Landmark Fisheries Research
213 – 2414 St. Johns Street
Port Moody, British Columbia
V3H 2B1

Sep 19, 2022

Prepared By: Samuel Johnson, Sean Cox, and Ashleen Benson, Landmark Fisheries Research

Attention: Anne-France Mattlet, Europêche Tuna Group

Regarding: Review of 2021 WPTT Indian Ocean yellowfin tuna stock assessment and feasibility of alternative assessment

Executive Summary

Recent IOYT Assessment. The most recent Indian Ocean Yellowfin Tuna (IOYT) stock assessment (Fu et al 2021) estimated the 2020 IOYT stock as slightly overfished at 78% of B_{MSY}, with 68% probability that overfishing occurring (fishing mortality at 127% of F_{MSY}). These results are found by integrating over a grid of 96 equally weighted stock assessment models (Stock Synthesis v3) that are intended to span the range of plausible states of nature for the IOYT stock.

Key issues. We identified four key issues with the IOYT stock assessment, some of which are known by the authors and mentioned in the introduction and discussion sections of the most recent assessment (Fu et al 2021). The four issues are (i) an overly complex spatial structure, (ii) a lack of fit to tagging data, (iii) overweighting some models in the ensemble assessment grid, and (iv) poor fits to length composition data. These four issues are largely a result of historical modeling choices, and a strong preference for including tagging data from the regional tuna tagging programme.

Proposed alternative. There are two alternative paths forward. The first, which we recommend, is overhaul the IOYT stock assessment model via a new custom IOYT stock assessment is developed from the ground up to suit the specific information available. The second option is to continue with Stock Synthesis and focus on incremental changes to improve the assessment. This would require less time, as most changes would be restricted to input data or assumptions about the spatial stratification of the stock, given that changes to the Stock Synthesis design and computer code are practically impossible for anyone who is not a Stock Synthesis developer. Major advantages of an overhaul are (i) it would be a small step from the assessment model to an operating model for management strategy evaluation or rebuilding plan development and (ii) all model elements can be crafted to suit IOYT specifically with far fewer compromises.

Budget. We estimated a maximum time budget of around 262 days (approximately 1 year of full time work for a single analyst) to complete a model overhaul resulting in a new custom fishery

stock assessment model (and almost operating model) for IOYT (Appendix A). This budget covers consultation, funding applications, model design, sensitivity analyses, and testing of various model components, and ends with estimates of biomass, productivity, stock status relative to biological reference points, and a framework for future management strategy evaluation. While the budget does not extend to closed-loop simulation work for management strategy evaluation per se, Landmark Fisheries Research has extensive experience with this type of work and can include estimates for those steps as well upon request. The final price depends on how each component is allocated among junior, intermediate, and senior analysts, with a current day-rate schedule included at the bottom of the appendix.

Section 1: Background

Fisheries stock assessment models are the scientific basis of fisheries management decisions. The quality of those decisions therefore is highly dependent on the quality of inferences extended by those models, which is related to both goodness of fit of the model to the input monitoring data and optimizing the parameterization to generate somewhat reasonable predictive inference. Biased or highly randomized model inferences (i.e., too many random effects parameters) could translate into biased management targets, such as target fishing mortality rates (F_{MSY}) and sub-optimal outcomes for the fishery.

The 2021 Indian Ocean yellowfin tuna (IOYT) assessment is done via an age- and spatially-structured (4 region) population model within Stock Synthesis (Methot et al 2020; Methot & Wetzel 2013). The model is fit to fishery dependent data, consisting of catch series, CPUE indices, catch-at-length compositions, and tag releases and recoveries (most recoveries were from the Purse seine fleet). This model estimates a 68% probability of IOYT being currently overfished and subject to overfishing (Fu et al. 2021).

This report reviews the 2021 IOYT stock assessment history, from its origins in 2008 when the first statistical catch-at-length model was fit to IOYT data (Langley et al 2008) and culminating with the 2021 base model, its underlying major assumptions, and how spatial stratification has changed since 2018. We then list a series of mostly known issues with the 2021 model, which we found to be primarily based on inertia from previous modeling choices and a strong preference for including tagging data that may not be as informative as originally hoped. Finally, we conclude with recommendations for major and minor changes to the IOYT assessment along with a proposed budget for a new stock assessment model addressing some of the major concerns in an appendix. In general, we recommend a shift to a simpler model structure that includes all sources of data but reflects the evidence of those data as well the associated limitations of how they are collected.

Known issues with current IOYT assessment model

There are several issues with IOYT assessments that are openly acknowledged by the authors. For example, recent models produce pessimistic estimates for IOYT population dynamics parameters, which, when used in forward projection models, leads to stock crashes within a few years even at low fishing pressure (Fu et al 2018b). Such behaviour was later attributed to assumptions about how quarterly recruitment is distributed among regions (IOTC 2020). Ultimately, IOYT assessment working groups decided that the model was not adequate to forecast rebuilding plans because model uncertainty had not been adequately captured in projections (IOTC 2018). Several other operating models based on the 2018 stock assessment that could be used to test possible IOYT harvest strategies also produced unrealistically high fishing mortality rates and pessimistic results (Kolody and Jumppanen 2021), indicating that there may be a fundamental mismatch between the assumed model structure and the input data for the stock assessment. More recent models have eliminated this behaviour, but there are still significant retrospective patterns in key model outputs as well as a tendency to estimate lower productivity as data are removed. Assessment authors ultimately concluded that the movement component of the model is more informed by the assumed spatial structure than the tagging data as originally intended (Fu et al 2021) because (a) the majority of tag releases and recoveries are in the same spatial region and (b) most tag recoveries outside of that region are close to the border and do not suggest large scale movement at all.

Section 2: Model History

Indian Ocean Yellowfin Tuna (IOYT) have been assessed using some form of statistical catch-at-length model since 2008 (Langley et al 2008). Prior to 2008, stock assessment models were either virtual population analyses or age-structured production models (see Fu et al 2021 for references). The statistical catch-at-length models used since 2008 are all age- and spatially-structured population dynamics models which are fit to fishery catch observations, fishery CPUE indices, catch-at-length observations, and tag release/recovery data as mentioned above.

The first catch-at-length model – MULTIFAN-CL (Langley et al 2008) – introduced the ability to fit to data from the large-scale Regional Tuna Tagging Programme conducted in the Indian Ocean (RTTP-IO), with tags released over the 2005 – 2009 period. Indeed, fitting to the tagging data was the main reason cited for migrating to MULTIFAN-CL (Langley et al 2008). To fit a movement model to the tagging data, five spatial strata were defined for IOYT, which are similar to the four strata used in the 2021 assessment. MULTIFAN-CL was used to assess the IOYT stock four more times after 2008 (Langley et al 2009, 2010, 2011, 2012).

In 2015, IOYT assessments switched to Stock Synthesis v3 (SS3; Langley 2015). SS3 is a pre-made fishery stock assessment package that is in common usage in many fisheries across the world (Methot and Wetzel 2013). One advantage of SS3 is that it is maintained by an active development team at a government institution; therefore, SS3 is well supported and its features have been thoroughly tested. On the other hand, SS3 is by design a generic fishery stock assessment model and that implies a trade-off between generic and specific use cases, so SS3 features do not always fully reflect the features of every fishery to which it is applied; however, MULTIFAN-CL was also a pre-made fishery stock assessment package with the same generic/specific use-case trade-offs. In any case, both assessment packages give qualitatively similar results, and SS3 is more actively developed and supported than MULTIFAN-CL, so it was wise to switch to a newer package (Langley 2015). With the switch to SS3, the spatial stratification was revised to 4 areas, with the new region 1 combining what was previously two separated regions for the Arabian Sea and the Western Equatorial Region (Langley 2015). Since 2015, there have been four additional assessments of IOYT with SS3 (Langley 2016; Fu et al 2018; Urtizbera et al 2019; Fu et al 2021), as well as a 2019 review of the IOYT assessment by the creator and lead developer of SS3 that identifies several issues that we include in our discussion below (Methot 2019).

For remainder of this review, we focus on the 2021 IOYT assessment using SS3 (Fu et al 2021). The assessment is an ensemble of 96 models, across which several model assumptions are varied to capture a range of structural uncertainties that affect the estimates of IOYT biomass and productivity. Some of the main features and assumptions used in the 'basic' model are listed below.

- The IOYT stock is spatially stratified into 4 regions, two equatorial/tropical regions (R1 and R4) and two temperate regions (R2 and R3). Bi-directional movement is estimated between R1/R2, R1/R4, and R3/R4. There is no movement between R2 and R3 (Fu et al 2021, Figure 3).
- SS3 is fit on a quarterly time-step, where each quarter is treated as a "year".
- Quarterly recruitment is distributed between tropical regions R1 and R4 only with deviations estimated annually.
- A total of 21 fisheries based on region, time-period, fishing gear, and for some fisheries set and vessel type were also differentiated (Fu et al 2021, Table 1).
- SS3 is fit to fishery dependent CPUE indices, fishery catch-at-length observations, and tag recoveries.
- Natural mortality is assumed to be a declining function of age (in quarters), going from around 1.3 at 1 quarter down to around 0.6 at 6 quarters (1.5 years), with a hump starting at 10 quarters (2.5 years), rising to 0.8 at 16 quarters (4 years) and returning to around 0.6 by 22 quarters (5.5 years) (Fu et al 2021, Figure 14).

Additional features and details for the basic model are given in Table 3 of Fu et al (2021). The ensemble grid comprised 96 model options, which were all combinations of the following (Fu et al 2021, Table 5):

- Spatial stratification (2 options): where the alternative was an expansion of R1 (Fu et al 2021, Figure 3),
- Stock-recruit steepness (3 options): 0.7, 0.8, 0.9,
- Tag weighting (2 options): tag likelihood weight of 1 or 0.1,
- Longline catchability in R1b (2 options): A single catchability for the full CPUE series, or two time blocks with 2007 – 2011 removed,
- Growth (2 options): A von Bertalanffy growth models with age-specific K parameters, based on observations from Fonteneau (2008) or Dortel et al (2014),
- Mortality (2 options): Natural mortality-at-age as described above, or at 70% of the basic model,

Section 3: Issues identified for 2021 model

1. **The spatial stratification assumed for the IOYT stock does not appear necessary, nor is it supported by the available data.**

The stratification of the IOYT stock into 4 spatial regions appears to be overly complex, and the assumption of connectivity between all 4 regions may in fact combine two distinct populations with very little exchange of individuals. This is supported by the following observations.

First, there does not appear to be any reason to model a separate region 3. Catch in region 3 is negligible, with only deep-water longline fleets operating there. Moreover, R3 fresh longline catches (LF fleets) are aggregated with R4 removals, and the only modeled fleet is LL3, which accounts for a recent average of around 2 kt, out of an average of around 400 kt across all four regions (i.e., 0.5% of total removals).

Second, there is not enough data to inform the estimation of movement model parameters. Tags that were included in the model were all released in regions 1 and 2 and out of 9916 tags recovered 91% were recovered within region 1. The remaining tags were recovered in region 2 (849 tags) and region 4 (27 tags), and there were zero recoveries in region 3. This data provides information for only 3 out of the 6 movement parameters needed for the assumed spatial structure: R1 -> R2, R2 -> R1, and R1 -> R4; the remaining movement parameters are then free to be influenced by the CPUE and catch-at-length data, making them unreliable at best and are most probably overfit to those data. For example, the movement parameters between R3 and R4 have no basis in the tagging data, and given the low fishing effort in R3, it is probably used by the model to 'store fish' for the fishing activity in R4 (Methot 2019).

Lastly, there does not appear to be any significant movement of IOYT from the western spatial strata (R1/R2) to the eastern strata (R4/R3). This is supported by very small sample of 27 tags recovered in R4, which leads to the estimate of zero longitudinal movement between R1 and R4 in several recent assessments (Fu et al 2018; Urtizbera et al 2019; Fu et al 2021).

Alternative spatial structures appear to have been explored by different groups (Urtizbera et al 2019; Kolody and Jumppanen 2021). It appears that 2-area models were rejected either because estimates of biomass and productivity were not sufficiently different to a 4-area model, or because they did not allow for longitudinal movement. Single area models were similarly rejected because it was harder to satisfy mixing assumptions for the inclusion of the tagging data at the larger spatial scale, which is required under the assumed SS3 tag model structure. However, in some exploratory models the 4-area configuration led to poor model fits, indicated by fishing mortality estimates at the upper bound of F = 2.9, or roughly a 95% harvest rate (Kolody and Jumppanen 2021).

Unnecessarily complex spatial structure will lead to biases in estimates of fishing mortality and stock biomass, and there are additional implications for estimates of biological reference points such as B_{MSY} and F_{MSY}. Indeed, it is not entirely clear how the spatial structure is included in estimates of fishery reference points. It is unclear whether the ability of the model to 'store' biomass within R3 as a spatial refuge is accounted for, or whether a spatial allocation of fishing mortality is assumed in the estimation of reference points.

2. **The SS3 model doesn't appear to fit to the tagging data well at all**

Tag recoveries appear to be on average under-estimated by SS3. The fits to tags recovered by the Purse-Seine fisheries are shown in Figures 21 – 24 of Fu et al (2021). It is sufficient to look at the summary panel (titled Total) to see the under-estimation, where the model predictions (red points) are often well below the data (blue points).

Such chronic under-estimation has implications for the model estimates of mortality. It could be that natural mortality is biased high, meaning that the 'simulated' tags used to predict the data are lost to natural causes before they are captured in fishing effort; there is some support for this in the likelihood profile in Figure B12 (Fu et al 2021), where the tagging data prefers a lower average natural mortality, while the length compositions and CPUE prefer a higher natural mortality rate. On the other hand, it could be that fishing mortality is biased low, which means that fish aren't being removed at a fast enough rate to capture the tags at the quantities observed in the data.

If natural mortality is over-estimated, then this affects the estimates of IOYT productivity via a lower average survival of age-1 fish to recruit first to the fishery, and ultimately to the spawning stock. Lower survival means that a higher average recruitment is necessary to maintain the same size stock, which would make the stock appear more productive than it truly is. On the other hand, if fishing mortality is being under-estimated then the abundance of the stock may be over-estimated, as a larger stock would be necessary to produce the same level of catch at a lower fishing mortality rate, making the stock appear larger than it truly is. In either case, there are significant implications for estimates of current stock status because bias in mortality rates affects biological reference points and the subsequent Kobe overfished/overfishing status.

3. **There appears to be over-weighting of some model configurations in the ensemble, which likely leads to biased estimates of stock status when integrated over all 96 models**

One of the assessment model axes of uncertainty tests natural mortality rates that are 70% of the level used in the basic model, and both natural mortality multipliers are weighted equally. However, when looking at the effect on the total model likelihood (i.e., goodness of fit) it appears that equal weighting may not be appropriate (Fu et al 2021, Figure B12). Indeed, M multiplier of 0.7, which corresponds to the Mlow model option, is approximately 190 likelihood units higher (i.e., less likely) than an M multiplier of 0.9, and about 50 likelihood units higher than the M multiplier of 1.0, which corresponds to the Mbase option. It is unclear, given the lower total likelihood, why the Mbase option does not correspond to the M multiplier of 0.9.

4. Fits to length composition data

There appear to be several issues with how the model fits to length composition data for several fisheries. Time-averaged fits appear to be quite acceptable, but the fits to individual yearly samples do not reflect the quality of the time-averaged fits (compare Figures 1 and 2). Presumably, the plots of time-averaged fits to composition data are weighted by effective sample size calculated based on goodness of fit (i.e., residual variance), which effectively down-weights years with poor fits and obscures model bias at the aggregate level. Poor fits to length composition data imply mis-specified fleet selectivity functions, which can lead to mis-assignment of fishing mortality among age classes, and therefore can have significant implications for estimation of stock productivity and biomass from the assessment, as well as the perception harvest strategy risk in model projections.

Much of the issue may be related to the multinomial likelihood function used for compositional data by SS3. Multinomial likelihood functions values are proportional to the total number of samples in the data, and if composition data are unscaled then the multinomial likelihood function value for length compositions can be very large, effectively dominating the total objective function used to optimise the model (Francis 2011). To mitigate this domination effect, IOYT assessment authors re-weight the total length composition likelihood function by scaling fleet length compositions for each quarter to a sample size of 5, with each length bin having the same proportions as the raw data (Fu et al 2021). However, this is not entirely successful as the total model objective function is still dominated completely by length composition data (Figure B12, Fu et al 2021). Furthermore, the scaling effectively weights each quarter of length data the same for every fleet, regardless of total sampling effort or catch; in some cases this scaling leads to 'smaller' fleets being weighted more heavily than larger fleets. For example, for IOYT the adjusted samples for longline fleets in all areas are weighed more heavily than the purse seine fleets, which have an order of magnitude more catch and several orders of magnitude more length samples (Table 1).

Section 4: Proposed changes to improve IOYT stock assessment model

This section proposes some changes to the IOYT model that we anticipate would improve the statistical properties of the assessment and possibly also projections of population dynamics for harvest strategy and management procedure development.

In general, there are two options. The first, which we recommend, is a major model overhaul as proposed in section 4a below. Such an overhaul would escape restrictive SS3 structure in favour of a custom modeling approach to IOYT. This has the benefit of being a made-for IOYT model, with population dynamics processes and data likelihoods defined to incorporate unique features of the IOYT fishery and address the issues outlined above. Moreover, it would be simple to extend a custom model into an operating model for closed loop simulation and management strategy evaluation (MSE), which is already being explored for IOYT (Kolody and Jumppanen 2021; IOTC 2018). In fact, it is very common for fisheries that are going through MSE processes to use custom built operating models for harvest strategy testing, even among trans-boundary fisheries that use SS3 for stock assessment such as Pacific Halibut (Stewart and Hicks 2022; IPHC 2022) or Pacific Hake (Johnson et al 2021; Jacobsen et al 2021). However, it is also common for operating model and assessment model structure to match, which has an additional benefit of making it easier to detect model mis-specification error between MSE cycles, such as in Southern Bluefin Tuna (CCSBT 2021), British Columbia Sablefish (Cox et al 2022), or Atlantic Halibut in Canada (Johnson et al 2022).

The second option, described in 4b below, would be to continue focusing on incremental changes to the SS3 model to improve how it fits to IOYT data. Given the nature of SS3, changing the code (i.e., as required to implement new likelihood functions) is likely outside of the scope for anyone who is not part of the SS3 development team, and therefore this option is largely limited to changing spatial stock structure, or adjusting the way that data is prepared for SS3. Data preparation could be adjusted by removing some sources of data that are difficult to fit under the current structure (e.g., tagging data), or changing the way that data are scaled/standardised before being input to SS3.

4a. Major model overhaul

Given the mismatch between the information content of IOYT data and the model structure assumed for the SS3 assessment of IOYT, it might be most straightforward to custom design a new IOYT stock assessment model from the ground up. The benefit of using a custom model for IOYT is that all elements of the model are specifically defined for IOYT data, so that there are no trade-offs between IOYT specific problems and defining a model for general use cases, as has been necessary with SS3 and previously with MULTIFAN-CL.

If the IOYT model were to be overhauled, we would advise the following steps. A proposed budget for these steps is included as Appendix A.

1. **Undergo an initial model planning stage including interviews with IOTC scientists.** There is a significant history of stock assessment work for the IOYT stock conducted by scientists who are very familiar with both the scientific limits of the IOYT data and other issues related to the trans-boundary nature of IOYT. We propose a series of interviews regarding historical IOYT stock assessments, the data collection process, and fishery management needs such as future management procedure evaluation and/or rebuilding planning requirements, with the aim to elicit features that would be included in the ideal IOYT stock assessment model, and to understand more fully the challenges associated with modeling IOYT.

2. **Seek funding from external sources for model development.** There may be significant resource requirements for developing a new stock assessment model for IOYT, depending on the scope. As such, we recommend developing funding applications for model development in partnership with stakeholders, including industry groups, NGOs, and government institutions. For example, the tagging effort for the RTTP-IO and several of the analyses of that data were enabled via the 9th European Development Fund (EDF; e.g. Hillary and Eveson 2015). The EDF is specifically designed for funding projects in countries outside of the European Union, and the current fund (12^{th}) spans 2021 – 2027 as part of the European Union Multiannual Funding Framework. Other sources of funding may also be available from other significant national interests (e.g., Australia, Japan, etc).

3. **Reduce the spatial complexity in the IOYT model.** There are a few different ways this could be achieved. The simplest would be to collapse the IOYT to either a single area model that aggregates R1 – R4, or to two independent populations split between R1/R2 and R3/R4. In both cases, fleets could continue to represent different spatial area and gear combinations via an areas-as-fleets approach like Pacific Halibut and Atlantic Halibut stock assessments (Stewart and Hicks 2022; Johnson et al. 2022), as well as others. The reduction in spatial complexity would remove the need to estimate movement model parameters and tagging data could be used for estimating mortality rates only.

4. **Adjust the tagging model to incorporate non-mixed tags for better estimation of mortality.** Standard tagging models typically assume that tagged fish are equally vulnerable to fishing given their size or age. If a spatial region is very large, as some IOYT regions are, then this assumption takes a long time to satisfy and could result in heavily biased fishing mortality rates (Kolody and Hoyle 2015). One option is to freely estimate 'unmixed' exploitation rates so that all tag recoveries can be included (Hillary and Eveson 2015). Another option, which may pair well with the spatio-temporal treatment of input data outlined below, would be to model individual fish movement as an advection-diffusion process, which could be fit to the tag release/recovery sites and the time-at-liberty (Sibert et al 1999; Senina et al 2020). Finally, the current ADMB implementation treats all parameters as fixed effects, but a treatment of individual fish as random effects would be better suited to allow individual fish to deviate from 'expected' population movement behaviour, especially before full mixing has occurred (Fu 2022).

5. **Adjust tagging model so that it is length based, to avoid biases associated with assigning an age to tag release groups.** The current tagging model in SS3 requires an age to be assigned to fish in each tag group, so that they can be tracked through time. Since there is very little age information for IOYT, the tag group age is assigned based on length, via the mean length-at-age relationship (Fu et al 2021). A common problem with this approach is that there is significant overlap in length distributions among age classes given variability in length-at-age, which leads to errors in fishery selectivity and fishing mortality on tagged fish. A length-based method that takes uncertainty in the length-at-age relationship into account via an age-length key (Fu 2022), or a fully length-based

model replacing age-based dynamics with size-class transition matrices could improve performance (Hillary and Eveson 2015).

6. **Use spatio-temporal standardisation for length data, CPUE indices, and possibly tagging data.** It is becoming less common for fisheries over large spatial scales to be assessed in discrete boxes (i.e., regions), and more common for data collected over the wide area to be standardised with respect to continuous space and time using Gaussian Markov Random Fields (GMRFs; Thorson, Maunder, and Punt 2020). Examples in high-profile fisheries include Pacific Halibut and Southern Bluefin Tuna (Stewart and Hicks 2021; CCSBT 2021). Spatio-temporal standardisation via GMRFs is especially useful for fishery dependent data, which is not design-based and generally collected during non-random fishing effort (Thorson, Maunder, and Punt 2020). Such continuous spatio-temporal methods may also prove useful for a continuous treatment of tagging data, such as in the advection-diffusion model suggested above.

7. **Replace multinomial likelihood for length compositions with logistic-normal likelihood.** As outlined above, and observed for IOYT, there are several issues associated with using multinomial likelihood functions for compositional data (Francis 2014). A better approach is to use a logistic-normal likelihood function (Schnute and Haigh 2007; Francis 2014), for which the likelihood function is a sum of squared logit-residuals. As a result, the function is self-weighting by the variance nuisance parameter, so there is no need to down-weight sample sizes as done for IOYT, and the likelihood function value is often on a similar scale to likelihoods for other sources of data. Finally, relative sizes of individual quarterly samples can be included as annual weights on the residual sum of squares to represent changes in sampling effort.

4b. Minor changes within SS3 structure (in lieu of a major model overhaul)

While we recommend a custom-designed assessment model for any stock of this importance, we also note the following options that may improve the statistical properties SS3 fits to IOYT data:

1. **Remove movement model and tagging data.** IOYT could be fit as a single stock in SS3 with an implicit assumption of uniform distribution across the Indian Ocean, or as 2 independent populations with no movement between areas. This is like the reduction in spatial complexity suggested for the model overhaul but fits within the SS3 structure already being used. This change was originally suggested by Methot (2019). The 2-area structure has the benefit of indirectly reflecting the tagging data (i.e., no longitudinal movement), while acknowledging the limits of that data's direct utility within the SS3 model.

2. **Remove movement model, but keep tagging data to estimate mortality within a combined R1/R2.** Similar to modification 1 but incorporates the tagging data more explicitly. This approach would estimate a new mixing period for a combined R1/R2

region, like what is used currently for mixing. This way, tags that are at liberty for longer than the mixing period can still be used in SS3 to help estimate natural and fishing mortality rates in R1/R2 without a major change to the SS3 tag data likelihood function. Then, the R1/R2 natural mortality rates may be used as informative priors for natural mortality rates in R3/R4.

3. **Adjust sample size scaling for length composition data.** The current approach to scaling all quarterly length samples to an adjusted sample size of 5 likely leads to model bias in selectivity, recruitments, and fishing mortality. We recommend avoiding this equal weighting scheme, and replacing adjusted sample sizes or some other weighting, such as by catch to reflect the relative influence of each fleet, or some iterative reweighting scheme that accounts for effective sample sizes based on residual variance and/or the raw sample sizes but manages to reduce domination of the objective function by compositional data. This way, the contributions of individual fisheries remain scaled to their relative influence (as measured by catch and/or sampling effort).

4. **Apply a spatio-temporal standardisation for input data.** There are legitimate reasons to separate tuna by temperate and tropical regions, but there is limited support in the data for doing so. Applying a spatio-temporal standardisation method such as VAST to length composition data could help reduce the bias in length compositions while simultaneously allowing for a simpler model structure. This would be similar to the approach described in the major model overhaul.

References:

CCSBT (2021). Report of the Twenty Sixth Meeting of the Scientific Committee.

Cox, S. P., Kronlund, A. R., Lacko, L., and Jones, M. K. (2022). A revised operating model for Sablefish (*Anoplopoma fimbria*) in British Columbia, Canada (In Press). Can. Sci. Adv. Sec. Res. Doc.

Francis, R. C. (2011). Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 68(6):1124–1138.

Francis, R. C. (2014). Replacing the multinomial in stock assessment models: A first step. Fisheries Research, 151:70–84.

Fu, D. (2022). A length-based extension to the Brownie-Peterson model and its performance. Fisheries Research, 249:106248.

Fu, D., Langley, A., Merino, G., Ijurco, A.U. 2018. Preliminary Indian Ocean Yellowfin Tuna Stock Assessment 1950-2017 (Stock Synthesis). IOTC–2018–WPTT20–33.

Fu, D., Langley, A., Merino, G., Ijurco, A.U. 2018b. Indian Ocean Yellowfin Tuna SS3 Model Projections. IOTC–2018–SC21–16.

Fu, D., Urtizberea, A., Cardinale, M., Methot, R. D., Hoyle, S., and Merino, G. (2021). Preliminary Indian ocean yellowfin tuna stock assessment 1950-2020 (Stock Synthesis). IOTC-WPTT23, 12.

Hillary, R. M. and Eveson, J. P. (2015). Length-based brownie mark-recapture models: Derivation and application to Indian ocean skipjack tuna. Fisheries Research, 163:141–151.

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

IOTC 2018. Report of the 21st Session of the IOTC Scientific Committee. Seychelles, 3 – 7 December 2018. IOTC–2018–SC21–R. 250 pp.

IOTC 2020. Report of the 23rd Session of the IOTC Scientific Committee. Seychelles, 7 – 11 December 2020. IOTC–2020–SC23–R[E]: 211pp.

IPHC (2022). Technical details of the IPHC MSE framework. Technical report, IPHC.

Jacobsen, N. S., Marshall, K. N., Berger, A. M., Grandin, C. J., and Taylor, I. G. (2021). Management strategy evaluation of Pacific Hake: Exploring the robustness of the current harvest policy to spatial stock structure, shifts in fishery selectivity, and climate-driven distribution shifts.

Johnson, K.F., A.M. Edwards, A.M. Berger and C.J. Grandin. 2021. Status of the Pacifc Hake (whiting) stock in U.S. and Canadian waters in 2021. Prepared by the Joint Technical Committee of the U.S. and Canada Pacifc Hake/Whiting Agreement, National Marine Fisheries Service and Fisheries and Oceans Canada. 269 p.

Johnson, S. D. N., Hubley, B., Cox, S. P., and den Heyer, C. E. (2022). Framework Assessment of Atlantic Halibut on the Scotian Shelf and Southern Grand Banks (NAFO Divs 3NOPs4VWX5Zc) Model Update (In Press). Can. Sci. Adv. Sec. Res. Doc.

Kolody, D, Jumppanen, P. 2021. Indian Ocean Yellowfin Tuna Management Procedure Evaluation Update March 2021. Working Paper prepared for the Management Strategy Evaluation Task Force of the Indian Ocean Tuna Commission Working Party on Methods Meeting, March 2021. IOTC-2021-WPM12(MSE)-03

Kolody, D. and Hoyle, S. (2015). Evaluation of tag mixing assumptions in western Pacific ocean Skipjack tuna stock assessment models. Fisheries Research, 163:127–140.

Langley, A. 2015. Stock assessment of yellowfin tuna in the Indian Ocean using Stock Synthesis. IOTC- 2012-WPTT-17-30.

Langley, A. 2016. An update of the 2015 Indian Ocean Yellowfin Tuna stock assessment for 2016. IOTC-2016-WPTT18-27.

Langley, A., Hampton, J., Herrera, M., Million, J. 2008. Preliminary stock assessment of yellowfin tuna in the Indian Ocean using MULTIFAN-CL. IOTC-2008-WPTT-10.

Langley, A., Herrera, M., Hallier, J.P., Million, J. 2009. Stock assessment of yellowfin tuna in the Indian Ocean using MULTIFAN-CL. IOTC-2009-WPTT-11.

Langley, A., Herrera, M., Million, J. 2010. Stock assessment of yellowfin tuna in the Indian Ocean using MULTIFAN-CL. IOTC-2010-WPTT-12.

Langley, A., Herrera, M., Million, J. 2011. Stock assessment of yellowfin tuna in the Indian Ocean using MULTIFAN-CL. IOTC-2011-WPTT-13.

Langley, A., Herrera, M., Million, J. 2012. Stock assessment of yellowfin tuna in the Indian Ocean using MULTIFAN-CL. IOTC-2012-WPTT-14-38.

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

Methot, R. D. and Wetzel, C. R. 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142:86–99.

Methot, R. D., Wetzel, C. R., Taylor, I. G., and Doering, K. (2020). Stock synthesis user manual: Version 3.30. 15.

Schnute, J. T. and Haigh, R. (2007). Compositional analysis of catch curve data, with an application to *Sebastes maliger*. ICES Journal of Marine Science: Journal du Conseil, 64(2):218–233.

Senina, I., Lehodey, P., Sibert, J., and Hampton, J. (2020). Integrating tagging and fisheries data into a spatial population dynamics model to improve its predictive skills. Canadian Journal of Fisheries and Aquatic Sciences, 77(3):576–593.

Stewart, I., and Hicks, A. 2022. Assessment of the Pacific halibut (*Hippoglossus stenolepis*) stock at the end of 2021. IPHC-2022-SA-01. 36 p.

Thorson, J. T., Maunder, M. N., and Punt, E. (2020). The development of spatio-temporal models of fishery catch-per-unit-effort data to derive indices of relative abundance.

Urtizberea, A., Fu, D., Merino, Gorka., Methot, R., Cardinale, M., Winker, H., Walter, J., Murua, H. 2019. Preliminary Assessment of Indian Ocean Yellowfin Tuna 1950-2018 (Stock Synthesis, V3.30). IOTC-2019-WPTT21-50.

Tables

Table 1: A summary of the catch and catch-at-length data for each of the 21 fleets in the IOYT stock assessment, generated from the data.ss file used for the assessment up to 2020 (Fu et al 2021).

Fleet Number	Name	No. of Quarters	Mean Catch (t)	Total length samples	Adjusted length samples	Mean Observed Length
1	GI1a	284	5901.03	115953	520	84.08
2	HD1a	284	5249.75	34600	250	113.86
3	LL1a	248	1698.03	7716	75	119.73
4	OT1a	284	21.56	24794	120	49.4
5	BB1b	284	2346.3	308116	675	48.66
6	PSFS1b	157	13698.23	32335175	760	104.64
7	LL1b	268	4533.1	183856.4	1005	123.34
8	PSLS1b	156	13394.97	170425447	765	55.21
9	TR1b	284	577.74	0	0	0
10	LL2	268	1541.72	131042	945	125.06
11	LL3	273	424.43	224413	920	124.75
12	GI4	284	1405.28	452003	195	67.76
13	LL4	273	2028.61	189532	945	121.29
14	OT4	284	878.65	8300	100	42.12
15	TR4	284	1052.45	34064	115	47.59
16	PSFS2	99	1211.03	3283561	425	82.22
17	PSLS2	120	1586.11	14398028	520	60.3
18	TR2	284	391.32	0	0	0
19	PSFS4	118	435.94	642546	145	100.07
20	PSLS4	153	304.65	1486702	250	57.8
21	LF4	284	6686.15	111133	180	126.18

Figures

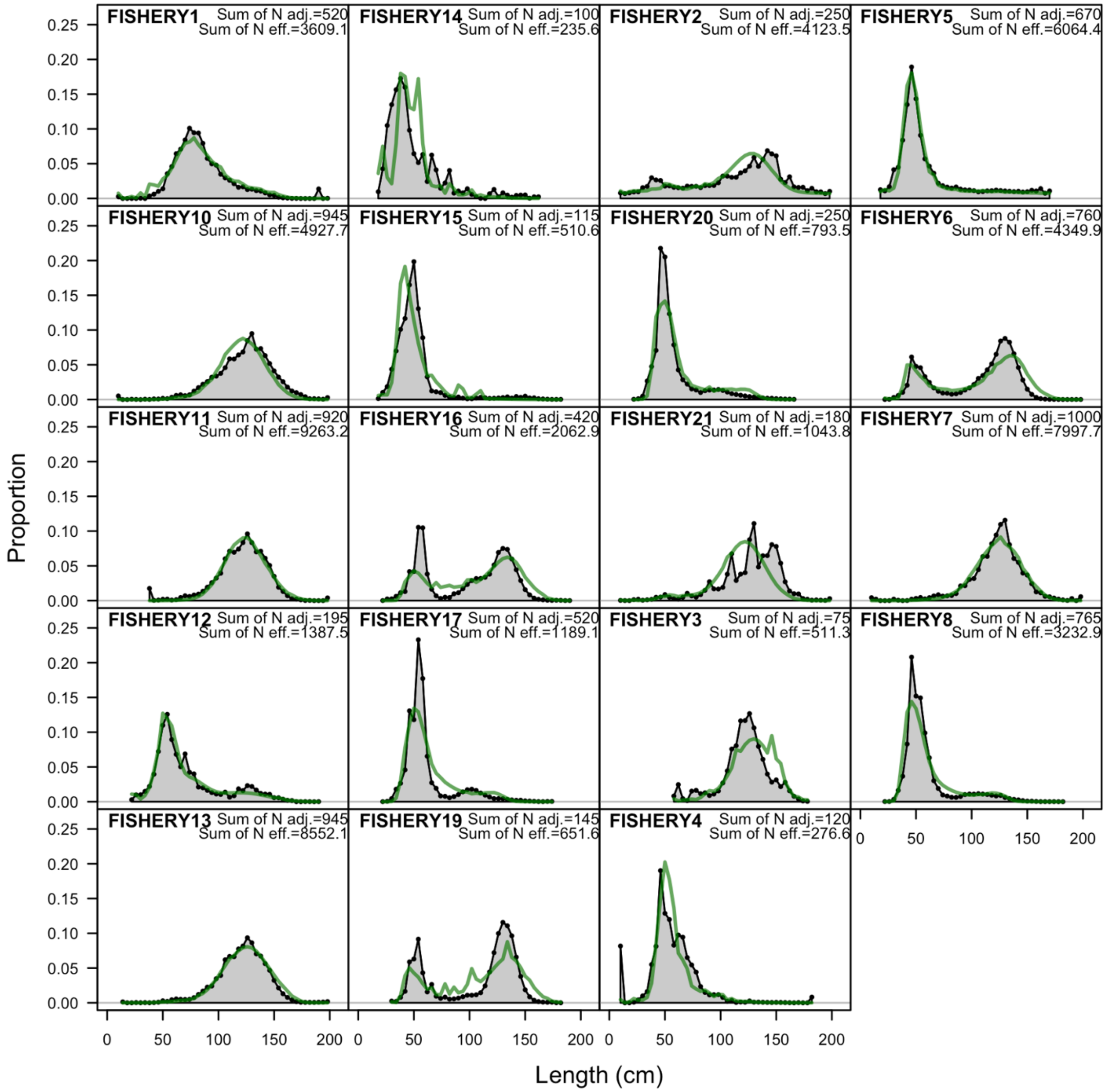

Figure 1: Time-averaged fits to length compositions for each fleet. Grey polygons are time-averaged data, while green lines are the model fits weighted by effective sample size.

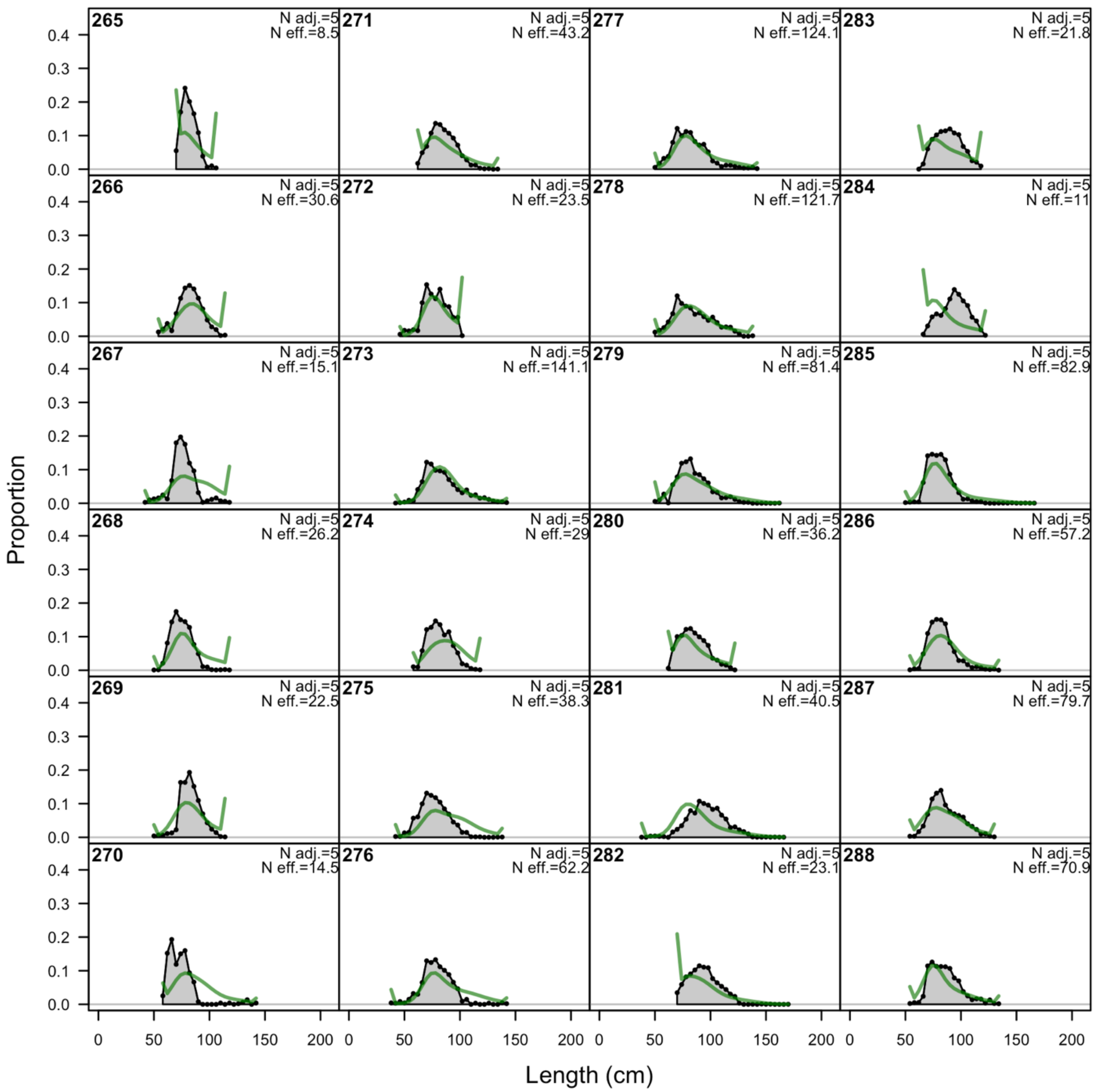

Figure 2: Fits to length composition data for Fishery 1 (GI1a) for individual quarters 265 – 288 (top right corner of each pane). Grey polygons show the proportion-at-age data, while the green line shows individual fits to those data. The numbers in the top right corner show the adjusted input sample size of 5, and the model estimates of effective sample size based on residual variance.

Appendix A: Budget to develop an alternative Bayesian IOYT stock assessment model ensemble. Estimated work days (in "[]") are approximated toward the upper end.

1. Consultation
    a. **[10 days]** Interview IOTC scientists and stakeholders: elicit desirable model features and must-haves in terms of included data sets, as well as a 'data guillotine' date
    b. **[5 days]** Funding application: prepare grant proposals for submission to external bodies for model development if desired
2. Model Design:
    a. **[10 days]** Acquire up-to-date IOYT stock assessment data (before the limit defined above) and conduct exploratory analyses
    b. **[10 days]** Develop mathematical definition of a new Bayesian IOYT stock assessment model including the following features
        i. A simpler spatial structure with two stocks split into current regions R1/R2 and R3/R4, with no movement between (DFO 20XX, attached to this review, **do not cite or circulate**).
        ii. Quarterly time-steps to acknowledge intra-annual seasonal effects as well as inter-annual variation, with quarterly recruitment to approximate continuous spawning of tropical tunas
        iii. Areas-as-fleets structure to acknowledge differences in availability of fish between equatorial and temperate regions, with possible seasonal and annual variation in fleet selectivity (Johnson and Cox 2022, attached to this review, **do no cite or circulate**)
        iv. A length-based Brownie-Peterson tagging model to estimate mortality only (no movement), with structure that accounts for non-mixed tags to increase sample sizes (e.g., Fu 2022 and/or Hillary and Eveson 2015)
        v. A self-weighting logistic-normal likelihood for length-composition data, potentially with lag-1 correlation in residuals for neighbouring length bins (Johnson and Cox 2022; Francis 2014)
3. Model coding and implementation:
    a. **[45 days]** Develop a new, or apply an existing (e.g., VAST), spatiotemporal CPUE standardisation that accounts for large scale and unique spatiotemporal dynamics of the fishery
    b. **[45 days]** Implement the final model design in Template Model Builder
4. Model testing:
    a. **[15 days]** Exploratory sensitivity analyses: Explore the sensitivity of new IOYT model maximum posterior density estimates of biomass and productivity to alternative assumptions
    b. **[15 days]** Code validation: code unit tests that externally validate model calculations of population dynamics processes as well as objective function components.
    c. **[20 days]** Simulation-estimation self-test: Evaluate new IOYT model bias and precision by fitting the new IOYT model to simulated pseudo-data sets and

recording relative estimation errors, thereby identifying sources of model bias and corrections if possible.

5. IOYT stock status estimation:
    a. **[10 days]** Ensemble grid definition and aggregation method: identify, in collaboration with IOTC stakeholders, key model sensitivities from 4a to be included as part of an ensemble grid for determination of IOYT biomass and productivity, and a method for combining those estimates across the grid (e.g., weighting or sampling scheme).
    b. **[20 days]** Model fitting and posterior generation: Fit the ensemble grid to IOYT data and tune model setting to produce convergent optimisations of maximum posterior density estimates, and fully mixed chains of samples from model parameter posterior distributions.
    c. **[2 days]** Determining stock status: Estimate reference points for each model in the ensemble grid, then use the aggregation method defined above to combine individual model posterior distributions of stock status.
6. Documentation and dissemination
    a. **[30 days]** Model documentation: In depth model description and instructions for other analysts to apply IOYT assessment model and estimate model parameters, generate posteriors, and combine ensemble model estimates
    b. **[15 days]** Working paper preparation: Writing a stock assessment report as a working paper for the IOTC secretariat
    c. **[10 days]** Response to reviews: Incorporating feedback from reviewer suggestions into the new IOYT model and its documentation.

Total of ***262 days***.

Please note: meetings and travel are not included in this budget given the variable nature of travel costs, but generally are a combination of expenses for transportation from Vancouver, Canada, accommodation, meals, and day-rates for travel and meeting participation.

Landmark consulting daily rates as of September 2022.

Role	Daily rate (CDN)	Responsibilities
Principal	$1,450	Strategy and model design for assessments and MSE. Consultation, communication, and facilitation
Senior scientist	$1,250	Model design, coding, testing, communications
Junior scientist	$1,000	Data modelling, survey design, statistical modelling
Analyst	$800	Data preparation, exploratory analyses, statistical modelling

Appendix References

DFO. 20XX. Evaluation of potential rebuilding strategies for Pacific Herring (*Clupea pallasii*) in Haida Gwaii. DFO Can. Sci. Advis. Sec. Sci. Resp. 20XX/999.

Johnson, S. D. N., Cox, S. P. (In Review). A modeling framework for stock assessment and harvest strategy evaluation for the NAFO 0 + 1 Greenland Halibut (*Reinhardtius hippoglossoides*) fishery. DFO Can. Sci. Advis. Sec. Res. Doc. 2022/XXX. iv + 61 p.