Geographical characteristics of CPUE for Pacific bluefin tuna caught by Japanese coastal longliners.

Yohei Tsukahara and Osamu Sakai

National Research Institute of Far Seas Fisheries,
Fisheries Research and Education Agency
5-7-1, Orido, Shimizu-ku, Shizuoka 424-8633, JAPAN

February 2017

Working document submitted to the ISC Pacific bluefin tuna Working Group, International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean (ISC), from 15 to 20 February 2017, Shimizu, Shizuoka, JAPAN.

Summary

It is considered that the CPUE for Pacific bluefin tuna (PBF) caught by Japanese coastal longliners has geographical trend. Ignoring geographical effects might lead to misunderstanding of the annual fluctuation of stock status. This study applied GAM to CPUE data in order to evaluate geographical effect. An interaction term of longitude and latitude showed remarkable positive effect on CPUE around Nansei Islands, in addition it was found somewhat a belt-like geographical effect offshore. For detailed analysis, year effect was included in the interaction term of geographical effect in order to investigate geographical differences by year. The results indicate that annual fluctuations are spatially correlated, but year effects have somewhat regional trend.

Introduction

CPUE for Pacific bluefin tuna (PBF) caught by Japanese coastal longliner is one of the important indices for their stock assessment. For the standardization of this CPUE, geo-statistical model has been suggested to consider the geographical effects in CPUE and this is considered as highest priority in the specific plan of PBF WG research (ISC/PBFWG 2016). Geographical trends on longline fisheries, such as number of catch and operation, probably exist. Lack of geographic perspective might lead to misunderstanding of the CPUE trend.

At the last stock assessment, CPUE was standardized by generalized linear model (GLM) (Sakai et al., 2016). The easiest and simplest way to attach the geographical effects to previous model is adding the term of geographical information on GLM, such as "area", "latitude" and/or "longitude" (Oshima et al. 2012). Operation area of longline, however, is so spread that it is difficult to evaluate geographical effects on CPUE trend as fixed categorical term. Moreover, it shouldn't be treated as continuous variable in GLM because it is thought that geographical effects are probably non-linear term. On the other hand, generalized additive model (GAM) is a kind of regression model same as GLM and is useful tools for model including non-linear effects, using interpolation by tensor product or spline curve.

This paper describes a way of estimating geographical effects on Japanese longline CPUE using GAM as a first step to address geo-statistical model. At first, geographical effects represented by an interaction term of longitude and latitude were estimated with the main terms that is referred to from previous GLM. In addition, year effect was also included in this geographical interaction term to evaluate the geographical differences by year.

Material and method

Data sources and filtering

The logbook data of Japanese coastal longliners was used for this analysis. The resolution of the data is set-by-set operational level, which have not only PBF catch but a lot of information, such as hook per basket (hpb), catch of other fish, longitude, latitude and so on. These data can be used as explanatory variables. A disadvantage of GAM is sensitivity to the distant values from the other data because of spread of interpolation range. Filtering of data, therefore, was different from previous criteria for the standardization used in Sakai and Tsukahara (2017) in terms of catch performance by area. Season was filtered same as the previous criteria in order to focus on major fishery for PBF as follows;

- ✓ Area: 120-140 degree of east longitude, 20-40 degree of north latitude,
- ✓ Season: April to June.

Generalized additive model

GAM can contain the terms that are both linear and non-linear effects because of smoothing term. This advantage is helpful for analysis of the geographic effects in this study. It enables to develop the models with continuous numerical explanatory variables easily, such as latitude and longitude. GAM with smoothing terms is very powerful to understand responses of CPUE to explanatory variables. On the other hand, GAM has some disadvantage compared to previous CPUE standardization by GLM. The effects estimated by smoothing in GAM are results of optimization. In other words, the effects at the data points have some margin from effects before smoothing, potentially affected by smoothing. The results don't represent proper estimation at the data points. This makes it difficult to calculate the LSMEANs that was used in precious standardization. Additionally, spread of poor or no data points lead to reliability degradation in smoothing. It requires picking data that have continuously a certain amount of data.

CPUE data for PBF have so many zero-catch data that previous standardizations took some measure to deal with. In this analysis, tweedie distribution (Peter and Gordon 2005) was adopted, using "gam" (available in R package "mgcv") for R software ver. 3.3.1. This distribution is potential candidate to suppress the influence by a lot of zero catch (Shono 2008) and has been used in several standardizations of CPUE recently (Humberto et al. 2011, Rui et al. 2013). The power parameter, p, in tweedie distribution was arbitrarily defined as 1.1 to accept many zero catch data. While GAM using tweedie distribution is a powerful tool for analyzing model that have non-liner effects, it has a risk of over-fitting. A parameter in "gam", *gamma*, of penalty for degrees of freedom in smoothing terms contribute to preventing model from over-fitting and it was set to 1.4 in accordance with previous studies (Simon 2006).

Candidate of explanatory variables were selected by reference from some previous studies (Sakai and Tsukahara 2017, Oshima et al. 2012). The advantages of GAM over GLM is to be able to represent the non-liner effects of continuous value. Explanatory variables were treated as non-linear terms using spline smoother allow flexibility for fitting model. The explanatory variables in final model were selected as non-linear terms by BIC through the backward method (decreasing variable). Only the longitude and latitude were treated as interaction term because combination of them are more meaningful as area information. After selection, main term was changed to linear effects or categorical effects to decide their appropriate forms using BIC. Candidate of explanatory variables are as follows;

- Year: 23 calendar years, from 1994-2016.
- Day10: Period during the spawning season, from April to June. defined by 10 days interval. (last period in May contained 11days)
- Ship-size: Gross ton of each vessel.
- Days for cruise: Duration per trip.
- Hook per basket (hpb)
- CPUE of each other species in logbook: CPUE of yellowfin tuna (YFT), albacore (ALB), bigeye tuna (BET), amount of the others (sharks and billfishes) (Other).
- Longitude and latitude: 1x1 degree of longitude and latitude.

Every effect including smoothing terms was discussed about similarities to and difference from GLM analysis. Especially, the term of geographical effects was focused and analyzed in detail, if there seems to be some kind of area structure in the result.

Analyzing geographical structure of year effects

The relations of year effects among area are required to formulate the geographical structure in geo-statistical model. To address this, the year effect was included into the interaction term of geographical effects. This change enables GAM to calculate year effects by 1x1 degree cells of longitude and latitude. The year effects in every 1x1 degree cell are so vast amount of data, thus they were classified broadly into some area groups according to cluster analysis using “hclust” and “ward.D2” method (available in R package “stats”) for R software ver. 3.3.1. The relations in and among each cluster were evaluated by correlation coefficients. The high coefficient means close relations in and among area groups and makes it easy to formulate geographical relations in geo-statistical model.

Results and discussion

Final model was as bellow,

__Final model__

log(μ) = (intercept) + Year + s(Longitude, Latitude) + s(Day10) + s(Days for cruise) + l(YFT) + (error term)

where l( ) means linear term of continuous value, s( ) means spline terms and no bracket mean factor term. In addition, maximum degree of freedom of day 10 was set to 8 because its default value, 10, was over the number of day10.

Estimated effects of each term are shown in Fig. 1 and Fig. 2. The trends of all effects were similar to LSMEANs estimated by GLM in standardized CPUE including year trend (Sakai and Tsukahara 2017). Moreover, there was some additional information due to their continuousness. For example, the smoother term of days per trip in GAM showed an inflection point at near 25 days, while current standardization showed a somewhat different threshold at 14 days in this categorical terms (Sakai and Tsukahara 2017). Results of smoothing could be useful information for deciding a categorization. With respect to area effects, the area around Nansei Islands had higher value than the other area. Furthermore, there was belt-like area of positive value off shore along Japan main island. It means catchability in offshore area is relatively higher than nearshore area. These results indicate area effects on CPUE vary according to area. In order to examine geographical difference of year effects, final model was modified in terms of interaction term among longitude, latitude and year as bellow.

__Temporal and spatial interaction effects model__

log(μ) = (intercept) + s(Longitude, Latitude, Year) + s(Day10) + s(Days for cruise) + l(YFT) + (error term)

Estimated effects with exception of interaction term between year and geographical effects were quite similar to estimated effects in final model. Moreover, BIC was little different from the final model (slight increase). This model could be used to discuss effect of interaction term between year and geographical effects. The year effects estimated by 1x1 degree cells seems to have some kind of regional tendency. Therefore, they were classified into some groups by cluster analysis, using the normalized year effects by 1x1 degree cells which have over 20 data of year effects, because poor number of data might lead to underestimation of distance among each data.

The positions and averaged year effects of each cluster are shown in Fig. 5 and 6, respectively. The cluster 1 was compose by area around Nansei Islands, the cluster 2 was composed by offshore area, and the cluster 3 was composed by nearshore area. This result was consistent with geographical tendency of smoothing effects in the final model. Additionally, the year effects in the cluster 1 didn't have long-term trend but moderate oscillation. On the other hand, the year effects in the cluster 2 and 3 show declining trend with small short-term fluctuation until around 2012. Moreover, there is a slight phase difference between them. These results indicate that it might be necessary to consider geographical structure in terms of not only target shift but also annual fluctuation of catchability in standardized CPUE.

The spatial correlation coefficients could be one of the important index for geo-statistical model. Table 1 shows correlation coefficients between each cluster and among components in each cluster. The cluster 2 and 3 had high correlation with each other. Moreover, correlation coefficient between cluster 2 and 1-year behind cluster 3 was higher value, 0.97. On the other hand, the cluster 1 had relatively low correlation coefficients both with the other cluster and among its components. These infer that relations in geographical structure aren't steady pattern spatially. The geo-statistical model is required to consider these unsteady patterns of geographical structure. It is necessary to figure out relations among these areas and among 1x1 degree cells cautiously and precisely.

These results were preliminary analysis for development of geo-statistical model. Considering the causations of different catchability by area with several perspectives, for example length composition of catch by area, migration and so on, will lead to deeper understanding of geographical effects on PBF abundance.

Reference

ISC/PBFWG 2016. Report of the Pacific bluefin tuna working group workshop, 29 February – 11 March 2016 La Jolla, California, USA

Sakai, O., Y. Hiraoka, and K. Oshima. 2016. Japanese coastal longline CPUE for Pacific bluefin tuna: Re-update up to 2014 fishing year for stock assessment. ISC16/ PBFWG-1/01

Oshima, K., Mizuno, A., Ichinokawa, M., Takeuchi Y., Nakano H., and Uozumi, Y. 2012. Shift of fishing efforts for Pacific blefin tuna and target shift occurred in Japanese coastal longliners in recent years. ISC/12/PBFWG-3/05.

Sakai O., Tsukahara Y. 2017. Japanese coastal longline CPUE and Catch at length for Pacific bluefin tuna: Update up to 2015 fishing year. ISC/16/PBFWG-1/

Peter K. Dunn, Gordon K. Smyth. 2005. Series evaluation of Tweedie exponential dispersion model densities. Statistics and Computing, Vol. 15, pp. 267-280.

Shoho H., 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fisheries Research, Vol. 93, pp.154-162.

Humberto G Hazin, Fabio Hazin, Paulo Travassos, Thierry Frédou. 2011. Standardized CPUE series of Blue marlin caught by Brazilian tuna longline fisheries in the southwestern Atlantic Ocean (1980-2008). Sci. Pap. ICCAT. Vol. 66.

Rui Coelho, Miguel N. Santos, Pedro G. Lino. 2013. Standardized CPUE series for blue and shortfin mako sharks caught by the Portuguese pelagic longline fishery in the Indian ocean, between 1999 and 2012. IOTC–2013–WPEB09–22

Simon N Wood. 2006. Generalized Additive Models: an introduction with R. Chapman & Hall/CRC.

Table 1 Correlation coefficients among each cluster and among components in each cluster.

(Diagonal elements are averaged correlation value among components in each cluster.)

	Correlation coefficient		
	Cluster 1	Cluster 2	Cluster 3
Cluster 1	0.64	-	-
Cluster 2	0.74	0.96	-
Cluster 3	0.71	0.95	0.95

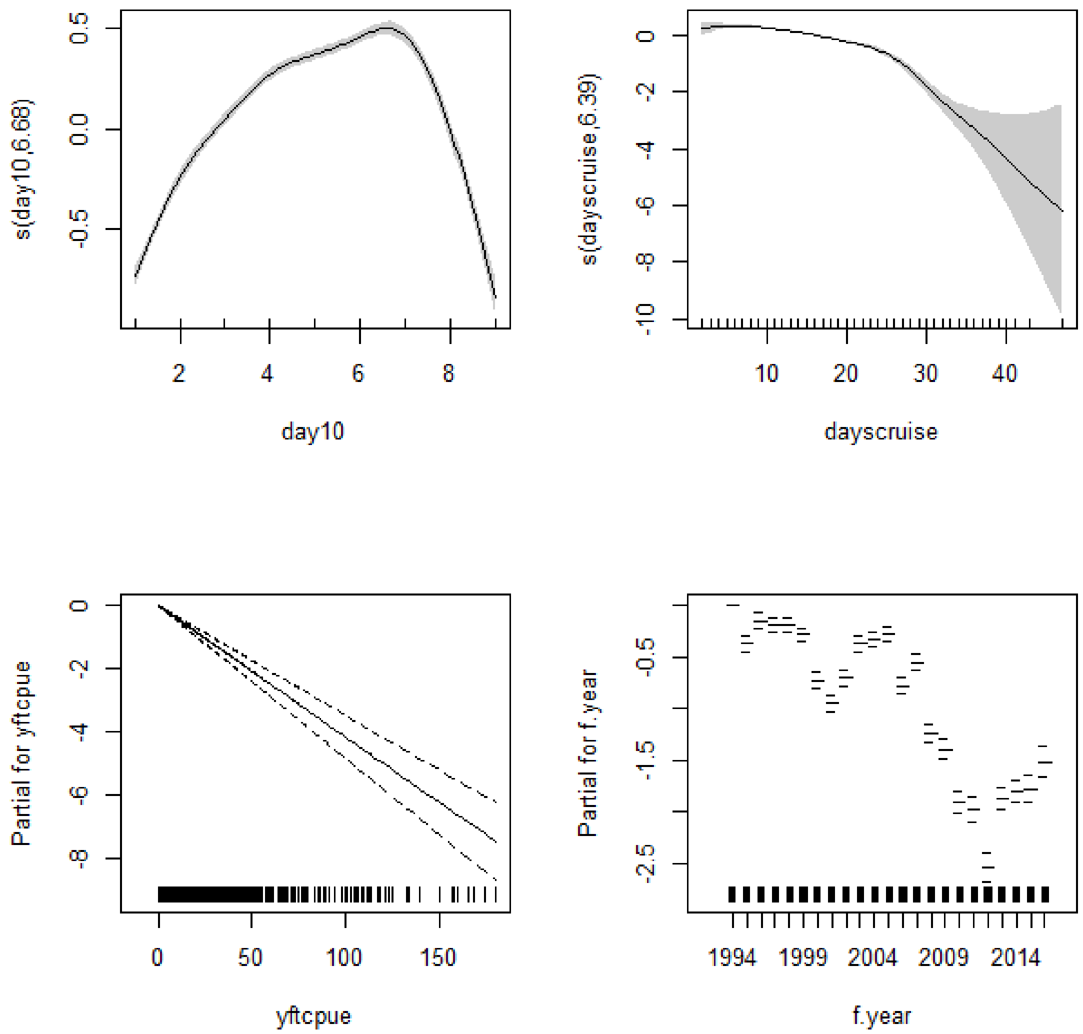

Fig. 1 Estimated main effects in final model.

(Note that “f.year” doesn’t mean fishery year but calendar year as categorical variable)

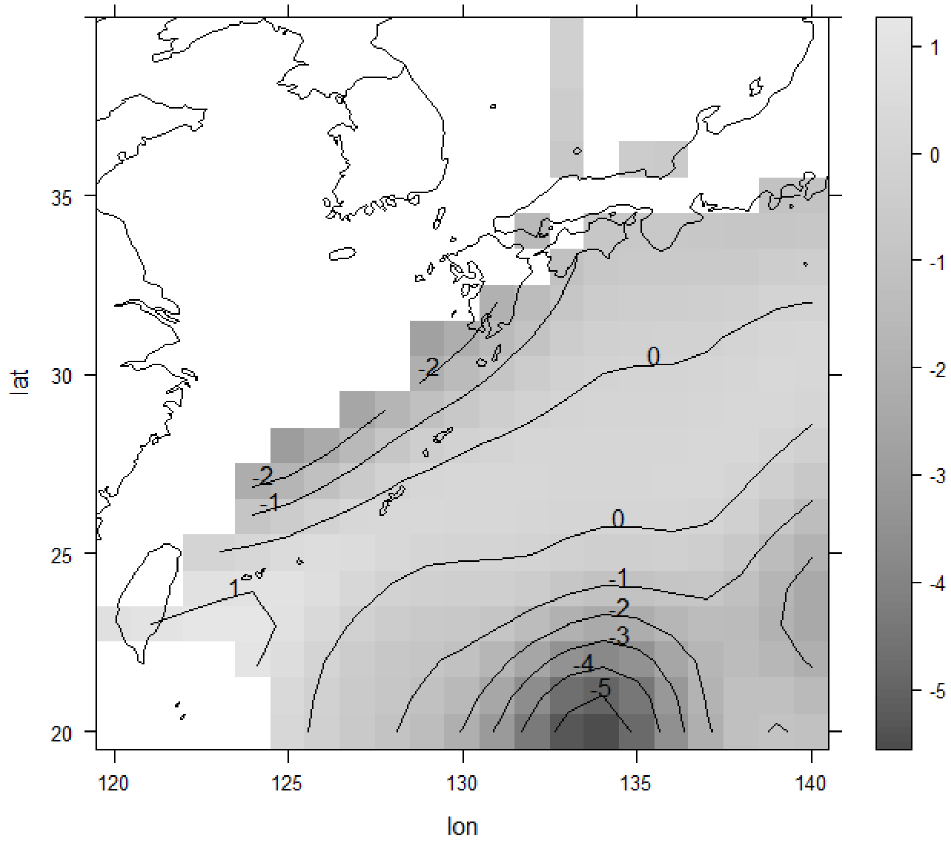

Fig. 2 Geographical effects in final model estimated by interaction term between longitude and latitude.

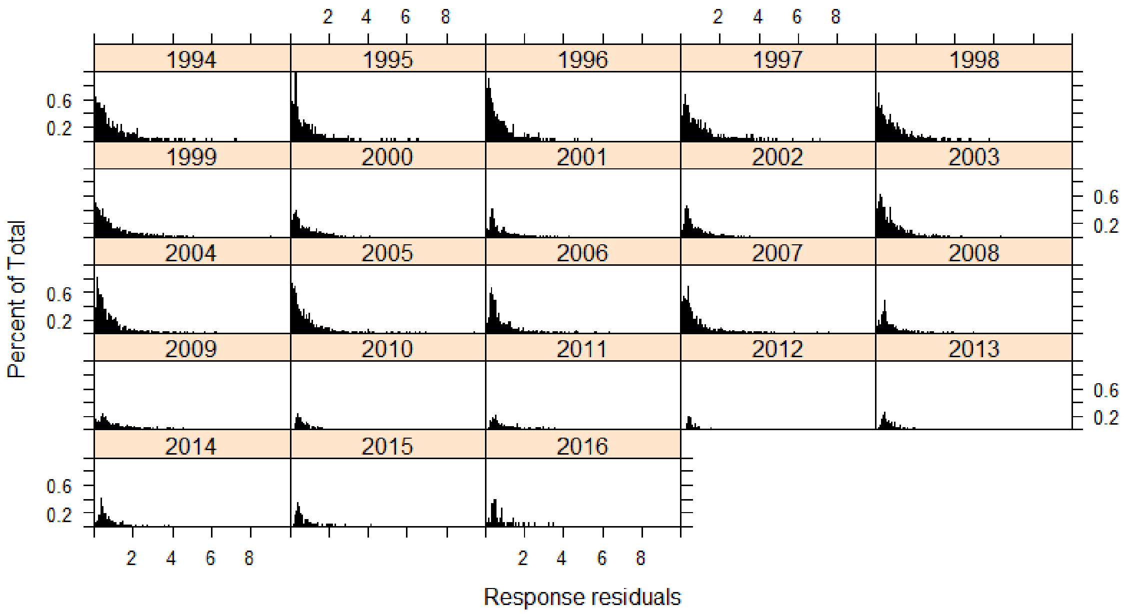

Fig. 3 The response residuals in final model by year.

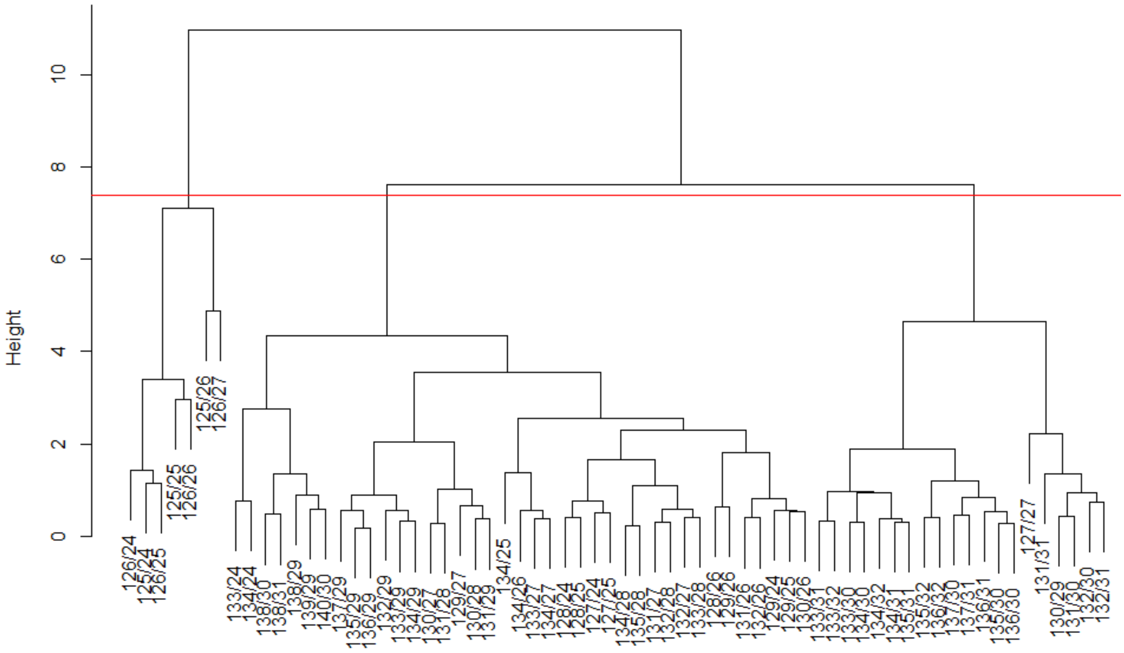

Fig. 4 The dendrogram of cluster analysis.
(Red line means incision of dendrogram in this analysis.)

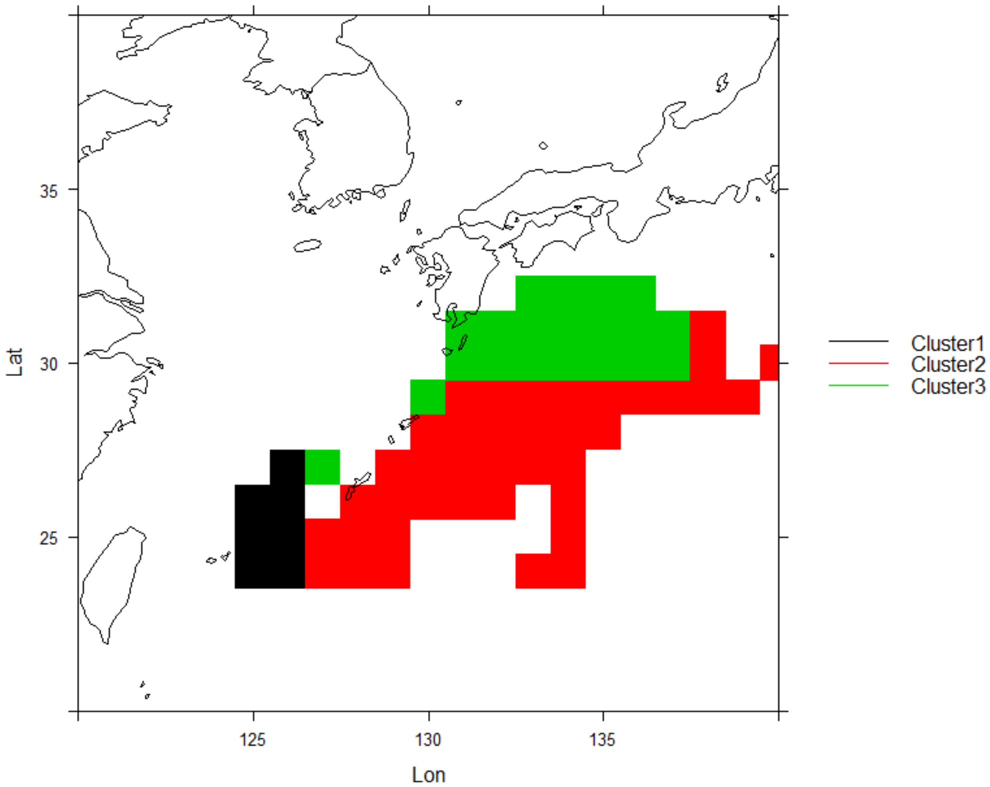

Fig. 5 The distribution of 1x1 cells classified by cluster analysis of year effects.

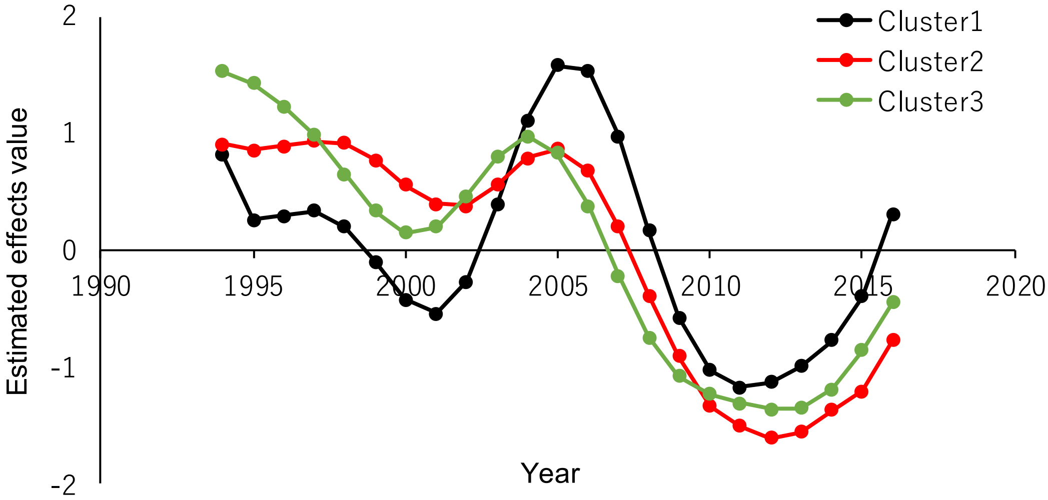

Fig. 6 Averaged year effects in each cluster.