DATA EXPLORATION AND CPUE STANDARDIZATION FOR THE KOREAN SOUTHERN BLUEFIN TUNA LONGLINE FISHERY (1996-2022)

Jung-Hyun Lim, Youjung Kwon, Heewon Park, Haewon Lee and Sung Il Lee

National Institute of Fisheries Science (NIFS)

216 Gijanghaean-ro, Gijang-eup, Gijang-gun, Busan 46083, Republic of Korea

ABSTRACT

In this study we standardized southern bluefin tuna (*Thunnus maccoyii*, SBT) CPUE from Korean tuna longline fisheries (1996-2022) using Generalized Linear Models (GLM) with set by set (operational) data. The data used for the GLMs were catch (number), effort (number of hooks), number of hooks between floats (HBF), fishing location (5° cells), and vessel identifier by year, quarter, and area. We explored CPUE by area and identified two separate areas (CCSBT statistical area 8 and 9) in which Korean vessels have targeted SBT. SBT CPUE was standardized for each of these areas. We applied two alternative approaches, data selection and cluster analysis, to address concerns about target change over time that can affect CPUE indices. Explanatory variables for the GLM analyses were year, month, vessel identifier, location (5° cells), number of hooks, and targeting (HBF and cluster). GLM results for each area suggested that year, month, location, and targeting effects were the principal factors affecting the nominal CPUE. The standardized CPUEs for both areas decreased until the mid-2000s and have shown an increasing trend since that time.

INTRODUCTION

Developing indices of abundance using catch per unit effort data requires decisions based on the understanding of both the fishery and the population dynamics of the species. This is particularly the case in a multi-species fishery, in which targeting behaviours change seasonally, spatially, and from year to year. Such analyses require careful data exploration, and methods to differentiate fishing practices.

Southern bluefin tuna (*Thunnus maccoyii*, SBT) are the target of a high value international fishery, managed by the Commission for the Conservation of Southern Bluefin Tuna (CCSBT). The stock has been assessed as highly depleted, but since a low point in 2005 has shown signs of recovery (CCSBT 2017).

Korean tuna longline fisheries began targeting SBT in the CCSBT convention area in 1991 (Kim et al. 2015). SBT was reported as bycatch before this time, starting in 1972. Catch was initially low but increased to 1,320 mt in 1996, peaked at 1,796 mt in 1998, and thereafter decreased to below 200 mt in the mid-2000s. In 2008, the catch increased again to 1,134 mt and thereafter fluctuated in a range of 705-1,268 mt due to the national catch limit. The catch in 2022 was 1,173 mt (Fig. 1).

In developing the index, we compare two alternative methods for differentiating targeting practices in the Korean distant water longline data. First, we explore the operational set-by-set data and develop data-based indicators of effort targeting SBT, using the number of hooks between floats (HBF), and the month. Secondly, we use cluster analysis to group the effort into fishing strategies based on the species composition of the catch.

We also apply two methods, the lognormal constant method and the delta lognormal method, for estimating indices.

DATA AND METHODS

Set by set (operational) catch and effort data were compiled by the Korean National Institute of Fisheries Science (NIFS). Data were selected with the criterion that when a vessel reported the capture of at least one SBT in a month, all effort for the vessel-month was included.

The Korean tuna longline vessels fishing for SBT have mainly operated in two locations to the south of 35ºS either between 10°W-50°E (within statistical area 9) or between 90°E-120°E (within statistical area 8) (Fig. 2). Effort has focused on western areas (statistical area 9) from March to September/October and shifted to the east (statistical area 8) from July/August until December (Fig. 3). In general, there has been more fishing effort in the west.

The fields reported in the operational data were catch (number), effort (number of hooks), floats (number of floats), vessel id, fishing location to 1° cell of latitude and longitude, fishing date, and catch in numbers of southern bluefin tuna (SBT), bigeye tuna (BET), yellowfin tuna (YFT), albacore (ALB), skipjack tuna (SKJ), swordfish (SWO), black marlin (BLM), blue marlin (BUM), striped marlin (MLS), sailfish (SFA), sharks (SHA), and other species (OTH).

Data from the period 1996-2022 were used in this study, because data prior to 1996 were not available due to insufficient data from vessels targeting SBT. Dates were converted to months and quarters. Moon phase was used to calculate the relative lunar illumination for each date, using the R package *lunar* (Lazaridis 2014). Spatial positions were classified into 5° cells and CCSBT statistical areas. The numbers of hooks between floats (HBF) were calculated by dividing hooks by floats and rounding to the nearest whole number.

For CPUE standardization, data were cleaned by removing sets in which there were fewer than 1,000 hooks and more than 5,000 hooks.

Data were plotted to explore trends in total catch over time; the spatial and seasonal distributions of effort; and patterns in operational characteristics such as HBF and hooks per set. We examined patterns over time and among species in the nominal catch rates by year-quarter and statistical area, and compared them with patterns in the proportions of sets with no catch of each species. We plotted maps of the species composition over time, to identify possible changes in fishing behaviour or population composition.

To explore changes in effort distribution and concentration over time, we plotted the numbers of 5°x5° and 1°x1° cells fished and the average number of operations per fished cell for each statistical area and for each year.

To address target changes over time, two approaches, data selection and cluster analysis, were applied as in previous study, and those methods are described in detail in Hoyle et al. (2019).

The operational data were standardized using Generalized Linear Models (GLM) in Microsoft R Open 3.3.2 (R Core Team 2016). Data were prepared by selecting operational data for vessels that had made at least 100 sets, for years in which there had been at least 100 sets, and for 5° cells in which there had been at least 200 sets.

SBT CPUE standardization was conducted by lognormal constant model and delta lognormal model, and the details are described in Hoyle et al. (2019).

The lognormal constant model is as follows.

$$\ln(CPUE_s + k) \sim year + vessid + latlong + \lambda(hooks) + g(month) + h(moon)$$

The constant k, added to allow for modelling sets with zero catches of the species of interest, was 10% of the mean CPUE for all sets. The functions λ, g and h were cubic splines with 5, 3, and 4 degrees of freedom respectively. The number of hooks was included in the model to allow for possible hook saturation or other factors associated with hooks per set. The variable *moon* was the lunar illumination on the date of the set. The variables *year*, *vessid*, and *latlong* (5° latitude-longitude cells) were fitted as categorical variables. For the clustering-based approach, models also included a categorical variable for the cluster.

Delta lognormal analyses (Lo et al. 1992, Maunder and Punt 2004) used a binomial distribution for the probability *w* of catch rate being zero and a probability distribution *f(y)*, where y was log(catch/hooks set), for non-zero (positive) catch rates. The index estimated for each year-quarter was the product of the year effects for the two model components, $(1-w).E(y|y \neq 0)$.

$$\Pr(Y=y) = \begin{cases} w, & y=0 \\ (1-w)f(y) & otherwise \end{cases}$$

$$\mathrm{g}(w) = (CPUE = 0) \sim covariates + \epsilon,$$

$$\mathrm{f}(y) = CPUE \sim covariates + \epsilon$$

where *g* is the logistic function.

Models did not include HBF because the ‘select’ method addresses HBF by only including values in the range 9-12. The ‘cluster’ method addresses targeting independently of HBF, and in any case only less than 1% of sets included HBF outside the 9-12 range.

Data in all models were ‘area-weighted’, with the weights of the sets adjusted so that the total weight per year-quarter in each 5° square would sum to 1. This method was based on the approach identified using simulation by Punsly (1987) and Campbell (2004), that for set *j* in area *i* and year-quarter *t*, the weighting function that gave the least average bias was: $w_{ijt} = \frac{log(h_{ijt}+1)}{\sum_{j=1}^{n} \log(h_{ijt}+1)}$. Given the relatively low variation in number of hooks between sets in a stratum, we simplified this to $w_{ijt} = \frac{h_{ijt}}{\sum_{j=1}^{n} h_{ijt}}$.

Model fits were examined by plotting the residual densities and using Q-Q plots.

RESULTS AND DISCUSSIONS

Data exploration

Almost all effort used between 9 and 12 hooks per float (HBF) (Fig. 4), while the majority of HBF outside this range came from north of 35°S, outside the main SBT targeting area. The number of hooks per set averaged less than 3000 in the period from 1990-95, but since that time has been relatively consistent, averaging a little over 3000.

Mean catch rates by species in the statistical areas 7, 8, and 9 are highest for SBT until the mid-2000s (Fig. 5). After this time in area 8 and 9 SBT catch rates decrease and other species, particularly albacore tuna, increase. However, in the most recent years the SBT catch rates are again higher than other species. Similarly, the proportion of sets reported with zero SBT catches was low through most of the time series in the areas 8 and 9, but area 9 shows an increase in the proportion of zero catches from 2004 to 2010 (Fig. 6).

In the statistical areas 13 and 1, the tropical bigeye and yellowfin tunas dominate with the highest catch rates, along with albacore tuna (Fig. 5). SBT catch rates are low throughout the time series, even though data were only selected for vessels reporting at least one SBT in the month.

Statistical areas 14 and 2 in the Indian Ocean are at temperate latitudes between 20S and 35S. Highest catch rates are for yellowfin and more recently albacore tuna in the area 14, and bigeye and albacore tuna in area 2 (Fig. 5). Since the mid-2000s albacore tuna catch rates have increased markedly and particularly in area 2, suggesting a trend towards targeting this species. Catch rates of SBT have been relatively low throughout the period, consistent with a high proportion of zero SBT sets, suggesting little or no deliberate targeting of SBT by the Korean longline fleet in these statistical areas (Fig. 6).

We mapped the proportions of SBT and ALB in total catch at south of 30°S by 5-year period (Figs. 7 and 8). The proportion of SBT in the catch was high in all periods, increasing further south, but declined steadily in all areas after 2005. Since the post-2010 period there is little SBT taken in statistical area 8 north of about 37°S, whereas a high proportion of the catch in this area is albacore tuna.

The total number of major (5° x 5° x month) cells fished has varied from year to year but declined steadily and considerably since the peak in 2009. Over the same period, effort has become more concentrated with more operations per cell. This increasing concentration is also apparent at the minor (1° x 1° x month) cell level (Fig. 9). The distribution of effort within major cells was more stable until recently, with similar numbers of minor cells per major cell on average, but the effort concentration in 2017-2019 increased to the highest level and in 2020 decreased to the level of 2016. After that, the effort concentration has fluctuated.

Target change

The data selection approach aimed to identify effort targeted mostly at SBT, by selecting area 9 data from March-October and area 8 data from July-December (Fig. 3).

Applying Ward's D hierarchical cluster analysis at the vessel-month identified strong separation among 2 to 3 groups in statistical areas 9 and 8 (Fig. 10). We chose to use 3 clusters in each area. We preferred to use more clusters where there was uncertainty because unresolved target change can cause bias in indices.

In area 9 (Figs. 11-13), clusters 1, 2, and 3 were more strongly represented in the late, middle, and early parts of the time series, respectively. Clusters 1 and 3 occur mostly in the period before August, while cluster 2 extends into October. Cluster 1 also has slightly more hooks between floats. Mean number of hooks is higher in cluster 1 and lower in cluster 2. Cluster 2 dominates the northeast of area 9, while cluster 1 dominates the southwest, and cluster 3 the southeast. The species composition of cluster 3 comprises almost entirely SBT, with small amounts of ALB and BET. Cluster 2 has significant SBT along with ALB, and some BET and YFT. Cluster 1 includes some SBT along with similar amounts of OTH, and also some SHA and ALB.

In area 8 (Figs. 11-13), cluster 1 dominates the early part of the time series, with clusters 2 and 3 more apparent after 2005. Cluster 1 averages fewer hooks between floats, while the hooks per set are similar for all clusters. Clusters 1 and 2 occur mostly in the second half of the year, while cluster 3 is represented during March to June. Cluster 1 is well represented at the east and the south of area 8, while cluster 2 occurs at the west of middle latitudes from about 38-42°S. Cluster 3 occurs almost entirely in the far north of the area. Cluster 1 is dominated by SBT, with little amounts of other species. Cluster 2 has significant SBT, along with similar amounts of SHA and OTH. Cluster 3 has more ALB than SBT, with some OTH, YFT and BET.

CPUE standardization

Table 1 shows the results of dropping each variable from the delta lognormal models, indicating that all explanatory variables were statistically significant, with the year, location, vessel, and month that are the largest factors affecting the model fit. For the cluster effect, it is also one of important factors of the model fit in the area 8, but the effect is less in the area 9.

The lognormal constant indices for the area 8 are broadly similar and indices for the area 9 are very similar for both approaches to addressing target change. And the delta lognormal indices are very similar for both targeting analysis methods (Fig. 14, Tables 2 and 3).

The lognormal constant model has important differences between the indices in the late 2000s for the area 9 and in the period of 2013-2014 for the area 8. For the area 9, the standardized indices are lower than the unstandardized indices in the late 2000s, and for the area 8, the clustered data indices show lower values than the other indices in that period. These would be because ALB has higher catch rates and lower zero catches, while SBT has lower catch rates and higher zero catches during that period of each area (Figs. 5 and 6). For the area 8, there are also differences between the indices in 2015 and 2016 when efforts are so low and concentrated (Fig. 9), and the index have a large decrease in 2020 and a slightly increase in 2021.

The delta lognormal indices are very similar for both targeting analysis methods. This may be because delta lognormal models are better than lognormal constant models at dealing with zero catches. However, the delta lognormal indices also had lower values for the clustered data than the selected data in the period of 2013-2014 for the area 8. In addition, they differ from the lognormal constant indices in several ways for the area 9. First, they are markedly lower in the period before 2005. Second, they are considerably higher than the lognormal constant indices in 2015.

Diagnostic frequency distributions and QQ-plots suggest that the data fitted the GLM adequately (Fig. 15).

Patterns in the indices differ somewhat between east and west. Both sets of indices decreased until the mid-2000s, and subsequently increased, particularly in the last few years.

REFERENCES

Campbell, R.A. (2004) CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. *Fisheries Research*, 70(2-3): 209-227.

CCSBT (2017) Report of the Twenty Second Meeting of the Scientific Committee.

Hoyle, S.D., Lee, S.I., Kim, Z.G. (2019) Data exploration and CPUE standardization for the Korean southern bluefin tuna longline fishery (1996-2018). CCSBT-ESC/1909/39.

Kim, Z.G., Kim, D.N., Lee, S.I., Kwon, Y., Cha, H.K. (2015) CCSBT-ESC/1509/SBT, 2015 Annual National Report of Korean SBT Fishery. *20th Extended Scientific Committee of the CCSBT*, Incheon, Republic of Korea, 1 - 5 Sep 2015.

Lazaridis, E. (2014) lunar: Lunar Phase & Distance, Seasons and Other Environmental Factors (Version 0.1-04). Available from http://statistics.lazaridis.eu. http://statistics.lazaridis.eu

Lo N.C., Jacobson, L.D., Squire, J.L. (1992) Indices of relative abundance for fish spotter data based on delta-lognormal models. *Can J Fish Aquat Sci* 49:2515-2526.

Maunder, M.N., Punt, A.E. (2004) Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*, 70(2-3): 141-159. 10.1016/j.fishres.2004.08.002

Punsly, R. (1987) Estimation of the relative annual abundance of yellowfin tuna, Thunnus albacares , in the eastern Pacific Ocean during 1970-1985. I-ATTC, LA JOLLA, CA ( ).

R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Table 1**.** Degrees of freedom, deviance, and delta AIC results from the delta lognormal models for CCSBT statistical areas 9 and 8, addressing target change using clustered data

Variable	Clustering analysis					
	Statistical area 9			Statistical area 8		
	df	Deviance	ΔAIC	df	Deviance	ΔAIC
<none>		177.9	0		57.6	0
year	26	207.3	2,731	17	66.5	1,305
latlong	19	197.1	1,831	10	59.7	307
hooks	5	179.0	103	5	57.8	17
vessid	29	185.3	681	24	59.7	291
month	3	180.6	264	3	59.2	244
moon	4	179.7	170	4	58.9	201
cluster	2	178.6	65	2	59.3	270

Table 1. Lognormal constant indices for statistical areas 9 and 8, for selected data (left) and clustered data (right)

Year	Selected data				Clustered data			
	Stat area 9	CV	Stat area 8	CV	Stat area 9	CV	Stat area 8	CV
1996	0.90	0.04	1.01	0.04	0.85	0.03	0.91	0.04
1997	0.66	0.03	0.50	0.02	0.74	0.03	0.60	0.02
1998	0.66	0.03	0.52	0.02	0.69	0.03	0.65	0.03
1999	0.71	0.03	0.46	0.02	0.67	0.03	0.59	0.03
2000	0.61	0.03	0.49	0.02	0.61	0.03	0.60	0.03
2001	0.70	0.03	0.55	0.03	0.69	0.03	0.60	0.03
2002	0.67	0.03	0.35	0.02	0.67	0.03	0.38	0.03
2003	0.55	0.03	-	-	0.58	0.03	-	-
2004	0.29	0.03	-	-	0.30	0.03	-	-
2005	0.16	0.05	-	-	0.17	0.05	-	-
2006	0.45	0.04	-	-	0.49	0.04	-	-
2007	0.36	0.03	-	-	0.38	0.03	-	-
2008	0.71	0.03	0.75	0.02	0.72	0.03	0.79	0.02
2009	0.53	0.03	0.54	0.03	0.54	0.03	0.50	0.03
2010	0.56	0.03	0.58	0.02	0.56	0.03	0.55	0.03
2011	1.35	0.04	0.75	0.02	1.26	0.04	0.74	0.03
2012	1.08	0.03	0.90	0.03	1.05	0.03	0.91	0.04
2013	0.92	0.04	1.18	0.04	0.91	0.04	0.84	0.04
2014	1.77	0.04	1.33	0.05	1.75	0.04	0.75	0.05
2015	0.94	0.04	0.78	0.04	0.94	0.04	0.79	0.05
2016	1.36	0.03	2.58	0.12	1.35	0.03	2.31	0.13
2017	1.41	0.04	-	-	1.41	0.04	-	-
2018	2.02	0.04	-	-	2.00	0.04	-	-
2019	2.43	0.04	-	-	2.42	0.04	-	-
2020	1.76	0.04	2.24	0.07	1.77	0.04	2.64	0.07
2021	1.95	0.05	2.51	0.07	1.97	0.05	2.85	0.08
2022	1.50	0.05	-	-	1.51	0.05	-	-

Table 2. Delta lognormal indices for statistical areas 9 and 8, for selected data (left) and clustered data (right)

Year	Selected data				Clustered data			
	Stat area 9	CV	Stat area 8	CV	Stat area 9	CV	Stat area 8	CV
1996	0.66	0.03	0.71	0.04	0.63	0.03	0.86	0.04
1997	0.43	0.03	0.48	0.02	0.50	0.03	0.55	0.02
1998	0.44	0.03	0.56	0.02	0.47	0.03	0.63	0.03
1999	0.52	0.03	0.43	0.03	0.46	0.03	0.50	0.03
2000	0.45	0.03	0.47	0.02	0.42	0.03	0.54	0.03
2001	0.53	0.03	0.52	0.03	0.51	0.03	0.57	0.03
2002	0.57	0.03	0.31	0.03	0.56	0.03	0.33	0.03
2003	0.53	0.03	-	-	0.55	0.03	-	-
2004	0.24	0.03	-	-	0.25	0.03	-	-
2005	0.20	0.05	-	-	0.23	0.06	-	-
2006	0.31	0.03	-	-	0.34	0.04	-	-
2007	0.30	0.03	-	-	0.31	0.03	-	-
2008	0.59	0.03	0.81	0.02	0.59	0.03	0.87	0.02
2009	0.47	0.03	0.52	0.03	0.50	0.03	0.60	0.03
2010	0.52	0.03	0.62	0.03	0.54	0.03	0.62	0.03
2011	1.21	0.04	0.80	0.02	1.13	0.04	0.80	0.02
2012	0.85	0.03	0.99	0.04	0.84	0.03	0.99	0.04
2013	0.90	0.04	1.32	0.04	0.90	0.04	1.04	0.04
2014	1.70	0.04	1.55	0.06	1.71	0.04	1.00	0.05
2015	2.74	0.05	0.88	0.04	2.79	0.05	0.81	0.05
2016	1.27	0.03	2.00	0.06	1.25	0.03	1.81	0.06
2017	1.50	0.04	-	-	1.50	0.04	-	-
2018	1.97	0.04	-	-	1.94	0.04	-	-
2019	2.62	0.05	-	-	2.60	0.05	-	-
2020	1.73	0.04	2.39	0.06	1.73	0.04	2.68	0.06
2021	2.05	0.05	2.63	0.06	2.06	0.05	2.82	0.06
2022	1.70	0.05	-	-	1.70	0.05	-	-

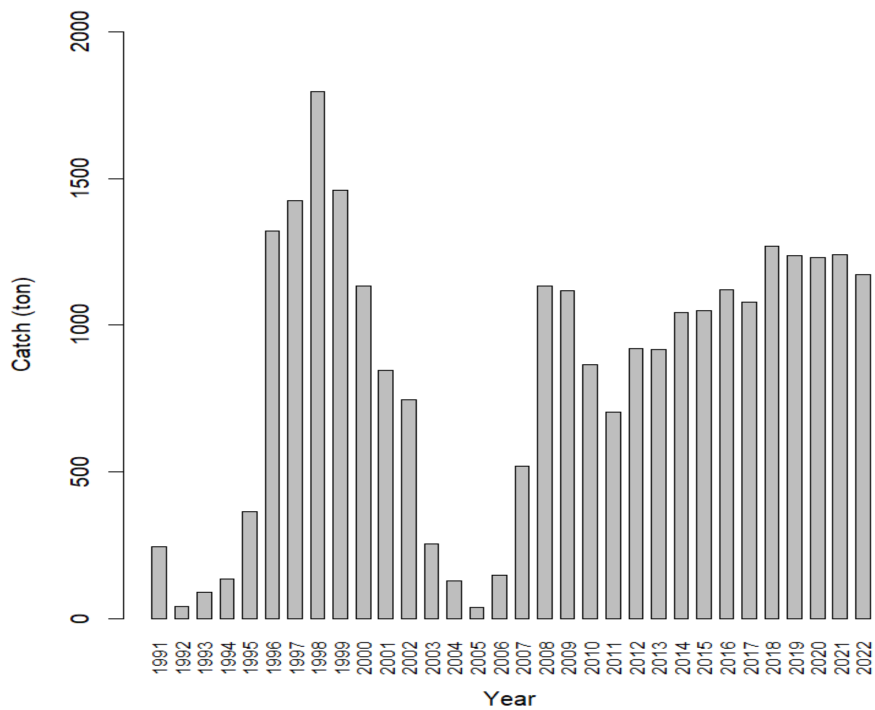

Fig. 1. The annual Korean SBT catches in the CCSBT convention area, 1991 - 2022.

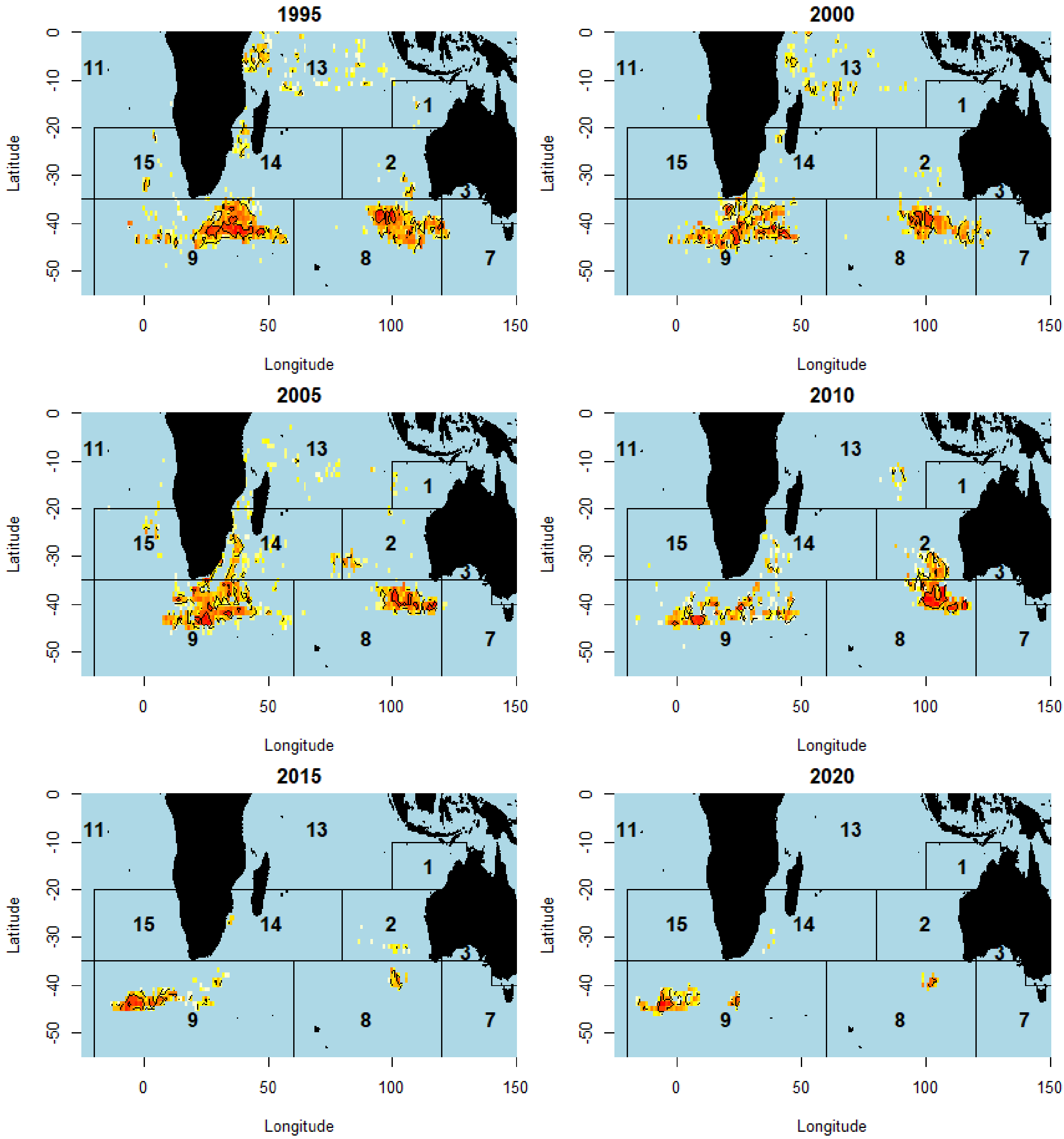

Fig. 2. Map showing the core areas of Korean tuna longline vessels fishing for SBT, aggregated by 5-year period. Red colour indicates higher fishing effort, n numbers of hooks.

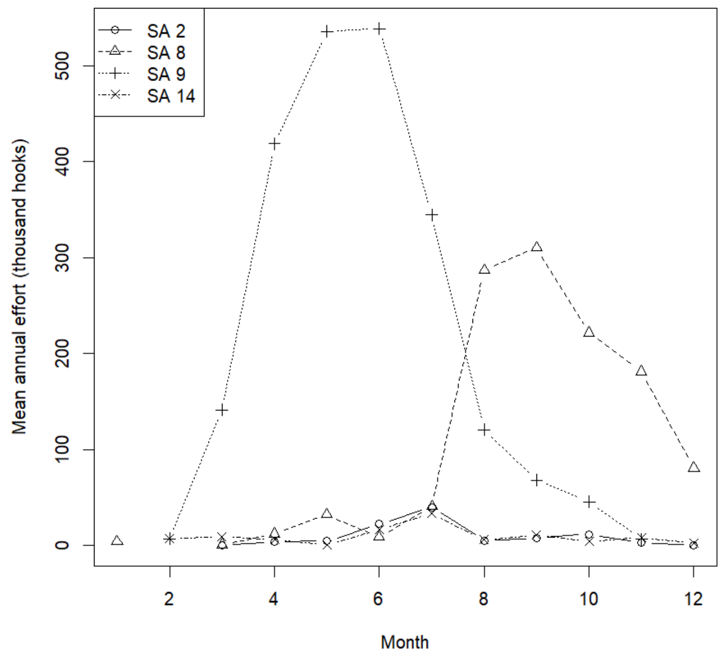

Fig. 3. Mean annual effort in thousands of hooks, by month and statistical area.

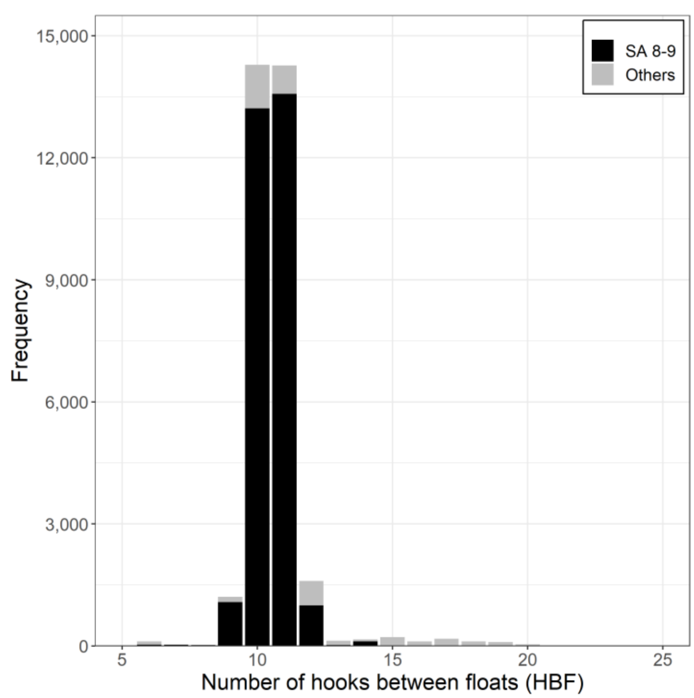

Fig. 4. Frequency table of HBF for the main fishing ground with the lighter shade for statistical areas (SA) 8-9, and the darker shade for other areas.

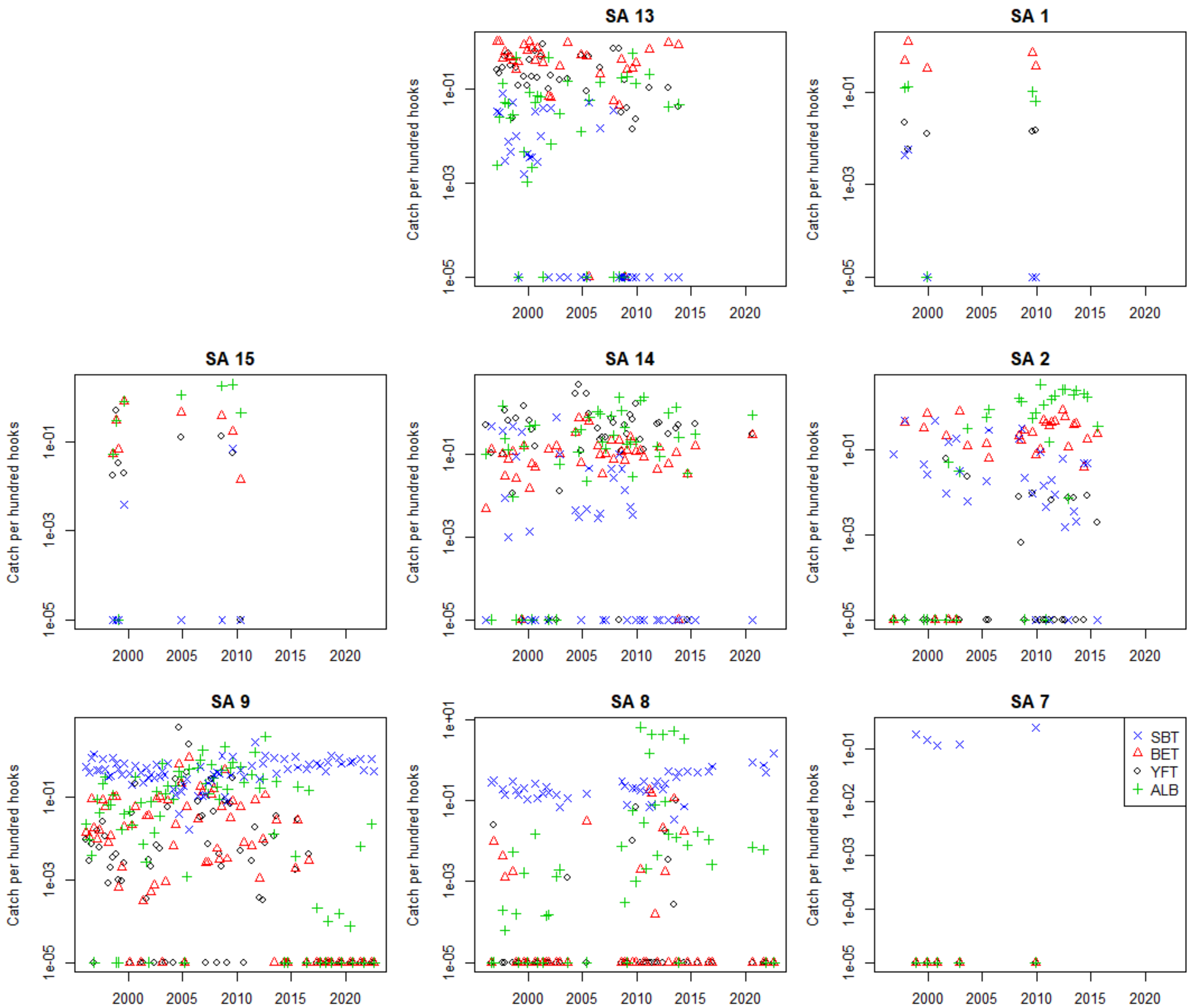

Fig. 5. Mean catch per hundred hooks by year-quarter, species, and statistical area, plotted on a log scale, for yellowfin, bigeye, albacore, and southern bluefin tuna. Each CPUE has 1E-5 added so that zero catches appear on the log scale.

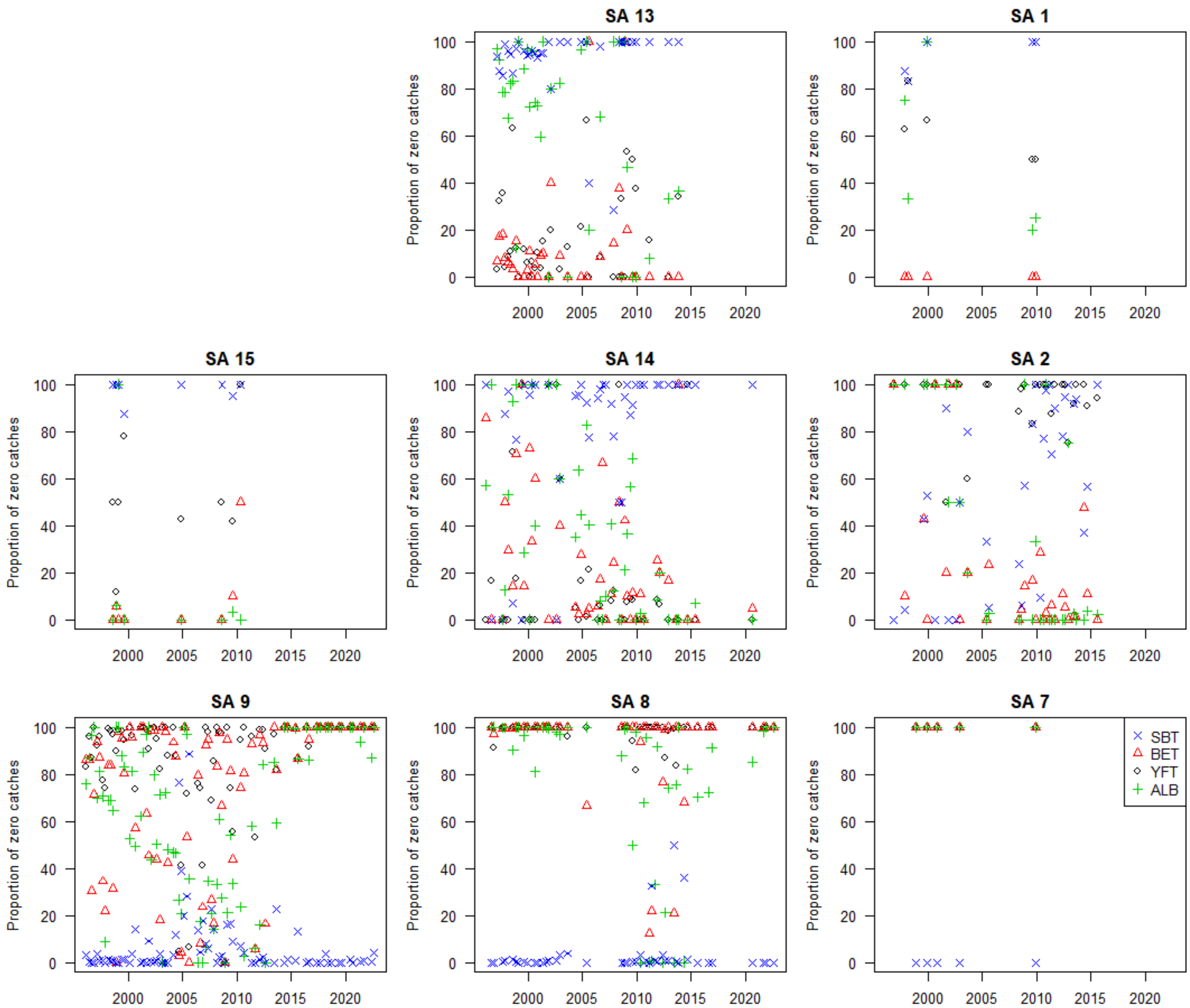

Fig. 6. Proportion of zero catches per set by year-quarter, species, and statistical area, for yellowfin, bigeye, albacore, and southern bluefin tuna.

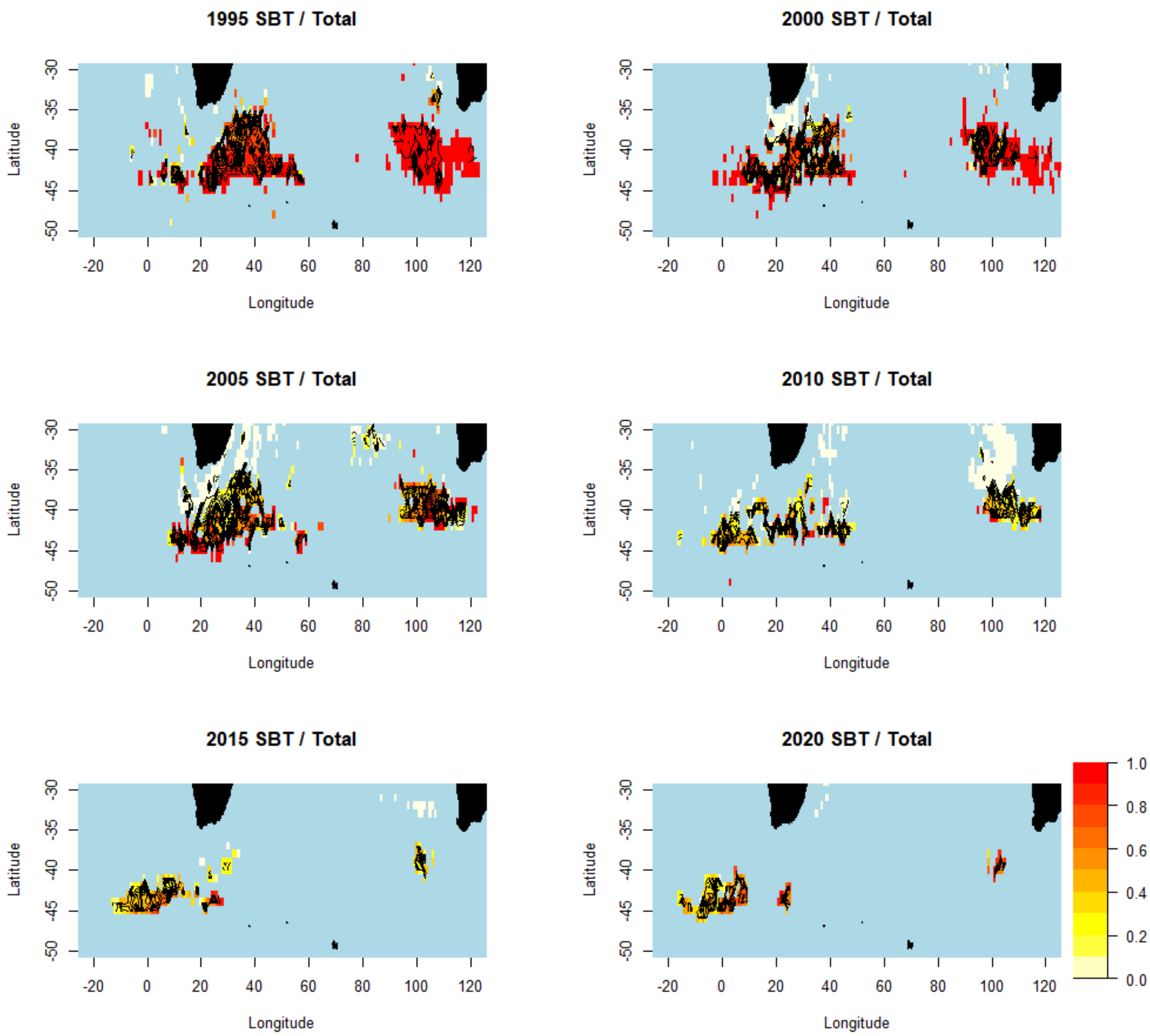

Fig. 7. Proportion of southern bluefin tuna (SBT) in the total reported catch in numbers by 1° cell, aggregated over 5 years within the period 1995-2022. Red colour indicates a higher proportion of SBT.

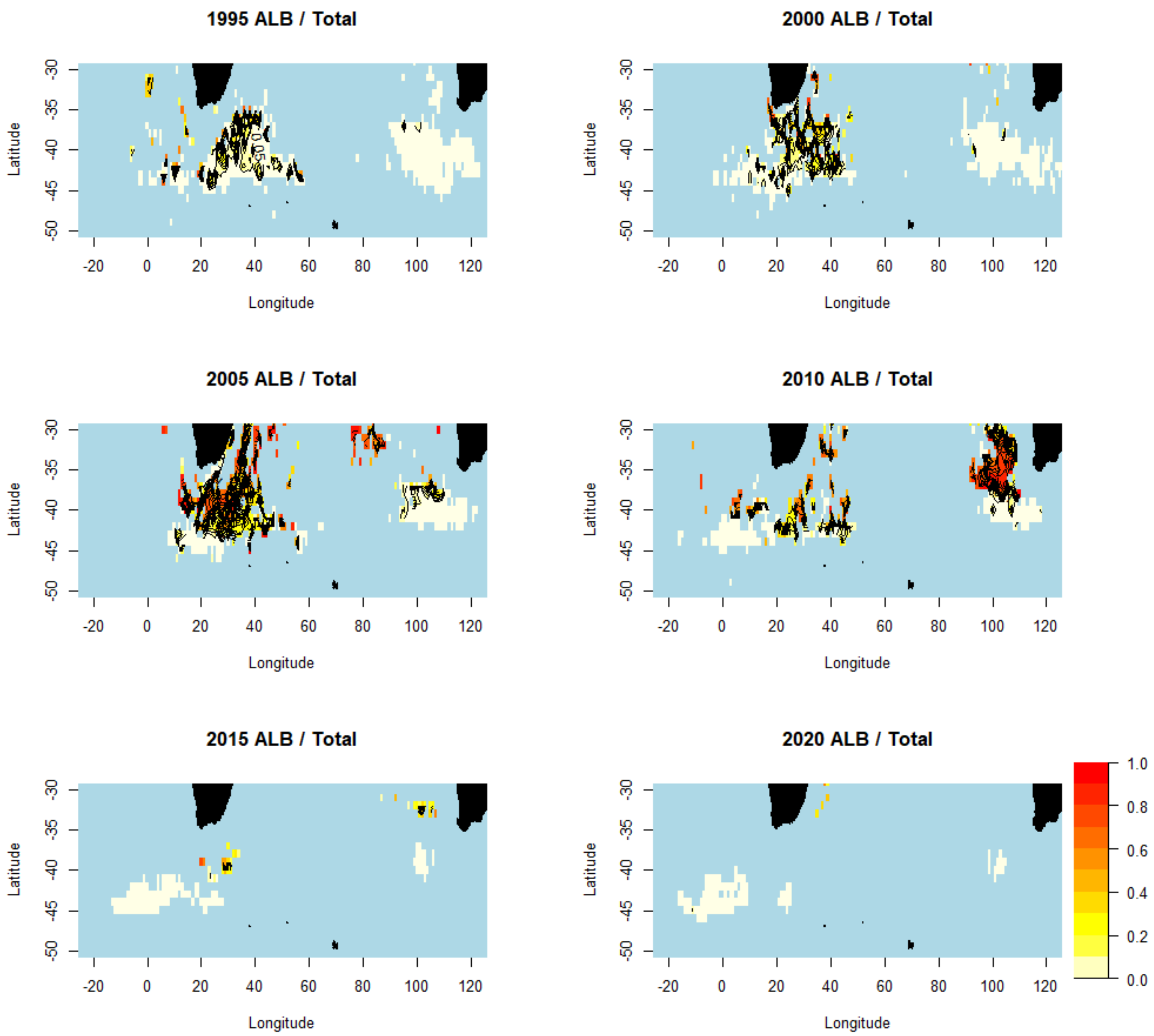

Fig. 8. Proportion of albacore (ALB) in the total reported catch in numbers by 1° cell, aggregated over 5 years within the period 1998-2022. Red colour indicates a higher proportion of ALB.

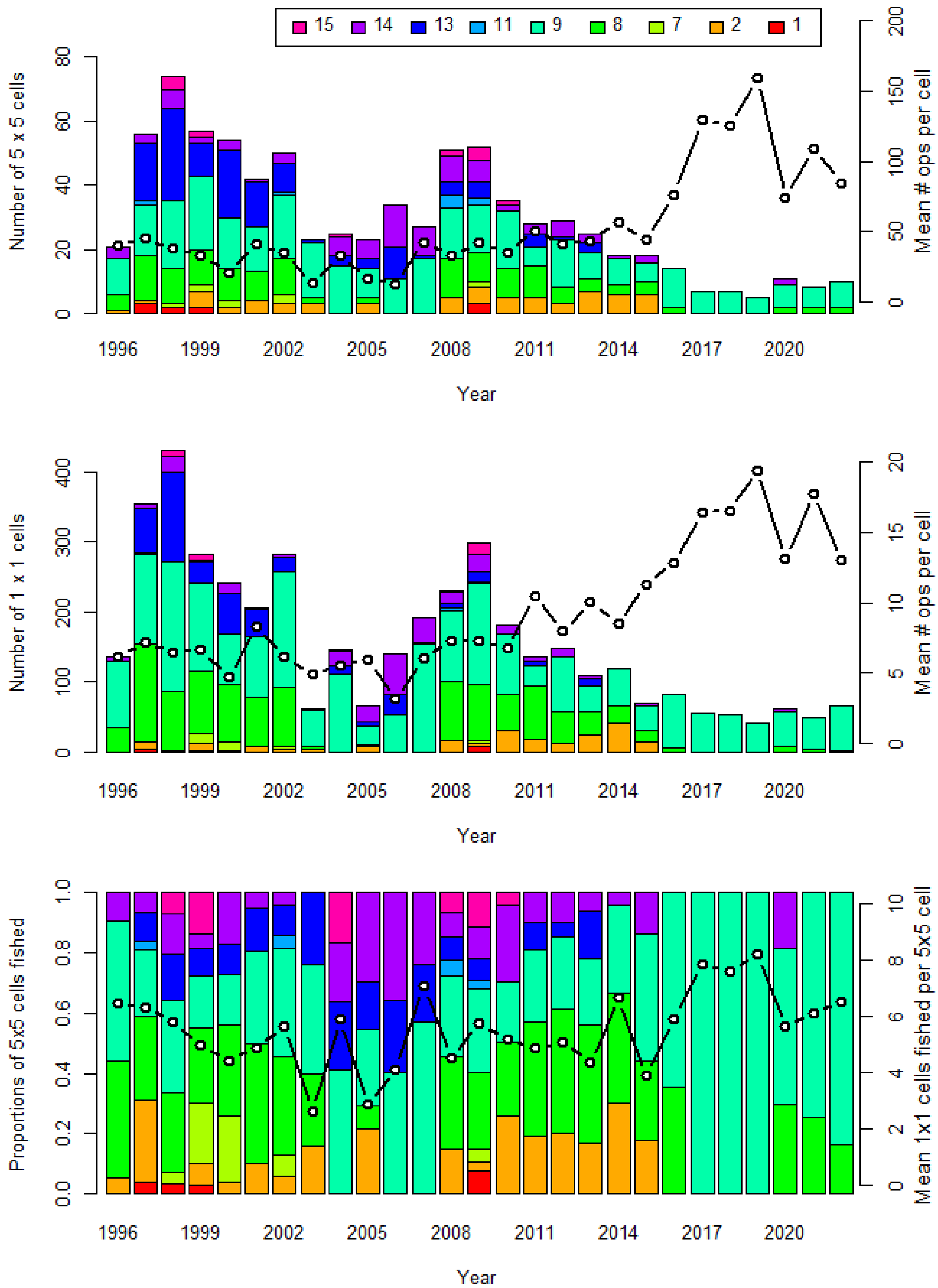

Fig. 9. (Upper) Bars represent the number of major cells (5°x5° by month) fished by CCSBT statistical area and year, see left y-axis. The line represents the mean annual operations per cell, see right y-axis. (Middle) As for upper plot, but with minor cells (1x1° by month) instead of major cells. (Lower) Relative distribution of fished major cells by the proportion of the cell fished, measured as the number of minor cells fished within each major cell, see left y-axis. The lowest (red) and highest (purple) bands represent major cells in which, respectively, 1 and 15 of the 25 minor cells were fished. The line represents the mean number of minor cells fished per major cell by year, see right y-axis.

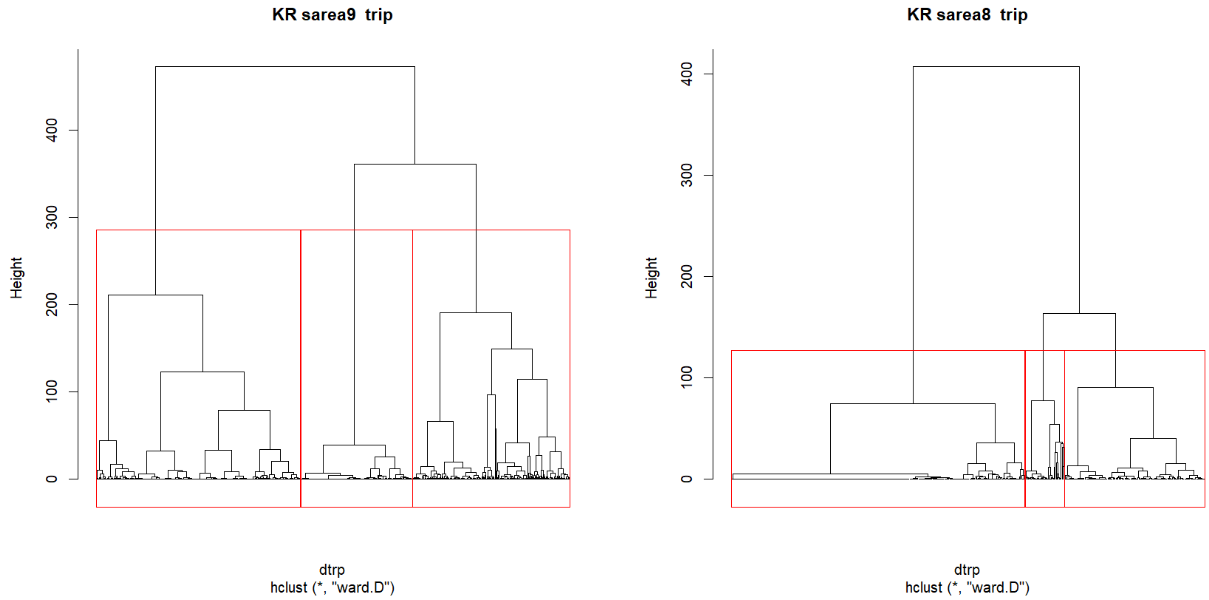

Fig. 10. Dendrograms for Ward’s D hierarchical cluster analyses of statistical areas 9 (left) and 8 (right), with the red lines indicating the separation into 3 clusters for each.

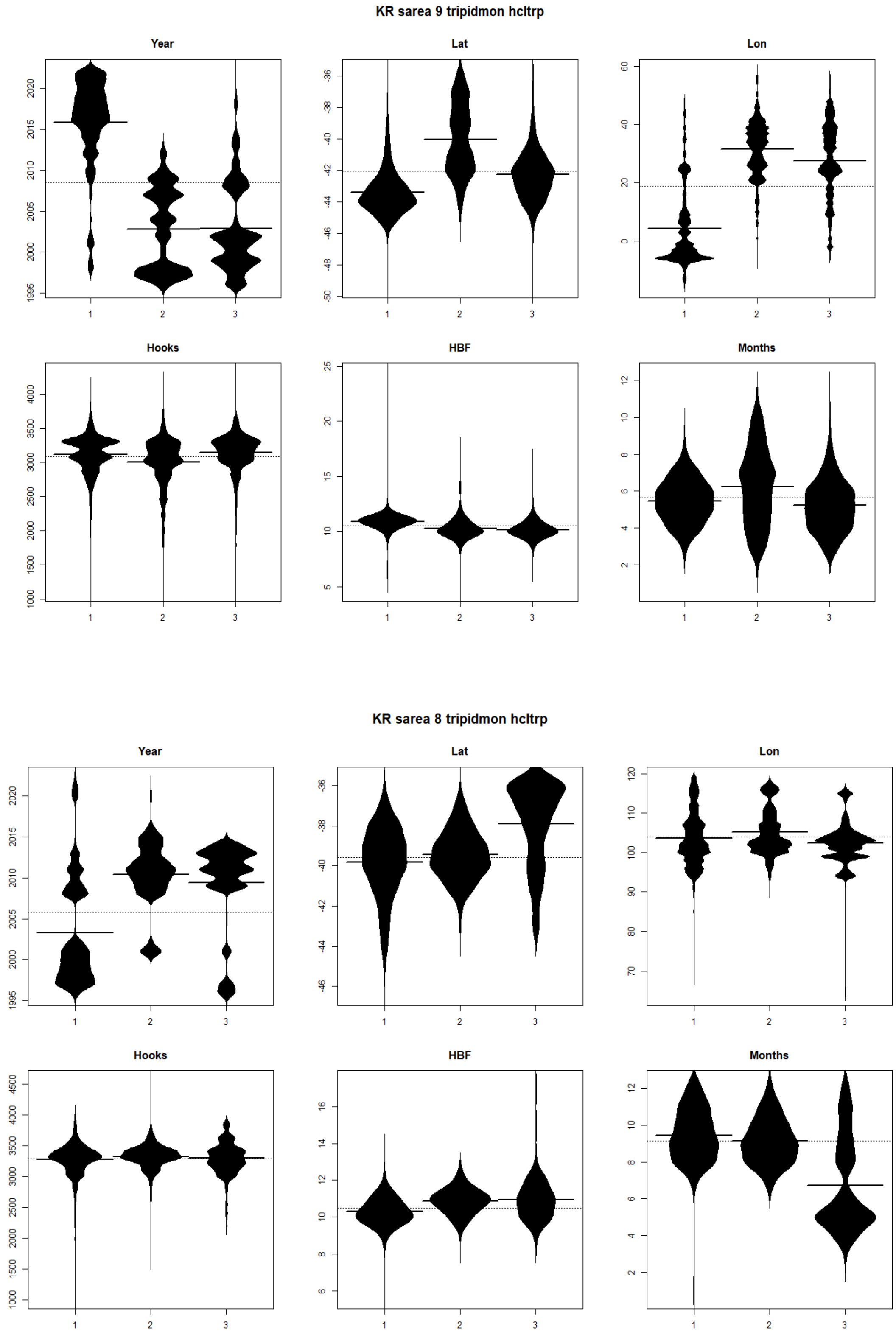

Fig. 11. Beanplots for statistical areas 9 (above) and 8 (below), showing the number of sets versus covariate by cluster. The horizontal bars indicate the medians.

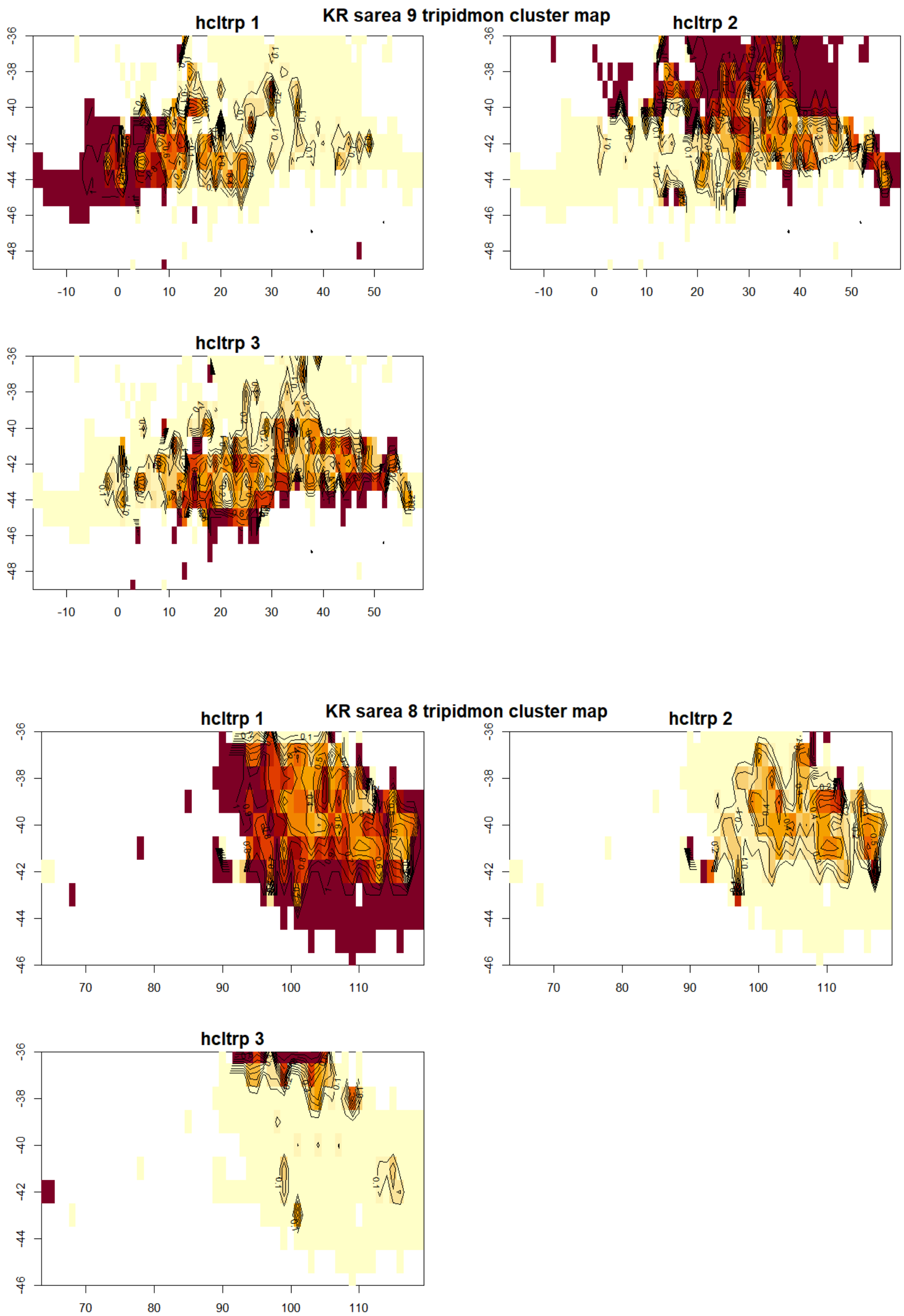

Fig. 12. Maps of the proportion of each cluster per 1 degree square in total effort for statistical areas 9 (above) and 8 (below). Higher proportions are shown in red. White space indicates no reported effort.

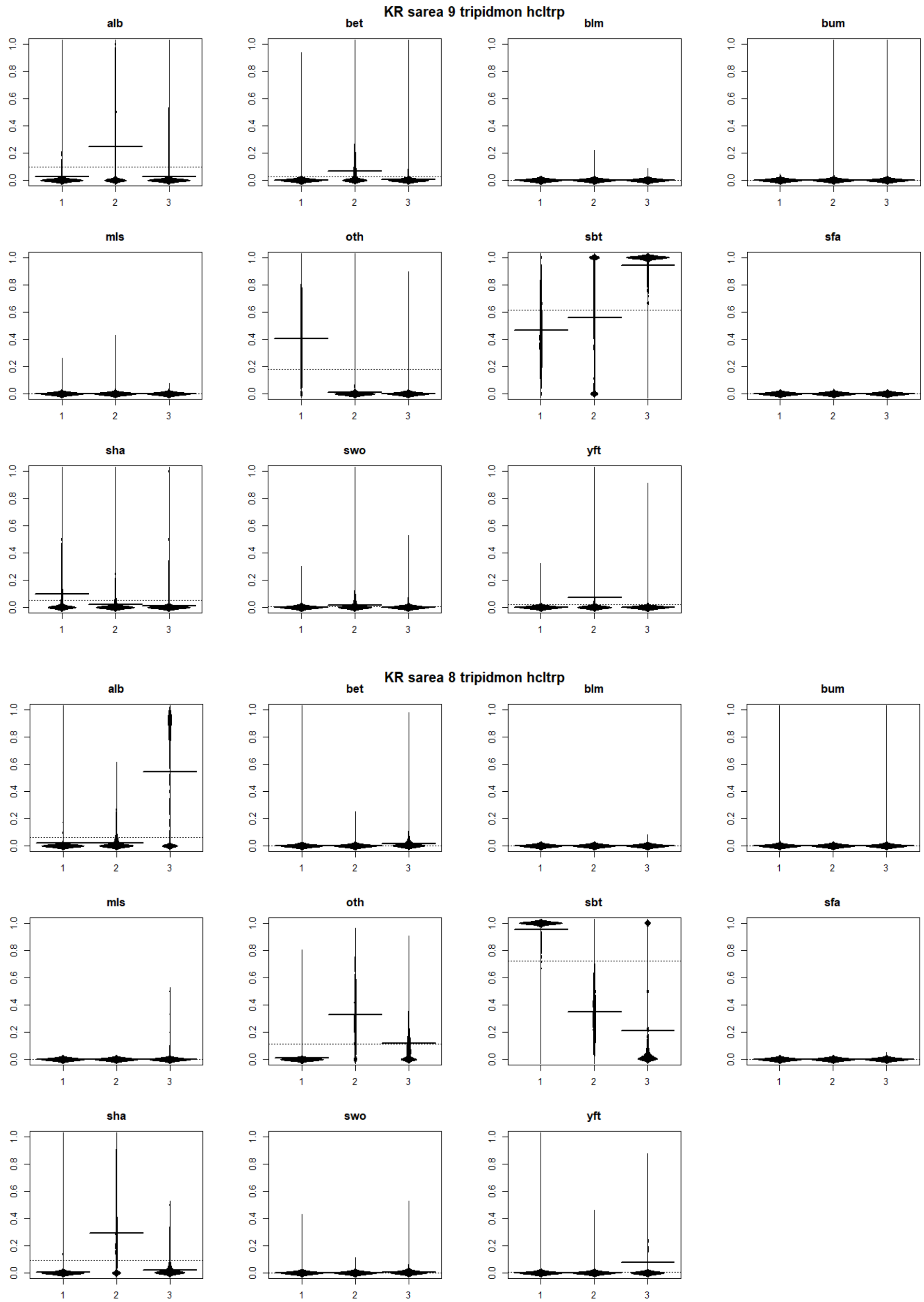

Fig. 13. Beanplots for statistical areas 9 (above) and 8 (below), showing species composition by cluster. The horizontal bars indicate the medians.

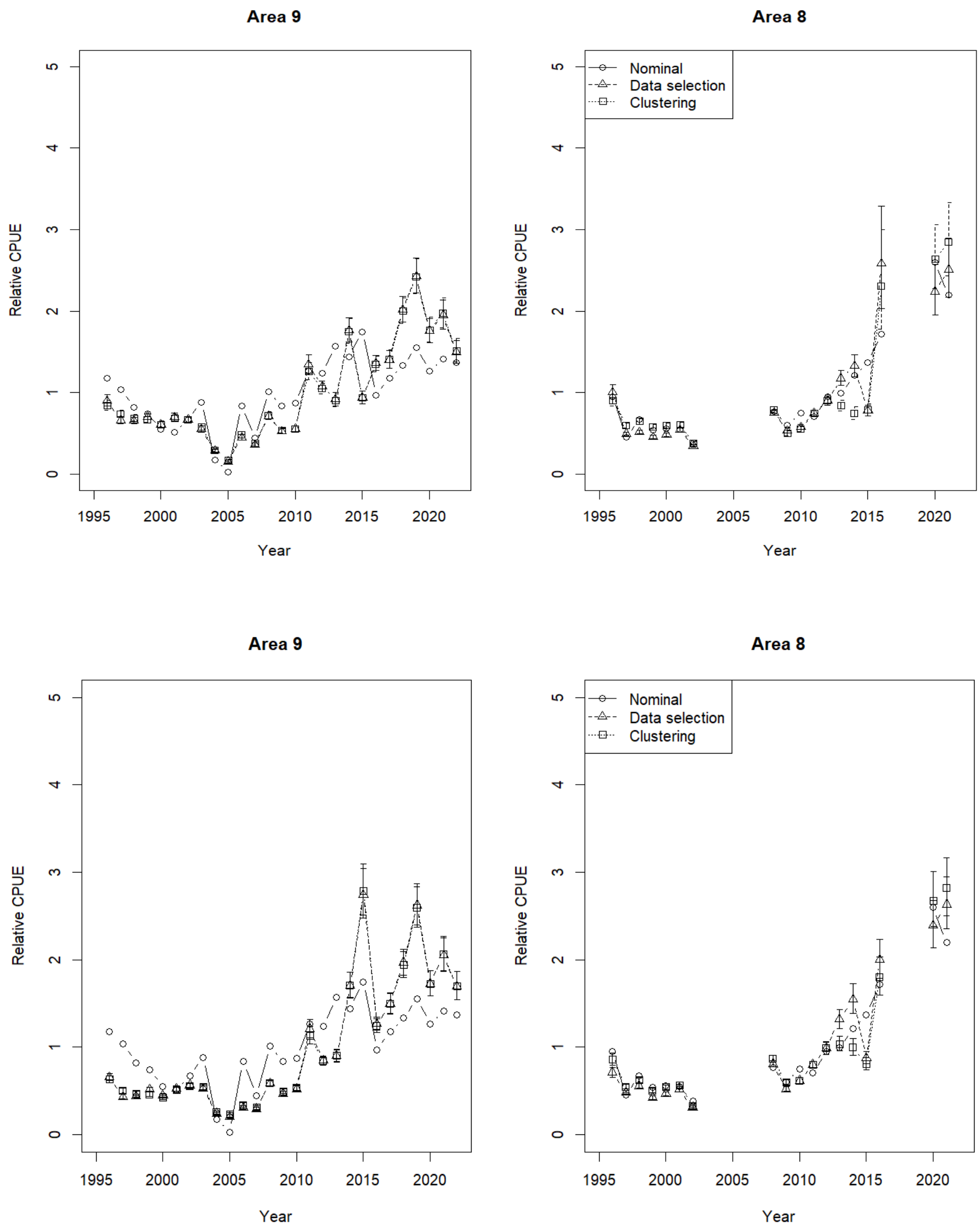

Fig. 14. Nominal and standardized CPUE indices based on lognormal GLMs with an added constant (above) and delta lognormal models (below), addressing target change using selected data (triangles) and cluster analysis (squares), for statistical areas 9 (left) and 8 (right).

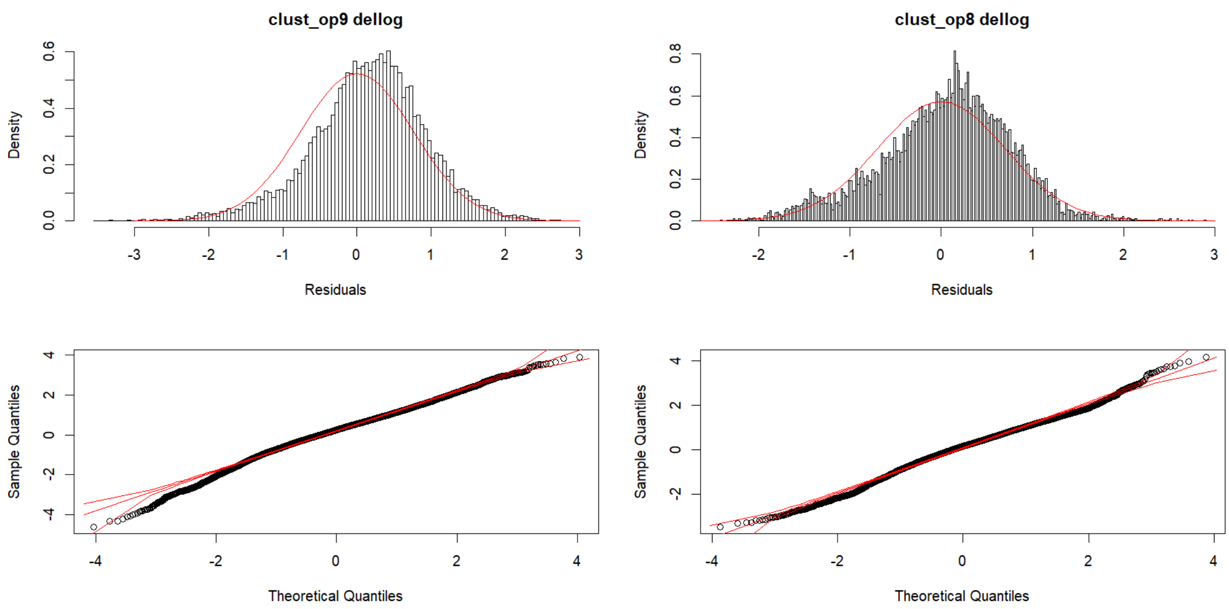

Fig. 15. Frequency distributions of the standardized residuals (above) and Q-Q plots of the standardized residuals for lognormal constant GLM analyses of statistical areas 9 (left) and 8 (right), based on the model with cluster as a covariate.