Model-based estimates of Length composition time series for Indian Ocean yellowfin tuna fisheries

DAN FU[1]

02 OCTOBER 2021

SUMMARY

This report presents model-based estimates of length compositions using a Bayesian Dirichlet-Multinomial (DM) model. The DM model was applied to the commercial catch samples to derive time series of length distributions for the Indian Ocean yellowfin tuna for the Longline and handline fisheries. The model incorporated spatial variability in the population length distributions at 5x5 longitude latitude (the level of reporting) while accounting for sampling errors amongst fleet and seasonal strata. The results suggested the DM model can be a potentially useful method to provide standardized length composition input for the IOTC stock assessments. However further work is required to better disaggregate and quantify various sources of uncertainty in the commercial length samples.

1. INTRODUCTION

Length composition data is an integral component in the stock assessments for Indian Ocean tuna and tuna-like species and provide critical information on fishing mortality, recruitment, and abundance. Bias in the length composition data can substantially affect the estimation of key management quantities. To date, the length composition data are compiled by IOTC Secretariat for inclusion into the species assessments and so far, it has involved pulling together the raw length samples from all sampling/reporting units for a fishery group (referred to as nominal LF). However, it is recognized that simple aggregation of nominal LF may introduce bias in the length composition if the spatial, seasonal, and fleet variabilities in the length samples are not appropriately accounted for (Hoyle et al. 2021). The design approach (e.g., stratified sampling) to estimate the length composition that takes into account weighting of individual sampling stratum is more widely used in national fishery applications (Stewart & Hamel 2014, Thorson 2014). However, for most t-RFMOs, the sampling design, coverage, data collection and reporting vary greatly among fleets, which makes analyses of the length compositions using a design approach impractical (Medley et al. 2021).

Model-based methods provide alternative means for standardizing length composition to estimate stock-level length distribution while accounting for appropriate weighting for the underlying data (Thorson 2014). The application of Model-based standardization of length compositions was rare but it has become more common recently where it is used in tuna assessments in the east Pacific Ocean (Minte at al. 2020). Thorson (2014) proposed two model-based methods for standardizing compositional data: a normal approximation and a Dirichlet-Multinomial (DM). The simulation by Thorson (2014) indicated the Dirichlet-Multinomial model was biased compared to the normal

---

[1] IOTC Secretariat, Dan.Fu@fao.org;

approximation. However, Neubauer et al. (2021) suggested the bias was likely to be due to that the DM model used a single dispersion parameter and not structured to adequately standardize spatial variability, and hence proposed a revised DM formulation that could attributed overdispersion to known sources of spatial variability. Harrison (2019) also showed Dirichlet-multinomial modelling performs better than alternative models in standardizing ecological count data. We explored the application of DM to the IOTC yellowfin length composition data, following a similar framework as Neubauer et al. (2021) and Harrison et al. (2019).

2. METHOD

The data used in the analysis were that provided to WPTT22 in 2020 (IOTC-2020-WPTT22-DATA11-SF_YFT_FL.xlsx). It consists of raw length samples of yellowfin tuna by gear, country, spatial grid (mostly 5 by 5 longitude latitude), year, and month, conforming to the IOTC reporting requirement (IOTC Secretariat 2014). The data were then aggregated by fishery group (see Fu et al. 2018 for the fisheries defined for yellowfin tuna), quarter, spatial grid, and fleet, assuming to represent a sampling unit for the analysis. The model-based approach was tentatively applied to the longline and handline fisheries to derive the respective time series of length distributions. The longline fishery contributes to the longest time series of length data for the yellowfin assessment with a relatively broad spatial and temporal coverage, whereas the quality of the length data from the handline were much poorer (Fu et al. 2018).

The model aims to estimate stock level length distribution $\theta_{f,y}$, which represents the true proportion at length for fishery *f* in year *y*. We assume the population varies spatially (i.e. 5x5) and there is a true length distribution $\theta_{f,y,s}$ in each spatial stratum. $\theta_{f,y,s}$ is assumed distributed according a a Dirichlet distribution with parameters $\alpha_s\theta_{f,y}$, representing random realization of the true stock proportion at length, where α_s determines the level of variability amongst spatial stratum and is assumed to be lognormally distributed. The i^{th} observed length frequency $\pi_{f,y,s,i}$ from sampling stratum s, follows a multinomial distribution with parameter $\theta_{f,y,s}$ where $n_{f,y,s,i}$ fish were sampled (capped at 200).

$$\begin{gathered}
\pi_{f,y,s,i} \sim \mathrm{MN}\left(n_{f,y,s,i}, \theta_{f,y,s}\right) \\
\theta_{f,y,s} \sim \mathrm{D}\left(\alpha_s \theta_{f,y}\right) \\
\alpha_s \sim logN(\alpha, \sigma_s) \\
\theta_{f,y} \sim \mathrm{D}\ (1/L)
\end{gathered}$$

Where MN is the multinomial distribution, D is the Dirichlet distribution, and logN is the lognormal distribution. A non-informative prior was used for $\theta_{f,y}$ (Dirichlet with parameter 1/L, where L is number of 4-cm length bin). σ_s is fixed at 5 as of Neubauer (2021). The DM model was implemented in Stan (Stan Development Team 2018) and the estimation used Markov Chain Monte Carlo (MCMC) algorithms. For each Fishery and year, MCMC were conducted to generate 1000 posterior samples of length distributions (first 500 samples were discarded).

3. RESULTS

Estimates of length frequency time series were shown in Figure 1 for the longline fishery (1960 – 2019), and Figure 2 for the handline fishery (2003 – 2019). The estimates of length distributions for the handline fishery have larger posterior intervals than those for the longline, apparently a result of the poorer sampling quality and larger variability for the handline fishery. For the longline fishery, the uncertainly appeared to have increased since around 2002, reflecting the patchier samples collected during recent years (Hoyle et al 2021). For both fisheries, the model estimates are remarkably similar to the nominal LF with few exceptions (e.g., HD 2003). In some instances, model-based estimates appeared have removed some sporadic large peaks in the nominal LF.

4. DISCUSSIONS

The analysis applied the DM model to estimate length distributions for Indian Ocean yellowfin tuna, and demonstrated the possibility of developing model-based length compositions that can be used in the stock assessments of IOTC species. However, there are a number of possible improvements to be considered in future iterations of the analysis.

Firstly, the current analysis has aggregated the length samples at a coarser level (i.e. Fishery, Quarter) than the original reporting units (Gear, month). The means the model may have underestimate the sampling variability that were exhibited in the raw samples. Further analysis should examine the choice of sampling units that can adequately capture sampling errors while balancing computational efficiency. Secondly, the model has only considered spatial variability in the true sub-population proportion at length, as they are generally considered to constitute the main source of variability for proportion at length (Hoyle et al. 2021, Medley et al. 2021). Seasonal and fleet variability are not explicitly modelled, although they have been included as part of the sampling errors. The model structure can be revised to better quantify different source of variability in the length samples. Thirdly, the current analysis is conducted for each year independently (a separate model is fit to each year's data). It's worth exploring a model where time series are fitted together to increase the precision of estimates (e.g., spatial variability may be constant or linked overtime).

5. REFERENCES

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

Hoyle, S.D. Chang, S.T., Fu, Dan., Geehan, J., Itoh, T., Lee, S.I. Lucas, J., Matsumoto, T., Yeh, Y.Y., Wu, R.F. 2021. Draft review of size data from Indian Ocean longline fleets, and its utility for stock assessment. IOTC-2021-WPTT23(DP)-08

Harrison, J.G., Calder, W.J., Shastry, V., Buerkle, C.A. 2019. Dirichlet-multinomial modelling outperforms alternatives for analysis of microbiome and other ecological count data. Molecular Ecology Resources doi: 10.1111/1755-0998.13128

IOTC Secretariat (2014). Guidelines for the reporting of Fisheries Statistics to the IOTC. IOTC Secretariat, Mahé, Seychelles, January 2014. 70pp.

Medley P.A.H., Defaux V., Huntington T. 2021. Exploratory analysis of tropical tuna longline selectivity and its implications for stock assessment. OTC-2021-WPTT23(DP)-11_Rev1.

Minte-Vera, C., Maunder, M.N., Xu, H., Valero, J.L. Lennert-Cody, C.E., Aires-da-Silva, A. 2020. Yellowfin Tuna in the Eastern Pacific Ocean, 2019: Benchmark Assessment. SAC-11-07 REV.

Neubauer, P., Tremblay-Boyer, L. 2019. Input data for the 2018 stock assessment of pāua (Haliotis iris) for PAU 5D. New Zealand Fisheries Assessment Report 2019/38

Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.17.3. Retrieve from http://mc-stan.org/.

Stewart, I.J., Hamel, O.S. (2014). Bootstrapping of sample sizes for length- or age-composition data used in stock assessments. Canadian Journal of Fisheries and Aquatic Sciences 71 (4): 581–588. doi:10.1139/cjfas-2013-0289.

Thorson, J.T. (2014). Standardizing compositional data for stock assessment. ICES Journal of Marine Science 71 (5): 1117–1128. doi:10.1093/icesjms/fst224.

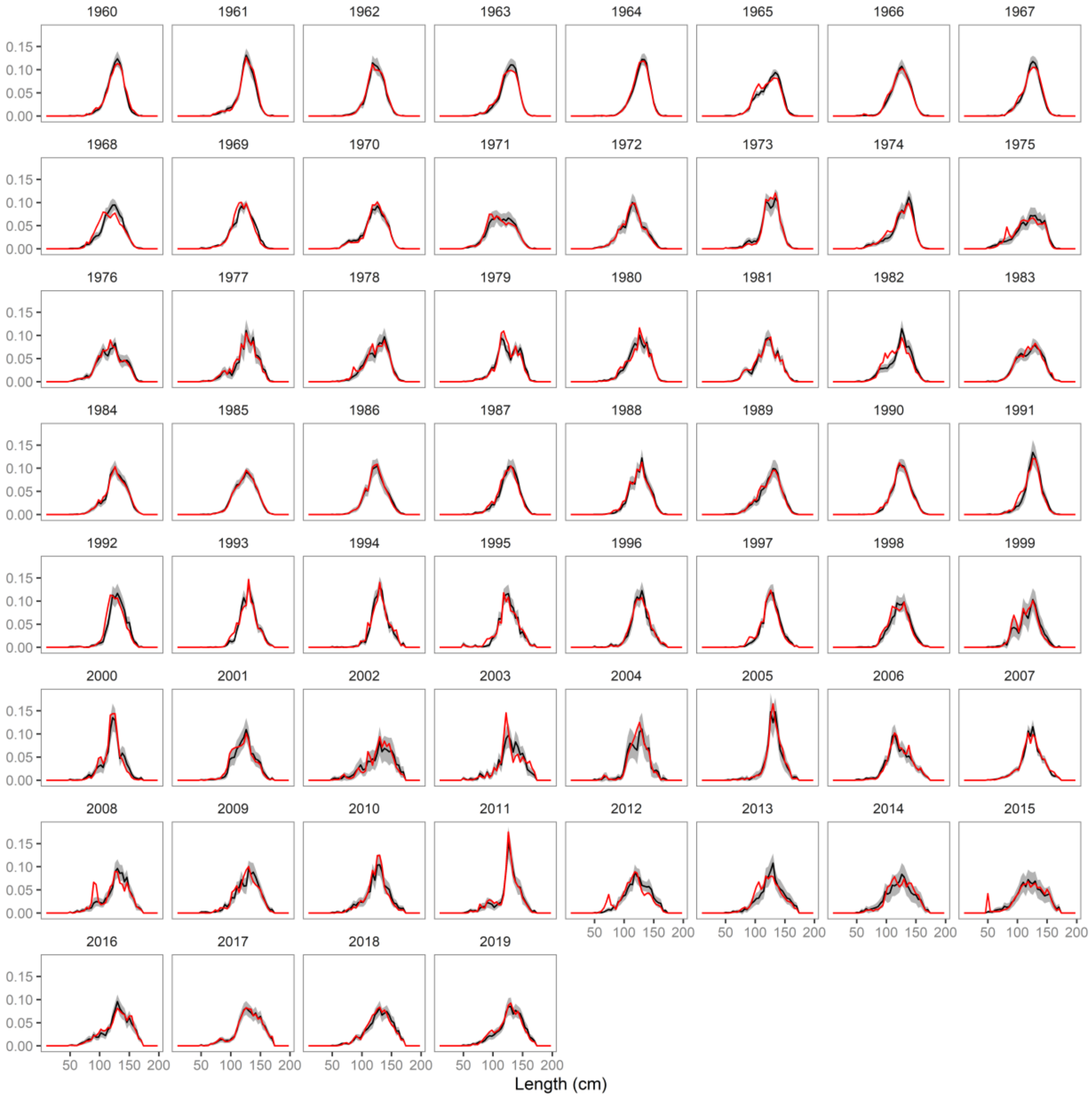

Figure 1: Model based estimates of annual length composition for the longline fishery 1960-2019 (black is the median, gray is the 95% credible interval from the MCMC), overlaid with the nomial LF (red) line

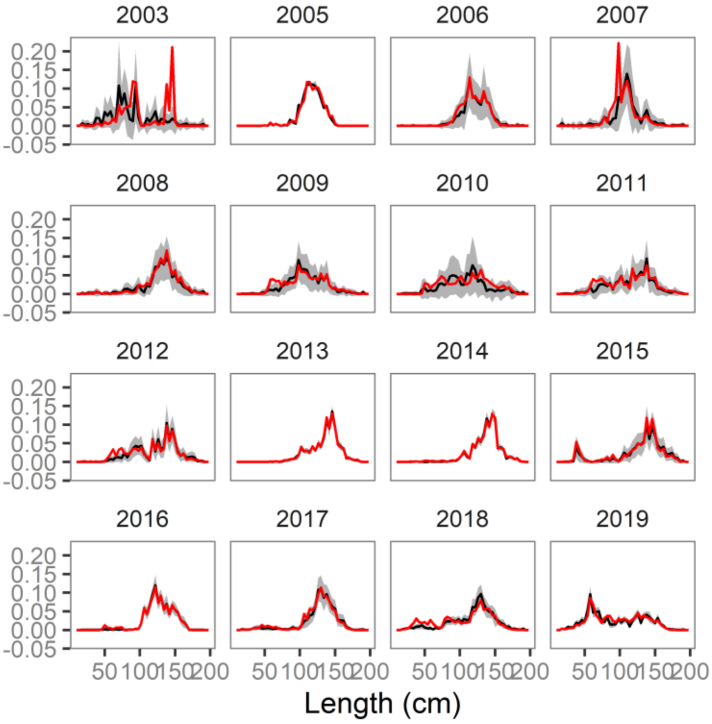

Figure 2: Model based estimates of annual length composition for the handline fishery 2003-2019 (black is the median, gray is the 95% credible interval from the MCMC), overlaid with the nomial LF (red) line