SCIENTIFIC COMMITTEE
NINETEENTH REGULAR SESSION

Koror, Palau
16 – 24 August 2023

Size-dependent distribution for bigeye and yellowfin tuna in the Pacific Ocean

WCPFC-SC19-2023/SA-IP-15

Hirotaka Ijima[1] and Keisuke Satoh[1]

[1] Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Yokohama 236-8648, Japan

SC19-SA-IP-15

Title: Size-dependent distribution for bigeye and yellowfin tuna in the Pacific Ocean.

Authors

Hirotaka Ijima[1] and Keisuke Satoh[1]

[1] Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Yokohama 236-8648, Japan.

Abstract

Yellowfin and bigeye tuna are highly migratory fishes found across the Pacific Ocean. Besides, the composition of their body size from fishery varies from region to region, possibly due to differences in stock structure or growth-dependent migration. To account for these ecological factors, the stock assessment model constructs a meta-population model, including fish migration, or assumes the area-as-fleet approach in which fishery is defined by delimiting a sea area. However, delimiting areas by specific latitudes and longitude may affect stock assessment accuracy because fish distribution continuously changes season and annually. This paper used a finite mixture model to classify Japanese longline logbook data for yellowfin and bigeye tuna all over the Pacific Ocean based on the mean body mass (semi-dress weight) of the fish caught. Additionally, we developed a multi-species geostatistical model using all latent size classes as response variables and attempted to standardize CPUE based on stock structure or cohort. Although the preliminary model converged, we could not obtain standard deviations of standardized CPUE for some years. We will continue this analysis to improve the problems.

Introduction

Yellowfin and bigeye tunas are called highly-migrate predator fish. Their distribution extends to equatorial areas to high latitude ranges, and they are caught throughout the Pacific Ocean by purse seiners, longlines, and coastal fisheries. To consider their ecological and fishery characteristic, the WCPFC has developed a stock assessment model assuming meta-populations in the Western Central Pacific Ocean (Hampton and Fournier 2001, Vincent et al. 2020, and Ducharme Barth et al. 2020). This model uses boundaries to estimate movement rates between areas. On the other hand, for these two tuna species, there have been reports of fish size varying depending on fishery location (Figure 1 and Figure 2). The differences in length composition in different areas could be due to various factors, such as differing distributions by growth stage (cohort) or stock structures. An area-as-fleet approach is practical when the distribution area varies by growth stage (Waterhouse et al. 2014). This approach involves separating the cohorts by areas based on length composition data. Thus, stock assessments often define areas for migration and fish growth stages. However, these approaches have their

challenges. Migration and distribution patterns can vary seasonally and annually, and their boundaries are not always linear. Additionally, different cohorts or stock structures may occupy the same space, which differs from the stock assessment model's assumptions and could largely impact the stock assessment results.

This study aims to find a new solution to these problems. We started by analyzing the mean body mass (semi-dressed weight) data from Japanese logbooks using a finite mixture model to identify different types of body masses. Then, divided logbook data based on the estimated latent classes. Finally, we constructed a geostatistical model with multiple response variables to help standardize the size-based CPUE for each class of body mass.

Material and Methods

Japanese logbook data

We used Japanese longline logbook data available since 1952 but only analyzed data from 1994 and onwards. Because the logbook data tabulation method had changed by 1994, the mean body mass by operation could not use before 1993. We were also able to include data from coastal vessels after 1994.

It is not easy to examine the effect of the Hooks Between Floats (HBF) during two periods (before 1994 and after 1994). It was considered that most vessels used nylon lines after 1994. On the other hand, before 1994, hemp lines were mainly used, and the depth of HBF seems different from late periods by their weight. Thus, the HBF could be treated as a proxy for gear depth in this study.

Our analysis utilized data from all Pacific oceans to examine how cohort and stock structure distribution varies seasonally and annually. The Japanese longline logbook recorded the mean body mass for each operation, or in cases where a trip calculated a mean body mass, the logbook on one operation recorded the value for body mass during that specific trip. To ensure accuracy, we calculated the standard deviation of body mass for each trip and excluded any trips with zero standard deviation data (Figure 3). We did not perform any additional data filtering.

Finite Mixture model

Finite mixture models are unsupervised machine learning that could classify latent classes in the Japanese logbook data (Ijima et al. 2023). This study constructed a simple model, assuming that the mean body mass follows a Kth mixture of the lognormal distribution as

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

where x mean-body mass from the logbook, π_k denotes mixing coefficients, and σ_k^2 is a variance of the lognormal distribution. The grouping factors were year, month, and 1° x 1°degree grid. The latent class k was set at 1-6, and the Bayesian Information Criterion (BIC) was calculated for each latent class model. We used the R software package, "flexmix," for all analyses (Leisch 2004).

Multi-species geostatistical model

The Finite Mixture model classifies the data in the logbook. However, each CPUE data class tends to be distributed in different ocean areas. If we construct a model for each class, we need to explore a large, biased zero-catch space. Handling zero-catch in the standardization of CPUE is challenging, and various methods have been proposed to address this issue. To tackle this problem, we have developed a multi-species model that complements the observed zero-catch values between classes.

We used the CPUE (Number of fish caught / 1,000 hooks) of data i by latent class k as the response variable, which is a continuous variable following a Tweedie distribution,

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

where $\mu_{i,k}$ is the mean of Tweedie distribution, p_k is the power parameter, and ϕ_k is the dispersion parameter. The mean CPUE follows linear regression,

$$\mu_{i,k} = \exp(\alpha_k + \beta_{y,k} y_{i,k} + v_{i,k} + m_{i,k} + h_{i,k} + \gamma_k^1 s_{i,k}^1 + \gamma_k^2 s_{i,k}^2 + \gamma_k^3 s_{i,k}^3),$$

$$v_{i,k} \sim N(0, \sigma_v^2),\ m_{i,k} \sim N(0, \sigma_m^2), \text{ and } h_{i,k} \sim N(0, \sigma_h^2),$$

where α_k is the intercepts, $y_{i,k}$ is categorical year variable with coefficients $\beta_{y,k}$, $v_{i,k}$ is random effect for vessel name, $m_{i,k}$ is random effect for the month, and $h_{i,k}$ is random effect for HBF. $s_{i,k}^*$ are spatiotemporal latent factors, and γ_k^* is the coefficient for the latent factors. We arbitrarily set several latent factors 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). We aggregated the logbook data by year, month, 1° x1° grid, vessel name, and HPB for faster computation. Finally, using estimated values, we calculated the "lsmeans" of each latent class CPUE.

Result

As the number of latent classes increased, the BIC of the Finite Mixture model decreased for both yellowfin and bigeye tuna (Figure. 4). We summarized the yellowfin tuna into different latent classes and observed a distinct spatial pattern and size distribution for classes three and four (Figure. 5). However, the size separation became unclear when the latent class was set to five (Figure. 5). The spatial pattern of the bigeye tuna became clearer when considering latent class assumptions (Figure. 6). However, there was no clear difference in the size distribution (Figure. 6). Especially the fifth body mass composition of the five latent classes overlapped almost entirely with the other distributions (Figure 6). Considering these results and computational cost, we aimed to develop a preliminary standardized CPUE for bigeye and yellowfin tuna by assigning the latent class a three-value.

The mean body mass of the three latent classes (K1, K2, and K3) are about 20, 25, and 40 kg, respectively (Figure 5 b). These classes exhibited distinct spatial distribution patterns, but some overlap was observed (Figure 5 a). Latent class K1 of yellowfin can be found throughout the western Pacific. However, spikes in annual CPUE between 2009 and 2019 prevented standard deviation

estimates for those years (Figure 7). Similarly, K2 is also present in the western Pacific, but its density tends to decrease in the western high latitudes and increase in the central high latitudes (Figure 7). The standard deviation of standardized K2 CPUE was unavailable for the period spiking 2017-2020 (Figure 7). The K3 is mainly distributed in the eastern Pacific, with increasing densities in southern and lower latitudes, while annual CPUE shows a decreasing trend. The standardized CPUE of K3 was also unavailable for a standard deviation for some years (Figure 7).

Like yellowfin tuna, bigeye tuna is divided into three categories according to weight (K1, K2, and K3). K1 weighs about 38 kg, K2 weighs 25 kg, and K3 weighs 20 kg (Figure 8). Each class has different distribution patterns, but there is some overlap (Figure 8). K1 is distributed widely in the eastern Pacific 20°S-20°N and the western Pacific near the equator (Figure 8). The trends in the standardized CPUE are flatter than nominal in all years (Figure 8). K2 has a low-density north of 20N in the western Pacific, and its trends in the standardized CPUE are also flatter than nominal (Figure 8). K3 has a lower density in the eastern Pacific Ocean from 20°S-20°N, and spotty waters with high CPUE characterize it (Figure 8). The standardized CPUE of K3 showed a big increase in 2019, but we do not have the standard deviation for that year (Figure 8).

Discussion

We have provided this preliminary analysis of difficulties as follows.

- The body mass distribution of bigeye tuna does not show apparent differences by latent classes, but there are distinct spatial patterns. Therefore, we also need to assess CPUE patterns between the latent classes. To compare the standardized CPUE for each latent class, check for an autoregressive pattern that indicates a cohort difference. If there is no time difference, it may be a stock structure difference, but there could still be a correlation even if the groups differ due to the Moran effect (Hansen et al. 2020).
- We must develop the methodology determining the optimal number of clusters in the Finite mixture model analysis. The commonly used BIC decreases as latent classes increase, but smaller BICs models with many latent classes did not accurately classify fish size groups. Hence, alternative methods such as cross-validation should be considered instead of relying solely on BICs.
- One major issue with the multi-species geostatistical model is its long computation time. The TMB program needs to be adjusted for multi-core computing to address this issue.
- The standard deviation of standardized CPUE cannot currently be determined, which may be due to several factors. To address this, we should examine the mesh settings, particularly in areas like southern New Guinea Island, where data is limited but estimated CPUE is high. We should also review the location of nodes used for standardization and check for any outliers in the data. In other words, we should consider filtering the data to avoid the risk of overfitting large catches.

- When conducting a multi-species geostatistical model analysis, it is essential to consider the model selection and diagnostic process. Specifically, utilizing MCMC to calculate Widely applicable information criterion (WAIC) and generating randomized quantile residuals through plotting are crucial steps in this complex study.
- The finite mixture and multi-species geostatistical models could work on one program. “flexmix” can also build such complex models, including linear regression models, but it cannot account for spatiotemporal effects, so there is room for further investigation using TMB.
- We can analyze the latent variables that result from the multi-species geostatistical model. These variables are then compared with environmental factors to determine which factors affect fish distribution (Ijima and Jusup 2023).

References

Hampton, J. and Fournier, D.A., 2001. A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (Thunnus albacares) in the western and central Pacific Ocean. Marine and Freshwater Research, 52(7), pp.937-963.

Vincent, M., Ducharme Barth, N., Hamer, P., Hampton, J., Williams, P., Pilling, G., 2020. Stock assessment of yellowfin tuna in the western and central Pacific Ocean (31July) - Rev.03. SC16-SA-WP-04.

Ducharme Barth, N., Vincent, M., Hampton, J., Hamer, P., Williams, P., Pilling, G., 2020. Stock assessment of bigeye tuna in the western and central Pacific Ocean (30July) - Rev.03. SC16-SA-WP-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.

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.

Leisch, F., 2004. Flexmix: A general framework for finite mixture models and latent glass regression in R.

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.

Hansen, B.B., Grøtan, V., Herfindal, I. and Lee, A.M., 2020. The Moran effect revisited: spatial population synchrony under global warming. Ecography, 43(11), pp.1591-1602.

Ijima, H. and Jusup, M., 2023. Tuna and billfish larval distributions in a warming ocean. arXiv preprint arXiv:2304.09442.

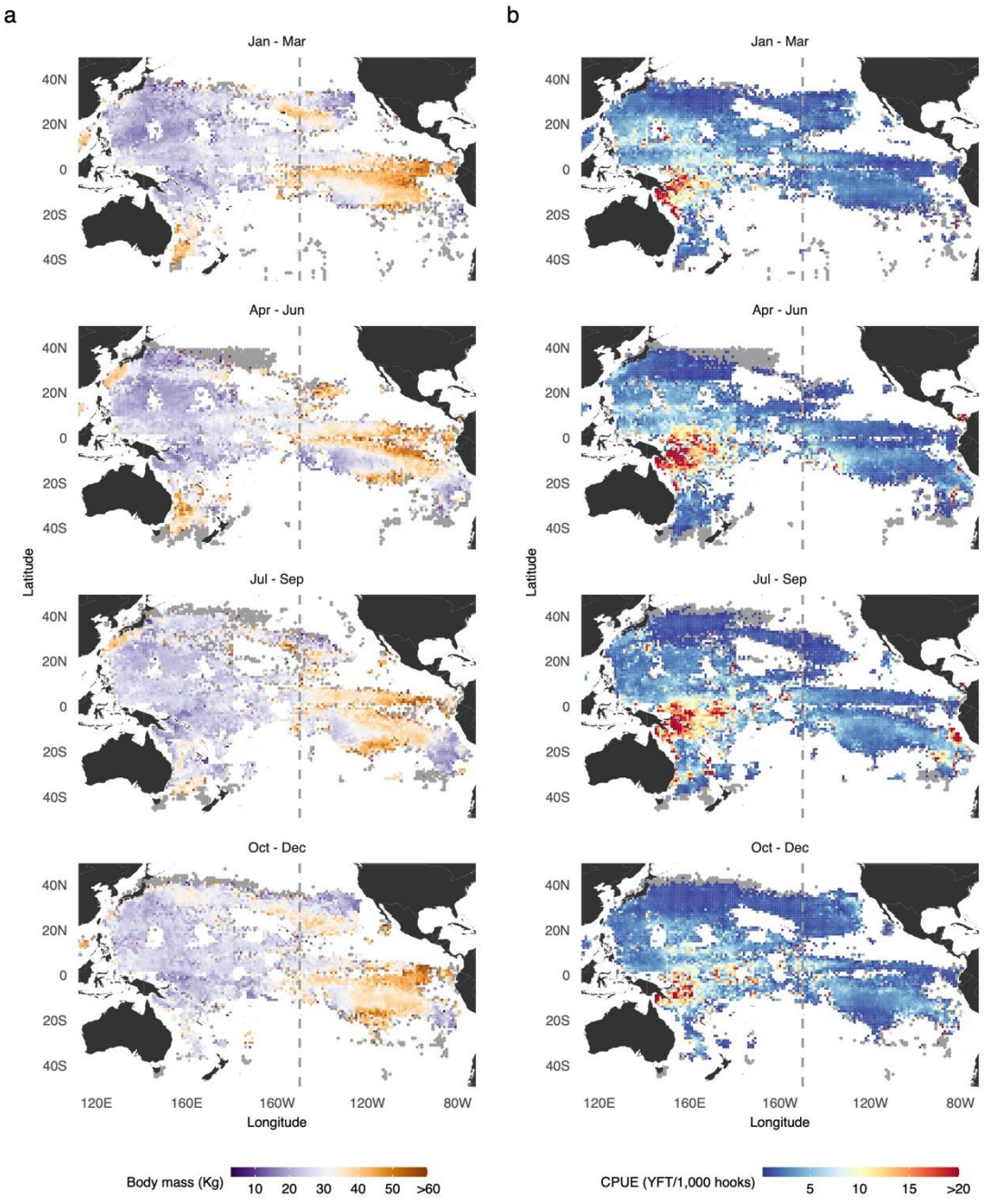

Figure 1. Spatial trends in yellowfin tuna caught by Japanese longline vessels. Dashed lines indicate a 150°E boundary and gray dots indicate no yellowfin catch. **a**: Average body mass (semi-dress weight) of yellowfin tuna caught aggregated by 1°×1° resolution. **b**: Spatial CPUE trends in yellowfin tuna aggregated by 1°×1° resolution.

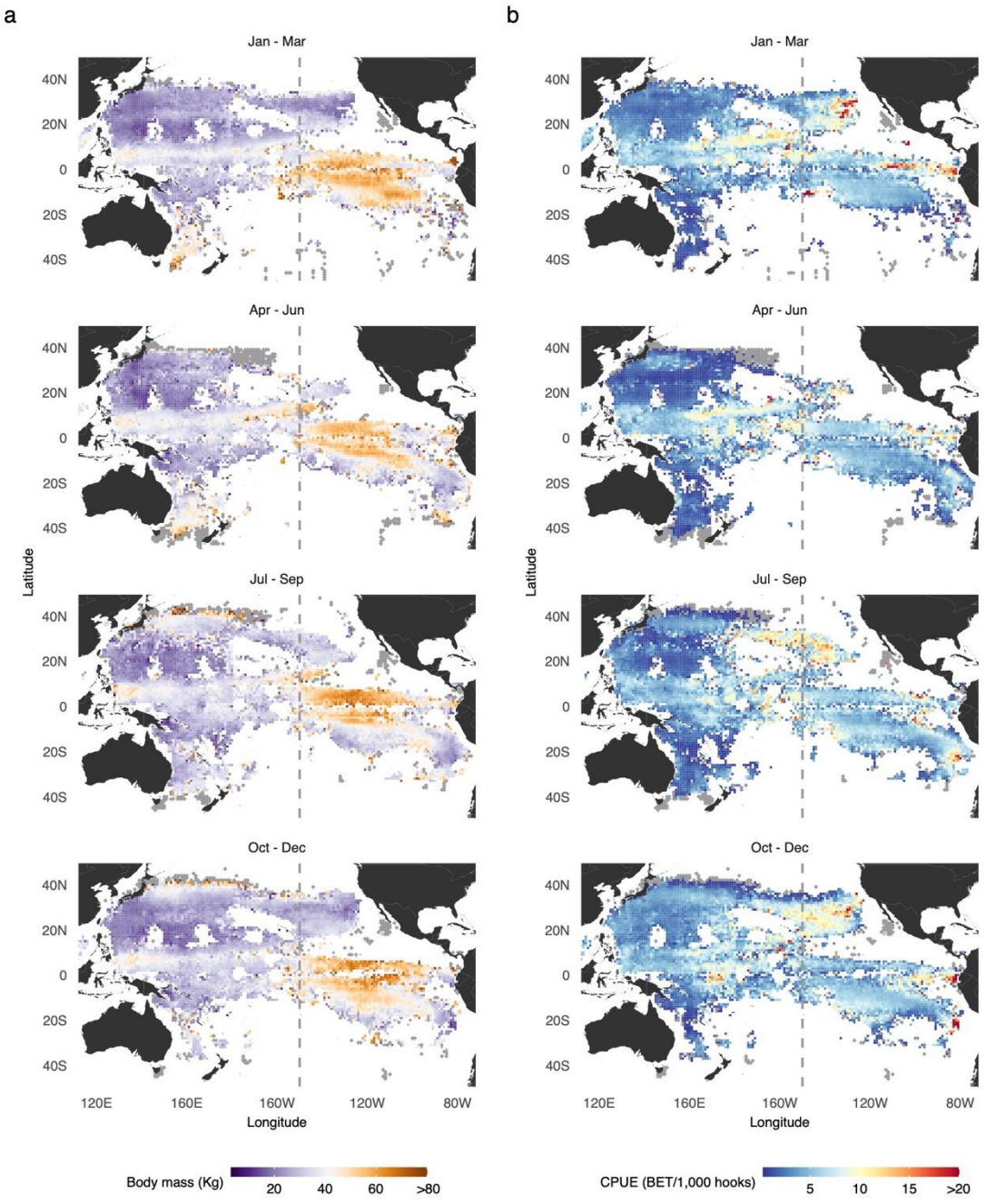

Figure 2. Spatial trends in bigeye tuna caught by Japanese longline vessels. Dashed lines indicate a 150°E boundary, and gray dots indicate no bigeye catch. **a**: Average body mass (semi-dress weight) of bigeye tuna caught aggregated by 1°×1° resolution. **b**: Spatial CPUE trends in bigeye tuna aggregated by 1°×1° resolution.

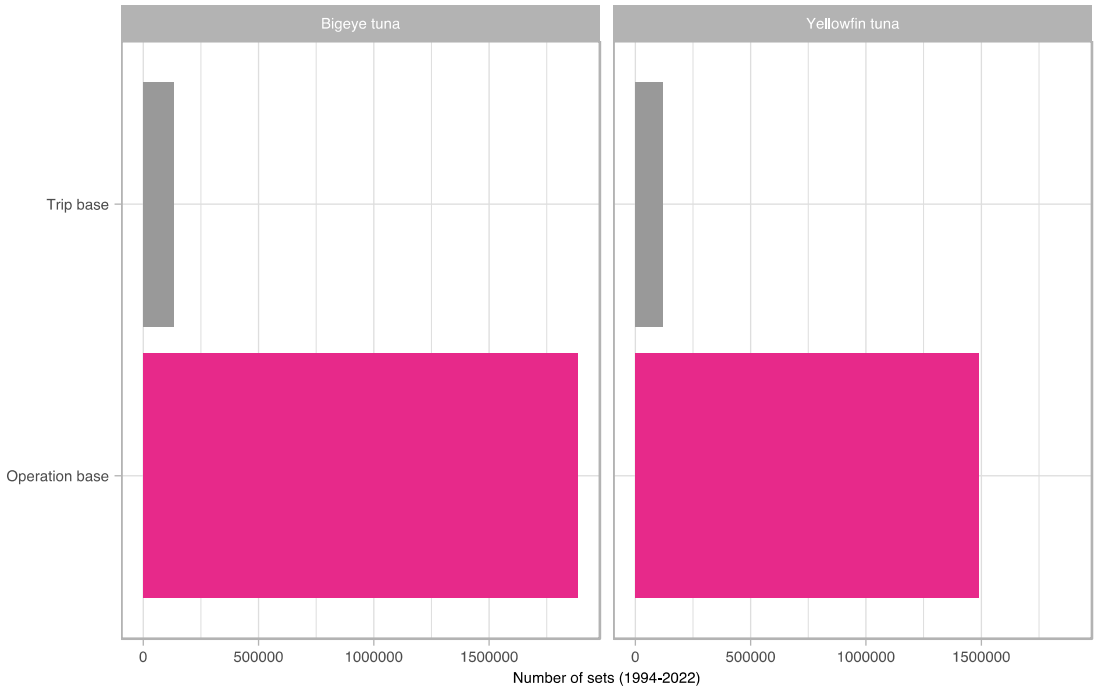

Figure 3. The Japanese longline logbook recorded the average body mass (semi-dress weight). The body mass has been recorded per trip or operation.

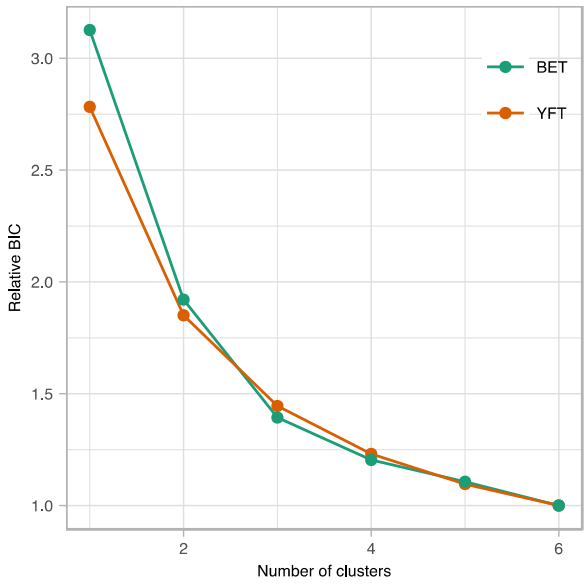

Figure 4. Result of Bayesian information criterion (BIC) from finite mixture model analysis.

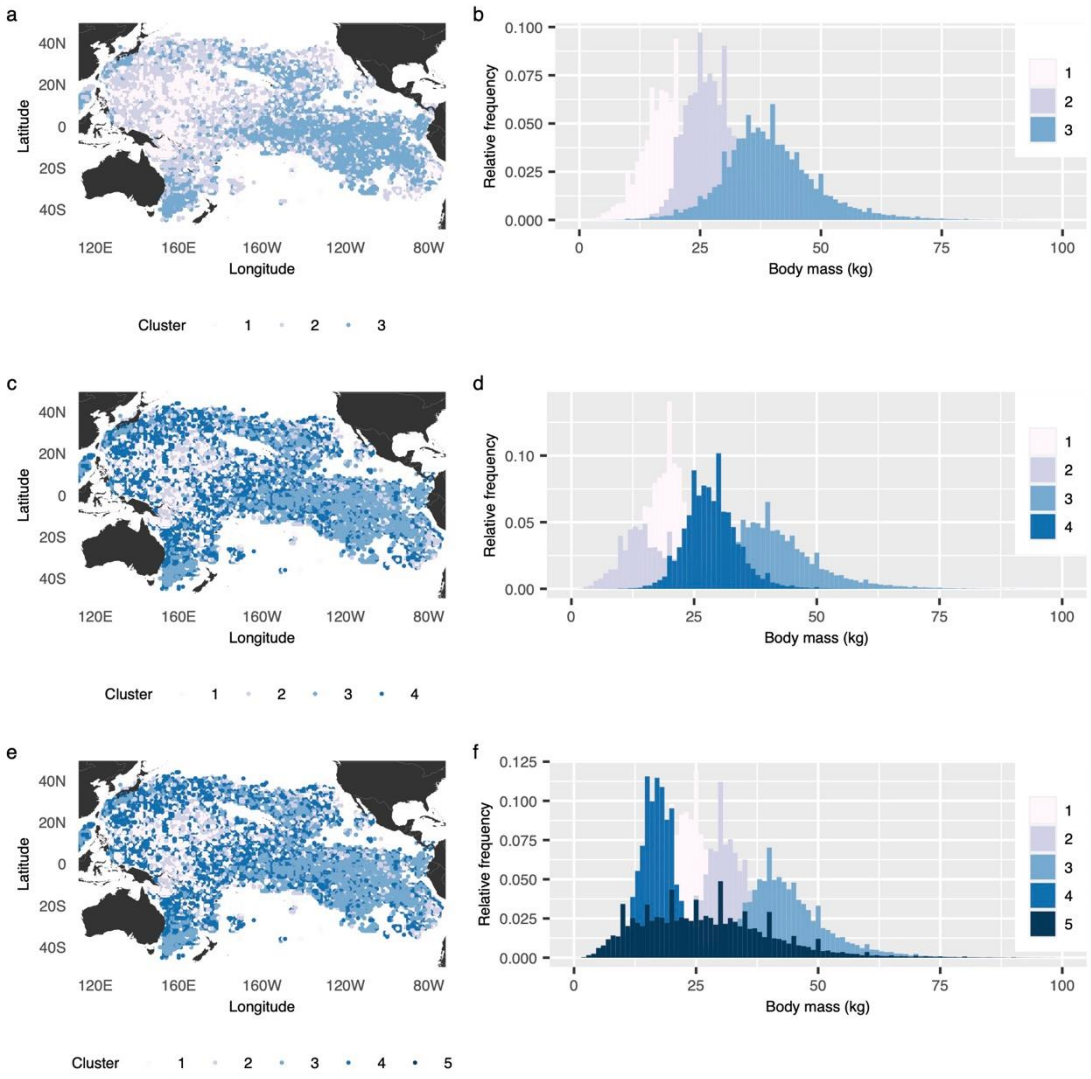

Figure 5. Finite mixture model analysis result. Japanese longline logbook data were classified into latent classes based on the average catch weight of yellowfin tuna (**b, d, f**). All data plots observed location by latent classes (**a, c, e**). Japanese longline logbooks were summed by year, month, and 1°×1° degree grid.

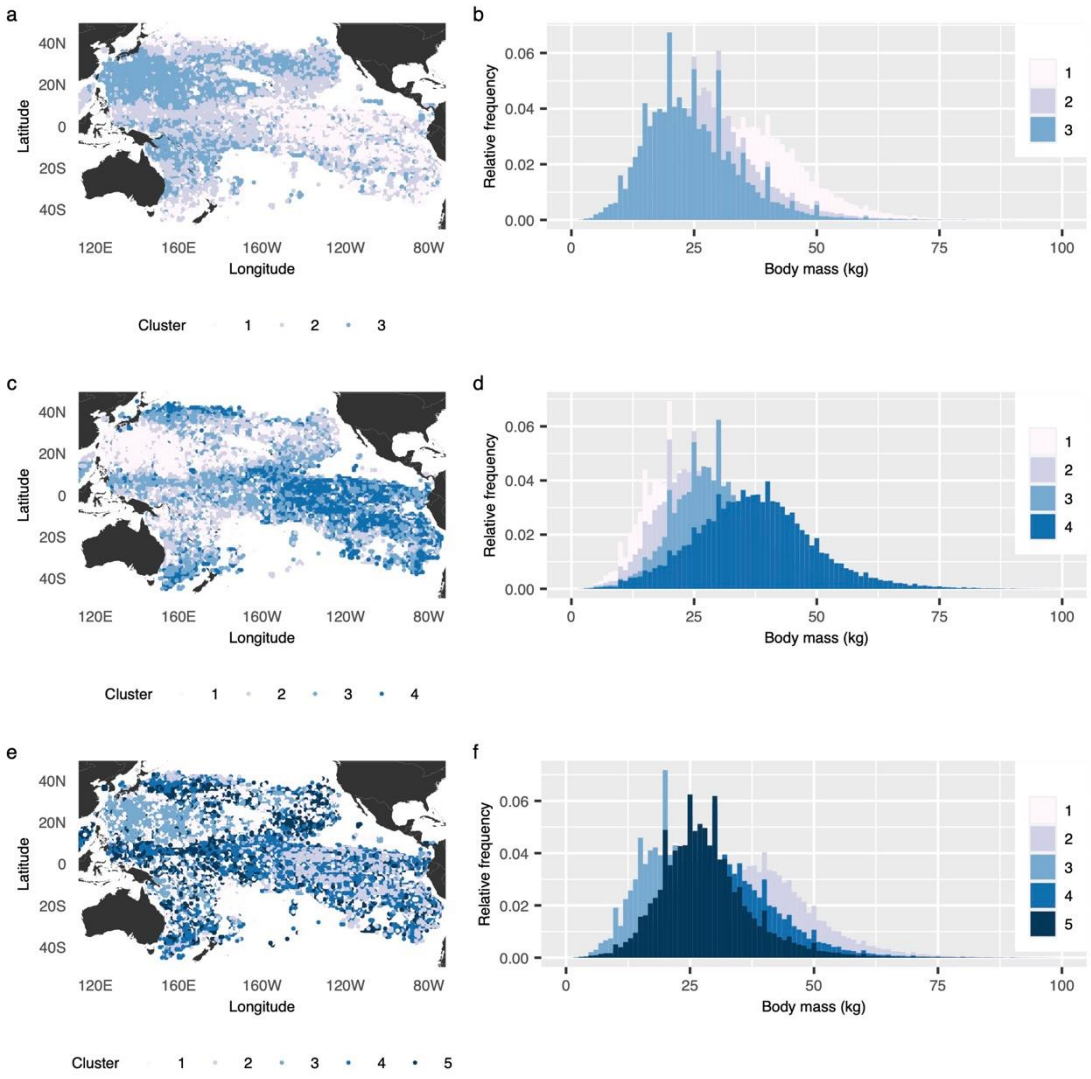

Figure 6. Finite mixture model analysis result. Japanese longline logbook data were classified into latent classes based on the average catch weight of bigeye tuna (**b, d, f**). All data plots observed location by latent classes (**a, c, e**). Japanese logline logbooks were summed by year, month, and 1°×1° degree grid.

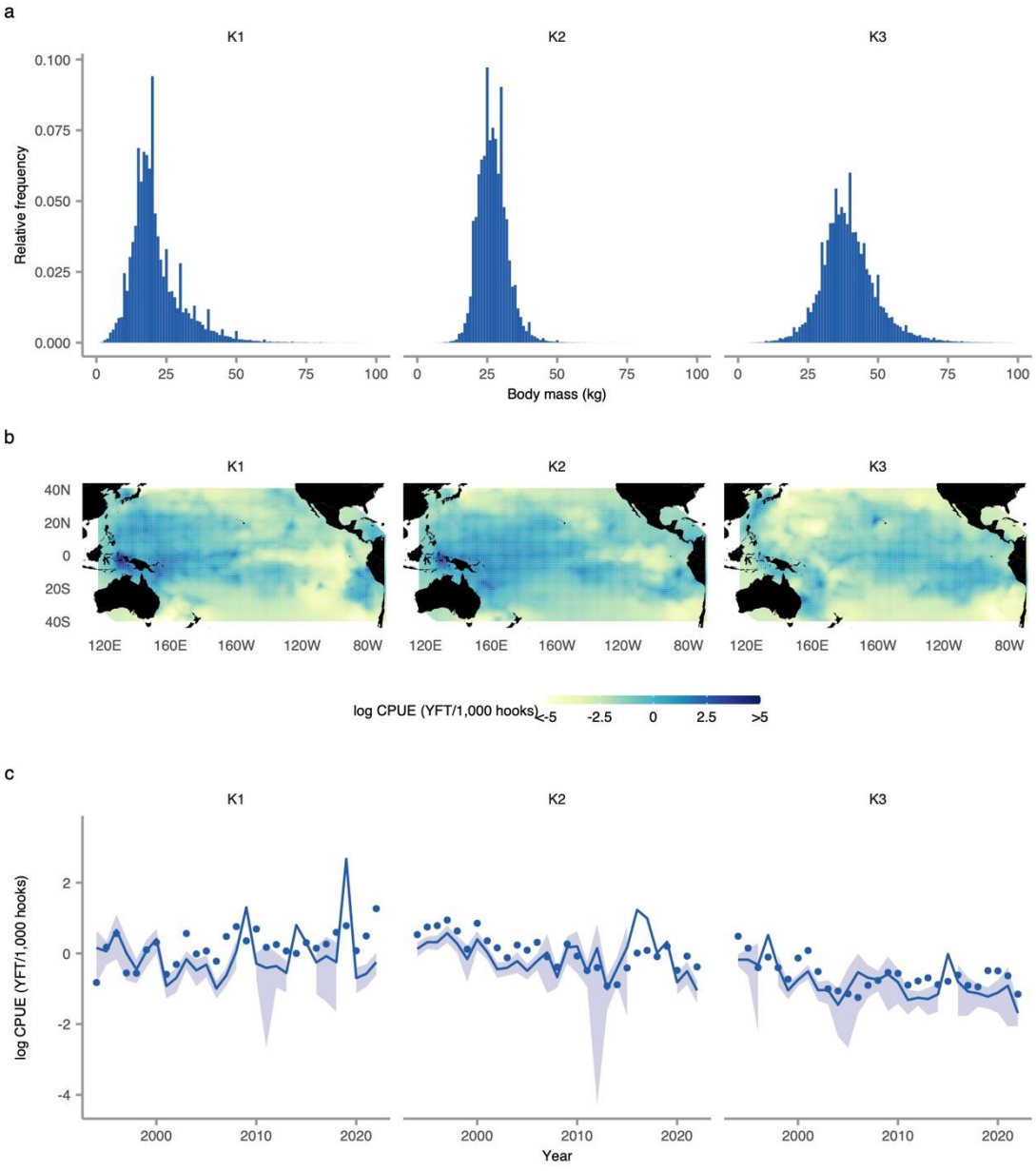

Figure 7. Results of multi-species geostatistical model analysis for yellowfin tuna. Three latent classes (K1, K2, and K3) classified by the finite mixture model were analyzed simultaneously in a multi-species geostatistical model. **a**: Frequency distribution of mean body mass (semi-dress weight) per class. **b**: Standardized spatial density per class. **c**: Trajectory of the yellowfin tuna CPUE. Blue circles denote nominal CPUE, the solid line shows the trends in standardized CPUE, and the filled area means 95% confidence intervals.

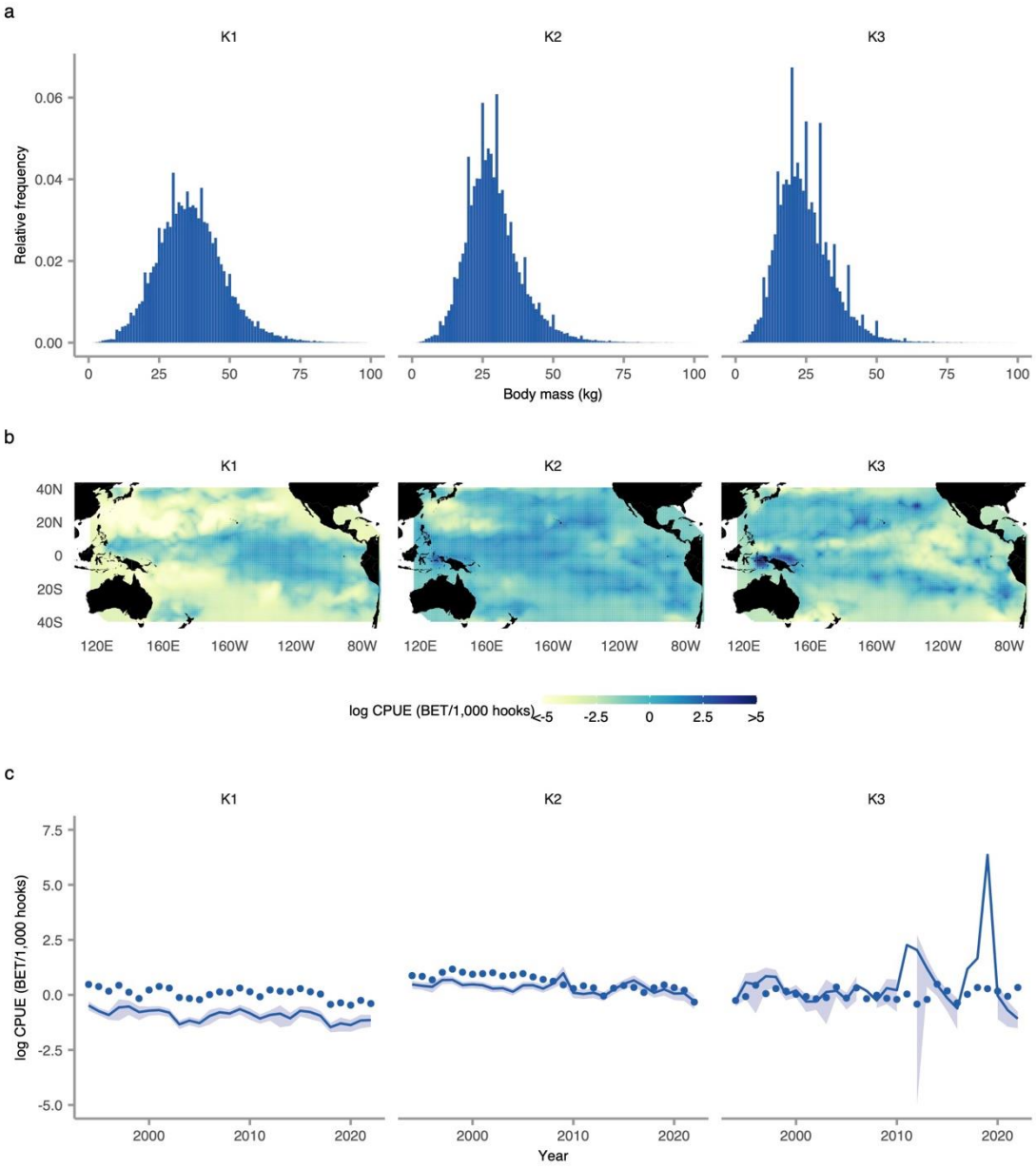

Figure 8. Results of multi-species geostatistical model analysis for bigeye tuna. Three latent classes (K1, K2, and K3) classified by the finite mixture model were analyzed simultaneously in a multi-species geostatistical model. **a**: Frequency distribution of mean body mass (semi-dress weight) per class. **b**: Standardized spatial density per class. **c**: Trajectory of the bigeye tuna CPUE. Blue circles denote nominal CPUE, the solid line shows the trends in standardized CPUE, and the filled area means 95% confidence intervals.