DATA EXPLORATION AND CPUE STANDARDIZATION FOR THE KOREAN SOUTHERN BLUEFIN TUNA LONGLINE FISHERY (1996-2016)

Simon Hoyle[1], Sung Il Lee[2] and Doo Nam Kim[2]

ABSTRACT

In this study we standardized southern bluefin tuna, *Thunnus maccoyii* (SBT) CPUE from Korean tuna longline fisheries (1996-2016) using Generalized Linear Models (GLM) with operational data. The data used for the GLMs were catch (number), effort (number of hooks), number of hooks between floats (HBF), fishing location (5° cell), and vessel identifier by year, quarter, and area. We explored CPUE by area, and identified two separate areas 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 through time which can affect CPUE indices. Explanatory variables for the GLM analyses were year, month, vessel identifier, 5° cell, and number of hooks. GLM results for the whole area suggested that location, year, targeting, and month effects were the most important 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

Korean tuna longline fisheries began targeting southern bluefin tuna, *Thunnus maccoyii* (SBT) in the CCSBT convention area in 1991 (Kim, Kim et al. 2015), although SBT were reported as bycatch before this time, starting in 1972. The 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 705-1,120 mt due to the national catch limit. The catch in 2016 was 1120 mt (Fig. 1).

Korean tuna longline vessels fishing for SBT have mainly operated in two locations to the south of 35°S either between 10°E-50°E (within statistical area 9) or between 90°E-120°E (within statistical area 8) (Figure 2). Effort has focused on western areas (statistical area 9) from March to July/August and shifted to the east (statistical area 8) from July/August until December (Figures 3 and 4). In general, there has been more fishing effort in the west, and this has been the case in 2015-16.

In this study, we first explored the data to better understand the fisheries, and then standardized the CPUE using several approaches to obtain a proxy for the abundance index.

---

[1] Hoyle Consulting, Nelson, New Zealand, simon.hoyle@gmail.com
[2] National Institute of Fisheries Science, Busan, Rep. of Korea, k.sungillee@gmail.com

DATA AND METHODS

Catch and effort data were selected with the criterion that when a vessel reported the capture of at least one 1 SBT in a month, all effort for the vessel-month was included.

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

Dates were converted to months and quarters, and to identify moon phase. Spatial positions were classified into 5° cells, and CCSBT statistical areas.

For CPUE standardization, data were cleaned by removing sets in which there were fewer than 1000 hooks.

Data were plotted to explore trends in total catch through time; the spatial and seasonal distributions of effort; and patterns in operational characteristics such as HBF and hooks per set. We examined patterns through time and among species in both the nominal catch rates and 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 through time, to identify possible changes in fishing behaviour or population composition.

To further explore changes in the fishery and identify periods of change, we plotted the participation of vessels in the fleet, sorted first by the start date and then by the end date of participation in the fishery.

Several approaches were used to explore changes in effort distribution and concentration through time. For each statistical area and for each year, we plotted the numbers of 5°x5° and 1°x1° cells fished and the average number of operations per fished cell. We defined two separate core SBT fishing areas: with statistical areas 9 in the west from March-October, and statistical area 8 in the east from July-December.

Indices of fishing effort concentration were also calculated, including the Gini coefficient (Gini 1912) and Gulland's index of concentration (Gulland 1956). The Gini coefficient is best known as an indicator of wealth concentration, but can be used to measure aggregation of any quantity. We use it to estimate the spatial aggregation of the catch of each species, and effort, in each region. A higher Gini coefficient indicates that more of the catch (or effort) is being taken from fewer spatial cells. We estimated values separately for each year, where the values y_i are catches or effort per 5° x 5° cell, ranked from lowest to highest, and including zeroes for unfished cells. Cell areas are assumed to be uniform.

$$Gini = \frac{2\sum_{i=1}^{n} iy_i}{n\sum_{i=1}^{n} y_i} - \frac{n+1}{n}$$

Gulland's index of concentration measures the extent to which a fleet has concentrated its fishing effort in areas with higher than average catch rates (Harley 2009). The weighted version of the index is calculated as follows, where y_i is the catch in the *i*th stratum, e_i is the effort in the *i*th stratum, and N is the number of exploited strata. from year to year depending on both the distribution of the effort, and the distribution of the catch rates. If effort is evenly distributed with respect to catch rate then the index will average 1, whereas it will be higher than 1 if effort is preferentially targeted to areas with higher than average catch rate (Hoyle 2014).

$$Gulland = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} e_i} . \frac{1}{\sum_{i=1}^{N} \frac{y_i}{e_i N}}$$

This index varies from year to year depending on both the distribution of the effort, and the distribution of the catch rates. If effort is evenly distributed with respect to catch rate then the index will average 1, whereas it will be higher than 1 if effort is preferentially targeted to areas with higher than average catch rate.

Given the spatial and seasonal separation of fishing in these two areas, and potentially different size distributions, we standardized data separately for each area.

Data from the period 1996-2016 were used in CPUE standardizations. Data prior to 1996 were not used in this study as they included insufficient reliable data from vessels targeting SBT.

CPUE standardization methods generally followed the approaches used by Hoyle and Okamoto (2011) and Hoyle, Okamoto et al. (2015), with some modifications. Parts of the methods text below are the same as these articles. R code is also used from examples presented in Hoyle, Langley et al. (2014).

Target change

Target change can be a significant problem for CPUE standardization since it can bias CPUE trends. Analyses were carried out using two main approaches to address target change. The first approach addressed targeting by removing effort considered unlikely to have targeted southern bluefin tuna. The second approach used cluster analysis to separate effort into groups that may have used different targeting methods, and included the categorical cluster variable in the standardization model.

The first approach used a data selection process. Data were cleaned by removing sets in which HBF was less than 9 or greater than 12. After examination, the data were cleaned to focus on the periods in which most SBT were caught in each statistical area. Effort for area 8 was included for the months July to December; and effort for area 9 was included for March to October.

Clustering

In the second approach, we clustered all data for areas 8 and 9 using the approach applied by Hoyle, Okamoto et al. (2015). We removed all sets with no catch of any of the species, and then aggregated by vessel-month. Set level data contains variability in species composition due to the randomness of chance encounters between fishing gear and schools of fish. This variability leads to some misallocation of sets using different fishing strategies. Aggregating the data tends to reduce the variability, and therefore reduce misallocation of sets. For these analyses we aggregated the data by vessel-month, assuming that individual vessels tend to follow a consistent fishing strategy through time. One trade-off with aggregation in this way is that vessels may change their fishing strategy within a month, which will result in misallocation of sets. For the purposes of this paper we refer to aggregation by vessel-month as trip-level aggregation, although the time scale is (for distant water vessels) in most cases shorter than a fishing trip.

We calculated proportional species composition by dividing the catch in numbers of each species by catch in numbers of all species in the vessel-month. Thus, the species composition values of each vessel-month summed to 1, ensuring that large catches and small catches were given equivalent weight. The data were transformed by centring and scaling, to reduce the dominance of species with higher average catches. Centring was performed by subtracting the column (species) mean from each column, and scaling was performed by dividing the centred columns by their standard deviations.

We clustered the data using the hierarchical Ward hclust method, implemented with function *hclust* in R, option 'Ward.D', after generating a Euclidean dissimilarity structure with function *dist*. This approach differs from the standard Ward D method which can be implemented by either taking the square of the dissimilarity matrix or using method 'ward.D2' (Murtagh and Legendre 2014). However in practice the method gives similar patterns of clusters to other methods, more reliably than ward.D2 (Hoyle et al 2015).

Data were also clustered using the kmeans method, which minimises the sum of squares from points to the cluster centres, using the algorithm of Hartigan and Wong (1979). It was implemented using function *kmeans* in the R stats package (R Core Team 2016).

Selecting the number of groups

We used several subjective approaches to select the appropriate number of clusters. In most cases the approaches suggested the same or similar numbers of groups. First, we considered the number of major targeting strategies likely to appear in the dataset, based on understanding and exploration of the data. Second, we applied hclust to transformed trip-level data and examined the hierarchical trees, subjectively estimating the number of distinct branches. Third, we ran kmeans analyses on untransformed trip-level data with number of groups *k* ranging from 2 to 25, and plotted the deviance against *k*. The optimal group number was the lowest value of *k* after which the rate of decline of deviance became slower and smoother. Finally, following Winker et al (2014) we applied the *nScree* function from the R nFactors package (Raiche and Magis 2010), which uses various approaches (Scree test, Kaiser rule, parallel analysis, optimal coordinates, acceleration factor) to estimate the number of components to retain in an exploratory PCA. Where there was uncertainty about the number of clusters, we selected the option with more clusters.

Plotting and data selection

We plotted the hclust clusters to explore the relationships between them and the species composition and other variables, such as HBF, number of hooks, year, and set location. Plots included boxplots of a) proportion of each species in the catch, by cluster; b) the distributions of variables by cluster; and c) maps of the spatial distribution of clusters, one map for each cluster.

GLM analyses

The operational data were standardized using generalized linear models in Microsoft R Open 3.3.2 (R Core Team 2016). Analyses were conducted separately for each of the two core areas, and for each of the two target change methods.

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.

Analyses were carried out using generalized linear models that assumed a lognormal distribution with an added constant. The following model, which we call the lognormal constant GLM, was used:

$$\ln(CPUE_s + k) \sim year + vessid + latlong + \lambda(hooks) + g(month) + h(moon) \qquad (1)$$

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 10, 4, 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 cell) were fitted as categorical variables.

For the clustering based approach, models also included a categorical variable for the cluster.

Data in the lognormal constant GLMs were 'area-weighted', with the weights of the sets adjusted so that the total weight per year-quarter in each 5° cell 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-qtr *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.

The effects of covariates were examined in exploratory analyses by using the package influ (Bentley, Kendrick et al. 2011) to show the influence of each covariate.

Changes in catchability through time were investigated by fitting to the operational data both with and without a term for individual vessel. The two models were designated respectively the 'base model' and the 'vessel-effects model'. Abundance indices were calculated for each model, and normalized to average 1. The indices estimated for each year-quarter were compared by dividing the base model by the vessel effects model, plotting the time series of ratios, and fitting a log-linear regression. The slope of the regression represented the average annual compounding rate of change in fishing power attributable to changes in the vessel identities; i.e. the introduction of new vessels and retirement of old vessels. Gradients are shown on the figures, together with confidence intervals.

Indices of abundance were obtained by running the lognormal constant GLM model with the standard settings, including vessel effects. Time effects were obtained by predicting the expected catch rate for each year, for (across all years) the vessel, month and cell with the most sets, lunar illumination of 0.5, and the median number of hooks. The uncertainty associated with the year effect was used as the measure of uncertainty. Indices were normalised by dividing through by the mean of the year effects, giving relative CPUE with mean of 1.

RESULTS AND DISCUSSION

Data exploration

Almost all effort used between 9 and 12 hooks per float (HBF) (Figure 5), while the majority of HBF outside this range came from north of 35S, 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 hooks per set (Figure 6).

Mean catch rates by species in the southern statistical areas 7, 8, and 9 are highest for southern bluefin tuna until the mid-2000s. After this time in area 9 SBT catch rates decrease and other species increase, some higher catch rates of albacore catch rates increase, particularly albacore. However, in the most recent years the SBT catch rates are again higher than other species (Figure 7). Similarly, the proportion of sets reported with zero SBT catches was low through most of the time series in the southern areas 7 to 9 (Figure 8), but area 9 shows an increase in the proportion of zeroes from 2004 to 2010. The majority of sets reported no yellowfin catch, and the same applied to bigeye and albacore. However, in area 9 there was a decrease in the proportions of zero albacore catches from about 2000-2010, followed by an increase in the late 2000's. There may have been some albacore targeting in area 9 during this period.

In the northern statistical areas 13 and 1, the tropical bigeye and yellowfin tunas dominate with the highest catch rates, along with albacore. Southern bluefin tuna catch rates are low throughout the time series, despite being inflated due to the selection of data only from vessels that report at least one SBT in the month. The existence of zero SBT catch rates is likely due to vessels being included due to reporting SBT catch during the month in a different statistical area, though some may be due to effort with SBT catch being removed during the cleaning process. Most sets in these areas catch no SBT (Figure 8), and there are few sets with zero catches of bigeye or yellowfin, while intermediate numbers of sets report no albacore catch. Given the low rate of SBT capture in the northern areas, misreporting, species misidentification, and data errors may be a concern, and could explain a sizable proportion of the observed catches.

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 in the western area 14, and bigeye and albacore in eastern area 2. Since the mid-2000s albacore 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 fleet in these statistical areas.

The proportions of zero catches for both sharks and other species reduced substantially between 1990 and 2014, particularly in southern areas 8 and 9. Sharks and other species have always been caught in most sets, so this change may be entirely due to increased reporting rather than increased catches, linked to stronger requirements to report catches of bycatch species.

We mapped the species composition of catch (proportion of SBT in the catch of all species) south of 30S by 5-year period (Figure 10). The proportion of SBT in the catch was high in all periods, increasing further south, but declined steadily in all areas after 2000. This partly reflects targeting of other species, but also reflects increased reporting of sharks and other species. In the post-2010 period there is little SBT taken in statistical area 8 north of about 37S, whereas a high proportion of the catch in this area is albacore (Figure 11). It is apparent from Figures 12 and 13 that this spatially differentiated targeting in area 8 also has a temporal aspect, with albacore targeting April-July, and SBT targeting July-December. In area 9 to the west, there is less spatial or temporal separation of SBT and ALB catch, with both species caught in the months March-October (Figures 12 and 13).

Only 38 Korean vessels have participated in the area 8 and 9 fishery since 1995 (Figure 14), with about half of the total number reporting their first participation before 2000. Arrival of new vessels has been slow but steady. Eight vessels stopped participating in 2009, and five more have stopped since then, but six others have joined the fishery in that time.

The total number of major (5° x 5° x month) cells fished has varied from year to year (Figure 15), but has declined considerably since the peak year in 2008. 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 (Figure 16). However, the distribution of effort within major cells has not concentrated significantly, with similar numbers of minor cells per major cell on average. In the period since 2008 the timing of effort in areas 8 and 9 has changed, gradually moving earlier in the year, though with different timing peaks in each area (Table 3).

Gulland's index of concentration indicates whether effort is concentrated in areas of high or low catch rate for a species, but estimates can be variable and uncertain where sample sizes are small. Plots for SBT in areas 8 and 9 suggest that effort is generally higher in areas with higher SBT catch rate, since most points are above 1 (Figure 17). The results for bigeye and yellowfin are considerably more variable, possibly reflecting the lower catch rates.

Gini coefficients are widely used in many fields to measure the distribution of quantity – with uniform to very uneven distributions represented by low to high Gini coefficients. Estimates for regions 8 and 9 and for SBT, bigeye and albacore tuna, and for effort, show similar patterns, with increasing concentration through time.

Target change

The data selection approach aimed to identify effort targeted mostly at southern bluefin tuna, by selecting area 9 data from March-October and area 8 data from July-December. This approach appeared to give reasonable results, but was not entirely successful, as indicated by the high proportions of zero catches in area 9 between 2004 and 2010 (Figure 19).

Applying Ward's D hierarchical cluster analysis at the vessel-month identified strong separation among 2 to 3 groups in statistical areas 8 and 9 (Figure 20).

In area 9, clusters 1, 2, and 3 were more strongly represented in the early, middle, and later parts of the time series (Figure 21). Clusters 1 and 3 occur mostly in the period before August, while cluster 2 extends into October. Cluster 1 also has slightly fewer hooks between floats and cluster 3 has more. Median number of hooks is higher in cluster 3. Cluster 1 dominates the southeast of area 9, while cluster 2 dominates the northwest, and cluster 3 the southwest (Figure 22). The species composition of cluster 1 comprises almost entirely southern bluefin tuna, with small amounts of albacore and bigeye tuna (Figure 23). Cluster 2 has the least southern bluefin of all clusters, with similar amounts of albacore, and also some bigeye and yellowfin tuna. Cluster 3 includes significant southern bluefin tuna along with similar amounts of 'other' species, along with some sharks and albacore tuna.

In area 8, cluster 1 dominates the early part of the time series, with clusters 2 and 3 more apparent after 2005 (Figure 21). 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 also represented during March to June. Cluster 1 is well represented across most of the fished area, while cluster 2 occurs at middle latitudes from about 38-43S. Cluster 3 occurs almost entirely in the far north of the area (Figure 22). Cluster 1 is dominated by southern bluefin tuna, with little reporting of other species (Figure 23).

Cluster 2 reports similar amounts of southern bluefin tuna, sharks, and ‘other’ species. Cluster 3 reports more albacore than southern bluefin tuna, with ‘other’ species, yellowfin and bigeye tuna also reported.

CPUE standardization

The models were initially applied to data from the first target change method, using the data selection process. Table 1 shows the results of dropping each variable from the lognormal constant GLMs, indicating that all explanatory variables were statistically significant, with the year, location, and month effects the largest factors affecting the nominal CPUE. It is common in CPUE standardizations for all variables to be statistically significant. However, lack of independence is to be expected in observational fisheries data, and tends to result in overestimation of statistical significance.

Patterns were broadly similar for both approaches to addressing target change (Figure 24). There were differences in the 2004-2006 period in area 9, with standardized catch rates higher for the clustered analysis. The clustering approach may be accounting better for the apparent switch towards targeting albacore during this time. The presence of more zero SBT catches in area 9 from 2004-2010 suggests that the data during that period may include more effort targeted at other species. Such ‘contamination’ of the effort would tend to bias the indices low, and this is a period when the indices are the lowest in the time series. It is therefore useful to separate sets with different fishing strategies. Applying cluster analysis to separate may be the best approach for this period.

Comparison of standardized and unstandardized CPUE series shows them to be quite similar. The largest change is for the area 8 indices in the most recent years, where the ‘selected data’ standardized indices first dips below the nominal and then climbs above it. The ‘clustered data’ indices are more stable overall, but also increase strongly in 2016.

The influence plots (Figures 25-29) showed the patterns of the parameter estimates at the top of each plot, and the influence of each parameter on the year effect on the right side of each plot. Note that the influence scales (bottom right) differ among plots. The influences of all variables are summarised in Figure 31. To save space the ‘selected data’ plots are presented except for the ‘hcltrp’ variable, which only occurs in the ‘clustered data’ analysis. .

Vessel effects (Figure 25) were quite variable, with a few vessels having significantly lower SBT catch rates. On average, the influence of vessel effects raised the average catchability at the end of the time series in both areas, but the low number of vessels resulted in significant variability.

Spatial effects (Figure 26) showed significant variation in catch rates, with more variation in area 9 than area 8. In area 9 there was a strong trend through time towards fishing in areas with lower average catch rates, particularly in the last two years. This trend has contributed to the higher catch abundance indices in 2014 and 2015. It would be useful to explore whether the areas of highest catch rate have moved through time. However, this would be difficult to determine since fishing activity is currently very concentrated spatially. Given the behaviour of the species, areas of highest catch rate are also likely to move during the year, and we do not account for this in the model. This is likely to increase uncertainty in the indices.

The effects of the number of hooks per set on catch rates (Figure 27) were difficult to interpret. In eastern area 8 there were relatively small differences by hook number across the range of data with most hooks, and minimal influence on year effects, apart from 2016 when effort was low and localised with only one vessel fishing. In area 9 there were larger differences, and apparent influence on the year effects, with catchability averaging about 5% above the mean in 2010-16. Sets with more than about 3150 hooks tended to catch more SBT than sets with fewer hooks. This may reflect a mixture of targeting methods in area 9, with different fishing methods using different numbers of hooks. In area 9 there were more sets with fewer hooks between 2004 and 2007, a period during which there were more zero SBT sets than at most other times.

The effect of month was strong in both the eastern and western areas (Figure 28). In both areas, the highest catch rates were obtained in July and August. The seasonality of fishing effort changed through time, with the model suggesting that mean catchability in area 8 was over 10% higher than the average in 2014-15, and almost 20% below average in area 9 in 2010-11 and 2013-15.

Catch rates appeared to vary moderately with lunar illumination (Figure 29). Longline catch rates of other pelagic fish such as bigeye tuna are known to be affected by moon phase (Poisson, Gaertner et al. 2010). The patterns we observed differed between the two areas, and may be artefacts of lack of independence in the data. Fishing effort is distributed relatively evenly across all phases of the moon, so moon phase has almost no influence on the year effects.

The distribution of the cluster variable *hcltrp* changed considerably though time, as the behaviour of the fleet changed with the abundance of the target species (Figure 30). There were also relatively large differences in catch rate between the clusters, so this variable was quite influential on the indices.

The summary plots showing the effects of all the influence plots shows how many factors affect the index trends (Figure 31). The variables have mostly similar effects in the ‘selected data’ and ‘clustered data’ analyses.

Like the abundance indices, the influence estimates are conditional on the model, which assumes that there are no interactions between the different effects. However, interactions may be expected, such as variation between years in the timing and location of higher catch rates, due to environmental variation affecting tuna movements. The small sample sizes limit the ability to model interaction terms, and there was limited time available to explore alternative models, but this would be worthwhile in future analyses.

Diagnostic frequency distributions and QQ-plots (Figure 32) suggest that the data fitted the GLM adequately.

Patterns in the indices (Figure 33) differ somewhat between east and west. Both sets of indices decreased until the mid-2000s, and subsequently increased, particularly in the last few years. Lack of data prevents the estimation for eastern area 8 from 2003-2007 (Table 2). The recent effort in region 8 is so low and concentrated that the estimates are likely to be highly variable, and not very reliable. Effort in region 9 is also becoming more concentrated, so indices are also likely to be getting less reliable.

This was the first application of cluster analysis to the data used in the southern bluefin tuna model. these results suggest that the approach may be useful, since it identified patterns that are consistent with our understanding of the fishery, and it changed the indices during the 2004-2006 period in a way that seems likely to be an improvement. However, it made other changes to the indices which are hard to characterise as improvements or otherwise, particularly given the uncertainty in all estimates. It seems reasonable to prefer the ‘clustered data’ indices, but either set of indices might be considered.

The ratios of analyses with and without vessel effects did not suggest significant trends in fishing power. Confidence limits on the trend estimates were wide relative to the potential effects, as expected since sample sizes and the numbers of vessels involved were small. Trends in fishing power estimated this way represent the effects of changes in the fleet composition. They do not account for changes in fishing power caused by vessels that stay in the fishery and change their equipment or their fishing behaviour.

REFERENCES

Bentley, N., T. H. Kendrick, P. J. Starr and P. A. Breen (2011). "Influence plots and metrics: tools for better understanding fisheries catch-per-unit-effort standardizations." ICES Journal of Marine Science **69**(1): 84-88.

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.

Gini, C. (1912). "Variabilità e mutabilità." Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi **1**.

Gulland, J. A. (1956). "On the fishing effort in English demersal fisheries." Fisheries Investigations Series II **20**(5): 41 pp.

Harley, S. (2009). Spatial distribution measures for the analysis of longline catch and effort data**:** 17.

Hartigan, J. A. and M. A. Wong (1979). "Algorithm AS 136: A k-means clustering algorithm." Journal of the Royal Statistical Society.Series C (Applied Statistics) **28**(1): 100-108.

Hoyle, S. D. (2014). Spatial considerations in bigeye and yellowfin CPUE from Japanese and Taiwan, China longline fisheries in the Indian Ocean, IOTC–2014–WPTT16–27 Rev_1. Working Party on Tropical Tunas 16, Bali, Indonesia, Indian Ocean Tuna Commission.

Hoyle, S. D., A. D. Langley and R. A. Campbell (2014). "Recommended approaches for standardizing CPUE data from pelagic fisheries."

Hoyle, S. D. and H. Okamoto (2011). Analyses of Japanese longline operational catch and effort for bigeye and yellowfin tuna in the WCPO, WCPFC-SC7-SA-IP-01. Western and Central Pacific Fisheries Commission, 7th Scientific Committee. Pohnpei, Federated States of Micronesia.

Hoyle, S. D., H. Okamoto, Y.-m. Yeh, Z. G. Kim, S. I. Lee and R. Sharma (2015). IOTC–CPUEWS02 2015: Report of the 2nd CPUE Workshop on Longline Fisheries, 30 April – 2 May 2015, Indian Ocean Tuna Commission**:** 126.

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

Murtagh, F. and P. Legendre (2014). "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?" Journal of Classification **31**(3): 274-295.

Poisson, F., J.-C. Gaertner, M. Taquet, J.-P. Durbec and K. Bigelow (2010). "Effects of lunar cycle and fishing operations on longline-caught pelagic fish: fishing performance, capture time, and survival of fish." Fishery Bulletin **108**: 268-281.

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

R Core Team (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

Raiche, G. and D. Magis (2010). "nFactors: Parallel analysis and non graphical solutions to the Cattell Scree Test." R package version **2**(3).

Winker, H., S. E. Kerwath and C. G. Attwood (2014). "Proof of concept for a novel procedure to standardize multispecies catch and effort data." Fisheries Research **155**: 149-159.

TABLES

Table 1: Degrees of freedom, Deviance, and delta AIC results from lognormal (CPUE + k) GLMs for statistical areas 8 and 9.

	Stat area 9			Stat area 8		
	Df	Deviance	ΔAIC	Df	Deviance	ΔAIC
<none>		143.7	0		36.3	0
Year	20	169.7	2294	14	42.0	1254
Latlong	18	156.2	1133	10	37.9	364
ns(hooks, 10)	10	146.7	278	10	36.6	71
Vessid	28	155.2	1024	21	37.2	179
ns(month, df = 4	4	154.8	1035	4	37.5	281
ns(moon, df = 4)	4	143.9	13	4	37.1	194

Table 2: Lognormal constant indices for statistical areas 9 and 8, for selected data (left) and clustered data (right).

	Selected data				Clustered			
Year	Stat area 9	CV	Stat area 8	CV	Stat area 9	CV	Stat area 8	CV
1996	1.21	0.04	0.95	0.04	0.97	0.04	1.23	0.04
1997	0.93	0.03	0.66	0.02	0.96	0.03	0.81	0.03
1998	0.88	0.03	0.73	0.02	0.81	0.03	0.89	0.03
1999	0.96	0.03	0.63	0.02	0.83	0.03	0.80	0.03
2000	0.81	0.03	0.65	0.02	0.70	0.03	0.82	0.03
2001	0.96	0.03	0.71	0.03	0.86	0.03	0.81	0.03
2002	0.94	0.03	0.46	0.02	0.82	0.03	0.52	0.03
2003	0.70	0.04	NA	NA	0.75	0.03	NA	NA
2004	0.40	0.03	NA	NA	0.49	0.03	NA	NA
2005	0.19	0.05	NA	NA	0.27	0.05	NA	NA
2006	0.55	0.04	NA	NA	0.72	0.04	NA	NA
2007	0.48	0.03	NA	NA	0.57	0.03	NA	NA
2008	0.90	0.03	1.00	0.02	0.83	0.03	1.07	0.02
2009	0.72	0.03	0.68	0.03	0.85	0.03	0.66	0.03
2010	0.76	0.03	0.75	0.02	0.84	0.03	0.75	0.03
2011	1.86	0.05	0.98	0.02	1.68	0.04	1.00	0.03
2012	1.57	0.03	1.14	0.03	1.73	0.03	1.21	0.04
2013	1.26	0.05	1.50	0.04	1.28	0.05	1.12	0.04
2014	2.17	0.05	1.69	0.05	2.16	0.05	0.97	0.05
2015	1.27	0.06	1.00	0.04	1.30	0.06	1.04	0.05
2016	1.49	0.05	2.45	0.06	1.57	0.05	2.29	0.06

Table 3: Numbers of operations per year, month and area. %10Y, %5Y, and %2016 denote the proportions of each area-month to the 10 year (2007-2016), 5 year (2012-2016), and 2016 totals, respectively. Where the %2016 is 3% larger or 3% smaller than the %5Y it is marked with grey shading or an outline, respectively.

		Year															
Area	**Month**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**%10Y**	**%5Y**	**%2016**
2	3	0	0	0	0	0	0	0	1	0	0	0	0	0	0.0%	0.0%	0.0%
	4	0	0	0	0	0	0	0	30	1	0	4	0	0	0.3%	0.1%	0.0%
	5	0	0	0	0	0	0	7	8	0	0	23	0	0	0.3%	0.4%	0.0%
	6	0	3	0	0	17	0	14	26	8	72	49	0	0	1.5%	2.5%	0.0%
	7	0	22	0	0	30	0	27	48	19	91	46	45	0	2.4%	3.9%	0.0%
	8	0	17	0	0	19	0	0	0	0	0	0	0	0	0.2%	0.0%	0.0%
	9	0	0	0	0	0	6	25	0	0	2	9	5	0	0.4%	0.3%	0.0%
	10	0	0	0	0	0	3	24	0	4	0	0	0	0	0.2%	0.1%	0.0%
	11	0	0	0	0	7	0	12	0	0	0	0	0	0	0.2%	0.0%	0.0%
8	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0.0%	0.0%	0.0%
	3	0	0	0	0	0	0	0	8	0	0	0	0	0	0.1%	0.0%	0.0%
	4	0	3	0	0	0	0	0	23	27	0	44	0	0	0.7%	1.4%	0.0%
	5	0	0	0	0	0	0	20	71	30	115	33	0	0	2.1%	3.5%	0.0%
	6	0	0	0	0	0	0	12	51	3	4	4	0	0	0.6%	0.2%	0.0%
	7	0	0	0	0	46	19	10	60	0	16	8	11	0	1.3%	0.7%	0.0%
	8	0	0	0	0	132	181	167	185	115	143	168	115	29	9.8%	11.1%	2.7%
	9	0	0	0	0	189	214	166	179	191	130	95	106	29	10.3%	10.8%	2.7%
	10	0	0	0	0	252	180	79	207	139	103	0	0	23	7.8%	5.2%	2.2%
	11	0	0	0	0	169	165	73	190	0	23	0	0	0	4.9%	0.4%	0.0%
	12	0	0	0	0	98	36	0	75	0	0	0	0	0	1.7%	0.0%	0.0%
9	3	62	5	0	64	6	19	0	0	87	0	0	0	0	1.4%	1.7%	0.0%
	4	103	20	6	127	76	258	172	72	125	97	136	122	54	9.8%	10.4%	5.1%
	5	69	9	18	130	189	329	173	86	130	115	168	194	246	13.9%	16.7%	23.0%
	6	63	3	5	125	198	297	138	65	156	105	152	137	305	13.3%	16.7%	28.5%
	7	45	34	34	169	144	186	69	17	122	22	83	30	270	8.8%	10.3%	25.3%
	8	75	21	133	138	71	60	2	0	0	0	0	0	91	2.9%	1.8%	8.5%
	9	71	24	41	154	0	8	0	0	0	0	0	0	22	1.5%	0.4%	2.1%
	10	119	0	17	142	2	0	0	0	0	0	0	0	0	1.1%	0.0%	0.0%
	11	44	0	0	1	0	0	0	0	0	0	0	0	0	0.0%	0.0%	0.0%
14	3	0	0	0	20	0	41	0	0	15	0	0	0	0	0.6%	0.3%	0.0%
	4	4	0	13	0	0	7	0	0	0	0	0	29	0	0.3%	0.6%	0.0%
	5	0	8	0	0	0	0	0	0	0	0	0	0	0	0.0%	0.0%	0.0%
	6	16	84	22	0	2	16	37	0	0	0	0	0	0	0.4%	0.0%	0.0%
	7	68	79	87	16	0	30	13	0	0	11	0	0	0	0.6%	0.2%	0.0%
	8	4	14	8	20	0	0	0	0	0	0	0	0	0	0.2%	0.0%	0.0%
	9	21	39	9	1	4	5	0	0	0	0	1	0	0	0.1%	0.0%	0.0%
	10	13	0	3	7	1	0	0	0	6	0	0	0	0	0.1%	0.1%	0.0%
	11	5	0	0	34	16	0	0	0	0	2	0	0	0	0.4%	0.0%	0.0%
	12	0	0	0	0	2	1	0	12	0	0	0	0	0	0.1%	0.0%	0.0%
		782	385	396	1148	1670	2065	1240	1414	1178	1051	1023	794	1069	12652	5115	1069

FIGURES

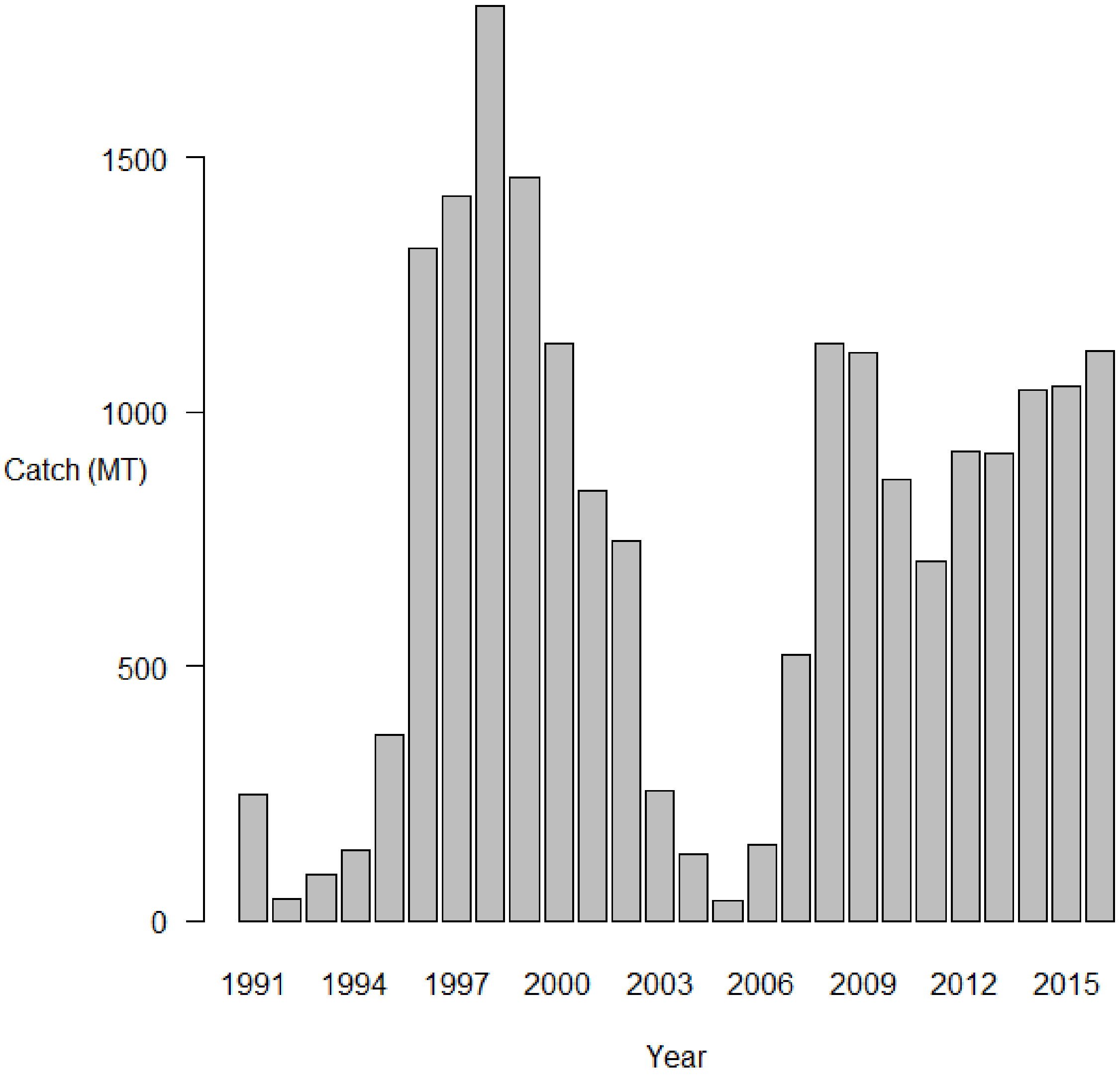


Figure 1: The annual Korean SBT catches in the CCSBT convention area, 1991 - 2016.

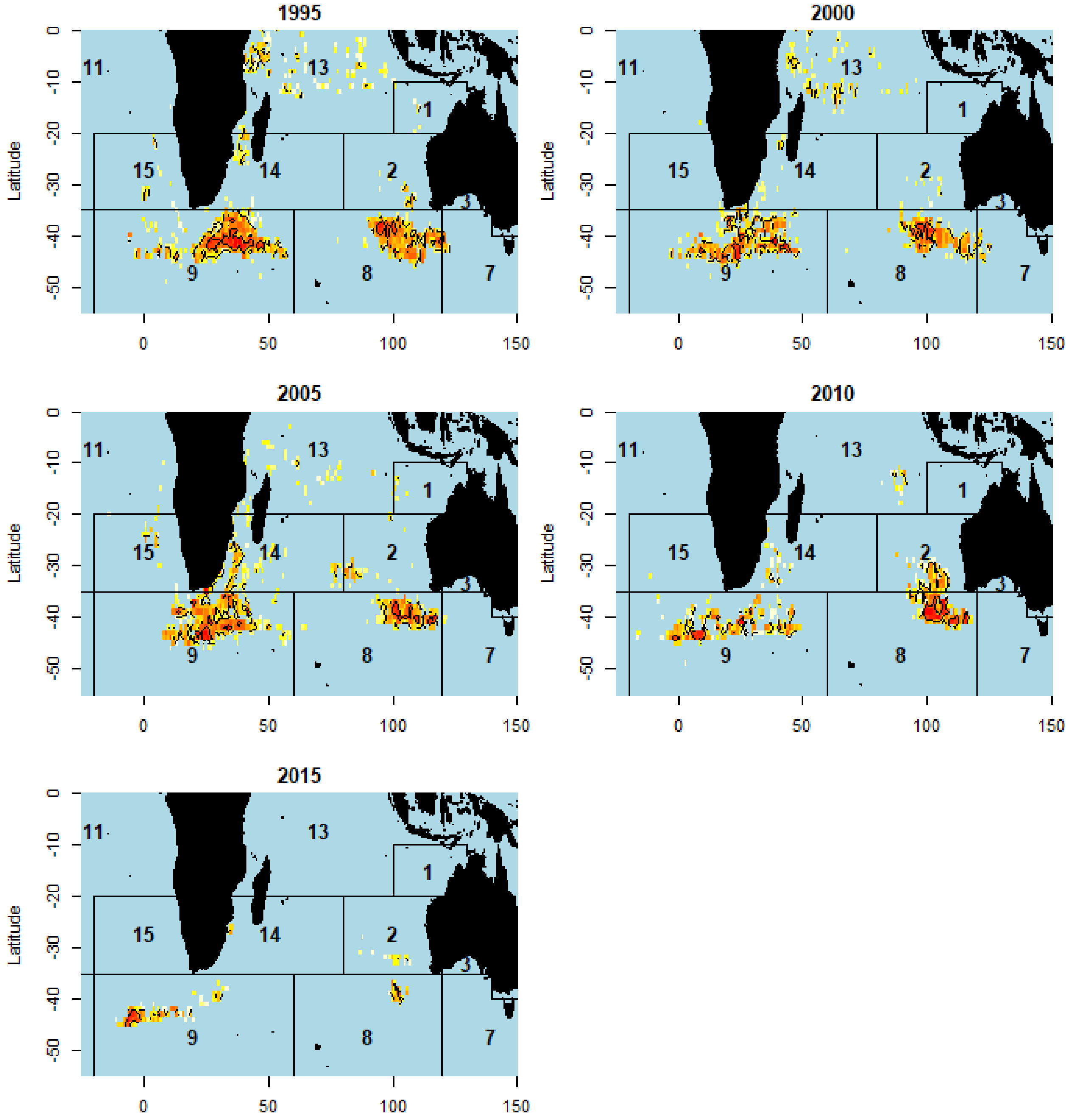


Figure 2: Map showing the core areas of Korean tuna longline vessels fishing for SBT, aggregated by 5 year period. Red colour indicates higher fishing effort, n numbers of hooks.

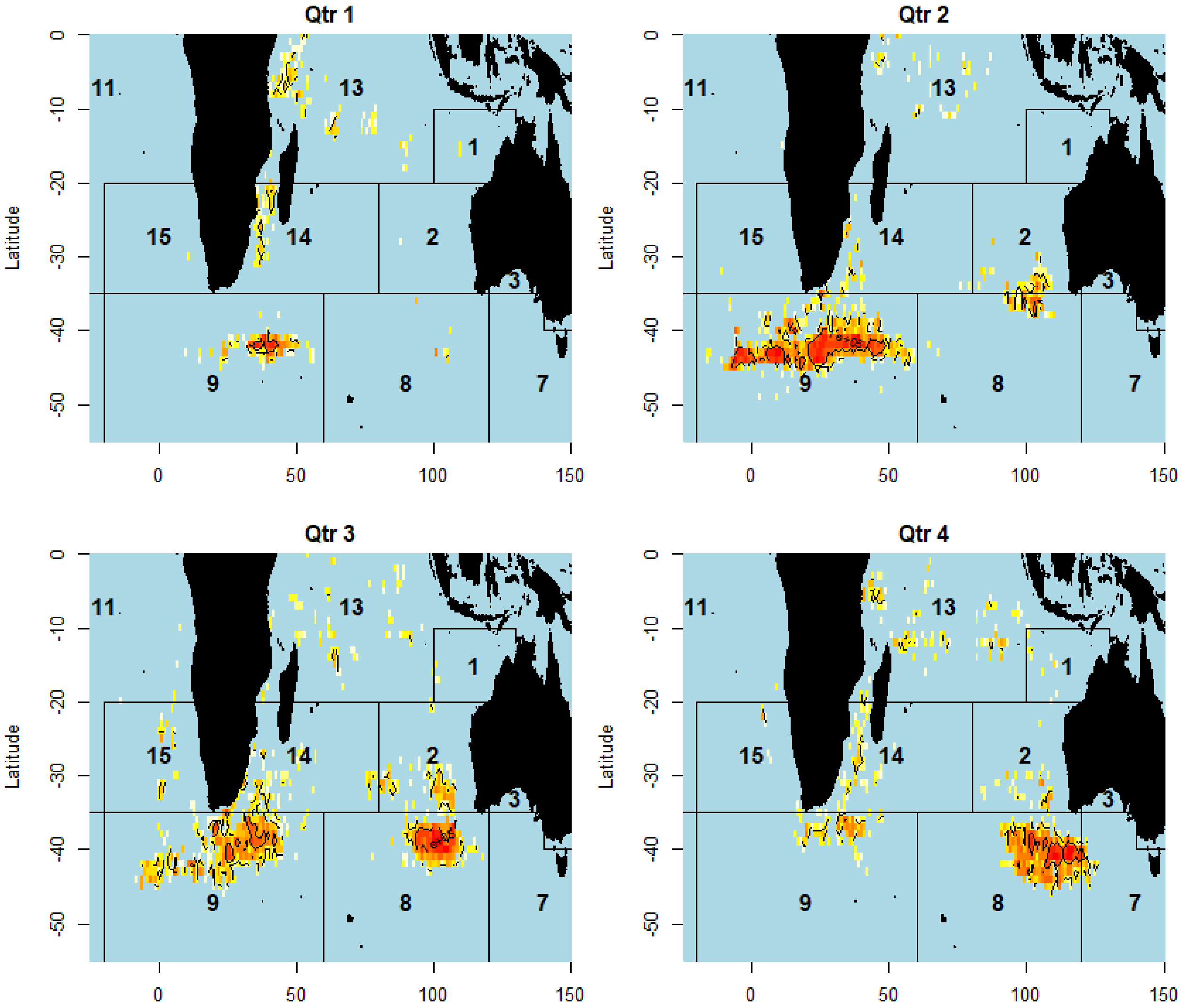


Figure 3: Map showing the core areas of Korean tuna longline vessels fishing for SBT, by quarter. Red colour indicates higher fishing effort, in numbers of hooks. Data are aggregated across the period 1994-2014.

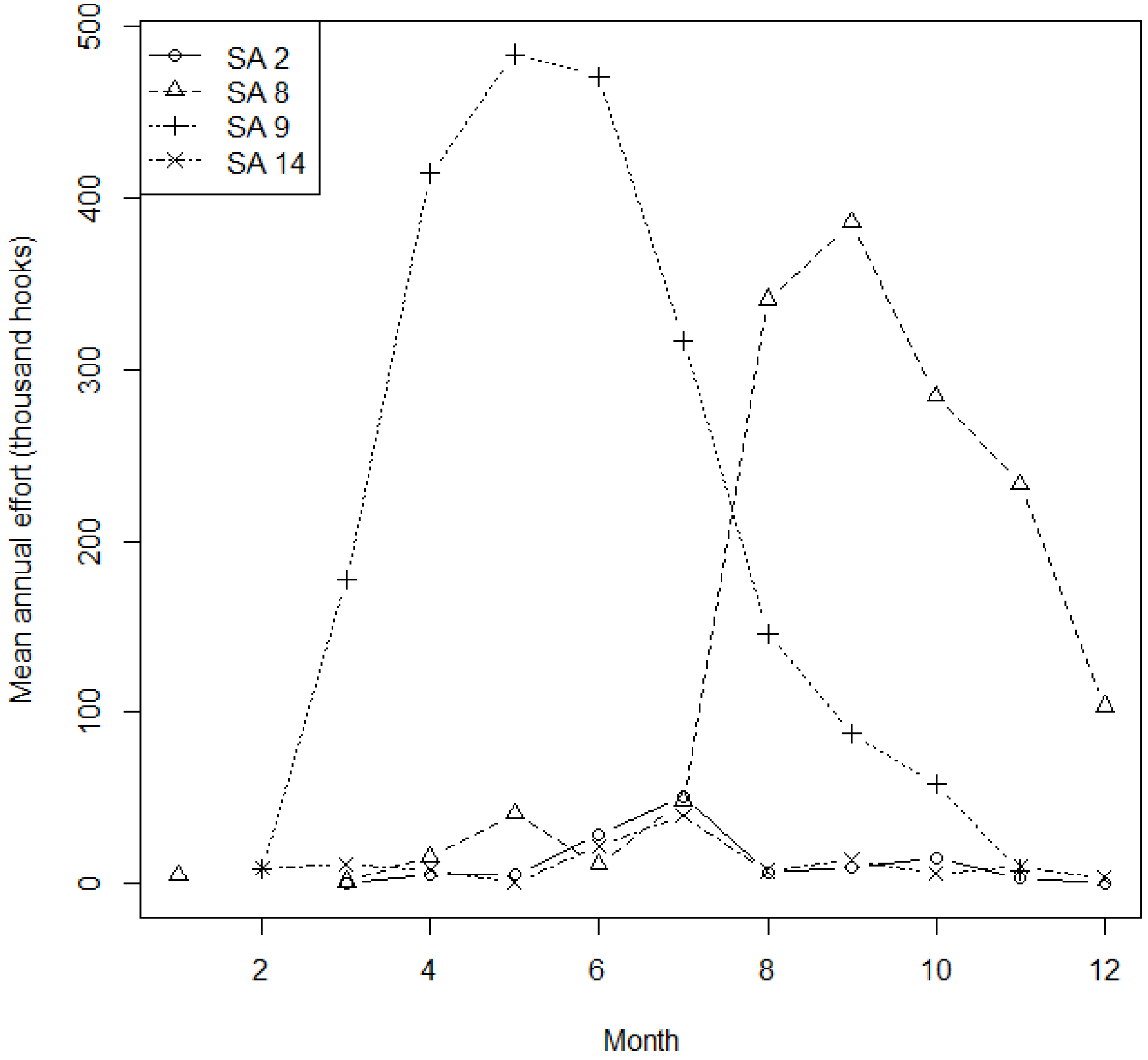


Figure 4: Mean annual effort in thousands of hooks, by month and statistical area.

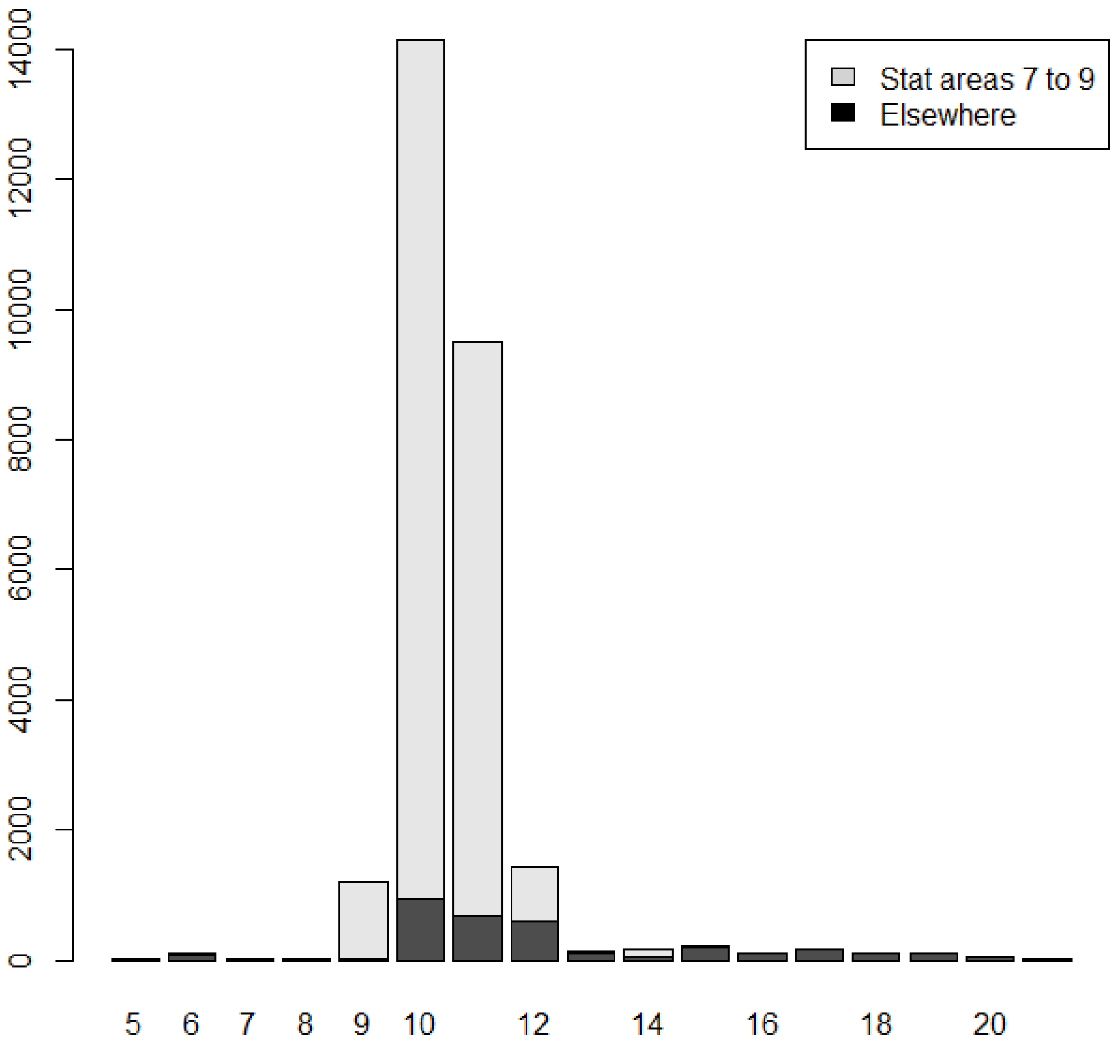


Figure 5: Frequency table of HBF for the main fishing ground with the lighter shade for statistical areas 7-9, and the darker shade for other areas.

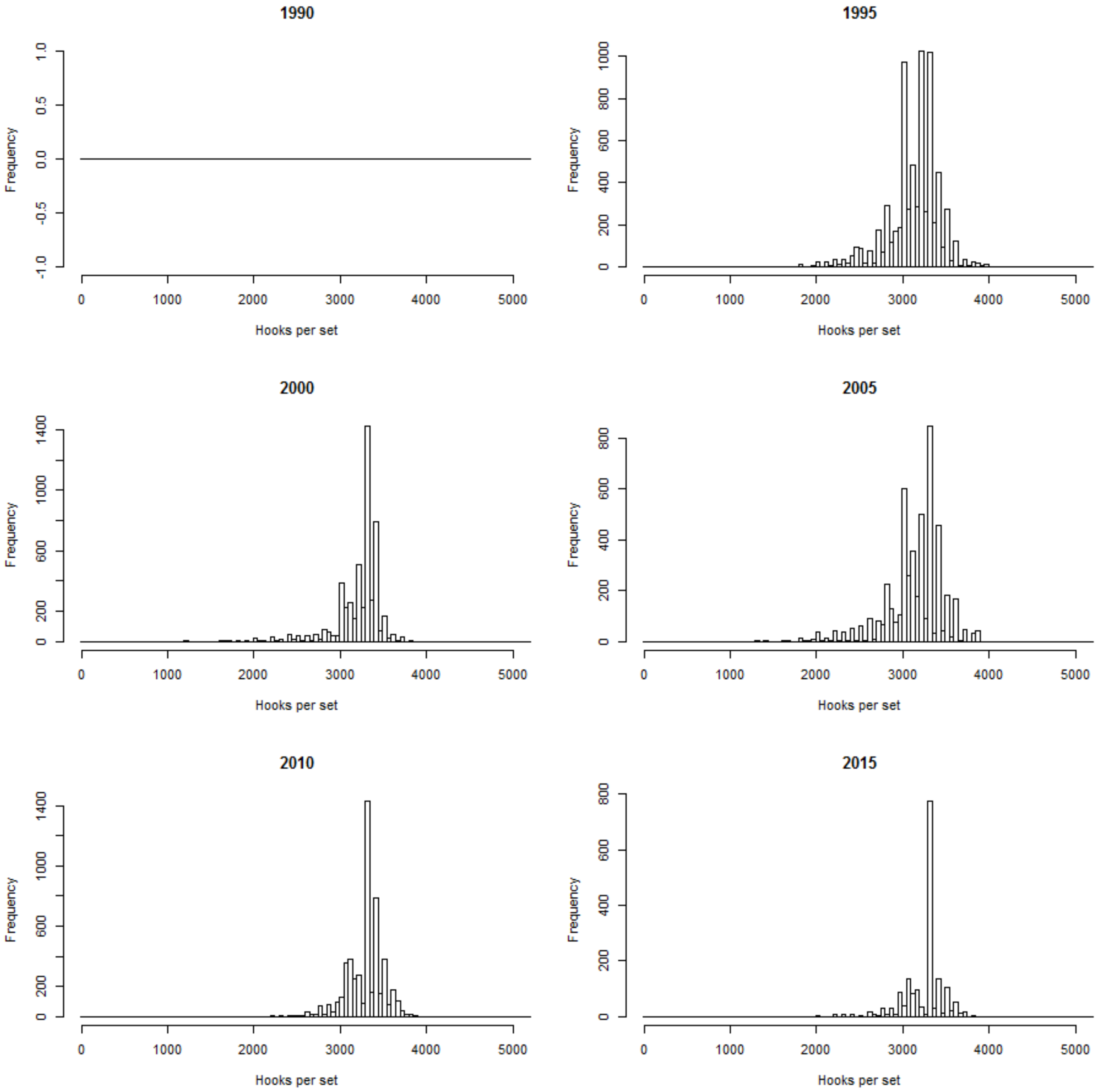


Figure 6: Distribution of hooks per set per 5 year period, for sets in the SBT-targeting areas south of 35S.

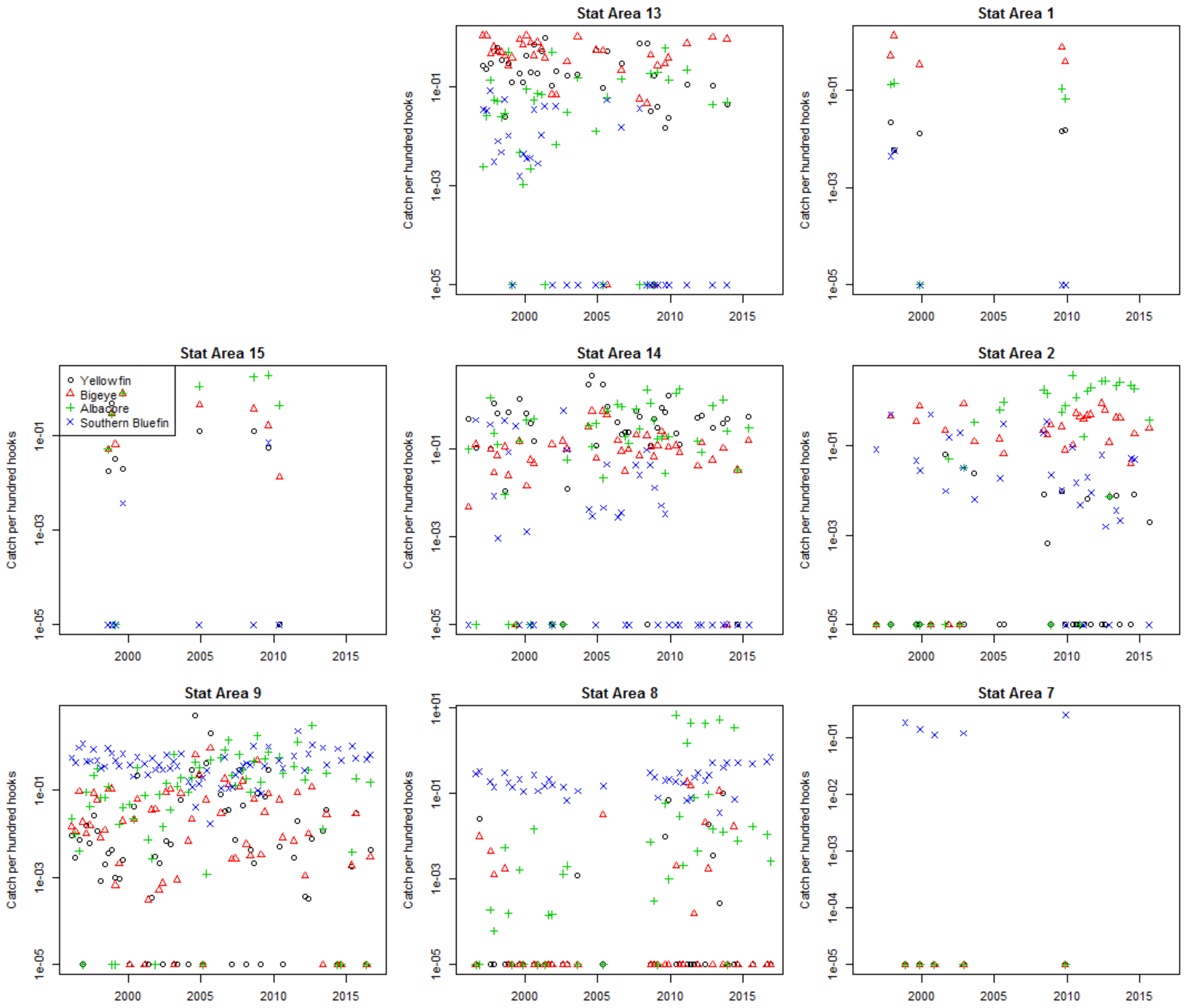


Figure 7: Mean catch per hundred hooks by year-qtr, species, and statistical area, plotted on a log scale, for yellowfin, bigeye, albacore, and southern bluefin tuna. Each CPUE has 1E-5 added so that zero catches appear on the log scale.

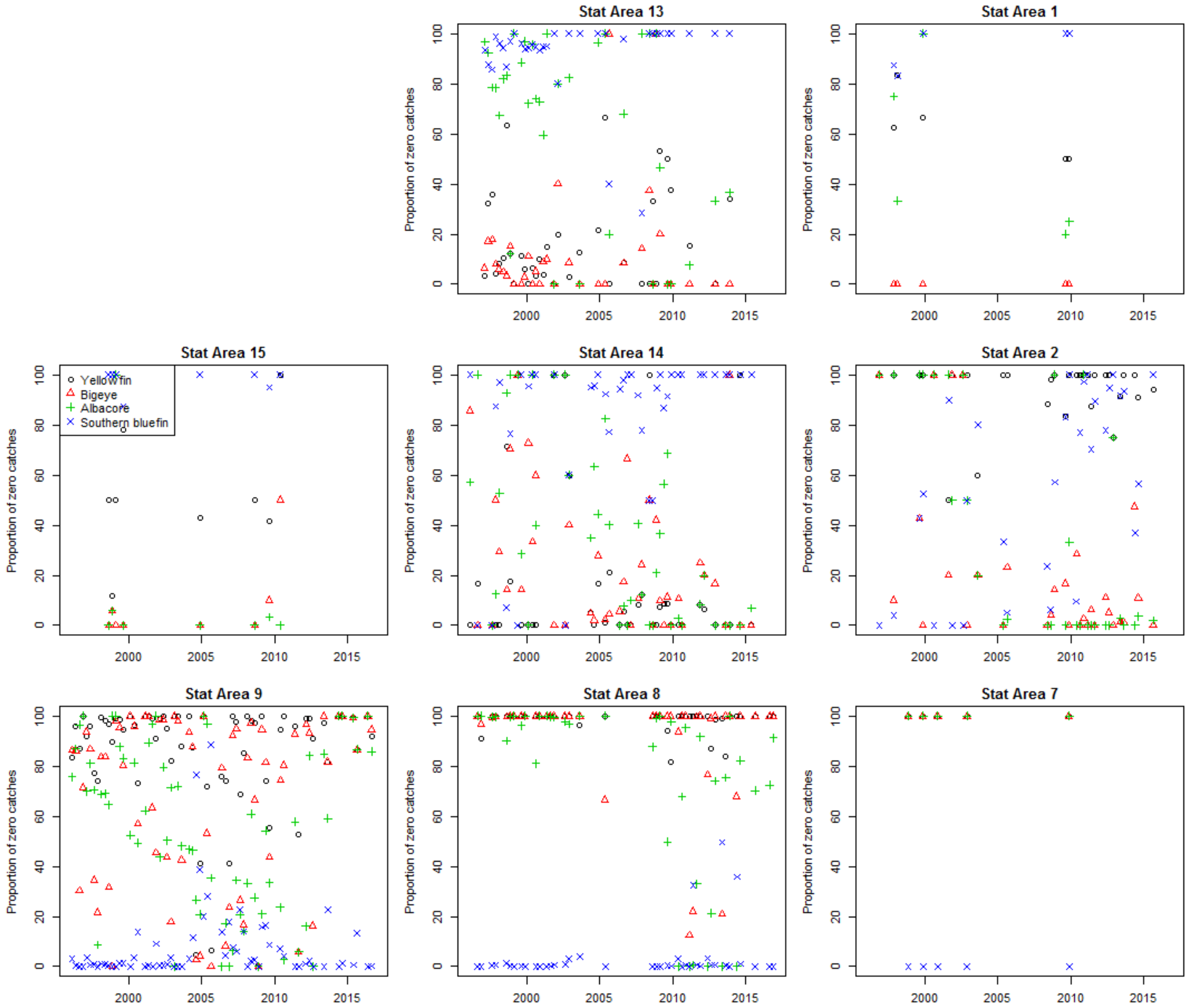


Figure 8: Proportion of zero catches per set by year-qtr, species, and statistical area, for yellowfin, bigeye, albacore, and southern bluefin tuna.

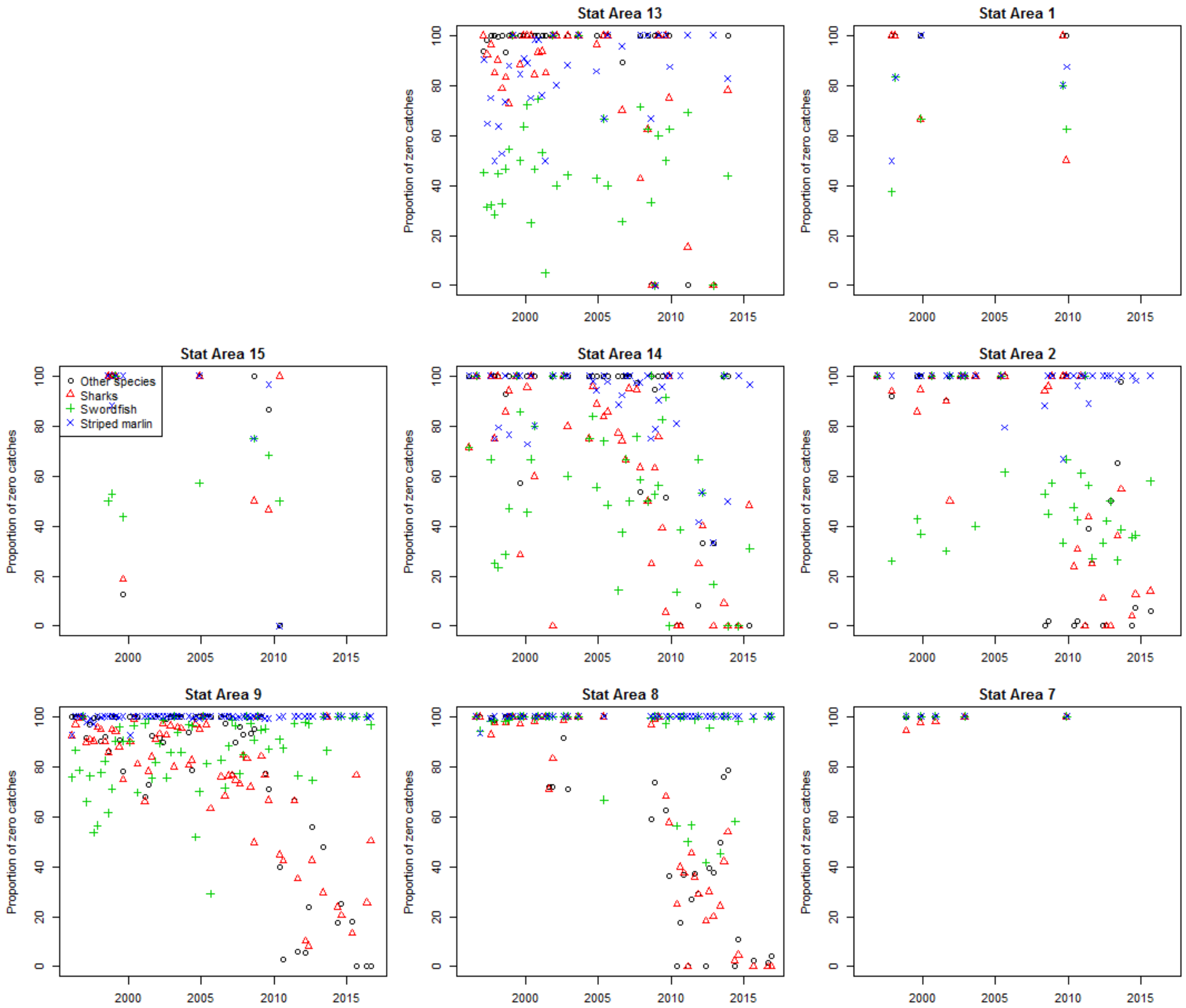
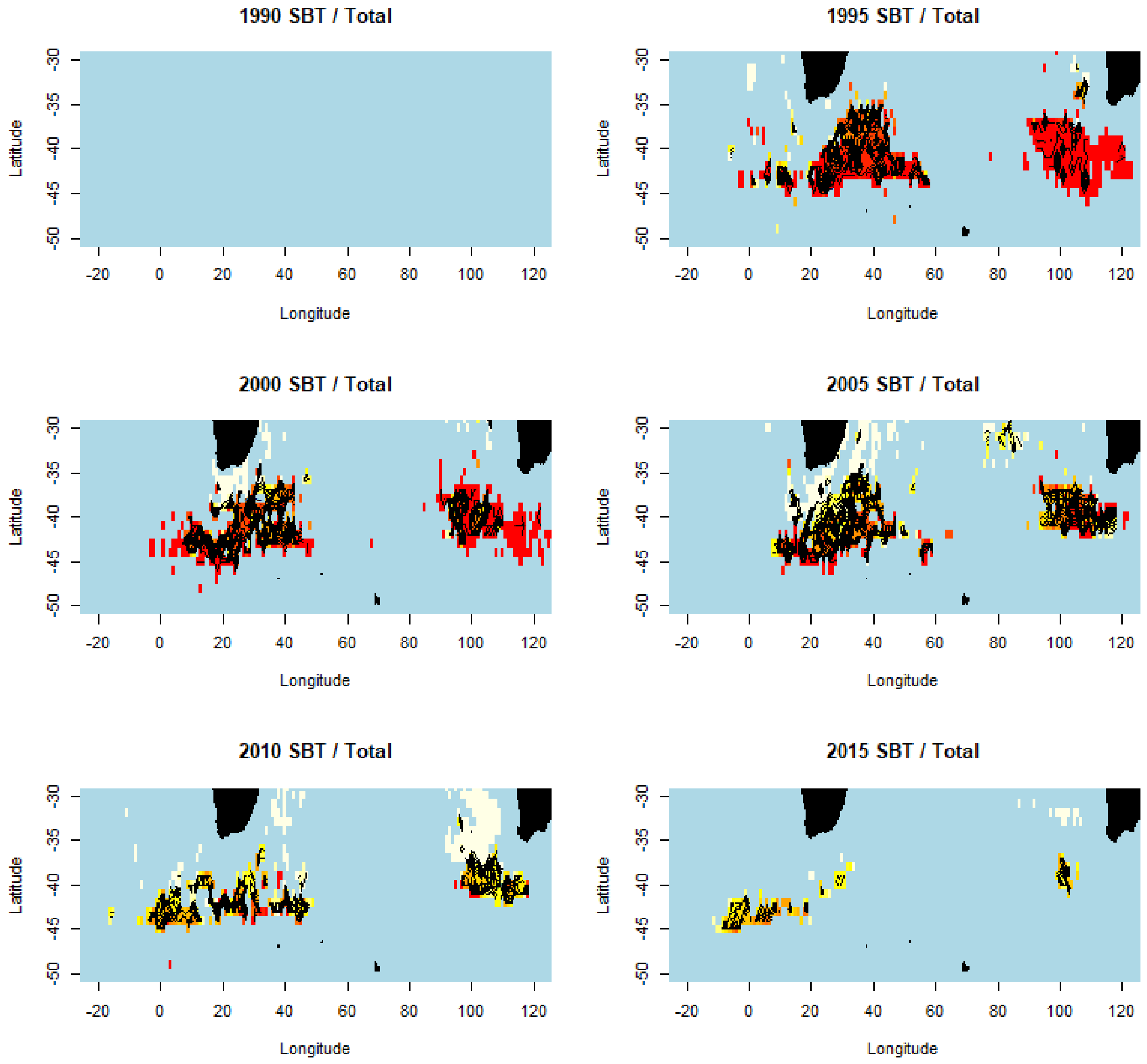


Figure 9: Proportion of zero catches per set by year-qtr, species, and statistical area for sharks, swordfish, striped marlin, and species not otherwise recorded (i.e. everything other than SBT, BET, YFT, ALB, SWO, BLM, BUM, MLS, SFA, SKJ, and SHA).


1990 SBT / Total

1995 SBT / Total

2000 SBT / Total

2005 SBT / Total

2010 SBT / Total

2015 SBT / Total

Figure 10: Proportion southern bluefin tuna (SBT) in the total reported catch in numbers by 1° cell, aggregated over 5 years within the period 1990-2014. Red colour indicates a higher proportion of SBT.

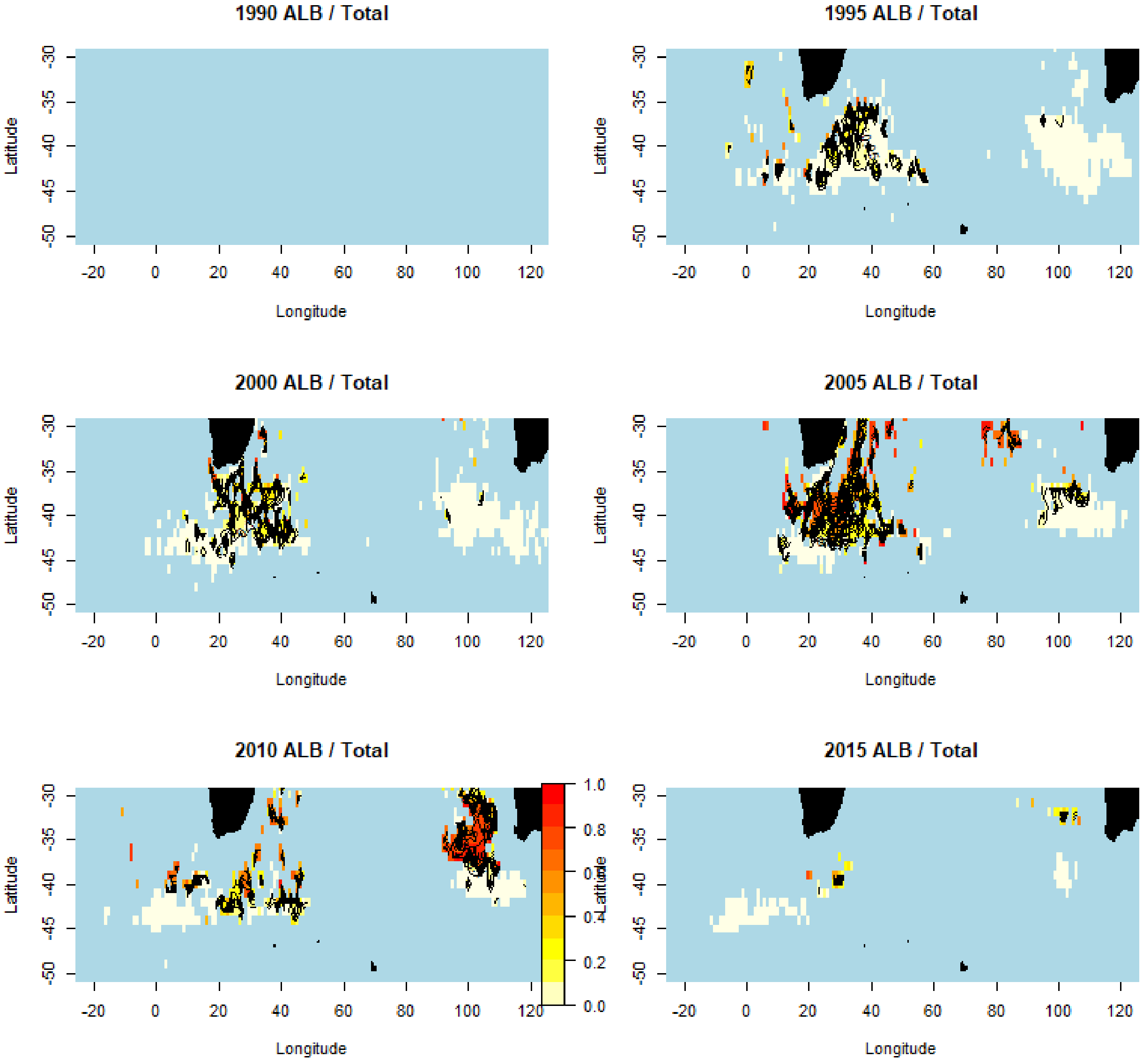


Figure 11: Proportion albacore (ALB) in the total reported catch in numbers by 1° cell, aggregated over 5 years within the period 1990-2014. Red colour indicates a higher proportion of ALB.

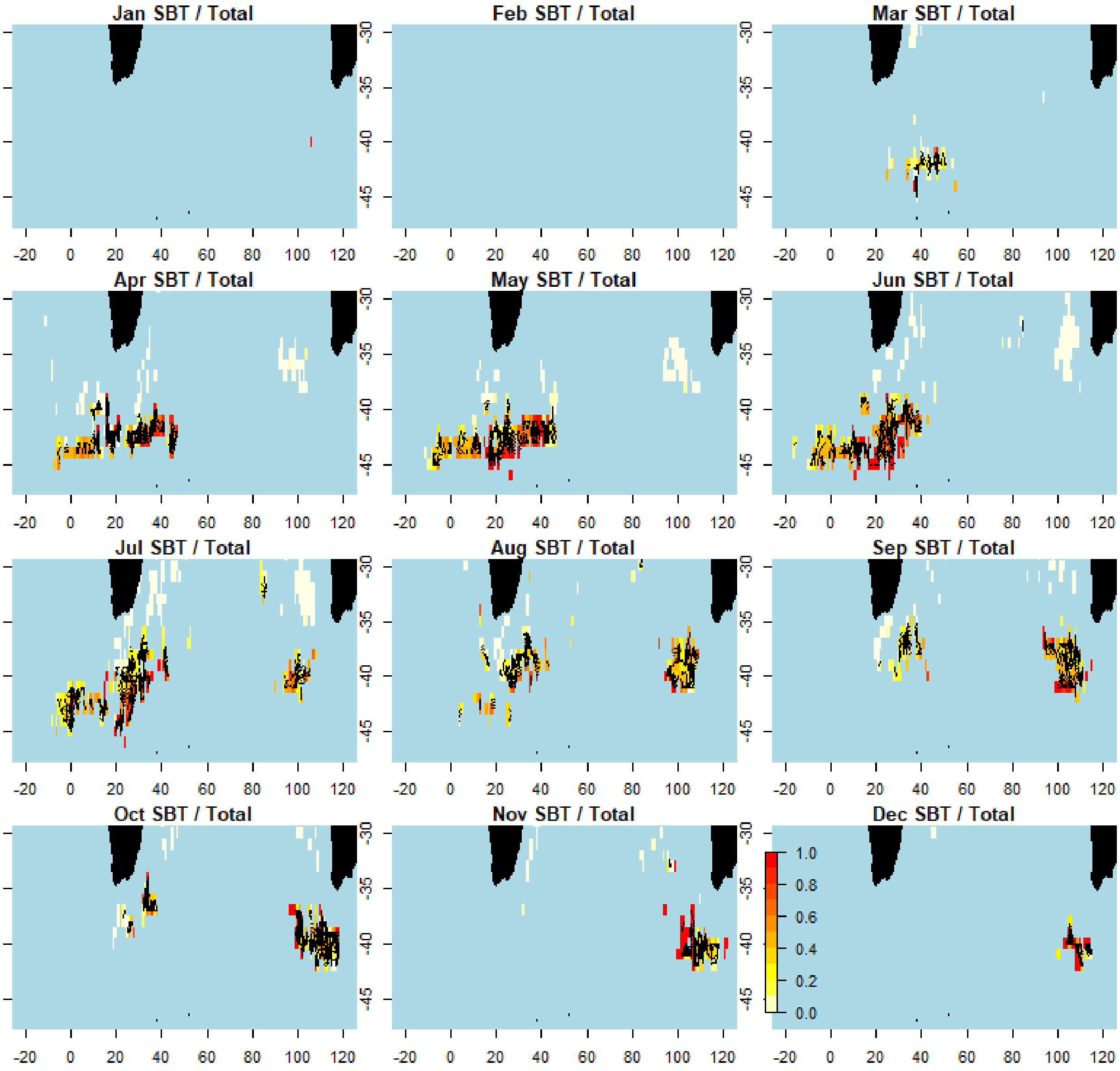


Figure 12: Proportion southern bluefin tuna (SBT) in the total reported catch in numbers by 1° cell, by month, aggregated over the period 2005-2014. Red colour indicates a higher proportion of SBT.

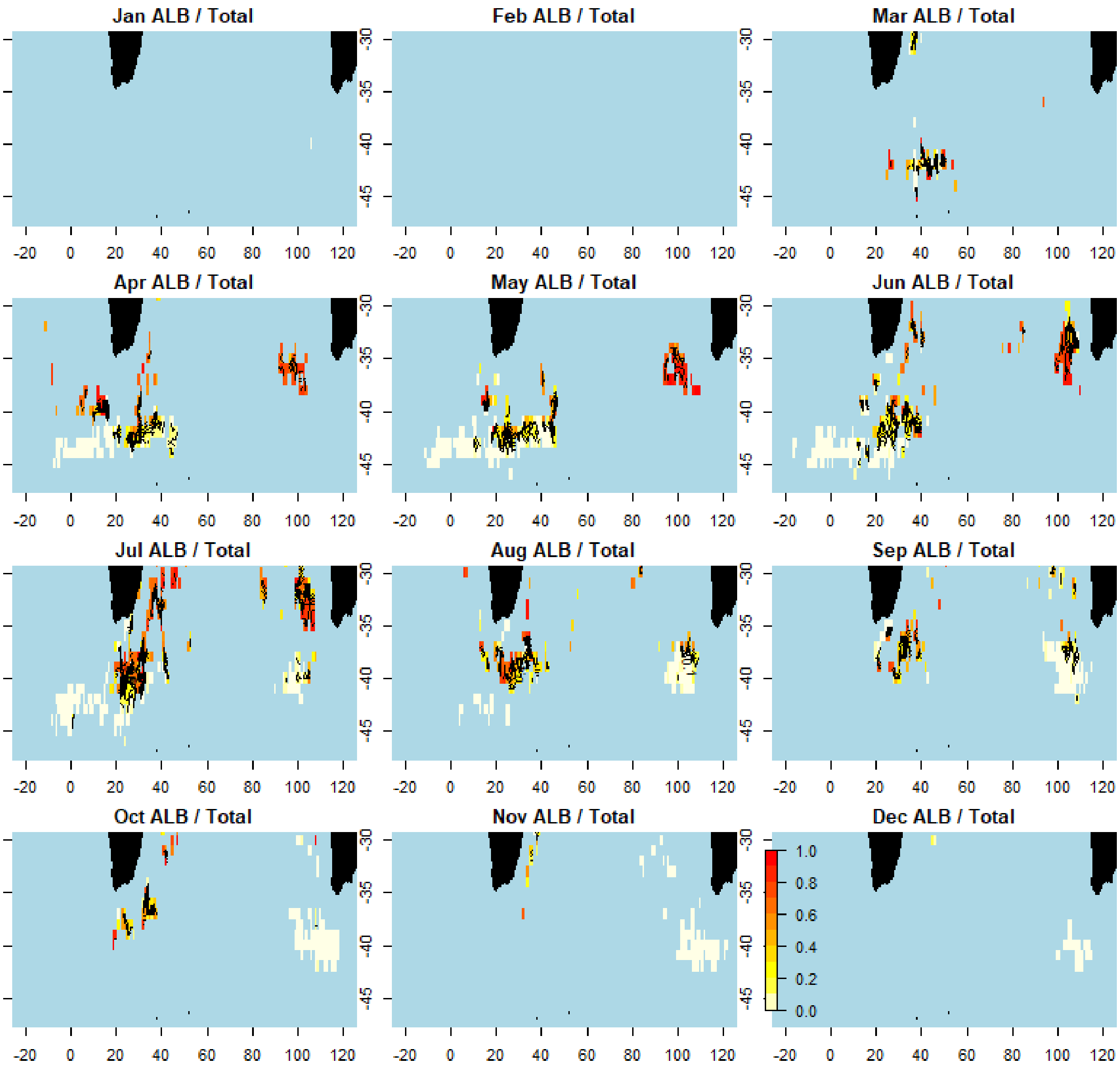


Figure 13: Proportion albacore (ALB) in the total reported catch in numbers by 1° cell, by month, aggregated over the period 2005-2014. Red colour indicates a higher proportion of ALB.

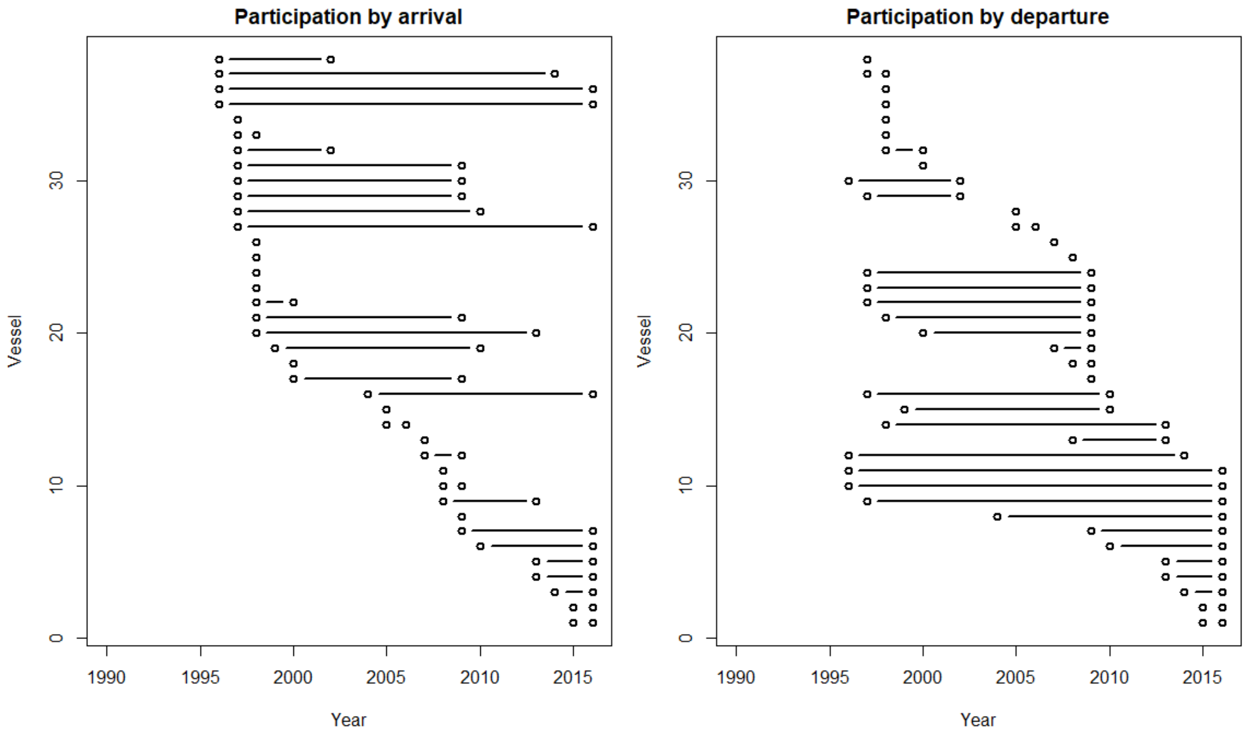


Figure 14: Plots of participation by vessel and year. Each row represents a vessel, sorted by the first year of participation, except for the top right plot which is sorted by the final year.

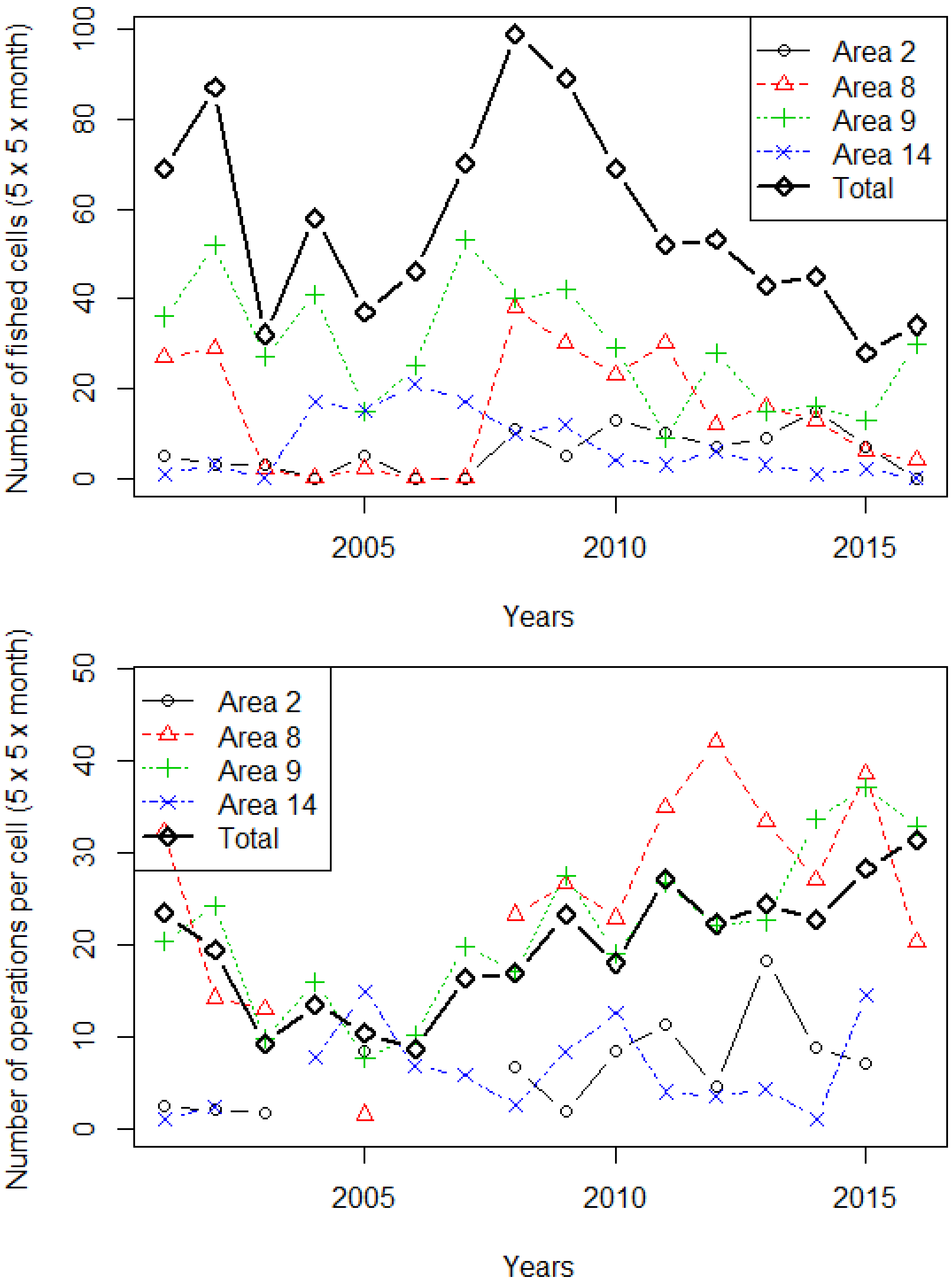


Figure 15: For fishing since 2000 in areas 2, 8, 9, and 14, the number of cells (5° latitude by 5° longitude by month) fished (above) and the number of longline operations per cell (below).

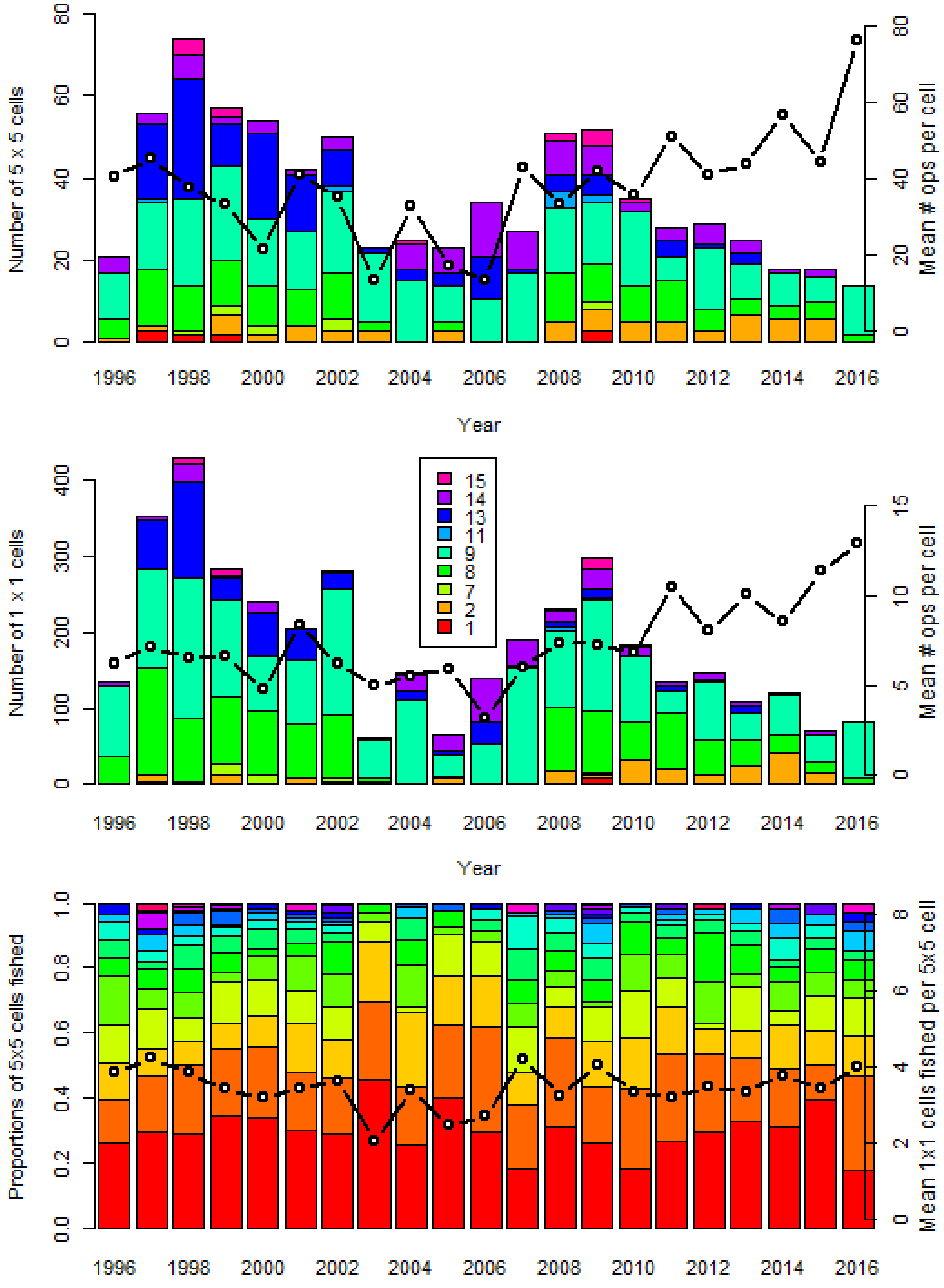


Figure 16: (Upper) Bars represent the number of major cells (5x5° by month) fished by CCSBT statistical area and year, see left y-axis. The line represents the mean annual operations per cell, see right y-axis. (Middle) As for upper plot, but with minor cells (1x1° by month) instead of major cells. (Lower) 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.

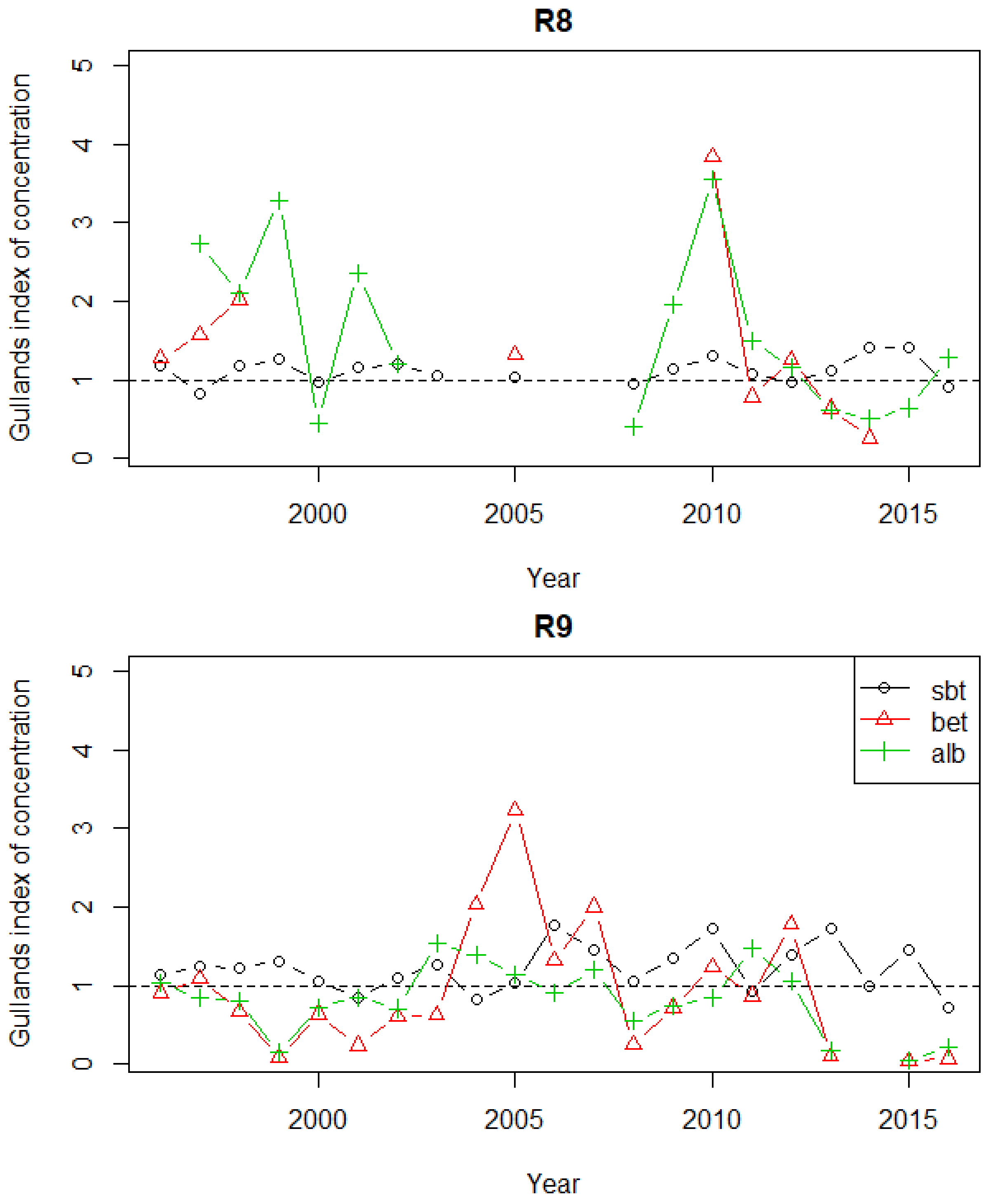


Figure 17: Gulland's indices of concentration estimated annually for southern bluefin tuna, bigeye tuna, and albacore tuna, in statistical areas 8 and 9.

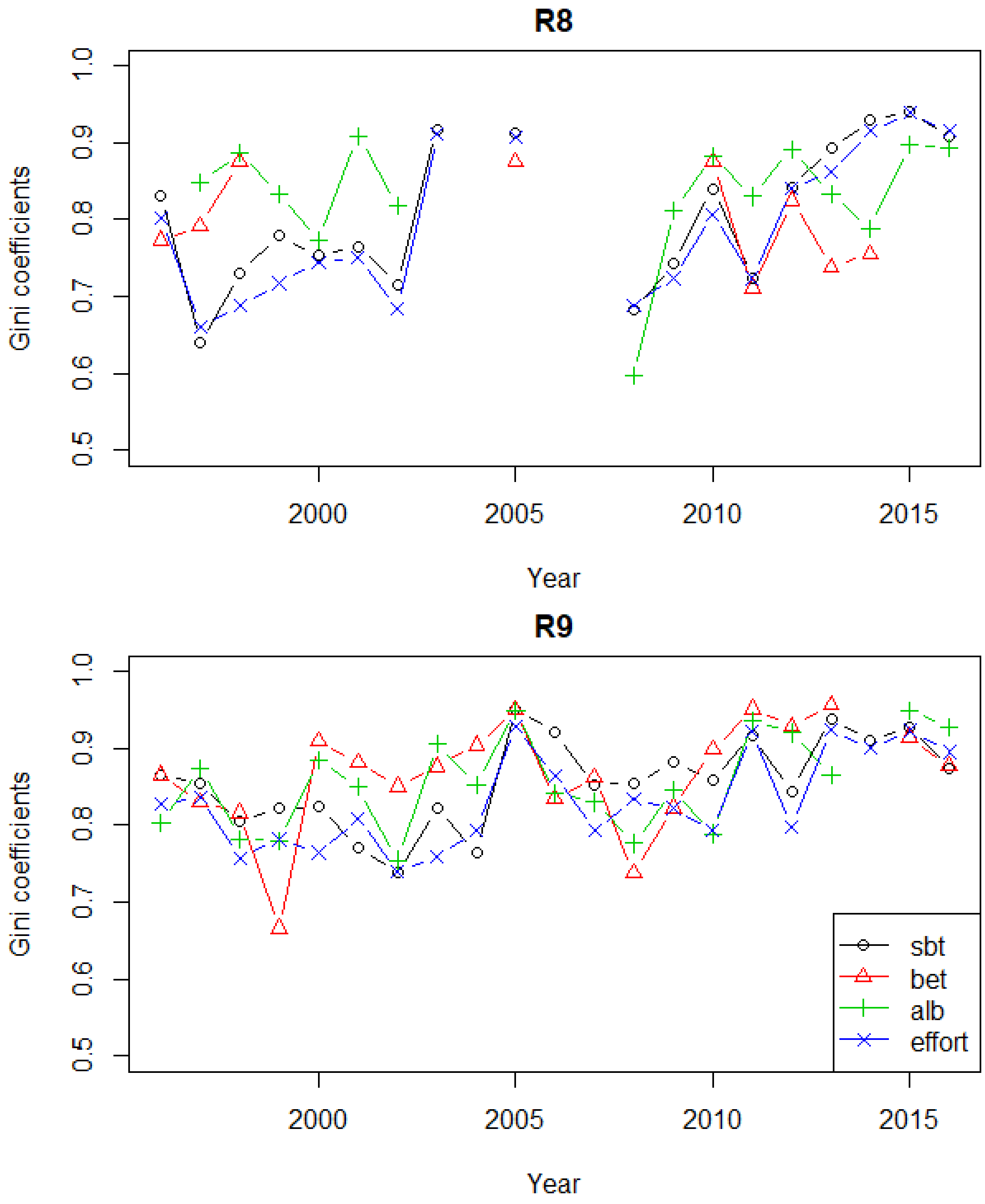


Figure 18: Gini coefficients estimated annually for southern bluefin tuna, bigeye tuna, albacore tuna, and fishing effort in statistical areas 8 and 9

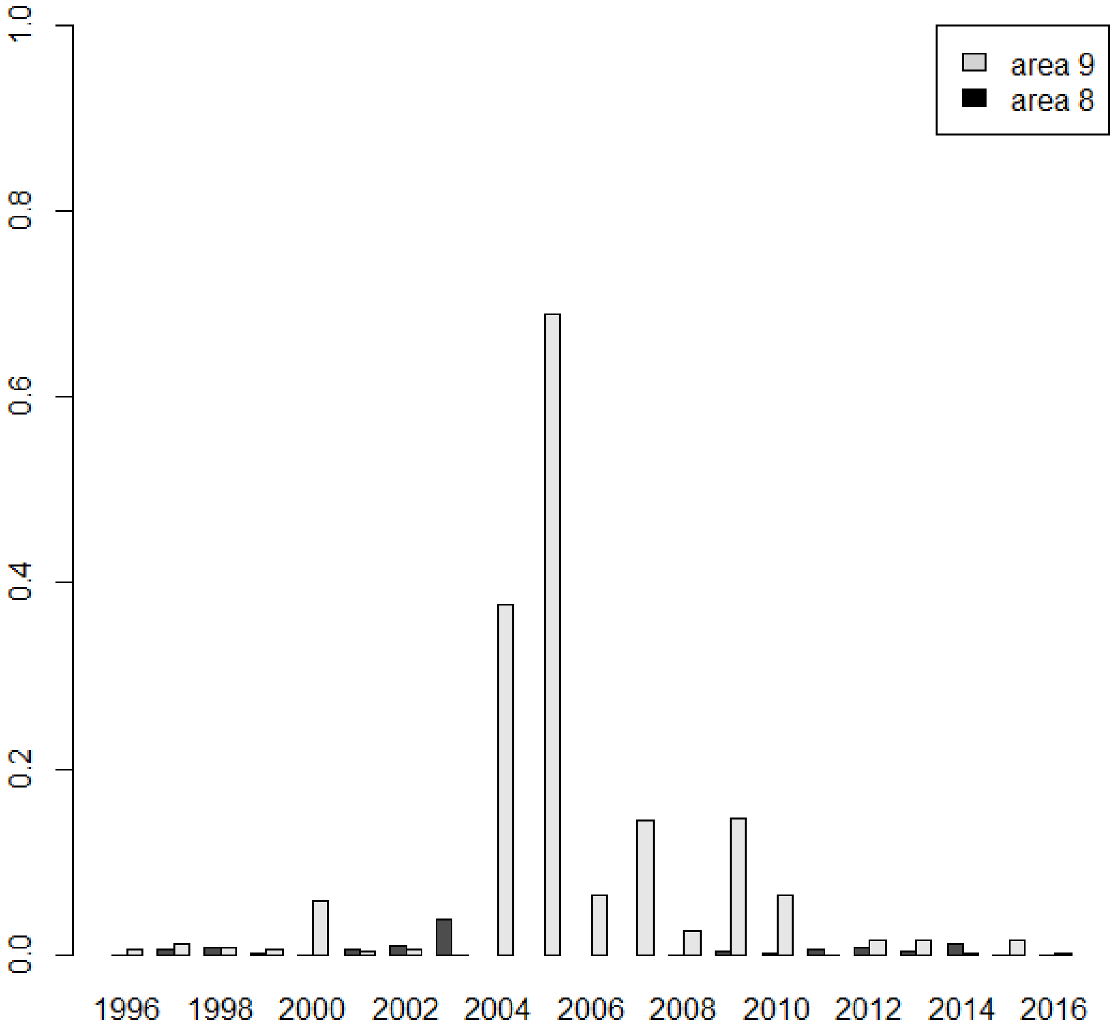


Figure 19: Proportions of sets with zero catches of SBT by year and statistical area, in the data used in the standardization models.

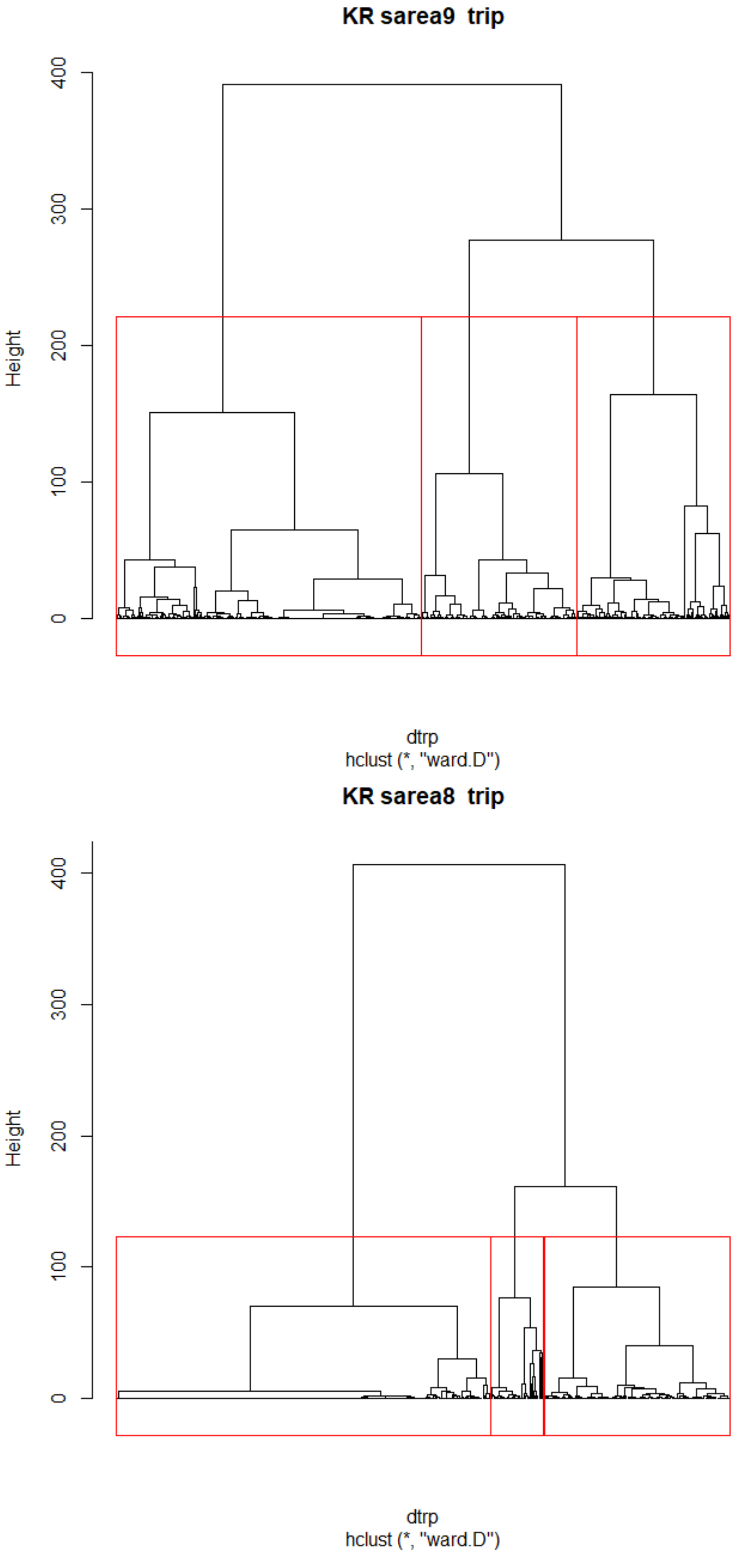


Figure 20: Dendrograms for Ward hierarchical cluster analyses of statistical areas 9 (above) and 8 (below), with the red lines indicating the separation into 3 clusters for each.

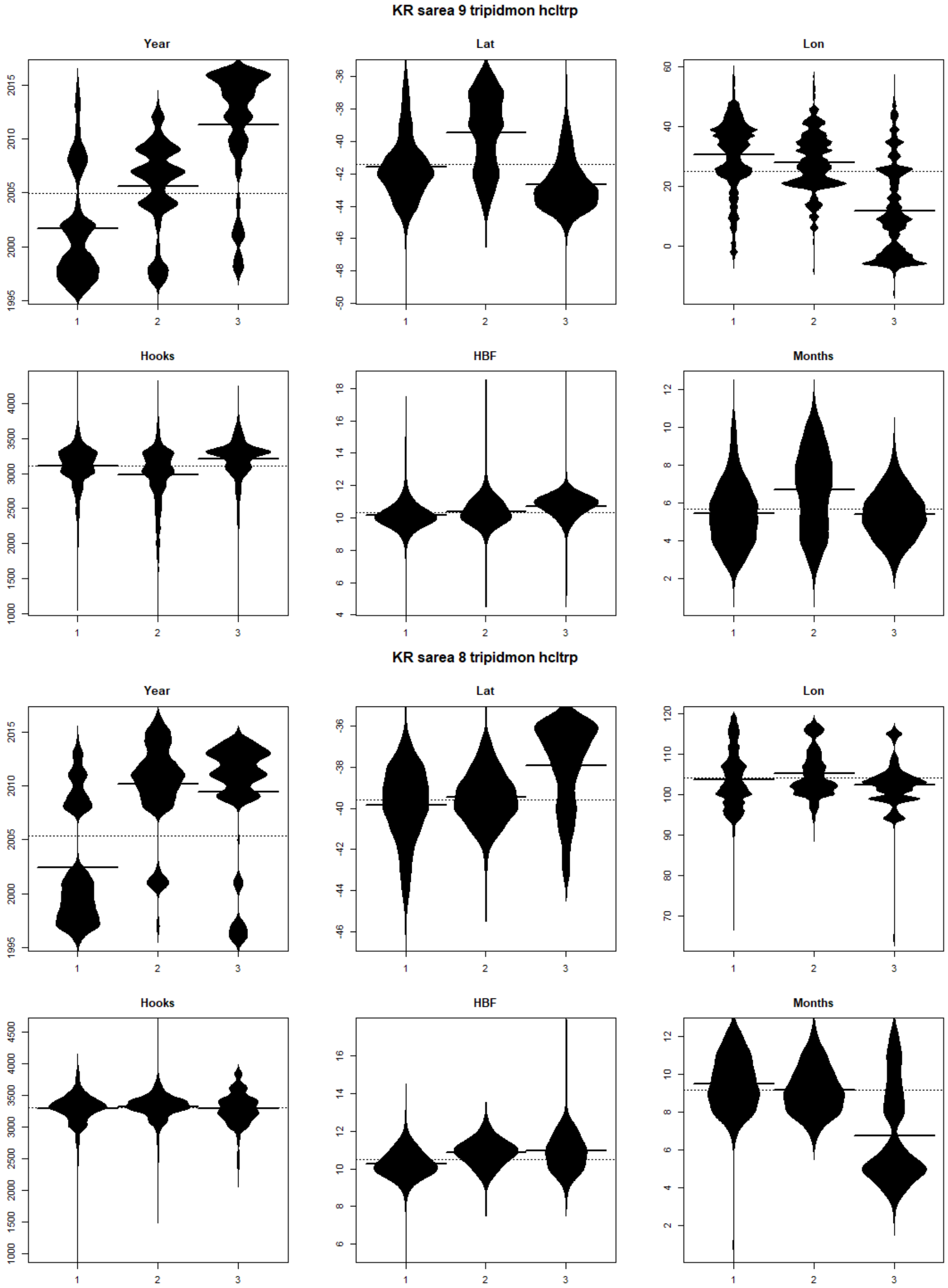


Figure 21: Beanplots for statistical areas 9 (above) and 8 (below), showing the number of sets versus covariate by cluster. The horizontal bars indicate the medians.

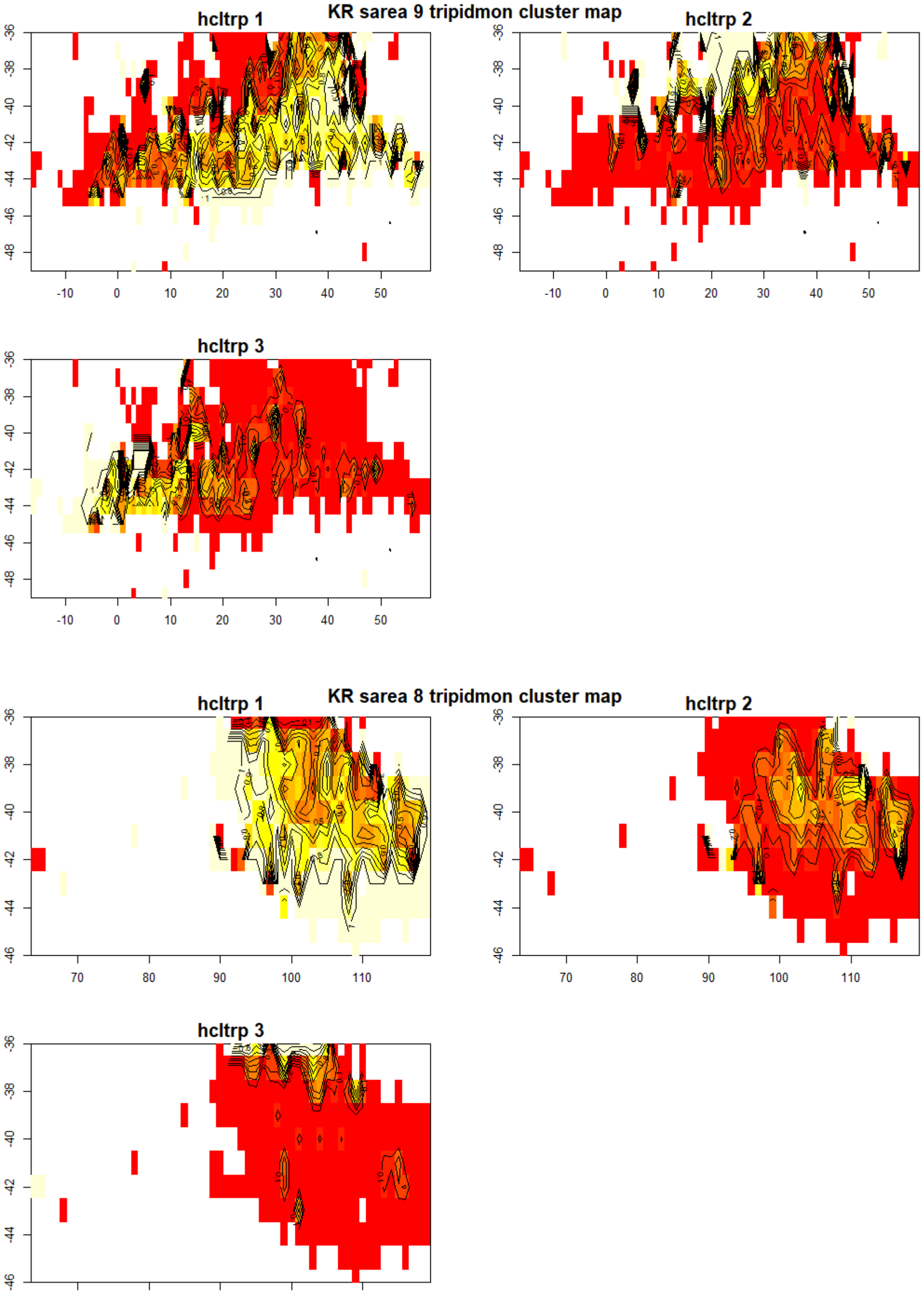


Figure 22: Maps of the proportion of each cluster per 1 degree square in total effort for statistical areas 9 (above) and 8 (below). Higher proportions are shown in yellow. White space indicates no reported effort.

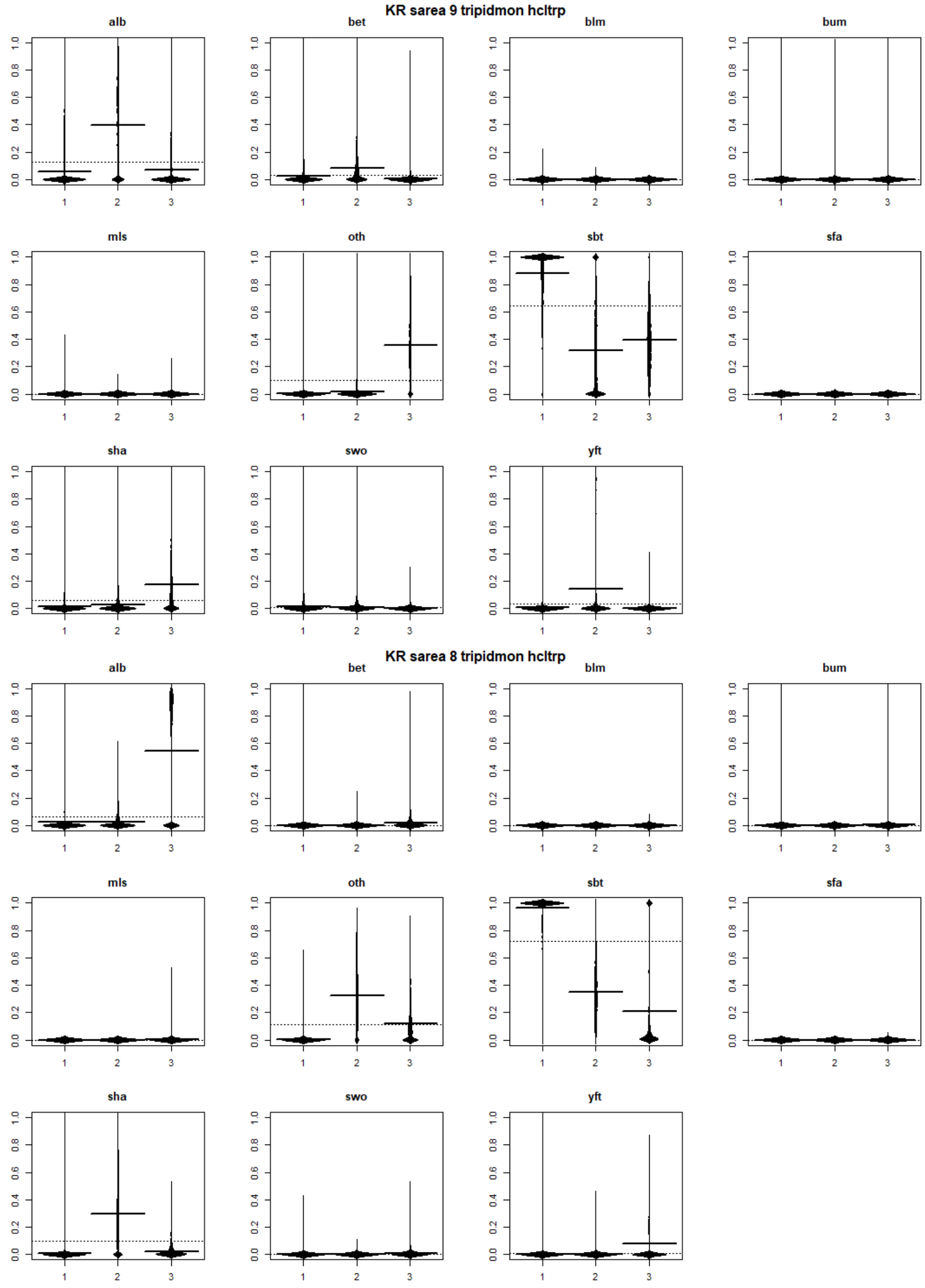


Figure 23: Beanplots for statistical areas 9 (above) and 8 (below), showing species composition by cluster. The horizontal bars indicate the medians.

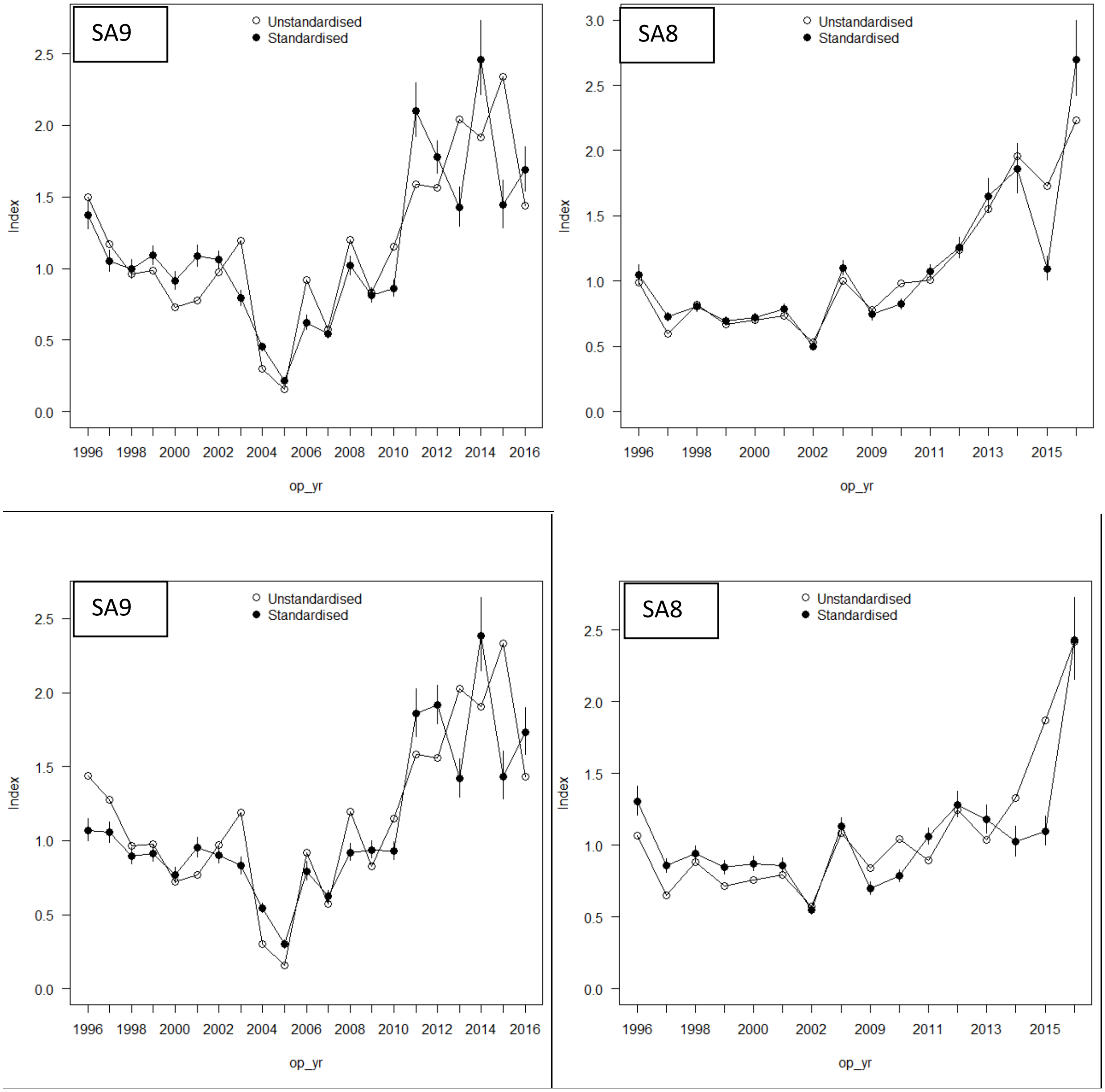


Figure 24: Unstandardized and standardized CPUE indices based on lognormal GLMs with an added constant, for statistical areas 9 (left) and 8 (right), and addressing target change using selected data (above) and cluster analysis (below).

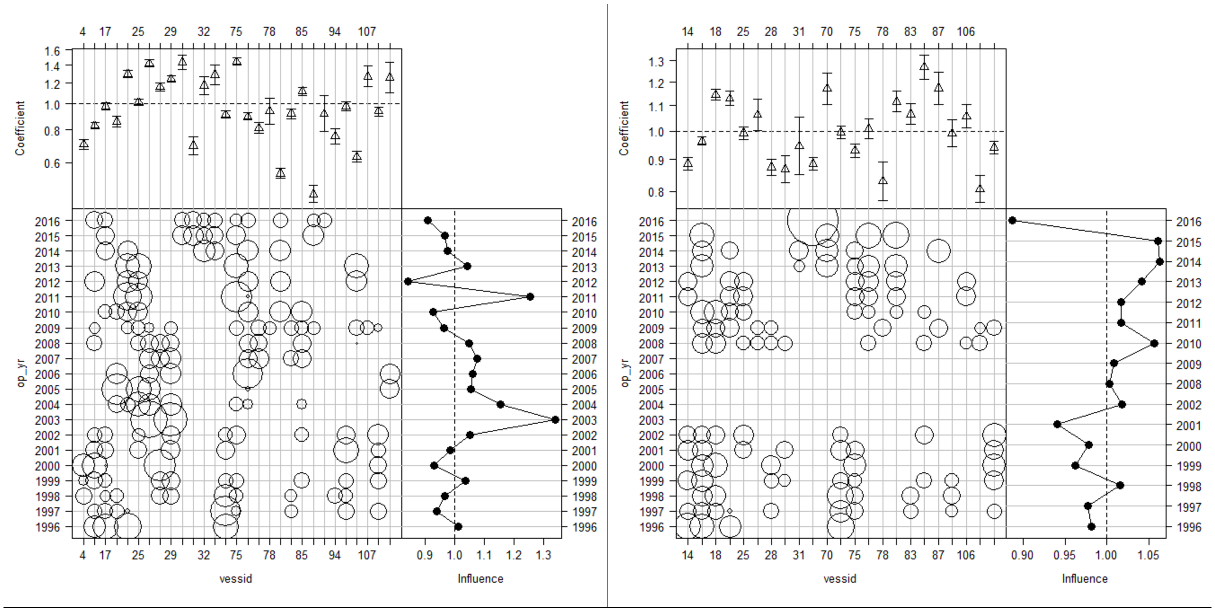


Figure 25: Influence plots for vessel effects for areas 9 (left) and 8 (right), addressing target change using selected data.

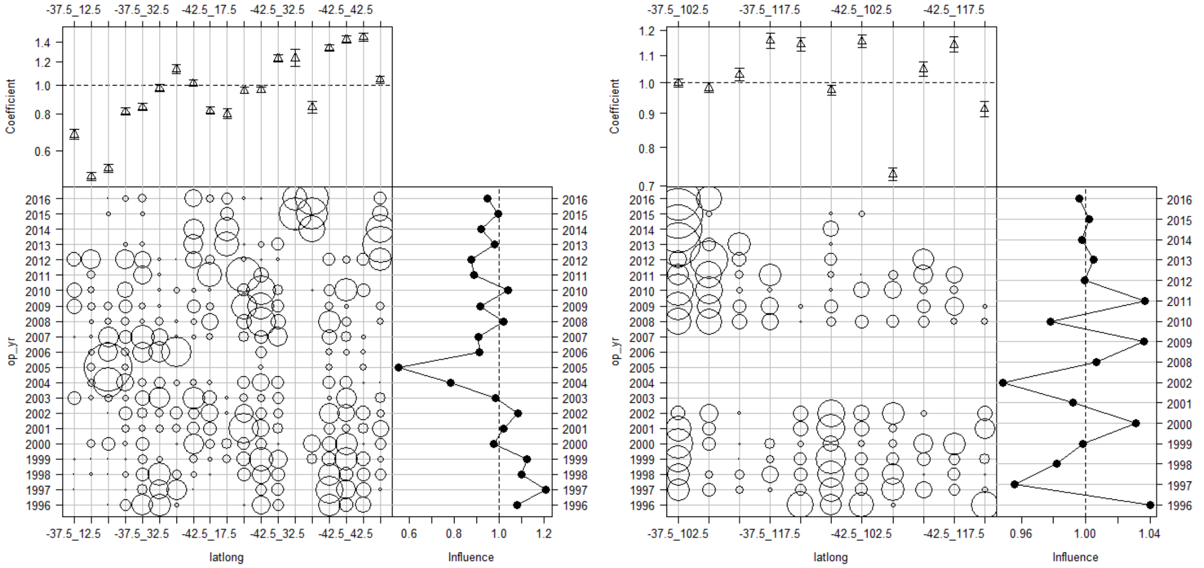


Figure 26: Influence plots for spatial latlong effects for statistical areas 9 (left) and 8 (right) , addressing target change using selected data.

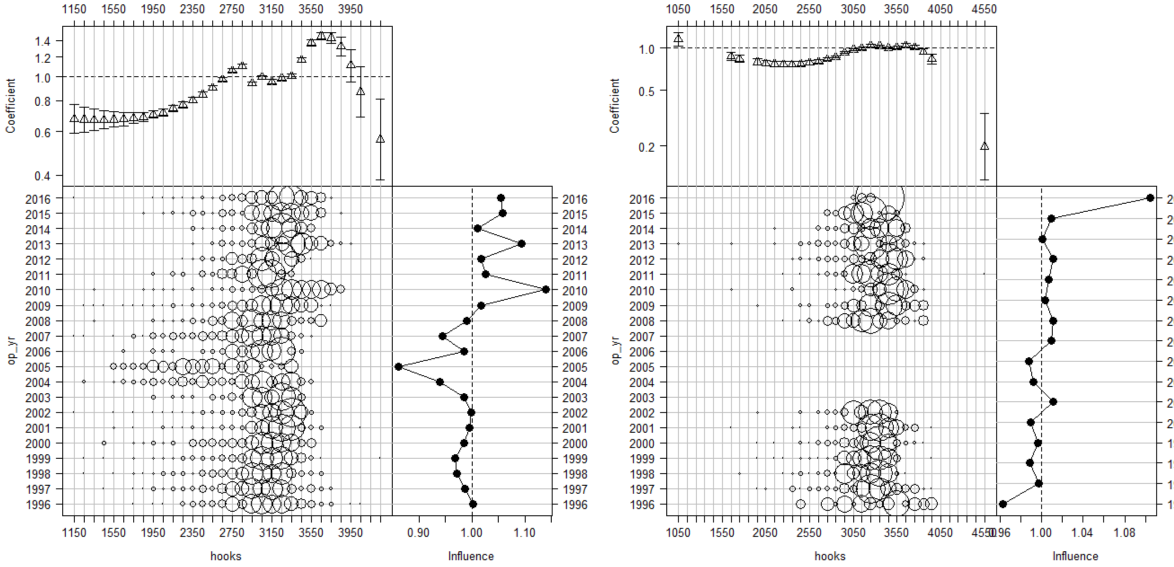


Figure 27: Influence plots for the effects of numbers of hooks for statistical areas 9 (left) and 8 (right), addressing target change using selected data.

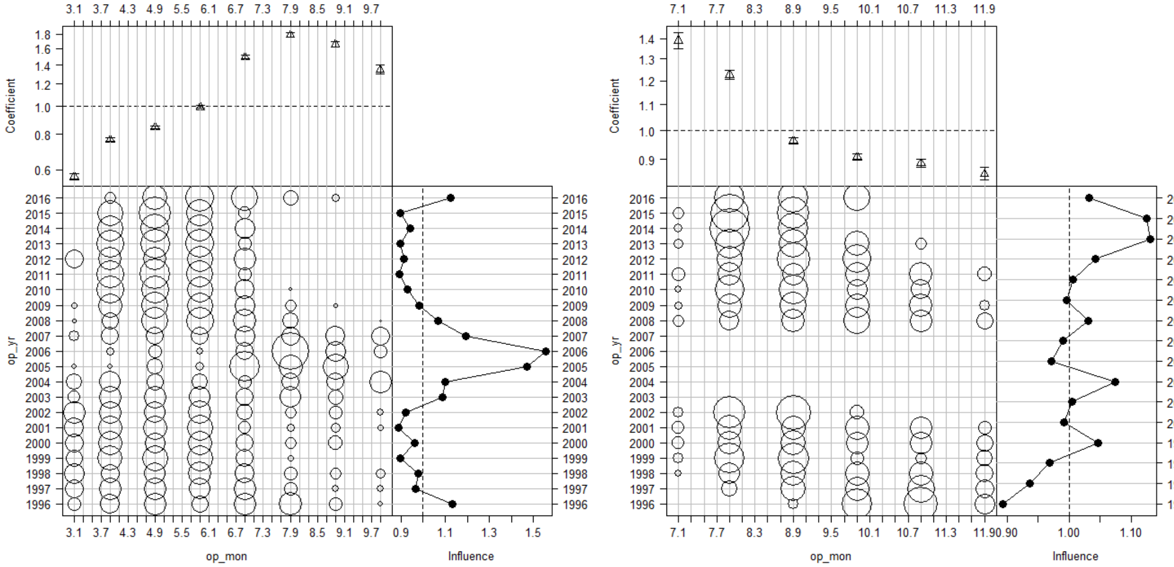


Figure 28: Influence plots for month effects for statistical areas 9 (left) and 8 (right), addressing target change using selected data.

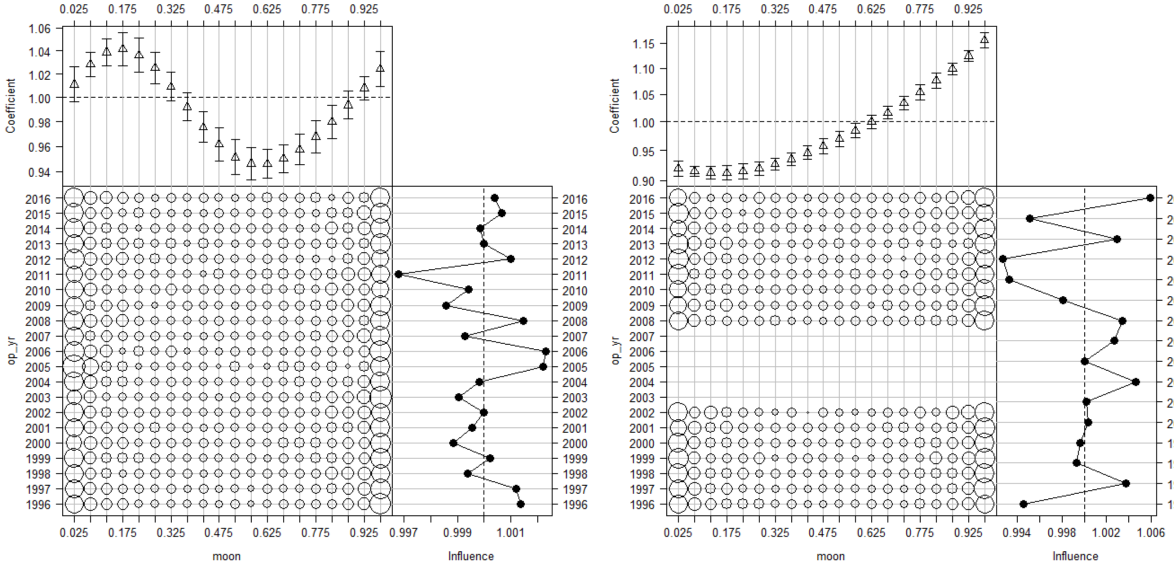


Figure 29: Influence plots for lunar illumination effects for statistical areas 9 (left) and 8 (right), addressing target change using selected data.

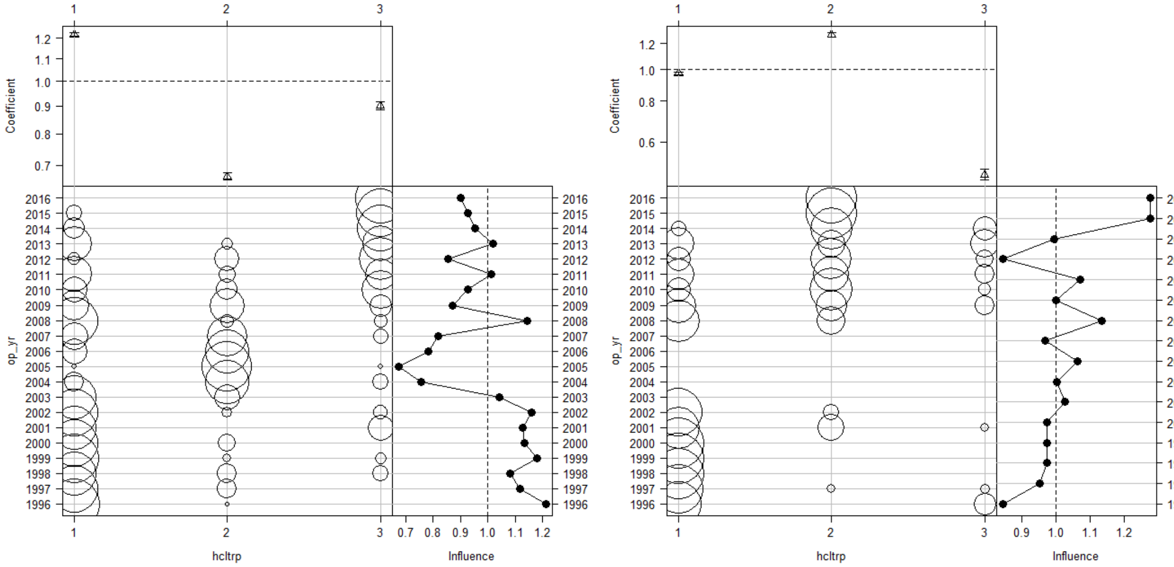


Figure 30: Influence plots for cluster effects for statistical areas 9 (left) and 8 (right), addressing target change using clustering.

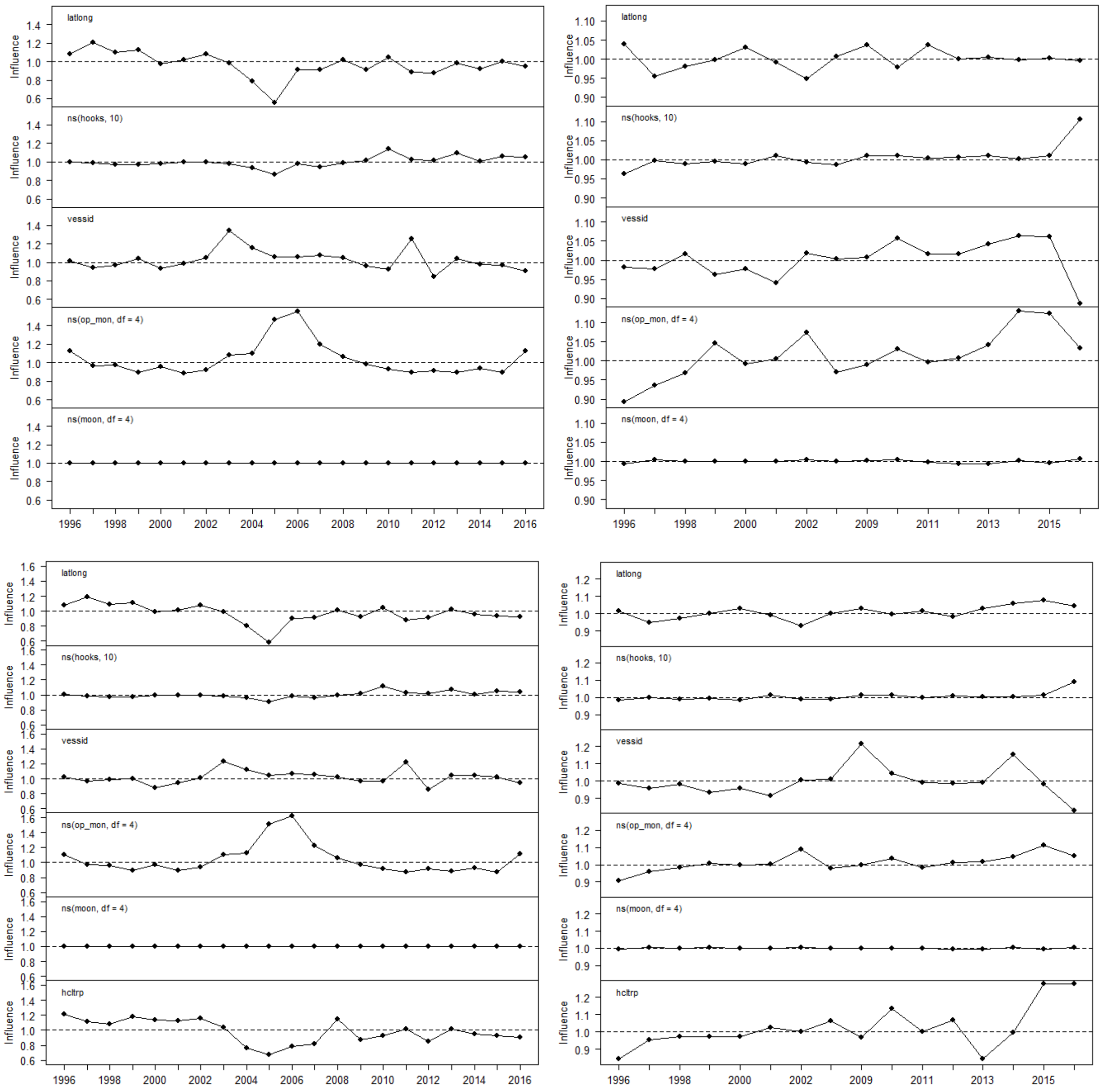


Figure 31: Compilation of influence plots for statistical areas 9 (left) and 8 (right), addressing target change using selected data (above) and clustering (below).

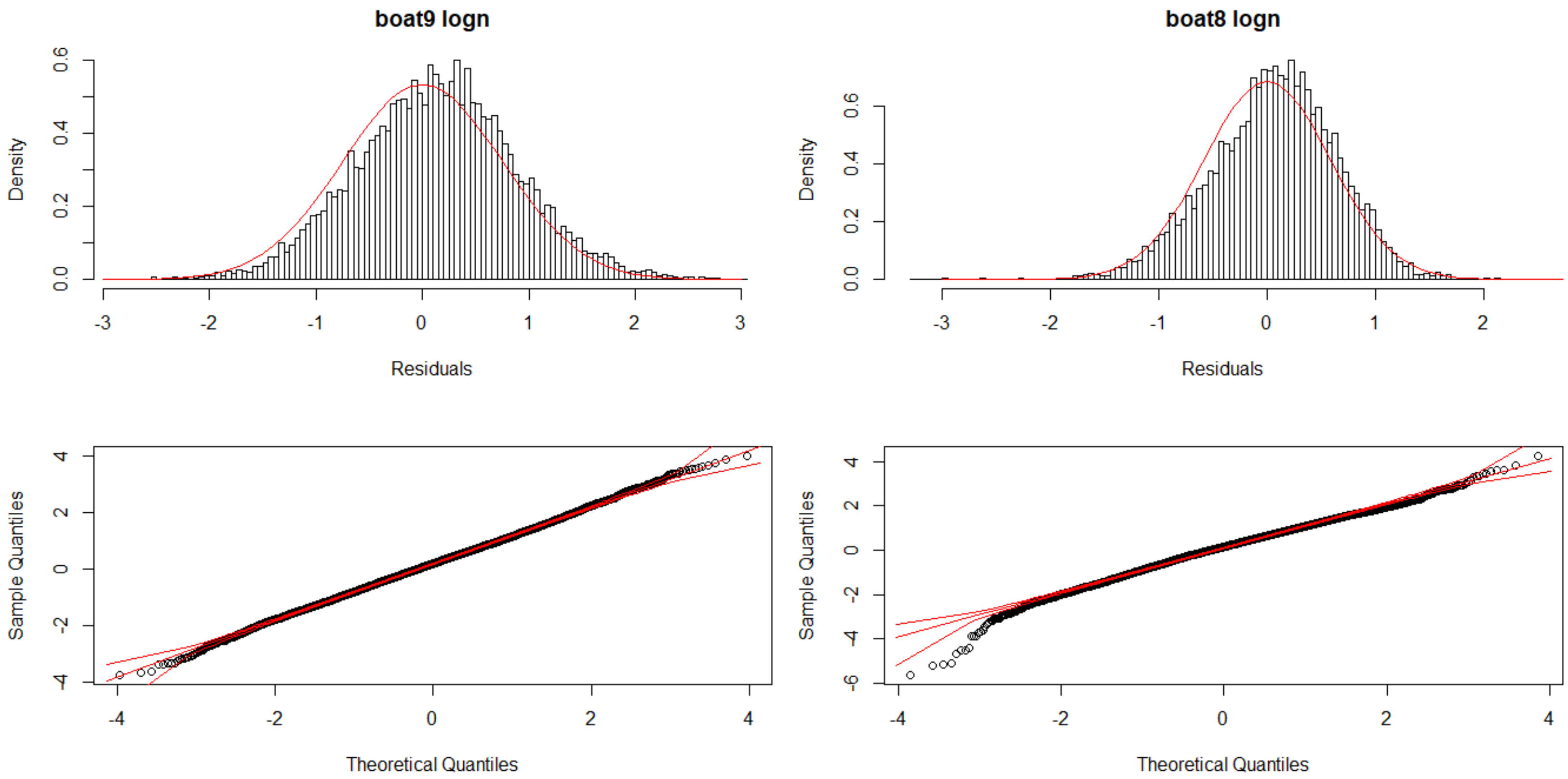


Figure 32: Frequency distributions of the standardized residuals (above) and Q-Q plots of standardized residuals for lognormal constant GLM analyses of statistical areas 9 (left) and 8 (right).

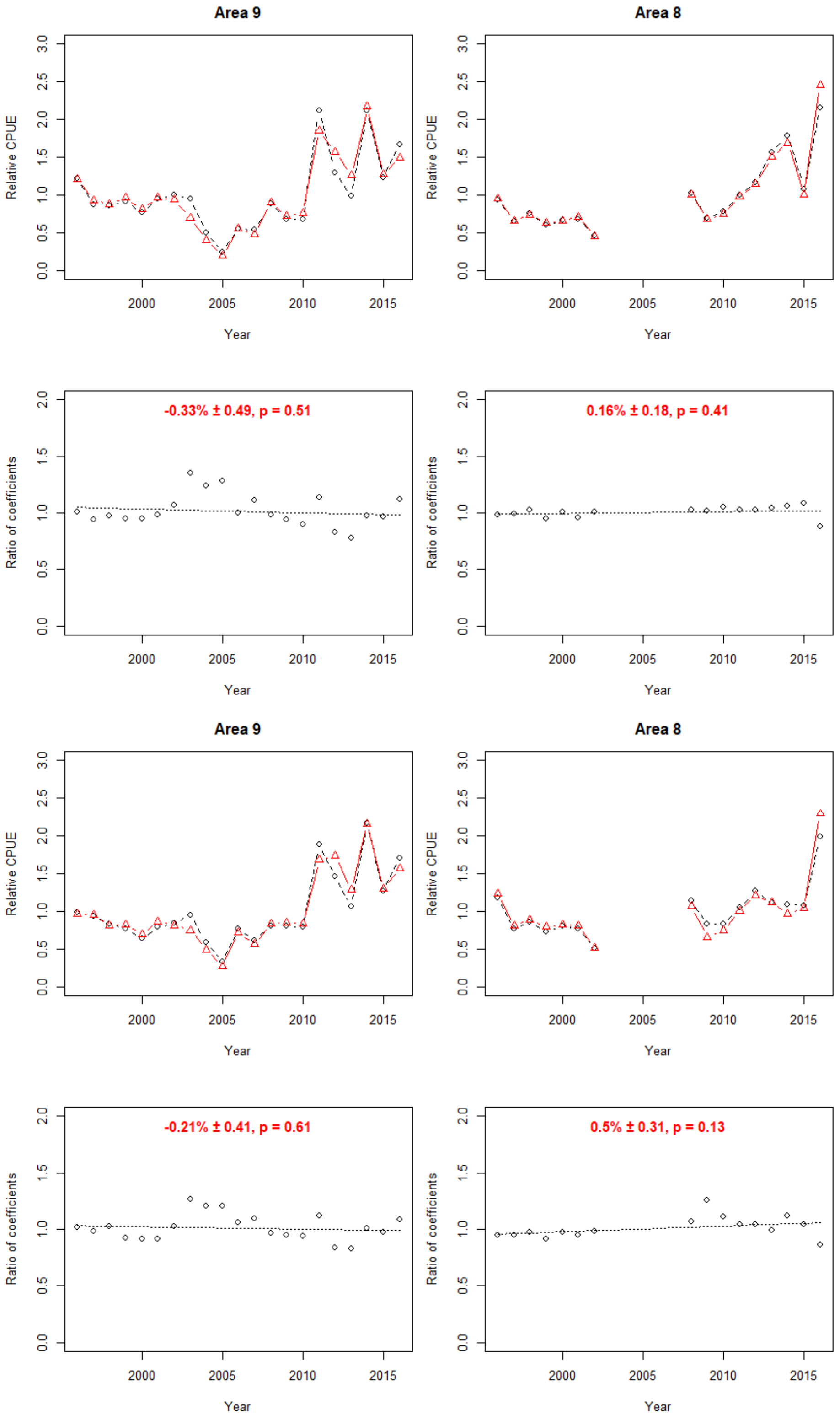


Figure 33: Annual abundance indices from standardizing SBT CPUE for statistical areas 9 (left) and 8 (right) using lognormal constant models, fitted either with (red triangles) or without (black circles) vessel effects. The second and fourth rows show the ratios of each pair of indices, with log-linear trends fitted. The numbers indicate the annual rate of change in the ratio. The top four plots use selected data, while the lower four plots use clustered data.

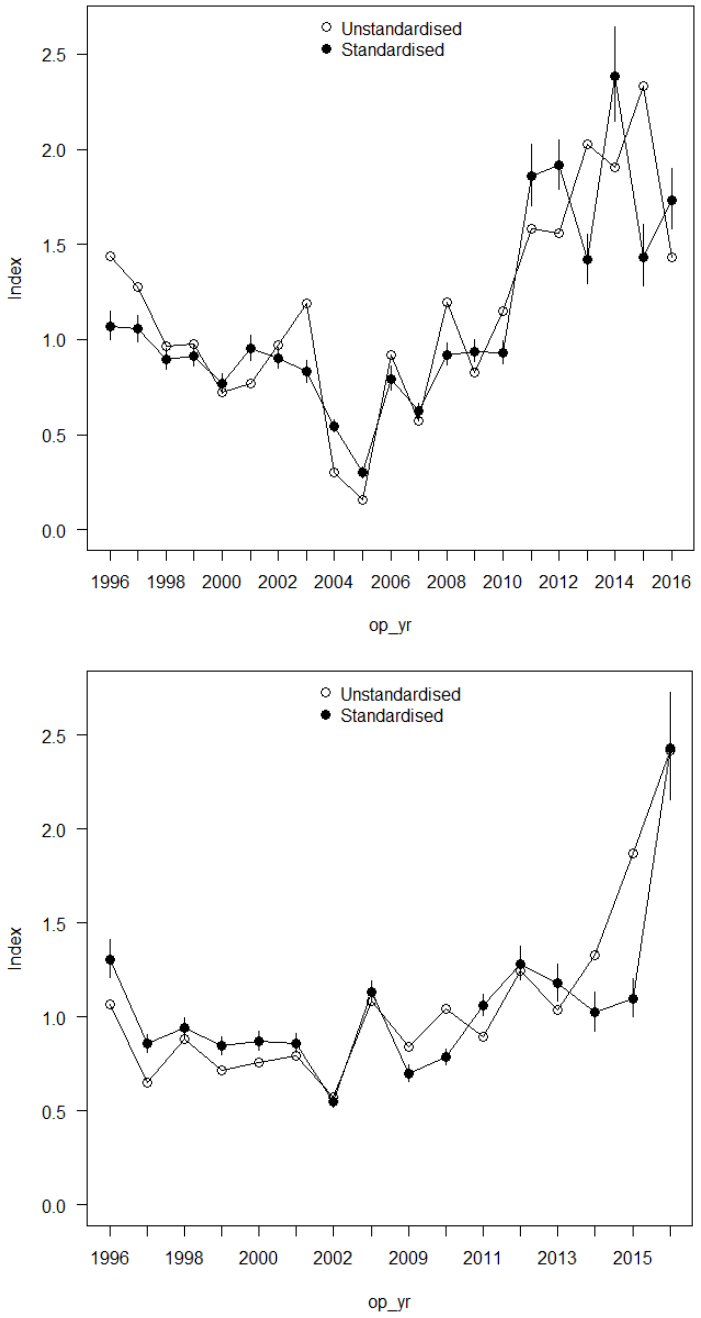


Figure 34: Unstandardized and standardized CPUE indices using clustered data for statistical areas 9 (above) and 8 (below), based on lognormal GLMs with an added constant.