Received: 10 & 16 November 2014

IOTC–2014–WPTT16–48 Rev_1

Exploration of area stratification for CPUE standardization of yellowfin tuna by Japanese longline.

by

Keisuke Satoh

National Research Institute of Far Seas Fisheries (NRIFSF), Japan

Summary

New spatial sub-area for yellowfin tuna CPUE standardization process of Japanese longline fishery in the Indian Ocean was proposed using the simultaneous tree method and examined performance for the present and new sub-area definition. Relative abundance indices using the two area definitions were compared. The analyses in this study included three components: Analysis 1; using only size data, Analysis 2; using only CPUE data, and Analysis 3; using both size and CPUE data. The trees of the three analyses appear to show agreement in two points: 1) the first split around 15S, and 2) the second split on around equatorial. The seasonal effect in the analysis 1 was not clear in the analyses 2 and 3. It looks like CPUE trends have more influence on the simultaneous tree structure than the size distributions do. The statistics, U(s), for ranking the candidate stratifications in the analysis 3 (0.154) was larger than the value of 0.117 for the present sub-area definition, which indicated that the new area definition presented more uniform in size composition and CPUE trend rather than the present sub-area definition. Comparison standardized CPUEs using the two area definition showed similar trend with some annual fluctuation.

Introduction

Catch-per-unit-effort (CPUE) is important information as relative abundance index in stock assessment. Many of the impacts of spatiotemporal changes in fishing effort on CPUE, or other factors that affect catchability, can be removed by applying statistical methods to standardize CPUE (Ichinokawa and Brodziak 2010). In the CPUE standardization process, we often separate whole target area into smaller spatial area (sub-area), since in such smaller area is a region in which fish density is homogeneous and CPUE is influenced by many factors in a similar manner (Bishop, 2006). Furthermore, catch size composition usually differs among fishing grounds even if the fishery is same. Large differences in the catch size composition could influence trends of the CPUE, at least cause time lag of CPUE trend according to the differences in the catch size composition. Spatial differences in the catch size composition indicate spatial differences in the population size composition if selectivity is constant over the area of the fishery (Lennert-Cody et al., 2013). Therefore, the smaller area defined by taking into account the catch size composition should provide the standardized CPUE with clearer catch size composition. Such information is useful for defining fishery and its selectivity in stock assessment.

Previous studies (e.g. Matsumoto et al., 2012, Satoh and Okamoto, 2012) to standardize CPUE appear to determine the sub-area in an ad hoc manner (Ichinokawa and Brodziak 2010). They refer spatial distribution of fishing effort and nominal CPUE or oceanographic conditions for making the sub-area. In recent years, several studies for the spatial pattern in CPUE standardization using tree regression model have been developed (Walsh and Kleiber, 2001, Ichinokawa and Broziak 2010, Lennert-Cody et al., 2010).

The aims of the present study were to (1) define new spatial sub-area for yellowfin tuna CPUE standardization process of Japanese longline fishery in the Indian Ocean using the simultaneous tree method (Lennert-Cody, CE et al., 2013); (2) to examine performance for the present and new sub-area definition; and (3) compare relative abundance indices between the two area definitions.

Methods and materials

Catch size composition data used:

Length frequency data in NRIFSF were used. The period for this study was from 1957 to 2008. Both catch size composition data measured by commercial vessels and training vessels were used. It is known there were no large differences between the two data source (Matsumoto 2013). The catch size composition data was adjusted to the middle month of each quarter by adding or subtracting a monthly length increment by the method described in Appendix 1. Length intervals were: 0–90 cm; –102 cm; –108 cm; –114 cm; –120 cm; –126 cm; –132 cm; –138 cm; -144 cm; >=144 cm.

Catch and effort data used:

The Japanese longline catch (in number) and effort statistics from 1957 up to 2008 were used. The data is same data source in previous Japanese longline CPUE standardization studies (e.g. Matsumoto et al., 2012, Satoh and Okamoto, 2012). Data in this study are the catch and effort data sets aggregated by quarter, 5° latitude x 10° longitude area before computing nominal CPUE (= sum of catches / sum of hooks). For each quarter of the year by 5x10 degrees square this gives a time series of nominal CPUE based on a maximum of 156 data points (= 3 months x 52 years). Grid cells were considered to have sufficient CPUE data for trend estimation if they contained at least 50 data points over 25 years and if the total catch was at least 0.01% of that for the entire data set. For each grid cell, the following simple smooth model was fitted to the CPUE data: log (CPUE+1/10 average) = f (year + error), where f is a smooth function, y indexes year. The natural logarithm transformation was applied for improvement of unequal variance of the CPUE. The nominal CPUE was smoothed by a generalized additive model (GAM) on year, the detail method for GAM was described in Appendix 1.

Defining sub area -The simultaneous tree method -

The method for defining sub-area, simultaneous tree method, applied in this study is almost same described in Lennert-Cody, CE et al. (2013). Therefore in this section there is no detail descriptions for the methods. The method assumed that spatial stratum indicating homogeneous CPUE trend and catch size composition will be good for sub-area for CPUE standardization. The method requires the following components: (1) an appropriate impurity measure for each data type; (2) an overall split criterion, based on the impurity measures of each data type, that is used to build the tree; and (3) a statistic that can be used to rank candidate stratifications produced from the tree. More information for the methods is briefly presented in Appendix 1.

CPUE standardization

CPUE standardized method and materials are almost same described in Ochi et al. (2014) except for the sub-area definition which produced by the simultaneous tree method in this study.

Results and Discussions

The simultaneous tree method applied to the yellowfin tuna data. The analyses in this study included three components: analysis 1; using only catch size composition data, analysis 2; using only CPUE data, analysis 3; using both catch size composition data and CPUE data. The partitions of the trees (**Fig. 1**) and final area stratifications (**Fig. 2**) of these analyses were presented. All values of U(s) for the last candidate of sub-area for analyses 1 and 2 were presented in **Table 1**. For each candidates of the analysis 3, and for the present sub-area definition were presented in **Fig. 3** and **Table 1**.

In the analysis 1 (size only), the partitioning was different in north of 15° S. The proportion for smaller fish in quarter 1 was greater than the other quarters in north of 15° S (**Fig. 4**). There was seasonal effect for the catch size was clear in the analysis 1. However the seasonal effect was not clear in the analysis 3. The partitioning for the analysis 2 was similar to the analysis 3. However the area of east of 100 E in north of 0N did not contain enough effort, therefore the area may not work well in CPUE standardization process.

Values of U(s) in the analysis 3 increased according to the number of area increased (**Table 1**). The highest number in the analysis 3 was 0.154 for the five sub-areas candidate (**Fig. 3**), which was larger than the value of 0.117 for the present sub-area definition. The new five sub-area definition generated by the simultaneous tree method in the analysis 3 presented more uniform in catch size composition and CPUE trend rather than the present sub-area definition.

U(s) values in the analysis 1 (0.027) was higher than any U(s) values (0.009- 0.018) in size component of the analysis 3 (**Table 1**). The U(s) of the analysis 2 (0.169) was also higher than U(s) (0.075-0.136) of CPUE component in the analysis 3. The simultaneous tree method slightly lost its performance rather than the situation using each data set separately. The spatial-temporal agreement between the two data types might is not good. However the size composition for each sub-area of the analysis 3 were well characterized, that is, wide size range fish in southeastern part of the Indian Ocean (new area 4), higher proportion of small fish in northeastern part (new area 5) and lower proportion of large fish in northeastern part (new area 1) (**Fig. 5**). Such clear differences of the catch size composition among sub-area indicate that the simultaneous tree method can group similar size composition in spatial manner.

Results of CPUE standardization of yellowfin tuna of Japanese longline in the Indian Ocean using the new sub-area and the present definition (Ochi et al 2014) showed very similar trend with some annual fluctuation (**Fig. 6**). Relative indices by sub-area were presented in **Fig. 7.** Each indices contained information of catch size composition, which were different from each sub-area.

In practical, same sub-area definition for stock assessment and CPUE standardization is useful because the standardized CPUE by sub-area is with size composition differed from other areas. Such information could be useful for defining fishery and its selectivity in stock assessment. In other hand, we need to focus on modeling of fish migration, which may require other area definition. In such case the standardized CPUE should be provided with flexibility.

Acknowledgement

I am very grateful for Dr. Lennert-Cody C.E. of Inter-American Tropical Tuna Commission (IATTC) for detail and useful instruction of the simultaneous tree method.

Reference

Bishop, J., 2006. Standardizing fishery-dependent catch and effort data in complex fisheries with technology change. Rev. Fish Biol. Fish 16, 21–38.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman & Hall/CRC, New York.

Eveson, J. P., Million, J., Sardenne, F., & Le Croizier, G. (2014). Estimating growth of tropical tunas in the Indian Ocean using tag-recapture data and otolith-based age estimates. Fisheries Research.

Ichinokawa, M., Brodziak, J., 2010. Using adaptive area stratification to standardize catch rates with application to North Pacific swordfish (Xiphias gladius). Fish. Res. 106, 249–260.

Lennert-Cody, C.E., Minami, M., Tomlinson, P.K., Maunder, M.N., 2010. Exploratory analysis of spatial–temporal patterns in length-frequency data: an example of distributional regression trees. Fish. Res. 102, 323–326.

Lennert-Cody, C. E., Maunder, M. N., Aires-da-Silva, A., & Minami, M. 2013. Defining population spatial units: Simultaneous analysis of frequency distributions and time series. Fish. Res. 139, 85-92.

Matsumoto. M. 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.

Matsumoto, T, Okamoto, H., Kitakado, T., 2012. Japanese longline CPUE for yellowfin tuna in the Indian Ocean up to 2011 standardized by general linear model. IOTC–2012–WPTT14–35.

Ochi, D., Matsumoto, T., Okamoto, H., Kitakado, T. 2014. Japanese longline CPUE for yellowfin tuna in the Indian Ocean up to 2013 standardized by generalized linear model. IOTC-2014-WPTT16-47.

Quinn II., T.J., Deriso, R.B., 1999. Quantitative Fish Dynamics. Oxford University Press, New York.

Satoh, K., Okamoto, H., 2012 Updated Japanese longline CPUE for bigeye tuna in the Indian Ocean standardized by GLM. IOTC–2012–WPTT14–26

Walsh, W.A., Kleiber, P., 2001. Generalized additive model and regression tree analyses of blue shark (Prionace glauca) catch rates by the Hawaii-based commercial longline fishery. Fish Res. 53, 115–131.

Wang, Q., Kulkarni, S.R., Verdú, S., 2005. Divergence estimation of continuous distributions based on data-dependent partitions. IEEE Trans. Inform. Theory 51, 3064–3074.

Table 1 U(s) values for the last sub area definition (Fig. 2) of analyses 1 and 2, for each candidates of analysis 3, and for the present sub-area definition.

number of sub-area (s)	Analysis 1 (size only)	Analysis 2 (CPUE only)	Analysis 3 (size + CPUE)			present sub-area definition*
			size	cpue	total	
2			0.009	0.075	0.084	
3	0.027		0.014	0.103	0.117	
4-1			0.016	0.129	0.145	
4-2			0.016	0.109	0.125	
5		0.169	0.018	0.136	0.154	0.117

*The present sub-area definition contained 75E as its criteria. Spatial resolution for the CPUE and size data in this study was 5 x 10 degrees. Therefore I used the criteria as 80E instead of 75E.

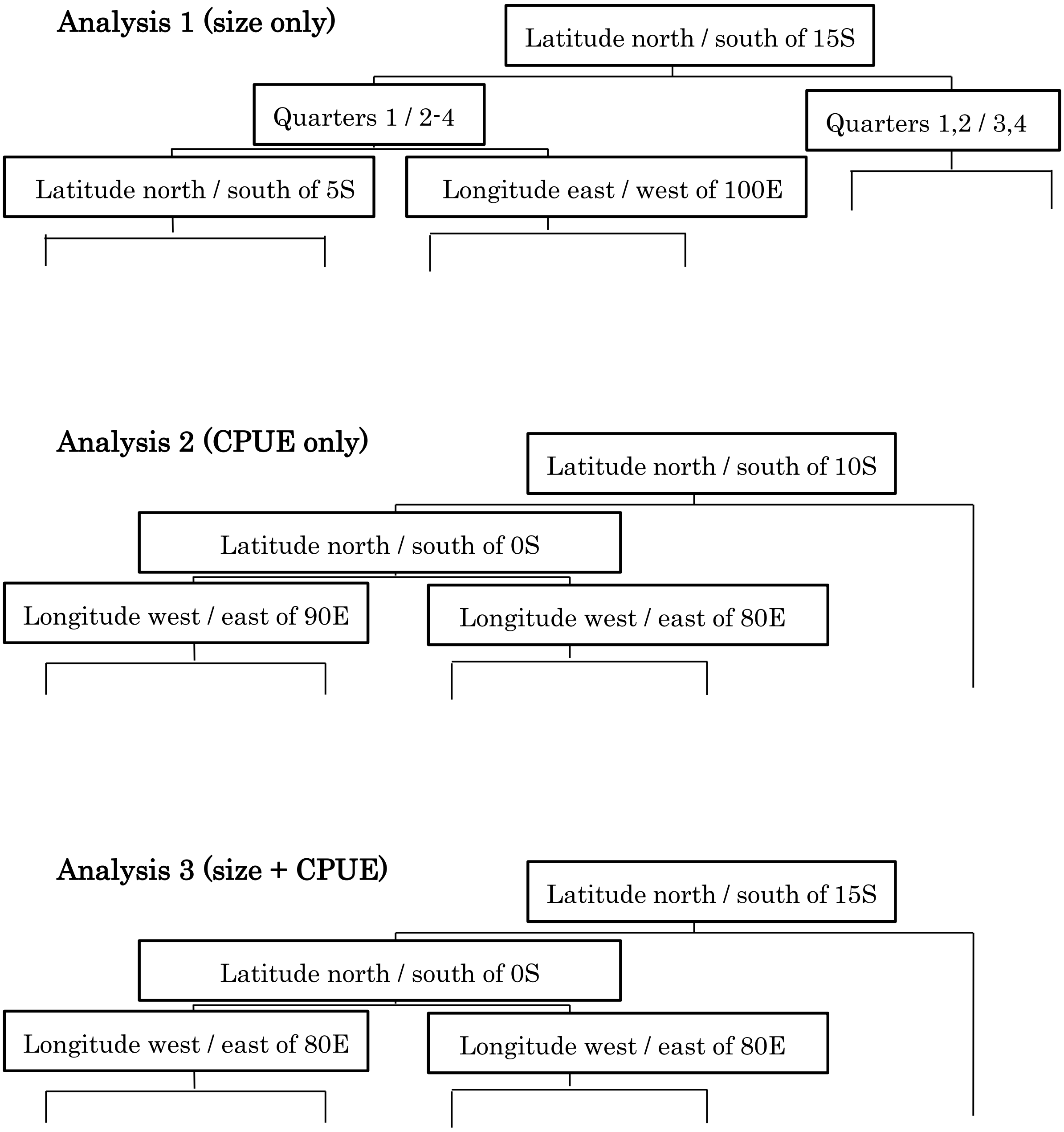

Fig. 1 Trees produced by the three analyses (upper panel; analysis 1 (catch size composition data only), middle panel; analysis 2 (CPUE data only), lower panel; analysis 3 (catch size compostion data + CPUE data)). Branch length has no meaning.

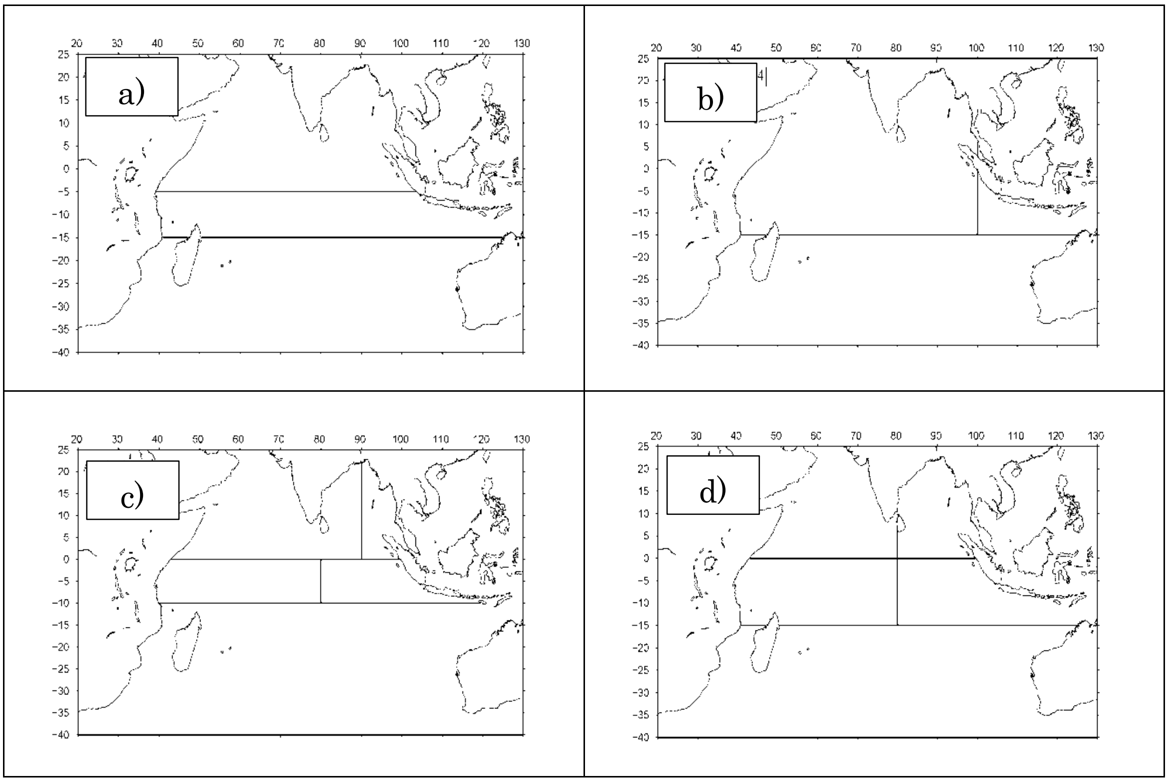

Fig. 2 Candidate stratifications generated by applying the method to the yellowfin tuna data in the Indian Ocean. (a) analysis 1 in quarter 1 (catch size composition data only), (b) analysis 1 in quarters 2 to 4, (c) analysis 2 (CPUE data only) and (d) analysis 3 (catch size composition data + CPUE data).

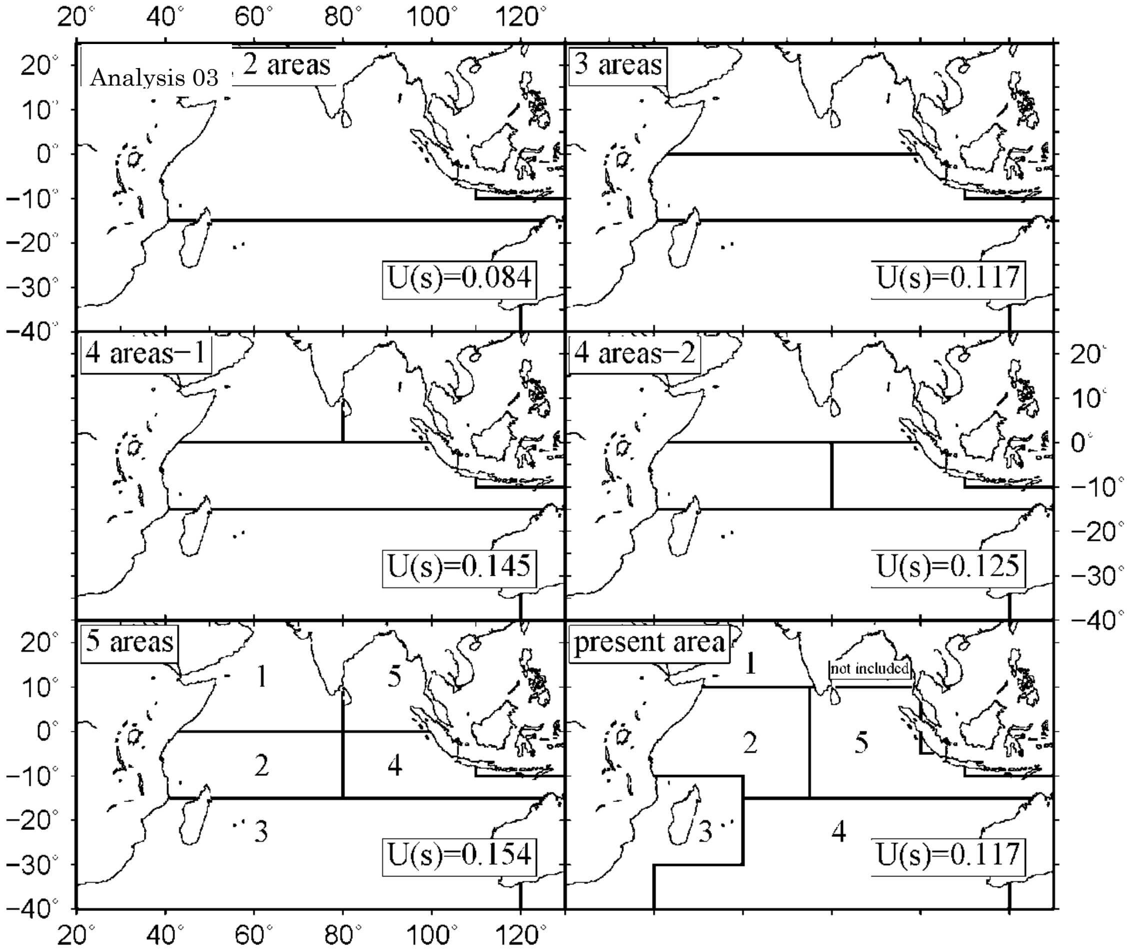

Fig. 3 Candidate stratifications and corresponding U(s) values generated by applying the simultaneous tree method to the yellowfin tuna data in the Indian Ocean for analysis 3 (catch size composition data + CPUE data). The value of U(s) for the present sub-area definition (bottom right).

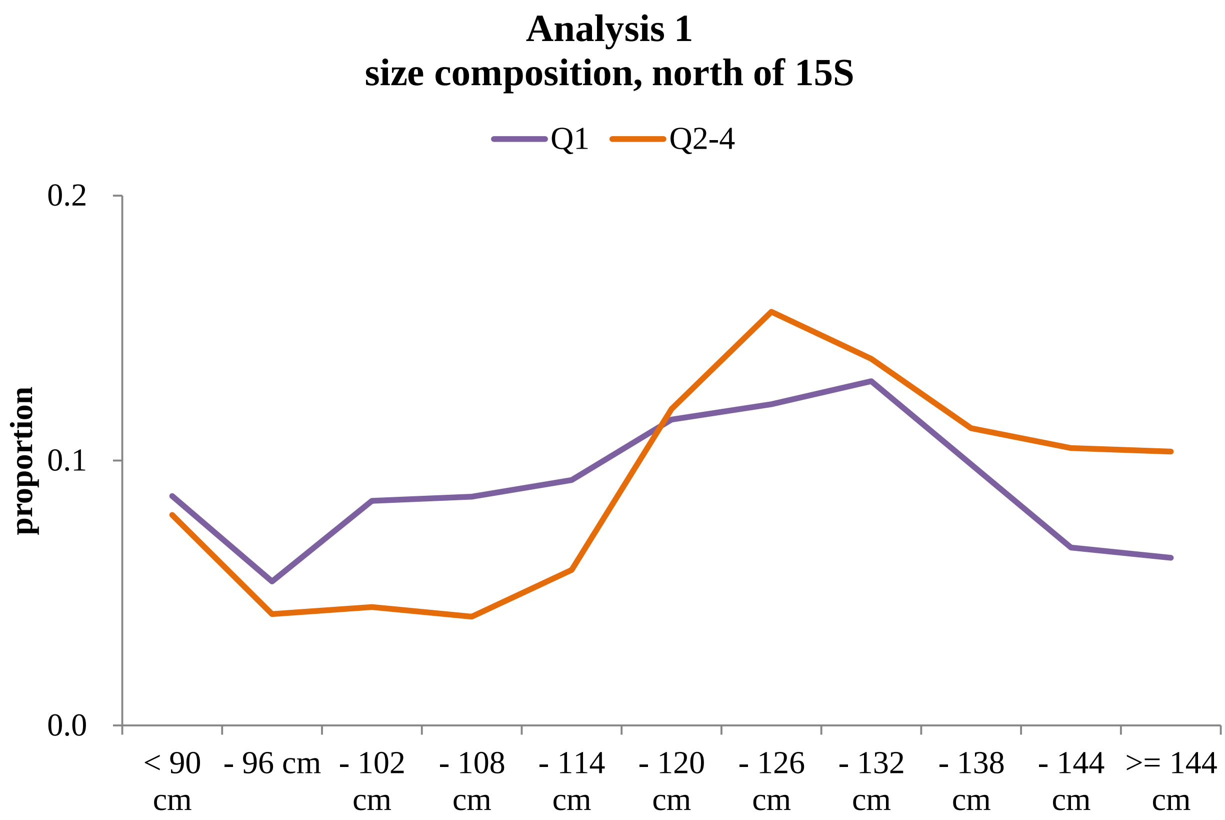

Fig. 4 Comparison between size composition quarter 1 and quarters 2 to 4 in north of 15S.

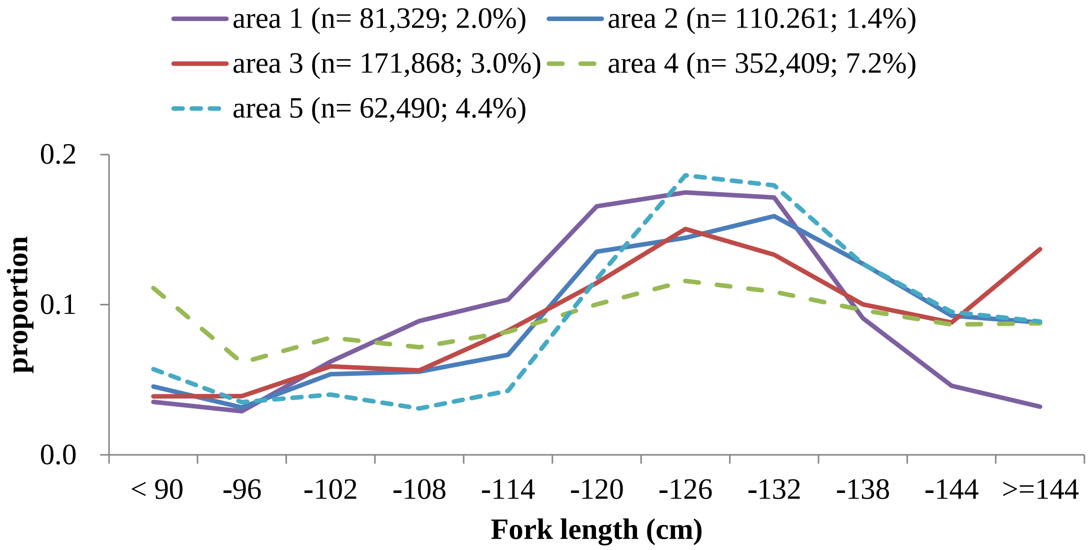

Fig. 5 Size composition for each area by applying the simultaneous tree method to the yellowfin tuna data in the Indian Ocean for analysis 3 (catch size composition data + CPUE data).

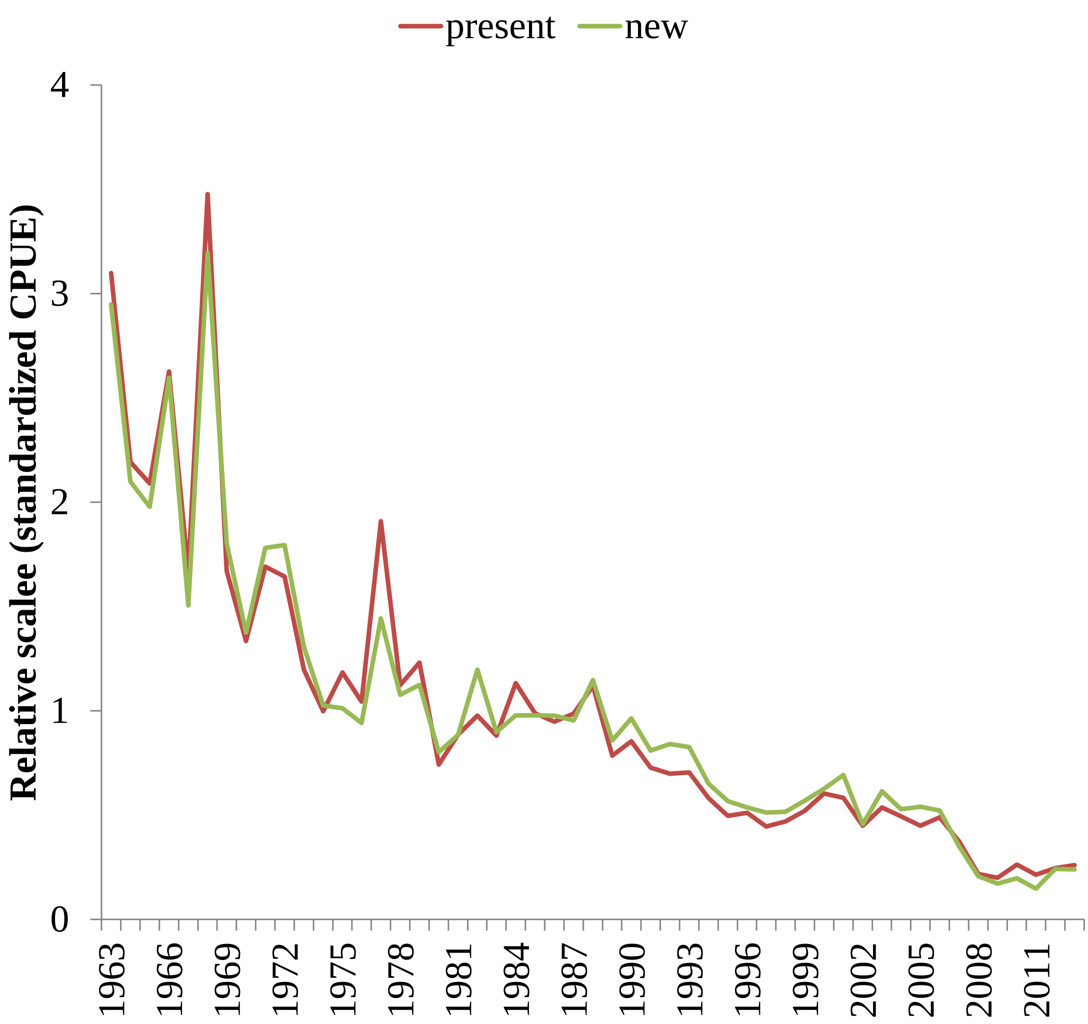

Fig. 6 Comparison of annual standardized CPUE of yellowfin tuna caught by Japanese longline in the Indian Ocean. Red (present sub-area definition), Green (new sub-area definition generated by applying the simultaneous tree method).

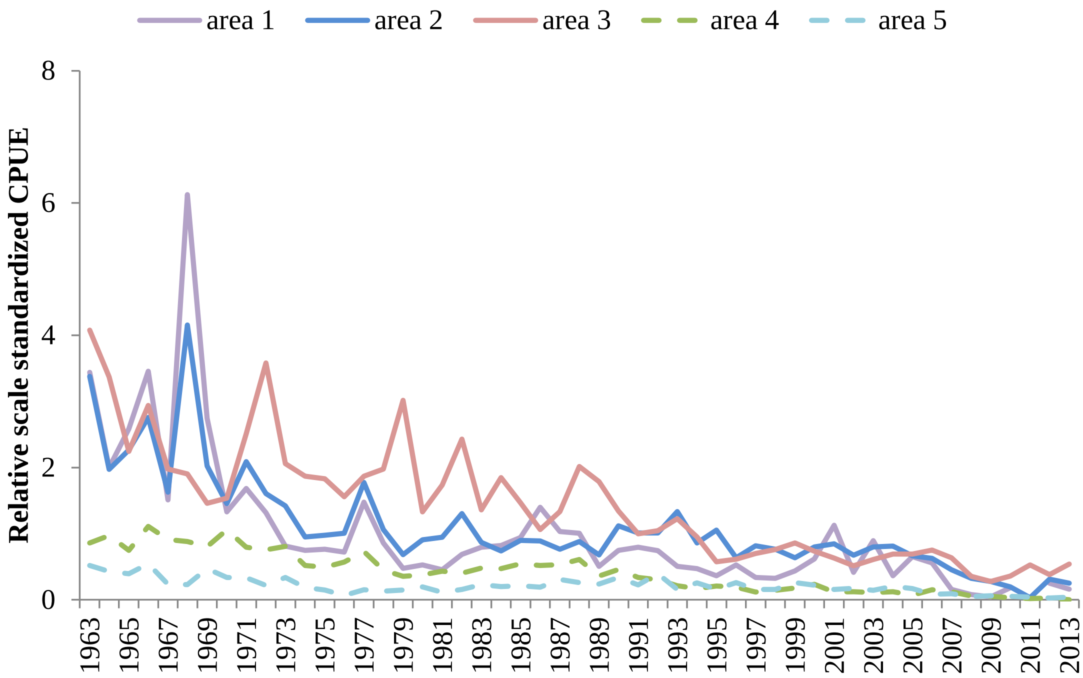

Fig. 7 Annual standardized CPUE by new area definition of yellowfin tuna caught by Japanese longline in the Indian Ocean.

Appendix 1

You can refer Lennert-Cody, CE et al., (2013) for all detail methods and more explanation for the method.

Impurity measure

Size data

The impurity measure for the size data was based on the Kullback–Leibler divergence (KLD; Wang et al., 2005). The KLD is a standard measure for quantifying differences between frequency distributions.

$$\mathrm{I_{KLD}} = \sum_{l}\sum_{j} pl(j) log \frac{pl(j)}{\bar{p}.(j)}$$

where pl(j) is the proportion of fish in the jth length interval of the lth sample, and ¯p.(j) is the average proportion of fish in the jth length interval (average computed over samples in the collection). IKLD puts more weight on the discrepancies of intervals with the greatest proportion of observations.

CPUE data

The impurity measure for the CPUE data was based on a multivariate sum of squares loss function (Yu and Lambert, 1999; Nerini and Ghattas, 2007), modified to address the fact that different grid cells may contain different amounts of data, both in terms of the number of years and in terms of the total number of data points. For a collection of grid cells on the node of a tree, the measure of impurity for the abundance data is given by:

$$\mathrm{I_{SS}} = \sum_{i}\sum_{y=1}^{m-1} ((\Delta\hat{C}i)_y - (\Delta\tilde{C})_y)^2$$

$$\Delta\hat{C}_i = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ & & \dots & & \\ 0 & \dots & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \hat{C}_{i1} \\ \vdots \\ \hat{C}_{im} \end{bmatrix}$$

where Δ is the m-1 x m first-difference matrix and ˆCi is the vector of m annual abundance estimates (the m-year time series of first-differenced relative abundance difference for grid cell i estimated from nominal CPUE and smoothed by a generalized additive model (GAM) on year. More information for GAM analysis is follows). ˜C is the time series of annual abundance estimated from all the data in the collection. $\left(\sum \widehat{\Delta \hat{C}\iota}\right)^{-1}$ is an m−1 by m−1 diagonal matrix with diagonal elements equal to the variance of the ˆCi. Iss is weighted by the variance for giving more influence to those grid cells with smaller variance of ˆCi.

Split criterion for the simultaneous tree method

Binary recursive partitioning (Breiman et al., 1984), applied simultaneously to each data set, was used to build the tree. The combined split criterion measures the improvement achieved by partitioning each of the two data types. The factor for splitting is latitude, longitude and quarter of year.

Size data

$$\text{Imp\_KLD} = n_{left} \sum_j \bar{p}left(j) log \frac{\bar{p}left(j)}{\tilde{p}.(j)} + n_{right} \sum_j \bar{p}right(j) log \frac{\bar{p}right(j)}{\tilde{p}.(j)}$$

where nleft and nright are the numbers of samples in the left and right subgroups, respectively.

CPUE data

$$\text{Imp}_{\text{SS}} = I_{ss;all} - (I_{ss;\,left} + I_{ss;\,right})$$

At each step in building the tree, the split-variable value selected to define a new partition of the data is that which maximizes the following split criterion, which is evaluated over the collection of candidate splits, {split}, for which Imp SS is non-negative:

$$\gamma \left[\frac{Imp\_KLD}{\max(split)(Imp\_KLD)}\right] + (1-\gamma) \left[\frac{Imp\_SS}{\max(split)(Imp_{SS})}\right]$$

where $0 < \gamma < 1$ is a weight that can be used to give added emphasis to one or the other of the two data types, for example, depending on their relative importance in the stock assessment or on data quality. In this analysis I used 0.5 for γ .

Comparison for candidates

It is convenient to rank the candidate stratifications for selecting area definition.

$$\mathrm{U(s)} = \gamma\left[\frac{I_{KLD}(null) - I_{KLD}(s)}{\mathrm{I}_{KLD}(null)}\right] + (1-\gamma)\left[\frac{I_{SS}(null) - I_{SS}(s)}{\mathrm{I}_{SS}(null)}\right]$$

where $I_{KLD}(s)$ is the impurity of the morphometric data for the candidate stratification of size data, $I_{KLD}(null)$ is the impurity from the size data with no stratification, $I_{SS}(s)$ is the impurity for the abundance data for the candidate stratification of CPUE data, and $I_{SS}(null)$ is the impurity for the abundance data with no stratification. U(s) increases to 1.0 as the explanatory ability of each of the two components increases.

GAM set-up for fitting the smooth model was a follows. The same cardinal spline basis, basis dimension (6 basis functions), knot locations and smoothing parameter (a value of 1.08) were used for each cell. The smooth model was fitted to the data using the gam function of the mgcv package in R (with bs="cr"; see Wood, 2006). The common smoothing parameter was selected in the method described in Lennert-Cody, CE et al., (2013).

Adjust length frequency

To be consistent with quarterly time step of the tuna stock assessment model, the length frequency of each monthly length-frequency sample were 'grown' or 'shrunk' to the middle month of each quarter by adding or subtracting a monthly length increment, where applicable (the middle month of each quarter requires no adjustment). Growth model in Eveson et al., (2014; Team 2 in Table 2) for the adjustment was used.