ISC/25/ALBWG-01/03

# CPUE standardization considering spatial fish-size distribution

Hirotaka Ijima, Naoto Matsubara, Marko Jusup 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

Email: ijima_hirotaka69@fra.go.jp


This working paper was submitted to the ISC Albacore Working Group Intercessional Workshop, 10-14 March 2025, held at the National Taiwan University, Institute of Oceanography in Taipei, Taiwan.

## Abstract

ISC albacore working group has used the area-as-fleet approach in North Pacific albacore stock assessment, assuming distinct length compositions in each region. However, annual fluctuations in age-specific distribution around area boundaries may impact CPUE standardization. We applied a mixture model to classify Japanese longline logbook data (1994–2023) based on mean body-mass groups. This method, defined at the year, month, and 5×5-degree block levels, provides a finer-scale alternative to the area-as-fleet approach. For CPUE standardization, we developed a spatiotemporal multi-species model using Tweedie and negative binomial distributions. Compared to the 2023 model and VAST model result, Tweedie model improved unrealistic spikes. The negative binomial model showed the steepest CPUE decline and large fluctuations, highlighting potential biases. Standardized CPUE for both juveniles and adult females declined, particularly around the Emperor Seamounts. Fishing effort has decreased in this region, requiring careful interpretation of CPUE trends. Future work includes refining models with GMRF-based intercepts, cross-validation for model selection, and residual validation. These improvements aim to enhance CPUE standardization accuracy and provide more reliable stock assessment indices.

## Introduction

The ISC Albacore Working Group (ALBWG) has adopted the area-as-fleet approach to define fleets for stock assessment model (ISC 2023). This method divides the ocean into different areas and assumes that fish with distinct length compositions are caught in each region (Waterhouse et al., 2014). Defining these areas as fleets allows the stock assessment model to account for distribution shifts as fish grow. Standardizing catch per unit effort (CPUE) by area is expected to provide indices corresponding to different growth stages. However, the age-specific distribution of North Pacific albacore appears to fluctuate annually (Ijima et al., 2023). Consequently, length composition within fixed areas may change over time, potentially impacting CPUE standardization. In recent years, CPUE has spiked sharply, possibly due to an influx of smaller fish into standardization areas, which may have artificially inflated CPUE (Nishimoto et al., 2024).

To overcome the limitations of the area-as-fleet approach, we propose an analysis using a mixture model. This method classifies fishing operations based on the size groups of captured fish, using catch weight and catch number recorded in logbooks as proxies for size composition (Ijima et al., 2023). Unlike the area-as-fleet approach, the mixture model defines clusters at the finer scale of year, month, and 1×1-degree blocks, avoiding classification into large spatial units. This approach enables more flexible tracking of annually variable size-based fish distributions.

Several challenges must be addressed before applying the mixture model approach to stock assessment for North Pacific albacore. First, length composition data and total catch are aggregated at the year, quarter, and 5×5-degree block levels. Therefore, the feasibility of classification with such coarse-resolution data must be examined. Second, CPUE standardization should be conducted using catch and effort records classified by size group. Conventional assessments have used Area 2 during the second quarter as the adult female index. However, since albacore fishing is most active during the first and fourth quarters, both the spatial definition and quarter selection should be reconsidered. Standardizing CPUE for each classified cluster is expected to provide more accurate CPUE estimates for each size group than the area-as-fleet approach. However, because some areas have overlap between size groups, an appropriate method for incorporating spatial effects must be developed. This remains a challenge.

In this study, we apply a mixture model to classify Japanese longline logbook data at the year, month, and 5×5-degree block levels to assess whether different size groups can be distinguished. For CPUE standardization, we develop a multi-species spatiotemporal model that integrates the clusters identified by the mixture model, allowing for improved representation of spatial patterns and reducing potential biases. This paper presents preliminary findings and examines key challenges and future research directions.

## Material and methods

### Japanese longline logbook data

We used Japanese longline logbook data from 1994 to 2023. The reports include the sizes of all vessels (coastal, offshore, distant water), along with catch numbers, catch weight, number of hooks, vessel names, operation locations, and gear types, such as Hooks Between Floats (HBF), for each operation. The mean body mass is calculated from the visually estimating recorded by fishermen and catch weight assumes catch number multiply mean body mass. This value is recorded either for each operation or for each trip. During the analysis, we did not use logs with eye measurements calculated per trip to apply coarse-scale information. To reduce computational costs, logbook data were aggregated by year, month, 1×1 location, vessel name, and HBF.

### Mixture model analysis

There is a spatial distribution of size class by each cohort, and the observed mean body mass data can be treated as a mixture distribution. Based on this, a mixture model was constructed under the assumption that mean body-mass follows a Lognormal mixture distribution as follows

$$p(x) = \sum_{k=1}^{K} \pi_k f(x|\sigma_k^2)$$

where $k$ is the number of clusters, $\pi_k$ is mixing coefficients vectors, and $\sigma_k^2$ is the standard deviation for the log-normal function $f$. The units of the clusters were set as year, month, and 5x5 degree blocks. We constructed mixture model using lat5lon5, Molat5lon5, Yrlat5lon5, and YrMolat5lon5, and the number of clusters was determined based on the Bayesian Information Criterion (BIC). The analysis was conducted using the R software package, Flexmix (Leisch 2004).

Using the results from the mixture model, each catch operation was assigned to a single size class bin, with all other size classes receiving NA. This classification allowed us to proceed with the CPUE standardization based on the identified size classes.

**CPUE standardization**

In the 2023 stock assessment, CPUE standardization was performed using second-quarter data. However, in this study, we used first-quarter data, which had the highest catch. To model the spatiotemporal distribution of albacore, we developed a multi-species model capable of incorporating multiple cohorts. We assumed that the CPUE for cluster $k$ in observation $i$ follows a Tweedie distribution:

$$CPUE_{k,i} \sim Tw(\mu_{k,i}, p_k, \phi_k),\ 1 < p_k < 2 \wedge \phi_k > 0,$$

where $p_k$ is the power parameter and $\phi_k$ is the dispersion parameter for cluster $k$. The expected value of CPUE ($\mu_{k,i}$) is defined as

$$\mu_{k,i} = \beta_k^0 + \varepsilon_{k,i}^{vessel} + \varepsilon_{k,i}^{hbf} + \mathbf{Z}_k\mathbf{L}_i$$

where $\beta_k^0$ is the intercept, $\varepsilon_{k,i}^{vessel}$ and $\varepsilon_{k,i}^{hbf}$ are random effects for vessels and HBF (Hooks Between Floats), and $\mathbf{Z}_k\mathbf{L}_i$ represents the spatiotemporal effect term with four latent spatial field.

As an alternative to the Tweedie distribution. We also considered the negative binomial distribution model with effort offset term in which the response variable is catch numbers.

We conducted parameter estimation using the R software package TMB (Kristensen et al., 2016). To capture the annual variation in distribution more accurately, we included the entire North Pacific in the analysis (Figure 1). Standardized CPUE was calculated by considering only nodes within the albacore distribution range (Figure 1). During this process, the effects of HBF and vessel names were treated as random effects, with their mean set to 0.

## Result and discussion

**Mixture model analysis**

Among all clustering configurations, YrMolat5lon5 had the lowest BIC values across all seasons (Figure 2). The BIC values continued to decrease as the number of clusters increased. To determine an appropriate number of clusters, we examined their distributions and visually assessed the results (Figure 3). Based on this assessment, we selected four clusters for the first quarter, three for the second quarter, and two for both the third and fourth quarters.

Catch volume was highest in the first quarter (Figure 4), and clusters were most clearly separated during this period (Figure 3). Given this representativeness, we propose using first-quarter logbook data for CPUE standardization. Additionally, we assume that the smallest size class in the first quarter represents juveniles, while the third largest size class corresponds to adult females. This suggests that separating logbook data by size group is feasible even at a coarse resolution.

**CPUE standardization**

This study estimated the standardized CPUE for juvenile and adult female albacore using a spatiotemporal multi-species model based on the Tweedie distribution (Figure 5). For juveniles, standardized CPUE differed significantly from the nominal CPUE (Figure 5a). However, the lower bound of the 95% confidence interval dropped below zero, indicating high estimation uncertainty. This highlights limitations in the accuracy of the standardization. Moreover, the CPUE for juveniles has been exceptionally low in recent years (Figure 5a). This is due to a decline in CPUE around the Emperor Seamounts (Figure 6). However, fishing effort in this region has decreased annually, with very few operations recorded in recent years. Thus, careful evaluation is needed to determine whether the CPUE decline reflects stock depletion or data insufficiency. The standardized CPUE for adult females also declined in recent years, while the nominal CPUE increased, creating a substantial discrepancy (Figure 5b). This likely results from removing targeting effects for albacore in the southern regions of Japan by CPUE standardization (Figure 7).

To assess the impact on stock assessment, we compared the results of the Tweedie model with the model used in the 2023 stock assessment (2023 model; Nishimoto et al., 2023), the VAST model (Matsubara et al., 2024), and the negative binomial model (Figure 8). These models differ in terms of input data and explanatory variables, thus direct comparisons are challenging. However, given that CPUE trends are generally assumed to correspond with stock trends, we visualized and compared their relative changes. The comparison was conducted by evaluating the relative CPUE from each model. The 2023 model was an updated version developed on first-quarter logbook data from Area 2. Similarly, the CPUE estimated using VAST model was standardized on first-quarter data from Area 2. The linear regression term of negative binomial model was same as the Tweedie model used in this study. The comparison revealed that the 2023 model and the VAST model exhibited similar overall trends. Both models adopted spatiotemporal frameworks and shared the same area, resulting in consistent fluctuation patterns. In contrast, the negative binomial model showed the steepest decline among the four models, with a pronounced spike in 2005 and an unresolved sharp increase in 2021. Previous studies have suggested that the negative binomial model may overestimate the rate of decline when applied to Japanese longline logbook data (Minami et al., 2007), indicating the need for caution in its use. The Tweedie model exhibited a steeper slope than the 2023 and VAST models but was less extreme than the negative binomial model. Notably, the sharp increase in 2021 observed in the negative binomial model was not present in the Tweedie model. Given that the Tweedie distribution accounts for zero-inflated data, it is expected to provide more accurate estimates than the negative binomial model.

## Future works

- To improve estimation accuracy in areas with no observations, consider a model where the intercept follows a GMRF, similar to VAST model.
- Construct models with different combinations of covariates and varying numbers of SF, and select the best model through cross-validation.
- Compute randomized quantile residuals for model validation.
- Summarize the results and publish them as a research paper.

## References


Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H. and Bell, B.M., 2016. TMB: automatic differentiation and Laplace approximation. Journal of statistical software, 70, pp.1-21.

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.

ISC. 2023. Stock assessment of albacore tuna in the north pacific ocean in 2023. 23rd Meeting of the International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean. Kanazawa, Japan July 12-17, 2023 Annex 8.

Leisch, F., 2004. FlexMix: A general framework for finite mixture models and latent class regression in R. Journal of Statistical Software, 11, pp.1-18.

Matsubara, N., Xu, H., Ijima, H., and Tsuda, Y., 2024. Standardized CPUE and size for North Pacific Albacore by Japanese Longline using VAST package. ISC/24/ALBWG-1/02

Minami, M., Lennert-Cody, C.E., Gao, W. and Román-Verdesoto, M., 2007. Modeling shark bycatch: the zero-inflated negative binomial regression model with smoothing. Fisheries Research, 84(2), pp.210-221.

Nishimoto M., N. Matsubara, H. Ijima, and Y. Tsuda., 2024. Update standardized CPUE for North Pacific albacore caught by the Japanese pole and line from 1972 to 2022. ISC/24/ALBWG-1/03

Waterhouse, L., Sampson, D.B., Maunder, M. and Semmens, B.X., 2014. Using areas-as-fleets selectivity to model spatial fishing: asymptotic curves are unlikely under equilibrium conditions. Fisheries Research, 158, pp.15-25.

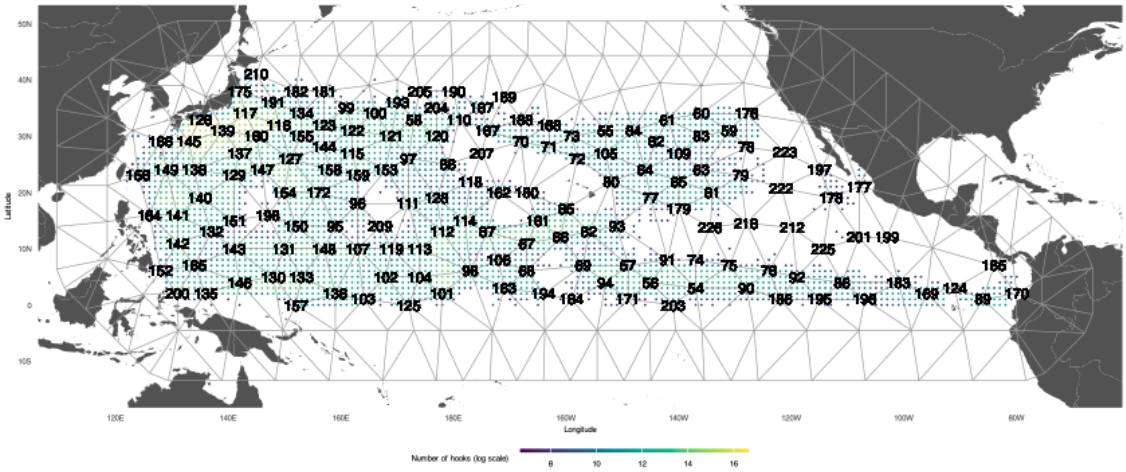

Figure 1. Area coverage for the North Pacific albacore CPUE analysis. We defined a total of 304 nodes in the Pacific Ocean. For CPUE standardization, we used data from nodes located in waters where albacore are present between 0°N and 40°N. Points indicate operation locations, and colors represent the total log number of hooks.

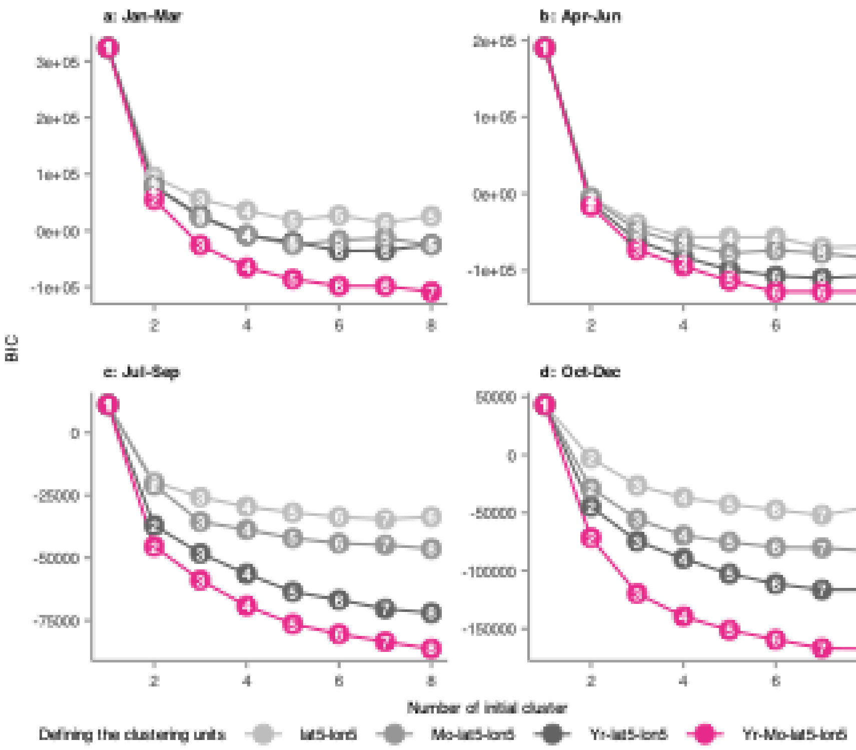


Figure 2. BICs for each mixture model.

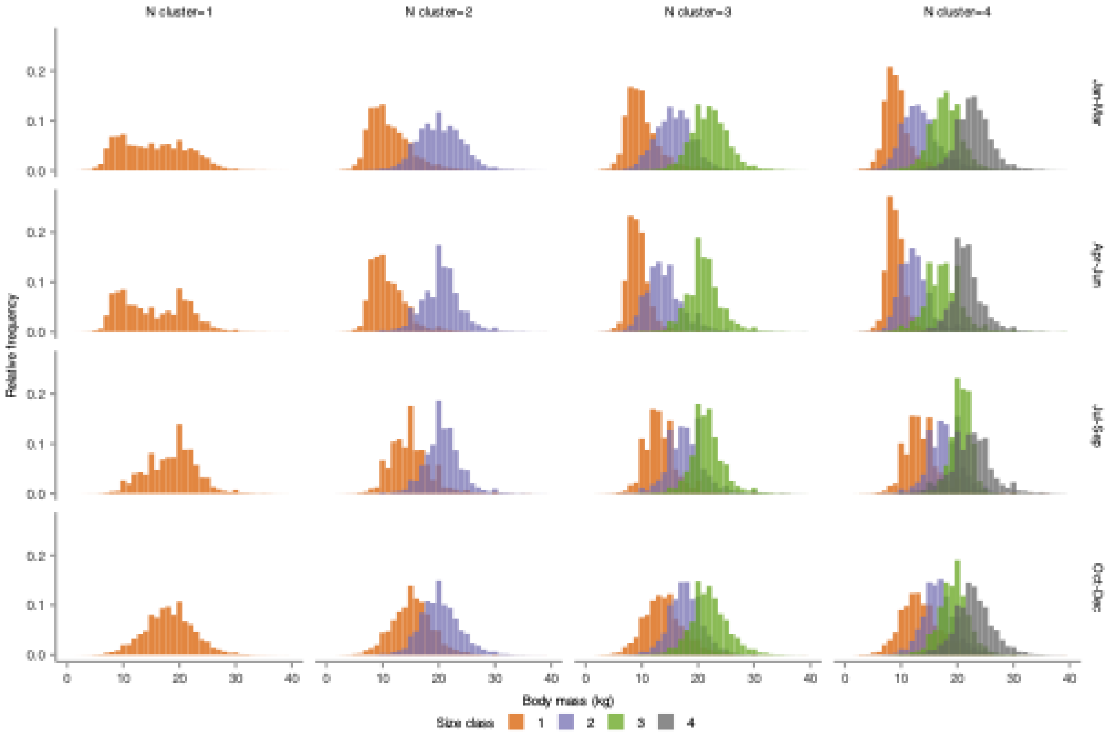


Figure 3. Distribution by classified size class.

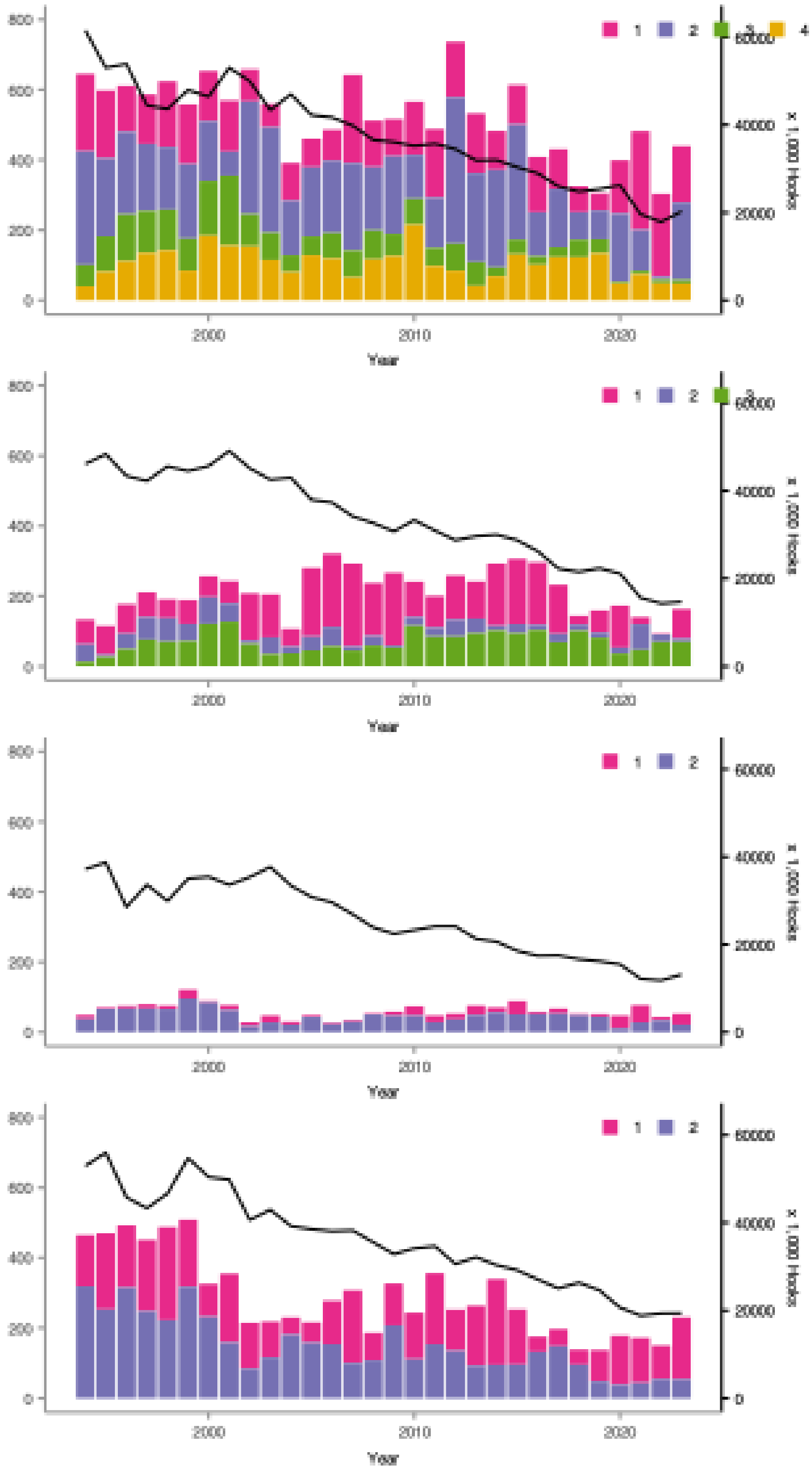


Figure 4. The number of fish caught per cluster (histogram) and the total fishing effort (black line).

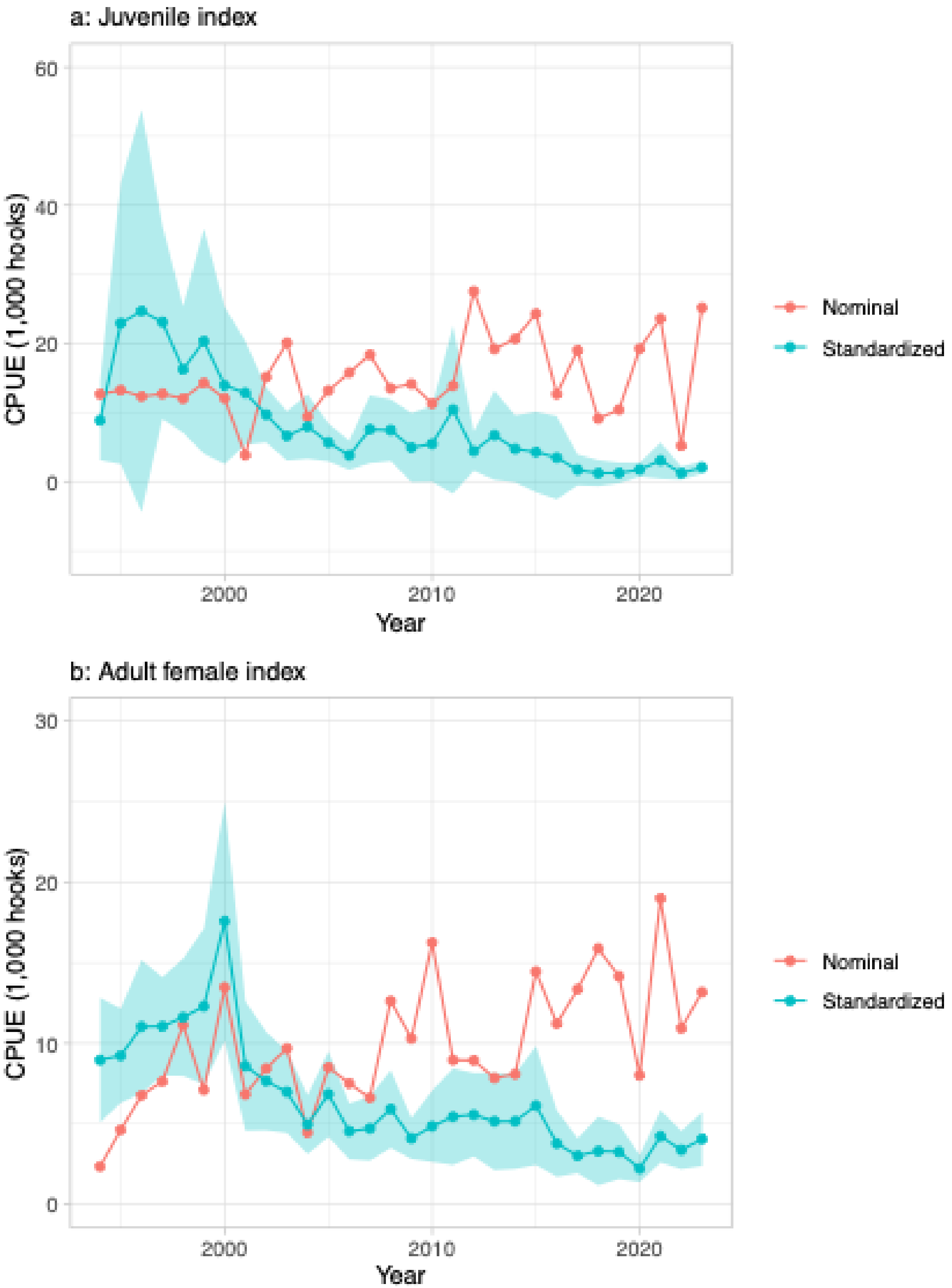


Figure 5. Size class based standardized CPUE. Red line indicated nominal CPUE, while blue line and shade means the standardized CPUE and confidence interval from 5 to 95%, respectively.

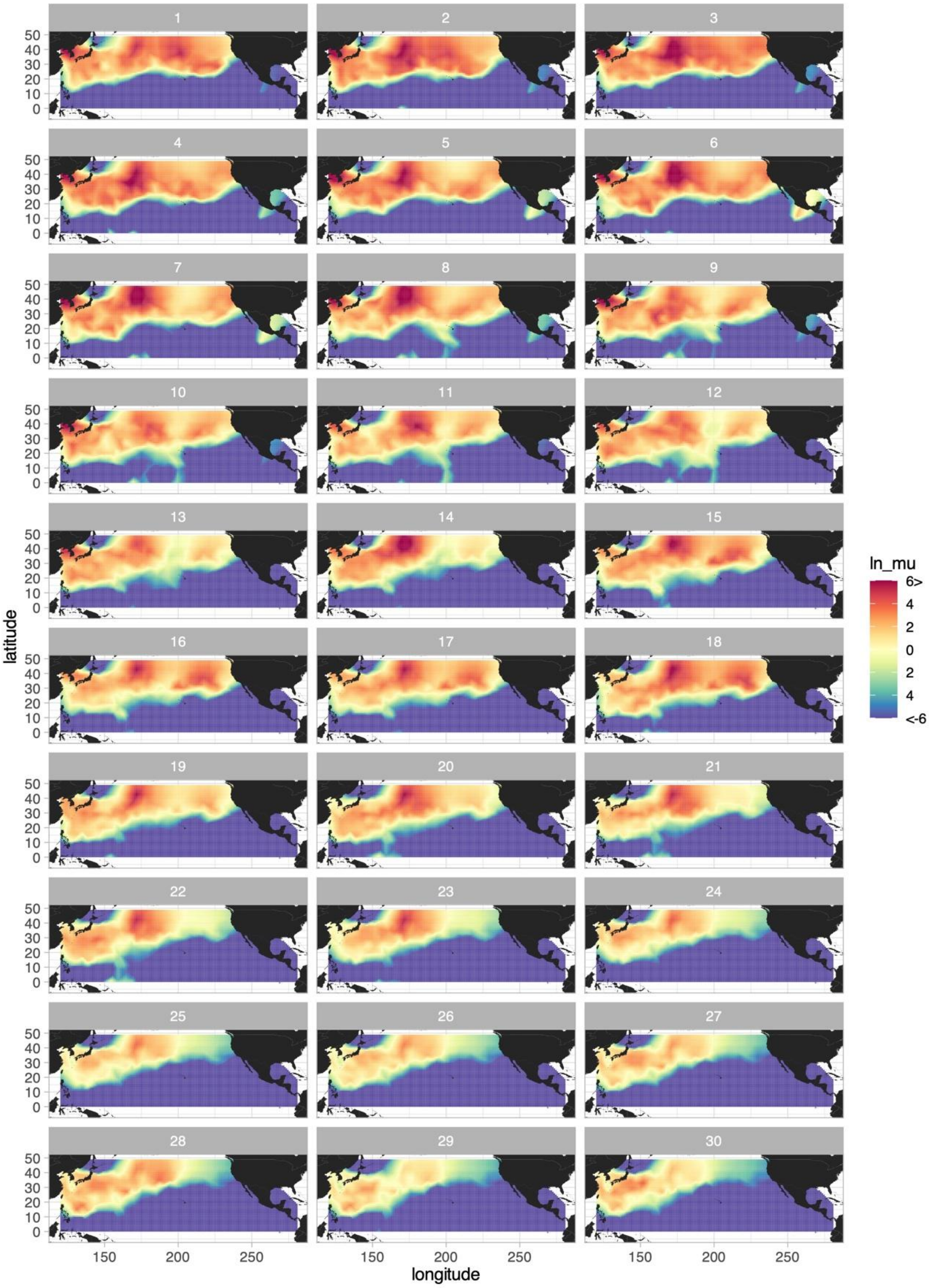


Figure 6. Spatiotemporal pattern of standardized juvenile log CPUE.

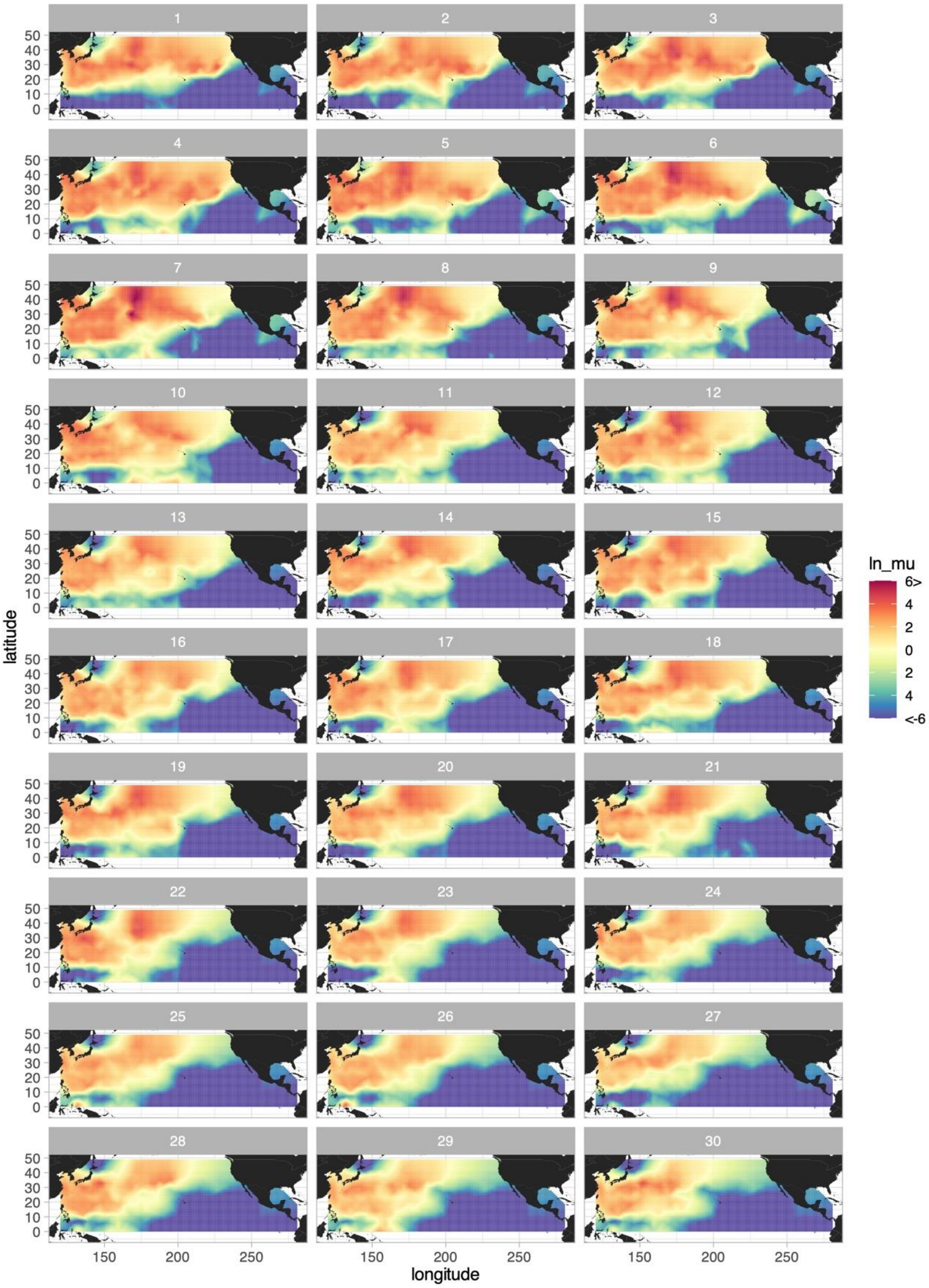


Figure 7. Spatiotemporal pattern of standardized adult female log CPUE.

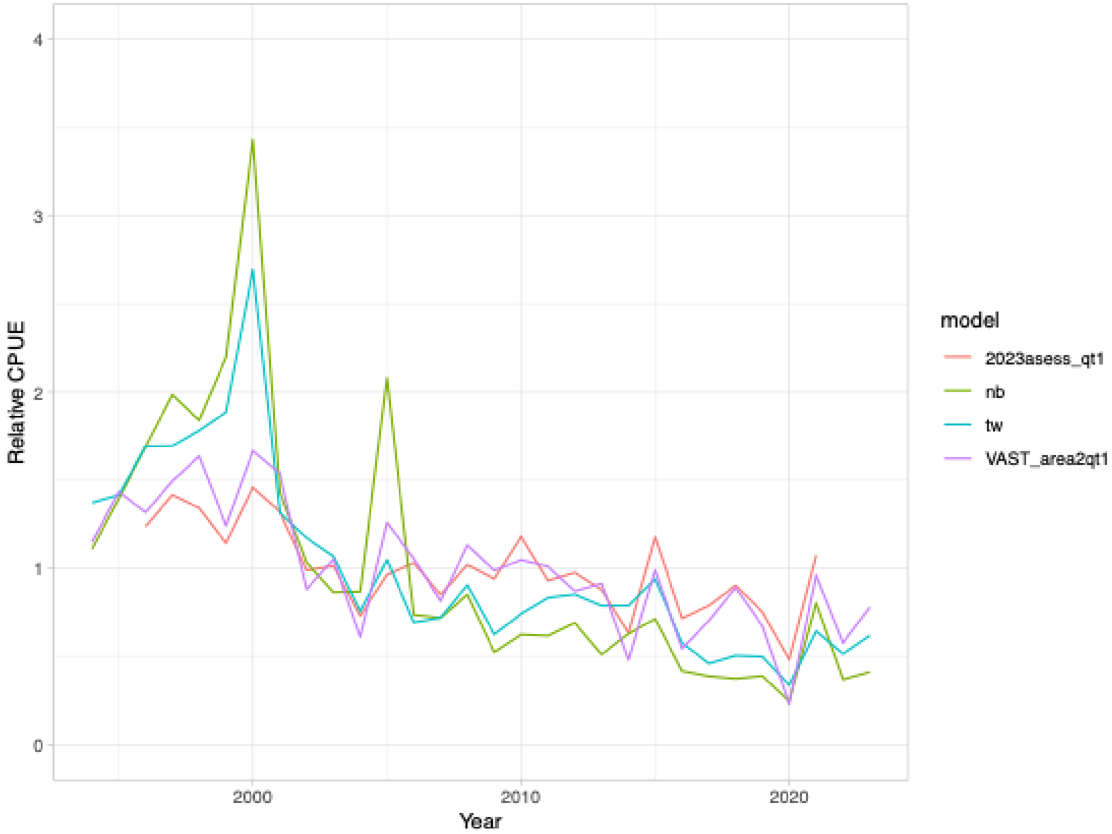


Figure 8. Comparison between different CPUEs. Red line showed the standardized CPUE which used for previous stock assessment in 2023. Green and blue lines showed the results of standardized CPUE with negative binominal and Tweedie distribution model, respectively. Purple line indicated the standardized CPUE which calculated using VAST package (Matsubara et al. 2025).