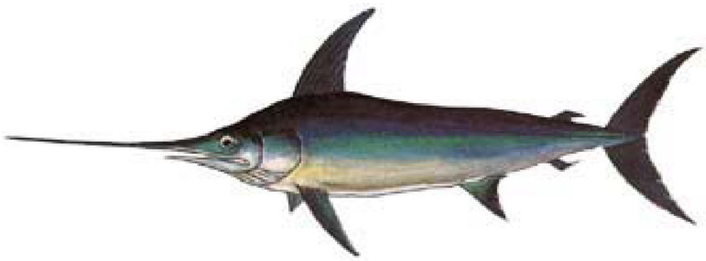
Pattern recognition of population dynamics for North Pacific swordfish (*Xiphias gladius*) : the operational data analysis of Japanese longline fishery using the finite mixture model. [1]

*Hirotaka Ijima and **Minoru Kanaiwa
E-mail:ijima@affrc.go.jp

*National Research Institute of Far Seas Fisheries, Fisheries Research and Education Agency
5-7-1, Orido, Shimizu, Shizuoka, 424-8633, Japan
**Mie University
1577, Kurima-Machiyacho, Tsu, Mie, 514-8507, Japan



[1]This working paper was submitted to the ISC Billfish Working Group Intercessional Workshop, 17-23 January 2018, held in the Pacific Islands Fisheries Science Center of the National Marine Fisheries Service, Hawaii USA. This document not to be cited the author's permission.

Abstract

Size selectivity is the fundamental assumption configuring the fishery definition of integrated stock assessment models such as the stock synthesis 3. The selectivity needs to define by fishing area because usually selectivity depends on the fishing operational ground corresponding to fish migration or distribution. To clarify spatial pattern of the Western and the Central North Pacific ocean (WCNPO) swordfish (*Xiphias gladius*) selectivity that was caught by Japanese longline fishery, we addressed the finite mixture model analysis. In this analysis, we used the Japanese longline operational data rather than the size composition data because there were large gaps between size composition data of commercial vessel and training vessel that have operated different area. Using R software package of the flex mix, we constructed 1 to 6 clusters with two-dimensional linear regression models that response are mean body weight and CPUE. Regarding the covariate of the linear model, we used year, quarter or gear(hooks between floats) effects. We also set grouping factor as $5^\circ \times 5^\circ$ area grid. BIC selected five cluster model which responses classified two type body weight group and five CPUE trends. Comparing with spatial cluster and several results, we suggest dividing WCPO into two fishery areas.

Introduction

The integrated stock assessment model such as Stock Synthesis 3 (SS3) (Methot and Wetzel 2013) have been widely used for Tunas Regional Fisheries Management Organization (Tunas RFMO). SS3 needs to define fisheries called fleets that depends on size selectivity. The size selectivity is estimated by size composition data and calculate catch at length. Thus, determining the size selectivity is one of the most critical configuration for SS3. However, estimated size selectivity is sensitive because of size selectivity changes by time-spatial effects (e.g., fish migration or recruitment pattern). SS3 can handle difficulties for selectivity estimation such as historical change of selectivity, but it is difficult to set the spatial difference. For example, two-mode size selectivity that depends on the juvenile distributed area made destabilize a result of the striped marlin (*Kajikia audax*) stock assessment in the Western and Central North Pacific ocean (WCNPO) (ISC 2015). Thus it needs to define the area dependent selectivity outside the integrated stock assessment model (Waterhouse, Sampson, Maunder, and Semmens 2014).

There is some research defining area based size selectivity. These studies applied the cluster analysis or the generalized additive model (GAM) using length frequency data (Ochi, Ijima, Kinoshita, and Kiyofuji 2016), (Langseth 2016). On the other hand, Catch pre-unit effort (CPUE) calculated by fishing operational data have a spatial information (Ichinokawa and Brodziak 2010). CPUE also includes time-spatial information of the length of catch on the fishing ground. In detail, CPUE changes with the fishing area where the cohort distribution is different and fluctuate annually. Thus, to define accurate fishery, it is better to use size information and CPUE simultaneously in the analysis.

The finite mixture model is the useful tool to address such a difficulties because the finite mixture model divides mixed distribution (Leisch 2004). Furthermore, the finite mixture model is one of the model-based cluster analysis which can use response and explanatory variables and can analyze the size and CPUE information simultaneously (Leisch 2004).

Here, we addressed the finite mixture model analysis to define spatial selectivity pattern of the WCNPO swordfish caught by Japanese longline fishery. Using the Japanese longline operational data, we set two response variables as mean body weight data and CPUE data.

Material and methods

Datasets

The available Japanese longline size composition data is between 1999 and 2016, and that coverage is wide-ranging. Size composition data can combine operational data. Thus it can use for the finite mixture model analysis. On the size composition data, smaller swordfish distribute around the North-central Pacific area (Figure.1 a). However, almost datasets were observed by the training vessel around the North-central Pacific area (Figure.1 b). We thought that it is difficult to clear the reason of size difference between West and Central Pacific ocean. Thus, we alternatively focused on longline operational data because operational data includes catch number and catch weight for all operations (Figure.1 c, d). The Japanese longline operational data is available between 1994 and 2016.

Finite Mixture Model

The Finite Mixture Model with K clusters and D-dimensional response $y = (y_1, ..., y_D)'$ are

$$\begin{aligned} h(y|x,\psi) &= \sum_{k=1}^{K} \pi_k f(y|x,\theta_k) \\ &= \sum_{k=1}^{K} \pi_k \prod_{d=1}^{D} f_d(y|x,\theta_{k,d}), \end{aligned} \tag{1}$$

where y_d are mutually independent the mixture density. x is an independent variables vector. The cluster k with the prior probability π_k is

$$\pi_k \geq 0, \sum_{k=1}^{K} \pi_k = 1. \tag{2}$$

$\theta_{k,d}$ is the clusters with the dimension specific parameter vector of the density function f_d, and ψ is the all parameters vector $\psi = (\pi_{1,1}, ..., \pi_{K,D}, \theta'_{1,1}, ..., \theta'_{K,D})'$.

In this analysis, we used two Generalized linear models (GLMs) as the density function f_d $(d = 1, 2)$ that responses are mean body weight by one operation and CPUE. Firstly, we assumed log normal density GLM (f_1) as follows:

$$\begin{aligned} \theta_{k,1} &= \beta_{k,1}, \sigma^2_{k,1} \\ \log(W_k) &\sim Normal(\mu_{k,1}, \sigma^2_{k,1}) \\ E(\log(W_k)) &= \mu_{k,1}, var(\log(W_k)) = \sigma^2_{k,1} \\ \log(W_k) &= X_k \beta_{k,1}. \end{aligned} \tag{3}$$

Where $\mu_{k,1}$ is the mean of normal distribution, $\sigma^2_{k,1}$ is the variance of normal distribution, W_k is the response vector of the individual mean body weight caught by one operation, X_k is the variable matrix, $\beta'_{k,1}$ and variance $\sigma^2_{k,1}$ is the parameter vector and scalar in cluster k. We assumed the variable as year, quarter and gear (hooks between floats) and treated as the categorical variables. Secondly, we constructed poisson GLM (f_2) for CPUE is

$$\begin{aligned} \theta_{k,2} &= \beta_{k,2}, \beta_{gear} \\ C_k &\sim Poisson(\mu_k) \\ E(C_k) &= var(C_k) = (\mu_k) \\ \log(\mu_k) &= X_k \beta_{k,2} + \beta_{gear} x'_{gear} - \log(1000 hooks). \end{aligned} \tag{4}$$

Where $\mu_{k,1}$ is the mean and variance of the poisson distribution, C_k is the response vector of CPUE, $\beta_{k,2}$ and β_{gear} are the parameter vector in the cluster k and $\log(1000hooks)$ are offset variable. We assumed variables year, quarter and gear but gear effects (β_{gear}) were not change by cluster. Area variable (5 ° × 5 ° grid data) was set as grouping factor because, our object is to define area dependent fishery definition for the SS3.

All parameters were estimated by R software package "flexmix" ver2.3-14. To chose the appropriate number of area cluster, we set 1-6 clusters for initial values on the flexmix. We use Bayesian information criterion (BIC) for the model selection. To define Japanese longline fishery, we plotted estimated clusters spatially and compared with variables that were used the Finite Mixture Model analysis.

Result and discussion

We made different initial cluster models using the flexmix. As a result, the five cluster model showed the lowest BIC (Figure.2). The Central Pacific area was always chosen as one cluster (Figure.3). In the Central Pacific area, the mean body weight of swordfish was smaller than another area (Figure.4). The summarized CPUE by each cluster showed a different trend, especially Central Pacific area (cluster 3 by 5 clusters) showed the lowest value and flat patterns (Figure.5). As above stated, Central Pacific area showed different trends both CPUE and mean body weight.

To classify the fishery definition for the SS3, we summarized the result of the cluster, total catch number, CPUE and mean body weight spatially (Figure.6). All information indicated that Central Pacific area has a different pattern (Figure.6). Hence, the WCNPO could divide Central Pacific area (Area 2) and other areas (Area 1)(Figure.6).

Following to the classified area, we summarized the size composition data by each season. The size composition of Area 2 is approximately same as Area 1 on quarter one and two, in contrast on quarter three and four, the size of swordfish of Area 2 is larger than Area 1 (Figure.7 b)). However, this is opposing to the operational data (Figure.7 a)). Furthermore, the coverage of size composition data observed by commercial vessel in Area 2 is quite low (Figure.8). We considered that this dataset might not represent Japanese longliner's selectivity. Thus, it need to discuss that to use size composition data of Area 2 for SS3 or to mirror the size selectivity using another countries data such as US Hawaii data that coverage larger than. The size composition data of EPO is also a little (Figure.8).

Summary and recommendations

Considering results of this analysis, we suggest the configuration of SS3 as follows:

- The size composition data which observed by the training vessel can't use for SS3 because training vessel might not represent commercial longliners.
- The result of the Finite Mixture Model analysis indicated that there are two different area in the WCNPO.
- The size composition data of Area2 that exclude training vessel data is inconsistent with mean body weight that comes from longline operational data. Thus, we recommend mirroring to Hawaii or Taiwanese longline selectivity.
- The number of size composition data observed in the EPO is a little. This data sets also do not represent Japanese longliner's selectivity after 1999.

References

Ichinokawa, M. and J. Brodziak (2010). Using adaptive area stratification to standardize catch rates with application to north pacific swordfish (xiphias gladius). *Fisheries Research 106*(3), 249–260.

ISC (2015). Stock assessment update for striped marlin (kajika audax) in the western and central north pacific ocean through 2013.

Langseth, B. (2016). Spatial and temporal patterns in striped marlin (kajikia audax) length in the hawaiian deep-set longline fishery. *ISC/16/BILLWG-01/08*.

Leisch, F. (2004). Flexmix: A general framework for finite mixture models and latent glass regression in r. *Journal of Statistical Software 11*.

Methot, R. D. and C. R. Wetzel (2013). Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research 142*, 86–99.

Ochi, D., H. Ijima, J. Kinoshita, and H. Kiyofuji (2016). New fisheries definition from japanese longline north pacific albacore size data. *ISC/16/ALBWG-02/06*.

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

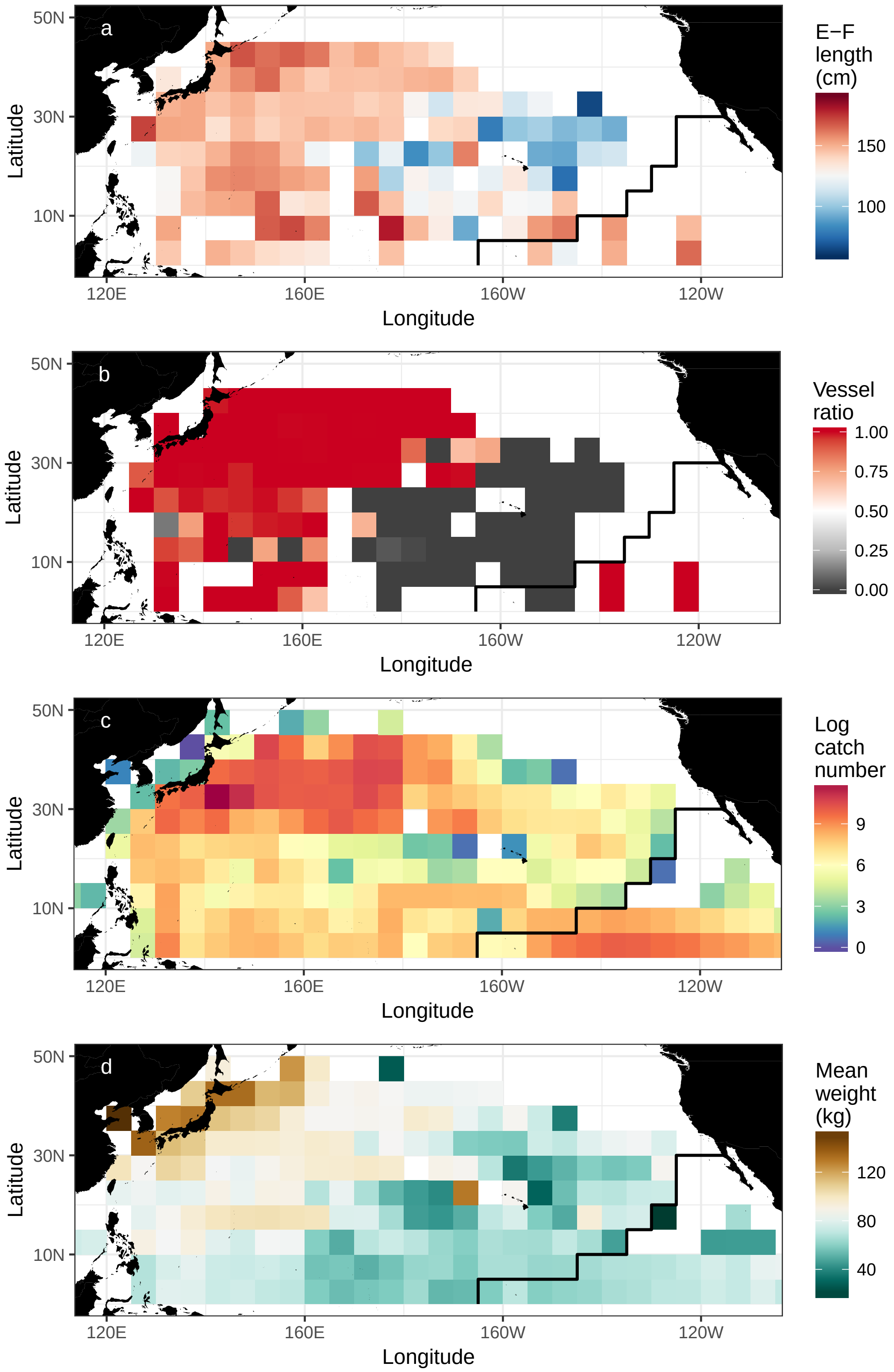


Figure 1: The spatial trend of Japanese longline fishery data (1999-2016). a) size composition data. b) the ratio of commercial vessel data in size composition data. c) aggregated number of swordfish catches calculated by Japanese catch statistics. d) mean body weight calculated by Japanese catch statistics.

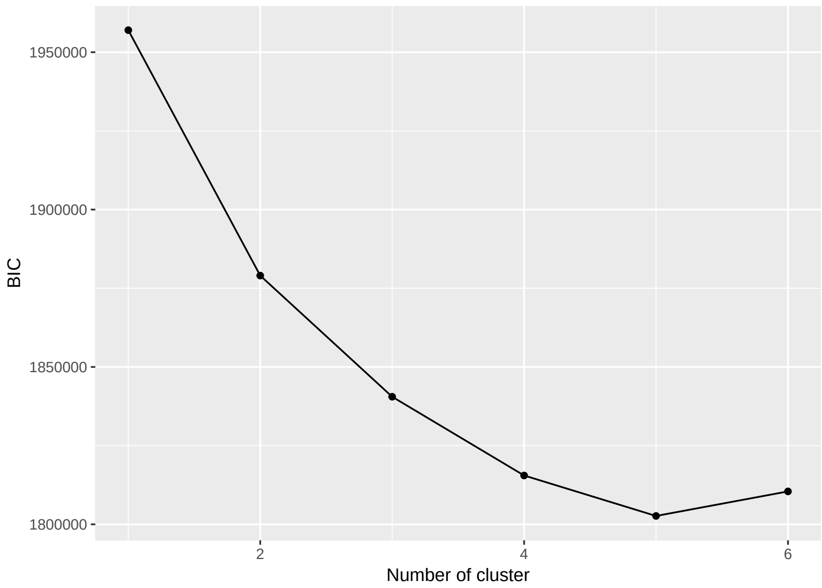


Figure 2: Bayesian information criterion (BIC) of different initial clusters. In this analysis, we set the reasonable regression models, and chose appropriate number of area cluster using BIC.

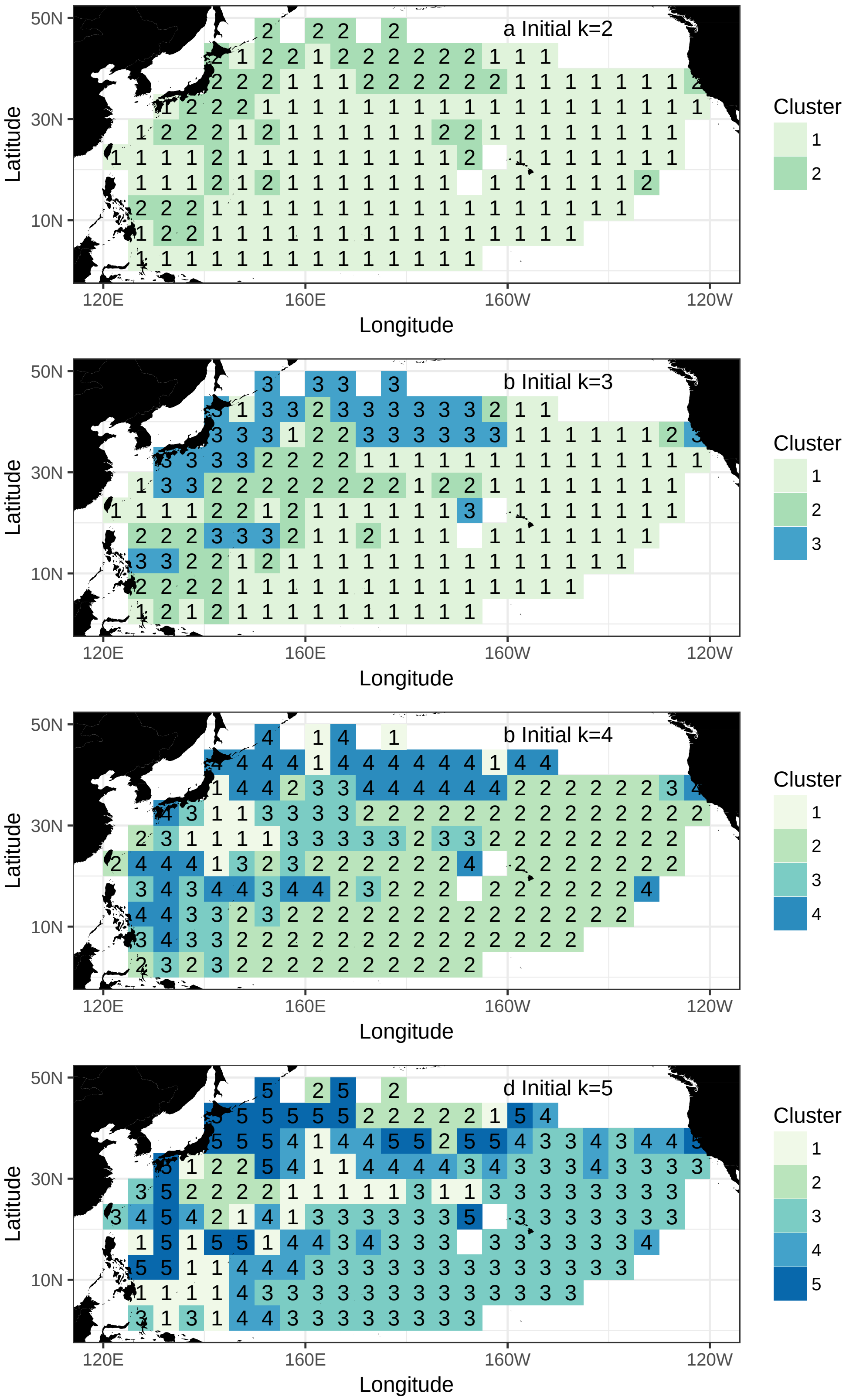


Figure 3: The area clusters defined by the finite mixture model (FlexMix). FlexMix needs to set different initial clusters because FlexMix is unsupervised learning.

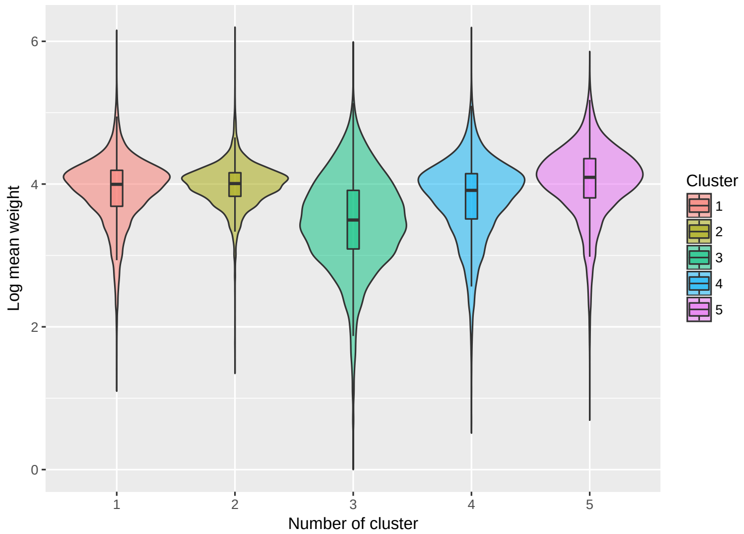


Figure 4: The distribution of mean body weight of one time fishery operation summarized by five clusters.

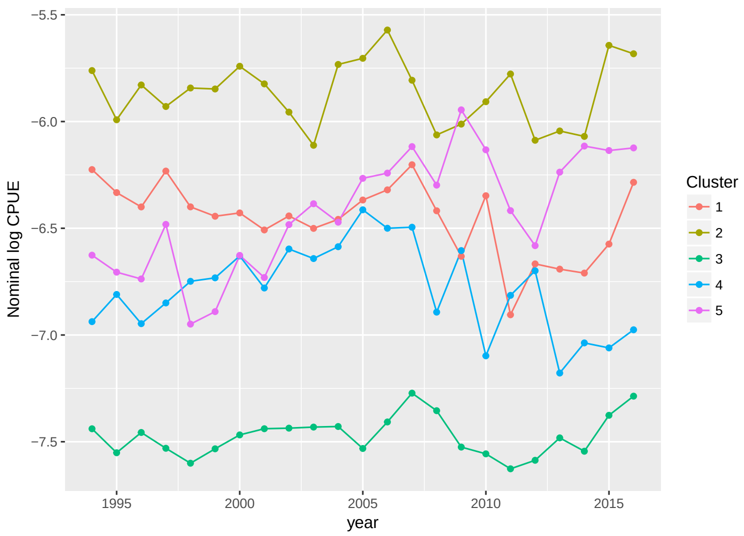


Figure 5: The nominal log CPUE summarized by five clusters. CPUE was calculated by positive catch.

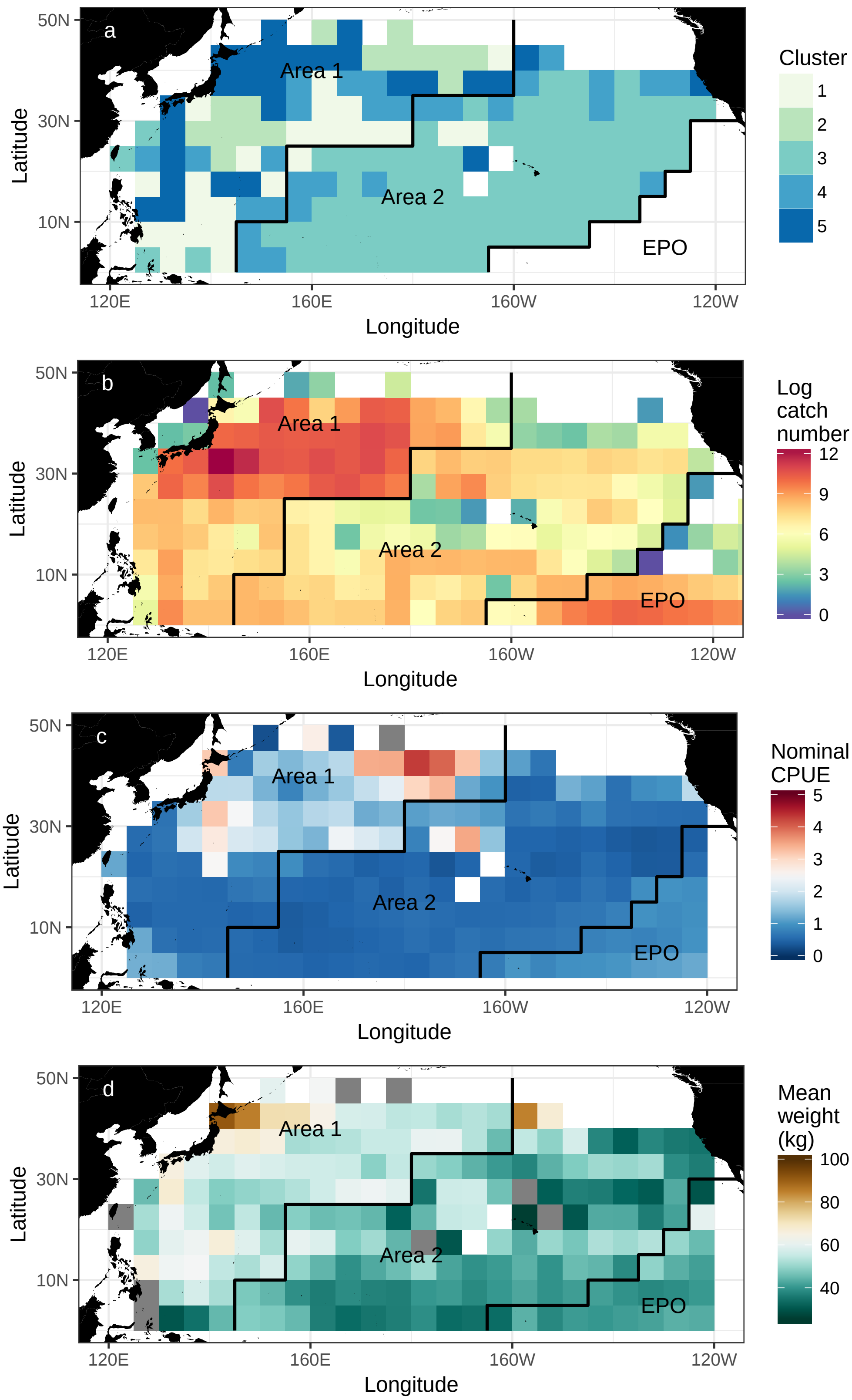


Figure 6: The candidate area definition and comparison of defined area by different information from Japanese longline operational data (1994-2016). a) Selected cluster, b) number of log catch, c) nominal CPUE, and d) mean body weight.

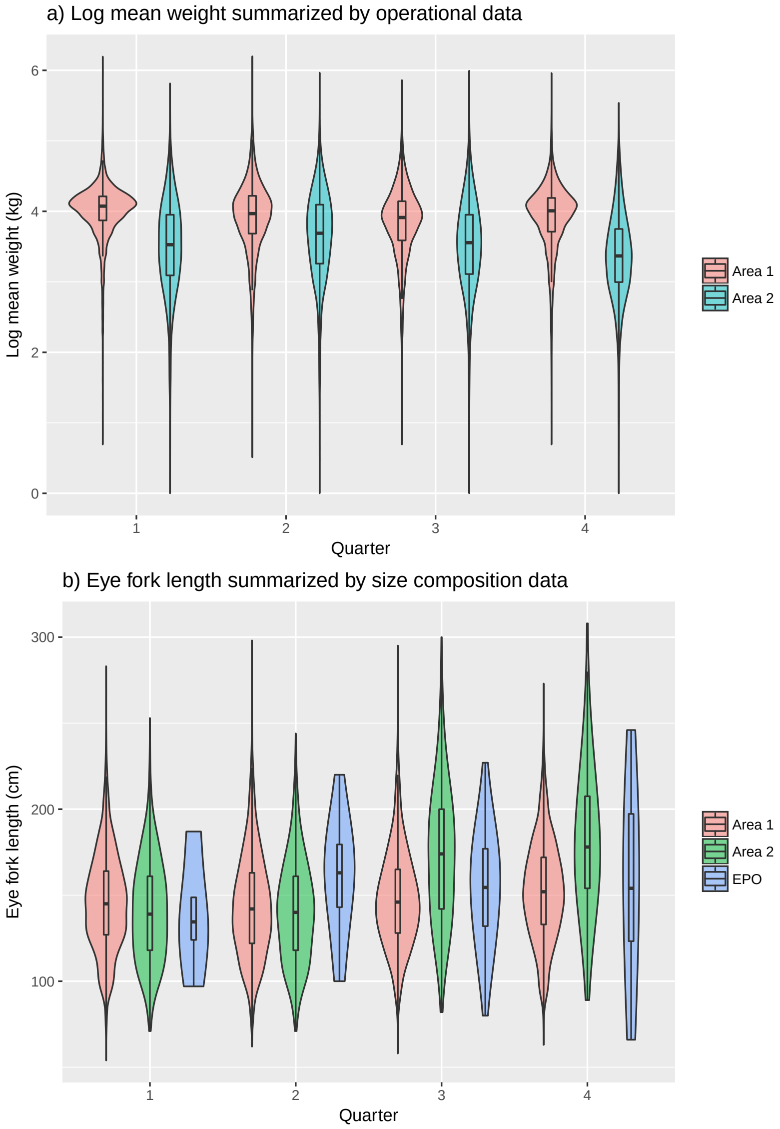


Figure 7: The comparison of the operational data and size composition data. a) log mean weight summarized by operational data. b) eye fork length summarized by size composition data. The data sets of EPO were removed in the flexmix analysis.

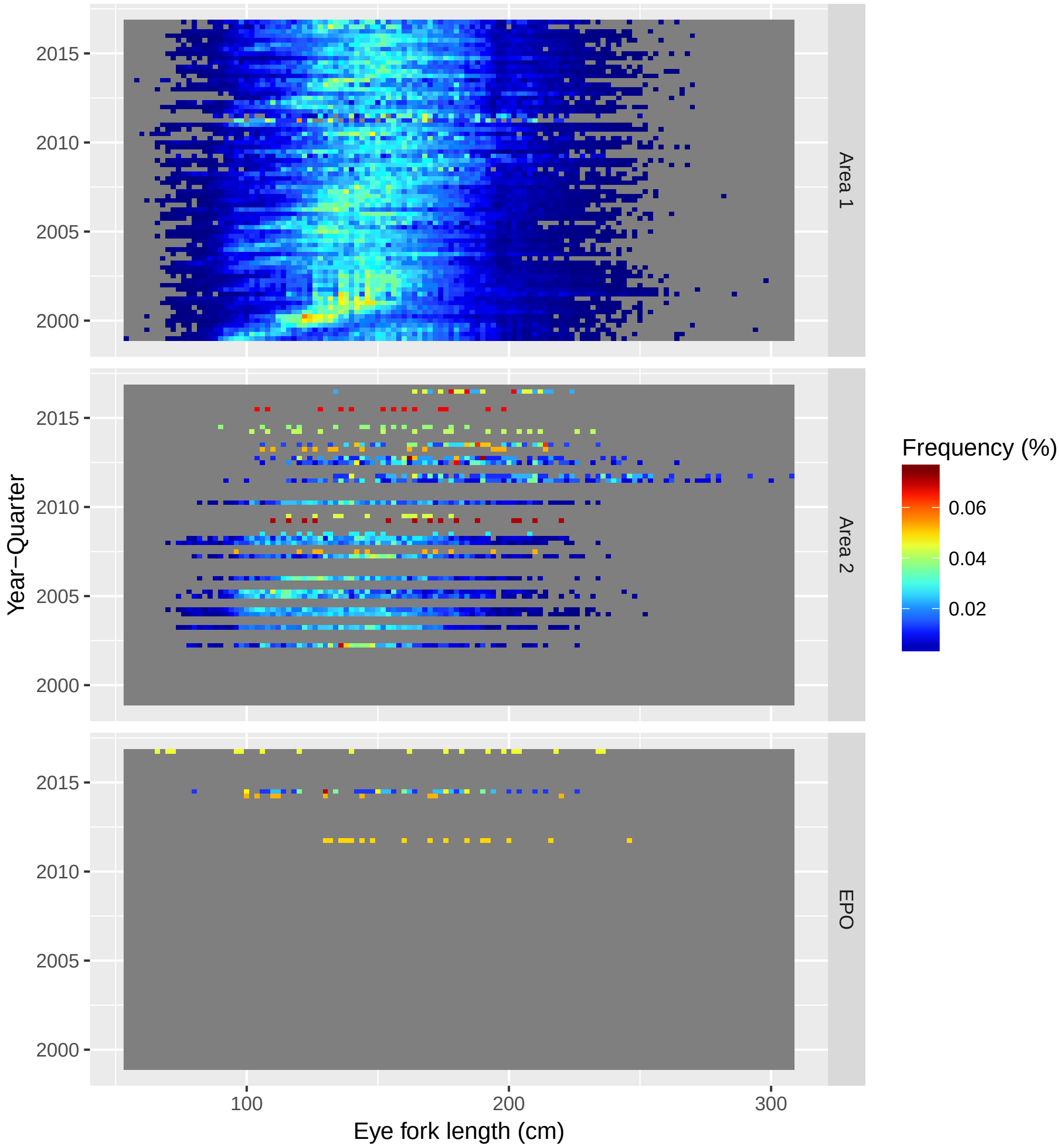


Figure 8: Historical change of size composition data which was aggregated by year, quarter and area. The training vessel data were excluded.