coverage of LF data: Provide a map for the longline data to evaluate where the data was taken compared to where the catch data comes from.

3. Data weighting: Adjust the effective sample size for the length distributions as described in the data weighting section, to allow better weighting between fleets. Ensure this is done in a way as to avoid unintended consequences in the relative weighting between length data and CPUE/tagging data.

4. Data treatment: Consider time-averaging the length distribution data for the fleets with the poorest data.

5.Time-varying selectivity: Investigate the potential for and impacts of relaxing the time-invariant selectivity assumption of the largest fleets (especially the longline fishery). This could be based on time blocks or by analyzing the spatio-temporal structure of the fishery

6. Selectivity assumptions: Consider moving to length-based selectivity for some of the other fleets, especially those catching smaller fish.

7. Model fits: Investigate all of the fleets for the mismatch between model and data. Look especially for signs of time trends in the mismatch and for mismatches which replicate across multiple fleets (and are therefore more likely to reflect an underlying reality, whereas a time trend in mismatch affecting a single fleet could indicate shifts in selectivity over time).

8. Growth estimation: If modal progression can be identified in any of the length frequencies, then this could be used to sanity check the assumed growth equations. This is already done to some extent but could be investigated further.

8. Presentation of results: Produce a table for each fishery for evaluation of the fisheries and for the assessment report including the following: name, amount of catch (average catch), amount of LF (sample sizes, number of years), reliability of the LF, selectivity type, whether or not the selectivity should be stable or change over time. This will help understand if the fishery composition data should be fit well or not and how the selectivity should be modelled.

9. Sampling: Identify fleets which make up a significant portion of the catch, but have poor sampling, and focus efforts to improve sampling on these. There is not much length frequency data for the FLL fishery in area 4 and more sampling for length frequency is needed.

10. Length composition estimation: The length composition data should be spatial raise to the catches for the fisheries and spatially weighted by the CPUE for the indices. The fisheries and the indices should have different selectivity curves, with the fisheries more likely to have temporal variation in selectivity.

7. Tagging data

Including the tagging data in the assessment model is problematic because the practicalities of tagging limit the spatial distribution of tags and therefore the tags are not initially fully mixed with the population and it is not clear after how many quarters, if at all, the tags become fully mixed. Using a mixing period of several quarters also greatly reduces the information content of the tag data. In addition, the model does not fit the tagging data well.

It is recommended that the initial exploratory analyses be conducted without the tagging data. Fine scale spatio-temporal tagging analysis should be conducted outside the stock assessment model as an exploratory analysis and as a possible way to maximize information while dealing with non-mixing (see Mildenberger et al. 2022). The results from this analysis (i.e., estimates of biomass, fishing mortality, and/or natural mortality) could then be used in the stock assessment.

Tagging related mortality, tag loss, and reporting rates all need to be estimated to ensure the tagging information is unbiased. Preferably, these should be estimated from data collected during the tagging program (e.g., double tagging, tag seeding, pen experiments). The cause of the higher return rate of larger fish should be determined (e.g., evaluate the possibility of higher tagging mortality for younger fish, comparing the time at liberty and size frequencies).

Recommendations:

1. Modelling: Conduct initial exploratory fit of the assessment models without tagging data

2. Treatment of the tagging data: Estimate tagging related mortality, tag loss, and reporting rates preferably from data collected during the tagging program (e.g., double tagging, tag seeding, pen experiments). Determine the cause of the higher return rate of larger fish (e.g., evaluate the possibility of higher tagging mortality for younger fish, comparing the time at liberty and size frequencies).

2.Future research: Conduct independent fine-scale spatio-temporal analysis of tagging data (see Mildenberger et al. 2022)

8. Growth

The 2021 stock assessment for yellowfin tuna in the Indian Ocean assumed a Von Bertalanffly growth curve in the base model with fixed age-specific k (deviates from the base k of 0.455 quarter^{-1} for age 2-13 quarters), to approximate the three stanza-growth and the mean length at age in Fonteneau (2008). An alternative 3 stanza-growth model (model 2 in Dortel et al. 2015) was used in the model ensemble. The Dortel et al. (2015) models use different combinations of the tagging, otolith daily ageing data (Sardenne et al. 2015) and length-frequency data from the purse-seine fisheries. The growth pattern assumed was linear and had slow growth up to 2 years old (between 30 and 60 cm), then accelerated growth to about 5 years old (120 cm) followed by slow down and eventually no growth after 5 years old. The length composition data shows a lack of fish from about 55 to 75 cm, which may be due to growth or availability. The empirical estimate of Fonteneau (2008) used in the assessment model has a Linf parameter value of 145 cm. This value is based on studies of the early life history growth of yellowfin tuna, i.e., of difference in size for immature versus maturing fish. The estimate of Dortel et al. (2015) is higher (156 cm), but the study is still based on samples from purse seine fishery which catches mostly small fish. These Linf estimates are low when compared to estimates from other oceans or estimates from many regional studies in the Indian Ocean.

Sardenne et al. (2015) did validation of daily aging using oxytetracycline (OTC) on about two thousand tagged and released fish of which 215 were recovered (Table 1). Team 2 seems to have an almost 1:1 correspondence of daily rings with time at liberty in days (Figure 6), which may be do positive bias in the low ages and negative bias in the older ages (Jessica Farley personal communication)

Table 1. The number and size ranges (in brackets; cm) of tunas tagged by conventional dart tags (DART) and oxytetracycline (OTC) tags and recovered through the Indian Ocean Regional Tuna Tagging Project, and the associated percent recovery rate (RR). Numbers are only for reliable data for species identifications and length measurements at tagging and recapture. From Sardenne et al (2015)

	Tagged	Recovered	Recovery rate (%)

Species	DART	OTC	DART	OTC	DART	OTC
YFT	51841 (32-144)	1993 (34-141)	7899 (35-168)	215 (44-135)	15.24	10.79

Two teams read the YFT otoliths with OTC marks, one estimated a deposition rate of 1 and the other of 0.9 (Figure 6) (Sardenne et al 2015).

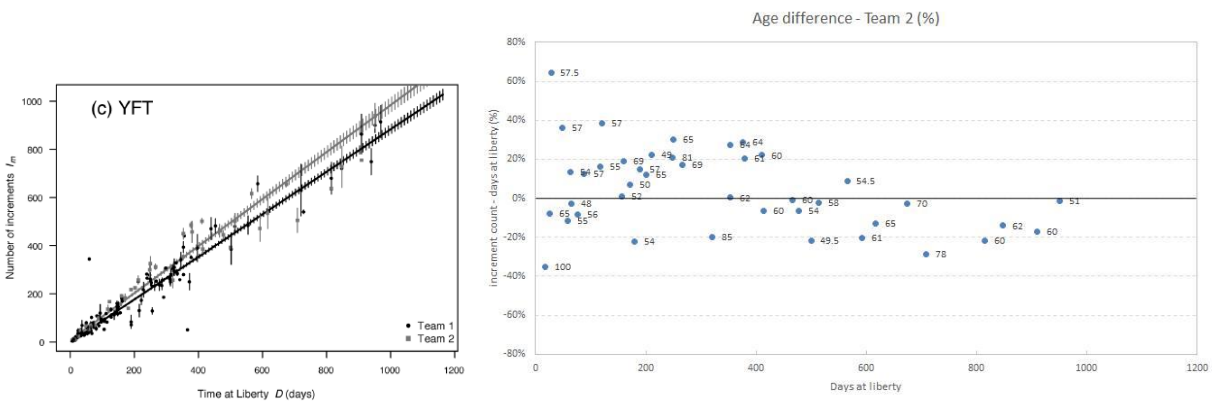


Figure 6. Right: Relationship between micro-increments counts made from the OTC mark (*Im*) to the otolith edge and time-at-liberty (D) for yellowfin tuna in the Indian Ocean (n=112). The small vertical lines around the individual symbols indicate the standard deviation for *Im* and the solid lines indicate the linear regression model fitted to the data for each team of readers, with 95% confidence interval. The estimated regression were: Team 1 (n=90) *Im*= -0.14+0.88*D*; Team 2(n=39) *Im*=7.39+0.98**D* (Sardenne et al.,2015).Left: increment counts minor days at liberty versus days at liberty (Jessica Farley personal communication)

In 2020, the collection of new biological samples was started and re-analysis of previous samples was done within the context of the European Union GERUNDIO project. The new estimates of age and growth for yellowfin use both daily and annual growth zones in otoliths (Farley et al., 2021). Only 3 fish with OTC marks were included in the study in addition to a few otoliths from fish with long time at liberty from the Sardenne et al (2015) study were also read by Farley et al (2021). The number of daily rings counted by Farley et al (2021) were systematically lower than Sardenne et al (2015) (Figure 7). Because the Sardenne et al (2015) study did validation of the daily rings and showed to be about 1:1 correspondence of the rings and the days at liberty of the tagged fish, one would expect that the readings from Farley et al (2021) should also correspond 1:1 to the Sardenne et al study. The systematic difference indicates that the Farley et al (2021) readings are biased low and the daily age may be underestimated. However, the comparison of counts for daily rings was done for fish with more then a year at liberty, for which the daily aging maybe biased, while for daily rings of OTC marked fish with less then a year of liberty the daily rings counts are expected to be more accurate (Jessica Farley personal communication).

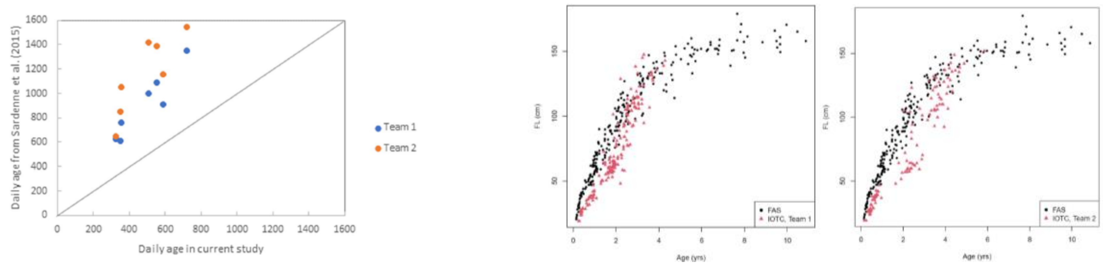
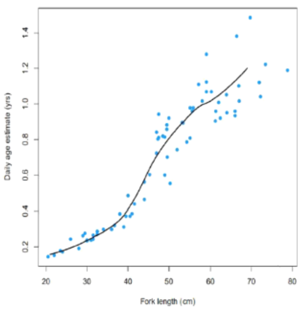
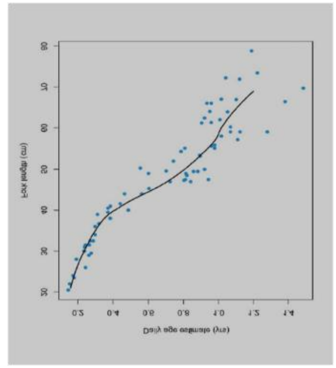


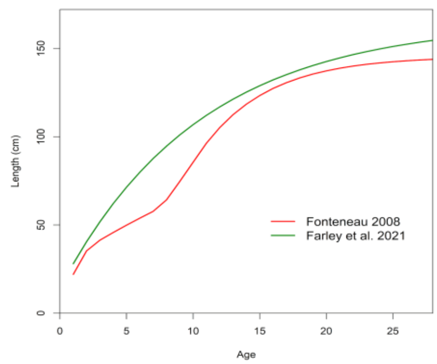
Figure 7. Left: comparison of putative daily ring counts for the same yellowfin tuna by FAS lab (Farley et al 2021) and by Team 1 and Team 2 in Sardenne et al. (2015). Right: Relationship between counts of daily rings and length from Farley et al (2021) and from Sardene et al (2015) Team 1 and Team 2. Source: Farley et al (2021)

Farley et al (2021) also use annuli and assume annual deposition rates, with limited validation (few fish and not done blind of the age). The annual aging method was corroborated based on bomb radiocarbon study (Jessica Farley personal communication). Also, the Tagging data was re-analyzed by Eveson and Farley (2021), using the relationship between fish length and the daily age estimates for the small sizes obtained in Farley et al (2021) to estimate age at release for fish in the tag-recapture data from their release lengths. The resulting age estimates are very different than those obtained from the random effects models in Eveson et al. (2015). Possible reasons for the difference discussed by the authors were that tagging slowed down the growth of the small fish, which after a period caught up with the growth of the larger fish tagged. However, another possible explanation is that the daily aging in Farley et al (2021) is bias low, to about half the rings. If the daily ring count were doubled, the length-age relationship could be easily overlayed in the 3-stanza growth functions as show below. Thus it is paramount that the daily increment be validated.





(**quarters**)



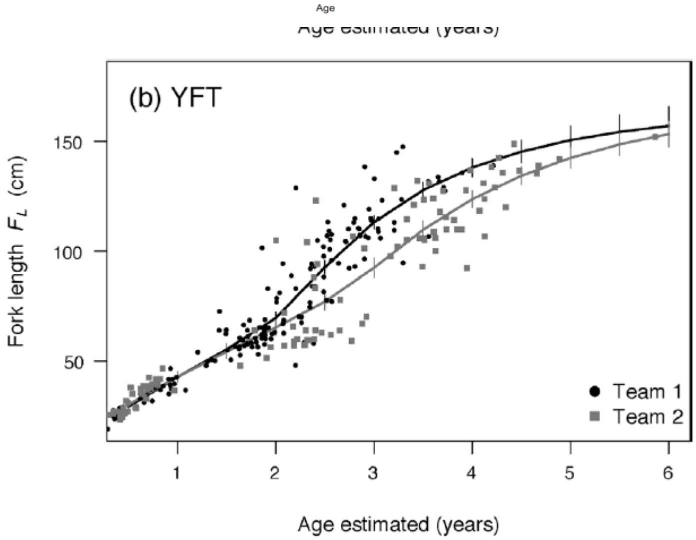


Figure 8. Top left: daily growth increment counts versus length for Farley et al (2021). Top center: the same plot flipped to be length versus daily growth increment counts. Top left pannels: growth curves from Fontaneau et al (2008) Bottom: growth estimates from daily otolith counts from Sardene et al (2015) Team 1 and Team 2.

Both the daily and the annual growth ringss have limited validation but there is corroboration the bomb radiocarbon study and tagged fish. It seems that there may be an opportunity to prepare the sister otoliths from the Sardenne et al (2015) fish marked with OTC to validate both the daily rings (fish at liberty for less than a year) and, at least for a few ages, the annual rings (fish at liberty for two years or more). It is strongly recommended that such studies be done. Some model runs were done during the review with the growth curve proposed by Farley et al (2021). The length composition data for fishery F6 (PS FS area 1b, small fish, ≤80 cm) was not fit well, on average the model predicts fewer fish, so the growth may be too fast. However, the growth of those fish seem also too slow with the Fontaneau (2008) growth assumption, which also problems fitting the large longline length frequency in area 4. Further exploration of growth in he assessment model should also be done.

Another aspect that was discussed during the workshop was the initial length (Lmin). Some length composition data showed smaller sizes that were not fit well by the model. The length frequency in Fishery F6 seem to be too variable and the growth too slow with the Fontaneau (2008) growth assumption. The birth date may also be off. Some runs estimating Lmin were done and they improved the fits to the length composition data (despite the estimates going to the lower bound). The age at Lmin is set to be 1 quarter.

Growth is currently estimated outside the assessment model using tagging data, otoliths data and length composition data.. The different age data sets (Sardenne et al 2015 Team 1, Team 2, Farley et al 2021 daily age and annual age) should be added to the data file (as conditional age at length), so that the estimation of growth within the assessment models can be explored. Ideally, the tagging data should also be integrated into the assessment model, this would require a modification of Stock Synthesis to include growth increment data.

The recent age sampling program is based on the collection of otoliths from the PS fishery. There is the potential to use the length-age observations to derive an age-length key specifically for the PS fishery and derive an estimate of catch-at-age for PS fishery from the relevant period. Initially, this would require a collation of the length-age observations to evaluate the intensity of sampling (by fishery and year/quarter).

The new age-and-growth study gives a strong indication that growth for males and females is different. To explore the implications of this hypothesis, a two-sex assessment model needs to be developed. This could be done initially in a simple one area model.

The variation of length at age assumed in all models of the ensemble was CV=0.1. The variation of length at age includes not only individual variation in growth, but also growth within the model time step (quarter), variation in date of birth,. For these reasons, the estimate of variability of size at age cannot simply be done externally by fitting to the age and length data from the growth study, but should be addressed within the assessment model. Generally, the coefficient of variation of size at age for younger ages maybe lower than at older ages. Values for CV for young and adult ages could be tuned to the observed variability in the length frequency data when one cohort is distinguishable, or to follow the right-hand size distribution of the length frequency of the large sizes.

Recommendations:

1. <u>Conceptual model</u>: Evaluate the data given the conceptual model and consider removing data that maybe from different stocks.
2. <u>Ageing data</u>: conduct validation studies of the daily and annual rings (e.g. by preparing the sister otoliths marked with OTC in Sardenne et al, 2015), iInclude age conditional at length data in the assessment model consistent with the conceptual model (i.e. remove data that maybe from different stock).
3. <u>Evaluate estimating growth in the assessment model:</u> by fitting to the age-conditional at length data

4. Estimate L1: Consider estimating L1 and explore different birth dates (that is different recruitment seasons)
5. Growth by sex: explore the implication of growth differences by sex
6. Growth curves: Inspect the fits to the age conditional at length data and try other growth curves if necessary (e.g. growth cessation model)
7. Variability of size at age: explore different assumptions for CV of length at age
8. Future research: Modify SS3 to include growth information from tag growth increment data.
9. Validation: Do more (OTC) validation work
10. Data collection: collect a wider spatial and temporal range of otolith data
11. Tagging data: Modify SS3 to use tagging growth increment data

9. Natural mortality

The natural mortality (M) in the 2021 assessment model was assumed to vary by age and to follow a shape similar to the WCPFC yellowfin stock assessment, which is based on estimation of M at length from tagging data. The M was high for juveniles, a decreasing trend towards a base level, followed by an increase and decrease back to the base M for adults. This shape accommodates the difference in natural mortality between males and females in a one-sex model. The base assumption for adult M was the level equal to the 2012 MFCL model (which included tagging data). An alternative scenario included in the uncertainty grid was *Mlow*, which assumed the adult M equal to 0.35year^{-1} , which is equal to the assumptions for M in the ICCAT yellowfin tuna stock assessment and it is based on a longevity of 18 years (Figure 9).

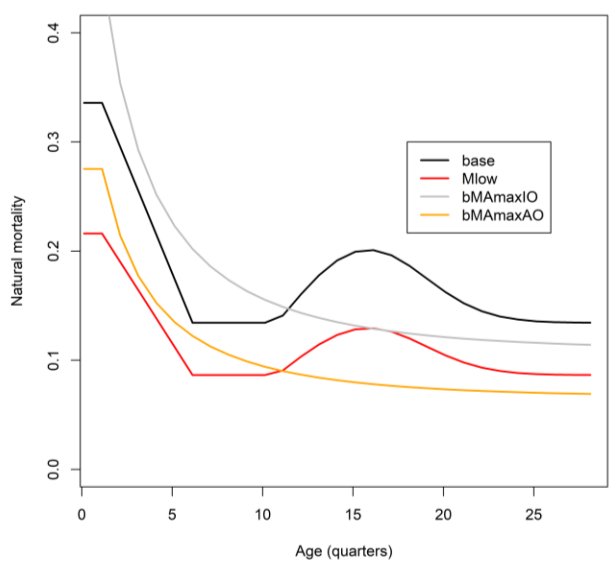


Figure 9. M at age in the base model and alternative scenarios in the 2021 yellowfin tuna assessment for the Indian Ocean.

Two other scenarios based on Hoyle (2021) were also included in the uncertainty grid. Those curves assume that M is high for younger fish and declines with age following the inverse of weight (Lorenzen 1996). Two alternative scenarios for adult M were considered, one based on the maximum age of 10.5 years observed for Indian Ocean yellowfin tuna (Shih et al. 2014) (bMAmaxIO), and the other based on the maximum age of 18 years recorded for the Atlantic Ocean (Pacicco et al. 2021 (bMAmaxAO). Hoyle et al. (2021) do not include a 'hump' to allow for higher female natural mortality because the change in sex ratio at length observed for yellowfin tuna was assumed to be explainable by differences in growth rather than M by sex. The results of the 2021 assessment model for the *bMAmaxAO* and the *Mlow* scenario were about the same.

Recommendations:

1. Parametrization of M: Assume a vector or natural mortality at age that considers higher natural mortality for juvenile fish that declines to smaller values in adult stages (e.g., Hoyle et al., 2022)
2. Adult M: use the longevity estimated for the Indian Ocean as base value and explore others as alternatives (e.g. 0.7*M_adult has the most support from the tagging data)
3. Data in the model: Include conditional age at length data in the assessment model even if not fitting to it, include sex ratio data or length composition data by sex (in a sex structured model).
4. Estimation of adult M: once conditional age at length and sex ratio data is added to the model, attempt to estimate adult M (and growth).
5. Differences by sex: Implement a two-sex model. Consider different longevities for males and females consistent with the latest aging studies (males seem to live longer) and with consistent with different growth between males and females (observed in several tuna species). Identify what would be the "expected bump" in a one-sex model if differences in adult M for males and females exist (but preferably avoid using a one sex model).
6. Sampling: include sex identification in the samples for length composition, especially in the GERUNDIO project, so that proportion of sexes at length could be used to estimate differences in natural mortality (note that there are now new genetic tests available to determine sex of tuna)

12. Maturity

Spawning occurs mainly in the equatorial area. The sizes exploited by the fisheries range from 30 to 180 cm, the length at 50% maturity is assumed in the model to be around 75 cm. According to the available maturity data, this length should be the length when fish start spawning. There are two maturity curves available: one is the onset of maturity; another is when the adults are ready to spawn. The first one is used in the model. The values are based on data taken year-round in 2013. Preliminary results from the GERUNDIO project (from year 2019) indicate that the length at 50% maturity maybe at smaller sizes. However, the samples may not represent the core of the spawning biomass as they come mostly from NW Indian Ocean and are taken from processing plant in Galicia. The model is based on a quarterly time step, the fish need to be ready to spawn in a particular quarter, thus the maturity ogive needs to provide the proportion of fish ready to spawn within a model time step. The current maturity ogive is a function of age and is obtained by transforming length into age using the growth function.

Recommendations:

1. Free schools: The free school length composition shows a persistent bimodal pattern, of small fish and large fish. The fish of medium size are not present. The presence of large fish schooling maybe related to spawning behavior. It is recommended to investigate the relationship between the occurrence of large fish in the free schools and the maturity to evaluate this hypothesis.
2. Maturity ogive: Use the maturity ogive by length rather than age, so changes in growth estimates will not require changes in the maturity ogive.
3. Develop a two-sex model: in a two-sex model the maturity for males and females can be input separately.
4. Future research: Determine batch fecundity and spawning frequency to produce a spawning output curve at length.

5. GERUNDIO project recommendations: The panel_recommends the collection and analysis of gonads to be continued, and that the coverage includes all regions of the Indian Ocean, but particularly from the northern and eastern areas (from all size classes and months) to improve the estimation of reproductive parameters. Fish >60 cm fork length (~minimum size at maturity) are particularly important to increase the sample size available for maturity, fecundity, and spawning fraction analyses. Monthly sampling is important in reproductive studies to obtain reproductive data throughout the year and to pinpoint the spawning period. If a clear spawning period is found, sampling within the spawning period should be more intense. The panel also recommends collecting additional gonad samples from different fishing gears (e.g., longline) to improve the size coverage and have better representation of the population spatial range.

13. Spawning and Recruitment

The current assessment estimates quarterly recruitments for the two equatorial regions from 1972 to 2019 (192+168 = 360 parameters). The model is over-parameterised given the limited data to inform the model regarding the magnitude of individual recruitments in combination with other model parameters (e.g., movement). Hence, the model has considerable freedom to fit the trends in the LL CPUE abundance indices via the recruitment parameters, particularly given the relatively low weighting attributed to the LF data sets. Limited analysis of the length composition data has been conducted to evaluate the consistency between the modal structure of the LF data and the model estimates of recruitment. The PS FS and GN fisheries provide the most comprehensive sets of length composition data for evaluating trends in recruitment within the western equatorial region (R1). Insufficient LF data are available to conduct a comparable analysis for the other regions, particularly for the eastern equatorial region (R4).

The review panel conducted a cursory examination of the length composition data from the PS FS (large fish) fishery, following an up-weighting of those data in an exploratory model run. There was some evidence of modal progression in the length composition data, although the data were not entirely consistent between consecutive year-quarters indicating a degree of sampling variability. Restructuring of the length composition data (following exploratory analyses) may reduce the variability in those data. Further, the fits to the PS FS (large) length composition data may be improved following refinements in the parametrization of the selectivity (see Section 6. Length frequencies and Selectivity) and growth (Section 8. Growth).

There are appreciable differences in the estimated recruitment trends from R1 and R4 over the last 20+ years with an increasing proportion of the recruitment attributed to R1 and a corresponding decline in recruitment in R4. These trends in the model are driven by an increasing level of catch from R1 and a greater decline in longline CPUE indices from R4. The estimated trends in recruitment are also influenced by the movement dynamics between regions. The current model estimates that there is negligible movement between the two equatorial regions (R1 and R4), i.e., the fish in the two regions are essentially being modelled as discrete populations. This may be due, in part, to the lack of data regarding movement between the two regions, the assumption of temporally invariant movement and the relatively low penalty associated with estimating separate sets of recruitment deviates for the two regions. Alternative stock hypotheses should be investigated to evaluate different assumptions regarding recruitment and movement. For example, using seasonal and temporal variation in movements. Another alternative is to estimate recruitment in all areas. If this model does not converge, try alternatives (e.g., two areas and using areas as fleets for the peripherical areas, etc.). and/or a single recruitment series or separate recruitments estimated for each of the 4 regions). It is also speculated that the declining trend in recruitment in R4 could be attributable to higher levels of fishing mortality of juvenile yellowfin within R1 and, hence, a reduction in the eastward migration of yellowfin tuna into the central equatorial region of the Indian Ocean.

Stock Synthesis applies a bias correction[1] to the recruitment deviates so that the time series of estimated recruitments are mean unbiased. The value assumed for the bias correction factor has been shown to be influential in the determination of the estimation of *R0* and, hence, *SB0* and associated reference points (e.g. B_{MSY}). It also may influence the value of recruitment used in projections. In the 2021 stock assessment (WPTT 23), the default bias correction adjustment factor (the scaler for $-\sigma_R^2 / 2$) of 1.0 was assumed for the models (basic and revised). Following the assessment, the bias correction factor was revisited to ensure projected recruitments were equivalent to the long-term average recruitment with a derived bias correction factor of approximately 0.4. The magnitude of the difference between the default and derived values of the bias correction factor was larger than expected by the review panel. Typically, the estimated values of the bias correction factor are about 0.8. Is the lower bias correction due to the relatively low variation in the recruitment deviates compared to the assumed σ_R value of 0.6? The review panel recommended a more thorough evaluation of the derivation of the bias correction factor.

The recruitment value used for reference points and projections should be compared to the average recruitment over the period for which historic recruitment is thought to represent that which should be used in reference points and projections. This can be used to evaluate if the bias correction has been applied correctly. The comparison is more complicated in a presence of a stock-recruitment relationship and should be based on the average of the bias-adjusted residual.

There may not be information on the recruitments for the early years because the model starts before the index and composition data are available. It is tempting to set the recruitment deviates for these years to zero (in combination with removing the log-normal bias-correction factor) and making the recruitment equal to that expected from the stock-recruitment relationship (or the average). However, this will ignore any uncertainty in recruitment and prevent the model from estimating any long-term trends in recruitment. On the other hand, if the recruitment deviates are estimated, they can compensate for a model misspecification. This approach is also associated with the selection of the start time of the model and the method used to create the initial age-structure. The best approach has yet to be determined generically, and requires further research.

The trend in the recruitment deviations was used as diagnostic for the models. Many tropical tuna stock assessments show trends in recruitment occurring when the PS LS fisheries increased. This is probably a consequence of the model trying to account for the increased catch of juveniles in the absence of a decline in the (longline) indices of abundance. It is still not known if this is due to the indices being hyper-stable,

---

[1] When using the penalized likelihood approach, a log-normal bias-correction factor is needed to ensure the deterministic equation represents the expected value (mean) of recruitment. However, when information is limited for a particular year, the full bias correction will bias the estimates. In the extreme case of no information, the bias will be equal to the bias-correction factor and the bias-correction factor should not be used. Lack of information can occur in early years due to the lack of composition data and in recent years because some cohorts are included in the composition data for only a few years. Since there is no data in the projections, the bias-correction factor should not be used for future recruitments. For years with partial information only a partial bias correction should be used and this is described by the bias-correction ramp in Stock Synthesis (Methot and Taylor, 2011) and should be determined for both the left hand (early years) and right hand (recent years) sides. Approaches that treat recruitment as a random variable (e.g., random effects, state-space, Bayesian) do not require the bias correction ramp. However, these approaches require integration and are often not practical for complex stock assessments or are not available in the software used.

a density dependent increase in productivity, or some other factor. Care is also needed when interpreting trends in recruitment because trends in the environment can also cause real trends in recruitment. Other factors, in addition to trends in recruitment, should be used to interpret whether they are real or indicate some form of model misspecification.

The steepness of the stock-recruitment model is always a main uncertainty in tropical tuna assessments and stock assessments in general. The steepness often does not influence the assessment results themselves, although it sometimes can, but can have a substantial influence on reference points. In general, there is little, if any, evidence of reduced recruitment in tropical tunas if the biomass is above 20% of unexploited, therefore when the stock does not go to low levels, using a steepness of 1 in the assessment may not be inappropriate. However, consideration of the steepness used to generate reference points, or the proxies used, is important. The IO YFT assessment uses a steepness 0.8, and uses values of 0.8, 0.7, 0.9 in the uncertainty grid. Given that there is essentially no information on the appropriate level of steepness, the review panel cannot provide any solid recommendations.

Recommendations:

1. <u>Length composition data</u>: restructure the fisheries using exploratory analysis to reduce noise and increase the information about recruitment in the data. Refine the parametrization of selectivity and growth to improve the fit.

2. <u>Stock structure:</u> Investigate alternative stock structure hypotheses to evaluate different assumptions regarding recruitment and movement.

3. <u>Bias correction</u>: Perform a thorough evaluation of the derivation of the bias correction factor.

4. <u>Projections</u>: Compared the recruitment value used for reference points and projections to the average recruitment over the period for which historic recruitment is thought to represent that which should be used in reference points and projections. This can be used to evaluate if the bias correction has been applied correctly.

5. <u>Early recruitment</u>: conduct research to determining an adequate method for the selection of the start time of the model and the method used to create the initial age-structure

6. <u>Trends in recruitment as diagnostic</u>: Other factors, in addition to trends in recruitment, should be used to interpret whether they are real or indicate some form of model misspecification.

7. <u>Steepness</u>: Given that there is essentially no information on the appropriate level of steepness, the review panel cannot provide any solid recommendations. The panel notes that when the stock does not go to low levels, using a steepness of 1 in the assessment may be reasonable.

8. <u>Recruitment patterns</u>: Evaluate recruitment patterns in R1 (NW) from the modal structure in LF data from PS LS, PS FS and GN fisheries. Conduct analysis using the single NW region model to evaluate coherence with LL CPUE trend from R1. Evaluate the declining trend in recruitment estimated for R4. Is there any information in the LF comp data or is the trend simply driven by the fit to regional CPUE indices?

9. <u>Regional recruitment distributions</u>: Avoid influence of movement assumptions. Preference is to estimate separate recruitment deviate parameters for each region (large number additional parameters). Reduce number of regions, excluding R3. Perform initial model testing with each region spatially discrete. Then add movement parameterization to examine interaction between recruitment and movement dynamics.

14. Movement

The spatial model includes movement among the regions and the movement parameters are estimated in the stock assessment. Tagging data provides the most direct information on movement, but because of the tag mixing issues mentioned above, this information is likely biased. The movement is also informed by the assumption of equal catchability among regions for the longline index of abundance. However, this is an indirect source of information and may be biased for a variety of reasons. In addition, the movement parameters may be confounded with the estimated recruitment. Movement may be influenced by the environmental conditions, which may add additional variability. Despite these environmental variations, the mean movement rates are the most important, unless there are substantial trends over time. There is some evidence that the movement of juveniles is related to ocean currents and the movement of adults is related to temperature.

Recommendation:

1. Model complexity: It is recommended that model development start with a single model in region 1 to solve issues with the assessment. This will isolate those issues from the confounding with movement and any issues with estimating movement. Models with multiple regions and movement can then be build using that model as a foundation.

15. Initial conditions

The model starts in 1950, when the fisheries commenced, in an unfished equilibrium. Recruitment deviations do not start until 1975 when the CPUE data becomes available. The initial conditions are difficult to estimate, and regime shifts in recruitment are common in tuna assessments. Therefore, care is needed when estimating the initial conditions before there is length composition data since it will impact the estimate of virgin recruitment (R0), which might be used in determining management quantities. It is recommended that models with alternative dates of the initial conditions (e.g. 1975) should be conducted and compare with longer-term models to determine if the R0 is the same, and if not, it is better to start in a later time, when the length frequency starts or adjust R0 used in the calculation of management quantities. The shorter-term models may require estimating and initial F for two fleets, one for large fish and one small fishery, but not fitting to the average catch, to have flexibility to approximate the exploited age structure.

Recommendations:

1. Initial year: Compare models with alternative dates of the initial conditions (e.g., 1975 versus 1950's) and determine if the R0 is similar, and if not, start in the model at a later time, when the length frequency starts or adjust R0 used in the calculation of management quantities (see also recommendation 11.5 – Early recruitments).

16. Data Weighting

The model tunes to the length composition for each fleet, to tagging data, and to CPUE indices for the longline fleets. The relative weighting between these datasets is important. Tagging data is effectively weighted by the overdispersion parameter, which has been set at a fixed value (7) with sensitivity analysis around this. CPUE is weighted by use of a constant CV, which has been estimated outside the model loosely based on the residuals from a smoother to the annualized CPUE. The length distributions are weighted individually using effective sample size, which has been sex at a fixed low value of 5 for all fleets.

At present, all the length compositions are set to have an effective sample size of 5. This was done to avoid downweighting the poorest data fleets to the point where the model was unable to estimate the selectivity. However, there is a considerable variability in the quality of the data between the fleets, and this equal weighting fails to adequately reflect this. By not allowing higher effective sample size for the more data rich fleets, the model is both underweighting these relative to the other fleets and may be underweighting the length data as a whole in relation to the CPUE and tagging data, and hence upweighting the CPUE and tagging data. The poor weighting of length compositions matters because the length composition data does not simply tune the selectivities in the fleets, but also gives the model information on population size, recruitment, and movement.

The length composition data provides important information on recruitment strengths, which is needed to reliably extract absolute abundance information from the indices of relative abundance through the depletion signal caused by catch. Therefore, including reliable length composition data, weighting it appropriately, and correctly specifying the processes is important. In general, fisheries with selectivities assumed to be asymptotic, as are the longline fisheries in this assessment, provide the most information on absolute abundance, and particular attention should be given to these fisheries. For fisheries with particularly problematic length composition data (e.g. highly variable over time), the composition data should not be used and the selectivity borrowed from another similar fishery or fixed appropriately. For fisheries with large catches additional effort should be taken to specify the selectivity correctly.

Data weighting is an important component of the stock assessment process and influences both the overall uncertainty and the influence of each data set on the results. Data weighting also must be considered in the context of model misspecification and unmodelled process variation. Overweighting of the composition data is dangerous, particularly if a process that influences the predicted composition is misspecified and/or the selectivity is asymptotic. The stock assessment has three main types of data, indices of relative abundance, length composition data, and tagging data, which all need to be weighted appropriately. Unfortunately, good practices for data weighting and modelling of process variation have not fully been developed.

Conclusions:

The externally derived weighting for tagging data and CPUE is likely appropriate, but the lack of relative weighting of composition data between fleets is a key deficiency in the model. It is not clear to what extent this is having impacts on the results of the model, and this needs to be investigated. Ideally the weighting of length distributions should be based on realistic estimates of effective sample size, although some proxy may be needed.

Recommendations:

1.<u>Length composition data weights</u>: We would strongly encourage assigning more realistic weighting for the fleets. It may be reasonable to keep a minimum effective sample size for the data-poor fleets, but the data-rich fleets should be upweighted to reflect the better data. Note that this will change relative weighting between the CPUE/tagging and the length distributions, so careful investigations are needed here. It is recommended that until good practices have been determined, the Francis method be used for fisheries with adequate years of data and the McAllister and Ianneli or the Dirichlet used for the other fisheries.

2. <u>Selectivity assumptions</u>: For fisheries with particularly problematic length composition data (e.g. highly variable over time), the composition data should not be used and the selectivity borrowed from another similar fishery or fixed appropriately. For fisheries with large catches additional effort should be taken to specify the selectivity correctly.

17. Diagnostics

Diagnostics for integrated models are an important component of stock assessment development. They help understand and fix issues with the assessment. A small number of diagnostics were used. The type of diagnostics applied to the yellowfin assessment should be expanded.

Recommendations:

Some diagnostics that should be added include:

1. Recruitment trends – plot the recruitment pattern from the retrospective models to see if there is an effect of the tagged data.
2. Age-structured production model (ASPM)
3. Retrospective
4. Hindcasting
5. Likelihood profile on M and movement parameters.
6. Empirical selectivity diagnostic

18. Uncertainty

Estimation of uncertainty is important to put the management advice in context. The representation of uncertainty should include all the sources including both parameter and model uncertainty. Using the conceptual model to set up the hypotheses and the models in the grid may also be useful. Improving the assessment will likely eliminate models from ensemble and new models are likely to be added as the assessment is further investigated. Only include models in ensemble that are reasonable (i.e., use diagnostics to eliminate unreasonable models). The approach used for IO yellowfin could be improved using the following recommendations.

Recommendations

1. Use diagnostics to discard unreasonable models
2. Consider adding additional models to the grid
3. Include different regional structures in the uncertainty grid
4. Use equal weighting until a better alternative has been identified
5. Include parameter uncertainty when combining models in the ensemble (e.g., use the multivariate normal distribution to represent estimate uncertainty from a single model and combine the normal distributions from multiple models in the ensemble)

19. Management (reference points, projections)

The ultimate goal of stock assessment is to provide management advice. Therefore, in addition to ensuring that the assessment is done correctly, care needs to be taken to calculate the management quantities derived from the assessment results correctly. For example, the recruitment used to define the reference points and projections should be checked to ensure that it is based on recruitment averaged over the desired period.

The spatial nature of tuna stocks and fisheries raise additional issues. The core area is usually estimated to be heavily exploited while peripherical areas are estimated to be much less exploited and these can contribute substantial biomass to the total stock. This could be an artifact of the model, and may be masking the actual depletion of the population.

20. Collaborative process and assessment development

The assessment was developed in a collaborative manner and documented well, but further collaboration could improve the assessment and better documentation could improve future reviews and understanding by stakeholders. Some recommendations include

1. Better document the assessment diagnostics
2. Share the assessment files and automatically produced SS html
3. Document any changes done in the working group meeting that modified the model ensemble used to provide management advice.
4. Post the html and model files of the final base model, share the control files, par file and report file for all the models in the ensemble.
5. Conduct a pre-assessment workshop to evaluate data and model assumptions. (there is an existing data preparation meeting to discuss CPUE input).

21. Future research and identification of priorities.

Several areas of priority research were identified. These include

1. Develop a conceptual model specifically to improve the understanding of spatial structure to use in the assessment.
2. The longline indices of abundance are the most important data in the model and further extensive work needs to be conducted to improve them. This should include a working group involving all the tuna RFMOs. In particular, techniques should be developed to integrate data from multiple longline fleets.
3. The tagging data needs to be evaluated outside the stock assessment using a fine scale spatial temporal model to address concerns with mixing. This approach may be able to estimate movement, natural mortality, biomass, and/or fishing mortality that can then be used in the stock assessment.
4. There are still concerns about the growth used in the assessment. Obtaining reliable estimates of growth is particularly important for assessments that use length composition data. More validation work is needed. More aging data with a wider spatial and temporal range is needed.
5. Natural mortality is a influential parameter in determining results from stock assessment models and the consequent management advice. Further work is need to better understand and estimate natural mortality including the use of a sex structured model.

22. References.

Fonteneau, A. 2008. A working proposal for a Yellowfin growth curve to be used during the 2008 yellowfin stock assessment. IOTC-2008-WPTT-4.

Xu, H., and Lennert-Cody, C.E. 2023. FishFreqTree: IATTC's regression tree R package for length frequency data. https://github.com/HaikunXu/FishFreqTree

Lennert-Cody, C.E. Minami, M., Tomlinson, P.K. and 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://www.sciencedirect.com/science/article/abs/pii/S0165783609003166

Le Blanc,J.-L. 2002. The Indian Ocean Climatology & Oceanography (IOCO) Gateway: I. Annual and Semiannual Variations on a Regional Scale. webpage http://indianocean.free.fr/annual1.htm Access on 9/08/2023

Mildenberger, T.K., Nielsen, A., Maunder, M.N. 2022. Length-structured spatiotemporal tagging model for skipjack in the EPO. Document SAC-13-INF-E https://www.iattc.org/GetAttachment/5d5a8b6b-8974-4d83-9072-4aeadeae43c2/SAC-14-INF-E_Spatiotemporal-tagging-model-for-skipjack-in-the-EPO.pdf

Appendix A. Panel requests

February 08th 2023

#	Theme	Request	Rationale	Results/Discussion
1	MSY	**MSY estimates based on each year in the model** Plot the equilibrium MSY computed for each year for the base model. There is no need to run the model again, simply to change the period for which the MSY is computed in the forecast file. **Sanity check on the MSY estimate** Also, a plot showing how do these time series of MSY estimates relate to the catches over the last few decades?	The last couple assessment models have shown a decrease in the MSY even with increase in catches. The most recent model estimates an MSY that is lower than the average annual catches for the model period. This request is to investigate whether the change in estimated MSY is related to the change in the mix of fisheries and/or in the selectivity of the fisheries	There is a big change in the early 1980's when the MSY decreases and increases back again in the early 1990's when the catches in area 4 increases. The MSY is very stable after 1995, but it is unclear why.
2	Recruitment	Plot the absolute recruitment by season and area (panel plot – seasons in rows and areas in columns)	This request is to investigate two things: 1) whether the recruitment is consistent with the hypothesis that the main spawning season is thought to be from December to March. It is expected that the main recruitment to the fishery that catches the smallest fish will be a couple of months after the main spawning season. 2) There is a trend in recruitment deviations over time in the two regions, it is not clear how important is this trend in absolute value. The trend in absolute recruitment by region is to evaluate how the model is accounting for trends in catch in each region mediated by the trend in abundance from CPUE indices	Recruitment is higher in Q3 and Q4 for R1 and in Q4 and Q1 for R2. Recruitment is higher in R1.
2	Longline Data	Plot the locations of the longline length frequency samples and compare it to the catches for each flag for which data is used in the model	The size data for the longliners is very variable in the recent years when several fleets are used. Japanese data shows decreasing trends in average size while other fleets show increasing trend. It is not clear if this is an effect of the spatial location of the samples or the coverage	There are not many samples for areas 1b Arabian sea

3	Longline catch data	Plot LL catches by season and area (panel plot – seasons in rows and areas in columns)	To be used to interpret the spatial model and the coverage of the size data	
4	Other catch data	Plot catches of other gears in the model by seasons and area (panel plot – seasons in rows and areas in columns)	To be used to interpret the spatial model	
5	Fisheries	Produce a table by fishery with name, Average catch whole time series, Average catch last 10 years, Reliability of catch data, Number of years with LF data, Average sample size of LF whole time series, Average sample size of LF last 10 years, Stability of selectivity over time, Form of selectivity used and whether it was estimated or mirror.	This request is to understand how important each fishery is, that is whether they should be fit well or not, and how much weighting should be given to the size composition data.	
6	Average size	Time series of average size by longline fleet	To be added to the SS3 data file but not used for estimation, so the model generates a predicted value. This could be used to evaluate how consistent is each data with the model (evaluate what size data would be consistent with the decline to 10% of the virgin biomass)	

February 9th 2023

#	Theme	Request	Rationale	Result
1	Length composition	One area model with new growth curve and size at age 1 estimated with bounds between 20 and 40	to see if the length composition modes can be tracked for fishery 4 (mixed gears), fishery 5 (bait boat) that show the smallest fish.	

February 10th 2023

#	Theme	Request	Rationale	Result/Discussion
1	Length composition	One area model with new growth curve and size at age 1 estimated with bounds between 10 and 40	to see if the length composition modes can be tracked of fishery 4 (mixed gears), fishery 5 (bait boat) that show the smallest fish.	The estimate went to 16 cm. The estimates of selectivity for fishery 4 went to a peak in age 10, which is wrong. It did not converge correctly.
2	Length composition	as Run 1 with the CV of variation in length at age for the large fish set to 10% (so all CV are increased)		Fitted the data better. The age-based selectivities converged to the correct ages for F4 and F5. The model is still predicting sizes that are slightly larger

				than the observed. F9 (PS FS large fish) the model cannot predict the intermediate sizes and cannot pick up the little mode. There is a consistent mode in F9, which may indicate that the fish stop growing. There are some cohorts visible, for example 264, 265, 266 , and seem to grow very fast. Some years the fishery catches the small fish and some years not. It may be a case for time-varying selectivity. Considerations are needed to see whether to use the information from in this fishery to inform the stock size. Do the tree analysis to see if you can explain the appearance of the modes. Is the size similar to the longline? To see if it needs asymptotic selectivity. Cut the LF at the peak on average mode size to get the decay of the larger size and if it is consistent with the longline. F1 – gillnet fishery. There is modal progression sometimes, it is hard to fit. F2 - handline – 293 impossible to fit. There may be a lot of spatial structure that we are not picking up (movement in and out of these areas).

Appendix B: Panel Biographies

Mark Maunder is the Head of the Stock Assessment Program at the Inter-American Tropical Tuna Commission. He received his B.Sc (Zoology and Computer Science), M.Sc (Zoology) at the University of Auckland and Ph.D. (Fisheries) at the University of Washington. Before joining the IATTC, Dr Maunder was a Quantitative Fisheries Scientist at the New Zealand Fishing Industry Board. His research interests include development of statistical methodology for fisheries stock assessment, protected species, and ecological modeling. He has coauthored over 100 papers in the peer- reviewed literature, along with many technical reports. Dr Maunder was co-founder and past president of the AD Model Builder Foundation, was a member of the Partnership for Mid-Atlantic Fisheries Science (PMAFS) Science Advisory Committee, and is Council Member of the Fisheries Integrated Modeling System (FIMS). Mark and his colleagues have been involved in extensive research into the development and application of fisheries stock assessment models. He was an early pioneer and advocate of the integrated assessment approach to stock assessment. His Phd dissertation involved integrating tagging data into stock assessment models. He was also the lead programmer of the general stock assessment model Coleraine that was an early ADMB based general model, and extensively used the integrated approach in a Bayesian framework, and codeveloped ASCALA that was used for assessing tunas in the EPO. He has also applied integrated analysis to protected species. In 2012, Mark co-founded the Center for the Advancement of Population Assessment Methodology (CAPAM; http://capamresearch.org/ [capamresearch.org]). The main activities of CAPAM revolve around the workshop series and associated special issues in the journal *Fisheries Research*. Mark has co-organized all the CAPAM workshops and chaired most of them. He has also been a guest editor for all the special issues is an Editorial Board Member with Fisheries Research. CAPAM has built an excellent reputation over the time it has been in existence, which has been recognized through being awarded the American Fisheries Society's (AFS) William E. Ricker Resource Conservation Award in 2018 and the American Institute of Fishery Research Biologists' (AIFRB) Outstanding Group Achievement Award in 2017.

Adam Langley

Daniel Howell

Carolina Minte-Vera is a Senior Stock assessment Scientist at the at the Inter-American Tropical Tuna Commission. At the IATTC, she is involved in research related to stock assessment, she is the lead assessor for yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean (EPO), and for the south EPO swordfish (*Xiphias gladius*). She collaborates in the EPO bigeye tuna (*Thunnus obesus*) assessment, as well as assessments for other species. She was part of the ISC (International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean) Albacore working group for 10 years, which does the north Pacific albacore tuna (*Thunnus alalunga*) assessment and did the management strategy evaluation for that stock. She currently is part of the ISC billfish working group, which assess the swordfish, the striped marlin and blue marlin stocks in north Pacific Ocean. She received her B.Sc. in Ecology at the São Paulo State University, Rio Claro (UNESP), Brazil and her M.Sc. in Ecology at the University of Campinas, Brazil, Specialization Applied Statistics, State University of Maringá, Brazil, PhD at the School of Aquatic and Fishery Science of the University of Washington, Seattle, USA, and Postdoctoral fellowship at the Oceanographic Institute, University of São Paulo, Brazil. Before joining the IATTC, she was assistant professor at the State University of Maringá, where she taught graduate and undergraduate courses, guided graduate students, and did research in stock assessment and fisheries. She has been advisor for several governmental and non-governmental organizations, I and has extensive experience on research related to artisanal fisheries, aquatic ecology, sampling design, data analysis and quantitative methods. Other past experiences include member of Brazilian Scientific Committee for Tuna and Tuna-like species and Brazilian representation to ICCAT; assistant editor in ecology for the journal "Neotropical Ichthyology"; visiting scientist New Zealand Seafood Industry Council Ltd., SEAFIC, New Zealand; research assistant, School of Aquatic and Fisheries Sciences, University of Washington; data analyst, at Research Center on Limnology Ichthyology and Aquaculture, University of Maringá; review Editor for the Assessment of the sustainable use of wild species of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), instructor for data analyses and stock assessment (including SS3 platform) courses to agencies/universities, assessor for fisheries against the Marine Stewardship Council Principles and Criteria; principal investigator for participatory monitoring and management of artisanal fisheries in the Abrolhos Bank (Brazil); researcher in the Long-Term Ecological Research at the Rio Paraná floodplain, and in ichthyological and fisheries studies in reservoirs such Itaipú, Manso, Billings. She has authored more than 100 publications, for complete list of publications please see: http://scholar.google.com/citations?user=wyVvP5gAAAAJ&hl=en. She is fluent in English, Spanish, Portuguese, and French.

Appendix C: Term of reference and agenda

Term of Reference for the 1st yellowfin stock assessment review workshop

Introduction and background

- IOTC's Working Party on Tropical Tunas and Scientific Committee have noted areas of uncertainty of the stock assessment of this stock that require follow-up investigation and expert advice.
- Outcome of the assessment affected by alternative model configurations, parameters and assumptions. Need to improve confidence on model results and developed management advice.
- Problems comparable to other tropical tuna stock assessments.

Objectives of the meeting

- External review of the 2021 stock assessment.
- Provide recommendations for improving the asesssment, including data inputs, model configuration, biological parameters, modelling approach and treatment of uncertainty.
- Identify improvement options for 2024 stock assessment and provision of management advice.

Scope (key areas for consideration)

Model inputs:

- Growth
- natural mortality
- catch
- tagging data
- size composition
- CPUE

Model configuration, assumptions and settings:

- Complexity
- selectivity assumptions
- treatment of uncertainty.

Model diagnostics:

- Review diagnostics used in 2021 and developed thereafter.

Future research and identification of priorities.

Expected output

Meeting report (due October 2023, earlier if possible)

- Summary of discussions
- Recommendations from review panel.
- To be delivered in the next WPTT (Annual meeting (October)) and SC (December) 2023.

Agenda of the meeting

1. **OPENING OF THE MEETING** (Chair)
2. **ADOPTION OF THE AGENDA AND ARRANGEMENTS FOR THE SESSION (**Chair**)**
    - Terms of Reference, key documents and functioning of the meeting.
3. **PRESENTATION OF THE 2021 STOCK ASSESSMENT (**Stock assessment team**)**
    - General introduction and summary of issues
    - Presentation of model inputs
    - Model specifications and assumptions
    - Model Diagnostics
    - Model outputs and projections
4. **REQUESTS FOR MODEL RUNS TO BE RUN DURING THE MEETING** (Chair)
5. **DISCUSSION ON ADDITIONAL RUNS** (Chair)
6. **EXPERT PANEL FEEDBACK**
7. **FINAL COMMENTS AND SUMMARY OF WORKSHOP OUTPUTS**
8. **OUTLINE OF REPORT**
9. **ANY OTHER BUSINESS**

Appendix D : Lists of documents and presentations

Annotated list of selected documents

On the 2021 stock assessment of yellowfin

- *Fu, D. et al. 2021. Preliminary Indian Ocean yellowfin tuna stock assessment 1950-2020 (stock synthesis). IOTC–2021–WPTT23–12.*

This is the latest stock assessment. Comparing to previous assessments (2015, 2016 and 2018), this model estimates a more optimistic stock status (BBmsy and FFmsy) but notably lower productivity (R0 and MSY) for this stock. The document contains a wide description of the model structure, data used, sensitivity analyses and justifications to the choices made.

- *Merino et al 2022. Investigating trends in process error as a diagnostic for integrated fisheries stock assessments. Fish Res, 256 (2022) 106478.*

This document evaluates the trends in recruitment deviates for tropical tunas worldwide with special focus on the 2021 stock assessment of Indian Ocean yellowfin. Trends in recruitment deviates are linked to extreme productivity scenarios and help explain the average low MSY estimated in the 2021 SA. The management implications of the trends in rec devs are explored in the document (*IOTC-2022-WPTT-24-15 management implications.pdf*) available in the general document folder.

- *Review of 2021 WPTT IOYT stock assessment and feasibility of alternative assessment (Landmark Fisheries Research)*

This review is part of the work that Landmark FR has started to develop an alternative stock assessment for yellowfin. This first document contains an initial review to the structure and data of the model.

Past stock assessment reviews

- *Sharma, R. 2018. Review of IOTC YFT in 2018. IOTC-2018-SC21-INF02.*

This document contains the review of the external expert of the 2018 stock assessment of yellowfin. The document was developed during the WPTT and evaluates different aspects of the 2018 SA. Many of the recommendations have been addressed in the 2021 SA but the document also identifies uncertainties that require further exploration.

- *Problems_with_projections_2018 (presented to the 2020 SC meeting)*

These few slides summarize the problems identified with the projections of the 2018 stock assessment. It shows how the model needs to change the initial regional distribution of recruits in order to compensate the increased catch in the NW area after the development and expansion of the purse seine fishery. In the projections, the model uses the initial rec distribution and therefore, the NW area runs out of fish to sustain current catches and crashes in 1-2 years. This was only identified when looking at each model of the grid individually and helps understand the model's way to support recent catch in the NW area. Besides the change in recruitment distribution, the model would increase the rec devs to compensate the increased catch (see Merino et al 2022).

- *Methot, R. 2019. Recommendations on the configuration of the Indian Ocean yellowfin tuna stock assessment model.*

This document was prepared by Dr Richard Methot after the problems identified with the projections in the 2018 stock assessment. It contains general recommendations about tagging data, recruitment distribution, scaling, selectivity and other parts of the model.

Documents related to CPUE

- *Hoyle, S.D., et al. 2018. Collaborative study of yellowfin tuna CPUE from multiple Indian Ocean longline fleets in 2018. IOTC–2018–WPM09–12.*

This document describes the development of the Joint Longline CPUE. The document describes the different indices aggregated in the joint index, the alternative modelling and data transformation methods explored, and all analyses for the development of the joint index, which is the basis of the assessments of Indian Ocean yellowfin, bigeye and albacore.

- *Hoyle, S.D., Langley, A. 2018. Indian Ocean tropical tuna regional scaling factors that allow for seasonality and cell areas. IOTC-2018-WPM09-13.*

The regional scaling factor has a high influence on the estimated relative abundance among the areas in the Indian Ocean yellowfin stock assessment and this document explains how the relative abundance among regions is estimated.

- *Kitakado, T. et al. 2021. Report of trilateral collaborative study among Japan, Korea and Taiwan for producing joint abundance indices for the yellowfin tunas in the Indian Ocean using longline fisheries data up to 2019. IOTC–2021-WPTT23(DP)-14.*

This document describes the development of the Joint LL CPUE for the 2021 stock assessment.

Documents related to size data

- *Hoyle, S. et. al. 2021. Review of size data from Indian Ocean longline fleets, and its utility for stock assessment. IOTC-2021-WPTT23-07.*

This report reviews the procedures used to collect and process longline size data for use in IOTC stock assessments. It describes the types of data collected, with a particular focus on data provided by the Japanese, Taiwanese, Korean and Seychelles fleets. It investigates the reliability of size data by comparing spatial and temporal patterns in median size among fleets and time periods. It explores reasons behind sudden changes in the shape of length frequency distributions for the Taiwanese fleet and recommends that stock assessments should in future omit Taiwanese length data but include weight data and observer data.

Documents related to tagging data

- *Hoyle, S.D. et al. 2015. Covariates of release mortality and tag loss in large-scale tuna tagging experiments. Fisheries Research 163, 106-118.*

This work analyses tagging data from the Indian Ocean and Western Pacifici tuna tagging programs to estimate tag mortality from differences on tag return rates among taggers. The tag mortality used in the IOTC stock assessments increased notably after this document (from 10% to 27.5%) and this parameter has large implications on the outcome of the stock assessment. In brief, with higher tag mortality, the model increases the estimated fishing mortality as there is less fish available for recapture.

- *Fu, D. 2020. Tag data processing for IOTC tropical tuna assessments. IOTC–2020–WPTT22(DP)–10.*

This report summarises how the tagging dataset were processed for incorporation into the recent Stock Synthesis assessments for yellowfin, bigeye, and skipjack tuna. The procedure includes filtering of dubious records, correction for potential tag loss, and adjustment for under-reporting of recaptures

- *Simon Hoyle during the WPTT2019_Options for progressing WPTT assessments*

These notes were developed during the 2019 WPTT annual meeting. It is suggested that the tagging data is a strong source of scaling information as CPUE data and size frequency data do not show enough contrast through time and SFD has problems (see Hoyle et al., 2021). It suggests additional information is necessary to help the model estimate biomass scaling. Some of these recommendations have already been applied in the 2021 SA (e.g. diagnostics).

List of all background documents

Most recent stock assessment report

Fu, D. et al. 2021. Preliminary Indian Ocean yellowfin tuna stock assessment 1950-2020 (stock synthesis). IOTC–2021–WPTT23–12.

Urtizberea et al. 2021. Indian Ocean yellowfin tuna SS3 model projections. IOTC-2021-WPTT23-22.

IOTC 2021. Report of the 23rd Session of the IOTC Working Party on Tropical Tunas. Online, 25 - 30 October 2021. IOTC–2021–WPTT23–R.

Merino, G. et al. 2022. Analysis of recruitment deviates of tropical tuna stock assessments. IOTC-2022-WPTT24(DP)-19.

Merino, G. et al. 2022. Management implications of trends in recruitment deviates of Indian Ocean yellowfin

Merino, G. et al. 2022. Investigating trends in process error as a diagnostic for integrated fisheries stock assessments. Fisheries Research 256.

Past stock assessment reviews

Sharma, R. 2018. Review of IOTC YFT in 2018. IOTC-2018-SC21-INF02.

Method, R. 2019. Recommendations on the configuration of the Indian Ocean yellowfin tuna stock assessment model.

Johnson, S. et al. 2022. Review of 2021 WPTT Indian Ocean yellowfin tuna stock assessment and feasibility of alternative assessment. IOTC-2022-WPTT24-17.

Stock structure and biology

Grewe, P., et al. 2020. Genetic population connectivity of yellowfin tuna in the Indian Ocean from the PSTBS-IO Project. IOTC-2020-WPTT22(AS)-12.

Fonteneau, A. 2008. A working proposal for a Yellowfin growth curve to be used during the 2008 yellowfin stock assessment. IOTC-2008-WPTT-4.

Dortel, E., et al. 2014. An integrated Bayesian modeling approach for the growth of Indian Ocean yellowfin tuna Fisheries Research.

Farley, J. et al. 2021. Estimating the age and growth of yellowfin tuna (*Thunnus albacares*) in the Indian Ocean from counts of daily and annual increments in otoliths. IOTC-2021-WPTT23-05_Rev1.

Eveson, J. P., Farley, J. 2021. Investigating growth information for yellowfin and bigeye tuna from the IOTTP tag-recapture data. IOTC-2021-WPTT23-21.

Hoyle, S. et al. 2021. Approaches for estimating natural mortality in tuna stock assessments: application to Indian Ocean yellowfin tuna. IOTC-2021-WPTT23-08_Rev1.

Hoyle, S. et al. 2022. Approaches for estimating natural mortality in tuna stock assessments: Application to global yellowfin tuna stocks. Fisheries Research. 257:106498; 2023.

<u>Yellowfin tuna statistical data</u>

IOTC Secretariat. Review of yellowfin tuna statistical data. IOTC-2021-WPTT23(DP)-07_Rev1

<u>Documents related to CPUE</u>

Hoyle, S.D., et al. 2017a. Causes of the historical discontinuity in Japanese longline CPUE. IOTC-2017-WPM08-19.

Hoyle, S.D., et al. 2018. Collaborative study of yellowfin tuna CPUE from multiple Indian Ocean longline fleets in 2018. IOTC–2018–WPM09–12.

Hoyle, S.D., Langley, A. 2018. Indian Ocean tropical tuna regional scaling factors that allow for seasonality and cell areas. IOTC-2018-WPM09-13.

Kitakado, T. et al. 2021. Report of trilateral collaborative study among Japan, Korea and Taiwan for producing joint abundance indices for the yellowfin tunas in the Indian Ocean using longline fisheries data up to 2019. IOTC–2021-WPTT23(DP)-14.

Guery, L., et al. 2021. European Purse Seine CPUE Standardization: Methodology and Framework for the YFT Stock Assessment. IOTC-2021-WPTT23(DP)-16

Baidai, Y., et al. 2021. Associative Behavior-Based abundance Index (ABBI) for yellowfin tuna (Thunnus albacares) in the Western Indian Ocean. IOTC-2021-WPTT23(DP)-15.

Medley, P., Ahusan, M. 2021. Bayesian Skipjack and Yellowfin Tuna CPUE Standardisation Model for Maldives Pole and Line 1970-2019. IOTC-2021-WPTT23(DP)-13.

Kolody, D. 2018. Estimation of Indian Ocean Skipjack Purse Seine Catchability Trends from Bigeye and Yellowfin Assessments. IOTC–2018–WPTT20–32.

<u>Documents related to size data</u>

Duparc, A. et al. 2019 Assessment of the species composition of major tropical tunas in purse seine catches: a new modelling approach for the tropical tuna treatment processing Case of the French fleet in Indian Ocean. IOTC-2019-WPTT21-10.

Hoyle, S.D., et al. 2017b. Selectivity changes and spatial size patterns of bigeye and yellowfin tuna in the early years of the Japanese longline fishery. IOTC-2017-WPTT19-34.

Hoyle, S.D., et al. 2017c. Exploration of Japanese size data and historical changes in data management. IOTC-2017-WPTT19-35.

Hoyle, S. et. al. 2021. Review of size data from Indian Ocean longline fleets, and its utility for stock assessment. IOTC-2021-WPTT23-07.

Documents related to tagging data

Hallier, J.P., Million, J. 2009. The contribution of the regional tuna tagging project – Indian Ocean to IOTC stock assessment. IOTC-2009-WPTT-24.

Hillary, R.M., et al. 2008. Reporting rate analyses for recaptures from Seychelles port for yellowfin, bigeye and skipjack tuna. IOTC-2008-WPTT-18.

Langley, A., Million, J. 2012. Determining an appropriate tag mixing period for the Indian Ocean yellowfin tuna stock assessment. IOTC–2012–WPTT14–31

Gaertner, D.; Hallier J.P. 2015. Tag Shedding by Tropical Tunas in the Indian Oceana and other factors affecting the shedding rate. Fisheries Research 163, 98-105.

Hoyle, S.D. et al. 2015. Covariates of release mortality and tag loss in large-scale tuna tagging experiments. Fisheries Research 163, 106-118.

Fu, D. 2020. Tag data processing for IOTC tropical tuna assessments. IOTC–2020–WPTT22(DP)–10.

List of presentations

Review of the IOTC statistical data available for yellowfin tuna

2021 IO yellowfin tuna stock assessment model inputs – observations and biological parameters

2021 IO yellowfin tuna stock assessment model structure – configurations and parameterization

2021 IO yellowfin tuna stock assessment model outputs – diagnostic, sensitivity, and uncertainty grid

Additional model runs to address Panel's request – Part one

Additional model runs to address Panel's request – Part two

Appendix E: List of participants

Manuel Baranger – Director Fisheries and Aquaculture division FAO gave welcome to the participants. Blue transformation strategy expand aquaculture in areas with food insecurity, put all fisheries into management, add value to products. Recognize fisheries as important for food security.

Name	Role	Institution	Email
Mark Maunder	Review Panel Chair	IATTC	mmaunder@iattc.org
Carolina Minte-Vera	Review Panel member	IATTC	cminte@iattc.org
Adam Langley	Review Panel member	NIWA	adam_langley@xtra.co.nz
Daniel Howell	Review Panel member	IMR, Norway	daniel.howell@hi.no
Toshi Kitakado	Chair of IOTC Scientific Committee		kitakado@kaiyodai.ac.jp

Gorka Merino	Chair of the IOTC working party of the tropical tunas	Azti Technalia	gmerino@azti.es
Agurtzane Urtizberea Ijurco	Analyst	Azti Technalia	aurtizberea@azti.es
Dan Fu	Analyst	IOTC	Dan.Fu@fao.org
Hilario Morua	Chair of the IOTC Working Party on Methods	ISSF	hmurua@iss-foundation.org
Rishi Sharma	Observer	FAO	Rishi.Sharma@fao.org
Henning Winker	Observer	GFCM - FAO: General Fisheries Commission of the Mediterranean	henning.winker@gmail.com