ISC/25/ALBWG-02/03

A Mixture model for defining fleet units from Japanese longline logbook data for North Pacific Albacore tuna (*Thunnus alalunga*)[1]

Hirotaka Ijima and Yuichi Tsuda

Fisheries Resources Institute (FRI) Japan Fisheries Research and Education Agency 2-12-4 Fukuura, Kanazawa-ku, Yokohama, Japan

Email: ijima_hirotaka69@fra.go.jp

1. This working paper was submitted to the ISC Albacore Working Group Intercessional Workshop, October 27 – November 2, 2025, held at the National Research Institute of Fisheries Science, Yokohama, Japan.

Abstract

This study proposes a framework for incorporating the spatiotemporal variability in catch size variability driven by movement and growth into stock assessments, independent of tagging data or pre-defined area boundaries. We focus on the Japanese longline fishery for North Pacific albacore (*Thunnus alalunga*). A mixture model for the mean weight per operational unit was executed quarterly on logbook data, automatically extracting size classes using Year, Month, and a 5° × 5° grid as keys. This consistent classification of the entire dataset achieved a non-pooling design, defining fleets via a statistical Area-as-Fleet approach. This resulting fleet definition was then applied to the length composition data, 5° × 5° catch amount, and Japanese yearbook statistics. Furthermore, a novel CPUE standardization using a spatiotemporal joint species model was applied to the data mapped to a 1° × 1° resolution. As a result, the existing 20 area divisions were simplified to 12, and length frequency data became usable across all fleets.

Introduction

The accurate representation of the spatiotemporal variability of growth and migration has long been considered one of the most challenging issues in stock assessment modeling. The use of migration-explicit stock assessment models is generally recommended for highly migratory species such as tuna, which exhibit rapid growth and extensive migration distances. However, constructing such models necessitates the availability of extensive tagging data, thereby limiting their applicability to resources where sufficient data are accessible. Furthermore, implementing migration often requires the definition of discrete geographical areas, which introduces significant problems. The resulting estimation can change depending on how these areas are delineated, and inter-annual variability in fish distribution and migration patterns can profoundly impact the stock assessment results. In practical assessment, the manual setting and validating multiple areas for assessment is often not feasible within time constraints. Consequently, there is a strong demand for stock assessment models that can capture the spatiotemporal variability of movement and growth without relying on arbitrary area definitions, particularly under data-limited conditions. For species whose distribution and migration patterns change across their life stages, extreme caution is required when estimating catch-at-age. The use of biased length composition data from a specific area to represent overall stock selectivity can introduce significant errors in the final stock size estimation.

The Area-as-Fleet approach is widely proposed as a countermeasure to this problem under no migration information. It involves explicitly defining areas to incorporate the assumption that the size distribution of fish caught differs by area. Methods such as clustering techniques (e.g., GLMTree) that utilize length composition data or Catch Per Unit Effort (CPUE) are typically used for this area classification. However, like migration-explicit models, inter-annual variability in distribution a key characteristic of highly migratory species introduces substantial uncertainty into the estimation results. Indeed, the 2023 North Pacific albacore stock assessment observed a significant change in the CPUE trend due to the expansion of small fish distribution into areas typically occupied by large fish (ISC 2023).

An approach based on the mixture model is a promising alternative that can potentially mitigate the before mentioned challenges. The core of this proposed method is to define a unit combined by Year, Month, and a latitude/longitude block as a single 'fleet', and to estimate each unit independently without any pooling with adjacent units in space or time. This method statistically recovers changes in size distribution those linked to movement and growth from the mixed size data. This approach eliminates the smoothing bias inherent in information borrowing or boundary-setting processes. Similar to conventional clustering methods, this approach uses length information, taking the latitude/longitude block and year/month as the basic units of analysis, and assumes the coexistence of multiple size distributions. Previous research has shown that the eye-to-fork length (or weight proxies) of albacore in Japanese longline data retains information on the catch size at the operational unit level (Ijima et al. 2023). Leveraging this finding, this study organizes the fishery data into fine-scale units to implement a "statistical Area-as-Fleet" that is not dependent on arbitrary, intentional boundary setting.

In this study, we focus on the North Pacific albacore (*Thunnus alalunga*) stock. The Japanese longline fishery will be classified using the mixture model. Based on this resulting classification, CPUE will be standardized using a statistical model, and the refined catch and length composition data will be subsequently constructed for use in the assessment. Furthermore, acknowledging that most official data are typically aggregated at a 5° × 5° grid resolution, the identical procedure will be applied at both the high-resolution (1° × 1° derived from logbooks) and the aggregated 5° × 5° resolutions. Ultimately, the goal is to propose a new statistically implemented methodology that avoids intentional area boundaries, unlike the conventional Area-as-Fleet approach.

Material and Methods

Data source

This study utilized Japanese longline logbook data, Japanese port sampling length composition data, longline catch data aggregated to a 5° × 5° grid, and Japanese government catch statistics (Yearbook). The logbook and length composition data spanned the period from 1994 to 2024. Catches from vessels that were not obligated to submit logbooks were complemented and aggregated using these statistics and the 5° × 5° catch data. Sea Surface Temperature (SST) was obtained from an external data source (Rayner et al. 2003), and monthly grid means were used for the analysis.

Fleet Definition

The analytical framework consisted of two main stages that are defining the statistical fleets and standardizing the CPUE. This research proposes a method for defining fleets within the stock assessment model based on the results of a mixture model. First, we constructed the mixture model quarterly using the average weight per operational unit from the longline logbook as the response variable. The grouping factors were Year, Month, and 5° × 5° grid. The initial number of clusters was explored within the range of 1 to 8. The catch number per operation was used for likelihood weighting. The final number of clusters was determined by comprehensively assessing the Bayesian Information Criterion (BIC), posterior probabilities, and the plausibility of the estimated distribution shapes. From this result, a key combining the corresponding Year, Month, and 5° × 5° grid was created for each size-cluster. This key was then used to classify the logbook, length composition, and 5° × 5° catch data consistently across all datasets.

CPUE Standardization

The Catch Per Unit Effort (CPUE) was standardized using the logbook data, with the unit effort defined as 1,000 hooks. Data were aggregated by Year, Month, 1° × 1° grid, vessel ID, and hooks between floats (HBF). The SST from the previously mentioned external source was re-gridded to the same resolution and merged with the data. A spatiotemporal joint species distribution model was applied to the analysis, treating CPUE as the response variable following a Tweedie distribution. SST was represented using a spline with k=4 degrees of freedom, while HBF and vessel effect were introduced as random effects. The latent spatial field (spatiotemporal effect) was expressed in three layers to explicitly capture the spatial and temporal correlation structure. The average weight was converted to length using a length-weight conversion formula that accounted for uncertainty, and resampling based on the catch number was performed to evaluate the propagation of conversion error and the effective sample size based on catch numbers.

Results and Discussion

We thoroughly examined the analytical results of the mixture model and successfully separated the Japanese longline fishery into 10 distinct statistical fleets (Figure 1). Although this analysis was conducted at a coarser spatial resolution than previous studies, the validity of the classification was generally good. For fisheries lacking logbook data, we aggregated catches combining the first and second quarters (Q1/Q2) and the third and fourth quarters (Q3/Q4), respectively. The resulting fleet definitions reduced the number from 20 to 12, thereby allowing the use of length composition data for every single fleet with available logbook data. The length composition data for the resulting fleets exhibited distinct spatial distribution patterns, with some overlap observed in certain regions (Figure 1a). Furthermore, the study’s data partitioning effectively captured spatiotemporal dynamics including seasonal migration and growth as evidenced by the significant seasonal variation in fish size distribution. We aggregated catch amounts by fleet using the classification key (Year, Month, 5° × 5° grid) derived from the mixture model. The resulting data were organized as a Year × Quarter time series (Figure 1b). Given that Q1 accounted for the largest catches and spatial coverage, CPUE standardization was limited to that quarter. The size classes were subsequently defined as three groups: Juvenile, Sub-adult, and Adult.

The standardized CPUE (Q1) from the spatiotemporal joint species model showed significant spatial overlap between Juvenile and Sub-adult classes. Adults, meanwhile, generally tended to be distributed within the range of 10° N to 30 ° N latitude and 120 ° W to 160°E longitude (Figure 2).The distribution pattern of adult fish is consistent with the known distribution of albacore larvae (ref), supporting the potential for the Adult CPUE class to function as a spawning stock index. Examining the annual trend of the standardized CPUE, a tendency to be relatively lower than the nominal CPUE was observed starting around the year 2000. The longline fleet selectively concentrated its operations in high-density areas as the overall fishing grounds gradually contracted. This selective targeting corrected the spatial bias observed during standardization. Notably, for the Juvenile and Sub-adult indices, the estimated spatial field tended toward zero after the period when fishing grounds in the north of Hawaii disappeared, which may lead to potential underestimation of CPUE. Additionally, the abrupt shift from high to low density at the northern distribution boundary causes the spatial field estimation to become unrealistically large in this analysis. This phenomenon contributes to the instability observed in the Juvenile and Sub-adult indices. Therefore, we decide to use only the Adult index for the stock assessment, excluding the Juvenile and Sub-adult CPUEs. The effect of Sea Surface Temperature (SST) on CPUE differed by size group, with a tendency for the optimal water temperature to shift toward the warmer side as size increased (Figure 3a). This relationship is generally consistent with the size-specific spatial distribution patterns (Figure 3b).

We investigated whether the size classification and CPUE derived from the mixture model approach were consistent with other information. The ISC Albacore Working Group assumes sex-specific growth (males tend to be larger). However, since individual operation data lack sex information, tihs study premise that only growth and natural mortality exhibit sex differences. We did not explicitly address sex differences. However, if the size-as-age classification is plausible, the temporal variability in class-specific CPUE should exhibit time lags corresponding to the age differences. Applying growth curve to estimated size frequency, age difference between Adult and Juvenile to be approximately four years, between Adult and Sub-adult about three years, and between Sub-adult and Juvenile about one year (Figure 4a). Furthermore, the cross-correlation analysis showed significant correlation around these estimated lags, confirming the age differences (Figure 4b). However, the significant range of the lag was somewhat broad and did not perfectly align with the growth equation adopted by the Working Group. Potential causes for this discrepancy include uncertainty in the growth equation, misidentification in size classification, fishing effects, and the previously mentioned instability of the Juvenile and Sub-adult indices.

To validate whether the mean weight reported in the fishing logbook is a valid proxy for catch size information, we compared port sampling length measurements with the converted fork length by mean weight per operation (Figure 5). Both results showed high concordance, suggesting that the mean weight accurately reflects size information in the albacore fishery, and that the port sampling data did not suffer from significant sampling bias. Although, mean weight is available from all operations and minimizes sampling bias, fully characterizing its measurement error and variance structure remains difficult. Conversely, port sampling offers high measurement precision but is susceptible to sampling bias and includes missing data. We suggest using the mean weight time series for SS3 because it is less susceptible to sampling bias and missing data.

References

ISC 2023. Stock Assessment of Albacore Tuna in the North Pacific Ocean in 2023. ISC/23/ANNEX/08

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(1), p.8742.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., Kaplan, A. 2003 Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century J. Geophys. Res.Vol. 108, No. D14, 4407 10.1029/2002JD002670

Table 1. Alternative Japanese longline fleet definition.

Fleet ID	Shortname	Description	Data
F1	F1_JPLL_Q1_J_num	Japan LL, Q1, Juveniles only, catch in 1000s of fish	CSI
F2	F2_JPLL_Q1_SA_num	Japan LL, Q1, Sub-adult only, catch in 1000s of fish	CSIA
F3	F3_JPLL_Q1_A_num	Japan LL, Q1, Adult only, catch in 1000s of fish	CSI
F4	F4_JPLL_Q2_J_num	Japan LL, Q2, Juveniles only, catch in 1000s of fish	CS
F5	F5_JPLL_Q2_SA_num	Japan LL, Q2, Sub-adult only, catch in 1000s of fish	CS
F6	F6_JPLL_Q2_A_num	Japan LL, Q2, Adult only, catch in 1000s of fish	CS
F7	F7_JPLL_Q3_SA_num	Japan LL, Q3, Sub-adult only, catch in 1000s of fish	CS
F8	F8_JPLL_Q3_A_num	Japan LL, Q3, Adult only, catch in 1000s of fish	CS
F9	F9_JPLL_Q4_SA_num	Japan LL, Q4, Sub-adult only, catch in 1000s of fish	CS
F10	F10_JPLL_Q4_A_num	Japan LL, Q4, Adult only, catch in 1000s of fish	CS
F11	F11_JPLL_Q1Q2_wt	Japan LL, Q 1 & 2, catch in mt	C
F12	F12_JPLL_Q3Q4_wt	Japan LL, Q 3& 4, catch in mt	C

C: catch, S: length composition, I: CPUE index and A: length at age.

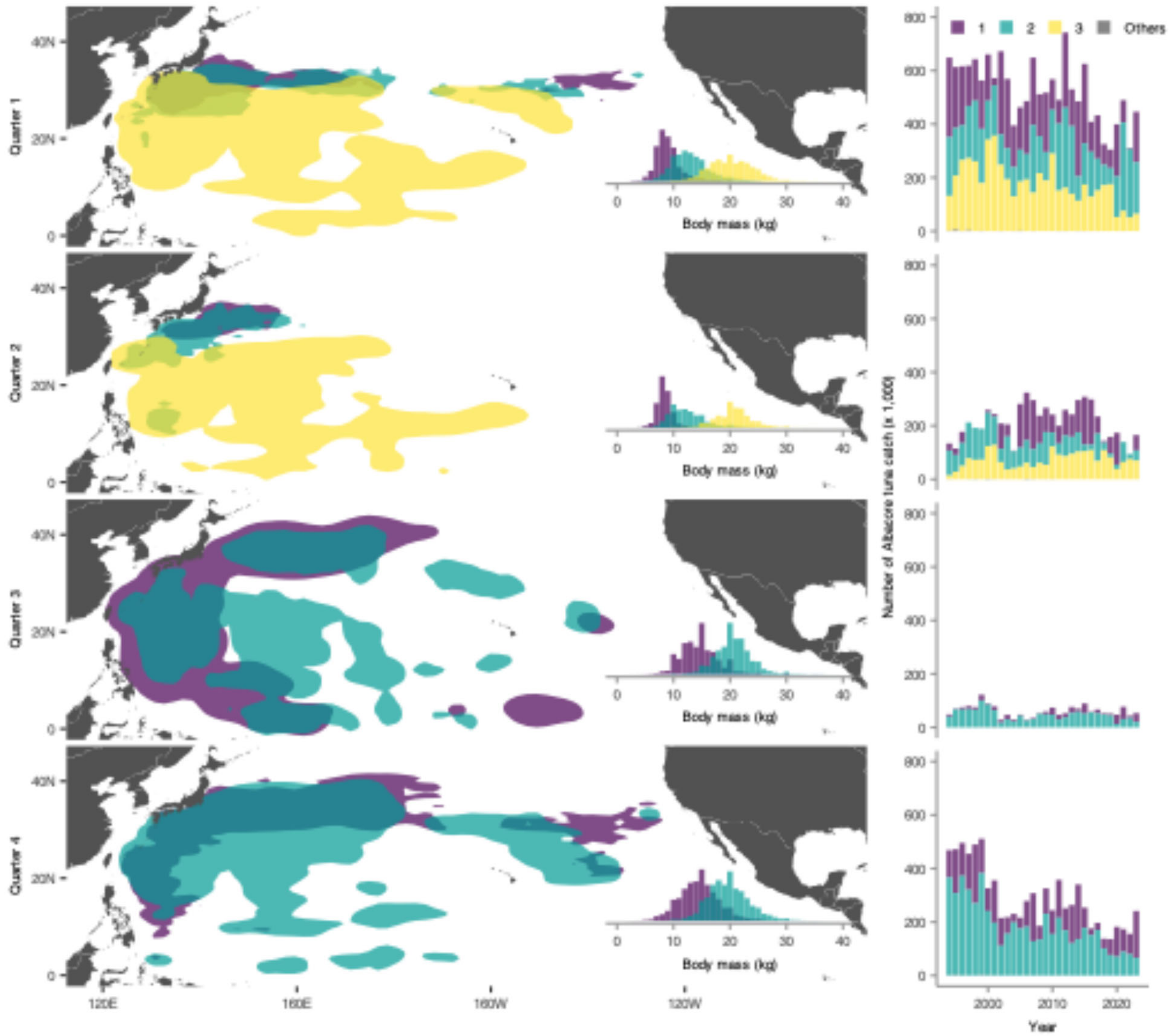

Figure 1. Visualization of the spatial distribution and catch amount per size group, obtained by analyzing logbook data using the mixture model. Areas where size groups are distributed independently and areas where they overlap changed seasonally. Panel a show the highest density region of the 95% probability mass for each classified size group. Panel b shows the number of fish caught per size group.

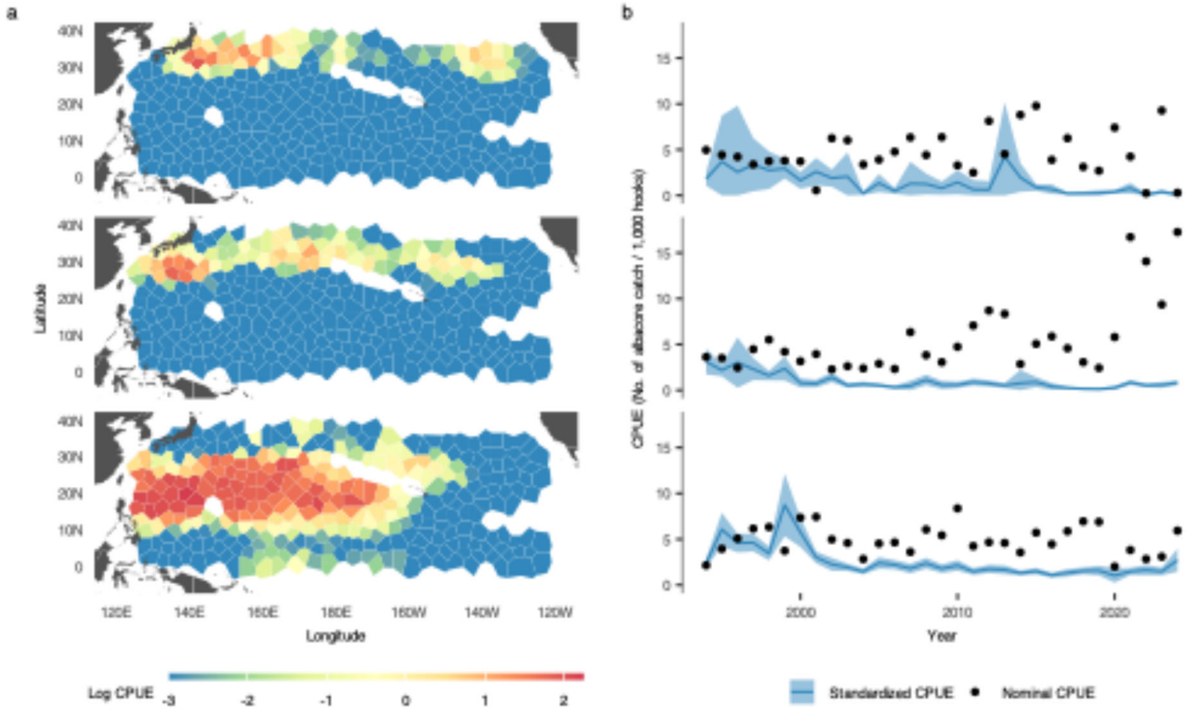

Figure 2. Standardized CPUE in the first quarter using the spatiotemporal joint species distribution model. Panel a show the spatial average log CPUE from 1994 to 2024, arranged from top to bottom by Juvenile, Sub-adult, and adult size groups. Panel b shows the temporal trend of the standardized CPUE for each size group. The standardized CPUE deviates significantly from the nominal CPUE for all classes, which is the result of the fishing grounds shifting toward high-density albacore areas over the years.

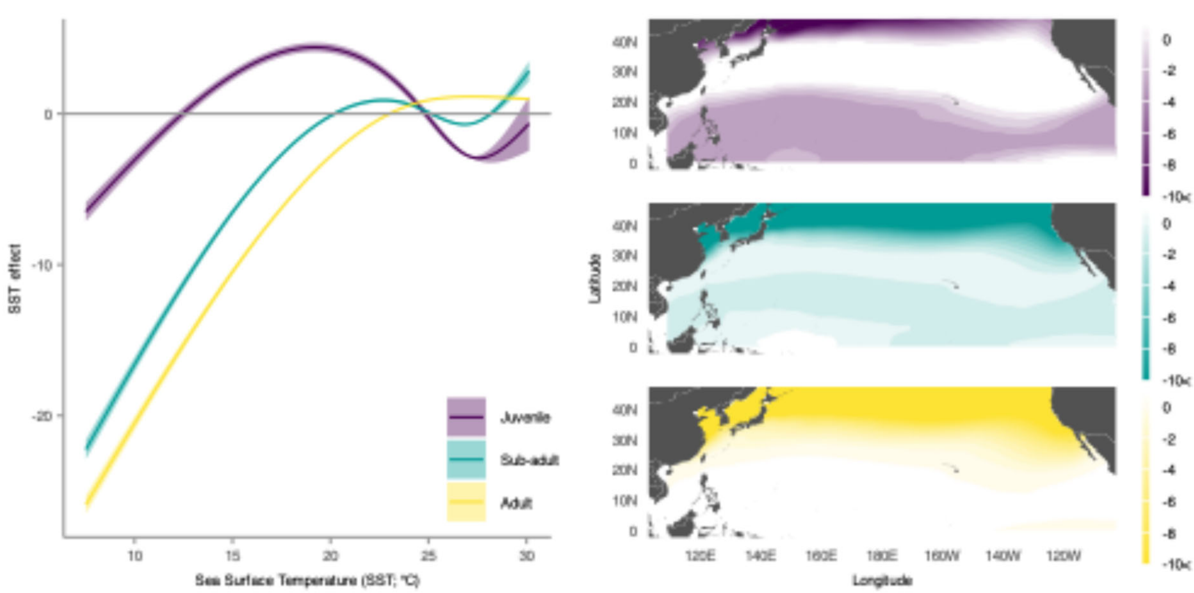

Figure 3. The influence of SST on the spatial distribution of albacore varies by body size. Panel a show the relationship between SST and albacore density in the first quarter, obtained from a spatiotemporal joint species model. Panel b shows the spatial effect of the average SST during the first quarter from 1994 to 2024.

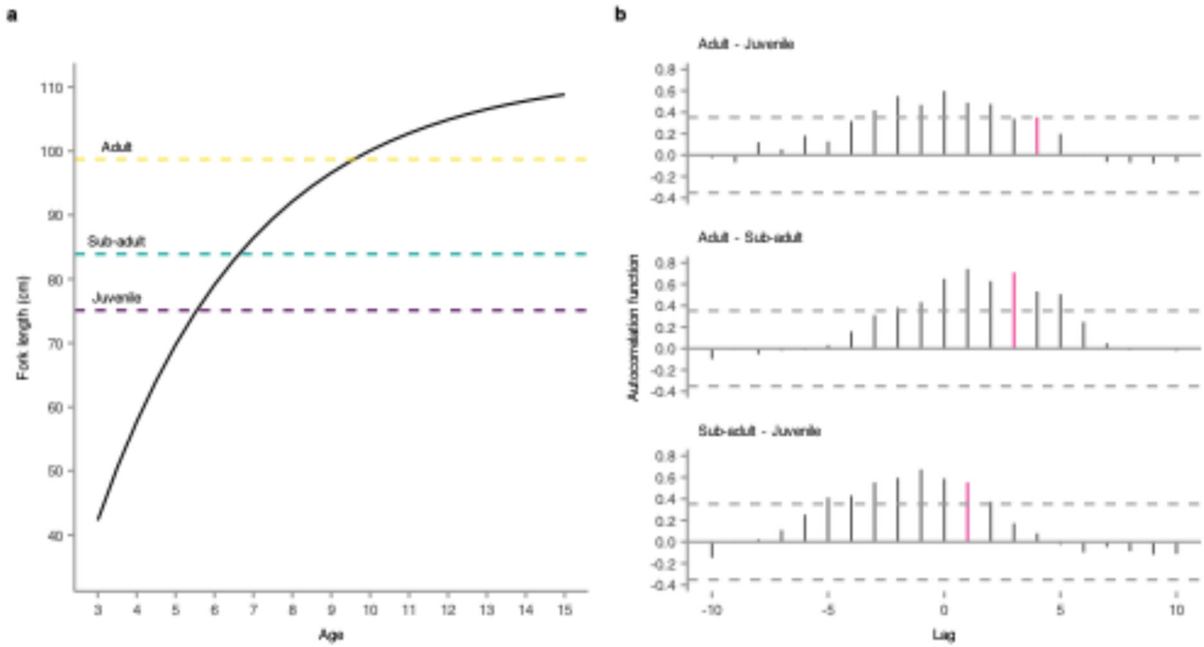

Figure 4. Consistency check of biological assumptions and standardized CPUE in the stock assessment.
Panel a show the albacore growth curve in the North Pacific (sexes combined; ref) as a solid line, and the average body length of the three size groups (juvenile, sub-adult, and adult) classified in the first quarter as dashed lines. The assumed age differences derived from the growth equation are 4 years between adults and juveniles, 3 years between adults and sub-adults, and 1 year between sub-adults and juveniles. Panel b shows the results of the cross-correlation analysis of the CPUE. The red bars represent the lag years expected from the growth equation. The dashed lines in Panel b represent the 95% confidence interval.

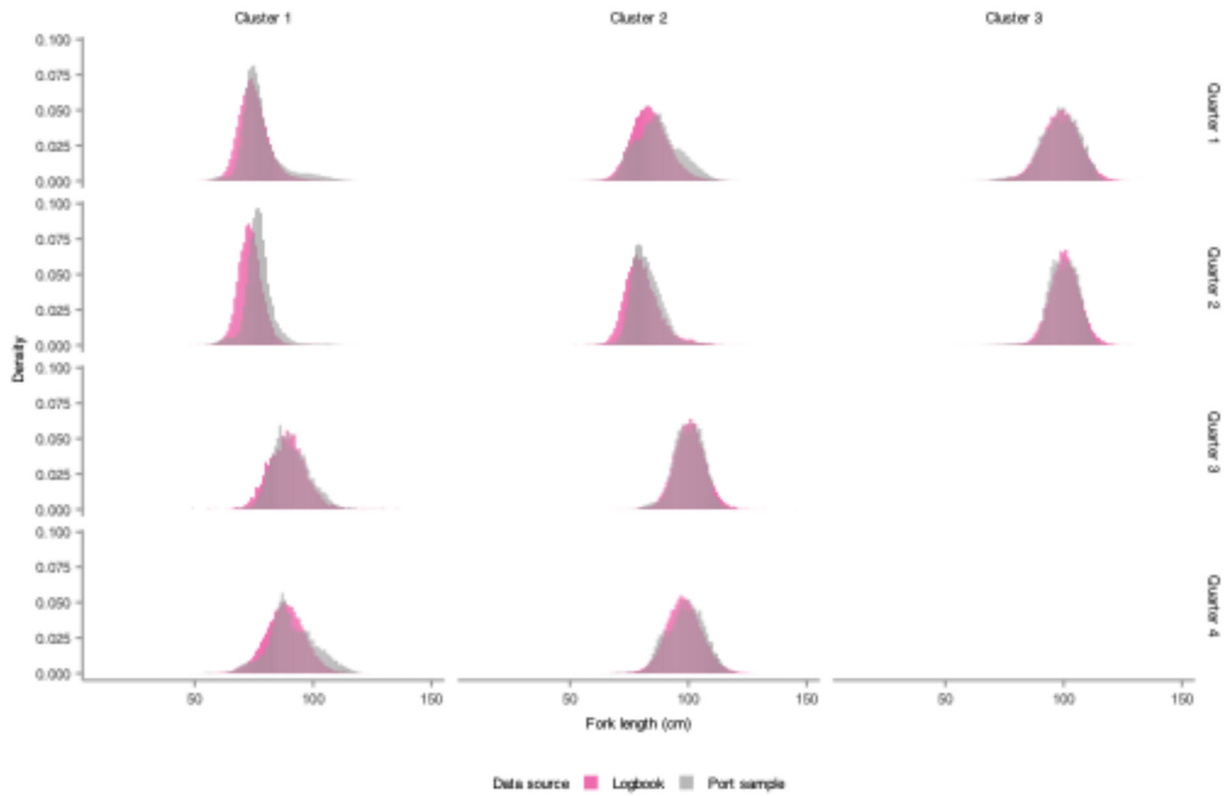

Figure 5. Comparison between the length composition obtained by port sampling and the mean weight used in the mixture distribution model. The mean weight was converted using a length-weight relationship equation that accounts for uncertainty, and length composition data were generated by resampling according to the number of fish caught.