SCIENTIFIC COMMITTEE
FOURTEENTH REGULAR SESSION

Busan, Republic of Korea
8-16 August 2018

Improvements in skipjack (Katsuwonus pelamis) abundance index based on the fish size data from Japanese pole-and-line logbook (1972–2017)

WCPFC-SC14-2018/ SA-WP-04
Rev 1 (27 July 2018)

Junji Kinoshita[1], Yoshinori Aoki[1], Hirotaka Ijima[1], and Hidetada Kiyofuji[1]

[1] National Research Institute of Far Seas Fisheries, Japan Fisheries Research and Education Agency

Improvements in skipjack (*Katsuwonus pelamis*) abundance index based on the fish size data from Japanese pole-and-line logbook (1972–2017)

Junji Kinoshita, Yoshinori Aoki, Hirotaka Ijima, and Hidetada Kiyofuji

National Research Institute of Far Seas Fisheries, Japan Fisheries Research and Education Agency, Shimizu, Shizuoka, Japan

Abstract

In this document, we report an analysis result of fisheries area definition of skipjack tuna in the western and central Pacific Ocean. To clarify area definition for CPUE standardization, we addressed model-based cluster analysis using average fish weight and CPUE recorded on Japanese pole-and-line (JPNPL) logbook. The logbook data includes the daily SKJ catch per vessel by 1°×1° resolution from 1972 to 2017. Catch and mean body weight larger than 0 were aggregated into 5°×5° resolution and then extracted grids with more than 10 vessels-day operations. The model was constructed with the average body weight and nominal log CPUE as response variables and with year, quarter, and gross register tonnage of fishing vessel as explanatory variables. As a result of the model, five clusters were selected by the minimum BIC and showed unique characteristics to the size composition and differences of operation mode on each cluster. One of the major achievement of our study is the result of area clustering analysis matches better the area definition suggested by Kiyofuji and Ochi (2016) based on tagging and larvae survey than that of reference case. This result suggests that biological information such as body weight will provide useful evidence for determining the stock assessment fisheries definition. Therefore, we recommend SC14 to consider a new area definition which we propose in this document as the reference case of spatial stratification in the next skipjack stock assessment.

Introduction

Application of fishery information based on underlying biology to the stock assessment model is a basic concept to achieve a better assessment. Skipjack, in particular, is a free-ranging species distributing from tropical to temperate areas through their life cycle (Matsumoto et al., 1984). Their highly migratory ecology makes the model assumption uncertain especially for their migration range, thereby the area definition for assessment of their habitat.

The biologically-driven area definition can give us a better explanation to the process of settings in the model. Given that the skipjack migration is linked to water temperature (Lehodey et al., 1997), it is more reasonable to change the area definition by year rather than applying the same definition with assumption of constant migration. An alternative area definition based on the tagging and larvae surveys' data has been proposed (**Figure 1**, Kiyofuji and Ochi, 2016), which has an explanatory power in terms of individual behavior but still lacks the evidence in terms of school behavior when it comes to the representation of the stocks.

Logbook in Japanese pole and line fishery (JPNPL) has a reliable and comprehensive dataset (i.e., the amount of catch, operational area, average weight of the skipjack, etc.) in the north western Pacific Ocean due to imposition of a duty report submission since 1972. Historical data can capture decadal changes in fishery areas from the past to the current. Furthermore, different types of data can be a good indicator to provide biological evidence of the area definition by applying them to classification analysis.

In this document, we classified the skipjack habitats based on body weight and catch per unit effort (CPUE) obtained from JPNPL logbook from 1972 to 2017 and discuss the validity of the application of the two distinctive area definitions in comparison of our results.

Materials and methods

Logbook in Japanese pole and line fishery (JPNPL)

JPNPL logbook data (1972–2017) for skipjack (SKJ) include the daily catch, vessel ID, operational area latitude and longitude with 1°×1° resolution, and mean body weight (BW) with accuracy of 0.1 kg. Records containing zero catch or unknown BW (expressed as 0 kg) were removed. We calculated weighted mean body weight (weighted BW) in each grid. BW of a grid was weighted according to the amount of catch in each grid as explained in an equation: weighted BW = $\sum_{i=1}^{99}$ BW_i×Catch at BW_i/Total catch, where i is a category of fish weight with 0.1 kg increment until the maximum of 9.9 kg. Finally, we aggregated the data into 5°×5° resolution and thereafter extracted more than 10 vessels-day operations data to classify the data by a finite mixture model.

Finite mixture model

Finite mixture models include a model-based cluster analysis that are very useful tools when applied to data where observations originate from various groups and the group affiliations are not known (Leash 2004). Thus, we used this method for area clustering of JPNPL operational data.

The likelihood of the finite mixture model with K clusters and with D-dimensional response$\boldsymbol{y} = (y_1, \dots, y_D)'$ are

$$y(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\psi}) = \sum_{k=1}^{K} \pi_k \prod_{d=1}^{D} f_d(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}_{k,d}),$$

where $y()$ is the density of mixture distribution and $f_d()$ are the density function for each component, and $\boldsymbol{x}$ is an independent variables vector (e.g. year, quarter, and gear). The cluster k with the prior probability π_k is

$$\pi_k > 0, \sum_{k=1}^{K} \pi_k = 1.$$

$\boldsymbol{\theta}_{k,d}$ is the cluster specific parameter vectors for the density function $f_d()$, and $\boldsymbol{\psi}$ is the all parameter vector $\boldsymbol{\psi} = (\pi_{1,1}, \dots, \pi_{K,D}, \theta_{1,1}, \dots, \theta_{K,D})'$.

In this analysis, we used two generalized linear models (GLMs) as the density function $f_d(d = 1,2)$ that responses are weighted BW by one operation on 5°×5° grid area and nominal CPUE, respectively. Firstly, we assumed gamma GLM (f_1) with inverse link function as follows:

$W_k \sim Gamma(s_k, r_k)$
$E(W_k) = \frac{s_k}{r_k} = \mu_{k,1}, var(W_k) = \frac{s_k}{r_k^2}$ and
$\mu_k^{-1} = \boldsymbol{\alpha}_k^{'} \boldsymbol{X}_k,$

where r_k is the scale parameter of gamma distribution, s_k is the shape parameter of gamma distribution, W_k is the response vector of the individual weighted BW by one operation on 5°×5° grid area, $\boldsymbol{\alpha}_k^{'}$ is the regression coefficient vectors corresponding to variables matrix $\boldsymbol{X}_k$, and scalar in cluster k. We assumed the variable as year, quarter and gross register tonnage. All variables are treated as the categorical variables. Secondly, we constructed log normal GLM (f_2) for CPUE is

$log(C_k) \sim Normal(\mu_k, \sigma_k^2)$
$E(log(C_k)) = \mu_{k,2}, var(log(C_k)) = \sigma_k^2$ and
$log(\mu_{k,2}) = \boldsymbol{\beta}_k^{'} \boldsymbol{X}_k + log(effort_k),$

where $\mu_{k,2}$ is the mean of normal distribution, σ_k^2 is the variance of normal distribution, C_k is the response vector of SKJ catch, $\boldsymbol{\beta}_k^{'}$ are the regression coefficient vectors corresponding to variable matrix $\boldsymbol{X}_k$. We assumed variables year, quarter and vessel size but vessel size effects were not change by cluster. Area variable (5°×5° grid area) was set as grouping factor because our goal is to define area dependent fishery definition for stock assessment.

All parameters were estimated by R software package “flexmix” ver 2.3-14. To choose the appropriate number of area cluster, we set one to eight clusters for initial values on the flexmix. We use Bayesian information criterion (BIC) for the model selection. To define JPNPL fishery, we plotted estimated clusters spatially and compared with variables that were used the Finite Mixture Model analysis.

Result and Discussion

Basic descriptions of skipjack catch by Japanese pole and line fisheries (JPNPL)

JPNPL operational area ranges from 120°E to 160°W in longitude and from 20°S to 45°N in latitude (**Figure 2**). High catch was found in the northeast offshore of Japan during quarters 2 and 3, and the areas between north tropical and subtropical areas through the quarters. The high catch was due to high efforts, so that the nominal CPUE was not so remarkable in the same area (**Figure 3**). To a reference, the times series of SKJ catch by region is shown in **Supplement figure 1**.

Distinct size difference was well observed around 28°N in the north western Pacific Ocean (**Figure 4**): large SKJ (4–8 kg) distributed in offshore areas from 5°N to 28°N, whereas the small SKJ (1–2 kg) distributed in northern offshore areas (>28°N) as well as coastal areas around the Nansei Islands and the Philippines.

The composition of weighted BW displayed two peaks at 2.5 kg and 4 kg with longtails extending to 9.9 kg (**Figure 5A, B**).

Classification of skipjack habitats based on the JPNPL fisheries

For intuitive understanding, the area distributions were visually confirmed for the serial of the clusters (*k*=1…8) (**Figure 6**), which helps to determine the potential dominancy of the classified area. Two major clusters were distinguished around boundary of 25–30°N (*k*=2 in **Figure 6**). This result would be derived from the size difference appeared around the 28°N as mentioned earlier (**Figure 4**). The consistency between the basic description of skipjack fishery and classified areas strengthened the validity of the classification. Highly dominant area in the classification found in the north and south of the boundary would suggest that the area of northeast offshore of Japan is utilized by juvenile SKJ during their seasonal northward migration.

Aside from the visual confirmations in eight types of classifications, the number of clusters was determined to have the minimum BIC by checking the relationship between BIC and cluster numbers from one to eight (**Figure 7**). As a result, the number of clusters was determined to be five in this study.

Characteristics of locations, body weight, and catch of five clusters

All characteristics related to area distributions and weighted BW in each cluster were described in **Figures 8** and **9** and summarized in **Table 1** and **Supplement figure 2**.

Cluster 1 (red tiles in the **Figure 8**) distributed in areas from 15°N to 25°N and was roughly located on the boundaries between clusters 4 and 5. Weighted BW in Cluster 1 was 1.9 kg with large deviation in total (**Figure 9**), and it increased from the 1980s to the 2000s. Cluster 2 (orange tiles) distributed in the southern hemisphere and a part of the north tropical areas through the quarters until the 1980s, but the area of Cluster 2 shrank from the 1990s to the 2010s (**Figure 8**). Weighted BW in Cluster 2 showed an almost constant value of 3.9 kg through the decades (**Figure 9**). Cluster 3 (yellow tiles) distributed around the west side of Japan as far as off Nansei Islands (**Figure 8**). Weighted BW (1.8 kg on average) in Cluster 3 was the lowest among clusters and it slightly increased from the 1980s to the 1990s (**Figure 9**). Cluster 4 (green tiles) distributed in the north subtropical areas through the quarters (**Figure 8**). Although the constant weighted BW at 4.0 kg was found until the 1990s, the weighted BW increased to 5 kg since the 1990s (**Figure 9**). Cluster 5 (blue tiles) distributed in the north temperate areas over 30°N mainly in quarters 2 and 3 (**Figure 8**). Weighted BW in Cluster 5 was 2.2 kg on average and it gradually increased.

Decadal trends of catch, effort, and nominal CPUE was explored from 1972 to the current. Drastic decrease in catch was found between the 1980s and the 1990s in all clusters except for Cluster 5 (**Figure 10**), which would reflect especially in regions 2 and 3 of reference case area definition, a change in fishing fleet in the classified area by entering of the extent number of purse seiners in the 1980s (McKechnie *et al*., 2016). Efforts at Clusters 1, 2, and 4 steeply decreased in the 1980s and have stayed low since then, on the other hand, those at Cluster 3 and 5 showed gradual decrease continuing from 1980s to the present.

Implications for the area definition in the stock assessment

Our area classification based on the biological information such as weighted BW and CPUE provide valid evidence for defining the fisheries area in the stock assessment. One concrete way to see whether the existing area definition is valid or not is to compare our results with the current and newly proposed area definitions. Here, a comparison was made possible by checking the catch amount occupied by the clusters in each region (**Figure 11**). In this comparison, the less the consisting number of clusters in each region, the more reasonable of the area definition in terms of biological explanation. One of the major achievement of our study was the validation of the alternative regions of 1, 6 and 7 to be reasonable, which corresponds to Region 1 of the reference case. Region 1 of the reference case mainly included Clusters 3 and 5 (**Figure 11A**). On the other hand, majority of Clusters 3 was removed in the alternative region 1 (**Figure 11B**). Removed cluster was found in the alternative region 7. This suggest that the subdivision of the reference case Region 1 into the alternative regions of 1, 6, and 7 would be successful in the context of the biological evidence for the area definition. Subsequently, Region 2 of the reference case mainly included clusters 2 and 4. This region was subdivided into the alternative regions of 2 and 6 (**Figure 1**). Cluster 4 in the alternative region 2 slightly decreased compared to Region 2 in the reference case because some of them were classified as the alternative region 6. Region 3 was not changed largely between the reference and alternative definitions. Region 4 of the reference case mainly included Cluster 4 but the alternative region 2 included Cluster 2 instead, although the catches in these regions were relatively low. Region 5 is the same definition between the reference and alternative definitions. Although the more quantitative comparison would be required, the first implication here is that our approach can be a good tool to examine the validation of area definition. In regard to Region 1 in the reference case, we emphasize that subdividing the region into the alternative regions 1, 6, and 7 is supported by the biological evidence and that the alternative definition suggested by Kiyofuji and Ochi (2016) based on tagging and larvae survey matches our result better than that of reference case. Because the clustering result has any mismatches even if the alternative area definition, we propose a new area definition which better explains our clustering result (**Figure 12**). We recommend SC14 to consider this new area definition as the reference case of spatial stratification in the next skipjack stock assessment.

Future work plan

1. It was found that latent differences of fishery form can be classified by using finite mixture model with mixing GLMs (likelihoods) of biological information (SKJ mean body weight from logbook) and of CPUE. However, the biological variable used in this analysis was not actually measured therefore we think that it is necessary to reanalysis using fork length composition data measured at many ports in Japan. When reanalyzing, it should be take into account for area covering rate of sample collection area to operational area and the similarity of fork length size composition compared to body weight composition converted with fork length-body weight (allometric) relationship.

2. Based on the result of reanalysis as mentioned at future work plan 1, an improvement of JPNPL abundance index would be addressed considering the classification.

References

Kiyofuji, H. and Ochi, D. (2016) Proposal of alternative spatial structure for skipjack stock assessment in the WCPO. WCPFC-SC12-2016/SA-IP-09.

Lehodey, P., Bertignac, M., Hampton, J., Lewis, A. and Picaut, J. (1997). El Niño Southern Oscillation and tuna in the western Pacific. *Nature*, 389(6652), p.715.

Leash, F. (2004) FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R. Journal of Statistical Software, 11(8). URL https://www.jstatsoft.org/article/view/v011i08

Matsumoto, W.M., 1984. Synopsis of biological data on skipjack tuna, *Katsuwonus pelamis*. *NOAA Tech*. *Rep*. NMFS Circ., 451, pp.1-92.

McKechnie S., Hampton J., Pilling G. M., Davies N. (2016) Stock assessment of skipjack tuna in the western and central Pacific Ocean. WCPFC-SC12-2016/SA-WP-04.

Table 1. Characteristics of each cluster. Fishing quarters at clusters 1, 2, and 4 shifted between the 1980s and 1990s, which would reflect the change of fishing form according to the large-scale entry of purse seiner in 1980s. Mean body weight is classified into two size categories, about 2 kg (cluster 1, 3, and 5) and around 4 kg (cluster 2 and 4). Median absolute deviation (MAD) is a more robust variation index than standard deviation which is sensitive to outlier data. The medians of mean body weight at clusters 1, 3, and 5 shifted larger between the 1980s and 1990s and its value at cluster 4 changed bigger between 1990s and 2000s. Catch and effort at clusters 1 to 4 drastically decreased from the 1980s to 1990s and these at cluster 5 has gradually decreased from 1990s to present. Clusters 1, 3, and 5 are mainly included in region 1 on reference case area definition but they are separated into three regions (1, 6, 7) on alternative area definition. Cluster 2 is mainly included in two regions (2 and 3) on both area definitions. Cluster 4 is included in three regions (2 to 4) on reference case area definition but in two regions (2 and 6) on alternative area definition.

Cluster	Main fishing quarter	weighted BW (kg) on peak density	Turning point of weighted BW	Turning point of Catch & Effort	Main region of reference case	Main region of alternative
1	~1980s: all 1990s~: 1, 2	1.9 MAD = 2.97	1980s | 1990s	1980s | 1990s	1	1, 6, 7
2	~1980s: all 1990s~: 4, 1	3.4 MAD = 0.74	-	1980s | 1990s	2, 3	2, 3
3	all	1.8 MAD = 0.74	1980s | 1990s	1980s | 1990s	1	7
4	~1980s: all 1990s~: 4, 1, 2	3.9 MAD = 1.48	1990s | 2000s	1980s | 1990s	2, 3, 4	2, 6
5	2, 3, 4	2.2 MAD = 0.74	1980s | 1990s	1990s~	1	1, 7

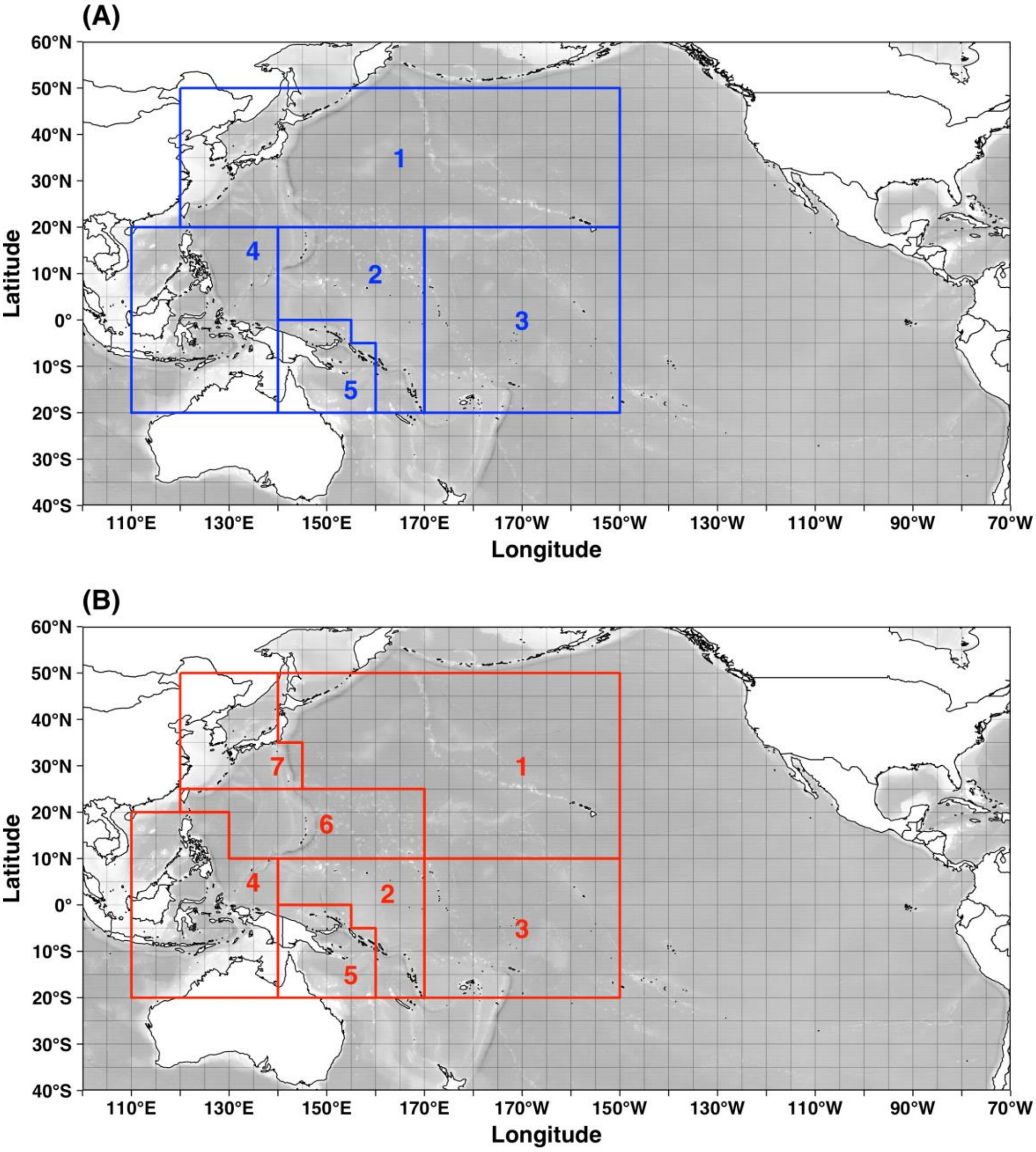


Figure 1. Area definition of (A) reference case and (B) alternative definition suggested by Kiyofuji and Ochi (2016).

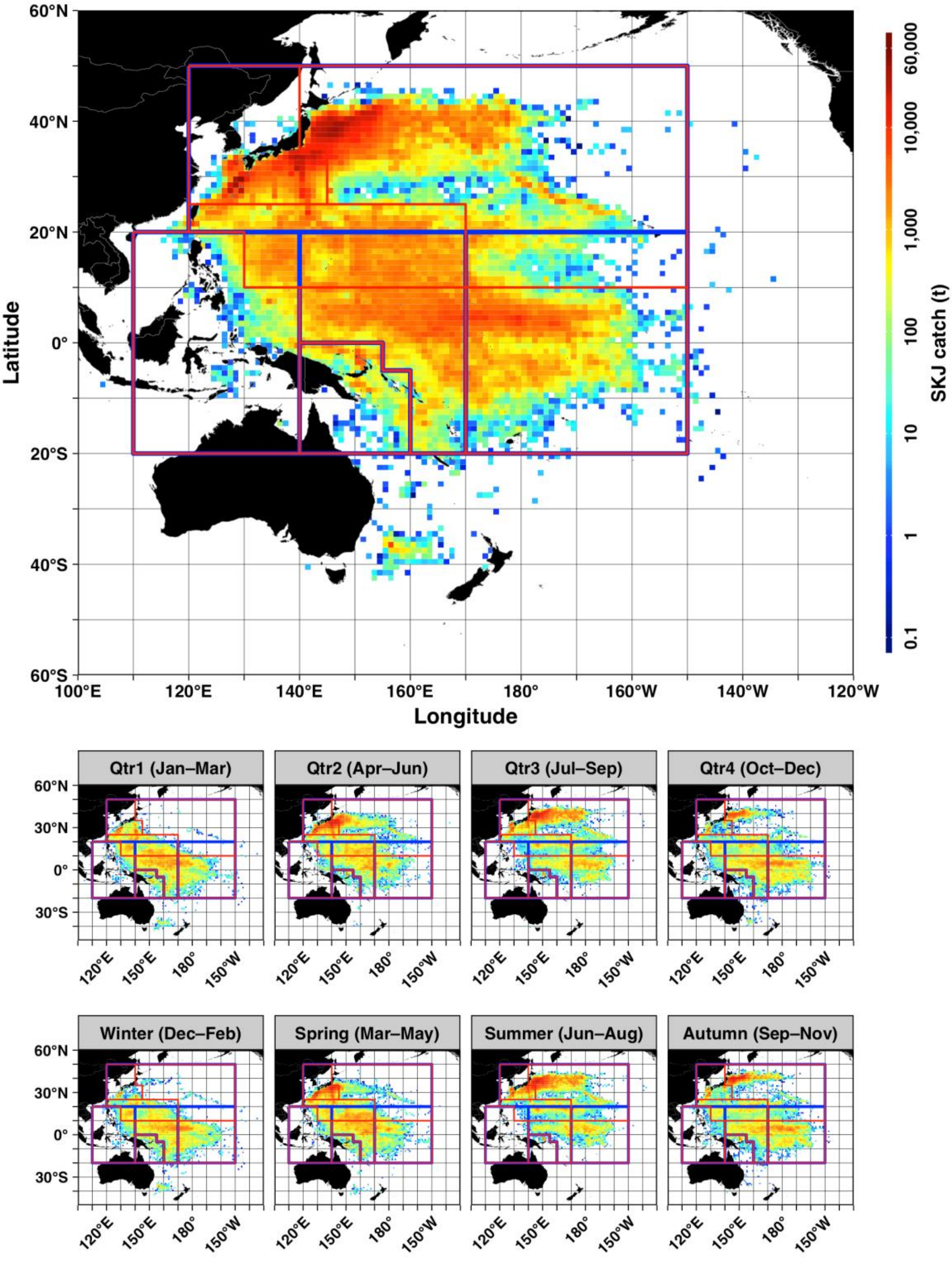


Figure 2. Distribution of SKJ catch from JPNPL logbook data (Top) Total (Middle) by quarter (Bottom) by season. Blue and red (thinner than blue lines) solid lines indicate reference case region borders and alternative region borders suggested by Kiyofuji and Ochi (2016), respectively.

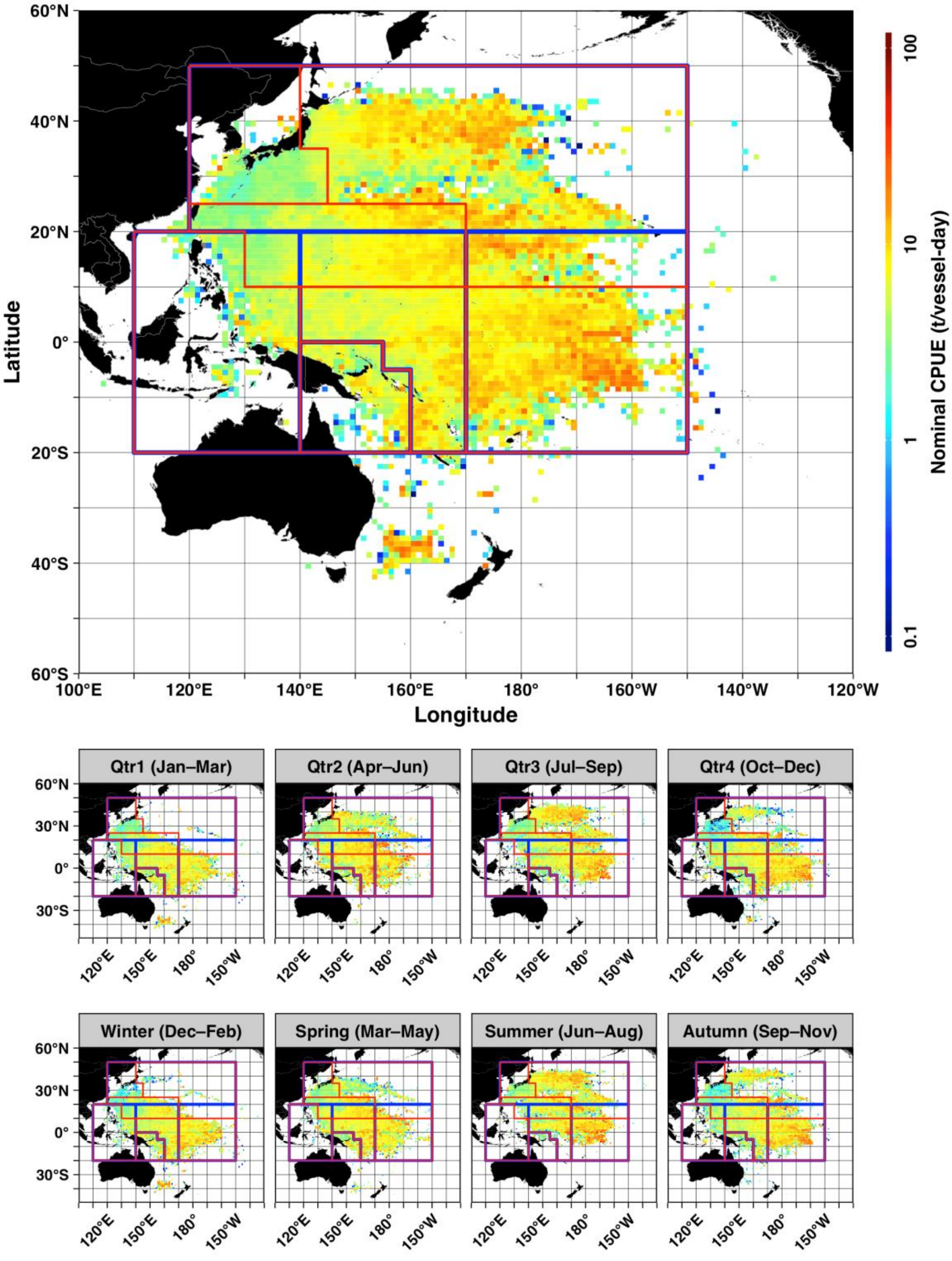


Figure 3. Distribution of SKJ nominal CPUE from JPNPL logbook data (Top) Total (Middle) by quarter (Bottom) by season.

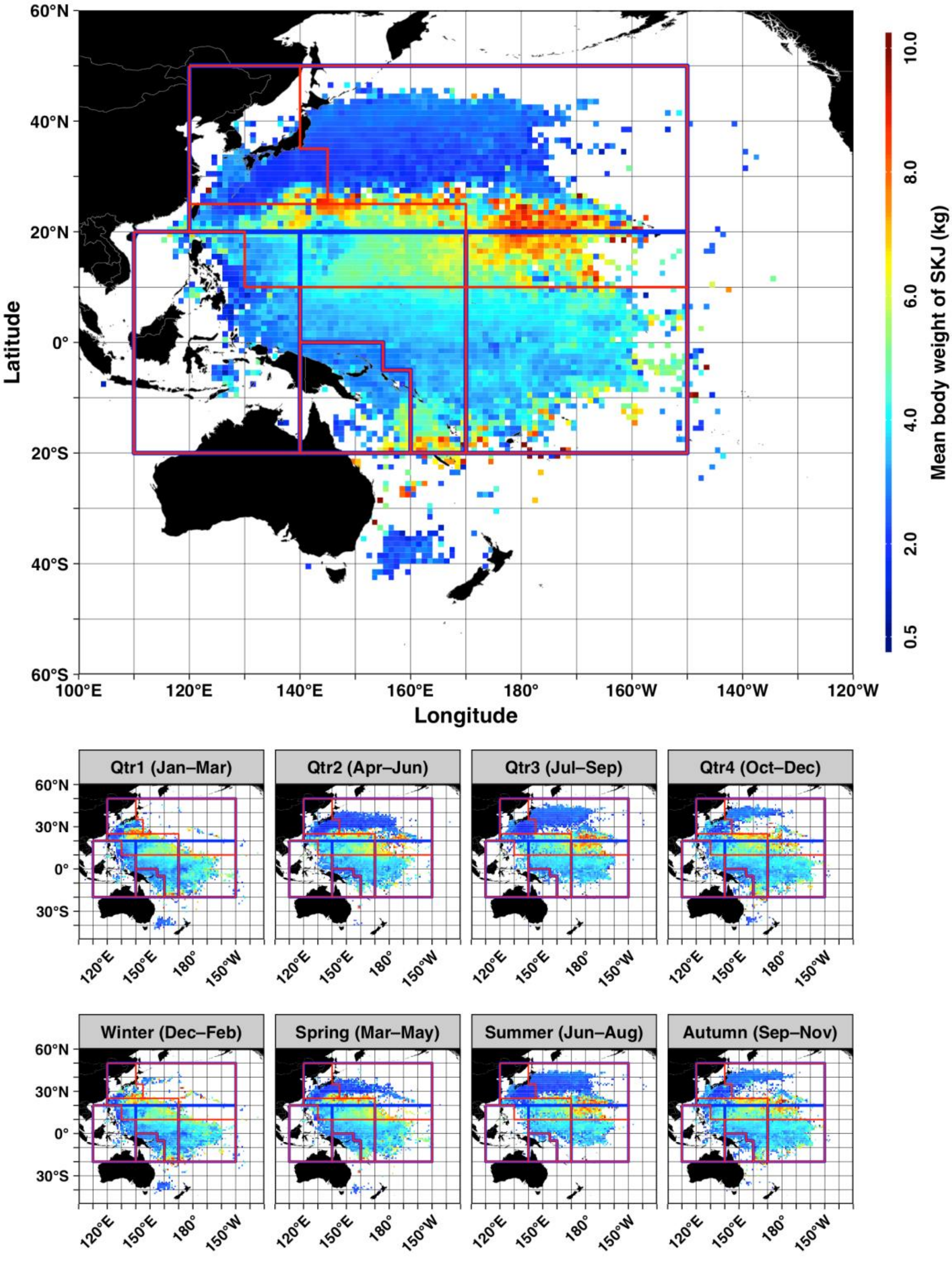


Figure 4. Distribution of weighted BW from JPNPL logbook data (Top) Total (Middle) by quarter (Bottom) by season. BW is weighted in each 1°×1° grid.

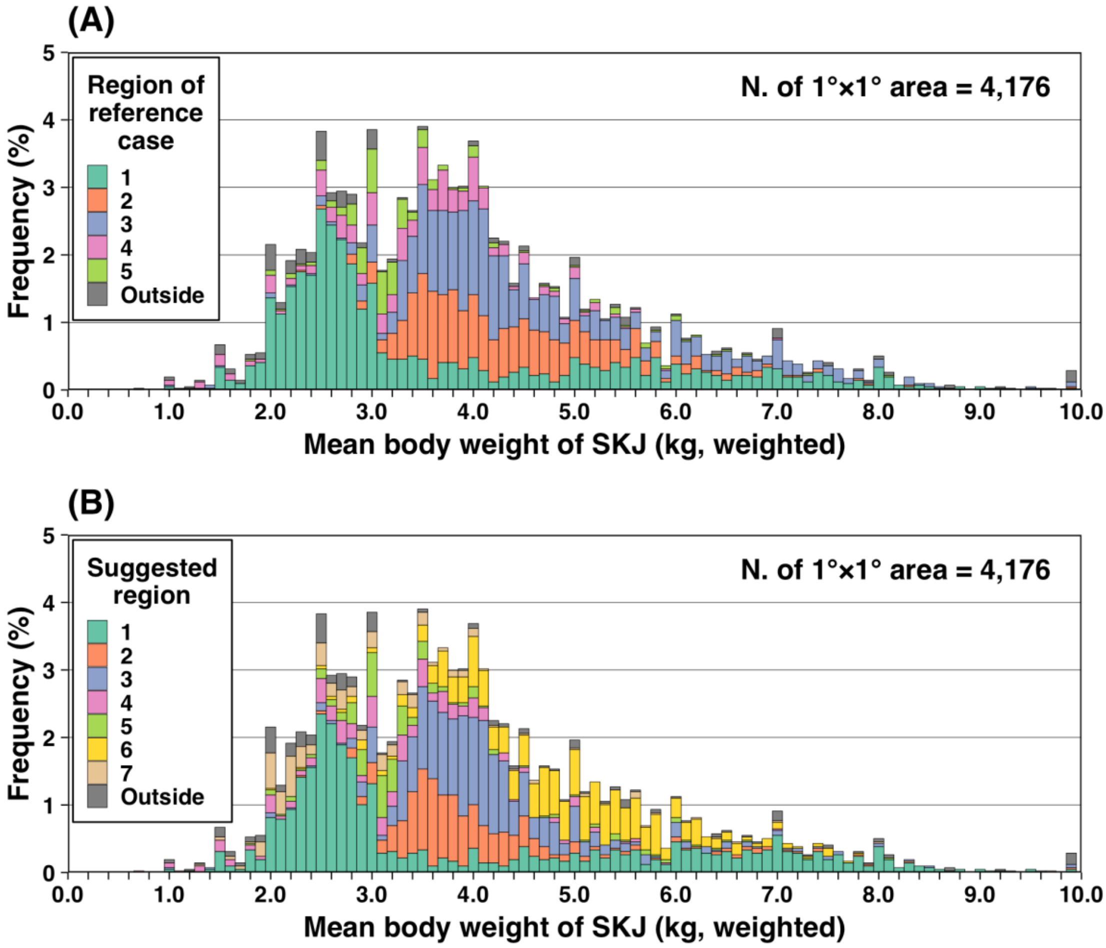


Figure 5. Composition of SKJ weighted mean body weight by region (A) reference case (B) alternative region suggested by Kiyofuji and Ochi (2016).

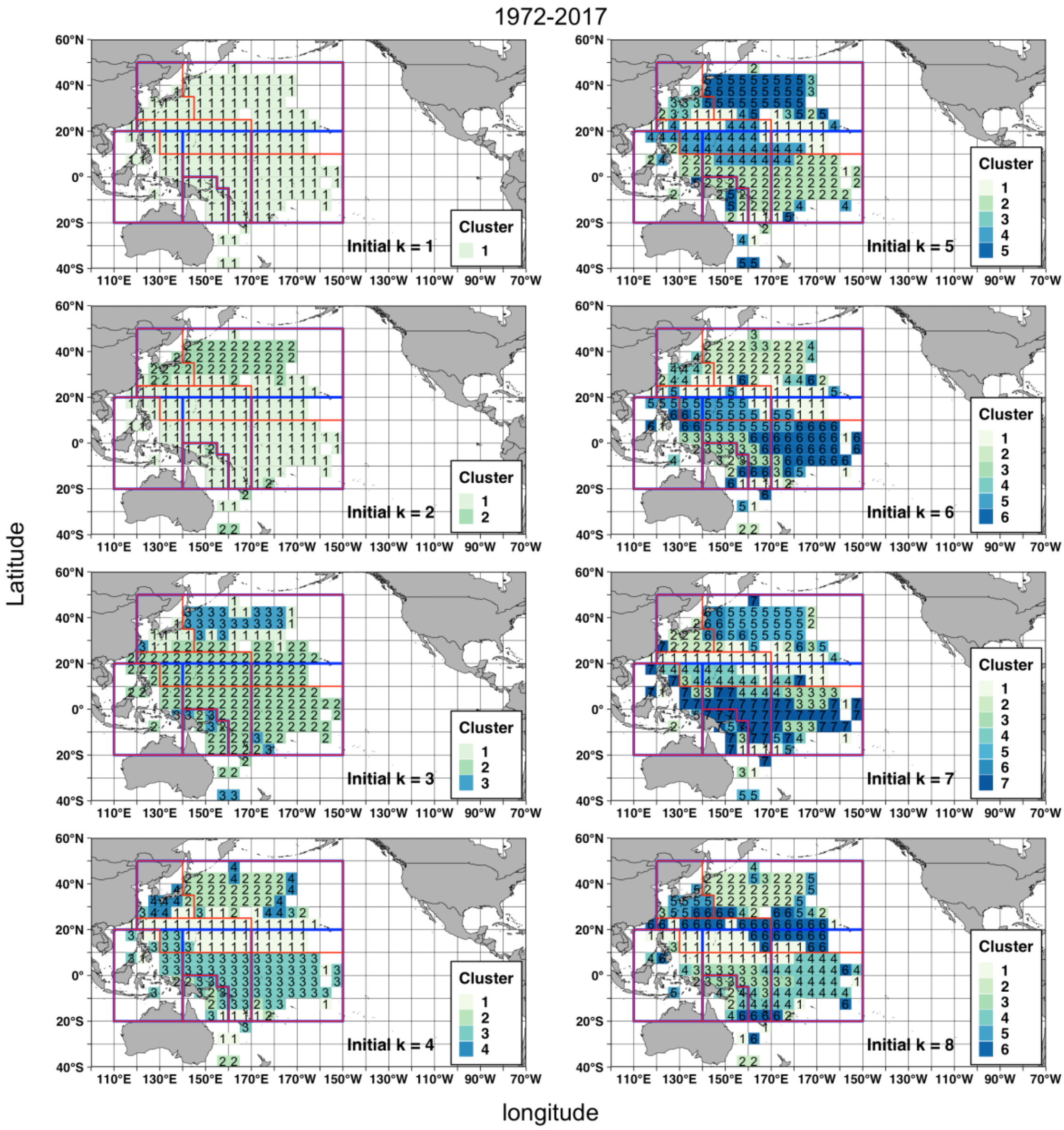


Figure 6. The area clusters defined by the finite mixture model (flexmix). R package 'flexmix' needs to set different initial clusters because it is unsupervised learning. The distribution of initial cluster five was shown at top-right.

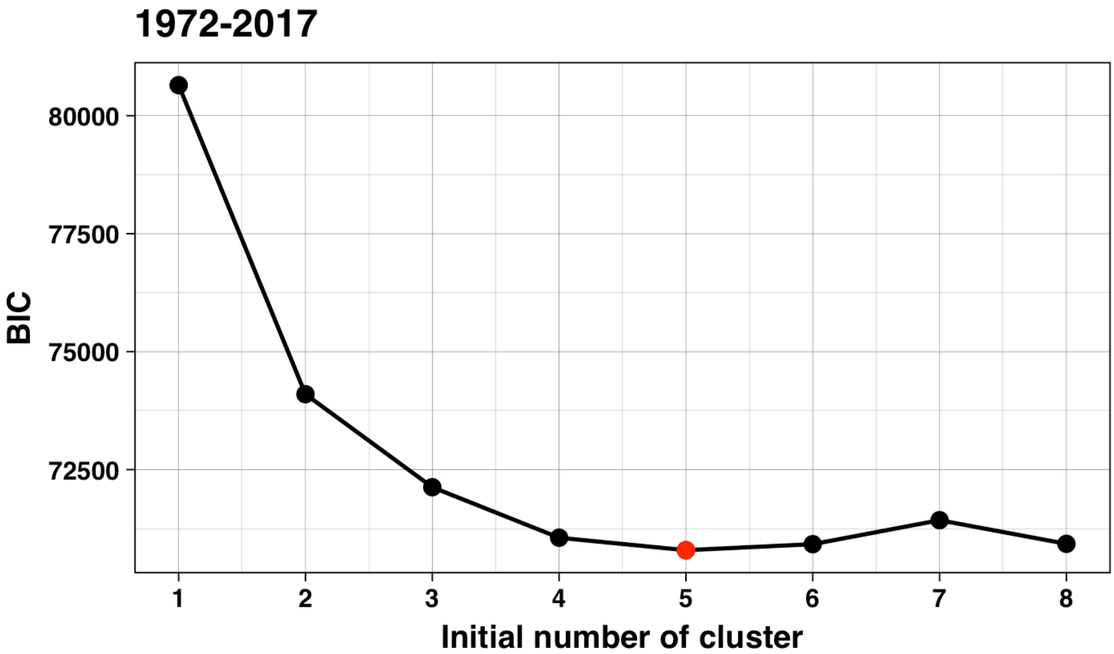


Figure 7. Bayesian information criterion (BIC) of different initial clusters. Initial clusters were set from one to eight in this analysis. Initial number five has minimum BIC and was selected.

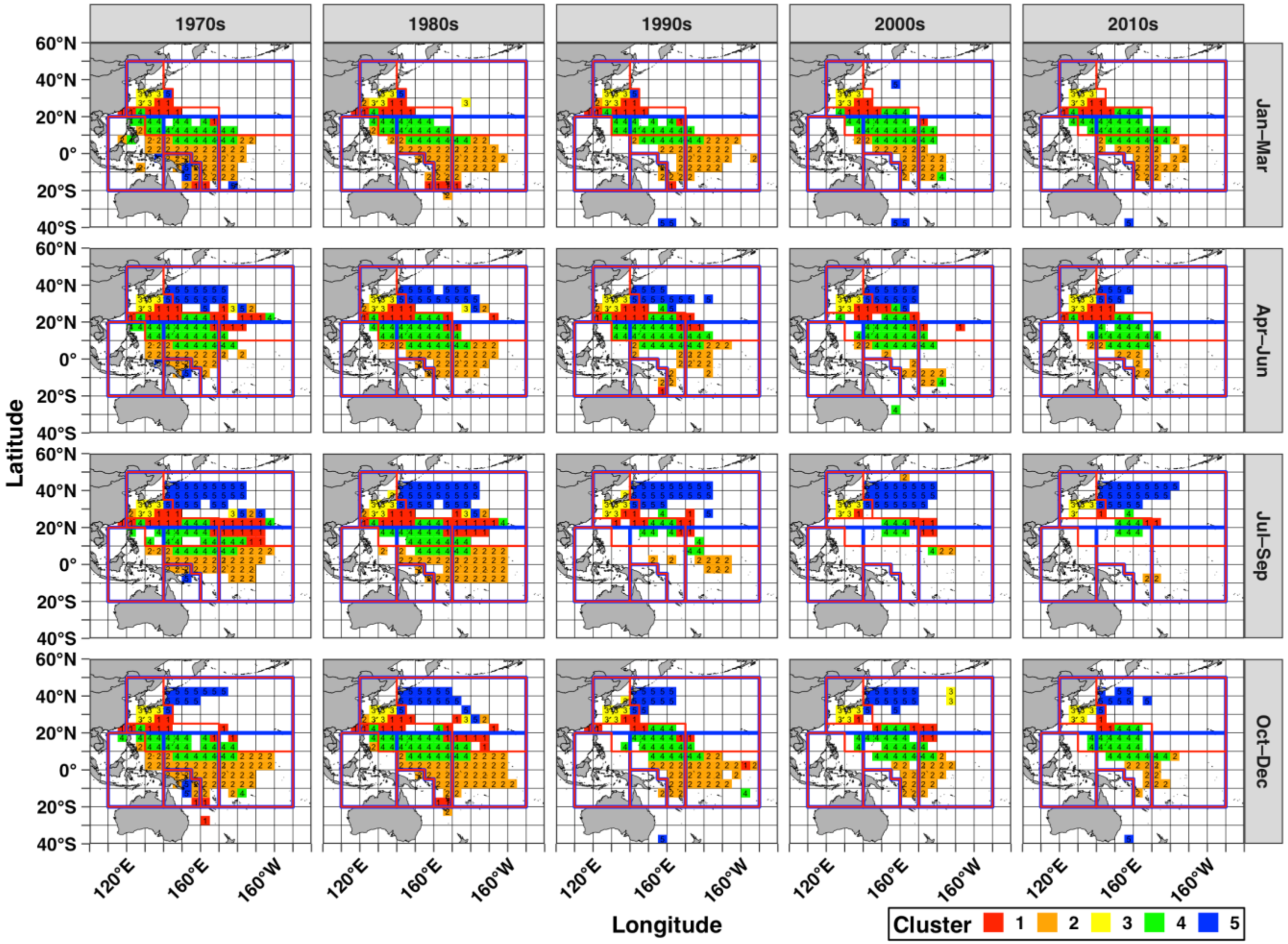


Figure 8. Distribution of each cluster by decade and quarter. Each cluster distributes almost same area from past to present. Cluster 1 distributes between cluster 4 and 5. Cluster 2 distributes around the tropical area all year round to 1980s. Cluster 3 (yellow tile) distributes around Nansei Islands of Japan. Cluster 4 distributes north subtropical area. Cluster 5 distributes north temperate area over 30°N in quarter 2 to 3 mainly. Considering SKJ seasonal migration, it will be important that the relationship among clusters 1, 4, and 5.

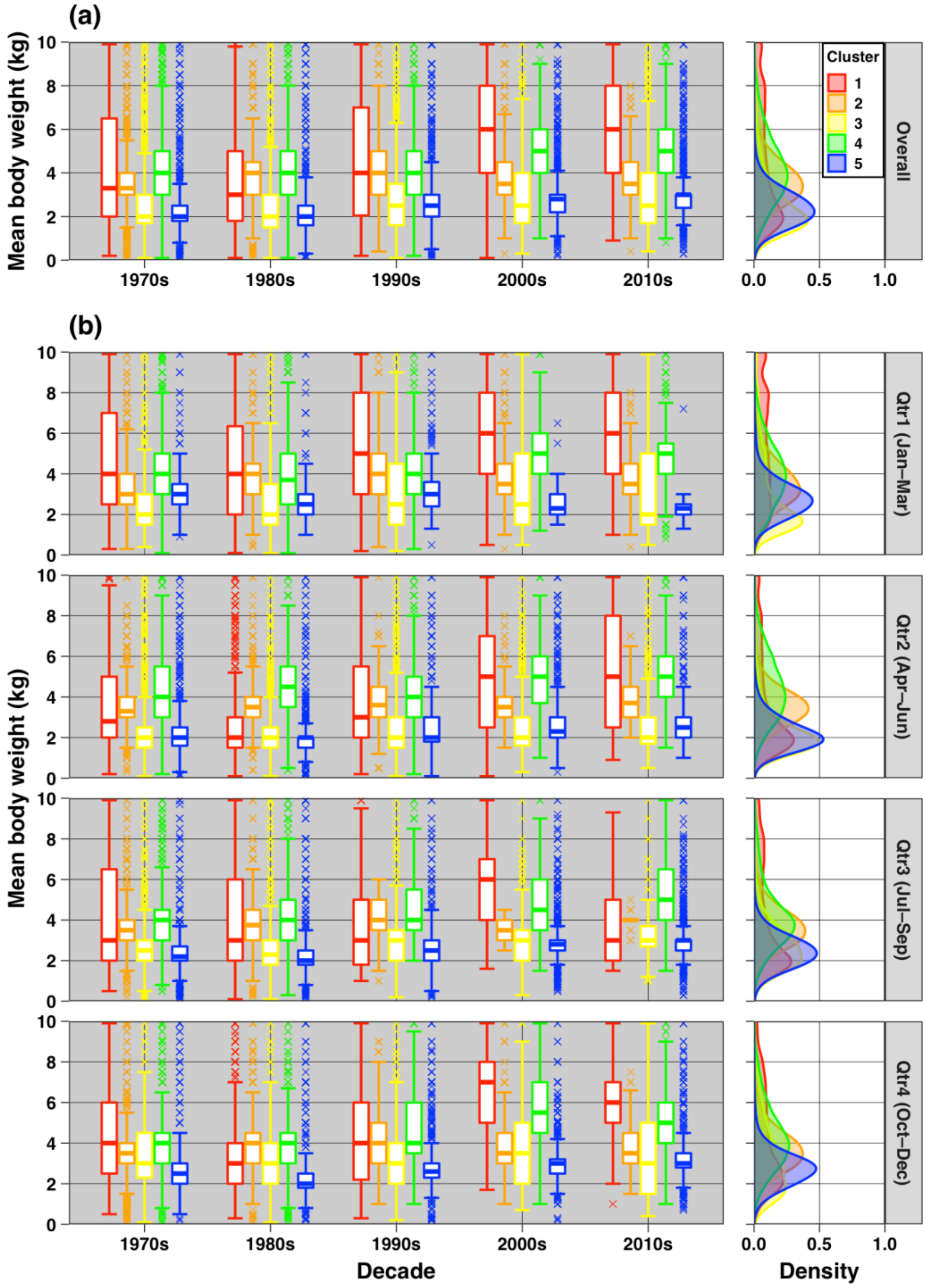


Figure 9. Box-plot and density-plot of weighted mean body weight by decades (a) all round year (b) by quarter. The density peaks of five clusters indicate two body size classes, which are about 2 kg class (cluster 1, 3, and 5) and about 4 kg class (cluster 2 and 4). Cluster 1, which horizontally distributes between cluster 4 and 5 displayed in **Figure 8**, has large deviation and shows the shift up of its median between the 1990s and 2000s. Cluster 2 shows no shift of its median of mean body weight. Cluster 3 shows the shift up between the 1980s and 1990s. Cluster 4 shows the shift up between the 1990s and 2000s. Cluster 5 shows the shift up between the 1980s and 1990s and its deviation getting smaller.

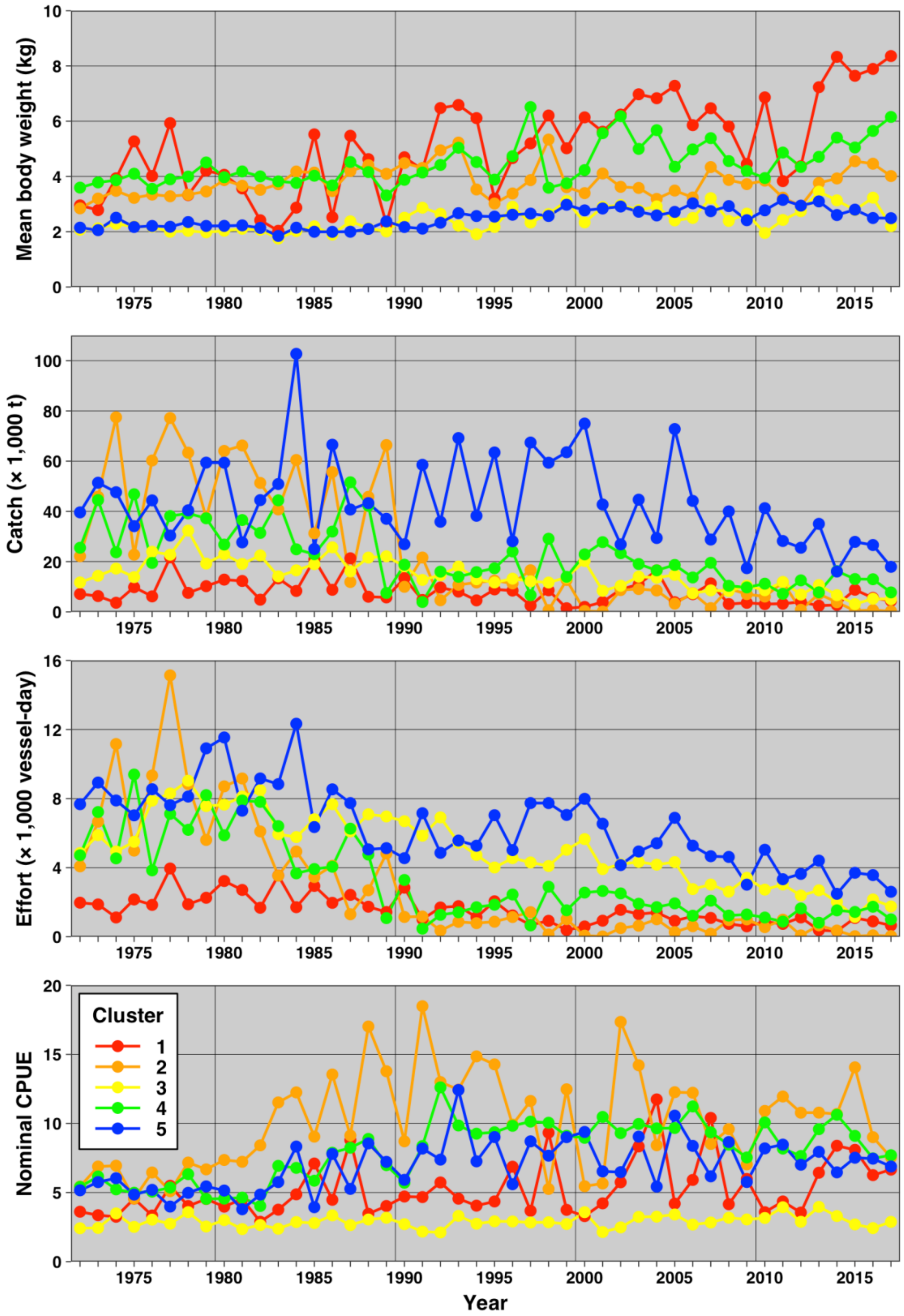


Figure 10. Time series of weighted mean body weight, total catch, effort, and nominal CPUE by cluster. Total catch drastically decreased between the 1980s and 1990s at all clusters except for cluster 5 (blue line). Efforts at clusters 1, 2, and 4 steeply decreased in the 1980s, on the other hand, those at cluster 3 and 5 gradually decreased from 1980s to present. Nominal CPUE indicates almost the same value for each cluster.

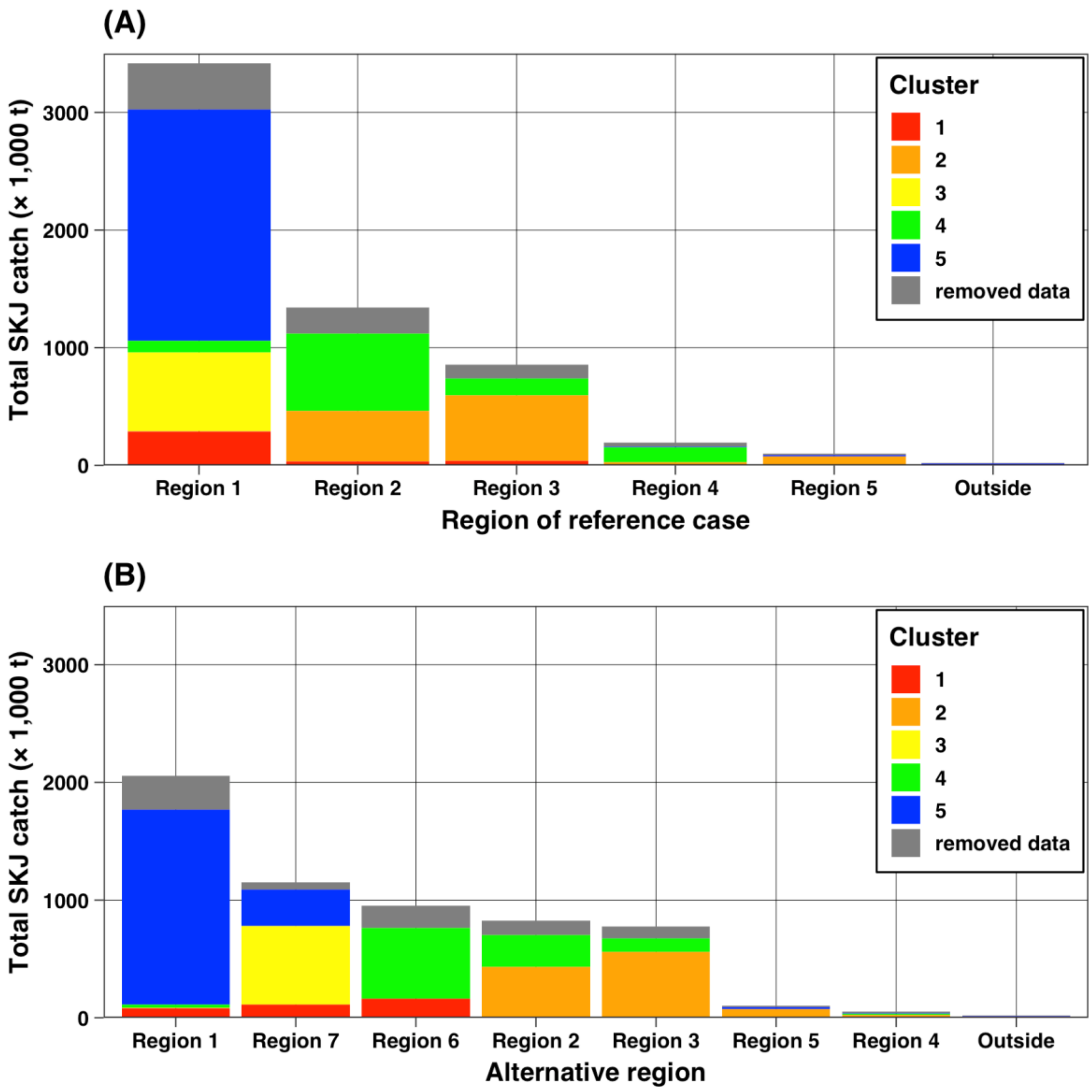


Figure 11. SKJ total catch of each region by cluster (A) reference case region (B) alternative region suggested by Kiyofuji and Ochi (2016). This bar-plot was sorted by the total catch of each region. Reference case region 1 mainly includes cluster 1, 3, and 5. On the other hand, alternative region 1 mainly includes cluster 5. Regions 2 and 3 of reference case and alternative mainly include cluster 2 and 4. Reference case regions 4 and 5 mainly include cluster 4 and 2 respectively but the values of total catch are low. Similarly, alternative regions 4 and 5 mainly includes clusters 2 but the values of total catch are low. Alternative region 6 mainly includes cluster 1 and 4. Alternative region 7 mainly includes cluster 1, 3, and 5.

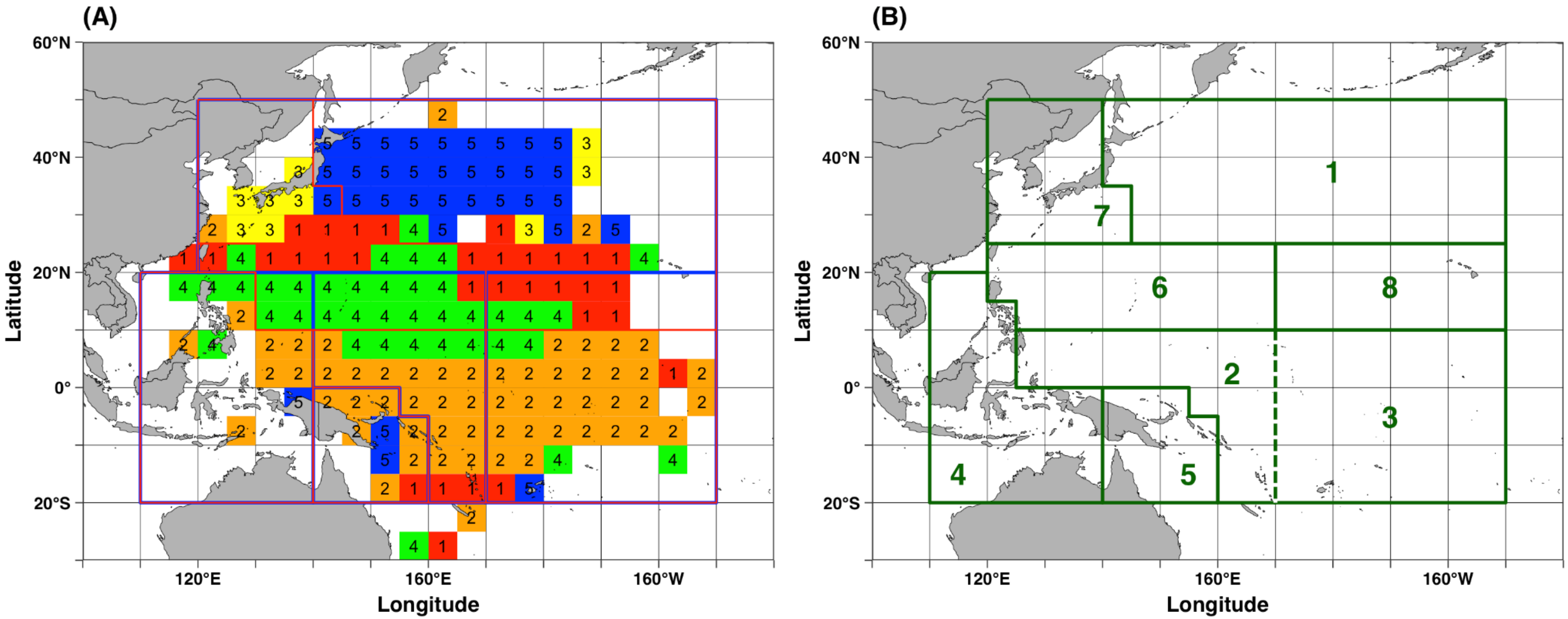
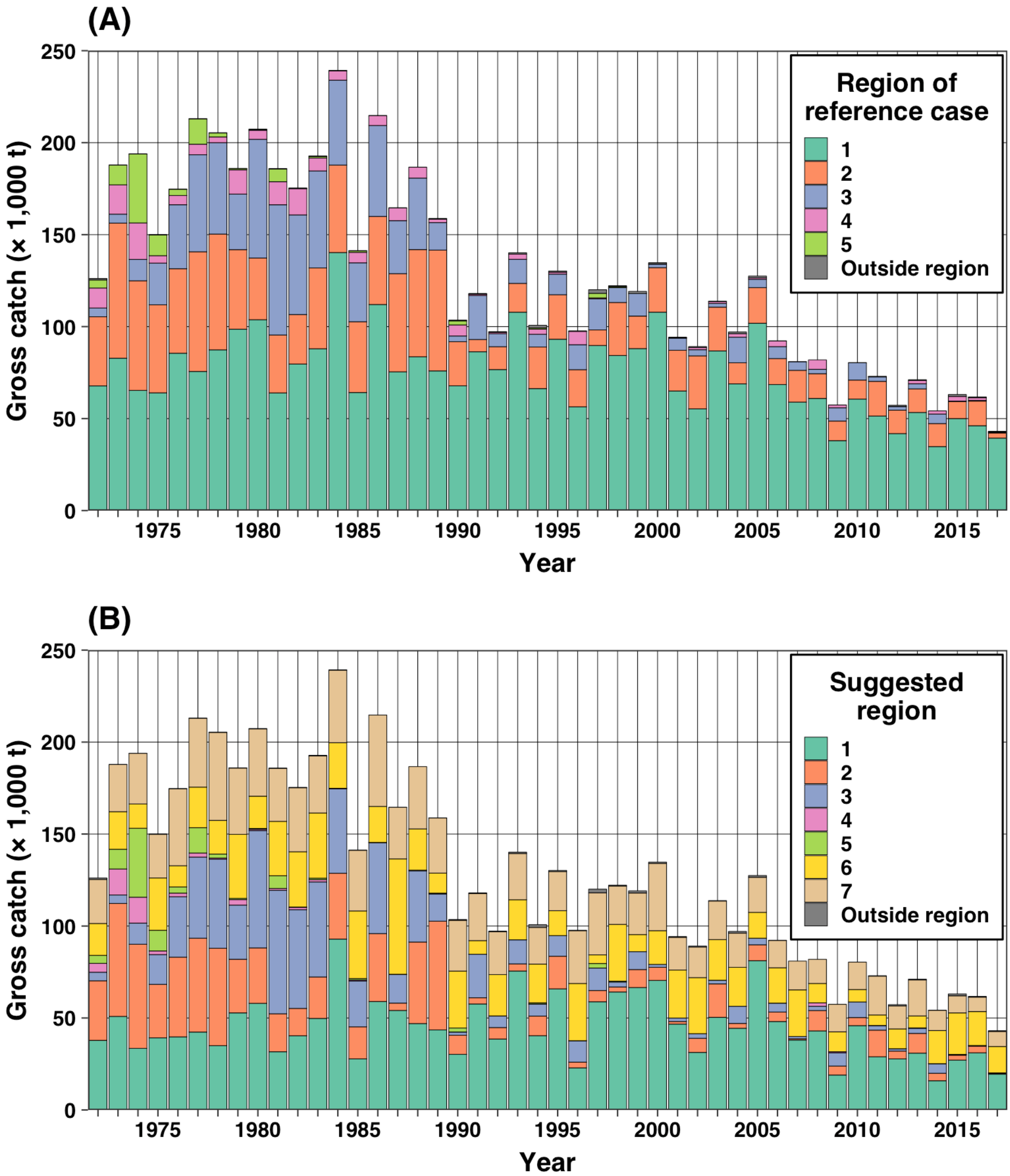
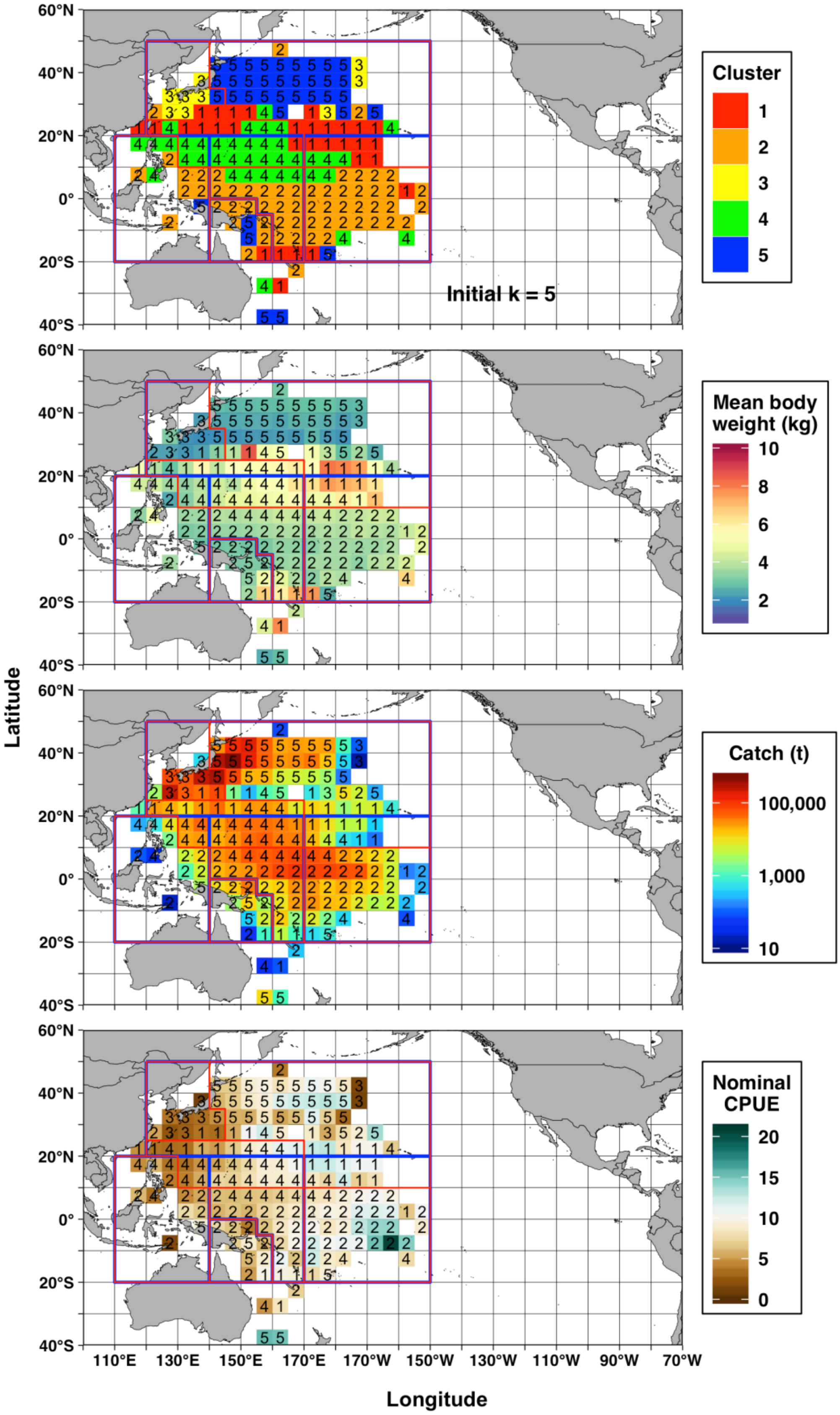


Figure 12. (A) Distribution of five clusters and (B) a new area definition which we propose. In the left panel, blue and red (thinner than blue lines) solid lines indicate reference case region borders and alternative region borders suggested by Kiyofuji and Ochi (2016), respectively.



Supplement figure 1. Time series of SKJ total catch by region (A) reference case (B) alternative region.



Supplement figure 2. Distribution, weighted BW, total catch, and nominal CPUE of five clusters (the minimum BIC).