ISC/24/ALBWG-01/01

Updated analysis for Japanese longline logbook data by finite mixture model

Naoto Matsubara, Makoto Nishimoto, Yuichi Tsuda and Hirotaka Ijima

Highly Migratory Resources Division, Fisheries Resources Institute,
Japan Fisheries Research and Education Agency (FRA)

2-12-4, Fukuura, Kanazawa-ku, Yokohama,
Kanagawa 236-8648, JAPAN

Email: matsubara_naoto84@fra.go.jp

This working paper was submitted to the ISC Albacore Working Group Intercessional Workshop, 11-19 March 2024, held at Fisheries and Oceans Canada, Government of Canada.

Summary
This document summarized for the information of new fleet definition by clustering of finite mixture model to input the future stock assessments. We reanalyzed Japanese longline logbook data using the finite mixture model to improve the fisheries definition in the North Pacific albacore stock assessment. In a previous document, we reported the results of our analysis focusing only on Area 1 and Area 3 (Ijima and Tsuda., 2022), but at previous stock assessment meeting, there were the trends of large fluctuations for Japanese longline CPUE and it was mentioned that it may be potentially reflected the mix of adult and immature albacore information in Area 2 due to the problems with the area division of the fleet definition. This document summarized about reanalysis of all areas by the several finite mixture model with updated logbook data from the previous document, in addition, statistically validate the model selection and number of clusters in the model.

Introduction
The last albacore stock assessments use the area-as-fleet approach to define fleets for each fishery (ISC, 2023). This approach considers the growth-dependent albacore distribution in the stock assessment model. However, the "area-as-fleet" approach cannot consider changes in the distribution of adult and immature albacores over time because the definition of each fleet is uniquely determined by latitude and longitude information. Therefore, some part in fleet definitions, especially related to Japanese longline fleets, have a length composition that includes both adult and immature albacore possibly. As a result, when using the current fleet definition in the stock assessment model, it is potentially difficult to Fit between the Japanese longline CPUE, which is only an adult index, and a fleet that includes both an adult and immature fish.

In the previously submitted document, the analysis focused on Area 1 and Area 3 (Ijima and Tsuda., 2022) Because these areas were confirmed, there is the possibility including both immature and adult albacore from length composition data (ISC 2023). On the other hand, there was large fluctuation in the last Japanese longline CPUE (Matsubayashi et al., 2023), and it was suggested that one of the factors for the fluctuated CPUE was the presence of a mixture of adult and immature fish in Area 2. Therefore, in this document, reanalysis was conducted using longline logbook data from all areas including Area 2. In addition, Verification of the number of clusters and the model selection were conducted statistically to input the fleet definition by clustering into future stock assessment models.

Data and Methods
Logbook data
This study used Japanese longline logbooks and length composition data updated from previous results. The details of two types of records (Trip base and Operation base) for mean body weight in Japanese longline logbook data are as described by the previous document

(Ijima and Tsuda 2022). As same as the previous method, the mean body mass per operation were organized by voyage, and voyages with zero variance were excluded from the analysis, as they were judged to have distributed the total catch weight equally. Consequently about 80% of logbook data recorded eye-measured albacore mass, comparing between Trip base and Operation base (Figure 1).

Size composition data

Size composition data was used to compare to finite mixture model analysis results from logbook data. The length composition data used for this study were included the data of fork length of the albacore and some data of body mass since 1962. After 1998, the database format was changed, and individual-based length and body mass measurements were available. Basically, the body mass data are minimal, therefore we used only length composition data as the length comparison data. Size composition data set was composed by multiple data sources: port sampling, training or prefecture vessels, observers, size measurement data conducted by each prefecture government. To compare the logbook and size composition data, 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.

Finite mixture model

The finite mixture model was used to separate the multiple cohorts potentially distributed in all area from logbook data. To construct finite mixture model, we analyzed by using "flexmix" package in R (Gruen et al. 2015). The variable used for analysis was the albacore mean body mass per operation, which was assumed to be the log-normal distribution because mean body mass is a continuous value greater than zero. We selected the best model from several grouping factor assumptions using 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 1-8 and selected visually by the clustering results and the posterior distribution of clustering.

Results and Discussions

Model selection

In this document, the model of the "year-month-5x5 grid" grouping factor was selected, with four clusters for quarters 1 and 2 and three clusters for quarters 3 and 4, as the reason discussed below.

Comparing the grouping factors with the BIC values, the "year-month-5x5 grid" was the smallest for all quarters (Figure 2). The values of BIC became smaller as the number of clusters was increased but tended to change little above a certain number of cluster; a similar trend was observed in the evaluation of BIC in Flexmix, which is thought to be due to overfitting to the data (Ijima et al., 2023). Therefore, as in the previous document, the

optimal number of clusters was determined by visually checking the posterior distribution and the histogram of length composition after clustering.
Confirming the length composition data with the results of clustering, the cluster covered the entire length composition data or the cluster mixing adult and immature albacore appeared when the number of clusters was greater than 5 in quarters 1 and 2 and greater than 4 in quarters 3 and 4 (Fig.3 and Fig. 4). This cluster was not fit to the objective of clustering between the adult and immature albacore. Also, the posterior distribution of clustering showed that the ratio of 0s and 1s tended to decrease after a certain number of clusters (Fig. 5). Thus, the accuracy of clustering tends to decrease when the number of clusters exceeds a certain number, as in the previous document (Ijima and Tsuda, 2022).

The results of Clustering
The annually averaged spatial distribution after clustering of catch, mean weight, and CPUE are shown in Fig. 6. The results for the distribution of mean weight show that most clusters are clearly separated by mean weight. Clusters 1 and 2 of Quarters 1, 2, and 4 were distributed with spatial overlap, with a particular tendency for high catch and CPUE to be concentrated in these clusters.
Based on the clustering results, the catch and weight composition data were compared (Fig. 7). The mode frequencies of body weight were generally consistent, and most of the clusters showed unimodal histograms. As a general trend, the weight composition data tended to have a wider range of data, but after resampling the data, there were more similarities between both data sets (Fig. 7).

This document provided the information for considering new fleet definition by clustering in future stock assessments. The document will be revised with the regard to the contents discussed at working group meeting.

References

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

Ijima, H., and Tsuda. (2022). Additional analysis for Japanese longline logbook data using finite mixture model. ISC/22/ALBWG-02/03.

Ijima, H., Minte-Vera, C., Chang, Y. J., Ochi, D., Tsuda, Y., and Jusup, M. (2023). Inferring the ecology of north-Pacific albacore tuna from catch-and-effort data. *Scientific Reports*, *13*, 8742.

Matsubayashi, J., H. Ijima, N. Matsubara, Y. Aoki, and Y. Tsuda. 2023. CPUE standardization for North Pacific albacore caught by Japanese longline fishery from 1996 to 2021 in Area 2 and Quarter 2. ISC/23/ALBWG-01/03.

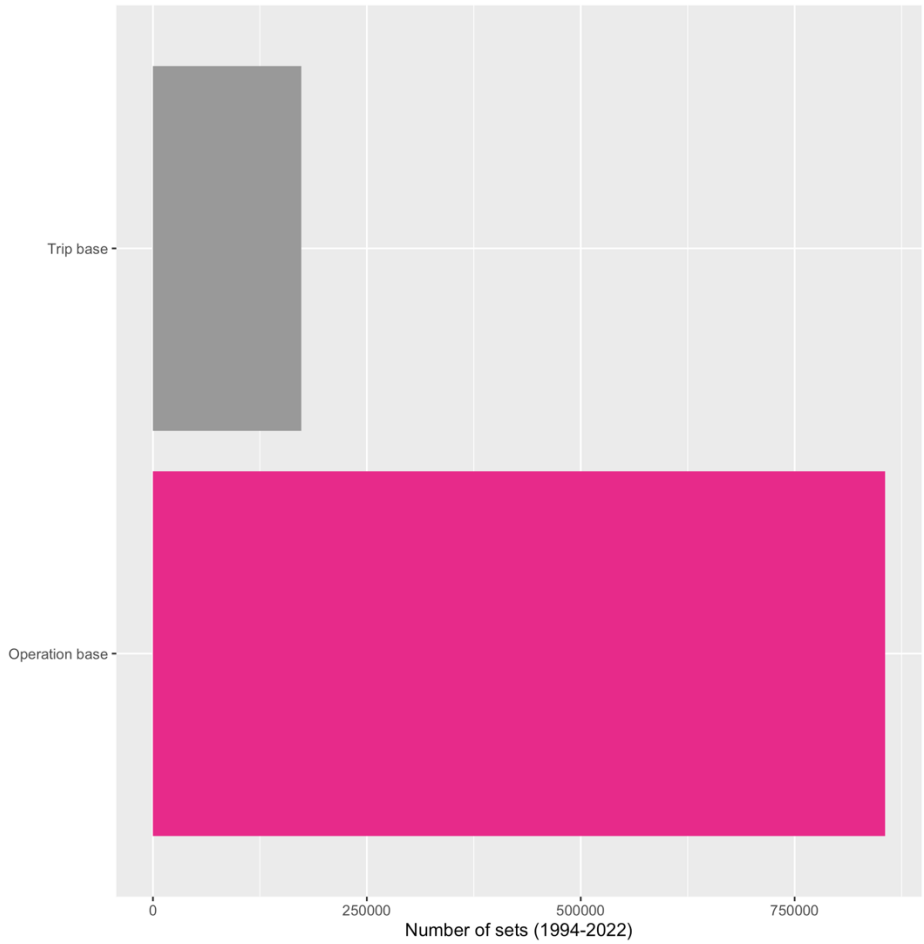

Figure 1. The number of sets in Trip base and Operation base from logbook data. Trip base: dividing the voyage's total catch mass measured at the time of landing equally among the number of operations. Operation base: eye-measured body mass by the fishing master during each operation.

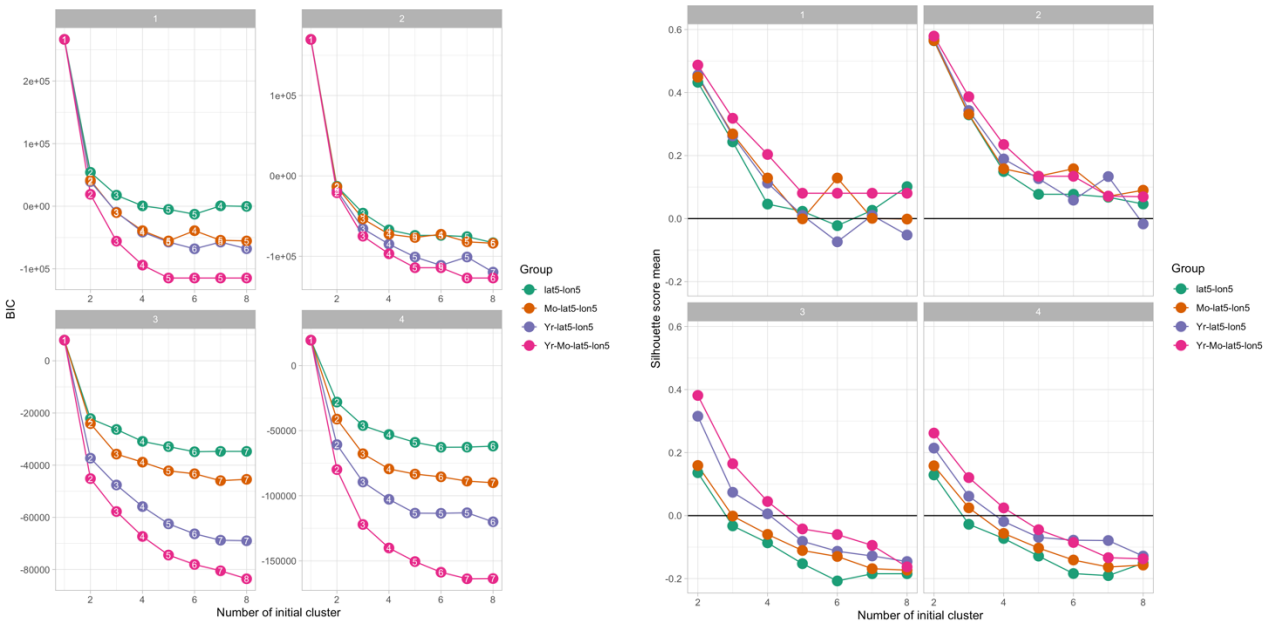

Figure 2. The Bayesian information criterion (BIC) and Silhouette score mean for the model selection.

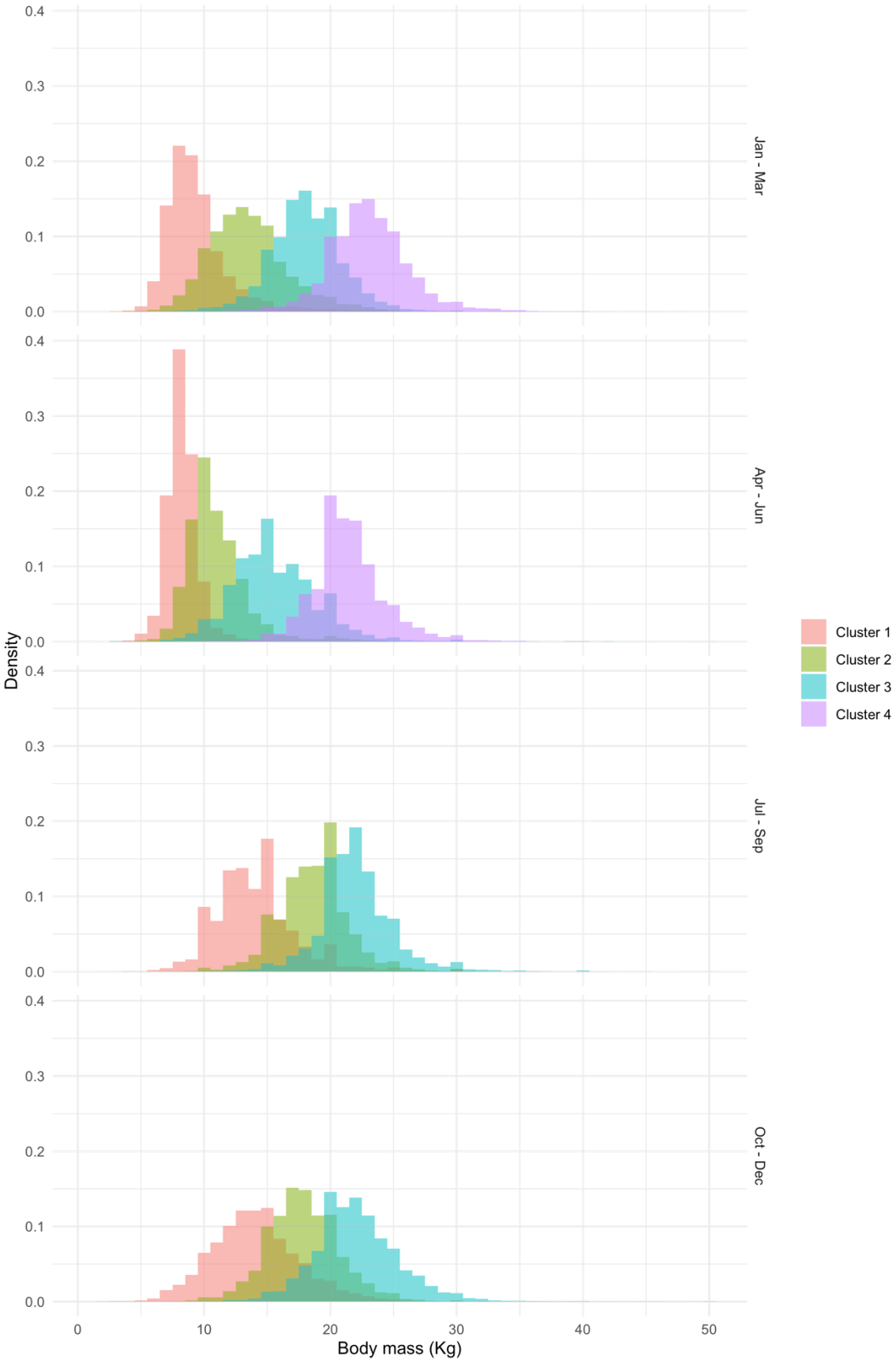

Figure 3. Albacore body mass by one operation summarized by the estimated clusters selected in this study.

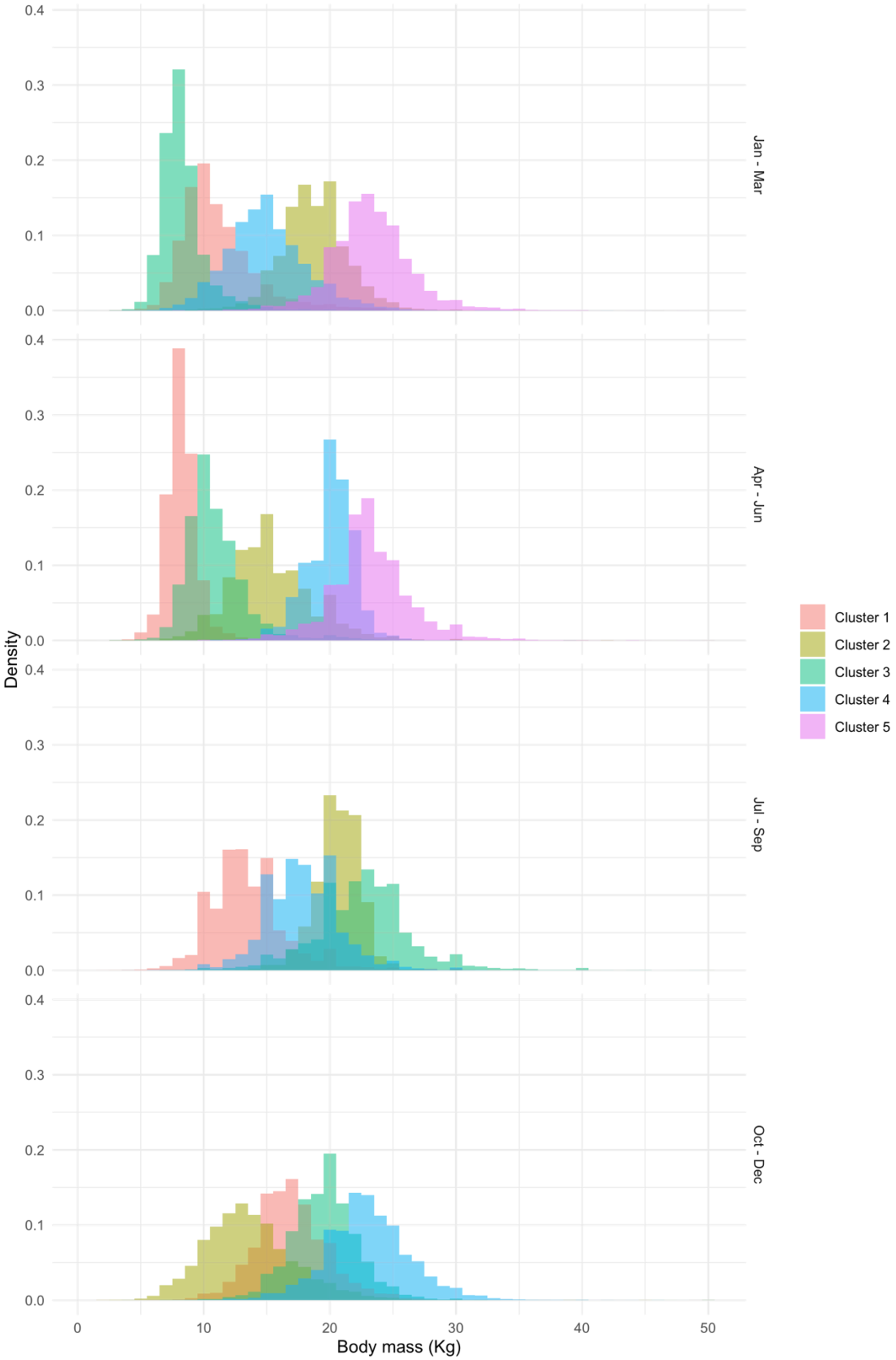

Figure 4. Albacore body mass by one operation summarized by the estimated clusters in case of applying more number of the cluster.

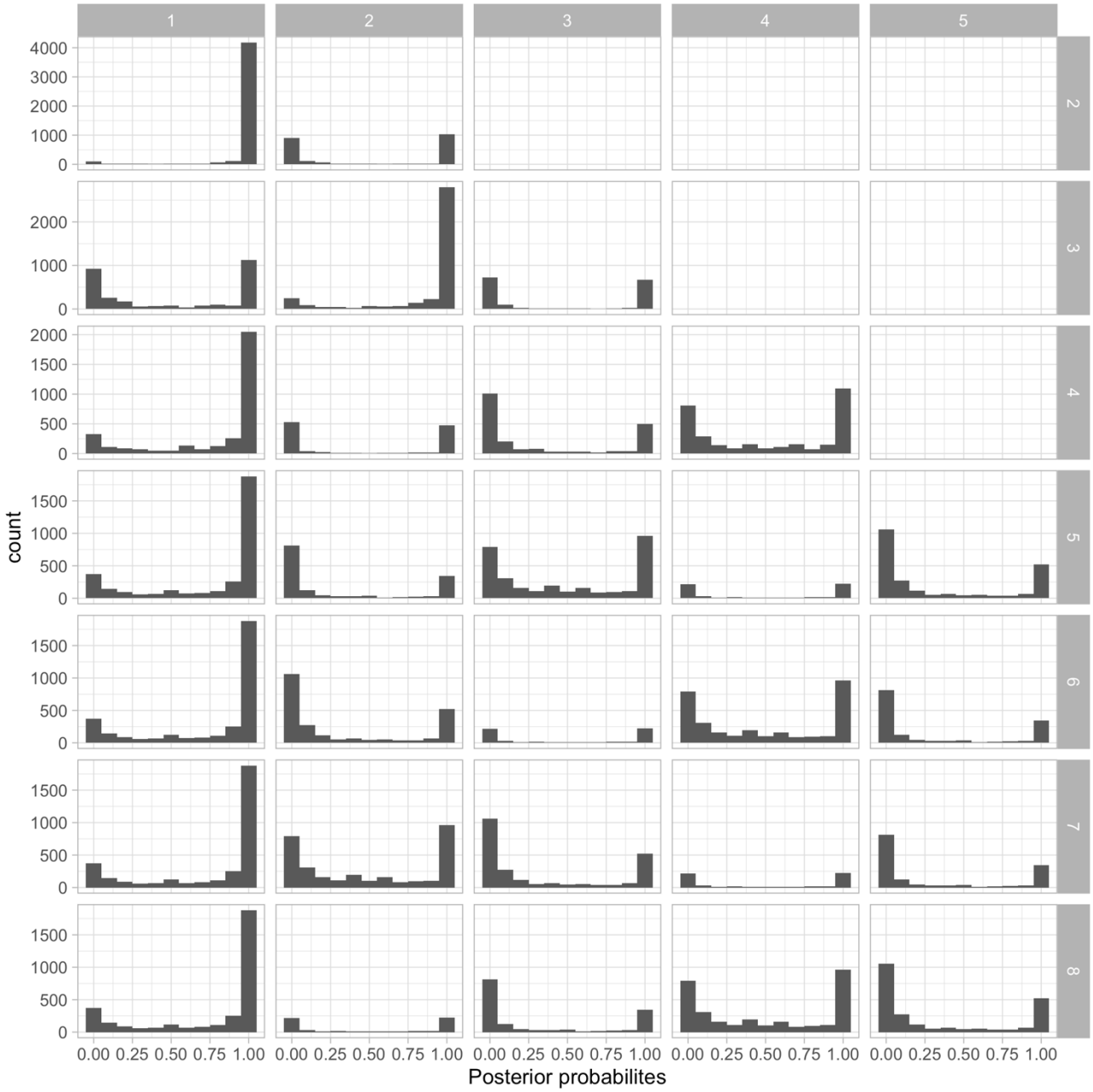

Figure 5. The example of result for posterior distribution in quarter 1

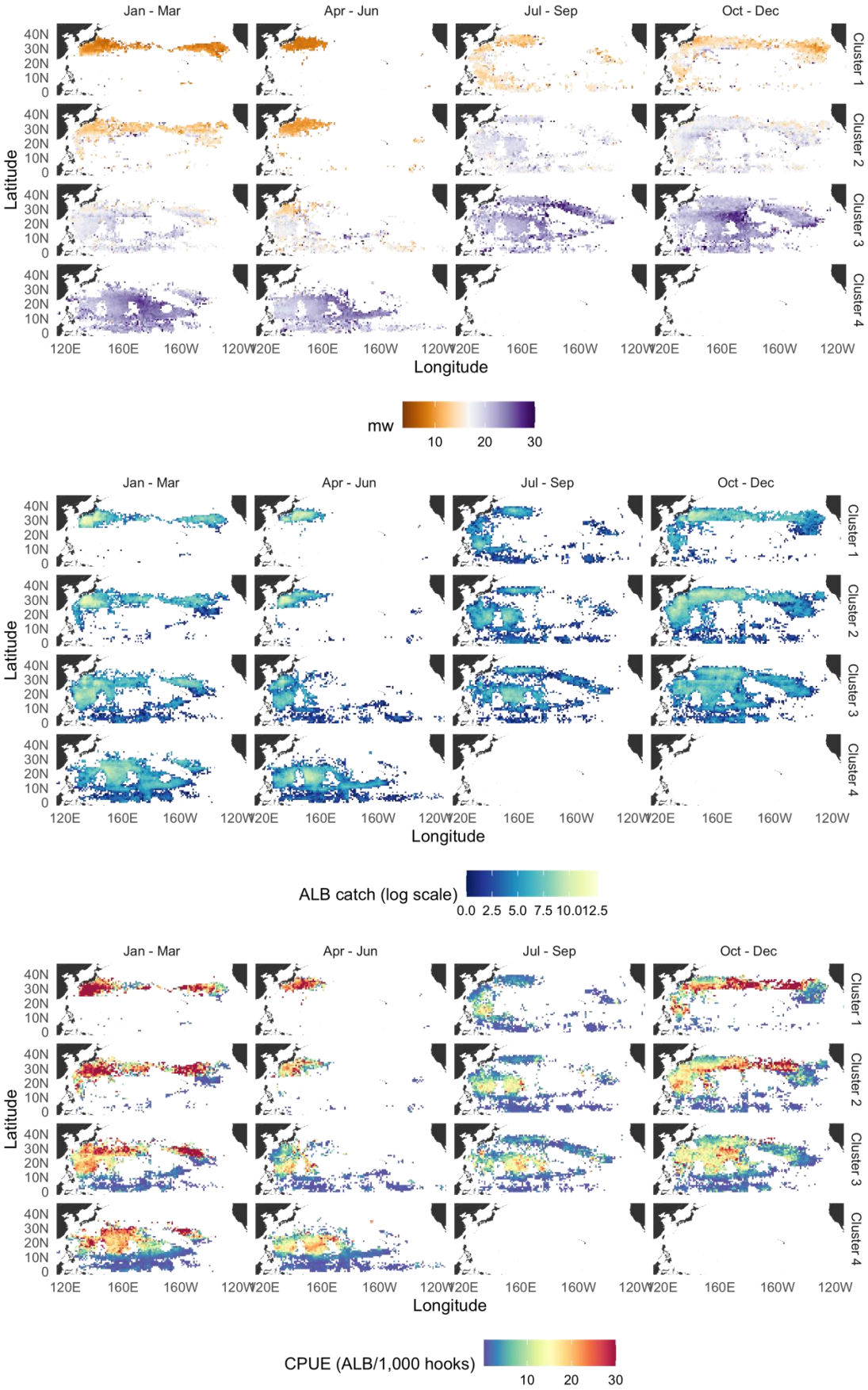

Figure 6. Albacore catch, CPUE, mean weight from Logbook data summarized by 1 x1 grid area and three or four clusters.

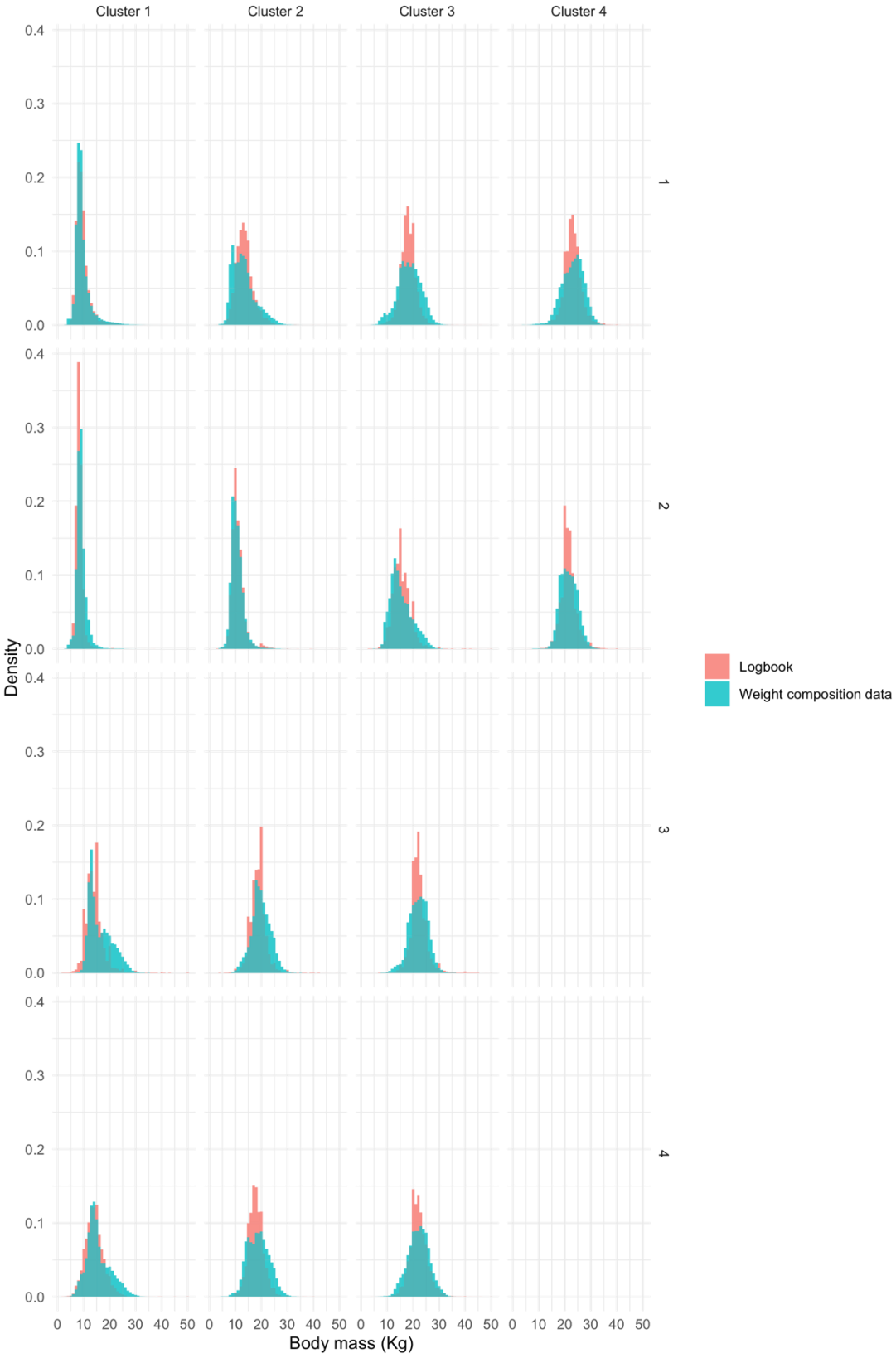

Figure 7. Length composition data summarized by different sampling method. Length composition data was sampled between 1994-2022. Four clusters for quarter 1 and 2 and three clusters for quarter 3 and 4 correspond to finite mixture model analysis using logbook data.

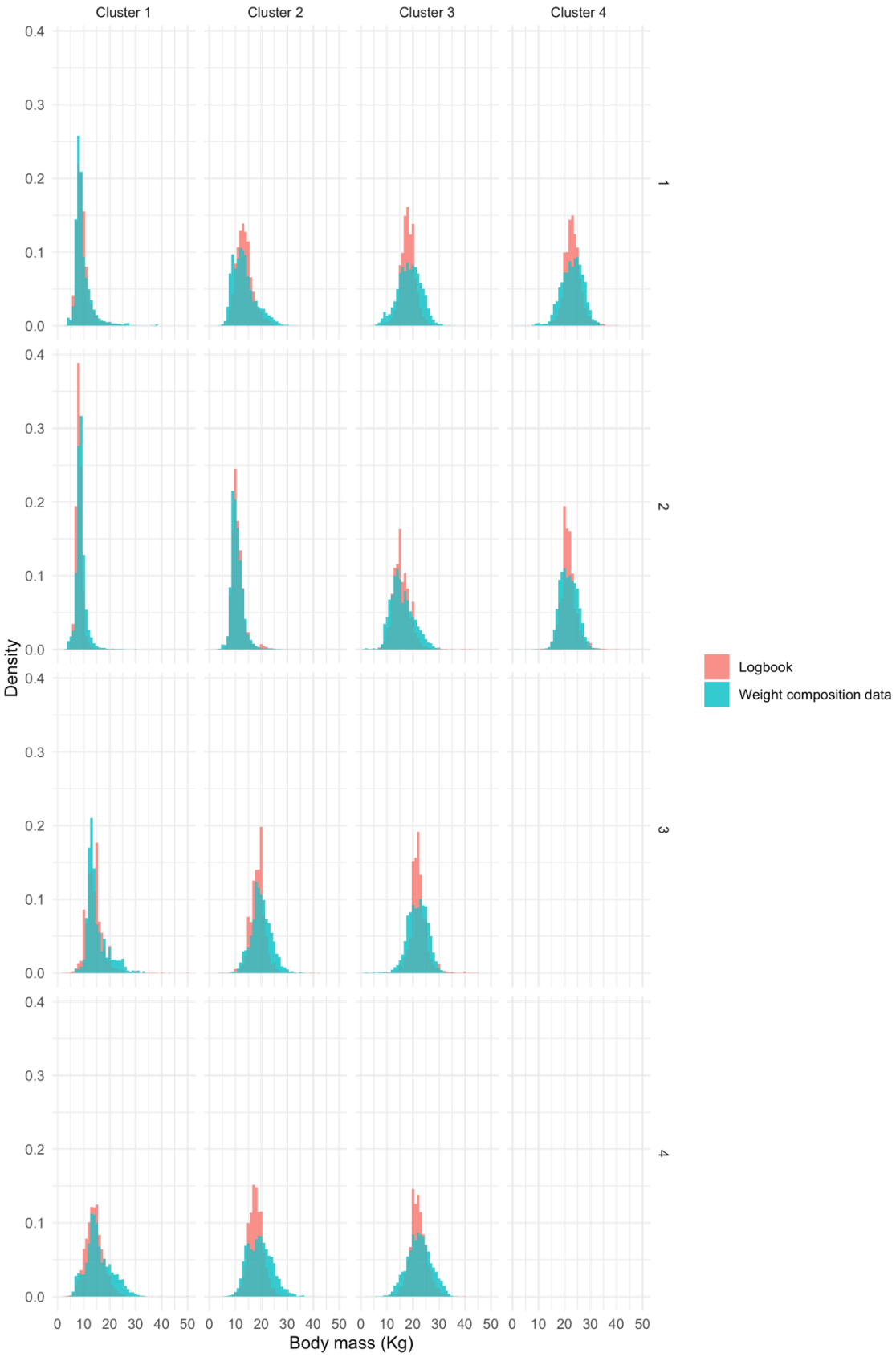

Figure 7. Continued. (Resampling one)