Exploratory analysis of tropical tuna longline selectivity and its implications for stock assessment

Version: 7 May 2021 (v. 1.3)

A working paper for the IOTC 23rd working party on tropical tuna (WPTT23):
Data preparatory meeting (10/05/2021 to 14/05/2021)

Authors: Medley P.A.H., Defaux V., and Huntington T.

Poseidon Aquatic Resource Management, fisheries consultants

Executive Summary

Different approaches to modelling fishing selectivity can have a profound impact on stock assessments. This exploratory work considers how this selectivity might be better modelled. Interest in this subject was inspired by the length composition changes observed in longlines over 2003/4 and how these changes may be best accounted for. Data used were the longline length frequencies for yellowfin and bigeye tunas within the tropical region (+/- 15 degrees latitude) covering the period from 1952 to 2018 and Regional Observer Program data. All the data are available to the public on IOTC's website. Results showed that there is a clear break in length frequency patterns around 2003/4, most apparent for yellowfin although a similar pattern occurs for bigeye. This pattern is solely found in the Taiwanese fleet length frequencies that makes up most of the data. It was originally hypothesized that the change in length compositions in 2003/4 could be due to change in discard practices. However, this appears to be inconsistent with the available data which does not show length frequency truncation that might result from size-specific discarding. The data are more consistent with some other cause such as a change in vessel operations which have affected the underlying selectivity. What this change might have been and how these changes may impact yellowfin and bigeye stock assessments remain unclear to date. Ways to proceed with this analysis that may provide further insight are discussed.

Introduction

This exploratory work considers how selectivity might be better modelled. It is useful to consider how selectivity is being modelled because it can have a profound impact on stock assessments. The original motivation was concerned with length composition changes observed in longlines through 2003/4 in the tropical region of the IOTC convention area and how these changes may be best accounted for, particularly with respect to the longline abundance index which is important source of information for the stock assessment. The potential relative misreporting of small size fish by industrial longlines since 2004 (Geehan and Hoyle 2013, Sharma 2018, Urtizberea et al., 2019, IOTC-2019-WPTT21-50) may have led to misspecification of selectivity which could particularly had a large effect on the joint longline CPUE used in the yellowfin stock assessment (Naunet, 2020). The associated with these uncertainties has been excluded from the longline abundance index (Hoyle et al. 2019). This paper evaluates possible causes of changes in the apparent size compositions reported by longlines during this period and explores alternative approaches to addressing this issue.

This study was funded by the Spanish Producer Organisation OPAGAC.

Data and Methods

The scope of data used were the longline length frequencies for yellowfin and bigeye within the tropical region (+/- 15 degrees latitude). These are the main large tropical tunas caught by longline. The length frequency data consisted of 2 cm precision length bins reported by month and quadrant area coded as 1°x1°, 5°x5° and 10°x20° grids (Figure 1). These data were made available by IOTC contracting parties to the IOTC secretariat for input into stock assessments, and have been pre-processed to remove errors and standardised various size measures to the 2 cm length bin measurement (including conversion of processed weight to length). For tropical tunas, the data cover the period from 1952 to 2018, but coverage of fishing activity varies. The reported length frequencies for longline are divided on the basis of categorical variables, relevant ones consisting of year/month, grid latitude/longitude location, fleet, and gear type. All data used in this analysis are publicly available and can be downloaded from the IOTC website:

'IOTC-2020-WPTT22_prep-DATA11-SF_YFT_FL.xlsx', and 'IOTC-2020-WPTT22_prep-DATA11-SF_BET_FL.xlsx'.

Because the initial motivation was to consider the apparent change in length compositions reported by longline in 2003/4, data were selected for 10 years on either side of this point to explore how selectivity might be modelled to account for the change. Irregular quadrants (coded '9') were excluded from this analysis as the area they are applied to are ill-defined. All other grids (coded: '5': 1°x1°, '6': 5°x5°, '2': 10°x20°) were mapped to their 5°x5° squares, including the small number of 10x20° squares for which the data were distributed evenly among the 5x5° squares they contained. This was done for the sampling data, and not related to catches which we considered separately.

This resulted in a subset of around 9 000 length frequency 'samples' each for yellowfin and bigeye for the period 1993-2013, mostly provided from Taiwanese vessels (see Anon. 2013). These 'samples' are defined by separate categorical variables, so each sample may combine one or more trips and may divide single trips among months and locations. The majority of data are self-reported (Anon. 2013), but scientific observer data are also included.

Catches were used as a relative statistical weight for the length frequency to evaluate their potential importance in the stock assessment. Catch-effort data were compiled separately ('IOTC-2020-WPTT22-DATA04-CELongline.csv') and treated similarly, aggregating data by month and 5°x5° square. Because catches were used as a statistical weight, catch was measured in numbers of fish. Where numbers were not reported, weight was converted to numbers using the average weight within the stratum, either within the month, or failing that over the year, for the relevant square.

For this exploratory analysis, length frequency data were combined based on their statistical similarity to see whether this reflected any categories and provide an alternative way to group data for modelling. In analysing length frequencies, a critical question is how data should be grouped into sufficient sets that will be consistent with the way selectivity is being modelled. Data are usually combined based on categorical variables, so for example fleets, target species, fishing location or time are used to define grouping for length frequency that can be added together to create homogeneous data sets with the same selectivity. Initial review of the data suggested that this may cover up underlying patterns that could help guide how length frequency data should be handled in the stock assessment.

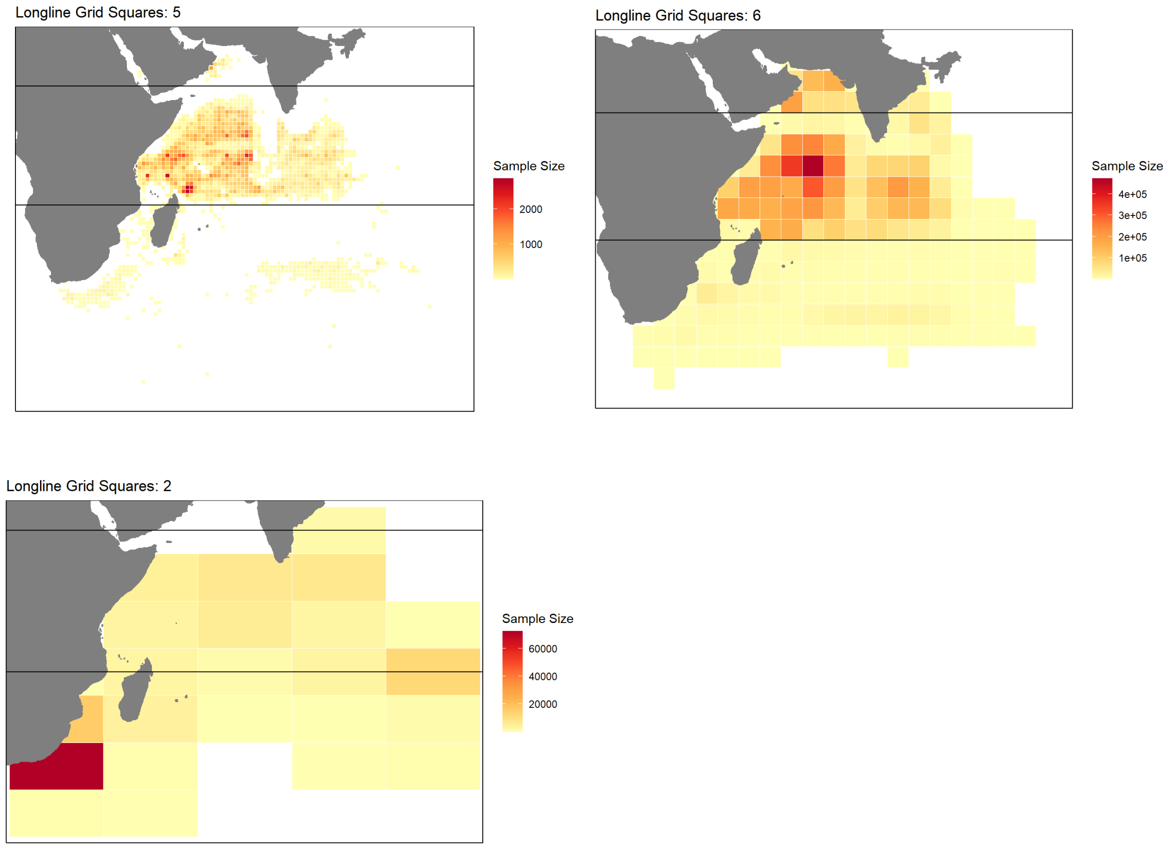

Figure 1 Spatial distribution of length frequency data for different grid sizes (coded: “5”: 1x1º, “6”:5x5º, “2”:10x20º). Data selected for this analysis were between the horizontal black lines

A simple clustering method was used based on the chi-squared statistic as a measure of distance between any two length frequency samples. This simple statistic is used to evaluate whether contingency tables are drawn from different underlying distributions, among other things. However, here we only used it as a measure of distance between length frequencies. To measure the difference, each pair of counts can be compared to the expected counts for the combined frequency using the chi-squared statistic:

$$\chi^2 = \frac{1}{d}\sum\frac{(O-E)^2}{E}$$

Where d=degrees of freedom, O=observed count and E=expected count based on the combined samples. This distance measure can be used in a cluster analysis to group samples based upon their similarity.

Some initial agglomerative clustering was carried out by combining the frequency data based on fleet, period, or area first and then conducting an agglomerative cluster analysis on the marginal data. Given around 9 000 samples for this subset of the data, it would not be easy to apply such agglomerative clustering for all samples simultaneously, so grouping was based on categorical variables.

For the individual 9 000 samples, a clustering method was applied where samples could only be combined with near neighbours. Samples could only be combined if they were within +/- 6 months of the start or end month of the sample, and shared a vertex with any 5x5º square, so the areas were at least adjoining each other. Each iteration, a pair of length frequencies which had the lowest chi-squared score were combined by adding them together to create a new length frequency 'cluster', incorporating the months and quadrants covered by both. The distance between this new length frequency and all other length frequencies was then calculated to complete the step. The initial time window was arbitrary but needed to start reasonably narrow because it rapidly expanded as length frequencies were combined. However, if too narrow, many clusters would be orphaned and would be unable to link through time.

Separately, the Regional Observer Program data ('IOTC-2020-WPEB16-DATA12-ROS_Rev2') was used to estimate the potential effect of fish size on discard rates. These data are compiled observer data that include observations on size and species of individual fish caught by longline, together with whether they were discarded or not. Data were available from EU France (2009-2018) and Japan (2012-2016) longliners and were filtered so that the data set (based on categories) had at least one recorded discard within the category based on flag, year-month, grid, and species. These data were used to estimate the discard rate conditional upon fish length using a standard binomial logit generalised linear model. This assumes that the observations are unaffected by discarding, and it is not clear that this is the case. For example, depredation is a major cause of discarding, may prevent a length measurement being taken and could be size dependent.

Interpreting length frequencies outside of an integrated stock assessment requires the use of data- limited models. The main aim is to attempt estimation of the length composition of the population alongside selectivity. There is no easy flexible way to do this, and models generally require hard assumptions if they include explicit growth, mortality, and selectivity. For this analysis, a length-based catch curve was used to account for the underlying negative exponential mortality dependent upon multiple double-normal selectivities. The objective was to see what form selectivity functions might need to take when separate length frequencies were considered together rather than provide reliable parameter estimates. In general, due to parameter aliasing, parameters need to be limited to reasonable ranges and cannot be estimated freely. Guidance for parameter values were taken from the 2018 yellowfin stock assessment (Fu *et al*. 2018).

Results

There is a clear break in length frequency patterns around 2003/4, most apparent for yellowfin although a similar pattern occurs for bigeye (Figure 2). This pattern is solely from the Taiwanese fleet which makes up most of the data. Other patterns are not clearly discernible from the standard cluster analysis.

Based on the natural clustering method, the length frequencies appear to come from different distributions that can be separated. Results suggest at least five groups, based only on statistical difference, might exist (Figure 3). For further exploration below, 10 or 8 clustering levels were examined.

Length frequencies consisting of similar sizes separate out into discrete clusters for both yellowfin (Figure 4) and bigeye (Figure 5). The clusters can be identified by their mean length. The small fish apparent as a separate mode in the combined length frequencies form their own separate group and are not integral with samples of larger fish. This is not consistent with, for example, changes in discarding practice where the change in presence of smaller fish might be detected as an increased left-sided truncation of the samples.

There is some correspondence between mean length of clusters between yellowfin and bigeye. Length frequencies containing small or large fish for both species are more likely to occur together within the same time and location (Figure 6).

In a minority of samples either the vessels are targeting swordfish or observers are present. When the longline is targeting swordfish, the samples tend to be more often allocated to clusters containing smaller fish (Table 1). Swordfish longline is generally set shallower and overnight compared to tuna longline, and this would likely affect selectivity. The implication is that variations in the way tuna longline are set may similarly affect the size of fish they catch. Observer length frequencies are absent from the small fish clusters, but no observer data exist for the earlier period which will explain this. Otherwise, observer data shows a reasonable spread among clusters, but data from training and research trips have not been identified within this period.

No patterns were evident based on flag, although it is worth noting only Japan, Korea and Taiwan cover the earlier years and Taiwan data dominate the observations for all periods.

The cluster groups chain through time in a complex way and there is no simple pattern in their distribution. For the period considered, different length frequency groups either extend throughout the time period or are limited in extent, mostly breaking around the 2003/4 mid-point for both yellowfin and bigeye (Figure 7).

Focusing on the smallest category of yellowfin (mean length 87 cm) and plotting its occurrence over time suggests that it sporadically appears throughout the 1993-2003 time period, with peak times in 1993, 1995 and through 1999-2001 (Figure 8). In terms of location, the majority of the samples come from the area bounded by 0°N 60°E and 10°N 65°E. However, this cluster is not seen after 2003, and it is still not clear why.

Weighting the samples by reported catches appears to raise the proportion of the small fish selectivity considerably compared to the raw samples for yellowfin (Figure 4) and bigeye (Figure 5) but is based on a relatively small overlap between the length frequency samples and the reported catches. For yellowfin, between 85 % and 96 % of length frequency samples can be linked to a reported catch/effort in a year/month/grid stratum. This is likely because length compositions were reported from training vessels, but not catches. The proportion of catch/effort that can be allocated in reverse to a length frequency sample varies from 4 % to 41 % in any year, and these catch/effort data only cover 13 %-36 % of the total reported catch. The proportion of the total catches that can be reliably allocated to a length frequency cluster is therefore very small. The proportion of catch effort data that can be linked to a length frequency, important for the abundance indices, is higher but still will present a significant problem for interpreting the time series data if length frequency data are broken out into smaller groups. The same problem occurs for bigeye, but the overlap is a little greater.

It is possible that the clustering process and the method using only statistical similarity may produce artificial discontinuities in what is more a continuous change in the size composition. To test this, smaller fish were removed from the data so that the clustering procedure is applied with a data set where these data have been censored (as might occur with discarding). All fish <80 cm length were excluded from the yellowfin clustering, to see whether the apparent break in 2003/4 would disappear or remain (Figure 9). The results seem robust to this test. If time series is split and the clustering applied separately to the periods before and after this break, very different combined length compositions result (Figure 10). The later data 2004-2013 are clearly much more homogeneous.

The Regional Observer Program data indicated that discarding would not provide an explanation for change in size composition (Figure 11). Even for the yellowfin smallest size cluster, fish are predominantly above the 83 cm 95 % retention rate. Interestingly, the probability of retention was not significantly different between bigeye and yellowfin, although the number of small fish in the samples was very low because these sized fish were rarely caught. It was also found that length was a better predictor for discarding than weight. In general, the relatively low discard rates indicated here are consistent with those reported for Taiwan vessels (Huang and Liu, 2010).

Fitted double-normal selectivity curves suggested that some of these length frequency clusters may be adequately modelled using a simple logistic (as currently used for longline in the stock assessment), that bimodal length frequencies still exist after clustering that cannot be well-explained with a single mode selectivity function, and that there may be some evidence for domed-shaped selectivity in some cases (Figure 12). However, as noted, fitting length-based catch curves is problematic and requires that various parameters be fixed, so these preliminary results are qualitative only. Selectivity fits to bigeye were poorer, although the general results were similar.

The same clustering analysis carried out on the public ICCAT length frequency data, revealed a similar pattern to that seen for the IOTC data (Figure 13). A separate length frequency cluster of smaller fish formed, consisting of Taiwan data from the period 1993-2003.

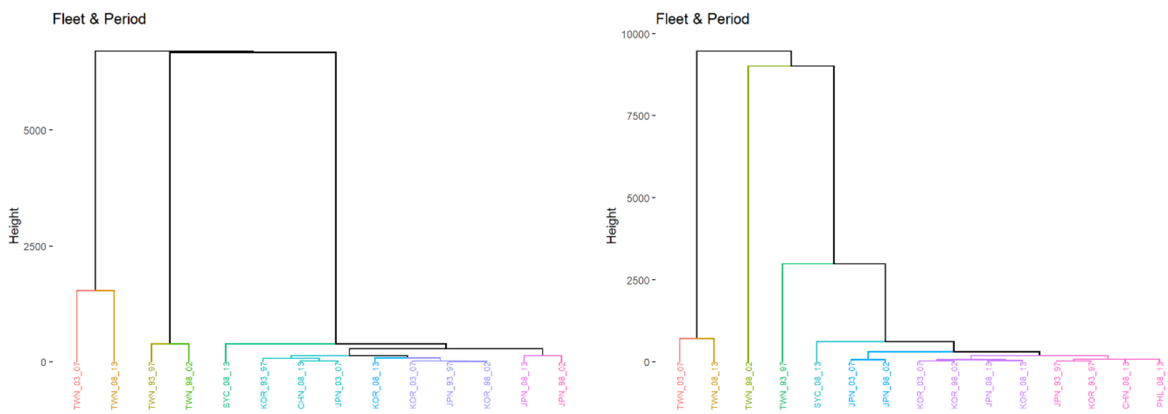

Figure 2 Hierarchical clustering for yellowfin (left) and bigeye (right) marginal distance measure (chi-squared statistic) for length frequencies combined into categorical groups based on fleet and 5-year period

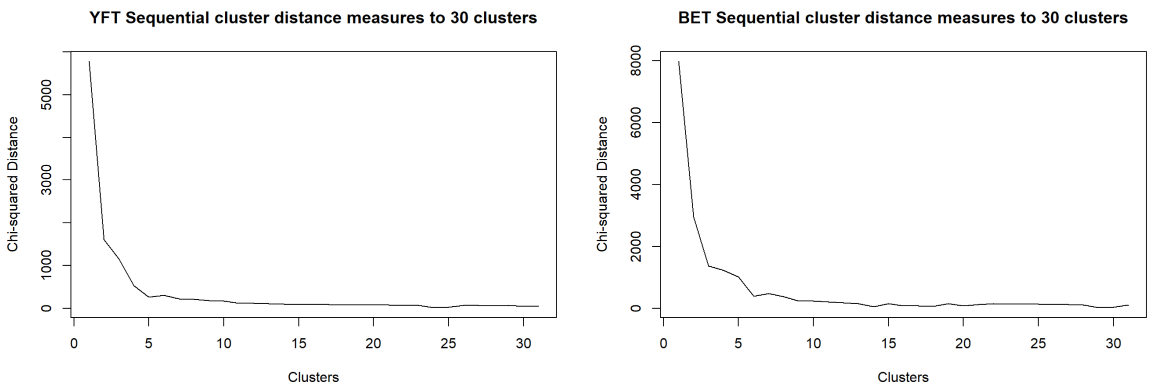

Figure 3 Sequential chain clustering distance for yellowfin (left) and bigeye (right) for the top 30 clusters

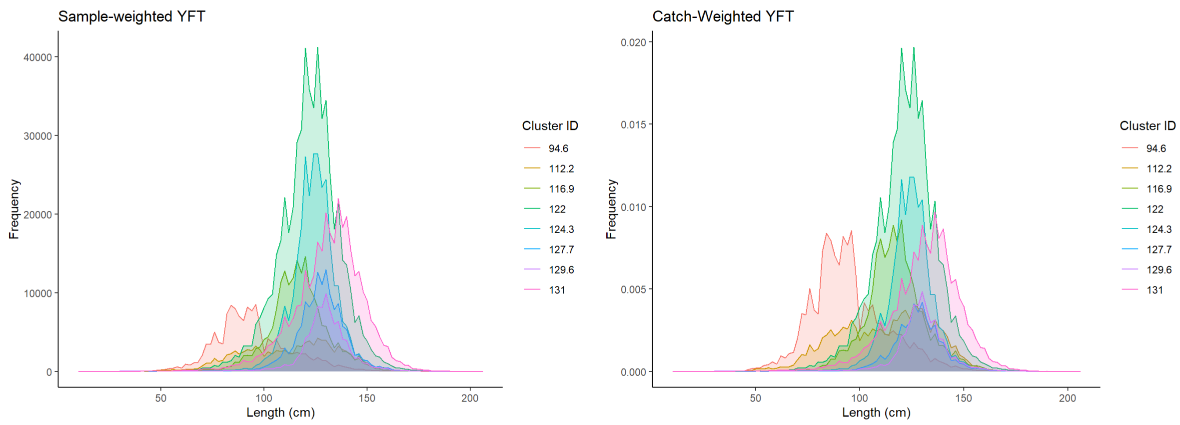

Figure 4 Yellowfin length frequencies for the eight chain clusters for the raw data (left) and raised based on the associated catch weights (right)

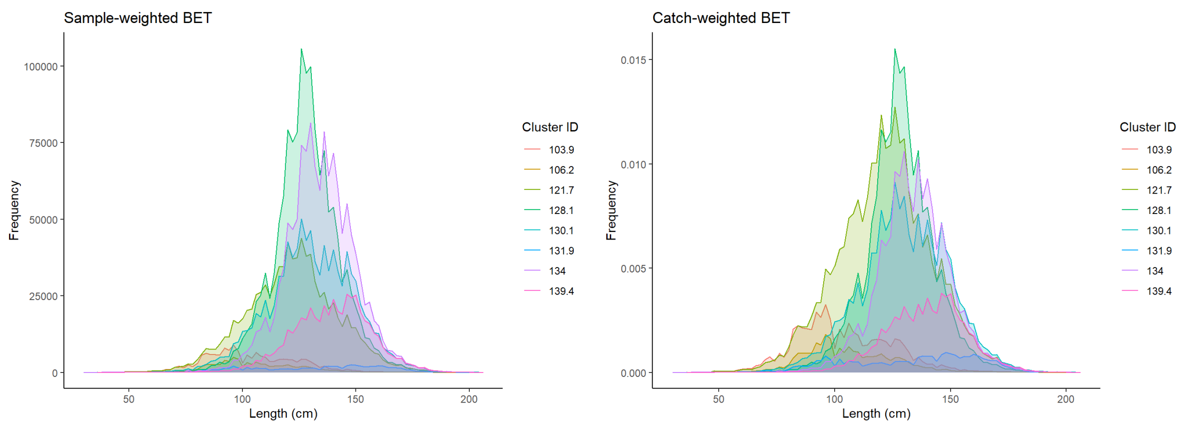

Figure 5 Bigeye length frequencies for the eight chain clusters for the raw data (left) and raised based on the associated catch weights (right)

Table 1 Yellowfin clusters with sample size (number of fish measured) broken down by longline type

Cluster Mean Length (cm)	Standard	Swordfish	Fresh tuna	With Observers
86.3	29 085	0	0	0
96.7	110 507	170	0	0
112.2	103 876	792	0	0
116.4	114 104	195	444	3 631
117.8	73 048	54	0	974
122.0	557 871	74	727	2 927
124.3	292 222	10	0	2 051
127.7	138 457	0	98	1 607
129.6	84 317	188	35	2 875
131.0	308 959	371	3 306	5 738

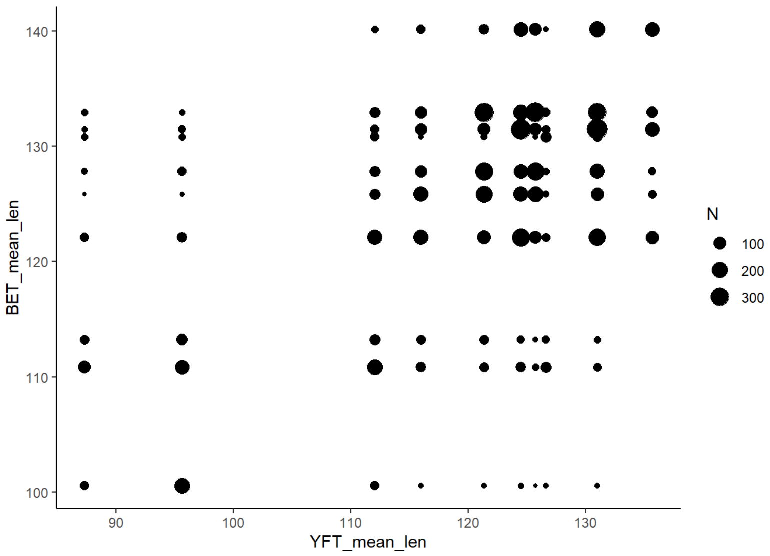

Figure 6 Cluster correspondence between bigeye and yellowfin

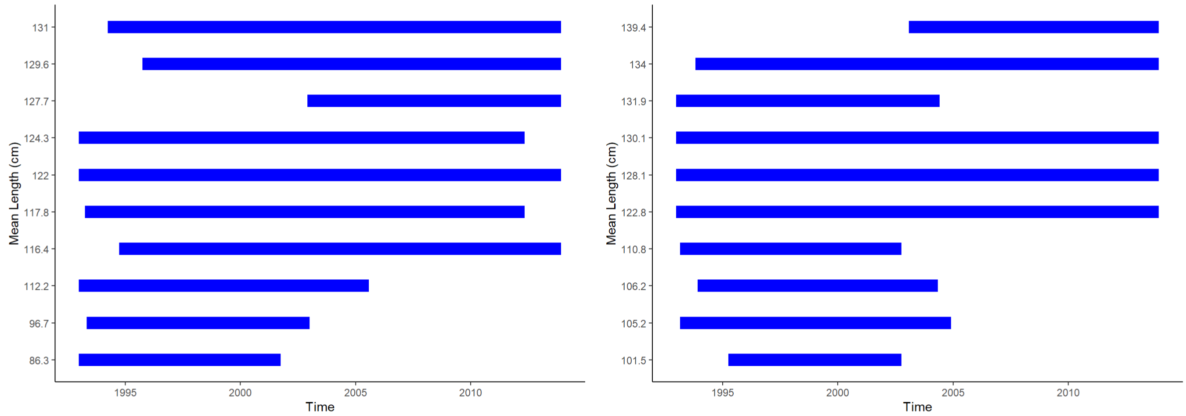

Figure 7 Ten length frequency clusters chains for yellowfin (left) and bigeye (right) distribution through time.

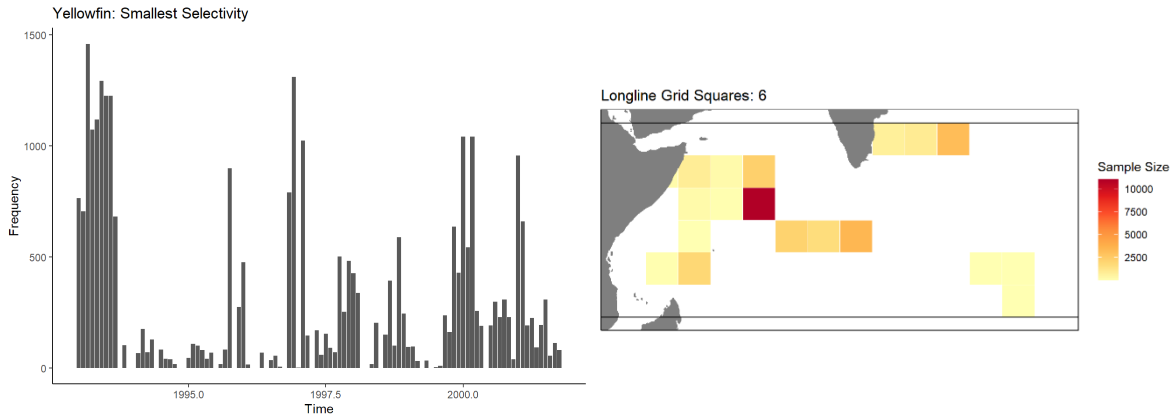

Figure 8 Distribution of the cluster chain for the smallest yellowfin through time (left) and in space (right)

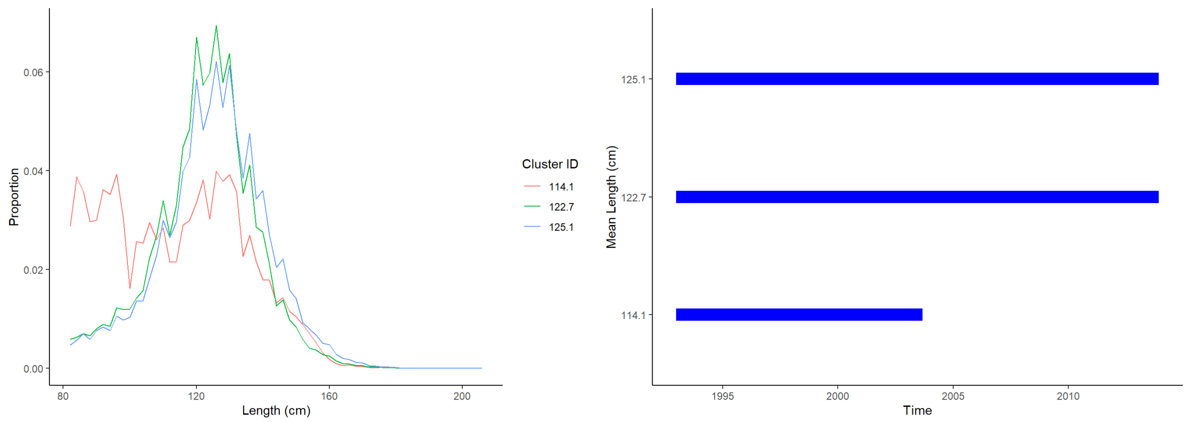

Figure 9 Yellowfin top three cluster chains for left-censored data which removes all length frequencies less than 80 cm

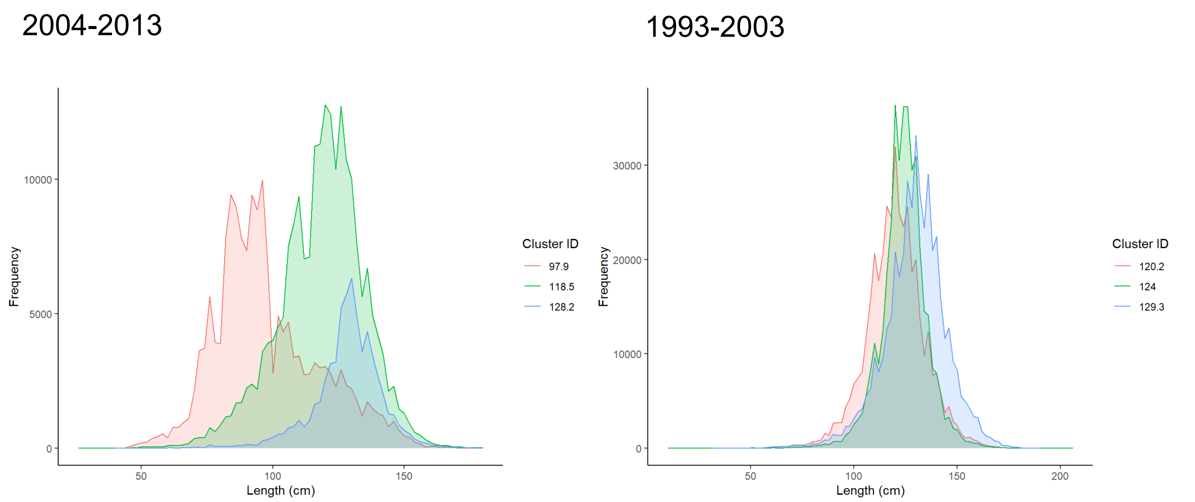

Figure 10 Yellowfin top 3 length frequency cluster chains conducted separately for the period 1993-2003 and 2004-2013

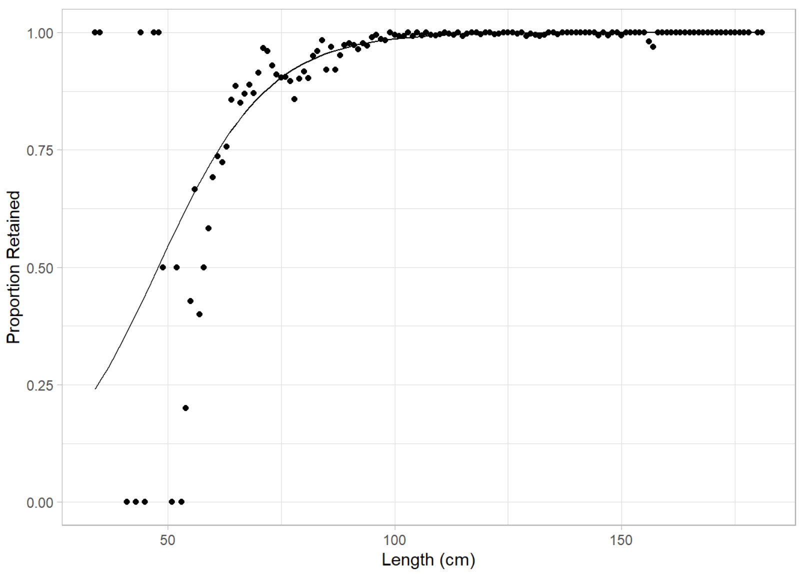

Figure 11 Retention proportion for regional observer scheme data with observed and expected based on a standard binomial logit model.

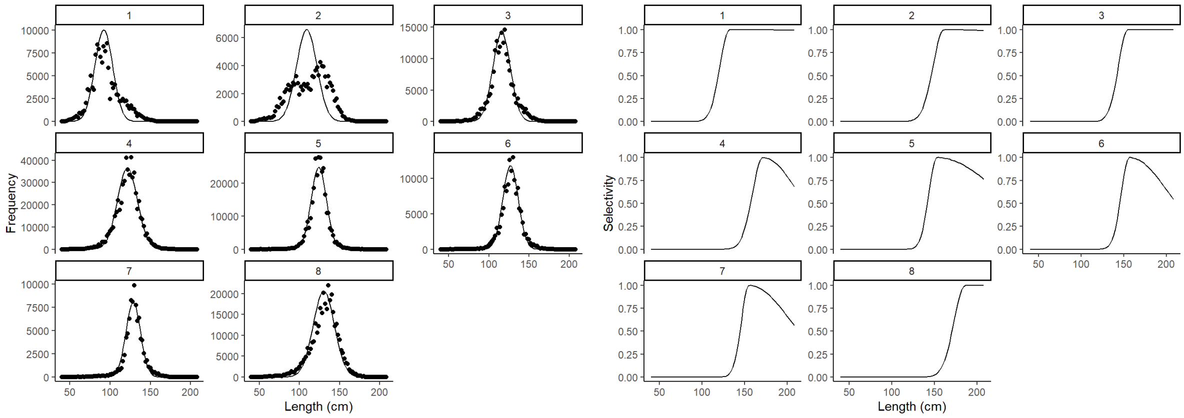

Figure 12 Yellowfin double-sided normal selectivity curves fitted to the length frequency clusters showing the fit (left) and the estimated curves (right).

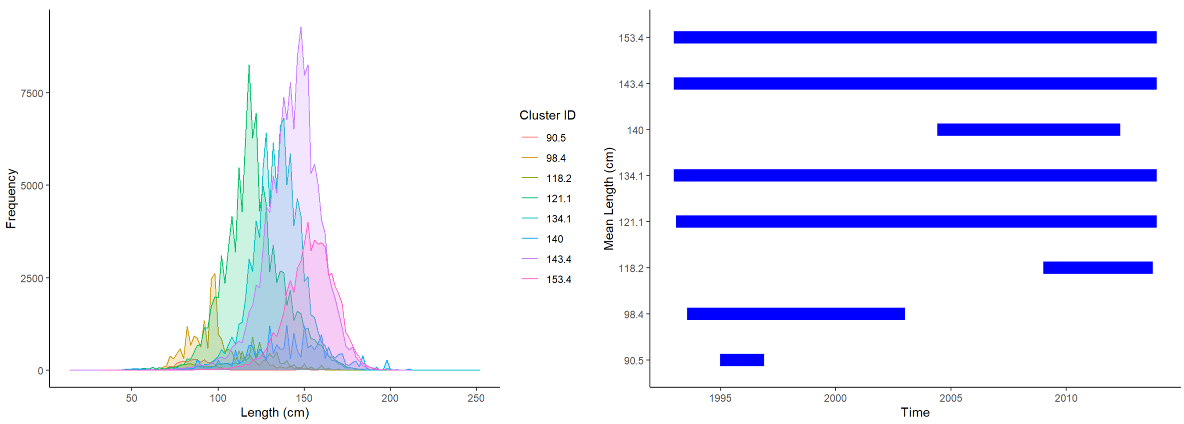

Figure 13 ICCAT yellowfin length frequencies for the eight chain clusters for the raw data (left) and distribution through time (right).

Discussion

An aim of this approach to combining samples was to provide an objective way to group data consistent with underlying selectivity models. Gear selectivity is likely the result of 1) various forms of sampling error 2) operational characteristics of the sets (e.g. set depth hook size) 3) relative availability of different size fish in the population. The results do not provide a solution to separating these issues but suggest that it may be possible to use the available information to guide combining and modelling length frequencies appropriately. Length frequency data appear to be lengths drawn from mixtures of simpler selectivity functions or at least might be reasonably modelled that way.

It was hypothesized that the change in length compositions in 2003/4 could be due to change in discard practice (Hoyle et al. 2019). This appears to be inconsistent with the available data which does not show length frequency truncation that might result from size-specific discarding. The data are more consistent with some other cause such as a change in vessel operations which have affected the underlying selectivity. However what this operational change might have been remains unclear. Because this pattern is seen in separate oceans it would seem to be mostly likely because of a common change in fleet operations sampling or data management. A change in sampling protocol cannot be ruled out because there remains an inconsistency with the catch-effort data (Geehan and Hoyle 2013) but it appears that the problem may still have only affected a discrete group of the samples. It therefore may be worth reconsidering how these Taiwan length frequencies which make up a substantial proportion of the data might be readmitted to the stock assessment.

Linking the length frequencies to catches appears to be problematic and the overlap between separate data sets is limited. While most length frequency samples can be linked to catch and effort data in the same month and 5°x5° square the reverse is not true particularly for yellowfin. Even if catches can be linked to the nearest length frequency it is not clear that length frequency samples at this fine scale would be representative of all catches in that area-time stratum. There are other data such as mean weights (Geehan and Hoyle 2013) which might help with resolving this problem.

The clustering approach used here was naïve and may not produce the desired grouping based upon selectivity. For example some clusters appear to be bi-modal and may be mixtures themselves. Clusters also appear to have a different spread with the right-hand side of each frequency potentially having a different slope that may reflect the way they have been selected to be combined rather than the underlying patterns in selectivity. Some samples consist of very few fish and these contribute to the ‘spiking’ around the frequency modes and some minimum sample size might be identified to exclude them. Grouping the length frequencies may need a more sophisticated approach to be reliable particularly as more heterogeneous frequencies are combined. It is worth noting that Taiwan uses cluster analysis to identify fishing strategy to standardise CPUE (Hoyle et al. 2019) so some correspondence might be found with their approach.

The results from this exploratory analysis suggest that there are latent categorical variables which might be inferred using more sophisticated classification techniques. These techniques might be able to identify an underlying cause for the apparent different selectivities by comparing different species oceans and fleets. Using classification to obtain more consistent length frequencies would improve modelling of the selectivity in the stock assessment.

There are several ways to proceed with this analysis that may provide further insight. These include:

- Expand the data outside the limited range used for this exploratory analysis. More recent data in particular should be included that have a higher proportion of observer coverage and the area should be extended to cover Oman in the north and the Mozambique Channel in the south.
- A model-based approached for grouping and combining length frequencies may improve results particularly as clusters become large. A model-based approach would explicitly try to identify the underlying mixture of length frequencies that make up the selectivity function and avoid for example multi-modal distributions forming.
- A standardisation approach for the size composition data may also be developed in a similar way as for the abundance index (Thorson 2014). This might be used specifically to account for the poor overlap between the size and catch-effort data. Mean weights could be used as an additional guide where direct length frequencies are unavailable as well as be used to detect inconsistencies.
- It would be useful to improve the distance measures incorporating time and location as distance measures. This would avoid the need to collapse 1°x1° squares into coarser 5°x5° squares and control better how clusters chain through time. If available finer scale data and raw reported data might be also used which could potentially identify differences between length frequencies more accurately as well as identify different sampling protocols (see Anon. 2013).
- The observer data can be used more explicitly to initiate clusters and mean weights can be used as an additional indicator of size composition (Matsumoto 2013). This might help identify data with sampling problems as opposed to other causes of differing length compositions.
- Length frequency data from other gears could be used to look for an availability effect on length compositions. The operation of free-school set purse seine gill net and troll size compositions may show size overlap to some degree with longline.
- Understand possible causes of change in length compositions from 2003/4 by consulting IOTC CPCs (that is IOTC members and cooperating non-contracting parties).
- The same analyses could be applied for other oceans (e.g. ICCAT data) to look for operational or sampling effects that they have in common.

References

Anon. 2013. Data Collection and Processing System of Statistics for the Taiwanese Deep-Sea Longline Fishery. IOTC-2013-WPDCS09-08

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

Geehan J. Hoyle S. 2013. Review of Length Frequency Data of the Taiwanese Distant Water Longline Fleet. IOTC–2013–WPTT15–41_Rev1

Hoyle S. D. Chassot E. Fu D. Kim D. N. Lee S. I. Matsumoto T. et al. 2018. *Collaborative study of yellowfin tuna CPUE from multiple Indian Ocean longline fleets in 2018.* Retrieved from https://www.iotc.org/documents/WPM/09/12

Hoyle S. Merino G. Murua G. Yeh Y. Chang S. Matsumuto T. et al. 2019. IOTC–CPUEWS–06 2019: Report of the Sixth IOTC CPUE Workshop on Longline fisheries. San Sebastian April 28th – May 3rd 2019. IOTC–2019–CPUEWS06–R[E]. Retrieved from https://iotc.org/sites/default/files/documents/2019/10/IOTC-2019-WPTT21-INF01.pdf

Huang H-W. Liu K-M. 2010. Bycatch and discards by Taiwanese large-scale tuna longline fleets in the Indian Ocean. Fisheries Research 106: 261–270

IOTC–2019–WPTT21 Report of the 21st Session of the Working Party on Topical Tuna. Donostia-San Sebastian Spain 21 - 26 October 2019

Matsumoto T. 2013. Comparison of size data and average weight for bigeye and yellowfin tuna caught by Japanese longline in the Indian Ocean based on different sampling or estimation methods. IOTC-2013-WPTT15-22

Naunet. (2020). Developing management advice to rebuild the Indian Ocean yellowfin tuna (Thunnus albacares) stock in two generations. A report for the Global Tuna Alliance (GTA). Retrieved from https://www.globaltunaalliance.com/wp-content/uploads/2020/07/IOTC-YFN-Report-FINAL.pdf

Sharma R. 2018. Review of IOTC YFT in 2018. IOTC Scientific Committee Seychelles December. 2018. IOTC-2018-SC21-INF02

Thorson J.T. 2014 Standardizing compositional data for stock assessment. ICES Journal of Marine Science doi:10.1093/icesjms/fst224

Urtizberea A. Fu D. Merino G. et al. 2019. Preliminary Assessment of Indian Ocean Yellowfin Tuna 1950-2018 (Stock Synthesis v3.30). IOTC-2019-WPTT-21-50