Standardized CPUE for Pacific Bluefin tuna caught by Japanese coastal longliners by a zero-inflated negative binomial model using aggregated cruise data

Yuko Hiraoka[1], Yaoki Tei[1], Minoru Kanaiwa[2] and Kazuhiro Oshima[1]

[1] National Research Institute of Far Seas Fisheries, 5-7-1, Shimizu-orido, Shizuoka 424-8633, JAPAN

[2] Tokyo University of Agriculture, 196 Yasaka, Abashiri, Hokkaido 099-2403, JAPAN

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), 20–24 April 2015, Shimizu, Shizuoka, Japan. **Document not to be cited without author's permission.**

Abstract

To process data that contain many zero-value observations, set-by-set datasets for Pacific Bluefin Tuna caught by the Japanese coastal longline fishery were aggregated according to each cruise. A cluster analysis for determining species composition by cruise was conducted to identify the target species, especially Pacific Bluefin tuna and Yellowfin tuna, and the three clusters were used as the target indicator in this study. A negative binominal model (NB model) was applied to select the explanatory variables of the aggregated data, followed by a zero-inflated negative binomial model (ZINB model), which converged successfully owing to the aggregated data. The two area definitions were applied as sensitivity analysis. The target shift indicated that the annual CPUE trend of Model B2 (ZINB model with three areas) would be the best available index of the population trend for adult PBF.

Introduction

Catch and effort data collected from the logbooks of Japanese coastal longliners are an extremely important source of information for stock assessments of Pacific Bluefin tuna (PBF), given that this fishery has been targeting the spawning population of PBF and thus its standardized CPUE has been used as an index for estimating adult abundance. Standardized CPUE estimates of the latest stock assessment were calculated by Hiraoka et al. (2014), which updated the CPUE estimates of Ichinokawa and Takeuchi (2012). Although the standardized CPUE calculated by the procedure described in Ichinokawa and Takeuchi (2012) was generally accepted to be an accurate index of abundance for the stock assessment conducted in 2012, concerns have been subsequently raised about the shift in target species (ISC 2013).

Oshima et al. (2012) described both the target shift of this fishery, from PBF to Yellowfin tuna (YFT), and the shift in the effort distribution over recent years. They noted that the PBF CPUE decline that was observed after 2005 would be overrepresented by the shift in target species and thus the uncertainty of the standardized CPUE has increased. As a result, the observation and process errors in the abundance index associated with this fishery are believed to have changed. Consequently, a relatively higher coefficient of variation was incorporated into the stock assessment model after 2005 (ISC, 2014).

The logbook data of Japanese coastal longliners included a very large number of zero-catch records of PBF (Fig. 1), especially in recent years, which might be due to a decrease in the PBF spawning stock biomass (ISC, 2014), as well as modifications in fishing strategy (Oshima et al., 2012). Because of the presence of so many zero-catches, the delta-type two-step method (Lo et al., 1992) was previously used to estimate the standardization of CPUE for this fishery (Ichinokawa and Takeuchi, 2012). More recently the zero-inflated negative binomial model has been applied to process data containing numerous zero-value observations (Minami et al., 2007; Brodziak and Walsh, 2013). For the data regarding PBF caught by this fishery, however, we attempted to apply the zero-inflated model but it failed to converge; the data were therefore aggregated by cruise to reduce the zero-value recordings and to increase the catch count per observation (Fig. 1). Aggregated data has previously been used to estimate the standardization of CPUE in other fisheries, such as the Southern Bluefin tuna fishery (Itoh and Takahashi, 2014).

Aggregated cruise data may enable fishing strategy to be identified more clearly than does set-by-set data.

Interviews with coastal longline fishermen revealed that they can specify their targets by cruise. In general, 12–14 operations are conducted per cruise during the PBF fishing season and the fishing strategy is not changed within a cruise.

The main purpose of this document is to develop a new standardization method for Japanese longline CPUE pertaining to PBF to fully account for the possible shift in target species over recent years. Aggregated cruise data were analyzed because target strategy is easily understandable and the zero-inflated model can be applied.

Materials and Methods

Data source and processing

The set-by-set data used in this study were compiled by the National Research Institute of Far Seas Fisheries (NRIFSF) from the logbooks of Japanese coastal longliners for the years 1994–2014. The data for estimating the standardized CPUE were filtered through the criteria described by Ichinokawa and Takeuchi (2012), which consisted of

- ✓ April to June (spawning season)
- ✓ 1x1 degree grids in latitude and longitude where at least one PBF per year has been caught

Cluster analysis

To classify fishing strategy by cruise, a hierarchical clustering using Ward's method (Ward, 1963) on Euclidean distance was performed on the aggregated cruise data. The proportion of each tuna species (PBF; YFT: yellowfin tuna; BET: bigeye tuna; ALB: albacore) or other fishes caught relative to total catch number per cruise was used.

Area definition

For estimation of standardized CPUE, the set-by-set data were selected to identify the fishing areas, defined as 1x1 degree blocks where at least one PBF per year has been caught in more than 10 years. This definition followed that set out in previous studies (Ichinokawa and Takeuchi, 2012; Hiraoka et al., 2014).

Two different types of areas were delineated (two and three areas) (Fig. 2A,B). The first corresponds with two areas used in previous studies (Ichinokawa and Takeuchi, 2012; Hiraoka et al., 2014), whereas the second was constructed for three areas as follows: the SW area was divided into the "SW" and "CORE" areas because the data regarding effort distribution classified as cluster 2 and cluster 3 have been aggregated for the region south of Okinawa Island since 2007. The "CORE" area was defined by Oshima et al. (2012), where the PBF CPUE was categorized into the highest class (Fig 2B). A pair of averages of latitude and longitude during a cruise was used as a representative point of operations to determine into which area each cruise was to be categorized.

Standardization of CPUE

The Explanatory variables used in this analysis were as follows:

- ✓ Year: 21 calendar years, from 1994 to 2014
- ✓ Day 10: Periods during the spawning season, from April to June, defined by 10 day intervals (last period

of May contained 11 days)

- ✓ Area: two regions (“NW” or “SW”; Fig. 1A) or three regions (“NW”, “SW”, “CORE”; Fig. 1B)
- ✓ Gear: “Shallow” (<16 hooks per basket) or “Deep” (≥16 hooks per basket)
- ✓ Ship-size: “Small” (<16 GRT) or “Large” (≥16 GRT)
- ✓ Days per trip: “Short” (<14 days) or “Long” (≥14 days)
- ✓ Length of move: three categories defined by combining the total moving distance per trip with the mean moving distance per day (“Not-moving”: both total and mean distance were zero; “Short-distance”: total distance <300 miles; “Long-distance”: total distance ≥ 300 miles)
- ✓ Cluster: three clusters derived from cluster analysis (Table 2, Fig. 2)

Because the annual CPUE trends were distinctly different, data from cruises longer than 28 trip days or that traveled more than 60 miles of mean moving distance per day were removed (0.91%, or 108 of 11,864 cruises).

GLMs with negative binomial (NB) and zero-inflated negative binomial (ZINB) distributions were constructed to calculate standardized CPUEs. To select the explanatory variables of the ZINB model, an NB model was applied to the aggregated cruise data with two or three areas, and the full model was defined as the main effects and the two-way interactions of all explanatory variables. The best model for NB was determined by stepwise model selection using the “stepAIC” function of the “MASS” R package with BIC. The full model for ZINB was then defined as the NB best model, and the best model for ZINB was determined manually via the following steps by comparison with BIC values.

(1) The initial model with only the main effect was constructed.

(2) The main effect was determined through the reduction of each variable by the main effect model for count and zero-inflation models.

(3) The interactions of the best NB model were added individually to the main effect model selected in step (2) for count and zero-inflation models.

(4) The model with the lowest BIC value was determined to be the best model for ZINB.

The GLM calculations were made using R3.0.3, the “MASS” packages and the “pscl” packages.

To reference the scaled CPUE against each factor, the least squared means (LSMEANS) of each independent effect and the two interactions were calculated from the best model using the same estimation procedure as the SAS package (Shono et al., 2002). The annual CPUE trends for Model B1 and Model B2 were estimated as area-weighting values (Shono et al., 2002) because the interaction between Year and Area was retained for those models.

Results

Cluster analysis

The cluster analysis identified three groups according to the percentage of target species (Table 2; Fig. 3). ALB and YFT were dominant in Cluster 1 (86.5%) and Cluster 3 (75.0%), respectively (Table 2). In Cluster 2, the highest proportion was composed of non-target species (45.7%), whereas a relatively high proportion of PBF was observed (7.7%) compared to the other clusters (1.1–5.6%).

The number of cruises for Cluster 3 initially increased rapidly, from 66 in 2007 to 301 in 2011, then decreased to 90 in 2014 (Fig. 4). In contrast, those of Cluster 2 showed a relatively stable trend. Cluster 1 showed the highest values (428) in 2000, followed by a gradual decrease.

The geographical distributions of fishing effort, PBF catch, PBF CPUE and YFT catch are shown in Figs. 5–8. Different distribution characteristics were observed among the clusters. There were no notable changes in the distribution of fishing effort for Cluster 1 from 1994 to 2014. On the other hand, the relatively high values of PBF catch and CPUE previously recorded for southern Honshu Island vanished in recent years. The distributions of fishing effort for Clusters 2 and 3 displayed a similar trend, with an aggregation south of Okinawa Island since 2007; however, effort for Cluster 3 stretched north–south in the fishing ground (Fig. 5). This region coincided with a high YFT catch region in both Cluster 2 and Cluster 3 (Fig. 8). For those clusters, high PBF catches have been distributed only in the seas off of southwestern Okinawa Island since 2006, which has been identified as a "CORE" area (Fig 2B).

Standardization of CPUE

As a result of explanatory variable selection, for the two area models (Model A1 and Model A2), the interaction of Year*Cluster was retained for the best model, whereas Year*Area was retained for the three area models (Model B1 and B2; Table 3). The same combination of explanatory variables remained in the zero inflation models for both Model A2 and Model B2 but there were differences in the selected variables for the two count models. The count model for Model A2 retained the main effects and the interactions of Year*Cluster, Area*Ship-size, and Area*Cluster, whereas the selected interaction for Model B2 replaced Year*Cluster with Year*Area and added Day10*Area. In addition, the main effect of Gear was dropped for the ZINB models for both Model A2 and Model B2.

The information criteria (BIC and AIC) indicated that the NB models (Model A1 and Model B1) had lower values than the ZINB models (Model A2 and Model B2; Table 4), but both NB models displayed overdispersion (ϕ = 2.28 and 2.58 for Model A1 and Model B1, respectively). The scaled values of standardized CPUEs were relatively similar between Model A1 and Model A2, as well as between Model B1 and Model B2 (Table 5).

Fig. 9 shows nominal CPUE and the standardized CPUEs of Model B2, Model A2 and the procedure applied in a previous study (Ichinokawa and Takeuchi 2012). Peaks of CPUE in the late 1990s occurred in different years among CPUEs. The standardized and nominal CPUE's of Model B2 peaked in 1996, but the CPUE of Model A2 and the CPUE based on the previous study peaked in 1997. All of the CPUEs hit a trough in 2001. The smallest difference in CPUE between the peak in the late 1990s and the trough in 2001 was recorded in Model B2. In 2007, when the second peak of CPUE occurred after 2000, the CPUE of Model B2 was the smallest.

Least squared means of all effects for Model A2 and Model B2 are shown in Figs. 10–11. For Model A2, the cluster effect varied by year, and the scaled CPUE for Cluster 3 showed the highest CPUE value (2.79) in 1997, which is different from those of Clusters 1 and 2 (Fig. 10A). The effect of Fishing season (Day 10) suggests a dome-shaped trend. The high CPUE continued from late May to early June (Fig. 10B). Compared to the influence of the Year*Cluster and Fishing season effects, those of the other effects (Area*Ship-size, Area*Cluster, Days per cluster and Distance of move) were relatively low (0.46–1.34; Fig. 10). For Model B2, the effects of Year, Fishing season, Ship size and Cluster differed by Area (Fig. 11), most notably with respect to the annual trend (Fig. 11A).

The CPUE of the "CORE" area indicated a relatively different trend from the "NE" and "SW" areas, whereas the trends of these two areas were similar to those of Model A2 (Figs. 10–11). The Pearson residual patterns were not distinctly different between Model A2 and Model B2 (Fig. 12).

Discussion

The aggregated cruise data enables a cluster analysis to be used to classify each cruise according to fishing strategy. The primary goal of using a cluster analysis was to divide the data into PBF and YFT targets, to examine the shift in species target from PBF to YFT that has occurred in recent years, which was reported by Oshima et al. (2012). The species composition of Cluster 3 was dominated by YFT, and the number of cruises for this cluster rapidly increased from 2007, with the fishing efforts of this cluster aggregated in the region with high YFT catch since 2007 (Figs. 3–4, 5, 8). These results indicate that Cluster 3 could be identified as the YFT target cruise. On the other hand, the species composition of Cluster 2 demonstrated that the relatively high PBF ratio and catches of this cluster have concentrated in the "CORE" area defined as a high PBF CPUE region (Oshima et al., 2012) since 2007. Thus, Cluster 2 could be identified as the PBF target cruise. Consequently, the cluster analysis effectively divided the data according to fishing strategy.

The aggregated cruise data were analyzed based on the assumption that fishers would not change their target strategy during a cruise. Although it is difficult to validate this assumption properly, supporting information was provided by interviews with fishermen. Furthermore, the species composition by set tend to record the zero-value observations even if the fisherman targeted PBF because of the extremely low catch rate per set. To take this into account, the species composition by cruise would be appropriate to use to identify the PBF target observation.

The CPUE trend of Model B2 could be an effective index of abundance for adult PBF. As noted above, cluster analysis can be used to identify target strategy. Because the models in this study included the cluster as a target indicator, the annual CPUE trend of these models should reflect a more realistic population trend than do previous models. In addition to target indicator, the new area definition was applied to Model B2. The addition of the "CORE" area would also effectively reflect the target shift. Because the effort distribution of Cluster 2 has been concentrated in "CORE" areas since 2007 and the Year*Cluster effect of Model A2 replaced Year*Area effect through the change of area definition for Model B2, the shift in targeted species could be explained by Area effect instead of Cluster effect. Although the lowest BIC was provided by Model B1, it was thought to be due to overdispersion ($\phi = 2.58$). Moreover, Minami et al. (2007) concluded that the NB model may overestimate model coefficients when using data that contains many zero-value observations. In conclusion, we suggest that the annual CPUE trend estimated by Model B2 would be the most appropriate index for determining the current population trend of PBF.

References

Brodziak, J. and Walsh, W.A. 2013. Model selection and multimodel inference for standardizing catch rates of bycatch species: a case study of oceanic whitetip shark in the Hawaii-based longline fishery. Can. J. Fish. Aquat. Sci. 70, 1723-1740.

Ichinokawa, M. and Takeuchi, Y. 2012. Standardized CPUE of North Pacific Bluefin tuna caught by Japanese coastal longliners: updates until 2011. ISC/PBFWG/12-1/8.

ISC, 2013. ISC Stock assessment of Bluefin Tuna in 2012. ISC/13/PLENAEY/INFO/12.

ISC, 2014. ISC Stock assessment of Pacific Bluefin Tuna in the Pacific Ocean in 2014. ISC/14/PLENAEY/INFO/18.

Itoh, T. and Takahashi, N. 2014. Update of the core vessel data and CPUE for southern bluefin tuna in 2014. CCSBT-OMMP/1406/13.

Lo, N.C.L.D. Jacobson, L.D. and Squire J.L. 1992. Indices of relative abundance from fish spotter data based on Delta-Lognormal models. Can. J. Fish Aquat. Sci. 49, 2515-2526.

Oshima, K., Mizuno, A., Ichinokawa, M., Takeuchi, Y., Nakano, H. and Uozumi, Y. 2012. Shift of fishing efforts for Pacific bluefin tuna and target shift occurred in Japanese coastal longliners in recent years. ISC/PBFWG/12-3/5.

Shono, H. Okamoto, H. Nishida, T. 2002. Standardized CPUE for yellowfin tuna (*Thunnus albacares*) of the Japanese Longline Fishery in the Indian Ocean by Generalized Linear Models (GLM) (1960–2000). IOTC Proceedings 5, 240-247.

Ward, J.H. 1963. Hierarchic grouping to optimize an objective function. J. Amer. Statistic. Assoc. 58, 236-244.

Table 1 Total number of cruises, hooks and PBF catch, and nominal CPUE, recorded in coastal longline logbooks for cluster analysis (1x1 degree grids in latitude and longitude where at least one PBF per year have been caught) and for standardization of CPUE selected as fishing area.

Calendar year	Fishing year(*1)	Data used for cluster analysis				Data used for standardization of CPUE selected as the fishing area			
		N of Cruise	N of hooks (x1000 hooks)	N of PBF catch	Nominal CPUE	N of Cruise	N of hooks (x1000 hooks)	N of PBF catch	Nominal CPUE
1994	1993	440	5734	2921	0.509	387	5273	2902	0.550
1995	1994	395	5095	1676	0.329	344	4582	1642	0.358
1996	1995	427	5704	2592	0.454	372	5051	2517	0.498
1997	1996	473	6130	2593	0.423	389	5185	2427	0.468
1998	1997	521	7041	3077	0.437	435	6038	2907	0.481
1999	1998	806	10706	4100	0.383	711	9119	3822	0.419
2000	1999	819	10857	2460	0.227	635	8097	2208	0.273
2001	2000	781	11856	1917	0.162	620	9306	1864	0.200
2002	2001	800	11636	2242	0.193	595	8418	1961	0.233
2003	2002	882	12434	3007	0.242	587	7128	2460	0.345
2004	2003	899	11999	3814	0.318	698	9046	3394	0.375
2005	2004	848	12703	4318	0.340	598	8339	3686	0.442
2006	2005	915	12953	2345	0.181	619	7858	1881	0.239
2007	2006	887	12044	3517	0.292	600	7664	3105	0.405
2008	2007	860	12089	1882	0.156	601	8160	1471	0.180
2009	2008	928	13438	1501	0.112	617	8339	1265	0.152
2010	2009	970	13958	814	0.058	661	8631	783	0.091
2011	2010	917	12247	773	0.063	671	7810	583	0.075
2012	2011	843	12626	508	0.040	609	8084	404	0.050
2013	2012	835	12411	888	0.072	581	7987	707	0.089
2014	2013	661	10506	808	0.077	426	5919	593	0.100

Table 2 Species composition (%) and number of cruises by each cluster estimated by the data for standardization of CPUE.

	Cluster		
	1	2	3
Yellowfin tuna	3.3	25.1	75.0
Albacore	86.5	19.9	9.6
Bigeye tuna	5.6	1.6	1.1
Pacific bluefin	1.0	7.7	2.3
Other species	3.6	45.7	11.9
N of cruise	2,014,209	249,276	135,272

Table 3 Explanatory variables of each model with the lowest BIC value.

			Explanatory variables
Two area model	Model A1: NB best model		year, day10, area, gear, shipsize, trip (days per trip), move (length of move), cluster, year*cluster, day10*cluster, area*shipsize, area*cluster
	Model A2: ZINB best model		
		Count model	year, day10, area, shipsize, trip, move, cluster, year*cluster, area*shipsize, area*cluster
		Zero inflation model	year, day10, area, shipsize, trip, move, cluster, area*shipsize
Three area model	Model B1: NB best model		year, day10, area, gear, shipsize, trip, move, cluster, year*area, day10*area, area*shipsize, area*cluster
	Model B2: ZINB best model		
		Count model	year, day10, area, shipsize, trip, move, cluster, year*area, day10*area, area*shipsize, area*cluster
		Zero inflation model	year, day10, area, shipsize, trip, move, cluster, area*shipsize

Table 4 Comparison of BIC, AIC, log-likelihood and degree of freedom by model.

	BIC	AIC	Loglik	DF
Model A1	46438	45723	-22764	97
Model A2	48186	47323	-23545	117
Model B1	45777	45032	-22415	101
Model B2	47503	46478	-23100	139

Table 5 Scaled values of nominal and standardized CPUEs estimated by each model.

Calender year	Fishing year	nominal CPUE	Updated*	Model A1	Model A2	Model B1	Model B2
1994	1993	1.922	1.997	1.773	1.649	1.933	1.897
1995	1994	1.253	1.444	1.199	1.137	1.301	1.270
1996	1995	1.767	1.805	1.658	1.476	1.947	1.836
1997	1996	1.642	1.901	1.960	1.913	1.702	1.655
1998	1997	1.689	1.649	1.976	1.670	1.645	1.595
1999	1998	1.454	1.176	1.582	1.575	1.342	1.380
2000	1999	0.957	0.912	0.888	0.856	1.106	1.077
2001	2000	0.688	0.706	0.735	0.800	0.806	0.863
2002	2001	0.809	0.794	0.713	0.816	0.980	1.060
2003	2002	1.182	1.333	1.134	1.088	1.353	1.345
2004	2003	1.308	1.440	1.394	1.413	1.448	1.407
2005	2004	1.538	1.687	1.593	1.646	1.598	1.640
2006	2005	0.842	0.849	0.708	0.731	0.753	0.787
2007	2006	1.397	1.181	0.949	1.087	0.889	0.932
2008	2007	0.641	0.652	0.674	0.799	0.618	0.648
2009	2008	0.524	0.411	0.465	0.525	0.329	0.321
2010	2009	0.316	0.215	0.328	0.375	0.233	0.252
2011	2010	0.258	0.215	0.325	0.339	0.202	0.209
2012	2011	0.172	0.139	0.204	0.236	0.173	0.184
2013	2012	0.305	0.231	0.359	0.413	0.290	0.291
2014	2013	0.335	0.263	0.387	0.457	0.351	0.350

*:Standardized CPUE updated by Ichinokawa and Takeuchi (2012)

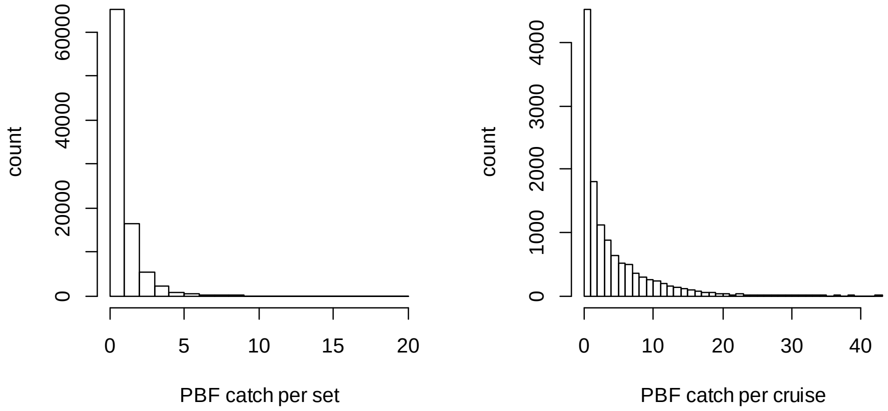


Fig. 1. Distribution of PBF catch number by set-by-set data (left) and aggregated cruise data (right).

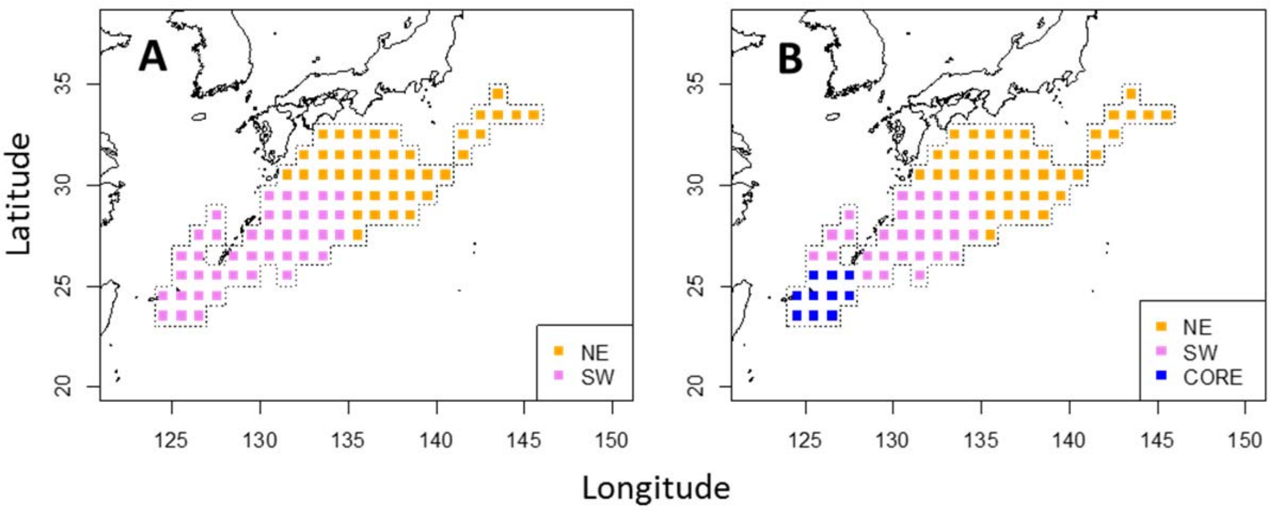


Fig. 2. The area surrounded by the dotted line represents the fishing area selected for standardization of CPUE derived from the logbooks of Japanese coastal longliners. Panel (A) shows the area defined by Ichinokawa and Takeuchi (2012) and panel (B) shows the area defined by this study.

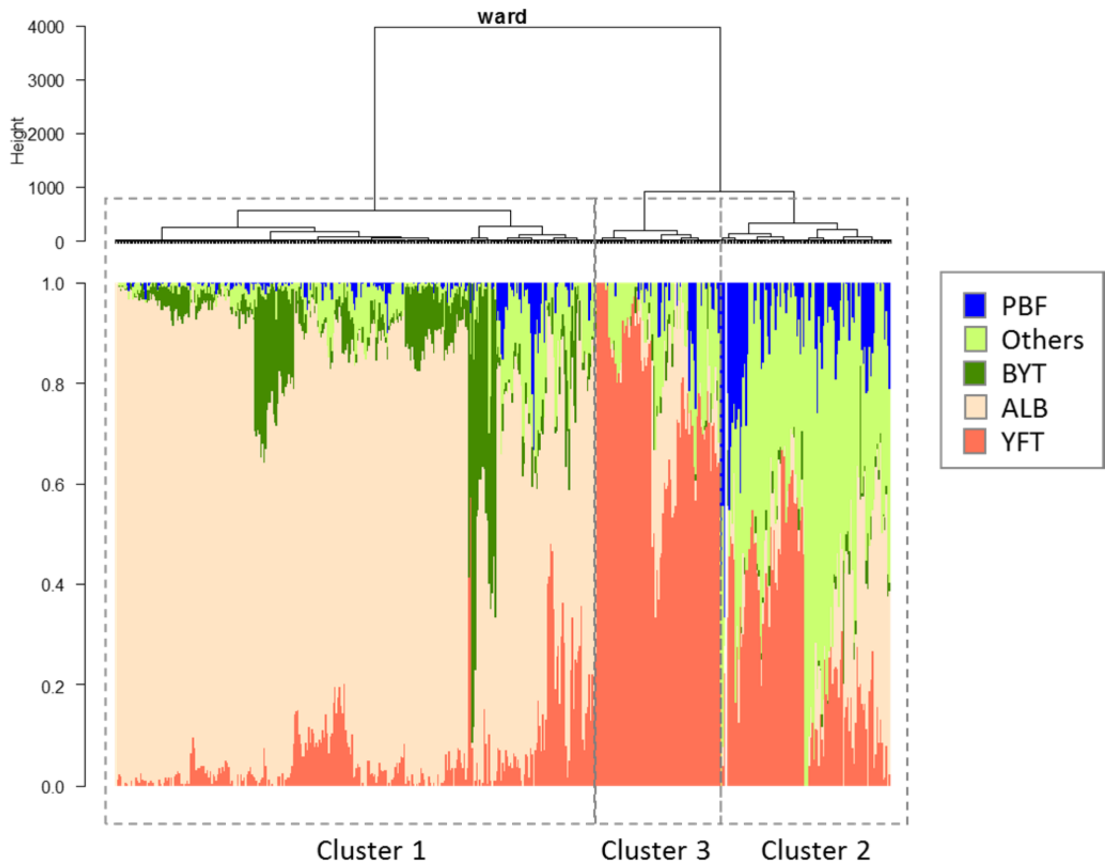


Fig. 3 Result of cluster analysis (Word's methods). The upper panel shows the dendrogram obtained by cluster analysis and the lower panel shows the species composition by cruise corresponding to each cluster.

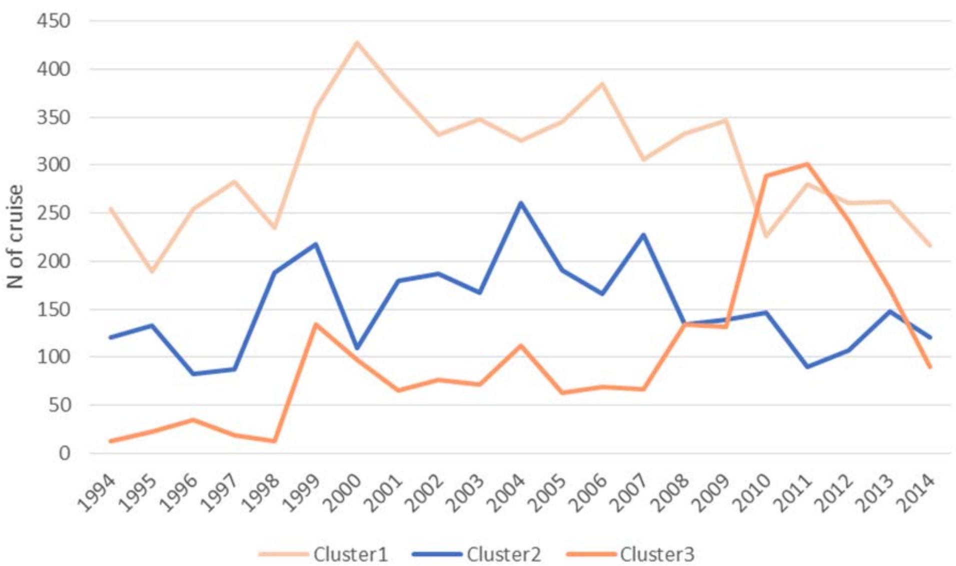


Fig. 4 Annual change of observation number (number of cruises) for the estimation of standardization of CPUE by cluster.

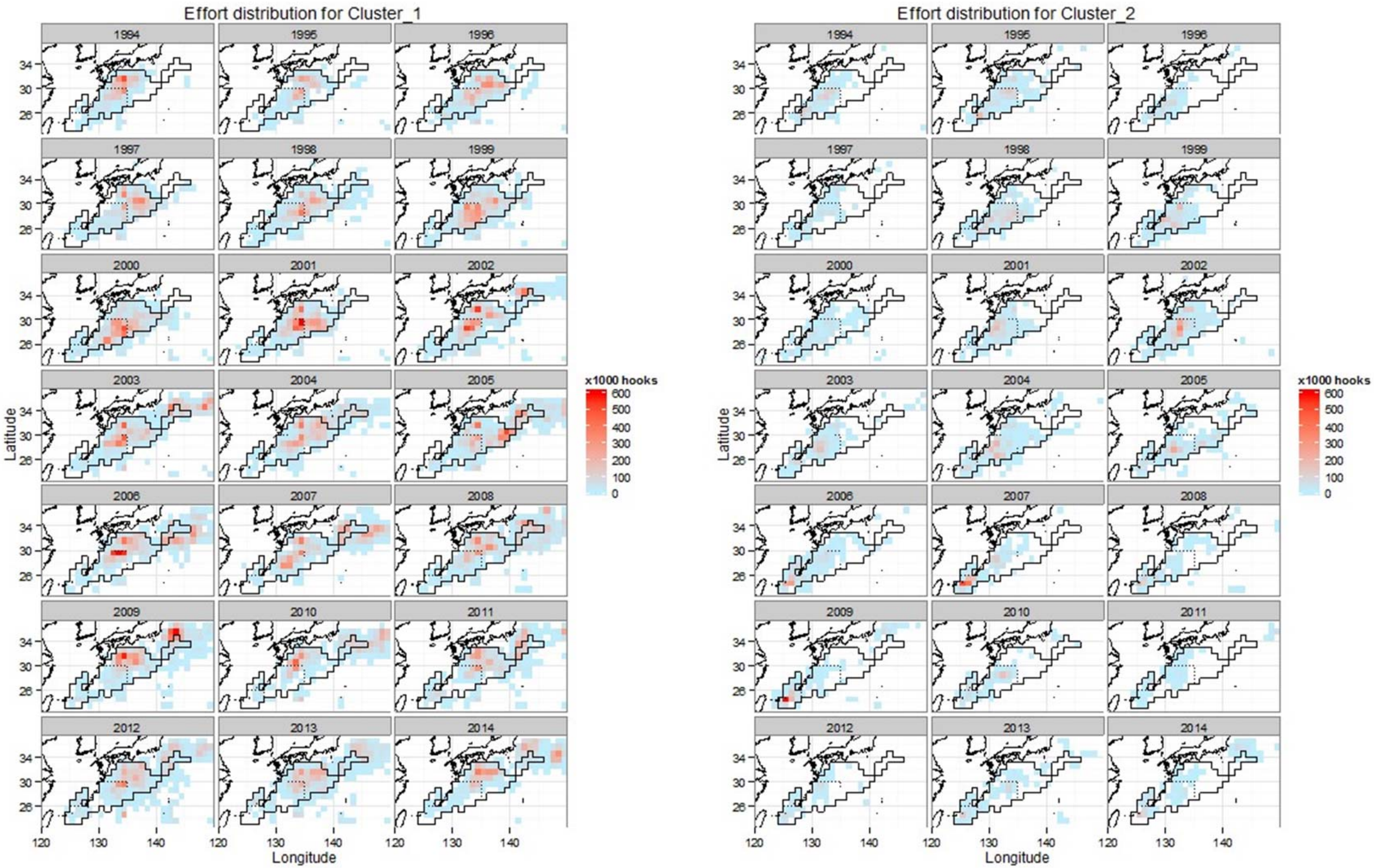


Fig 5. Geographical distribution of fishing effort from April to July by year and cluster. The red color indicates the grid where over 600,000 hooks were used.

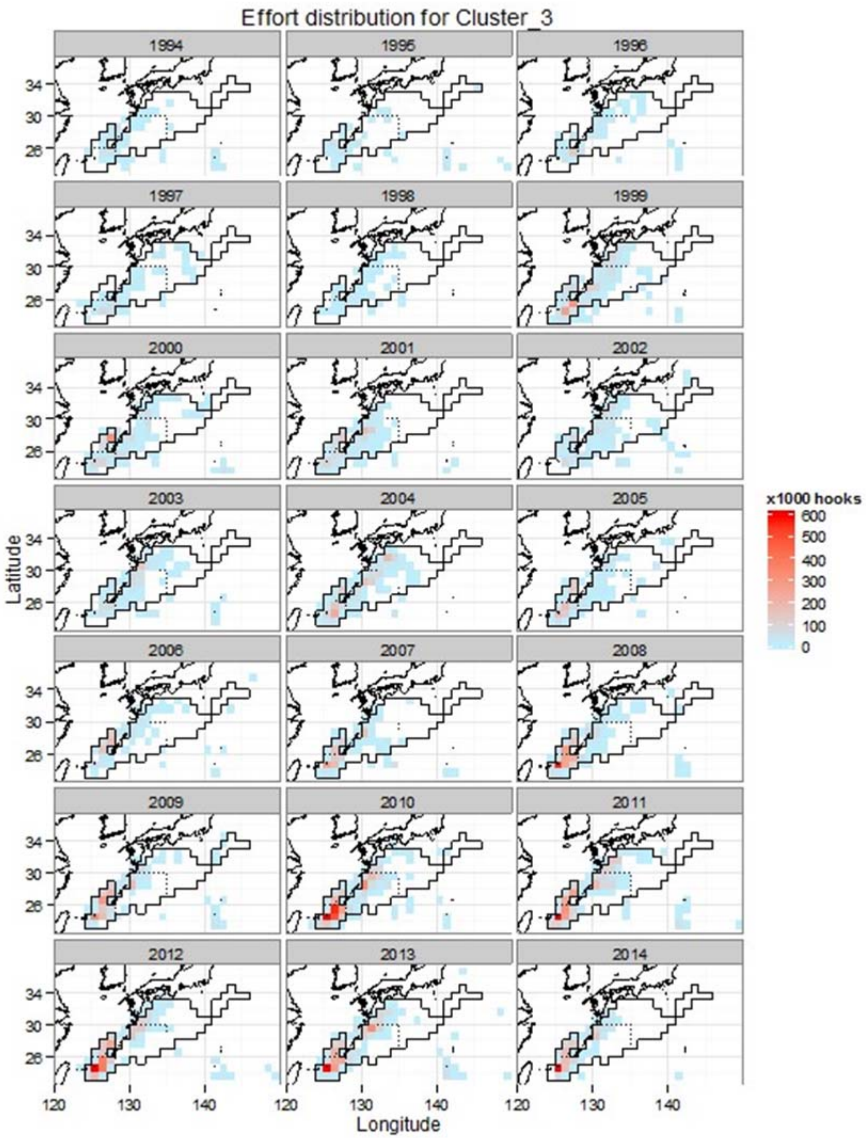


Fig 5. Continued.

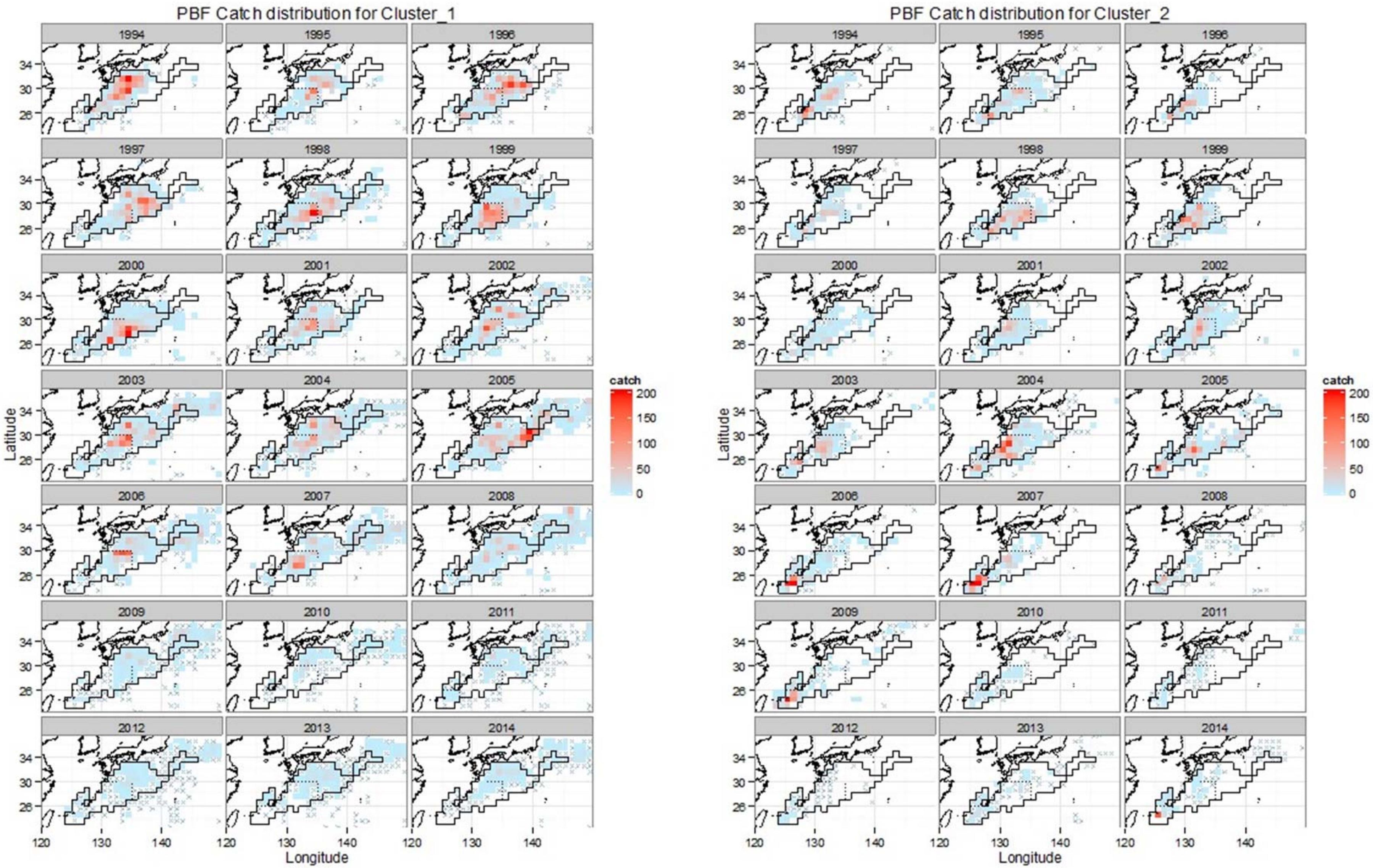


Fig 6. Geographical distribution of PBF catch from April to July by year and cluster. Red indicates grids where over 200 individuals were caught and the blue crosses indicate zero catch.

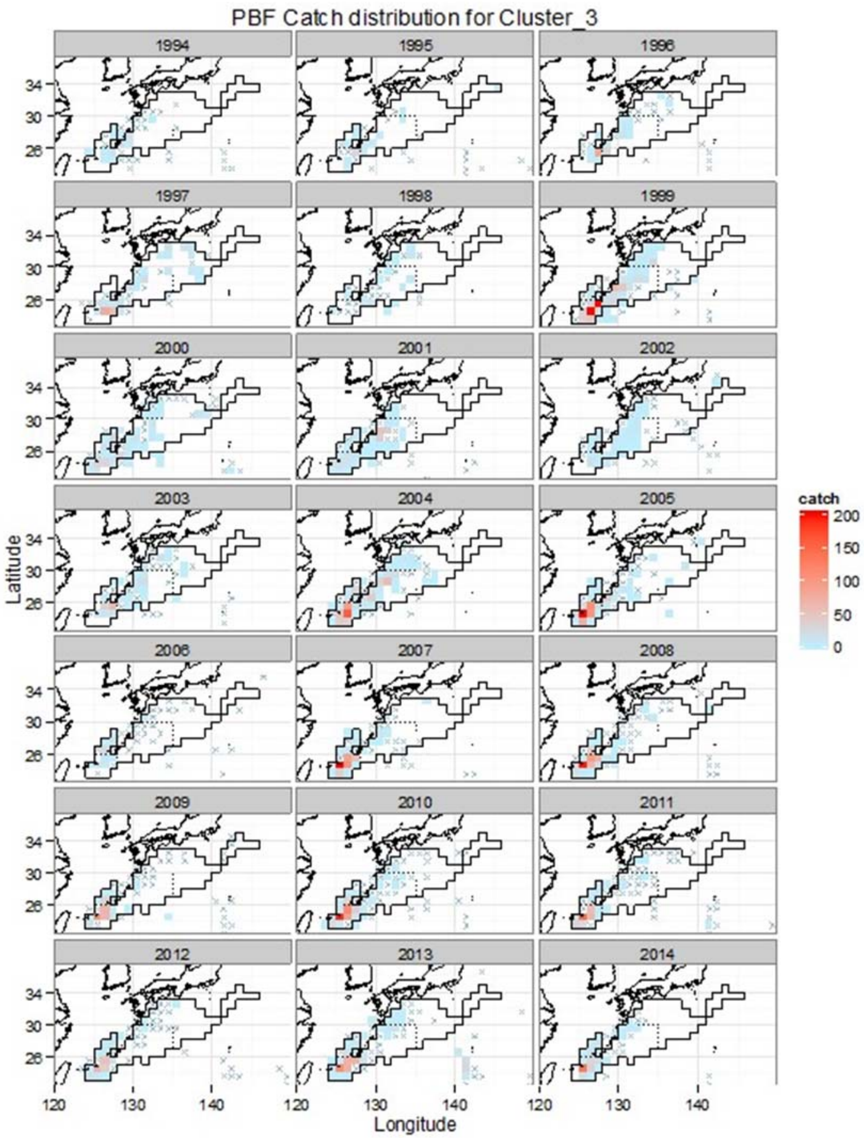


Fig 6. Continued.

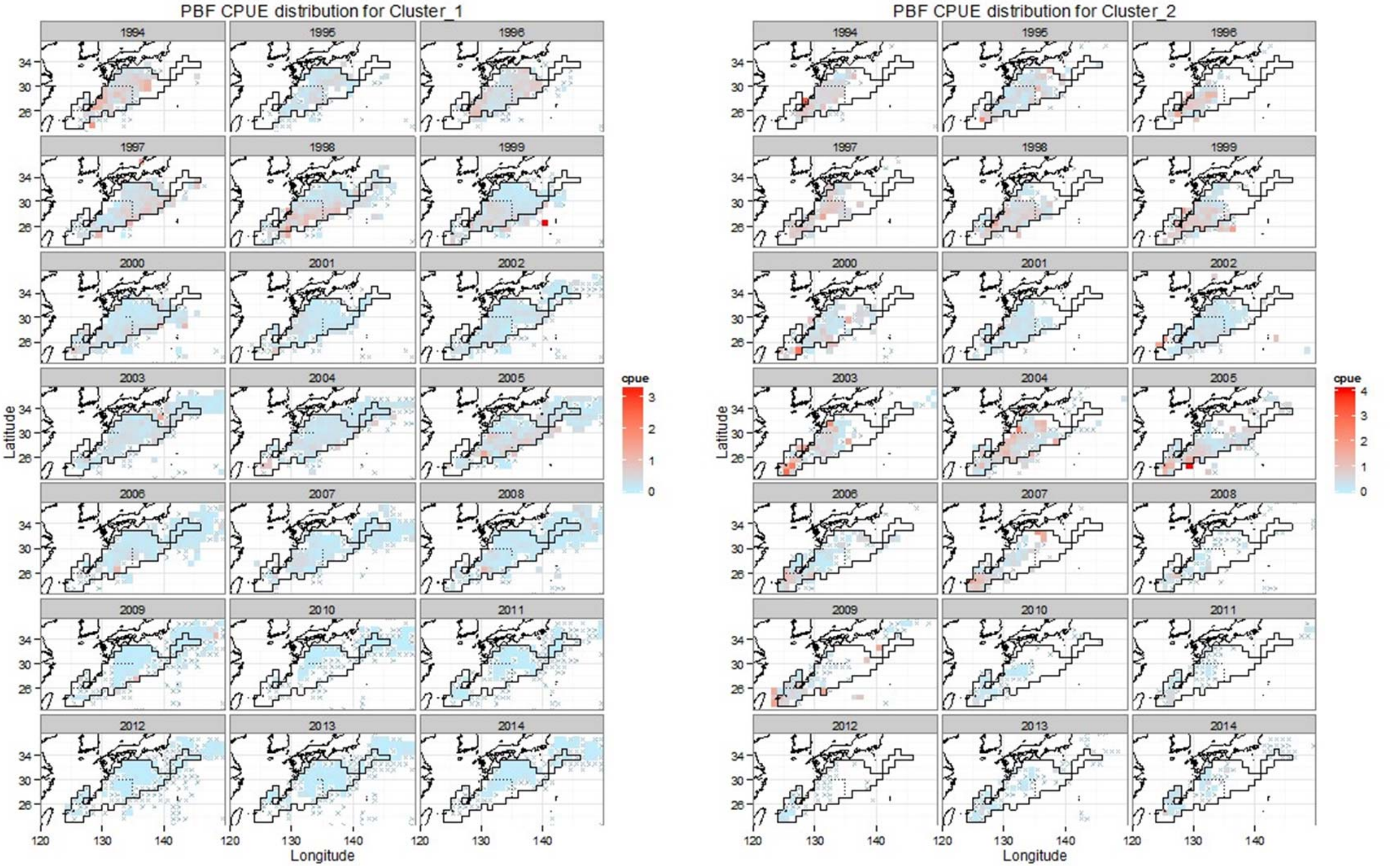


Fig 7. Geographical distribution of PBF catch from April to July by year and cluster. The blue crosses indicate zero catch.

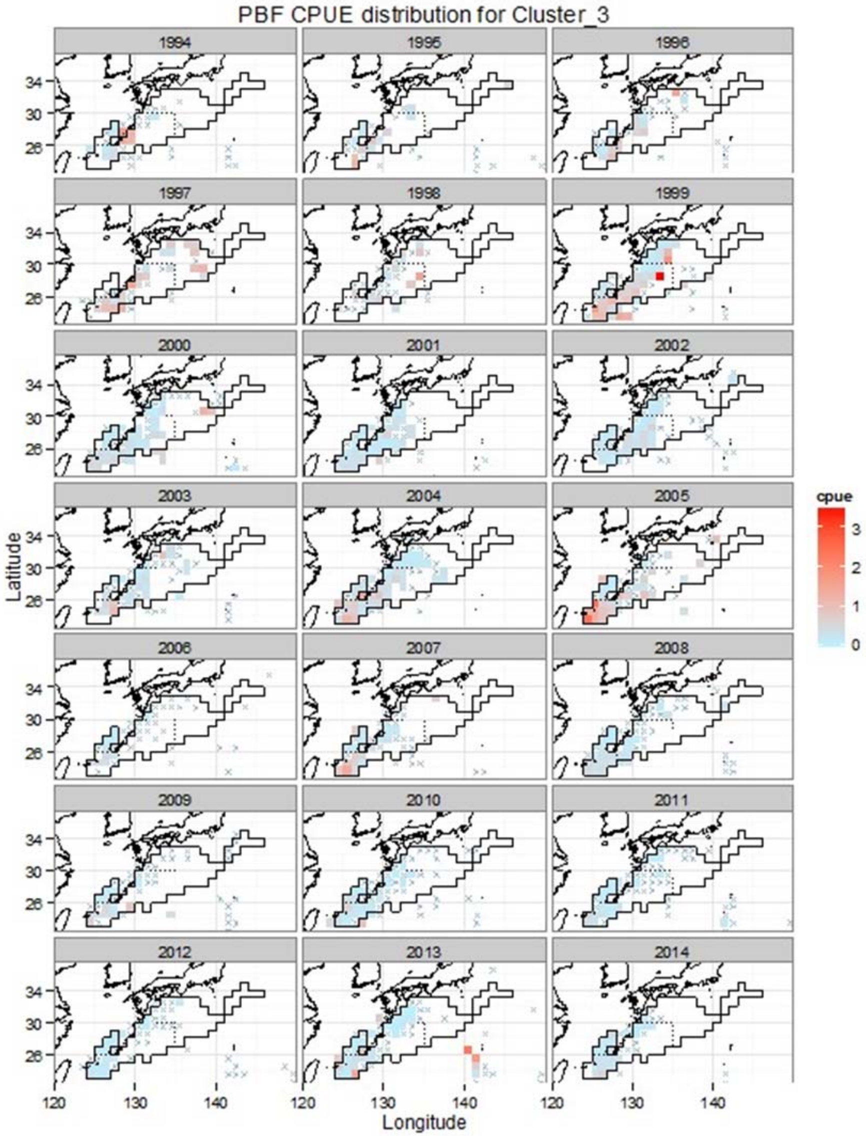


Fig 7. Continued.

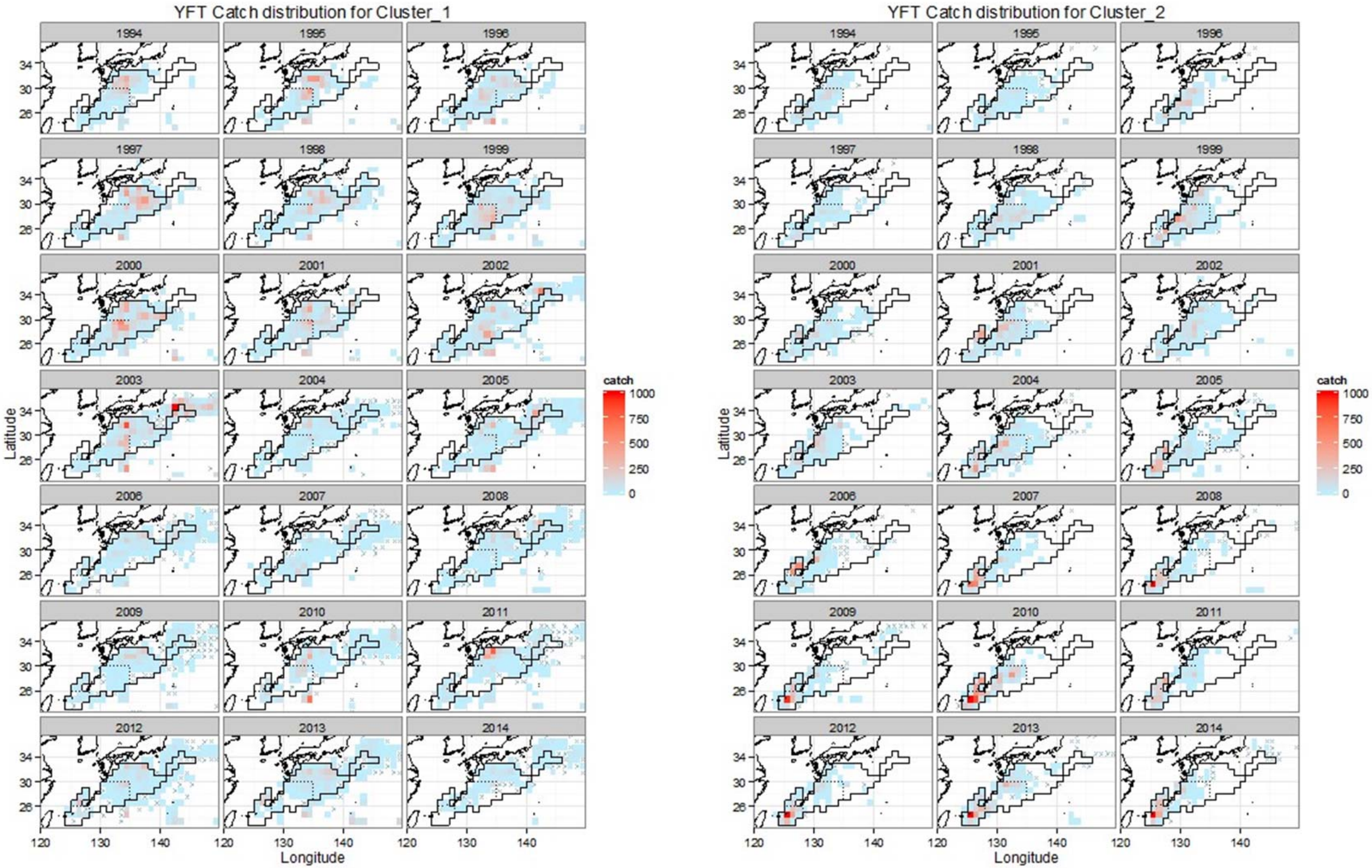


Fig 8. Geographical distribution of YFT catch from April to July by year and cluster. The blue crosses indicate zero catch.

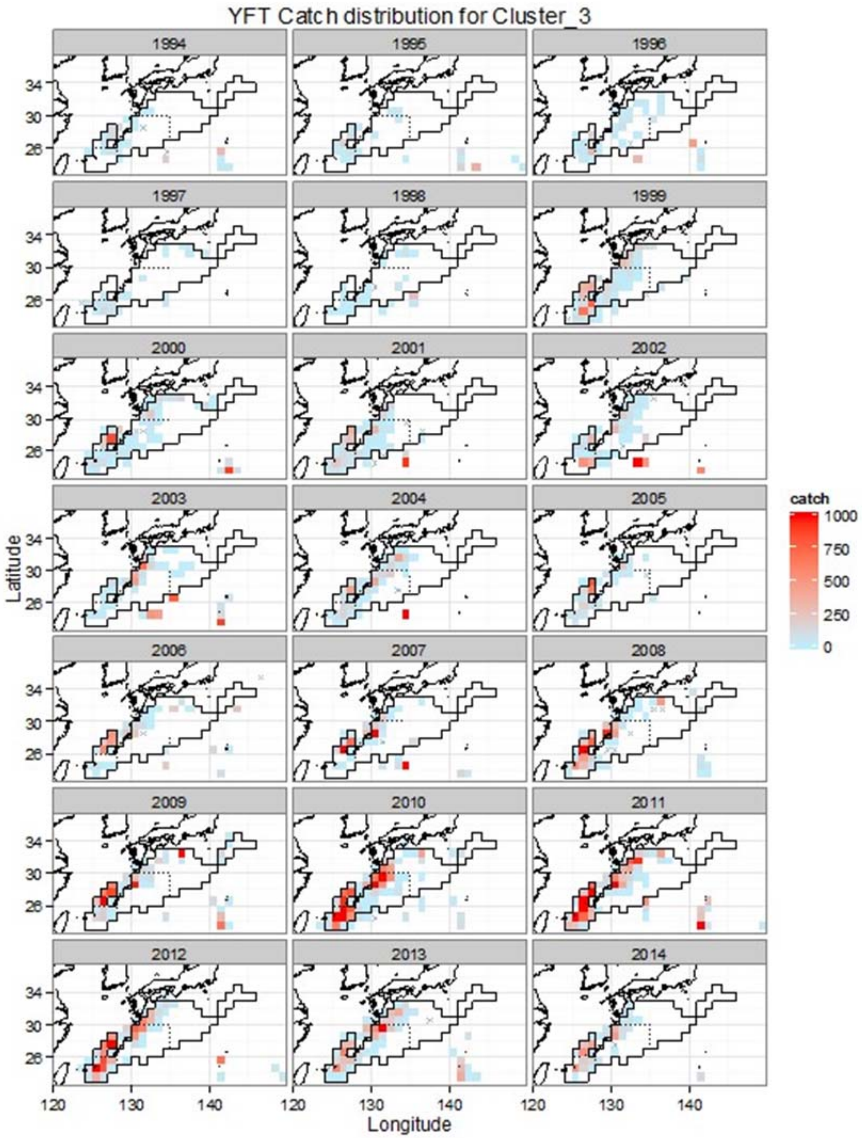


Fig 8. Continued.

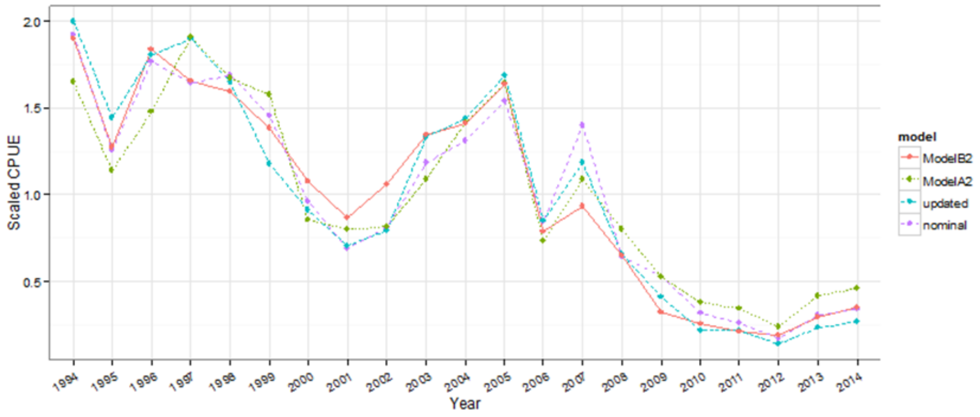


Fig 9. Comparison of scaled CPUEs among nominal CPUE, updated CPUE of previous study, Model A2, and Model B2.

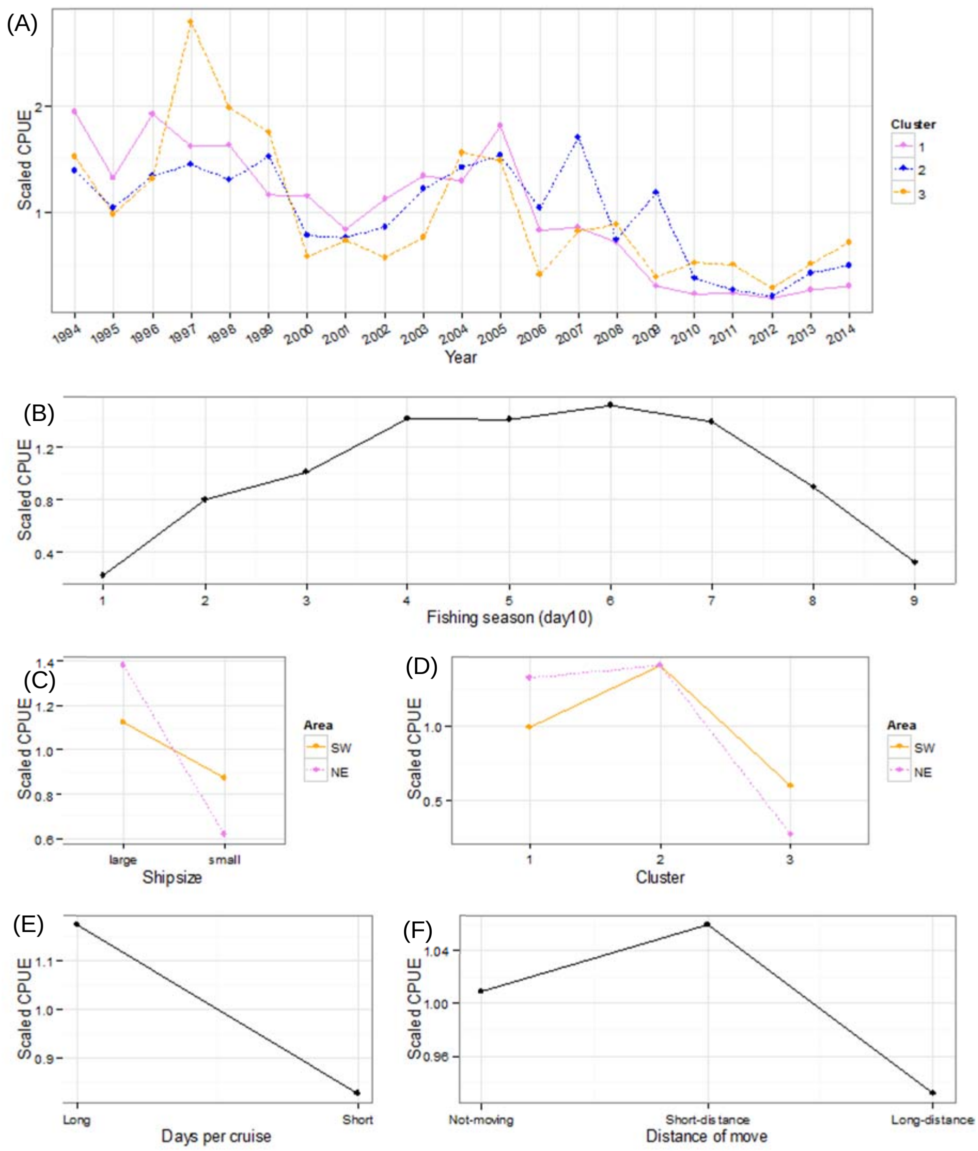


Fig 10. Least squared means of standardized CPUE estimated by Model A2.

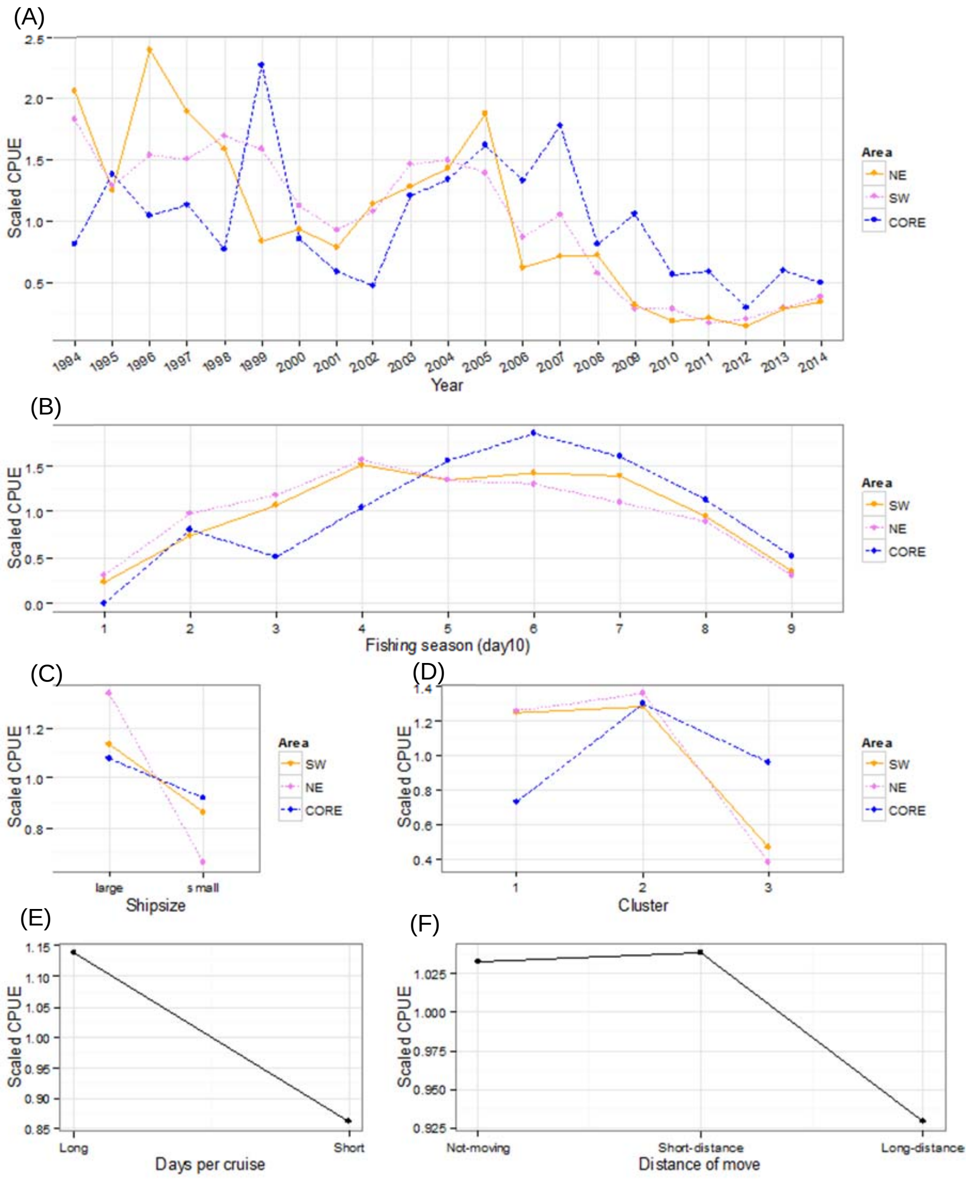
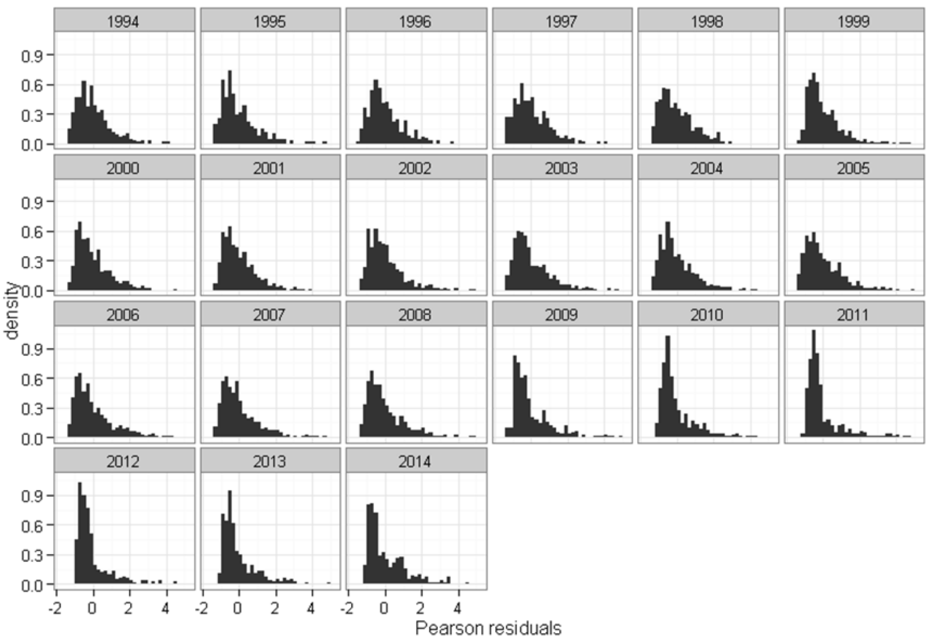


Fig 11. Least squared means of standardized CPUE estimated by Model B2.

(A)



(B)

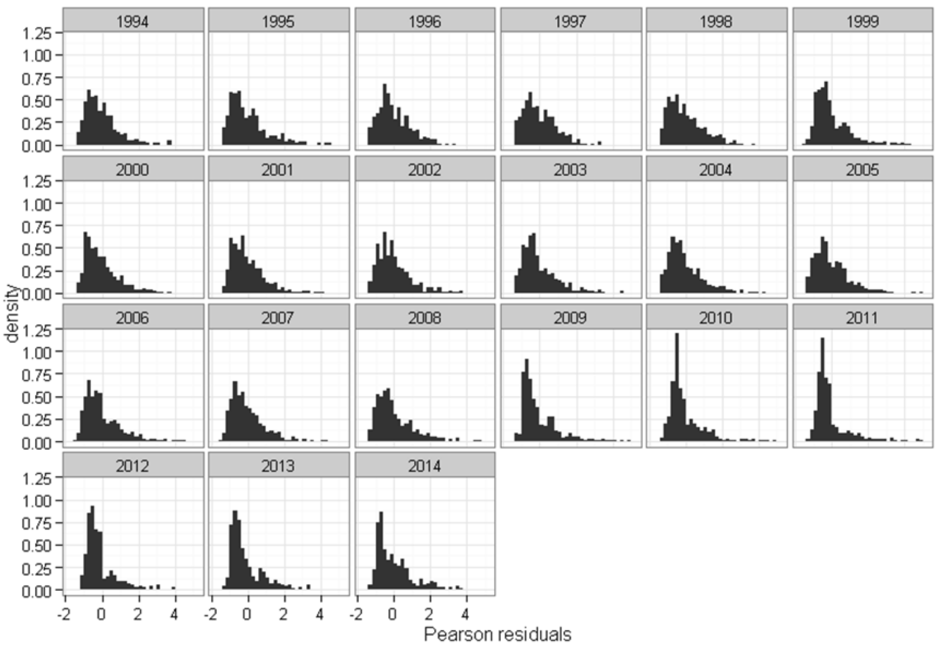


Fig 12. Residual distributions by year (A: Model A2; B: Model B2).