**ISC/24/ALBWG-01/06**

# Preliminary analysis for size-based abundance indices considering multiple latent spatial fields[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**

---

[1] This working paper was submitted to the ISC Albacore Working Group Intercessional Workshop, 11-18 March 2024, held at the Fisheries and Oceans Canada Institute of Ocean Sciences, Victoria, Canada.

## Abstract

The ISC Albacore Working Group adopts an Area-as-Fleet approach, considering that Japanese longline vessels catch albacore of different sizes and genders in different areas. However, multiple cohorts are caught in the same area, and the main distribution of albacore may extend beyond the designated areas by year. Japanese researchers have developed a new fisheries definition methodology using a finite mixture model. This approach allows the classification of multiple cohorts in the same area without the need for boundary setting, thus improving the accuracy of stock assessment. In this study, we attempted to standardize CPUE using logbook data sorted by fish size. As a result, the preliminary model converged, and standardized CPUE for four size groups was obtained. We plan the additional analysis, including discrete distribution models with offset terms, adding an average spatial field, developing an alternative model for three-size groups, and excluding fixed effects by year. We will summarize these results and prepare a manuscript for submission.

## Introduction

The ISC Albacore Working Group (ALBWG) adopts an Area-as-Fleet approach, considering that Japanese longline vessels catch albacore of different sizes and genders in different areas (ISC 2023). This method assumes that albacores of different sizes and genders are caught in each area. However, it has been pointed out that multiple cohorts may be caught in the same area or that the primary distribution of albacore may extend beyond designated areas (ISC 2023). Japanese researchers have developed a new fisheries definition methodology in response to these challenges using a finite mixture model (Ijima et al. 2023, Matsubara et al. 2024). This approach allows for capturing multiple cohorts in the same area without the need for area designation, thus improving the accuracy of stock assessment. However, due to the precise spatial distribution of fish size based on classification, there is an issue of generating significant zero catches when attempting to standardize Catch Per Unit Effort (CPUE) for specific fish sizes. To address this problem, we attempted to standardize CPUE using Japanese logbook data sorted by a finite mixture model, constructing a spatiotemporal model combining regression models for all cohorts. This study is a first-cut report. Thus, we will outline work plans based on these analytical results.

## Material and Methods

### Japanese longline logbook data

The ALBWG assumes that albacore spawning in the North Pacific peaks in the second quarter. To gauge adult albacore abundance, the group relies on CPUE data from areas predominantly inhabited by female albacore in the second quarter (Matsubayashi et al. 2023). Consequently, our analysis focuses exclusively on Japanese longline logbook data from the second quarter. In the Japanese longline logbook, fish caught in each operation are sorted into size groups. These groupings are determined based on flags assigned for each year, month, and five-by-five degree grid, derived from finite mixture model analysis results.

Furthermore, the logbook data is aggregated by year, month, size group, one-by-one degree grid, vessel name, and number of hooks between floats (HBF). CPUE is expressed as the number of albacores caught per 1000 hooks. Records with fewer than 500 hooks are considered unrealistic, and operations with CPUE exceeding 200 are excluded.

### Multi-cohort model

According to the finite mixture model analysis result, the logbook data can be divided into four size groups with a significant bias towards certain areas where the albacore of each size group was caught (Figure 1). This study constructed a multi-species model to address this issue. This model can complement areas with a low distribution density for each size group.

The response variable CPUE at the data point *i* of size (cohort) class *j* follows a continuous Tweedie distribution, $CPUE_{i,j} \sim Tw(\mu_{i,j}, p_j, \phi_j), 1 < p_j < 2 \wedge \phi_j > 0$ , where $\mu_{i,j}$ is the mean of Tweedie distribution, $p_j$ is the power parameter, and $\phi_j$ is the dispersion parameter of Tweedie distribution. The mean of CPUE $\mu_{i,j}$ depends on the linear regression model as follows;

$$\mu_{i,j} = \exp\left(\boldsymbol{\beta 0}_j + \mathbf{X}_i \boldsymbol{\beta}_j + \mathbf{V}_i \mathbf{v}_j + \mathbf{H}_i \mathbf{h}_j + \mathbf{Z}_i \mathbf{z}_j\right).$$

Where $\boldsymbol{\beta 0}_j$ is the intercepts, $\mathbf{X}_i$ is fixed effect variables with coefficients $\boldsymbol{\beta}_j$. This study, we set year and month as fixed effect variable. $\mathbf{v}_j$ is random effect for vessel name, $\mathbf{h}_j$ is random effect for HBF. $\mathbf{V}_i$ and $\mathbf{H}_i$ are the design matrix for each random effects variable. $\mathbf{Z}_i$are spatiotemporal latent spatial fields, and $\mathbf{z}_j$ is the coefficient corresponding to the latent spatial fields. We arbitrarily set several latent spatial fields and applied the stochastic partial differential equation (SPDE) approach for latent factor estimation (Cameletti et al. 2013). All parameters and latent variables are estimated by an R software package TMB (Kristensen et al. 2015). Using estimated values, we calculated the "lsmeans" of each size class CPUE.

## Result and Discussion

The preliminary model converged, yielding standardized CPUE for four size groups (Figure 2a, b). The spatial trends of standardized CPUE for each size group reflect the results of the finite mixture model, but high trends are observed in areas with no catches, and the estimated scales tend to be larger than reality (CPUE > 1,000 fish / 1,000) (Figure 2b, c). Additionally, due to mistakes in the program, adjustments are needed for the 95% confidence intervals of estimated annual trends. Significant spatial fluctuations are observed in the presumed size group of adult females, and the distribution may extend beyond the previously assumed areas (Figure 3). Therefore, this new method would be more accurate for understanding the annual trends of adult female albacore. However, investigation and improvement are necessary to address the issue of high CPUE in areas with no catches or low effort.

This analysis represents an initial attempt, and future improvements are expected. We will continue model developing. When we obtain the best model, we will summarize the new findings and prepare the manuscript for journal submission. The following are the planned next steps.

**1. Exploring models with offset terms:** We will investigate models incorporating offset terms, such as the negative binomial distribution, to assess the impact of observed high CPUE in low-effort operations on standardization. We noted that CPUE reached a maximum value of 500 despite minimal effort in this analysis.

**2. Addition of average spatial effects:** Incorporate VAST's standard settings to improve scale estimation accuracy by assuming average spatial distributions (Thorson et al. 2019). Additionally, explore scenarios involving two or four latent spatial fields alongside the current analysis, which assumes three latent spatial fields.

**3. Model development for three size groups:** The current analysis is limited in providing complementary estimates of latent spatial fields due to overlapping two juvenile distributions. In addition, small juveniles were not classified in 2000 (Figure 2, b). Thus, we will apply a simplified model consolidating similar juvenile data into a single group. As the finite mixture model effectively separates adult albacore data, simplifying the model is viable.

**4. Development of models without fixed effects**: None of the estimated fixed year effects were statistically significant (Figure 4). Including both temporal-spatial and fixed year effects in the model may lead to redundancy in accounting for yearly variations. Furthermore, incorporating fixed year effects may not necessarily enhance predictive accuracy. Given the hierarchical complexity inherent in such models, the conventional AIC or BIC criterion is not applicable, and cross-validation emerges as a more suitable method for evaluating estimation accuracy (Watanabe 2021). However, cross-validation poses computational challenges, particularly in terms of processing time. Therefore, we will utilize TMBstan (Monnahan and Kristensen, 2018) to generate posterior distributions and assess the necessity of fixed effects using WAIC (Watanabe 2013).

## References

ISC., 2023 Stock assessment of Albacore tuna in the north Pacific ocean in 2023. ANNEX08 23rd Meeting of the International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean Kanazawa, Japan July 12-17, 2023.

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.

Matsubara, N., Nishimoto, M., Tsuda, Y., and Ijima., H. 2023 Updated analysis for Japanese longline logbook data by finite mixture model. ISC/24/ALBWG-01/XX.

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.

Cameletti, M., Lindgren, F., Simpson, D. and Rue, H., 2013. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. AStA Advances in Statistical Analysis, 97, pp.109-131.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H. and Bell, B., 2015. Template model builder TMB. J. Stat. Softw, 70, pp.1-21.

Monnahan, C.C. and Kristensen, K., 2018. No-U-turn sampling for fast Bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan R packages. PloS one, 13(5), p.e0197954.

Thorson, J.T., 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research, 210, pp.143-161.

Watanabe, S., 2013. A widely applicable Bayesian information criterion. The Journal of Machine Learning Research, 14(1), pp.867-897.

Watanabe, S., 2021. WAIC and WBIC for mixture models. Behaviormetrika, 48(1), pp.5-21.

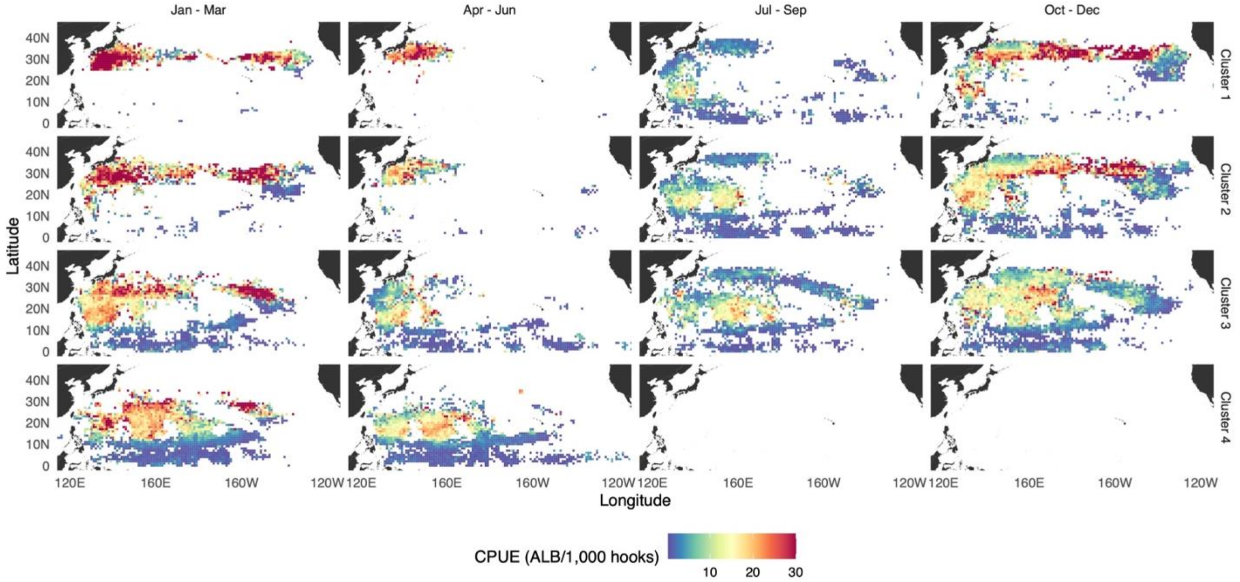

Figure 1. Spatial distribution of north Pacific albacore CPUE organized by size class.

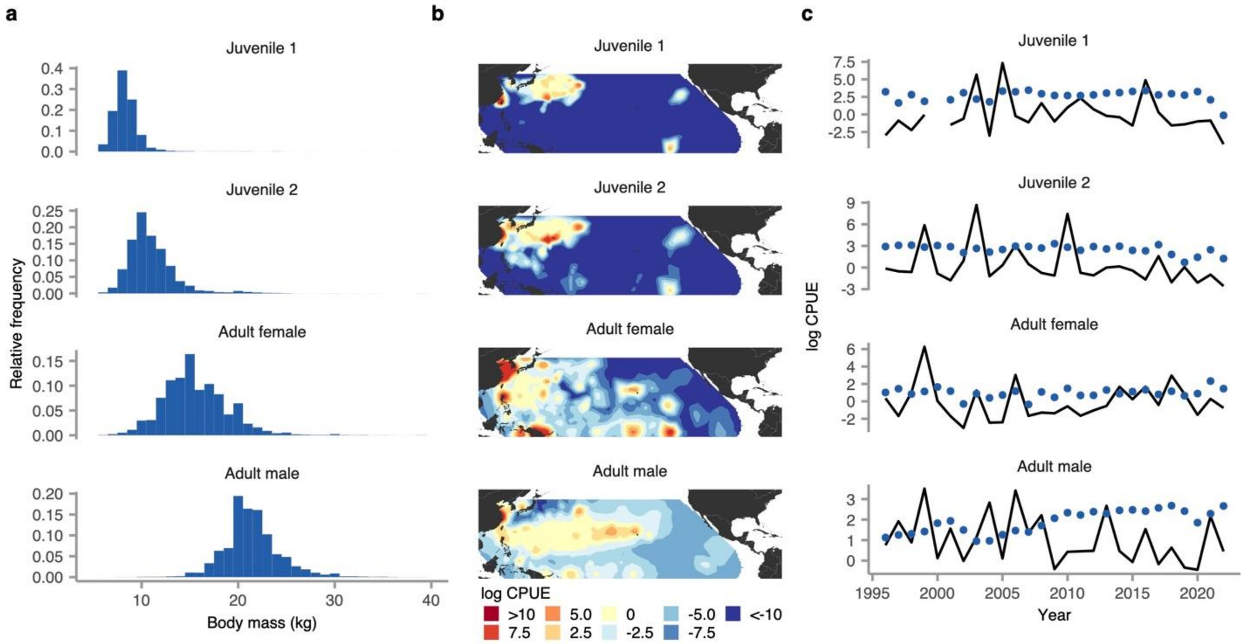

Figure 2. Standardized albacore CPUE in the second quarter. **a**. Histograms classified by a finite mixture model. **b**. Standardized spatial CPUE by size classes. **c**. Annual trends in size-based CPUEs, solod line is standardized CPUE, blue circle denotes nominal CPUE.

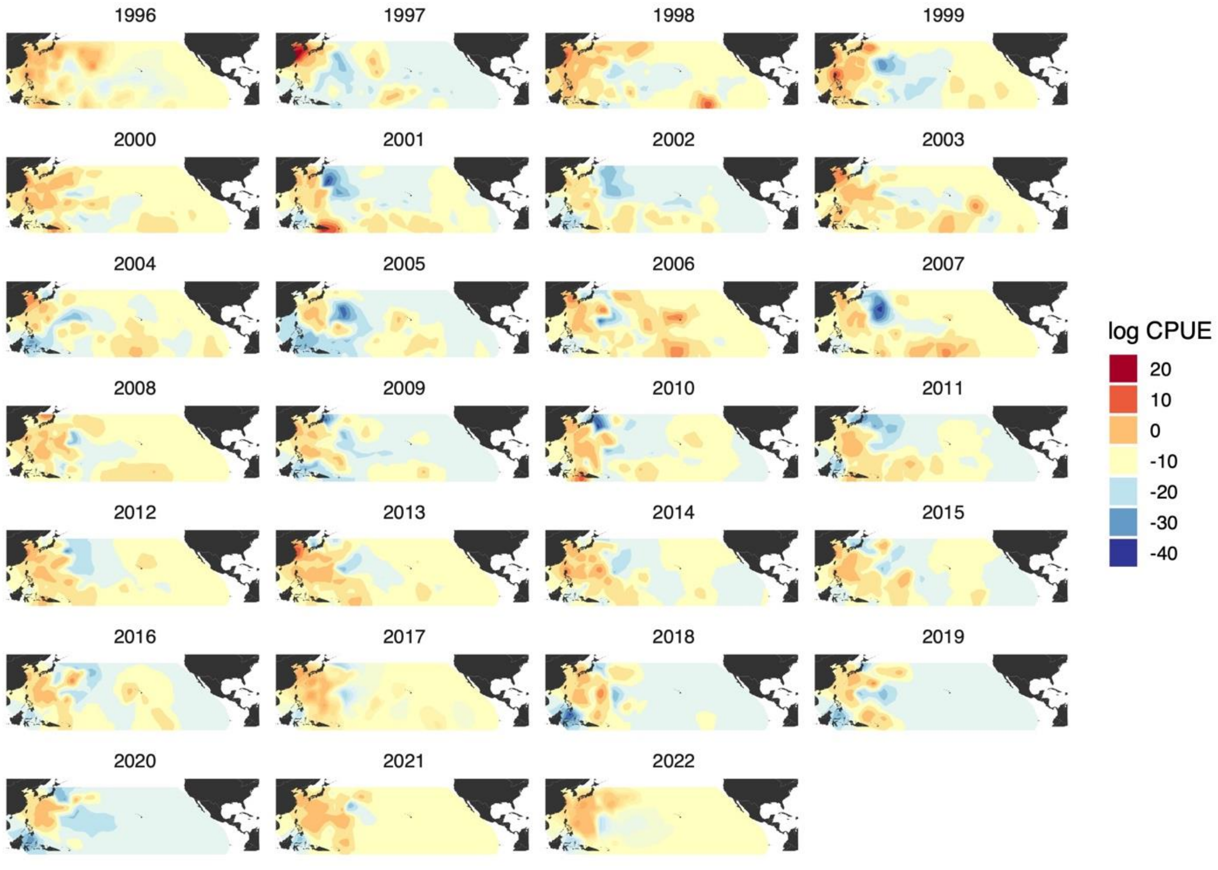

Figure 3. Estimated spatiotemporal pattern of adult female albacore CPUE.

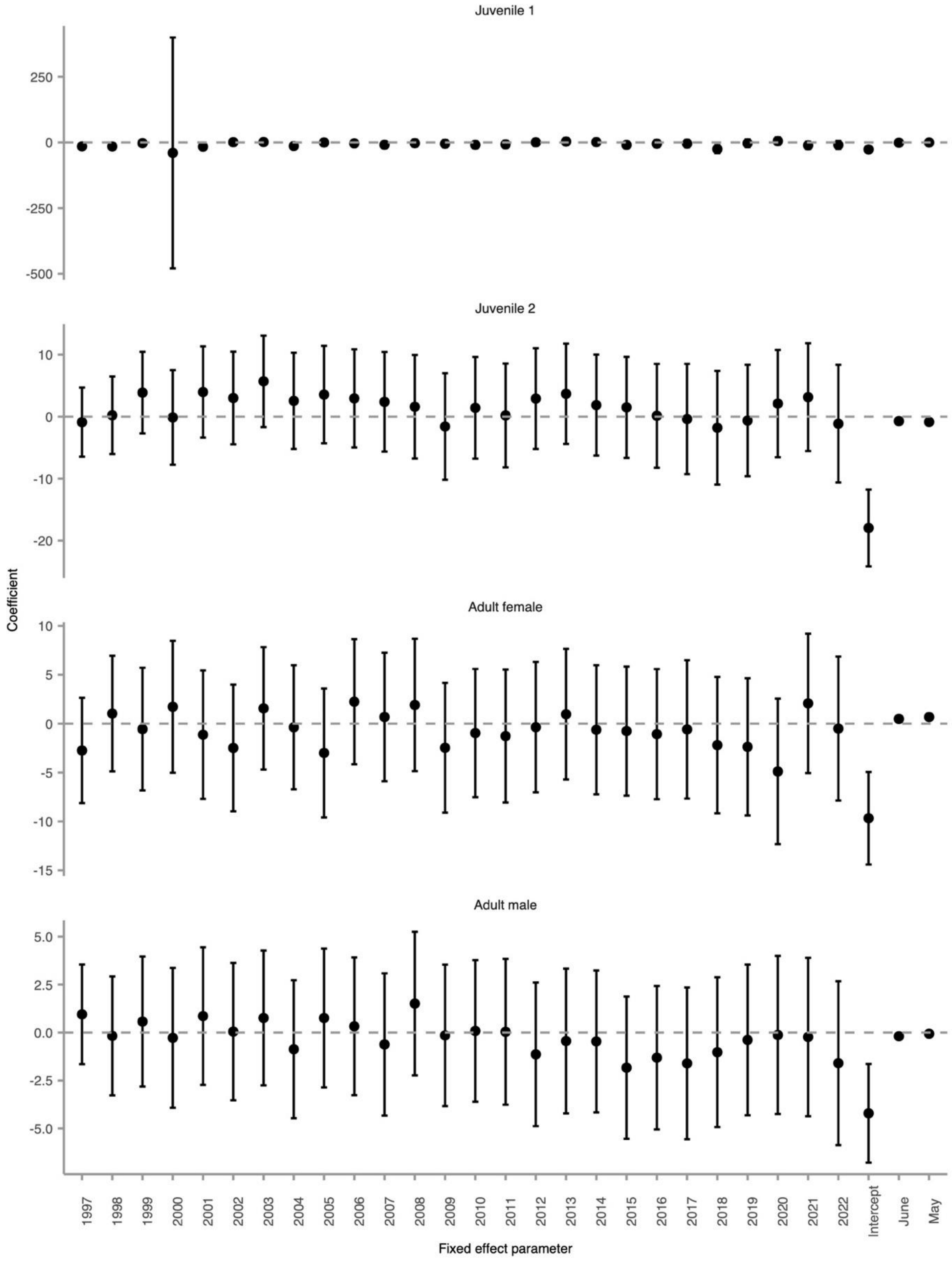

Figure 4. Estimated fixed effect parameters. The black circle is the point estimate value, and the error bar denotes 95% confidence intervals.