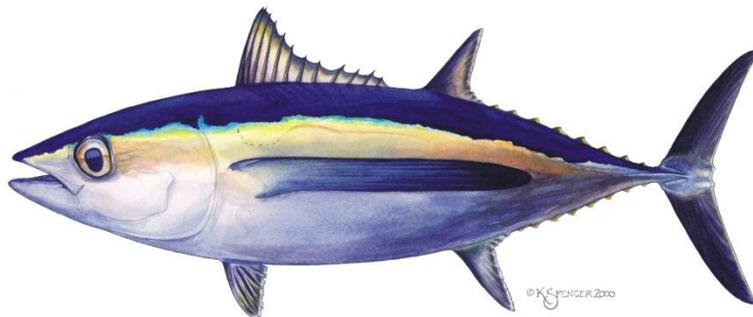


**Spatiotemporal modelling for size-specific CPUE standardization of
albacore tuna in the north Pacific Ocean caught by Taiwanese longline
fisheries**

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

Institute of Oceanography, National Taiwan University, Taipei, Taiwan

yjchang@ntu.edu.tw



Abstract

Understanding the spatial and temporal variations in the size structure of highly migratory fish populations is critical for improving abundance estimates and stock assessments. This study focuses on the North Pacific albacore tuna (*Thunnus alalunga*) and aims to develop a size-specific spatiotemporal model using Taiwanese longline fishery data to generate standardized size compositions for stock assessment purposes. We used fishery CPUE data and length measurements from 2004 to 2023, categorizing albacore into juvenile, intermediate, and adult size groups based on maturity schedules. Spatiotemporal modeling was performed using two modelling packages, VAST and sdmTMB, to estimate both abundance and size-specific CPUE across various years and locations. Results showed that both modelling packages produced consistent abundance trends, with a slight upward trend in juvenile abundance since 2017 and a significant decline in adult abundance since 2010. Additionally, size-specific abundance estimates indicated distinct spatial distributions for juveniles and adults, with juveniles concentrated in higher latitudes. Our analysis highlights the efficacy of the Stepwise method for integrating both CPUE and size composition data, providing a more comprehensive size-specific abundance index. This study offers valuable insights for improving stock assessments and supports the inclusion of Taiwanese data in future abundance indices for albacore tuna in the North Pacific Ocean.

1. Introduction

Understanding the spatial and temporal variations in the size structures of highly migratory fish populations is essential for improving abundance estimates, as these species often exhibit size-based spatial distribution patterns throughout their migratory ranges (Kai *et al.*, 2017; Satoh *et al.*, 2021). This pattern is particularly evident in North Pacific albacore tuna. For example, Ijima *et al.* (2023) found that smaller body mass pseudo-cohorts primarily inhabit higher latitudes, ranging from subtropical to temperate regions, with hotspots off the coast of Japan. In contrast, larger body mass pseudo-cohorts are more commonly found at lower latitudes, from tropical to subtropical regions, with hotspots in the western and central North Pacific.

With increasing fishing effort and data availability in the eastern Pacific Ocean (EPO) from the Taiwanese longline fishery in the recent decades, the ISC ALBWG recommended that data from Taiwan could serve as a valuable source for juvenile abundance indices in this region (ISC, 2024). Thus, this study aims to develop a size-specific spatiotemporal model to generate standardized size compositions for the index fishery. The results will be provided to the ISC ALBWG as a candidate CPUE index for stock assessment.

2. Methods and materials

2.1 Data sources

This study utilized two data sources from the Taiwanese distant-water large-scale longline (DWLL) fishery, obtained by the Overseas Fisheries Development Council (OFDC), to estimate the relative abundance and size composition of albacore in the North Pacific Ocean (NPO).

(1) Fishery CPUE data from albacore-targeted fleets

Daily operational logbooks from Taiwanese longliners in the NPO from 1995 - 2023 were used in this study. The operational data included information on fish species, catch numbers, number of hooks, and location (longitude and latitude), with a resolution of $1^\circ \times 1^\circ$ grid. CPUE is expressed as the number of fish caught per 1,000 hooks (N/1,000 hooks).

This study employed a two-stage cluster analysis as described by He *et al.* (1997) to distinguish between albacore-targeted (ALB fleets) and non-albacore-targeted (non-ALB fleets) fleets within the Taiwanese DWLL fishery in the NPO during 1995 - 2023. Detailed methodology and procedures are presented in Hsu *et al.* (2022).

The cluster analysis indicated that ALB fleets mainly operated north of 25°N during the first (January - March) and fourth (October - December) quarters of the year. Based on these findings, this study focused on CPUE data from ALB fleets operating in these regions and seasons to estimate the abundance of albacore. The spatiotemporal

distribution of nominal CPUEs recorded by ALB fleets between 1995 and 2023 is shown in **Figure 1**.

(2) Length dataset

The length dataset (measured in cm) consists of length measurements of fish caught by Taiwanese DWLL fleets during daily fishing operations in the NPO. Since 1995, regulations have allowed for the measurement of up to 30 fish per day, regardless of species. The length data are recorded at a spatial resolution of $1^\circ \times 1^\circ$. The length data (fork length in cm) for albacore was categorized into three size groups according to the albacore maturity schedule based on the most recent stock assessment (SC19-SA-WP-08): juvenile (<81 cm), intermediate size (81 - 92 cm), and adult (>92 cm).

To estimate size-specific CPUE for albacore, the logbook data from albacore-targeted fleets were combined with the length dataset. Specifically, the daily CPUE for albacore was allocated to the three size groups based on their respective proportions in the length dataset. However, due to data quality considerations, only data from 2004 to 2023 are included in this study. **Figure 2** shows the spatiotemporal distribution of size-specific CPUEs for albacore.

2.2 Spatiotemporal analysis

This study applied two spatiotemporal modeling packages: VAST (Thorson, 2019) and sdmTMB (Anderson *et al.*, 2022) to develop the abundance model and size composition model.

(1) Abundance model

This study applied both packages to estimate the relative abundance of albacore using CPUE data from the ALB fleet. Specifically, the spatial knots, which are used to estimate the spatial and spatiotemporal autocorrelation, were defined slightly differently between the two packages (**Figure 3**).

The model structure with the log-normal error distribution for the CPUE data was identical for both VAST and sdmTMB:

$$\log CPUE(s, t) = \beta(t) + \omega(s) + \varepsilon(s, t) + \delta(v)$$

where $\beta(t)$ is the fixed effect for each year t ; $\omega(s)$ is the spatial random effect for each knot s ; $\varepsilon(s, t)$ is the spatiotemporal random effect for each knot s in year t ; $\delta(v)$ is the random effect for vessel v .

(2) Size composition model

sdmTMB and VAST were used to estimate size-specific CPUEs across different years and locations. We used a delta-lognormal approach (encounter rate and positive catch components) in both models to account for zero CPUE in the size data. The

treatment for incorporating different size groups in the spatial and spatiotemporal random effects differed slightly between the sdmTMB and VAST.

(i) sdmTMB:

Encounter rate component (binominal distribution):

$$p(s, t, l) = \beta_p(t, l) + \omega_p(s, l) + \varepsilon_p(s, l, t)$$

Positive catch rate component (log-normal distribution):

$$q(s, t, l) = \beta_q(t, l) + \omega_q(s, l) + \varepsilon_q(s, l, t)$$

where $\beta(t, l)$ is the fixed effect for group l in each year t ; $\omega(s, l)$ is the spatial random effect for each knot s and group l (interaction term of spatial and group); $\varepsilon(s, l, t)$ is the spatiotemporal random effect for each knot s and group l in year t (interaction term of spatiotemporal and group).

(ii) VAST

Encounter rate component (binominal distribution):

$$p(s, t, l) = \beta_p(t, l) + L_{\omega, p}(l, f)\omega_p(s, f) + L_{\varepsilon, p}(l, f)\varepsilon_p(s, f, t)$$

Positive catch rate component (log-normal distribution):

$$q(s, t, l) = \beta_q(t, l) + L_{\omega, q}(l, f)\omega_q(s, f) + L_{\varepsilon, q}(l, f)\varepsilon_q(s, f, t)$$

where $\beta(t, l)$ is the fixed effect for group l in each year t ; $\omega(s, f)$ is the spatial random effect for knot s and factor f ; $\varepsilon(s, t)$ is spatiotemporal random variation by year t and knot s ; $L(l, f)$ is the loading matrix that leads to spatial (ω) and spatiotemporal (ε) covariation among groups (i.e., a loadings matrix multiplied by its transpose is equal to the covariance among species resulting from the factor), representing the association of group l with factor f (Thorson *et al.*, 2015).

2.3 Derived quantiles

(1) Annual relative abundance indices

The estimated area-weighted abundance ($\hat{CPUE}(t)$) for albacore was calculated except for the vessel effect by sdmTMB and VAST (from the abundance model) as follows:

$$\hat{CPUE}(t) = \sum_{i=1}^{n_s} A(s) \times \exp(\beta(t) + \omega(s) + \varepsilon(s, t))$$

where $A(s)$ is the surface area (in km²) of knot s .

(2) Size-specific CPUE and proportion

The estimated albacore's CPUE for size group l at s knot in t year ($d(s,t,l)$) is estimated by sdmTMB and VAST (from the size composition model) as follows:

$$d(s,t,l) = p(s,t,l) \times q(s,t,l)$$

The estimated size-specific CPUE ($d(s,t,l)$) was further converted into the proportions of size group l at s knot in t year ($\alpha(s,t,l)$) as follows:

$$\alpha(s,t,l) = \frac{d(s,t,l)}{\sum_{i=1}^l d(s,t,i)}, \quad \sum_{i=1}^l \alpha(s,t,i) = 1$$

(3) Size-specific relative abundance

This study used two estimations to calculate the size-specific relative abundance ($B(l,t)$) with area-weighting: (i) conventional and (ii) stepwise estimation calculated by sdmTMB and VAST.

(i) Conventional estimation:

$$B(l,t) = \sum_{s=1}^{n_s} A(s) \times d(s,t,l)$$

(ii) Stepwise estimation:

$$B(l,t) = \sum_{s=1}^{n_s} A(s) \times \hat{CPUE}(s,t) \times \alpha(s,t,l)$$

where $\hat{CPUE}(s,t)$ is the estimated relative abundance for knot s in year t (from abundance model); α is the proportion of l size group at a specific t year and s grid.

3. Results

3.1 Cluster Analysis

This study applied a two-stage cluster analysis to examine targeting species composition in the tuna and billfish catch data from the North Pacific longline fishery between 1995 and 2023. As shown in **Figure 4**, the catch composition can be classified into two targeting fleet types: one primarily targeting albacore tuna (ALB fleets) and the other targeting bigeye and yellowfin tuna (non-ALB fleets). The fishing characteristics of these fleets are presented in **Figure 5**, indicating that ALB fleets have higher albacore tuna catch, catch composition, and CPUE compared to non-ALB fleets. The spatial distribution of seasonal catch composition for both fleets is illustrated in **Figure 6**,

showing that the ALB fleets primarily operate north of 25°N, with fishing activities concentrated in the first and fourth quarters. Based on these findings, this study uses only the CPUE data from the ALB fleets operating north of 25°N during the first and fourth quarters for standardization analysis.

3.2 Abundance Model

The predicted abundance (density) and spatial knots configuration for the sdmTMB and VSAT packages, respectively, are shown in **Figure 3**. Although the spatial knots configuration differs between the two packages, the overall trend in predicted abundance remains similar. The standardized annual abundance trends derived from the two modelling packages are presented in **Figure 7**, showing a high degree of consistency in trend estimates and confidence intervals. In the early years, the estimated abundance exhibited greater fluctuations, whereas from 2014 to 2023, it stabilized with a slight upward trend.

3.3 Size Composition Model

Since VAST requires additional estimation of size-class correlation parameters in factor analysis, it has more parameters compared to sdmTMB. In terms of model fit, the AIC value for sdmTMB is 311,362 whereas for VAST, it is 313,392, indicating that sdmTMB achieves better model fit with fewer parameters and a more parsimonious structure.

Figure 8a shows the averaged probability-integrated transform (PIT) residuals by 1×1 degree in the studied area for the size composition model by using sdmTMB. The mean residuals exhibit no distinct spatial pattern and generally varied around 0.5. The residuals frequency distribution exhibits an approximately uniform distribution (**Figure 8b**), suggesting that the model provides a good fit.

3.4 Size-Specific Abundance Index

Figure 9 presents the spatial density distributions of different size classes (adult, intermediate, juvenile) by year, estimated using various spatiotemporal modelling packages (sdmTMB vs. VAST) and estimation methods (Conventional estimation vs. Stepwise estimation). The predicted high-density areas differ between juveniles and adults, regardless of the model or estimation method used. Juveniles show higher relative densities primarily between 30°N and 40°N, extending eastward from the Japanese coast to 180°E, as well as along the entire North American coast up to 140°E. In contrast, the relative density of adults is lower in these areas. South of 30°N, juveniles exhibit lower relative densities, whereas adults display higher relative densities.

Figure 10 presents the abundance estimates for different size classes based on various spatiotemporal modelling packages and estimation methods. The two spatiotemporal packages show similar trends, with VAST exhibiting greater interannual variability. Likewise, the two estimation methods produce comparable trends. Overall, size-specific abundance estimates indicate a slight increase in juvenile abundance since 2017, while the estimated abundance of intermediate-sized fish in 2023 was notably lower. In contrast, adult abundance has declined significantly since 2010 and has remained relatively low over the past four years.

4. Discussion

Given the similarity between the results of sdmTMB and VAST, as well as the more intuitive coding structure of sdmTMB, which offers greater flexibility and convenience compared to VAST's control file-based settings, this study recommends considering sdmTMB as one of the methods for future CPUE standardization analyses. Additionally, sdmTMB has already been used in the WCPFC SC20 meeting this year for CPUE standardization of South Pacific albacore tuna and Southwest Pacific striped marlin, with the results incorporated into stock assessments (WCPFC SC20-SA-IP-05, 06).

For estimating the size-specific abundance index, the Stepwise method allows for the use of more observational data compared to the Conventional method. This is because the Conventional method relies solely on size composition data, meaning that the number of observations available for model estimation is limited by the sample size of length measurements (a maximum of 30 fish can be measured per fishing day on each vessel, regardless of species), often resulting in fewer total records than those in the logbook data. In contrast, the abundance model in the Stepwise method incorporates the complete logbook dataset, while the size composition model also utilizes the length data used in the Conventional method. As a result, the final length-specific abundance estimates are derived by integrating both datasets, making full use of all available fishery data.

Furthermore, the Stepwise method enables the abundance model to capture overall abundance trends, while the composition model focuses specifically on length data. Additionally, the input data for the composition model are not limited to CPUE values; they can also include species composition proportions, thereby enhancing the model's flexibility to accommodate different data formats. This study also recommends future simulation analyses to further assess whether the Stepwise method can more accurately estimate the size-specific standardized index.

References

- Anderson, S. C., Ward, E. J., English, P. A., and Barnett, L. A. (2022). sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. *BioRxiv*, 2022-03.
- He, X., Bigelow, K. A., & Boggs, C. H. (1997). Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery. *Fisheries Research*, 31(1-2), 147-158.
- Hsu J., C.-H. Yi, C.-W. Chang and Chang Y.J. (2022). Catch, length composition, and standardized CPUE of the North Pacific albacore caught by the Taiwanese distant-water longline fisheries in North Pacific Ocean from 1995 - 2021. *ISC/22/ALBWG-02/07*.
- Ijima, H., Minte-Vera, C., Chang, Y. J., Ochi, D., Tsuda, Y., & Jusup, M. (2023). Inferring the ecology of north-Pacific albacore tuna from catch-and-effort data. *Scientific Reports*, 13(1), 8742.
- ISC. (2024). Plenary Report of the 24st meeting of the international scientific committee for tuna and tuna-like species in the North Pacific Ocean.
- Kai, M., Thorson, J. T., Piner, K. R., & Maunder, M. N. (2017). Spatiotemporal variation in size-structured populations using fishery data: an application to shortfin mako (*Isurus oxyrinchus*) in the Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(11), 1765-1780.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: automatic differentiation and Laplace approximation. *Journal of statistical software*, 70, 1-21.
- Liu, H. I., Chang, S. K., & Wu, R. F. (2015). Overviews on fishery development, data collection and processing system of Taiwanese distant-water longline fishery in the Pacific Ocean.
- Maunder, M. N., Hamel, O. S., Lee, H. H., Piner, K. R., Cope, J. M., Punt, A. E., Ianelli, J. N., Castillo-Jordán, C., Kapur, M.K., & Methot, R. D. (2023). A review of estimation methods for natural mortality and their performance in the context of fishery stock assessment. *Fisheries Research*, 257, 106489.
- Satoh, K., Xu, H., Minte-Vera, C. V., Maunder, M. N., & Kitakado, T. (2021). Size-specific spatiotemporal dynamics of bigeye tuna (*Thunnus obesus*) caught by the longline fishery in the eastern Pacific Ocean. *Fisheries Research*, 243, 106065.
- Thorson, J. T., Scheuerell, M. D., Shelton, A. O., See, K. E., Skaug, H. J., & Kristensen, K. (2015). Spatial factor analysis: a new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution*, 6(6), 627-637.

- Thorson, J.T., 2019. Guidance for decisions using the Vector Autoregressive SpatioTemporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fish. Res.*, 210, 143-161.
- WCPFC SC19. (2023). Stock assessment of albacore tuna in the North Pacific Ocean in 2023. WCPFC-SC19-SA-WP-08.
- WCPFC SC20. (2024). Background Analyses and Data Inputs for the 2024 South Pacific Albacore Tuna Stock Assessment. SC20-SA-IP-05.
- WCPFC SC20. (2024). Background analyses for the 2024 stock assessment of southwestern Pacific striped marlin. WCPFC-SC20-SA-IP-06.
- Yuan, T. L., Xu, H., Lu, B. J., & Chang, S. K. (2024). Comparison of linear and nonlinear modeling approaches to develop an abundance index based on voyage and market data for a data-limited fishery. *Frontiers in Marine Science*, 11, 1344181.

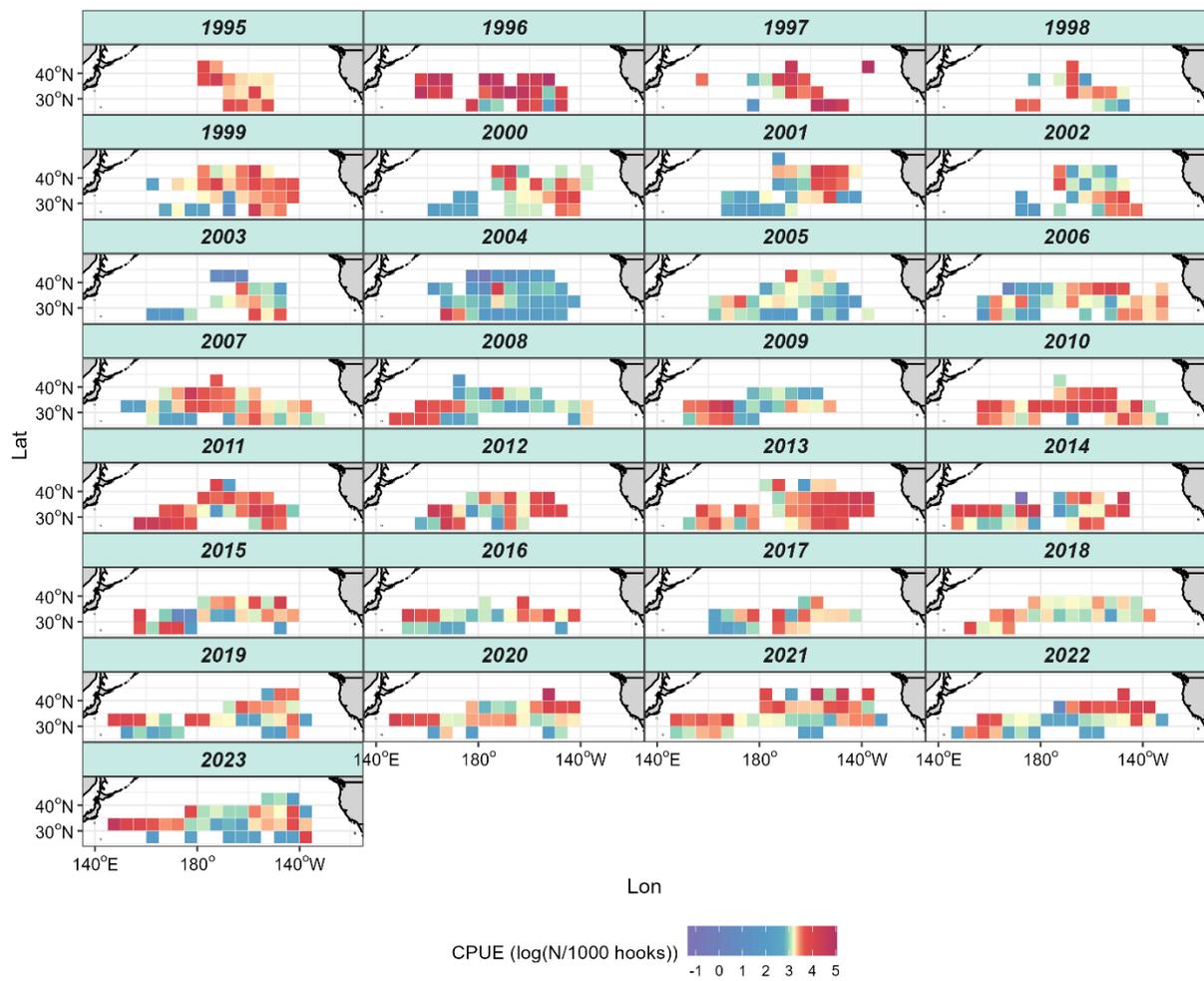


Figure 1. Spatial distribution of logarithm nominal CPUE (N/1,000 hooks) for albacore in the North Pacific Ocean by the Taiwanese large-scale tuna longline fishery in the first and fourth quarters from 1995 to 2023.

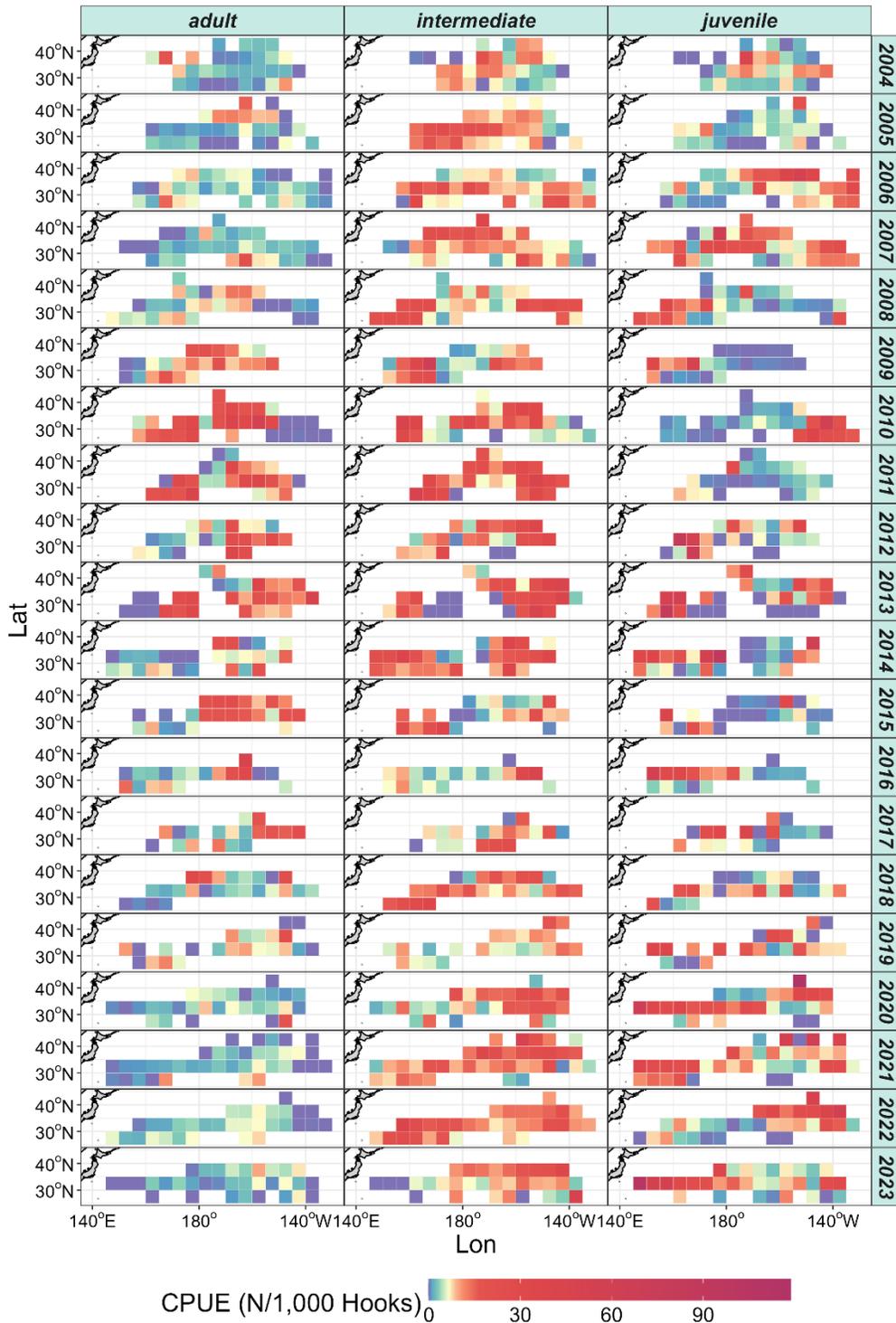


Figure 2. Spatial distribution of nominal CPUE (N/1000 hooks) for albacore in the North Pacific Ocean, categorized into three size groups - juvenile (<81 cm), intermediate (81 - 92 cm), and adult (>92 cm) - based on data from the Taiwanese large-scale tuna longline fishery during the first and fourth quarters from 2004 to 2023.

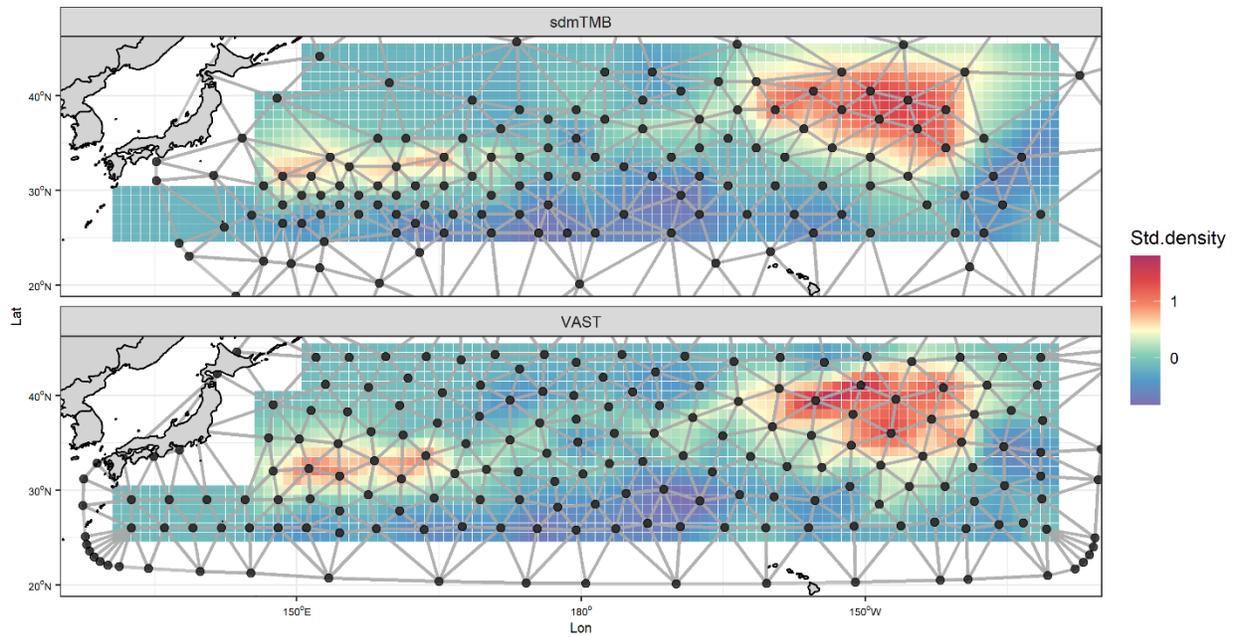


Figure 3. Spatial distribution of knots and estimated density by sdmTMB and VAST for albacore in the North Pacific Ocean. Different colored grids represent estimated density, and black dots indicate the spatial distribution of spatial knots used in sdmTMB and VAST.

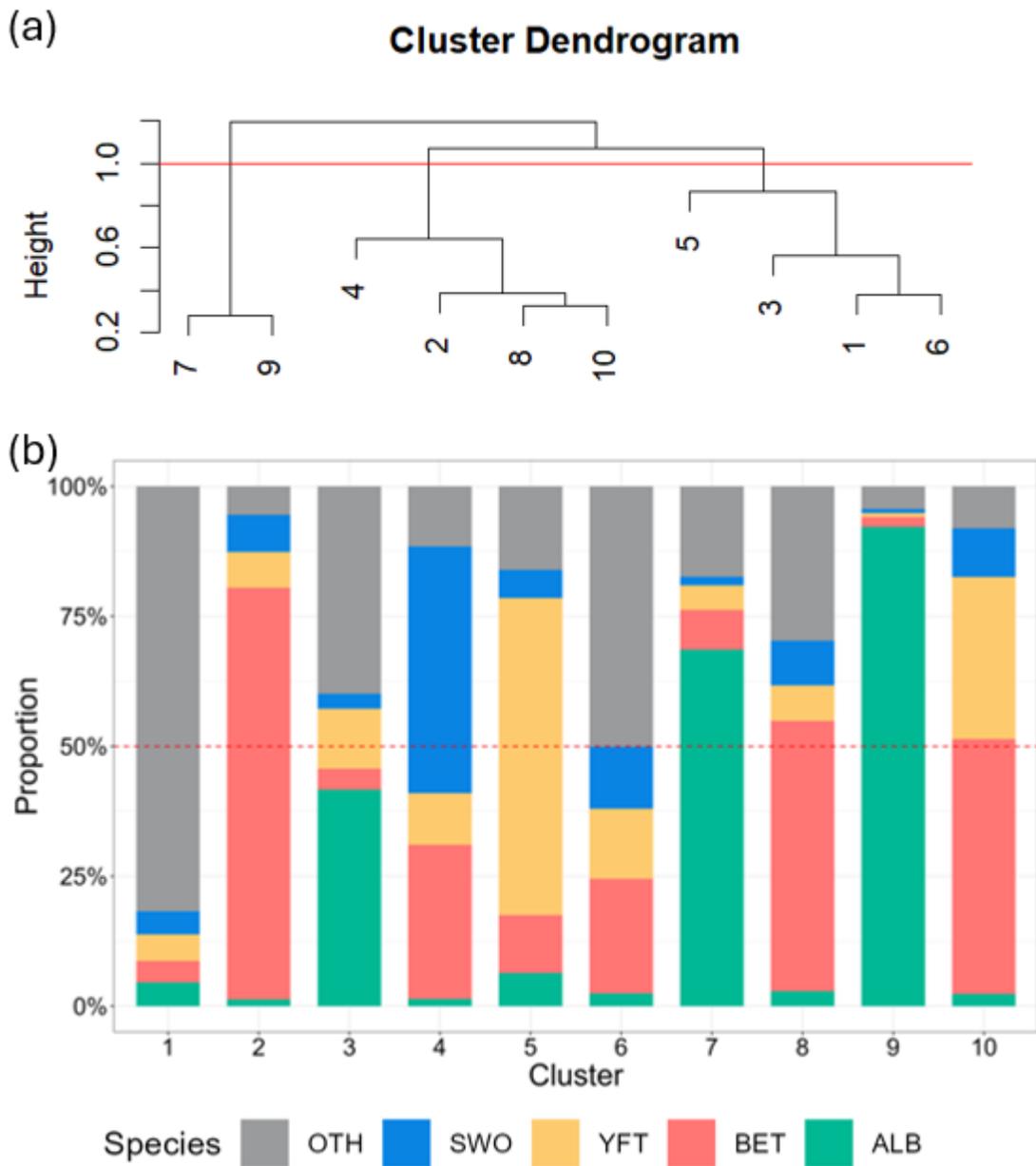


Figure 4. The proportions of catches (number) by species for the 10 non-hierarchical (K-means) clusters based on the Taiwanese distant water longline fishery logbook data from 1995 - 2023.

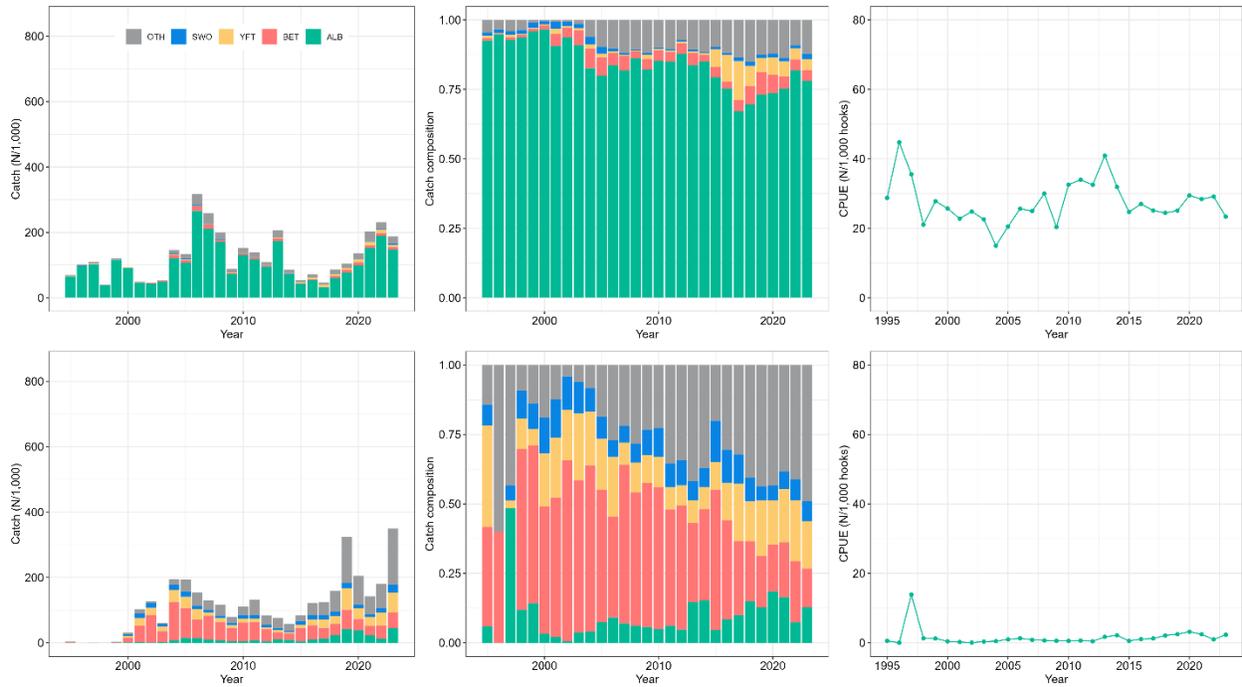


Figure 5. Catch (in numbers), catch composition, and albacore CPUEs (N/1,000 hooks) of the two clusters from 1995 - 2023. The top figure represents the albacore targeting fleets, and the bottom figure represents the non-albacore-targeting fleets.

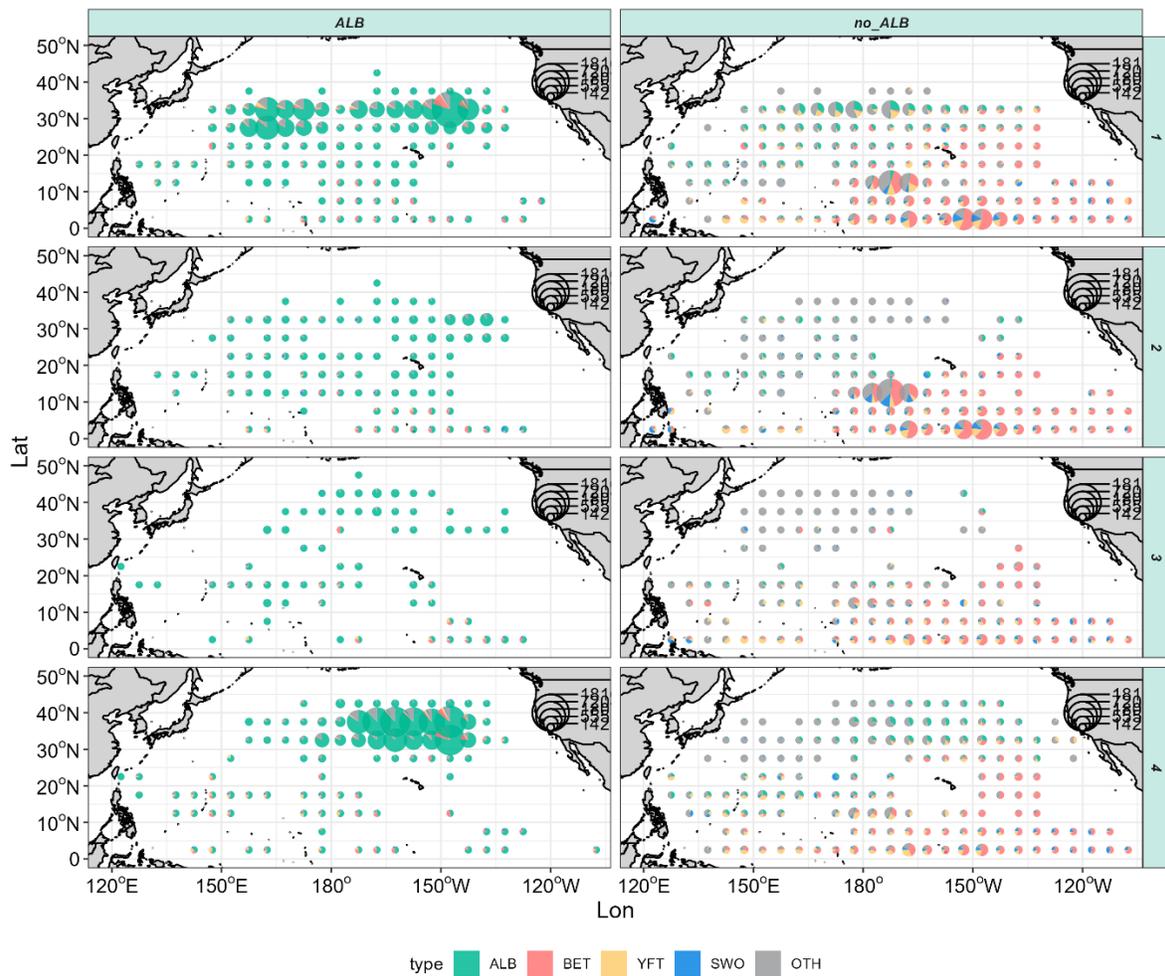


Figure 6. The spatiotemporal distribution of the species composition of fish catches in different fleets and seasons. Different colors represent different species; the size of the pie chart represents the catch amount.

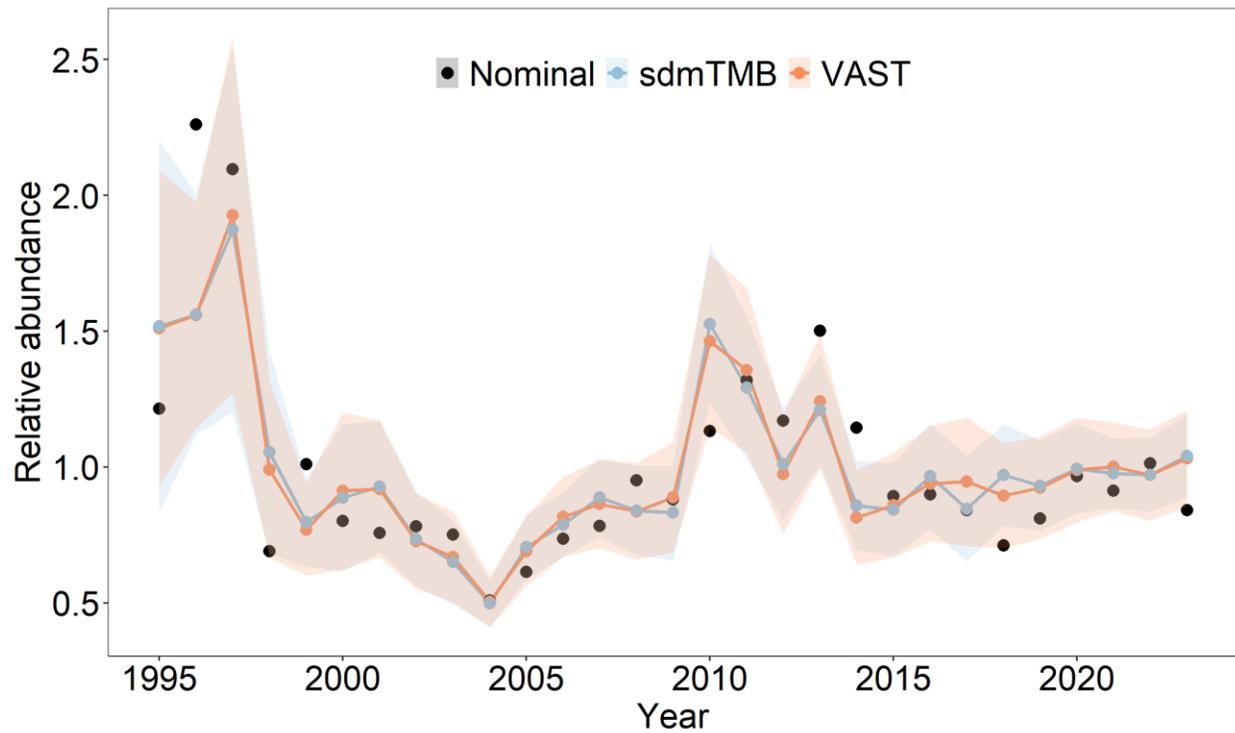


Figure 7. Annual relative abundance trends of albacore in the North Pacific Ocean during 1995 - 2020 estimated by sdmTMB and VAST. Black dots represent nominal CPUE, and the polygon represents the 95% confidential intervals.

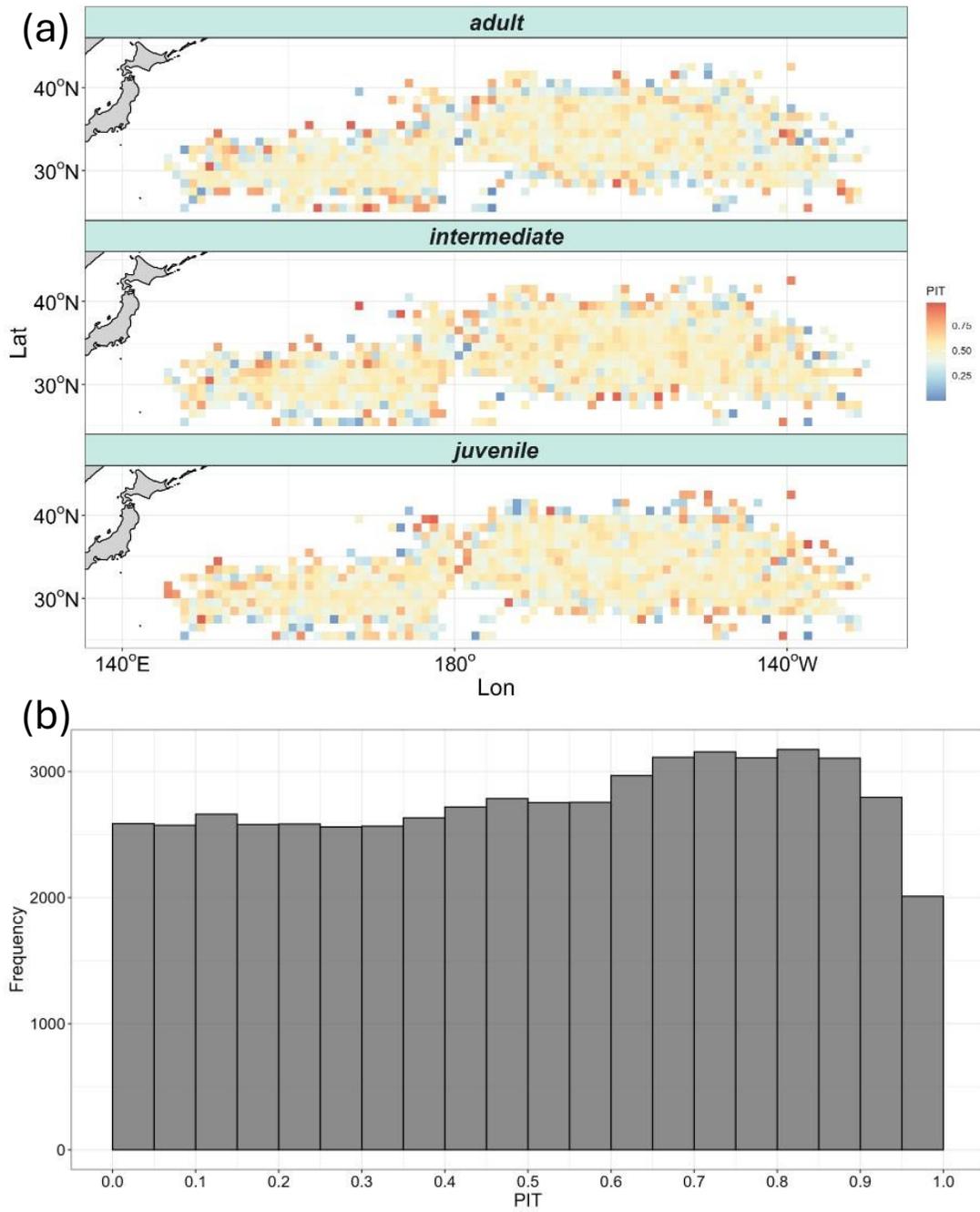


Figure 8. (a) Spatial distribution of aggregated average probability-integrated transform (PIT) residuals at a $1^\circ \times 1^\circ$ resolution for each size group based on sdmTMB, and (b) frequency distribution of model PIT residuals.

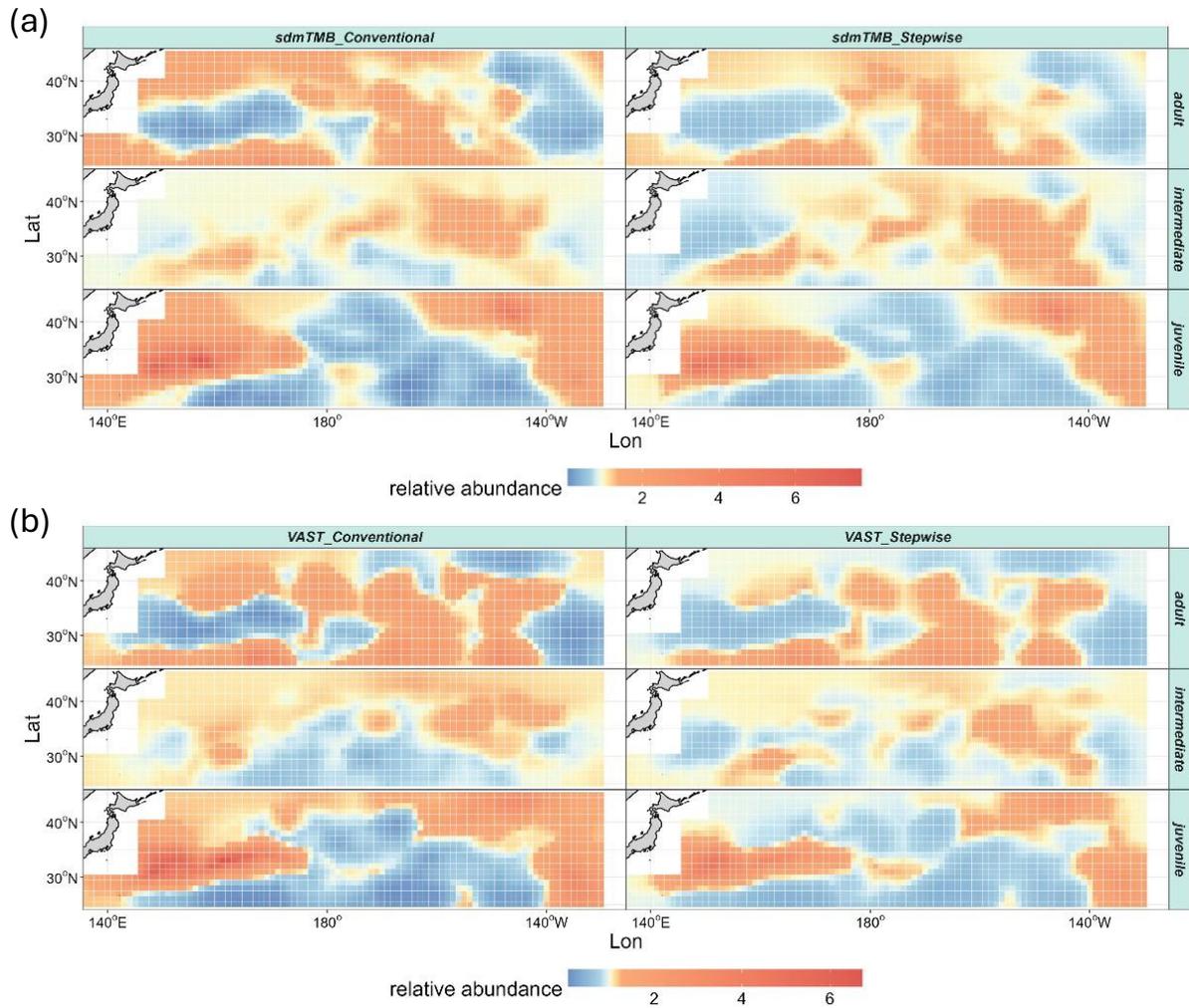


Figure 9. Estimated spatial density distributions by size group based on two estimation methods (conventional and stepwise) for the two modeling packages: (a) sdmTMB and (b) VAST.

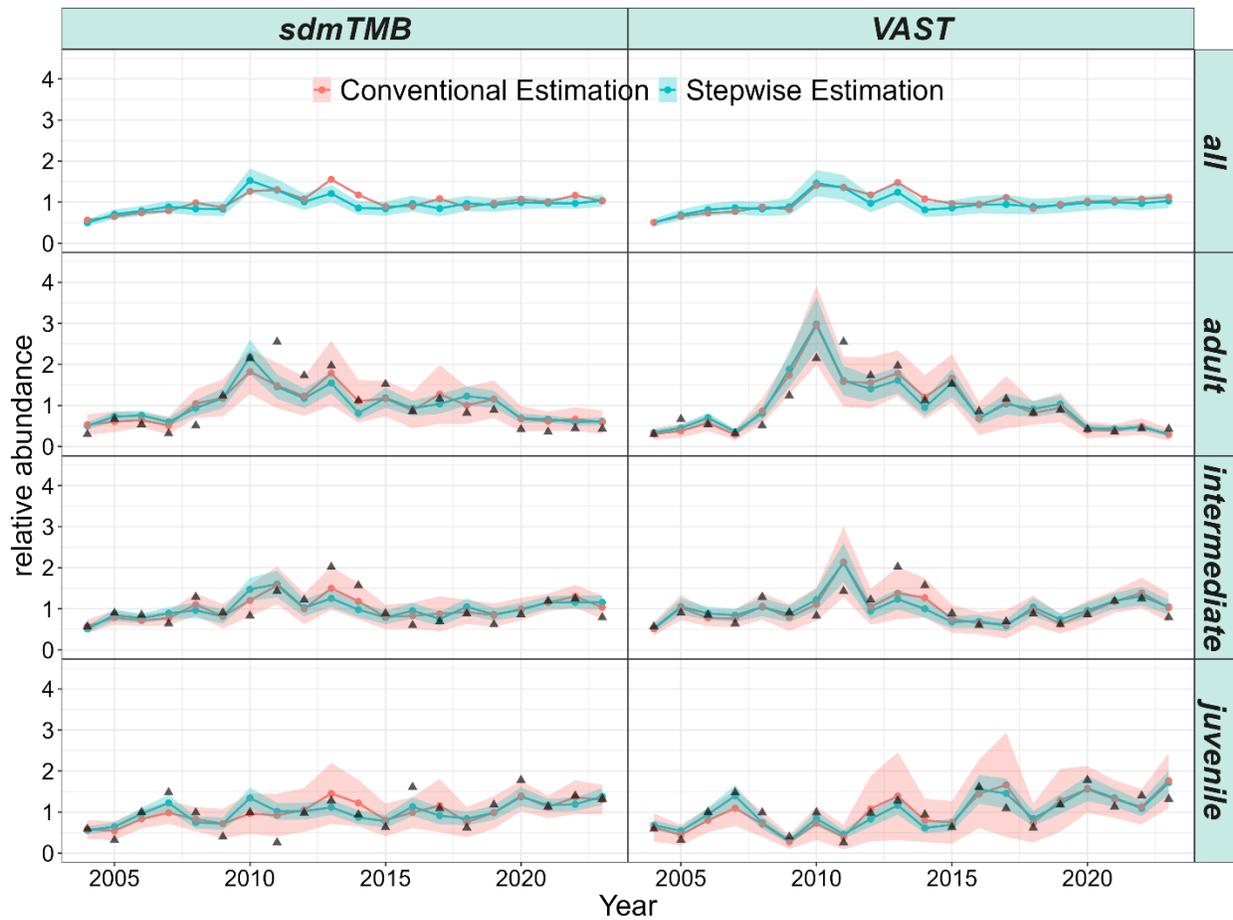


Figure 10. Annual relative abundance indices of albacore by different size classes, based on various spatiotemporal modeling packages (sdmTMB and VAST) and estimation methods (conventional and stepwise).