Additional analysis for Japanese longline logbook data using finite mixture mode.[1]

Hirotaka Ijima and Yuichi Tsuda

Highly Migratory Resources Division, Fisheries Stock Assessment Center, Fisheries Resources Institute (FRI), Japan Fisheries Research and Education Agency. 2-12-4 Fukuura, Kanazawa-ku, Yokohama, Kanagawa, 236-8648, JAPAN

ijima_hirotaka69@fra.go.jp

This working paper was submitted to the ISC Albacore Working Group Intercessional Workshop, 6-12 December 2022, held at the Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Kanagawa, Japan.

Abstract

We analyzed Japanese longline logbook data using a finite mixture model to improve the fisheries definition in the North Pacific albacore stock assessment. Before the data-preparatory meeting, we reported the progress of this analysis to the ISC ALBWG, and the ALBWG has requested several issues to improve the analysis. This paper reports our work and the results based on this request from the ISC ALBWG. The additional analysis isolated potential cohorts distributed in Area 1 and Area 3. In addition, we updated the length composition data and compared it to the analysis result. The increase in the length composition data has made the distribution of length composition data relatively similar to the distribution of body mass data in the logbooks. Both the body mass from the logbook and length composition data has different problems. However, at least the length composition data are measured length by individual fish. Thus, we suggest using length composition data for the SS3 model.

Introduction

Conducting the North Pacific albacore stock assessment, the ISC ALBWG uses length composition data to perform cluster analysis and define the fishery using an area-as-fleet approach (ISC 2020). However, the ALBWG recognizes the potential inclusion of multiple cohorts in some Japanese longline fisheries and the need to remedy this problem (ISC 2020). To divide such latent cohorts, we attempted to construct a finite mixture model using mean albacore body mass per operation, obtained from logbook data rather than length composition data, to establish clusters by operation. We first addressed a preliminary analysis and reported it to the ALBWG via web meeting. After a discussion via web meeting, we received the following request from the ALBWG.

1. Given that the WG is uncertain about the new data analysis presented and more work is required before using it in the 2023 assessment, the WG recommends that Japanese scientists also repeat and update the data preparation performed for the 2020 assessment and present that during the upcoming data preparation meeting.

2. The WG recommends that the Japanese scientists use a consistent area over time for catch and effort when developing abundance indices.

3. The WG recommends that the Japanese scientists explain the source of the total weight by set in the logbooks.

4. If the WG does agree to use the average weight data from logbook, the WG recommends that the WG consider whether to use that data as weight or length data.

5. If the average weight data can be used, the WG recommends that the cluster analysis concentrate on Areas 1 and 3 to develop the juvenile fleet. The fleet structure for the other areas will remain the same for the 2023 assessment.

6. The WG recommends that the Japanese scientists examine the posterior probabilities in the cluster analysis.

7. The WG recommends that the Japanese scientists examine the port sampling data from Yaizu port in more detail and report to the WG at the next data prep meeting.

8. The WG recommend that the average weight data from the logbooks be compared with available observer data from the same sets.

9. The WG recommend that all three length data sources (avg weight, observer, & port sampling) be compared on an aggregate (raised to the catch) and/or 5x5xquarter basis to understand the representiveness of the data. This working paper summarized the methods and overview of the preparation of the logbook and length composition data to answer these requests. We also explored additional finite mixture model analyses using Area 1 and 3 data.

We addressed answer requests 3 – 9. In detail, this working paper summarized the methods and overview of the preparation of the long line logbook and length composition data to answer these requests and added new length composition data. We also explored additional finite mixture model analyses using Area 1 and 3 data.

Material and methods

Data sources

This study used Japanese longline logbooks and length composition data. The Japanese longline logbook reports the number of albacores caught and the total catch mass per operation. The number of fish caught is counted directly during an onboard take. Catch mass per operation is calculated in two ways. First, the vessel captain determines the eye-measured mean body mass of the albacore during the operation and multiplies it by the total number of fish caught. The second method is dividing the voyage's total catch mass measured at the time of landing equally among the number of operations. The logbook data did not indicate which method was used to calculate catch mass. In this paper, we organized the mean body mass per operation by voyage, and voyages with zero variance were excluded from the analysis, as they were judged to have distributed the total catch weight equally (Figure 1). About 80% of logbook data recorded eye-measured albacore mass (Answer to Request 3). In addition, analyses with the finite mixture model were conducted only for areas 1 and 3 (Answer to Request 5, Figure 2).

We summarized size composition data to compare to finite mixture model analysis results (Answer to Request 9). The length composition data logged the fork length of the albacore and body mass since 1962. After 1998, the database format was changed, and individual-based length and body mass measurements were available. Each data set includes multiple data sources: port sampling, training or prefecture vessels, and observers (Answer to Request 7). The port name has been recorded since 1986. In this study, we added length composition data that the prefecture government has observed. We used length composition data recorded at a resolution of 5 X 5 degrees grids or better to directly evaluate the accuracy of the mean body mass in the logbook data. The body mass records are minimal in this study area. Thus, we used only length composition data for this comparison. The observer data was unreported in Areas 1 and 3 (Answer to Request 8).

Finite mixture model

We constructed a finite mixture model to separate the multiple cohorts potentially distributed in Areas 1 and 3 (Answer to Request 9, Figure 2). We used R software package "flexmix" to construct finite mixture model (Gruen et al. 2015). The variable used for analysis was the albacore mean body mass per operation, and that distribution was assumed to be the gamma distribution because mean body mass is a continuous value greater than zero. We selected the best model from several grouping factor assumptions by the Bayesian information criterion (BIC). The candidate grouping factors were 5 x 5 grid, Month-5 x 5 grid, Year-5 x 5 grid, and Year-Month-5 x 5 grid. The initial values of the clusters were set to 1-6. We also plotted histograms for each estimated cluster, and length composition data were aggregated with the same grouping factor and compared to the estimated length distribution. The posterior probabilities of clusters were further illustrated to determine the optimal number of clusters (Answer to Request 6).

Result and discussion

Based on the comparison of BIC, estimated cluster distribution, and posterior probabilities, we selected a grouping factor of Year-Month-5x5 grid and two clusters model. The BIC became smaller as the initial number of clusters was increased, with all seasons having the grouping factor Year-Month-5x5 grid being the smallest (Figure 3). When we organized the results for the grouping factor Year-Month-5x5 grid, the histograms had extensive overlapping, and the differences were not apparent except for the results with the number of clusters set to 2 (Figure 4). Year-Month-5x5 grid model, the posterior probabilities of clusters decreased in the frequency of zeros and ones as the number of clusters increased (Figure 5 - Figure 8). This result suggests that the accuracy of the estimated results decreases as the number of clusters increases. A spatial plot of the mean body mass for each classified cluster shows that the two size classes appear to be intermixed in each season (Figure 9).

Clusters based on the Year-Month-5x5 grid grouping factor obtained from this analysis were applied to mean body mass from the logbook and the length composition data from several sources (Figure 10). Comparing the two data sets' sizes showed similar distributions, unlike the preliminary analysis results (Figure 10). This may be due to the increase in length composition data. Summarizing the size data by port showed a similar trend to the aggregated results. However, cluster 2 in the third quarter is not mono-modal and did not adequately separate the cohorts (Figure 11).

When a similar analysis was conducted for the North Pacific marlin, the body mass information in the logbook and the port sampling results were almost identical (Ijima 2019). The discrepancy may be smaller in the marlin case because most catches are zero or one. Currently, it is impossible to determine which data source is more accurate, the port sampling data or the mean body mass data from the logbook. Both data sources have their problems. For example, the body mass information in the logbook provides an approximate, individual weight by wider-range fishing operations. However, it is based on visual inspection and only sometimes yields the same size albacore in one operation. Port sampling has repeated measures problems that are not sampled associated with the catch amount by season and area. Such repeated measures problems might be solved by resampling Bootstrap (Stewart and Hamel, 2014).

Based on the results of these analyses, we propose the following SS3 setup.

1. The body length data used in SS3 is based on port sampling statistics because the measurement error is low, although there is a possibility of sampling bias.
2. If the ALBWG uses the result of this analysis to define the SS3 fleets, the ALBWG has 7 or 8 fleets in areas 1 and 3.

References

ISC 2020. Stock assessment of Albacore tuna in the North Pacific Ocean in 2020. ISC/20/ANNEX/12

Gruen, B., Leisch, F., Sarkar, D., Mortier, F., Picard, N. and Gruen, M.B., 2015. Package 'flexmix'.

Ijima., 2019.Size-dependent distribution of Pacific Striped Marlin (*Kajikia audax*) : The analysis of Japanese longline fishery logbook data using the finite mixture model. ISC/19/BILLWG-1/09.

Stewart, I.J. and 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), pp.581-588.

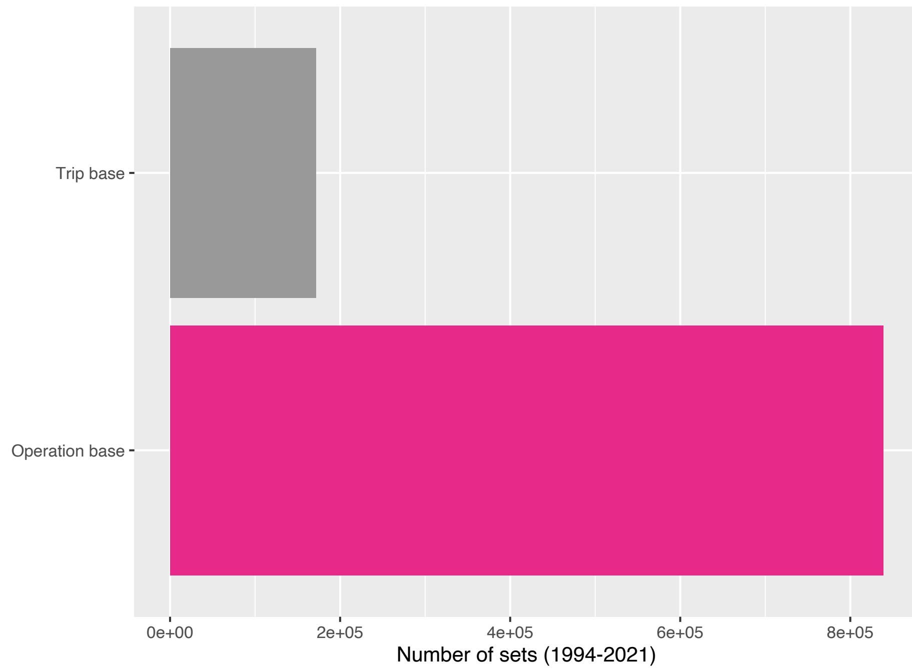

Figure 1. Calculation methods of catch mass by one operation. Operation base: eye-measured body mass multiplied by the number of fish caught. Trip base: dividing the voyage's total catch mass measured at the time of landing equally among the number of operations.

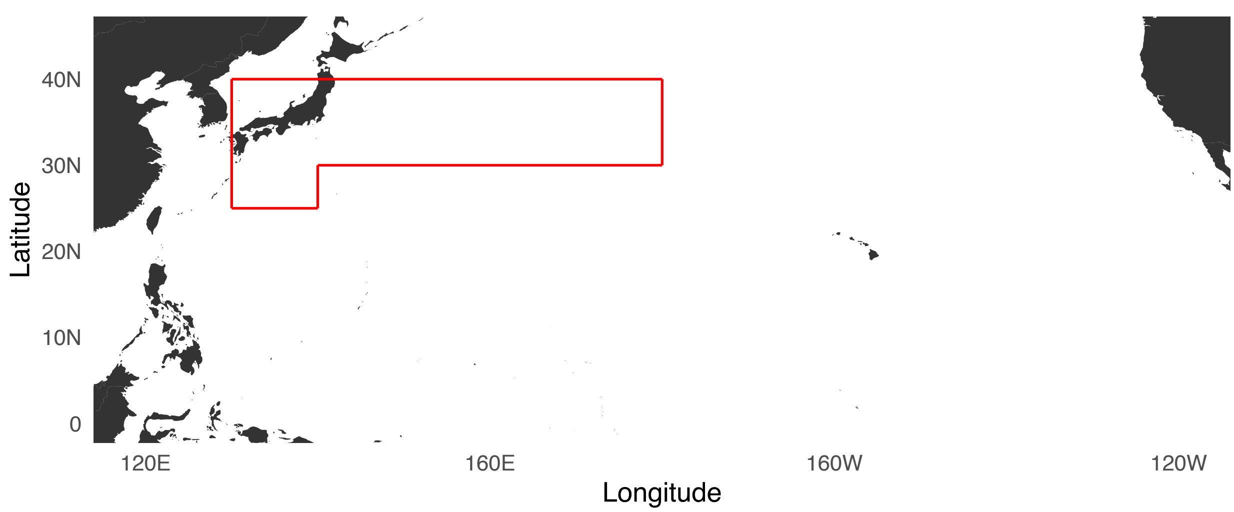

Figure 2. Analysis area for the finite mixture model analysis.

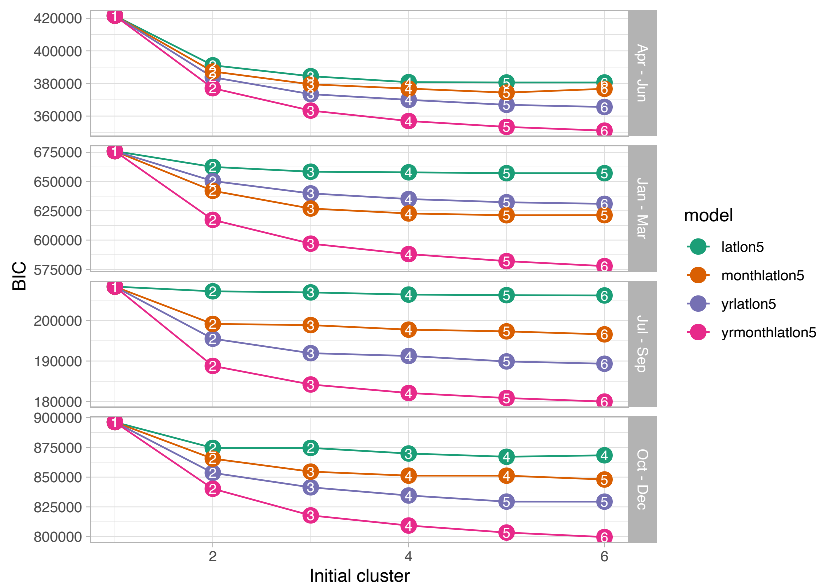

Figure 3. The Bayesian information criterion (BIC) for the different model assumptions.

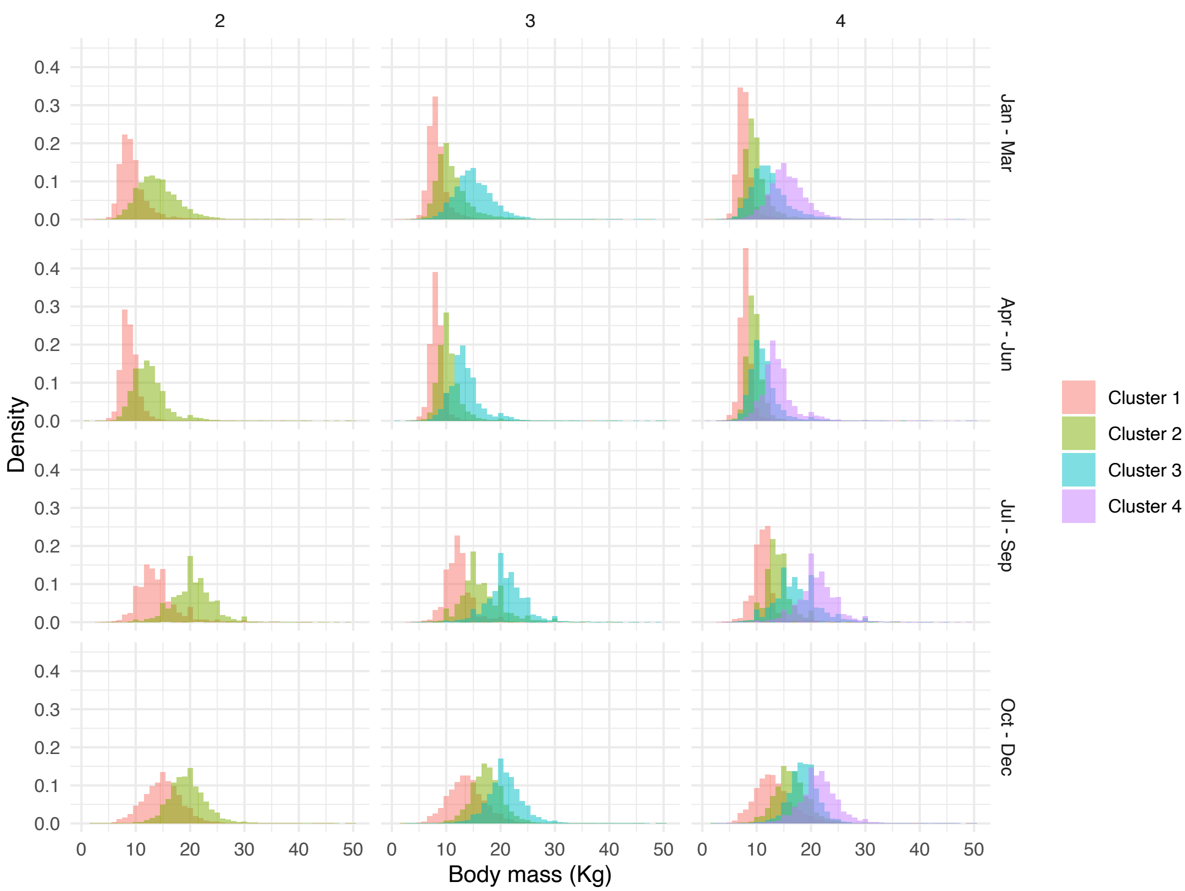

Figure 4. Albacore body mass by one operation summarized by the estimated clusters.

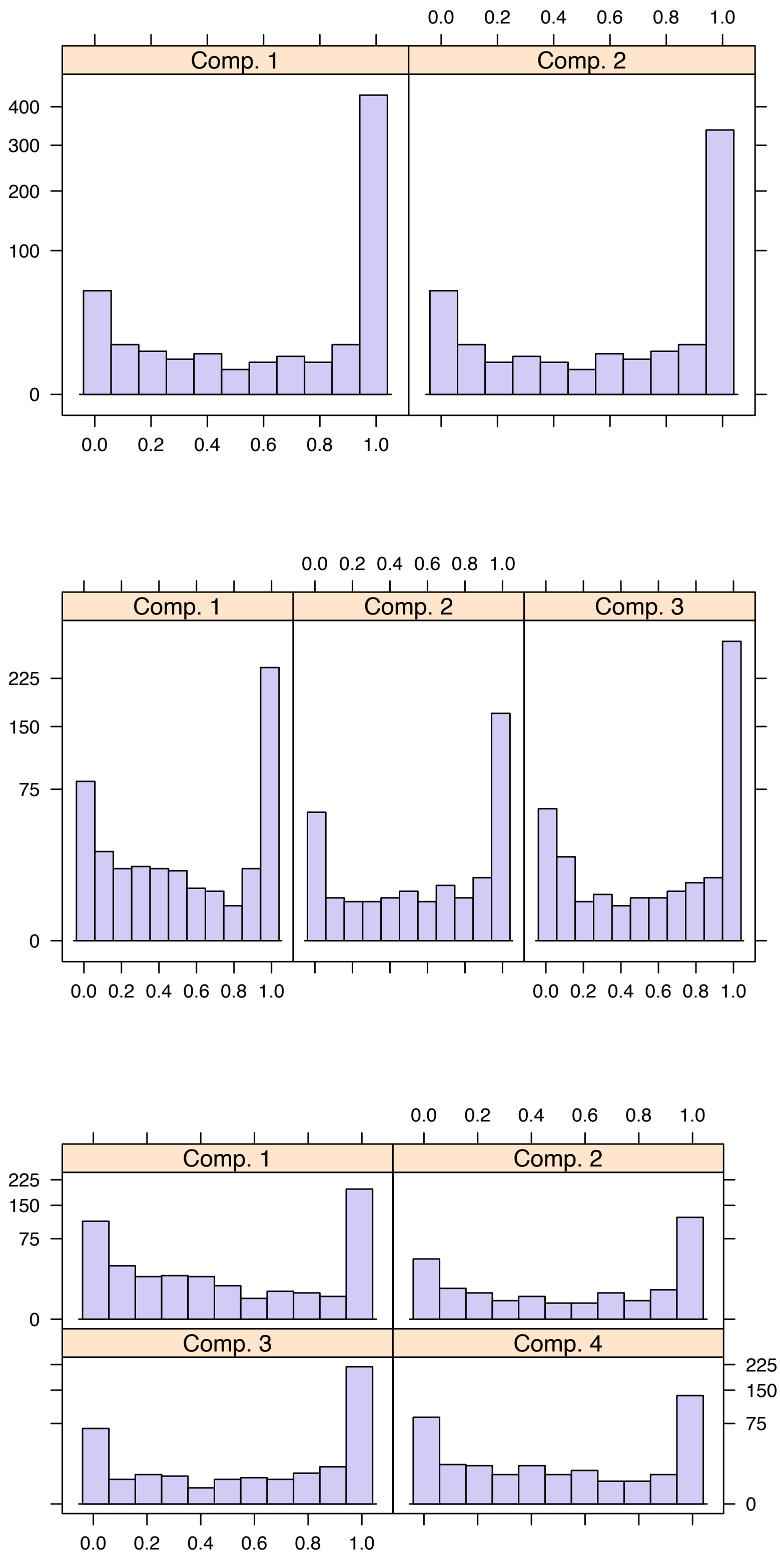

Figure 5. The posterior probability for each cluster in quarter 1.

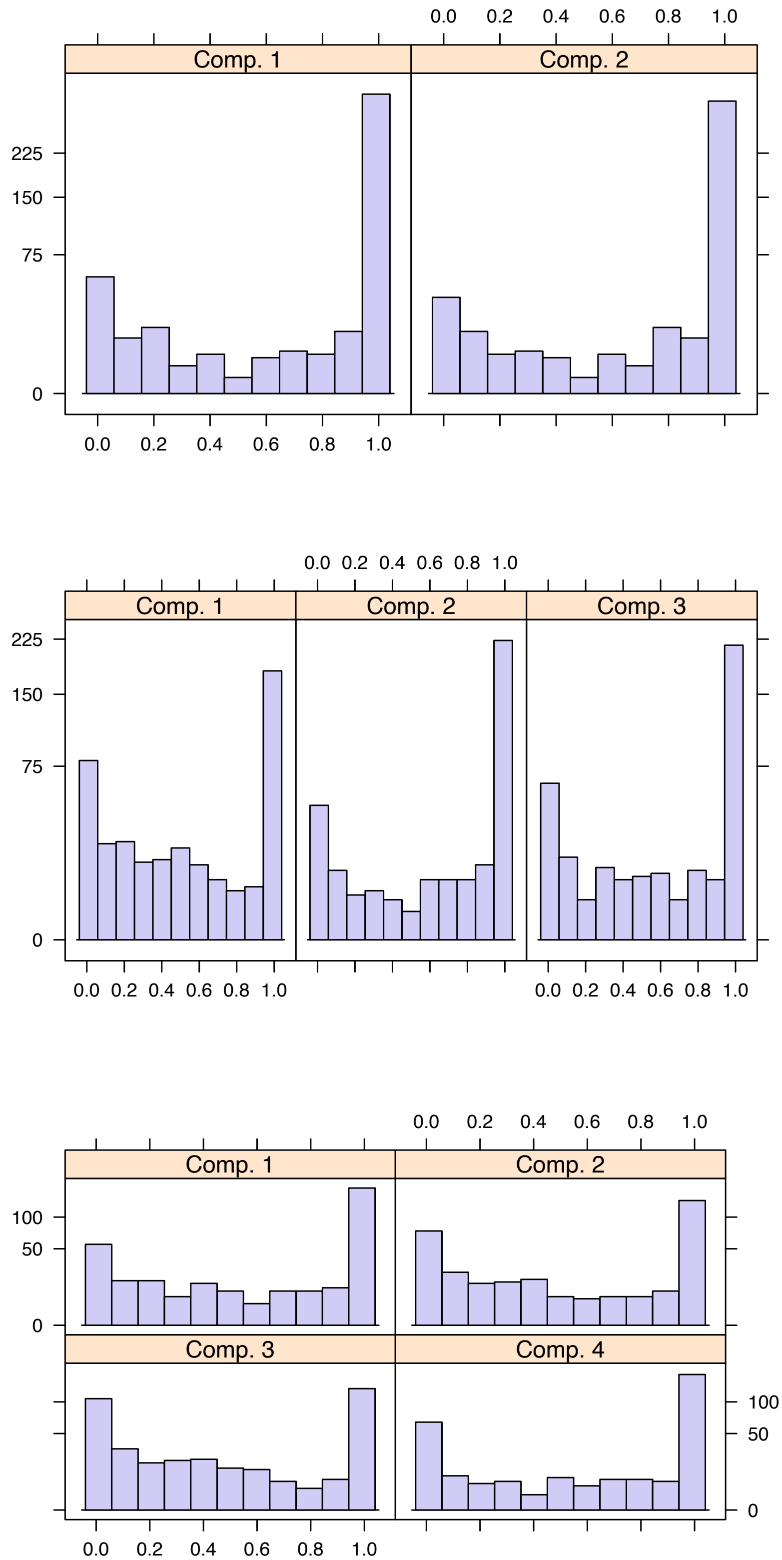

Figure 6. The posterior probability for each cluster in quarter 2.

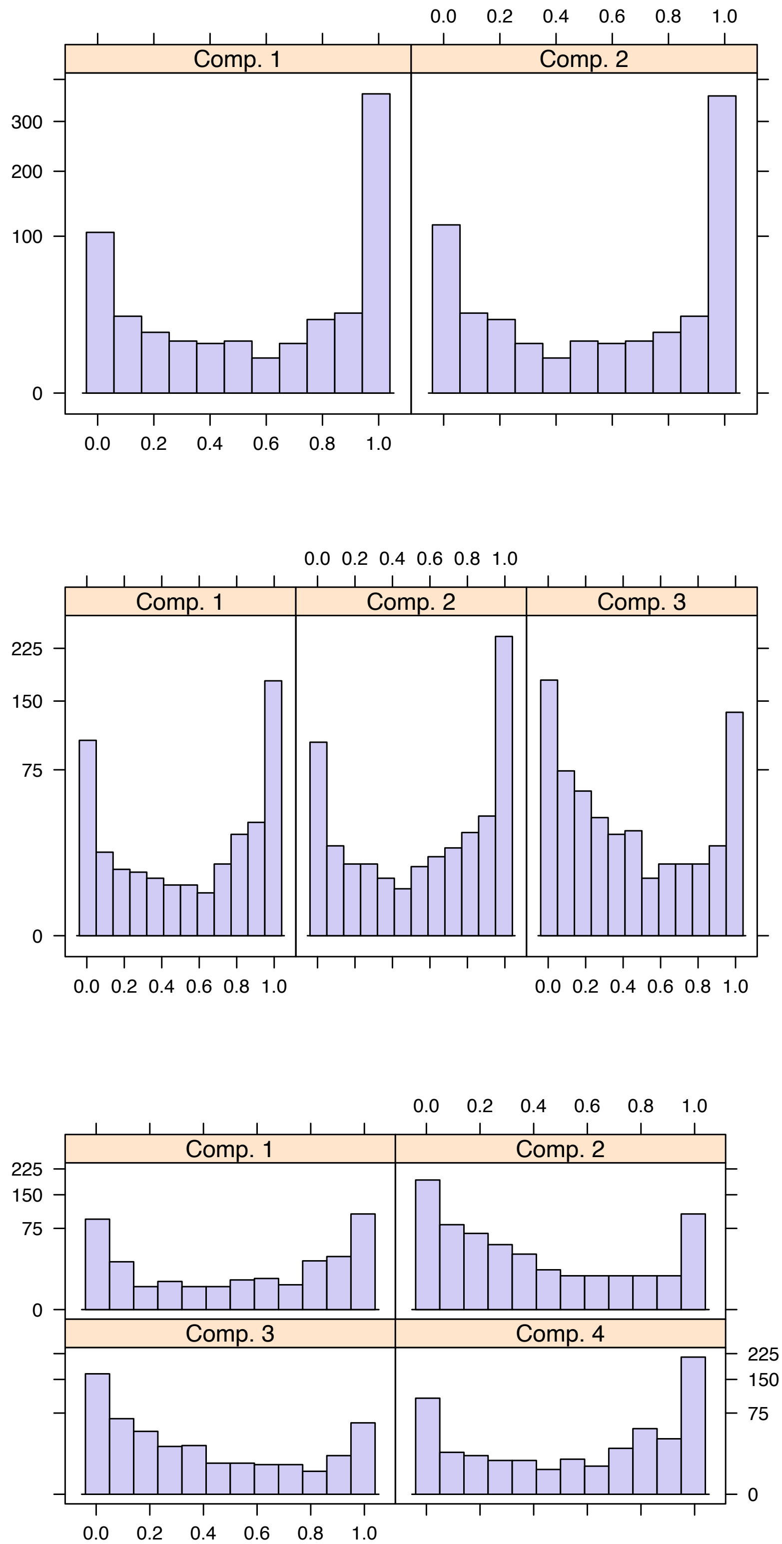

Figure 7. The posterior probability for each cluster in quarter 3.

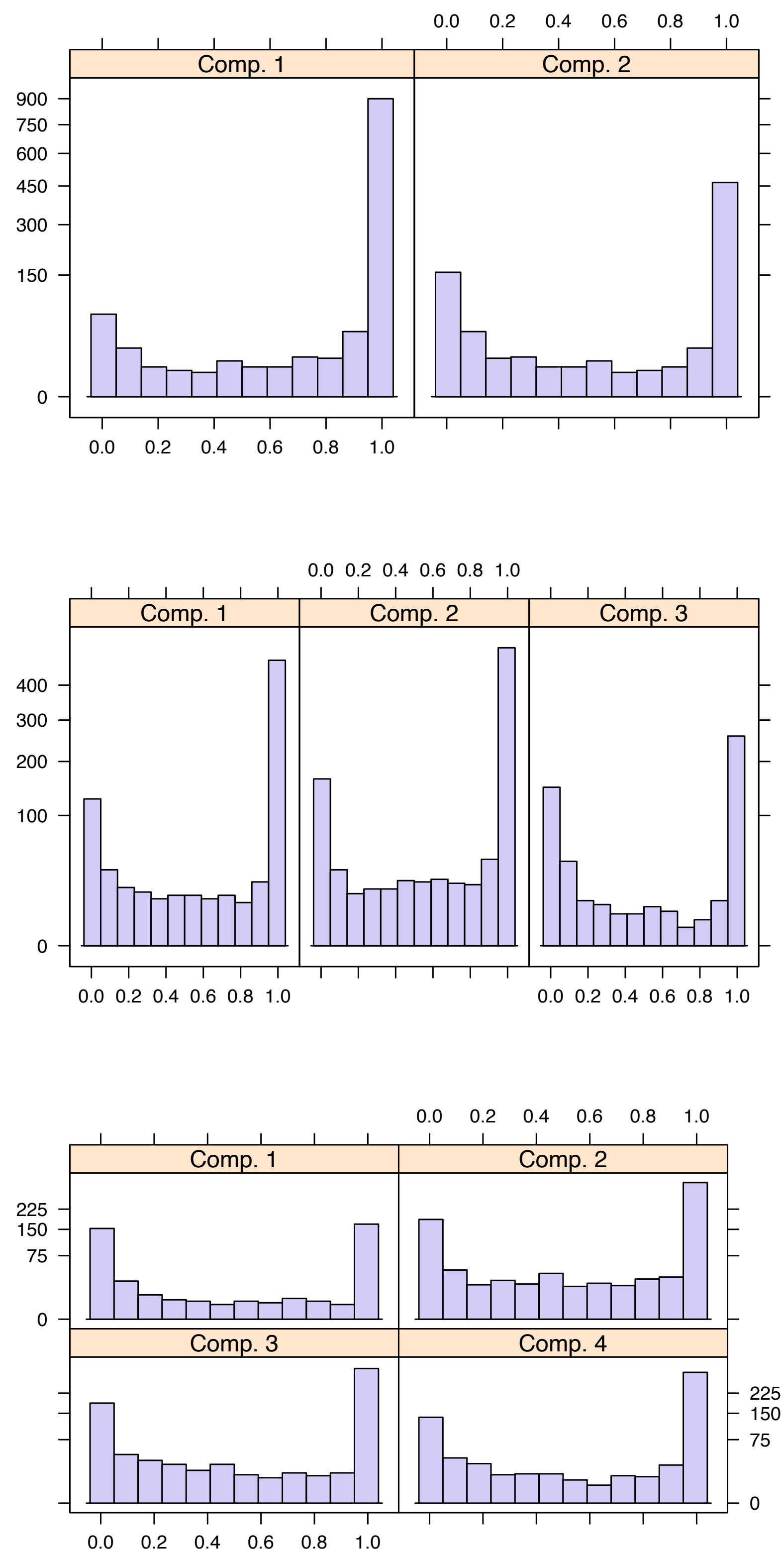

Figure 8. The posterior probability for each cluster in quarter 4.

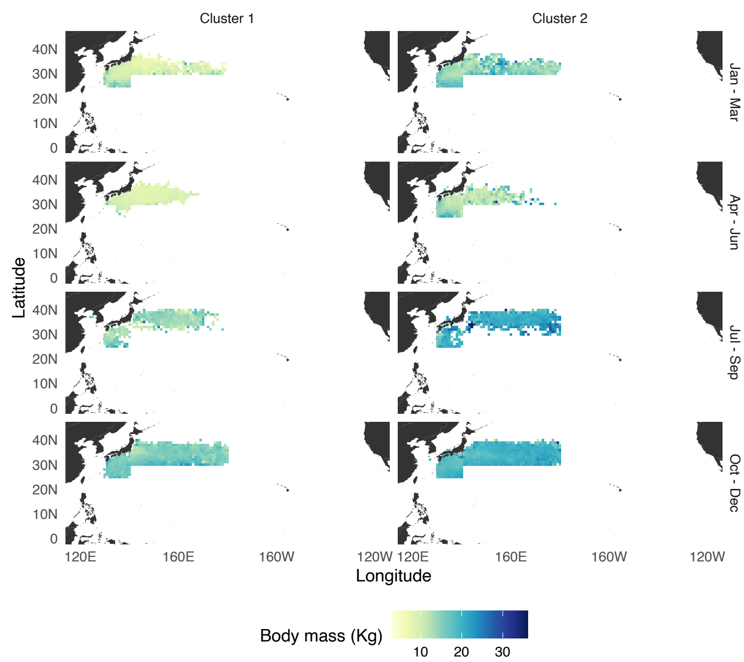

Figure 9. Albacore body mass from Logbook data summarized by 1 x1 grid area and two clusters.

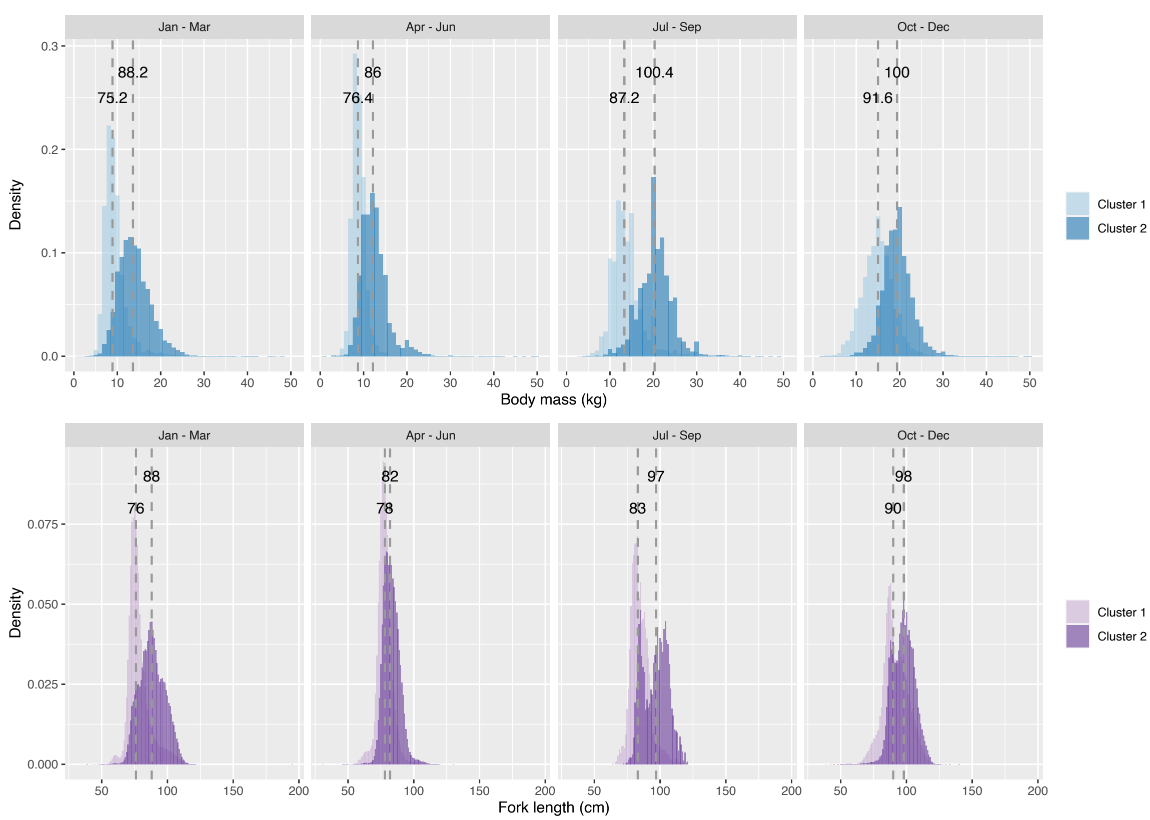

Figure 10. Comparison logbook data and length composition data summarized by Finite mixture model analysis result cluster.

Figure 11. Length composition data summarized by different sampling method. Length composition data was sampled in Area 1 and 3 between 1994-2021. Two clusters correspond to finite mixture model analysis using logbook data.