ISC/23/ALBWG-01/06

Spatiotemporal definitions of Taiwanese albacore longline fishery in the North Pacific Ocean based on a regression tree analysis of size data

Zi-Wei Yeh, Yi-Jay Chang, and Jhen Hsu

Institute of Oceanography,
National Taiwan University, No.1, Sec. 4,
Roosevelt Road, Taipei 106, Taiwan
Email: yjchang@ntu.edu.tw

Abstract

The objective of this study is to assess the suitability of the 25°N boundary used to delineate the Taiwanese longline fleet in the albacore tuna assessment. The study analyzed the size composition of albacore tuna caught by the Taiwanese longline fishery from 1995 to 2021, utilizing a multivariate regression tree model. The analysis was conducted with a minimum spatial-temporal resolution of season and 5◦ area. The study findings revealed that using latitude 25°N as the boundary had the highest explainable variation and was most effective in capturing the spatial differences in the size composition of the albacore. The southern group exhibited larger albacore sizes, with an average weight of 22.3 ± 5.8 kg, while the northern group had smaller albacore sizes, with an average weight of 13.8 ± 4.3 kg. Therefore, the study concluded that using latitude 25°N as the boundary could best reflect the spatial differences in the size composition of albacore caught by the Taiwanese longline fleet.

1. Introduction

The spatial stratification of the previous albacore assessment was defined as the northern (areas 3 and 5) and southern (areas 2 and 4) areas separated at 30°N based on the analyses of fishing operations and size composition data from Japanese and US longline vessels (ISC, 2020). Such of a definition was justified by a relatively consistent size distribution of albacore among various fleets. A previous study has shown that the Taiwanese longline fishery targeting albacore tuna mainly operated in the north of 25°N (Chen and Cheng, 2013). The boundary of 25°N was used to define the Taiwanese longline fleet in the assessment (ISC 2020). However, there is still a lack of scientific argument about the used boundary. This study aims to analyze the size data of albacore tuna in Taiwanese longline fishery using a regression tree approach to examine whether the 25°N boundary is appropriate.

2. Materials and methods

2.1 Data sources

This study used the longline fishery logbook data (1995 - 2021) obtained from Oversea Fisheries Development Council (OFDC), Taiwan. To calculate the mean weight for each fishing set, the total weight of each vessel during a fishing day, as recorded in the logbook data, was divided by the number of albacore tuna catch. In order to explore the distribution of weight composition in space and time, the minimum spatial-temporal resolution used for the analysis is by season and 5◦ area. Mean weight data were binned into: ≤ 5 kg , 6 -10 kg , 11-15 kg , ..., > 40 kg. The proportion of fish per weight interval was computed by each seasonal 5◦ × 5◦ area sample, which creates a multivariate response with 9 values per sample.

2.2 Analysis

In this study, referring to the size composition analysis method of Lennert-Cody et al. (2010) for yellowfin tuna, the multivariate regression tree model was used to classify the mean weight of albacore tuna for the fisheries definition of the Taiwanese longline fleet and to examine whether the 25°N boundary is appropriate. The method explores a series of binary decision rules to divide the data into smaller and more similar groups using cyclic partitioning. The grouping frequency is measured by the Kullback-Leibler divergence of node heterogeneity. There is no pruning step in this method because the purpose of this method is to explore data on a large scale and build small tree models, rather than to find the best predictive model (Breiman et al., 1984).

To explore the spatiotemporal variation of the albacore size composition, this study considered four variables: year, season, longitude, and latitude in the regression tree analysis. In this study, we evaluated various numbers of fisheries definitions (*k* = 2 - 4) based on the explainable variation to find the most appropriate number of groups and group boundaries. All data analysis in this study was performed through R 4.2.2 (R Core Team, 2022), and the "FishFreqTree" package was used for multiple regression tree analysis (Xu, 2022).

3. Results and discussion

The mean weight frequency distribution and the mean weight under 5◦ ×5◦ areas was shown in Figure 1 and Figure 2, respectively. The results of the multiple regression tree model showed that when the number of fisheries is 2 (*k* = 2), using the latitude 25°N (Figure 3) as the boundary line had the highest explainable variation (Table 1).

The mean weight distribution by groups was shown in (Figure 4) that fish size in the southern area is larger the that of the Northern area. The model explainable variation is 19.4% when *k* is equal to 3 (Table 1) in which the southern group can be further divided into two groups at the longitude 165°W (Figure 5). For the southern group, fish size in the area west of 165°W is larger than that of east of 165°W. The model explainable variation could be further improved (20.5%; Table 1) by increasing the number of fisheries (*k* = 4) that the northern group can be further divided into two groups based on longitude 175°W (Figure 7). However, the mean weight composition between the two northern groups are relatively similar (Figure 8).

Overall, none of the year and season predictors was selected in the analysis given the number of fisheries of 4. All of the fisheries definitions (i.e., when *k* = 2, 3, 4) supported the split of the northern and southern groups at the latitude of 25°N. The southern group has a larger size, with a mean weight of 22.3 ± 5.8 kg compared to that of the northern group (13.8 ± 4.3 kg). The result of this study is consistent with the previous study by Chen and Cheng (2013). We concluded that using latitude 25°N as the boundary could best reflect the spatial differences in the size composition of albacore caught by the Taiwanese longline fleet.

4. References

Breiman, L., and R. Ihaka. (1984). Nonlinear discriminant analysis via scaling and ACE. Davis One Shields Avenue Davis, CA, USA: Department of Statistics, University of California.

Chen, C. Y. and F. C. Cheng. (2013). Development of Taiwanese albacore-targeting longline fisheries in the North Pacific Ocean, 1995-2010. ISC working paper, ISC/13/ALBWG/16.

Hastie, R. (2001). Problems for judgment and decision making. Annual Review of Psychology 52(1), 653-683.

ISC albacore WG. (2020). Stock assessment of albacore tuna in the North Pacific Ocean in 2020.

Lennert-Cody, C. E., M. Minami, P. K. Tomlinson, and M. N. Maunder. (2010). Exploratory analysis of spatial–temporal patterns in length–frequency data: An example of distributional regression trees. Fisheries Research 102(3), 323-326.

Maunder, M. N., P. R. Crone, L. J. Valero, and B. X. Semmens. (2014). Selectivity: theory, estimation, and application in fishery stock assessment models. Fisheries Research 158, 1-4.

Team, R. C. (2022). R: A Language and Environment for Statistical Computing, Reference Index, R Foundation for Statistical Computing.

Xu H, C. Lennart-Cody. (2022). FishFreqTree: IATTC's regression tree R package for fisheries' size frequency data_. R package version 3.3.1.

Table 1. The geographic boundary and the explained variation for each of the fisheries definitions explored in this study by the multiple regression tree analysis based on the Taiwanese longline data during 1995 - 2021.

k	Split 1	Split 2	Split 3	Var_explained
2	lat 25			16.58%
2	lat 20			13.40%
2	lat 30			11.96%
2	lat 15			9.35%
2	lat 10			4.92%
3	lat 25	lon 195		19.38%
3	lat 25	lon 200		19.20%
3	lat 25	lon 190		19.14%
3	lat 25	lon 205		18.80%
3	lat 10	lon 195		18.52%
4	lat 25	lon 195	lon 185	20.52%
4	lat 25	lon 195	lon 190	20.43%
4	lat 25	lon 195	lon 200	20.38%
4	lat 25	lon 195	lon 205	20.37%
4	lat 25	lon 200	lon 185	20.34%

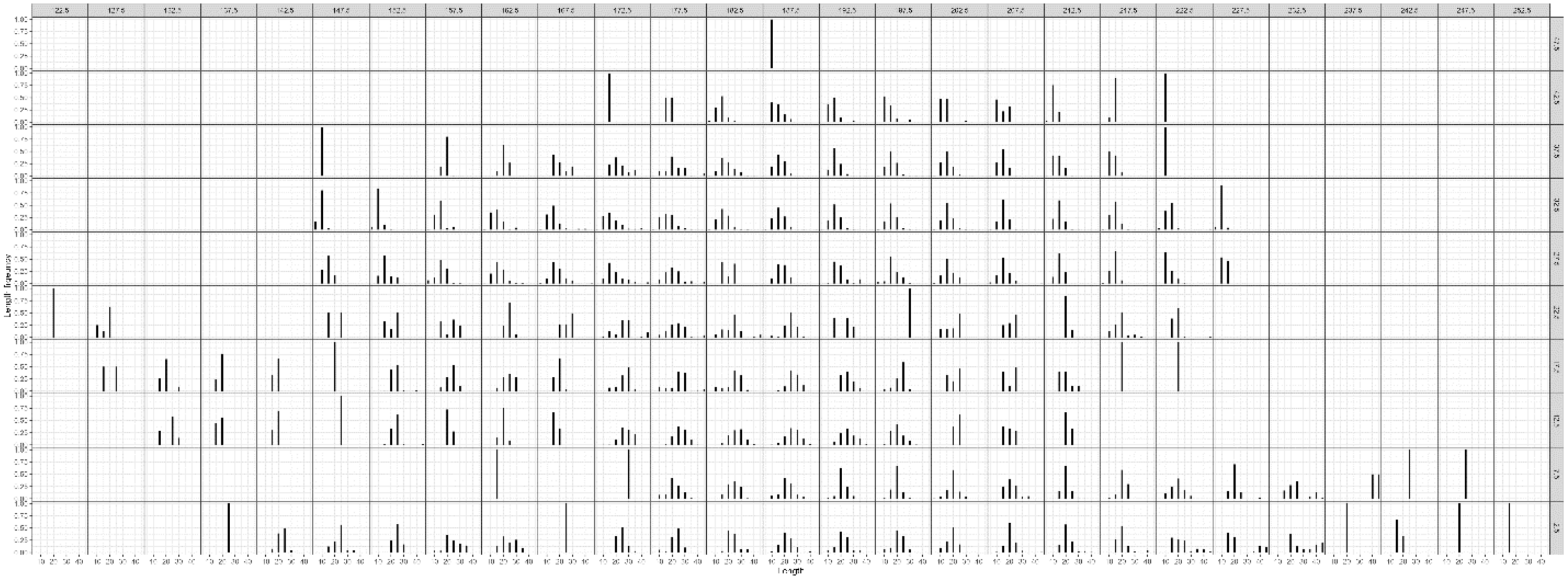

Figure 1. Spatial distribution of mean weight composition of albacore tuna based on the Taiwanese longline data during 1995 - 2021.

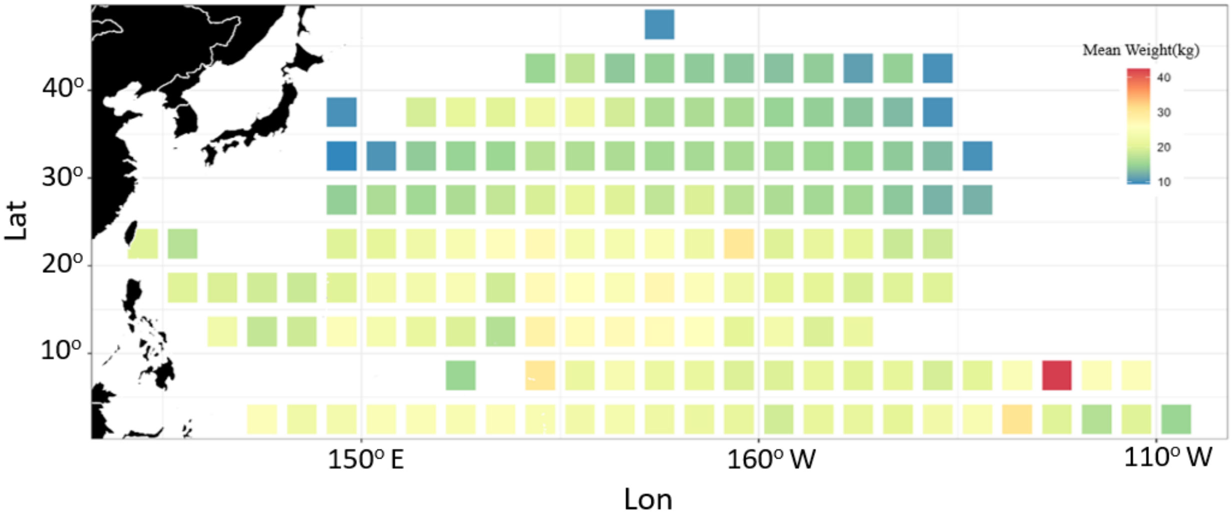

Figure 2. Spatial distribution of mean weights of albacore tuna based on the Taiwanese longline data during 1995 - 2021.

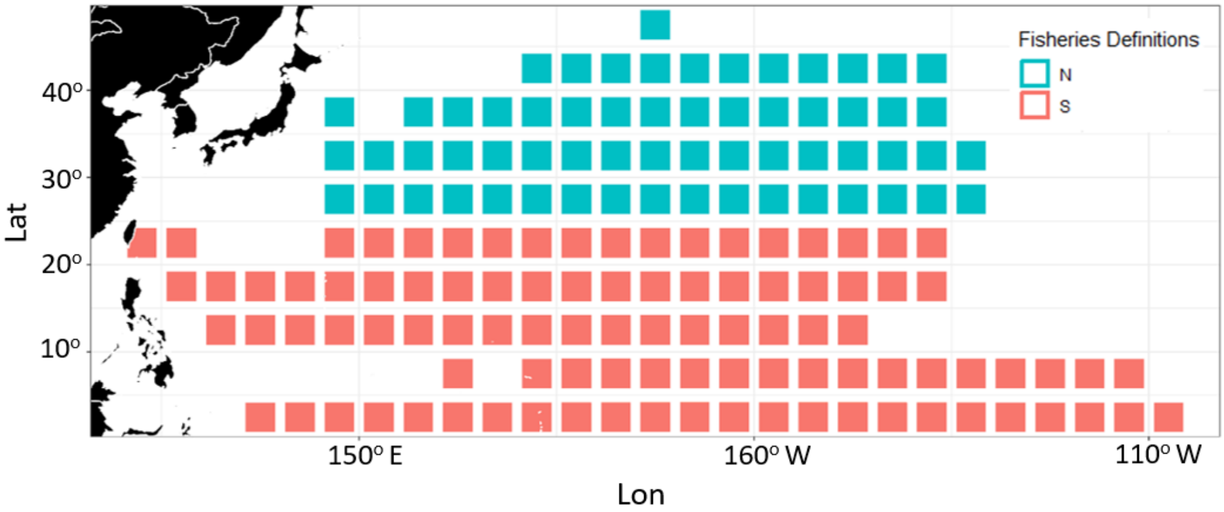

Figure 3. Spatial distribution of the classified fisheries by the multivariate regression tree ($k = 2$) based on the Taiwanese longline data during 1995 - 2021.

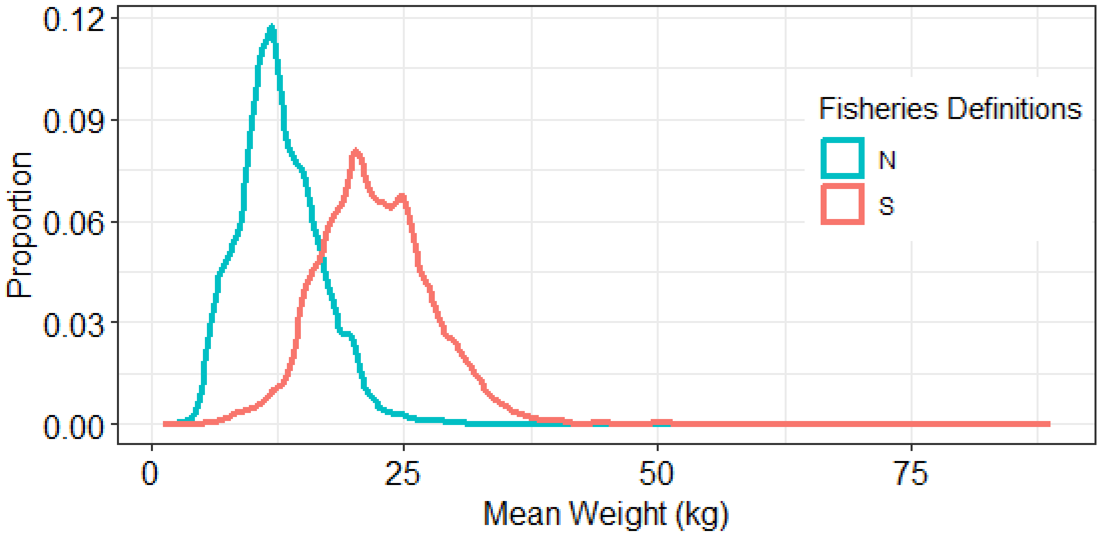

Figure 4. Distribution of the mean weight of albacore tuna of the classified fisheries by the multivariate regression tree (*k* = 2) based on the Taiwanese longline data during 1995 - 2021.

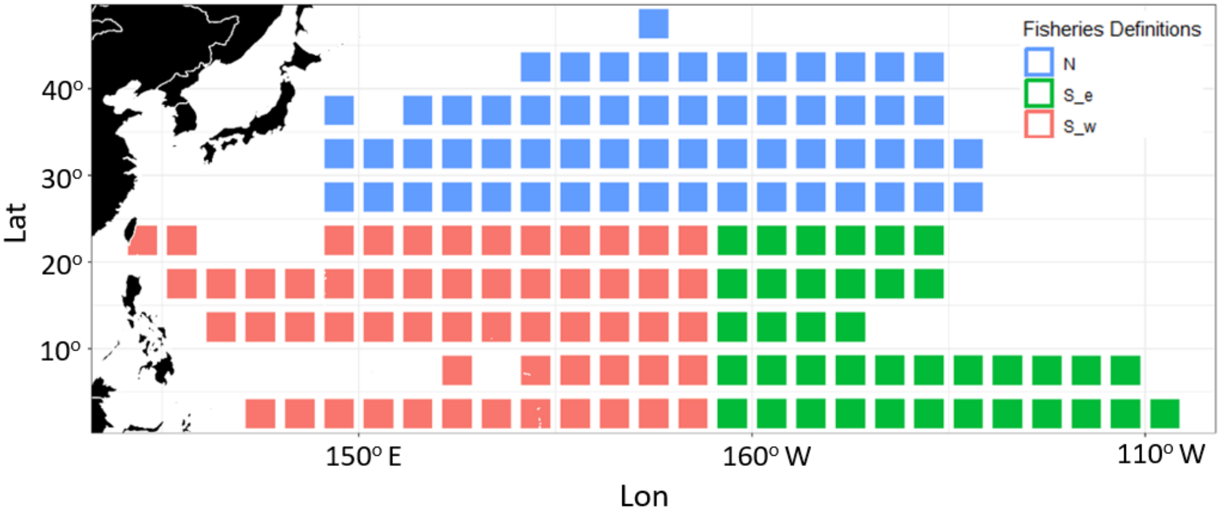

Figure 5. Spatial distribution of the classified fisheries by the multivariate regression tree (*k* = 3) based on the Taiwanese longline data during 1995 - 2021.

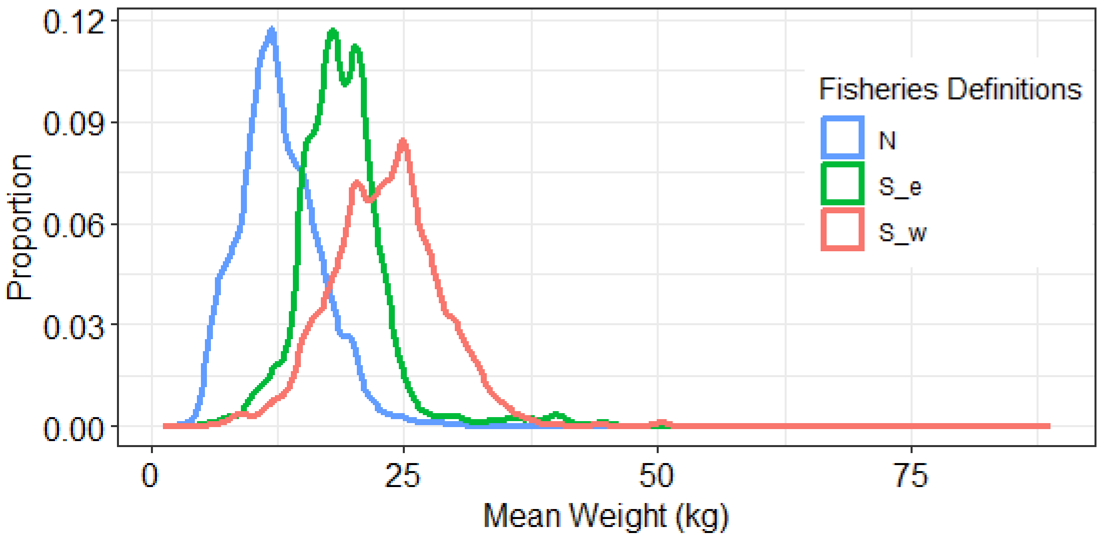

Figure 6. Distribution of the mean weight of albacore tuna of the classified fisheries by the multivariate regression tree (*k* = 3) based on the Taiwanese longline data during 1995 - 2021.

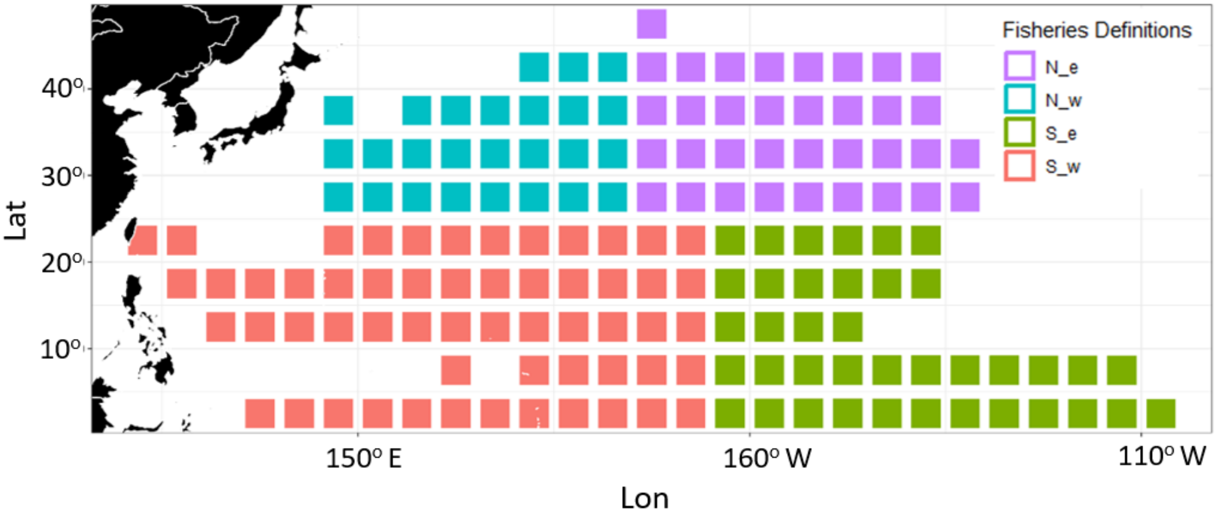

Figure 7. Spatial distribution of the classified fisheries by the multivariate regression tree (*k* = 4) based on the Taiwanese longline data during 1995 - 2021.

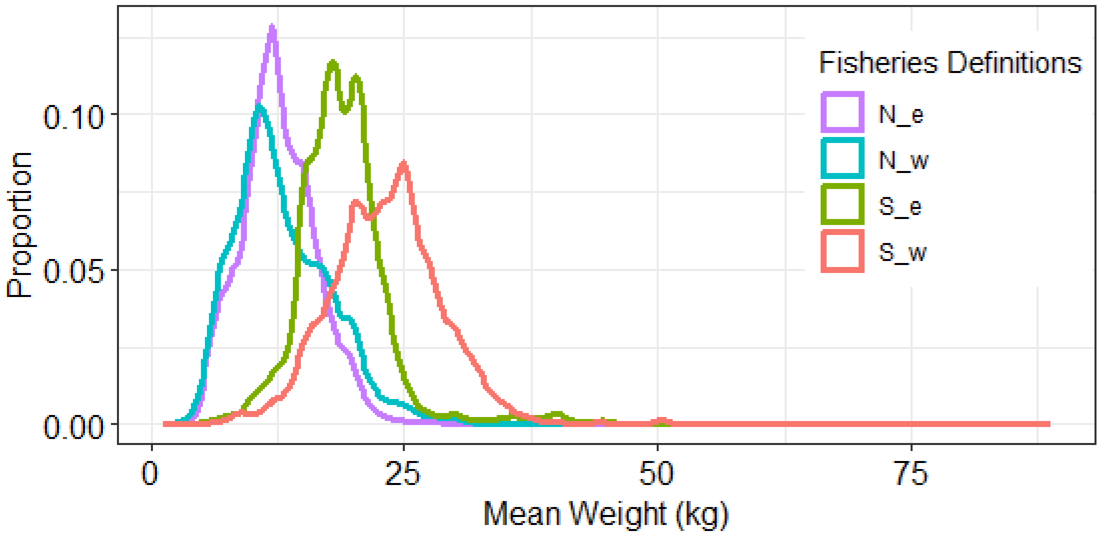

Figure 8. Distribution of the mean weight of albacore tuna for each group classified by the multivariate regression tree (*k* = 4) based on the Taiwanese longline data during 1995 - 2021.

Appendix

Table S1. Mean, standard deviation (SD), the maximum (Max), the minimum (Min) of albacore mean weight, and the number of observations (*n*) by various fisheries definitions and without classification (all) derived from the regression tree analysis based on the Taiwanese longline data during 1995 - 2021.

Fisheries definitions	Mean (kg)	SD (kg)	Max (kg)	Min (kg)	*n*
N	12.84	4.30	88.78	1.20	31,932
S	22.25	5.76	74.29	2.20	13,175
S_e	19.12	5.10	74.29	5.00	3,199
S_w	23.26	5.59	60.52	2.20	9,976
N_e	12.69	3.87	88.78	2.67	22,265
N_w	13.18	5.14	59.90	1.20	9,667
all	15.59	6.41	88.78	1.20	45,107