DATA EXPLORATION AND CPUE STANDARDIZATION FOR THE KOREAN SOUTHERN BLUEFIN TUNA LONGLINE FISHERY (1996-2014)

Simon Hoyle[1], Sung Il Lee[2] and Zang Geun Kim[2]

ABSTRACT

In this study we standardized southern bluefin tuna, *Thunnus maccoyii* (SBT) CPUE from Korean tuna longline fisheries (1996-2014) 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° square), 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. Explanatory variables for the GLM analyses were year, month, vessel identifier, 5° square, and number of hooks. GLM results for the whole area suggested that location, year, 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 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 705-1,117 mt due to the national catch limit. The catch in 2014 was 1044 mt (Fig. 1).

Korean tuna longline vessels fishing for SBT have mainly operated in two locations to the south of 35oS 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.

[1] Hoyle Consulting, Nelson, New Zealand, simon.hoyle@gmail.com
[2] National Fisheries Research and Development Institute, Busan, Korea, k.sungillee@gmail.com

In this study, we first explored the data in order to better understand the fisheries, and then standardized the CPUE data of Korean tuna longline fisheries (1996-2014) using Generalized Linear Models (GLM) to obtain a proxy for the abundance index.

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 degree square 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° squares, and CCSBT statistical areas.

For CPUE standardization, data were cleaned by removing sets in which HBF was less than 9 or greater than 12. Sets with fewer than 1000 hooks were removed from the dataset.

Data were plotted to explore CPUE by species by year-qtr, for each statistical area.

We examined patterns through time and among species in the nominal catch rates by year-qtr and statistical area, and compared them with patterns in the proportions of sets with no catch of each species.

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

Data from the period 1996-2014 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 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 *et al*. (2014).

GLM analyses

The operational data were standardized using generalized linear models in R 3.2.1 (R Core Team, 2014). Analyses were conducted separately for each of the two core areas. 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° squares in which there had been at least 200 sets.

Initial exploratory 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:

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

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 variable *moon* was the lunar illumination on the date of the set. The variables *year*, *vessid*, and *latlong* (5° latitude-longitude square) were fitted as categorical variables.

Data in the lognormal constant GLM, and the lognormal positive GLM below, 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-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 $w_{ijt} = \frac{h_{ijt}}{\sum_{j=1}^{n} h_{ijt}}$.

In models used to generate indices the delta lognormal approach to standardization (Lo *et al.*, 1992, Maunder and Punt, 2004) was used. This approach uses a binomial distribution for the probability *w* of catch being zero and a probability distribution *f(y)* , where y was log(catch/hooks set), for non-zero catches. An index was estimated for each year-quarter, which was the product of the year effects for the two model components, $(1 - w).E(y|y \neq 0)$.

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

$$w = p(CPUE = 0) \sim year + vessid + latlong + \lambda(hooks) + g(month)$$

$$f(y) = CPUE \sim year + vessid + latlong + \lambda(hooks) + g(month)$$

The categorical variables *year*, *month* and *latlong* (5° latitude-longitude square) were fitted in all analyses. The number of hooks in the set was included as a covariate using a cubic spline with 10 degrees of freedom. Models were run both with and without the vessel identifier (vessel id) as a categorical variable. Lunar illumination (*moon*) was not included in this group of analyses. 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 *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. For example, for the lognormal positive approach the following GLM was used, where α_t are the abundance indices, β_i are the coefficients for the 5° lat-long squares, and γ_{vessel} are the vessel effects.

$$log\left(\frac{sbt}{hooks}\right) = c + \alpha_t + \beta_i + g(hooks) + \gamma_{vessel} + \epsilon_{set}$$

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.

For all model comparisons, 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 delta lognormal GLM model with the standard settings, including vessel effects. Binomial time effects were obtained by taking the time effects from the glm and setting their mean to the proportion of positive sets across the whole dataset. Alternatively, the mean could be set to the mean of the average annual proportions of positive sets. However, the main aim with this approach is to obtain a CPUE that varies appropriately, since variability for a binomial is greater when the mean is at 0.5 than at 0.02 or 0.98, but the multiplicative effect of the variability is greatest when the mean is low. Lognormal positive time effects were obtained by exponentiating the time effects from the glm. This approach does not provide an uncertainty estimate for the base temporal effect, but comprehensive estimates of observation error were of lesser interest to us in this study. The outcomes were reported as relative CPUE with mean of 1.

RESULTS AND DISCUSSION

Data exploration

The great majority of effort employed between 9 and 12 HBF (Figure 5), and 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 year the SBT catch rates are again higher than other species (Figure **7**). Similarly, the proportion of sets reported with zero SBT catches was low throughout the time series in the southern areas 7 to 9. The majority of sets reported no yellowfin catch, and the same applied to bigeye and albacore. However in area 9 the proportion of sets with zero albacore catches increased in the late 2000's, and there was a decrease in the proportions of zero albacore catches from about 2000-2010. 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. The majority of 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 significant 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 is 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).

CPUE standardization

Table 1 shows the results of dropping each variable from the lognormal constant GLMs. These results suggest 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 this results in overestimation of statistical significance.

Comparison of standardized and unstandardized CPUE series shows them to be quite similar (Figure **14**). Similar patterns are also observed for the eastern (SA8) and western (SA9) indices.

The influence plots for vessel effects (Figure **15**), spatial effects (Figure **16**), hooks (Figure **17**), month (Figure **18**), and lunar illumination (Figure **19**) 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. Vessel effects were quite variable with a few vessels having significantly lower SBT catch rates. There was no clear trend to the influence on the year effects, but on average the vessel effects were higher at the end of the time series than at the start. Spatial effects showed more variation among areas with low and high catch rates, and in the western area 9 there was a trend through time towards fishing in areas with lower average catch rates.

The effects of hook numbers on catch rates were difficult to interpret. In eastern area 8 there were only small differences by hook number across the range of data with most hooks, and minimal influence on year effects. In western area 9, however there were larger differences, and the appearance of significant influence on the year effects. 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.

The effect of month was quite strong in both the eastern and western areas. In both areas the highest catch rates were obtained in July and August.

Catch rates also appeared to vary moderately with lunar illumination. Longline catch rates of other pelagic fish such as bigeye tuna are known to be affected by moon phase (Poisson et al., 2010). However 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. We omitted moon phase in the delta lognormal analyses.

Figure **20** shows diagnostics for the lognormal constant GLM, with frequency distributions and QQ-plots suggesting that the data fit the GLM adequately.

Delta lognormal indices (Figure **21**) are quite similar to the lognormal constant indices, and the patterns are also relatively similar 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 of indices for eastern area 8 from 2003-2007, or western area 9 in 2005 (Table **2**).

REFERENCES

Bentley, N., Kendrick, T.H., Starr, P.J., Breen, P.A. (2011) Influence plots and metrics: tools for better understanding fisheries catch-per-unit-effort standardizations. *Ices Journal of Marine Science* **69,** 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,** 209-227.

Hoyle, S.D., Langley, A.D., Campbell, R.A. (2014) GUIDELINES FOR PRESENTING CPUE INDICES OF ABUNDANCE FOR WCPFC STOCK ASSESSMENTS.

Hoyle, S.D., Okamoto, H. (2011) Analyses of Japanese longline operational catch and effort for bigeye and yellowfin tuna in the WCPO. No. WCPFC-SC7-SA-IP-01.

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

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. In: *20th Extended Scientific Committee of the CCSBT*. Incheon, Republic of Korea.

Lo, N.C.H., Jacobson, L.D., Squire, J.L. (1992) Indices of relative abundance from fish spotter data based on delta-lognormal models. *Canadian Journal of Fisheries and Aquatic Sciences* **49,** 2515-2526.

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

Poisson, F., Gaertner, J.-C., Taquet, M., Durbec, J.-P., Bigelow, K. (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,* Vol., I-ATTC, LA JOLLA, CA ( ).

R Core Team (2014) R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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

	Stat area 8			Stat area 9		
	Df	Deviance	ΔAIC	Df	Deviance	ΔAIC
<none>		34.859	0		87.419	0
op_yr	13	40.248	1189	17	106.257	2090
latlong	10	36.338	331	14	98.212	1239
ns(hooks, 10)	10	35.116	41	10	89.557	243
vessid	13	35.803	199	17	90.337	324
ns(op_mon, df = 4	4	36.285	330	4	97.787	1212
ns(moon, df = 4)	4	35.549	157	4	87.655	22

Table 2: Delta lognormal indices for statistical areas 8 and 9.

Year	Stat area 8	Stat area 9
1996	1.08	0.79
1997	0.61	0.36
1998	0.62	0.40
1999	0.63	0.56
2000	0.71	0.10
2001	0.63	0.54
2002	0.28	0.45
2003		0.80
2004		0.02
2005		
2006		0.13
2007		0.04
2008	1.11	0.23
2009	0.58	0.08
2010	0.79	0.23
2011	0.95	2.11
2012	1.07	0.64
2013	1.70	1.65
2014	1.55	3.36

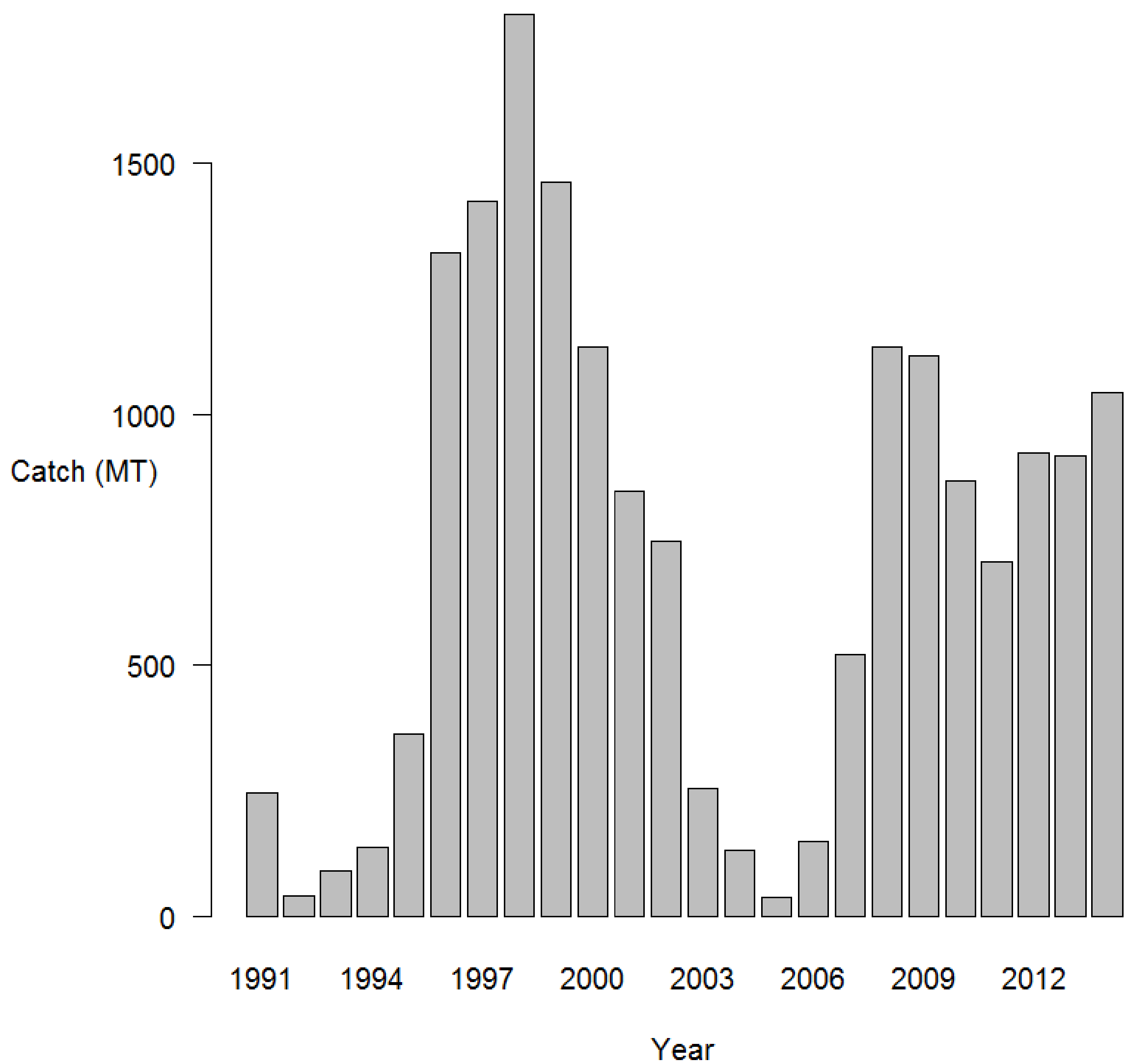

Figure 1: The annual Korean SBT catches in the CCSBT convention area, 1991 - 2014.

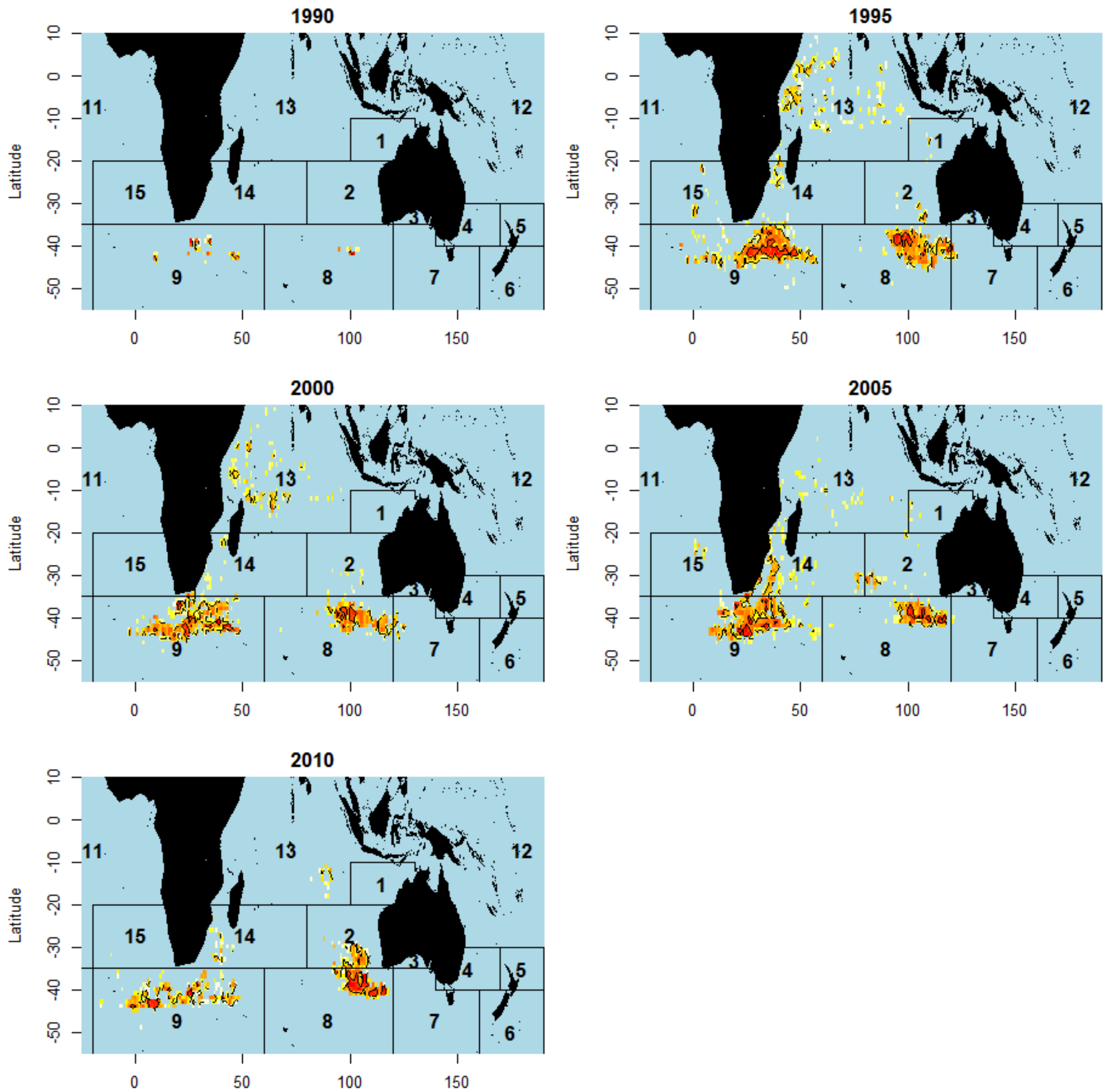

Figure 2: Map showing the core areas of Korean tuna longline vessels fishing for SBT, aggregated by 5 year period. Red colour indicates higher fishing effort, in numbers of hooks.

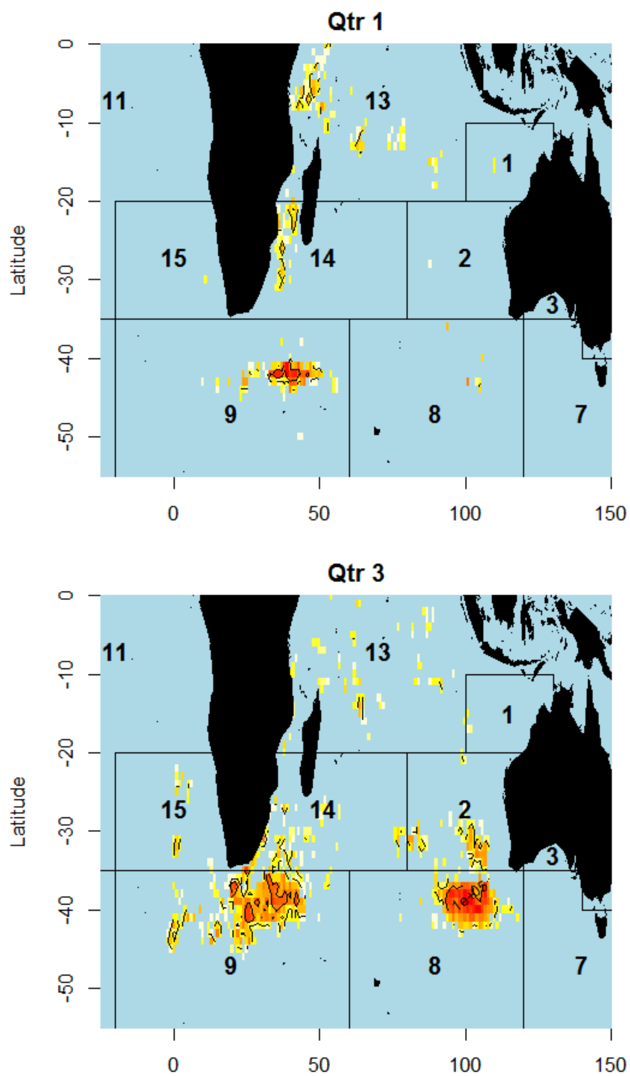

Figure 3: Map showing the core areas of Korean tuna longline vessels fishing for SBT, by quarter. Red colour indicates higher fishing effort, in numbers of hooks. Data are aggregated across the period 1994-2014.

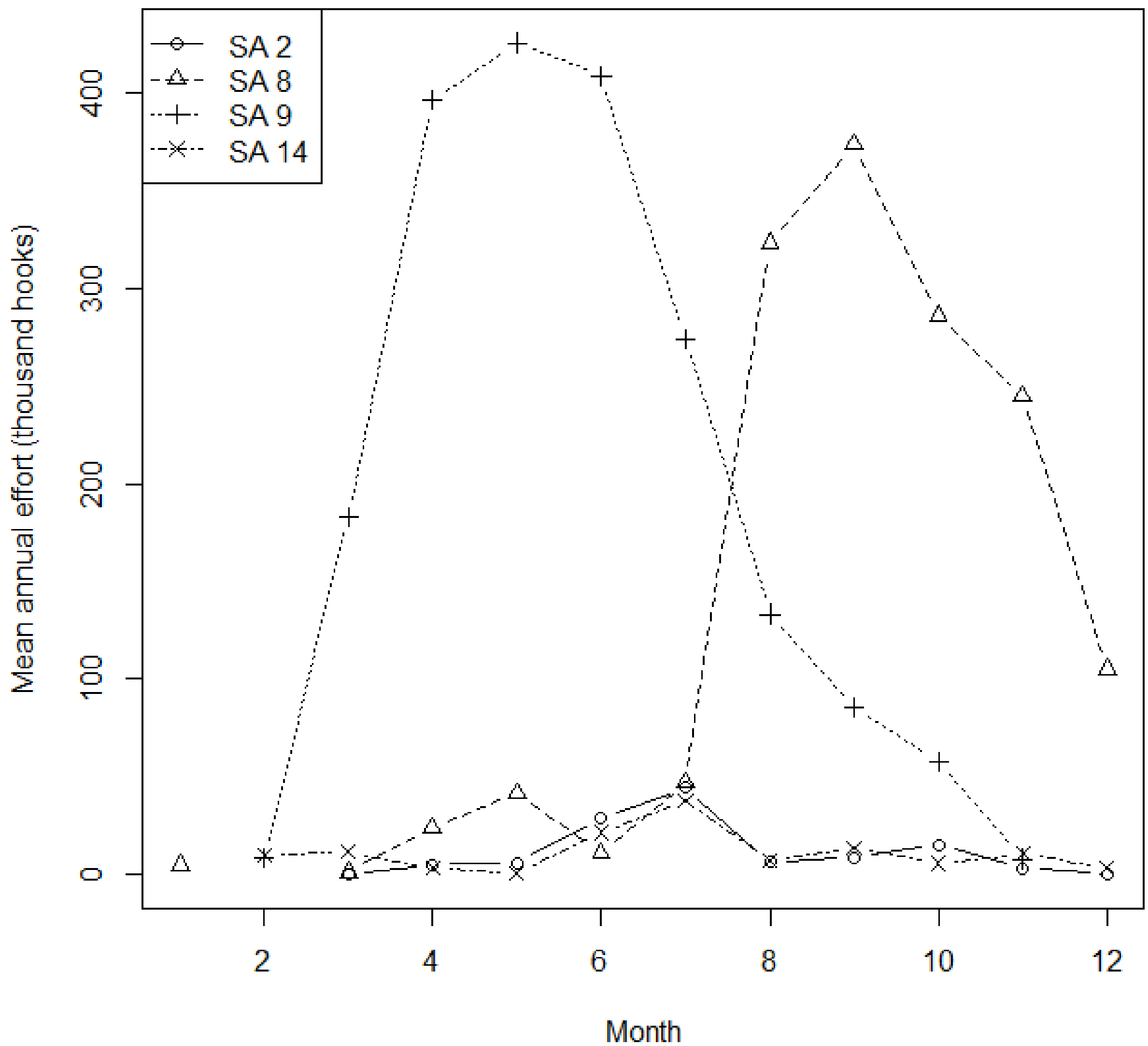

Figure 4: Mean annual effort in thousands of hooks, by month and statistical area.

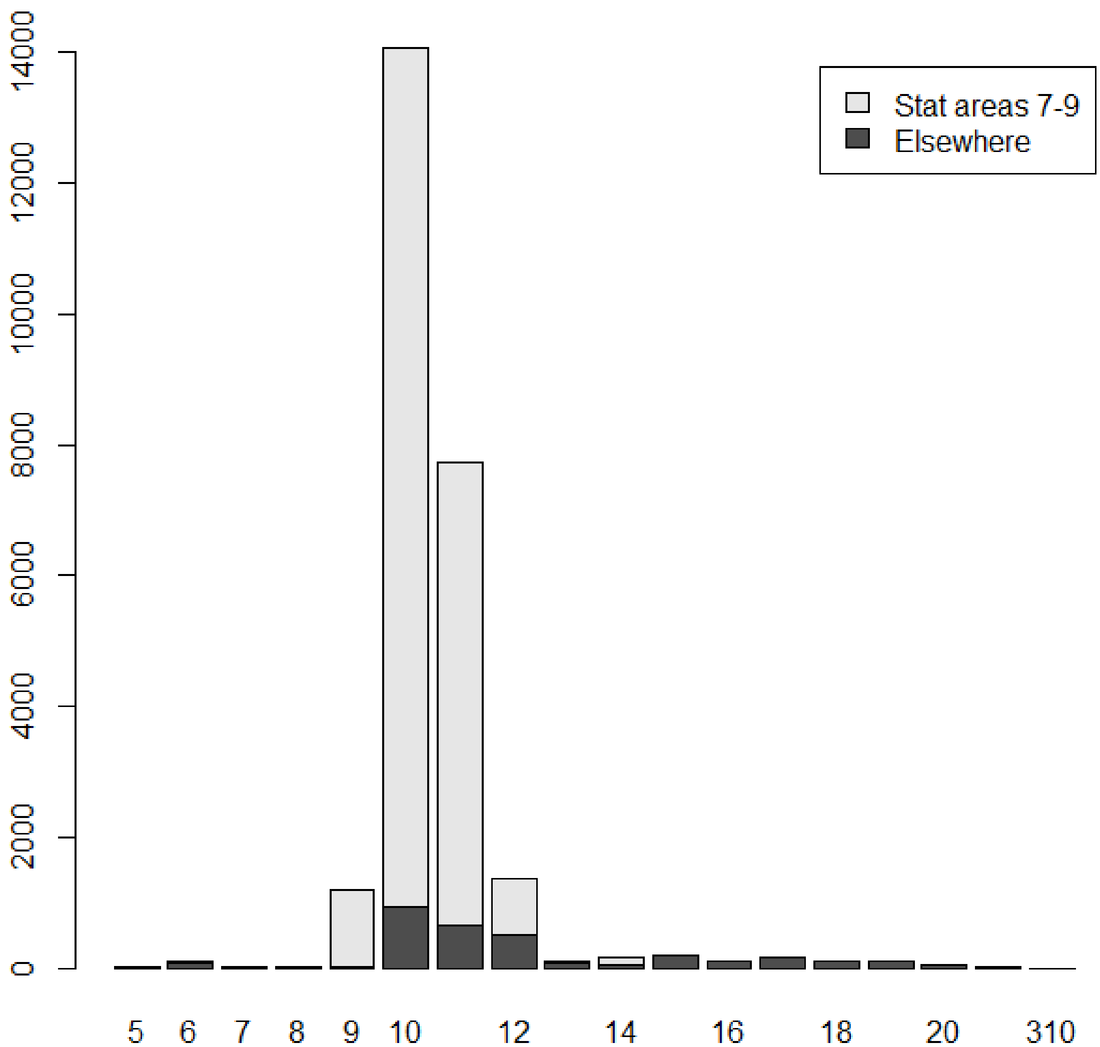

Figure 5: Frequency table of HBF for the main fishing ground with the lighter shade for statistical areas 7-9, and the darker shade for other areas.

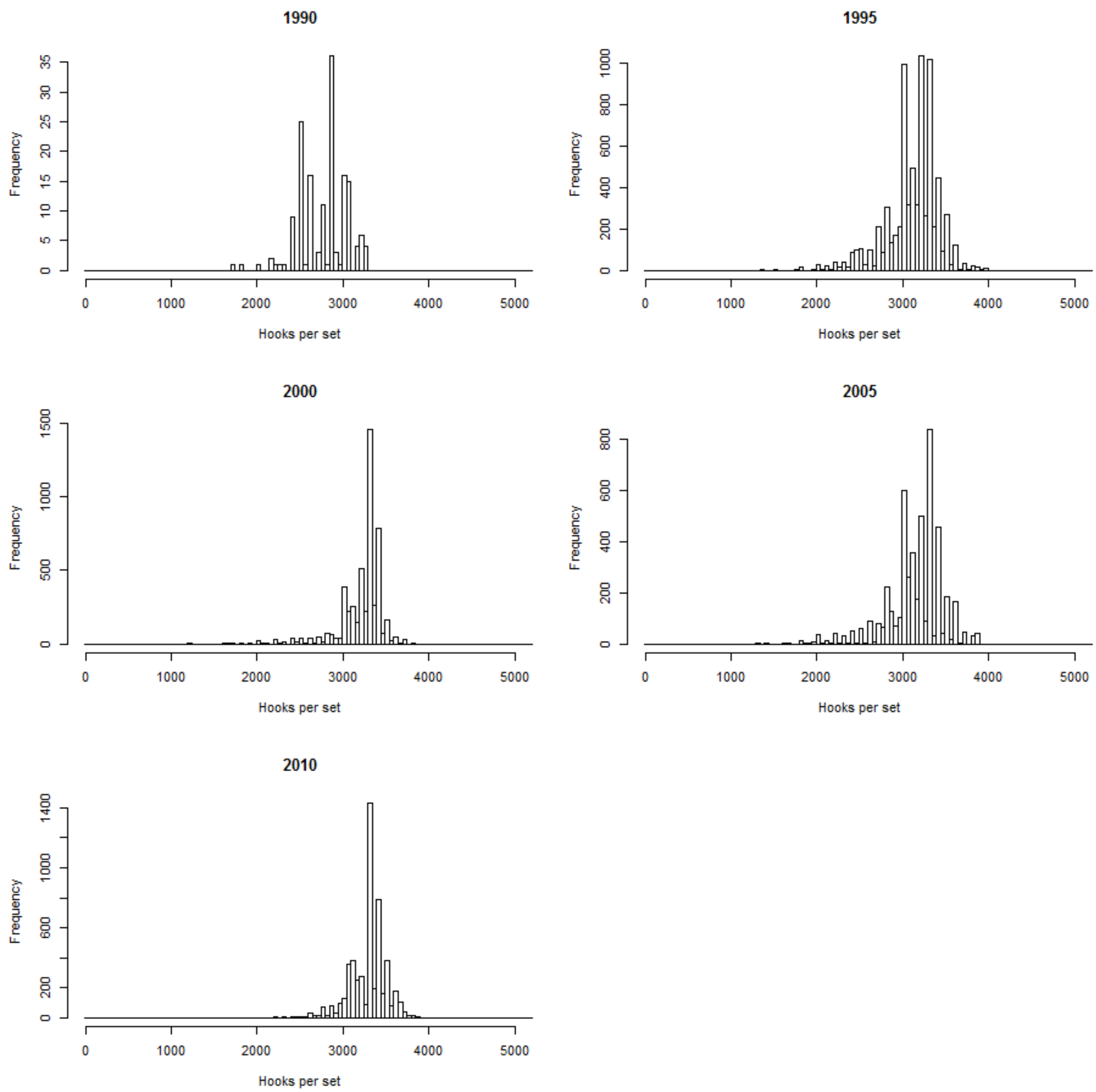

Figure 6: Distribution of hooks per set per 5 year period, for sets in the SBT-targeting areas south of 35S.

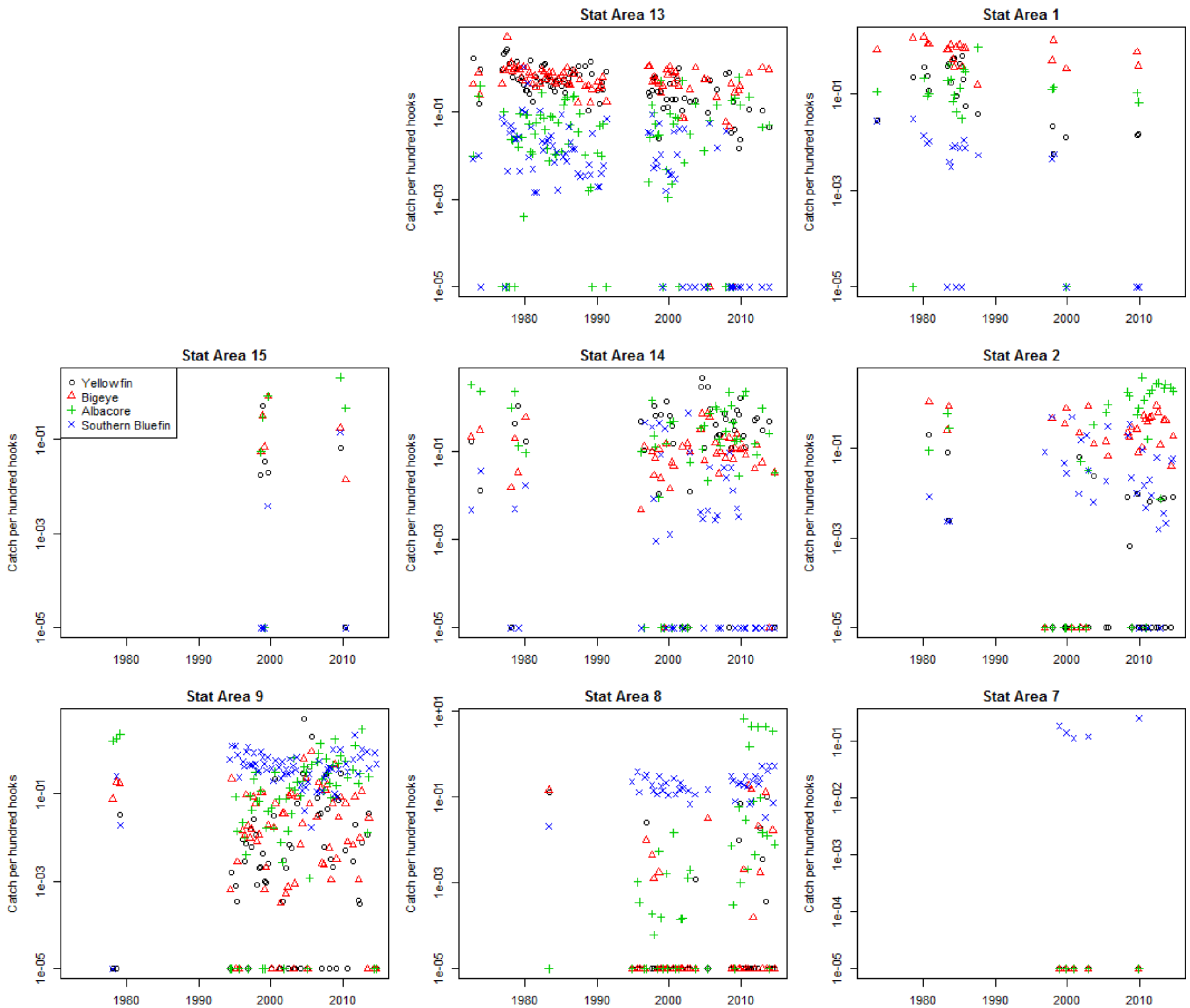

Figure 7: Mean catch per hundred hooks by year-qtr, species, and statistical area, plotted on a log scale, for yellowfin, bigeye, albacore, and southern bluefin tuna. Each CPUE has 1E-5 added so that zero catches appear on the log scale.

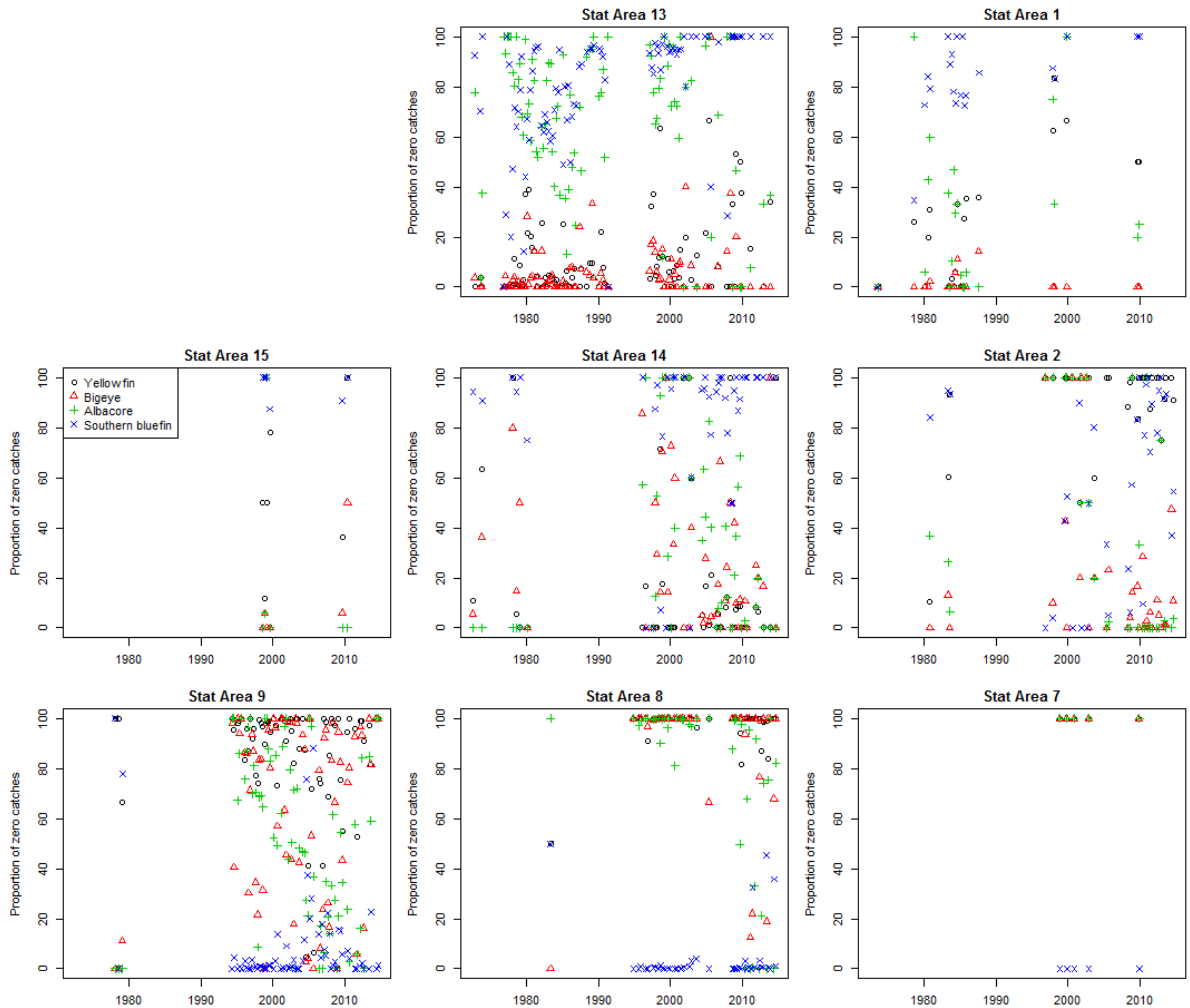

Figure 8: Proportion of zero catches per set by year-qtr, species, and statistical area, for yellowfin, bigeye, albacore, and southern bluefin tuna.

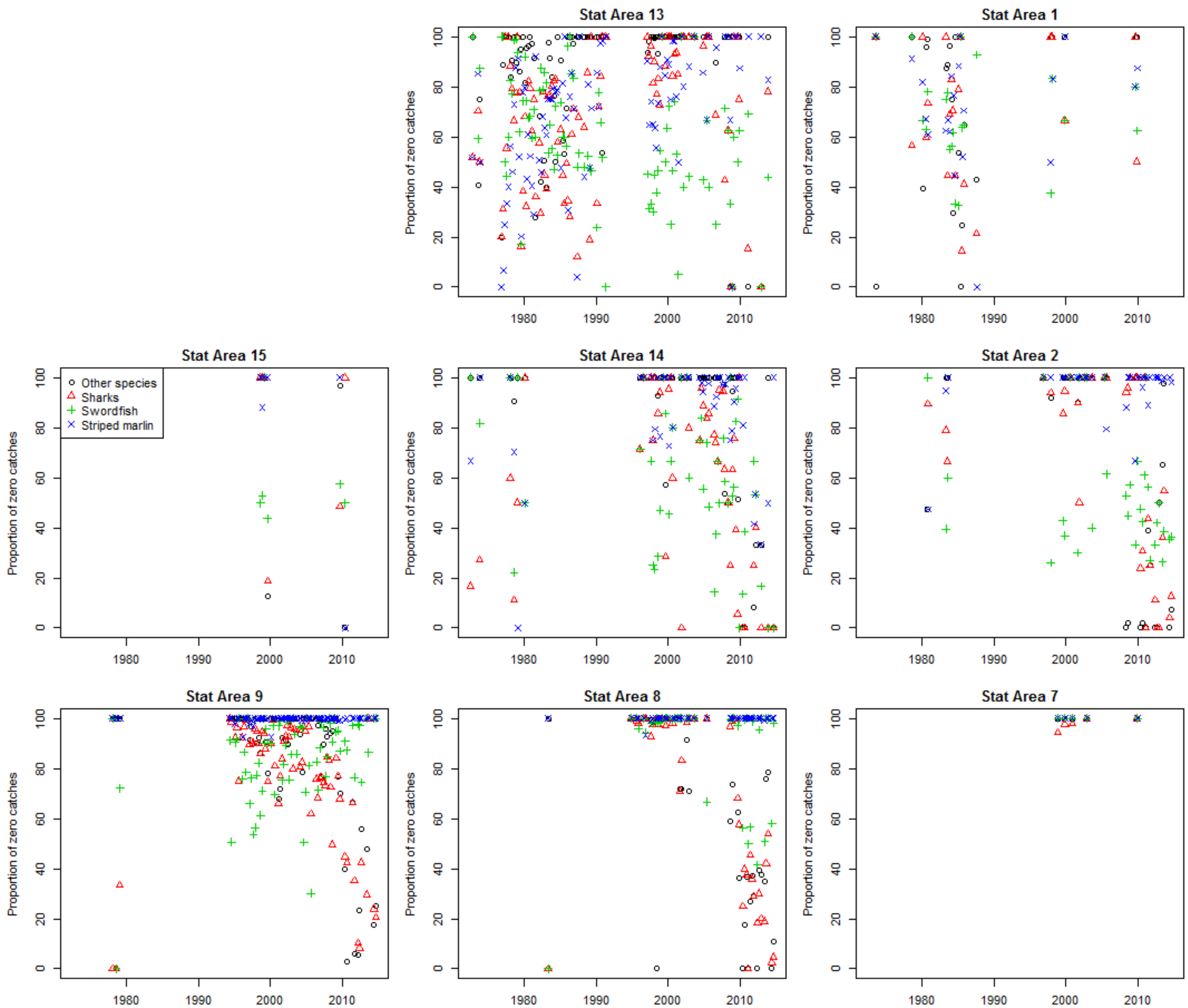

Figure 9: Proportion of zero catches per set by year-qtr, species, and statistical area for sharks, swordfish, striped marlin, and species not otherwise recorded (i.e. everything other than SBT, BET, YFT, ALB, SWO, BLM, BUM, MLS, SFA, SKJ, and SHA).

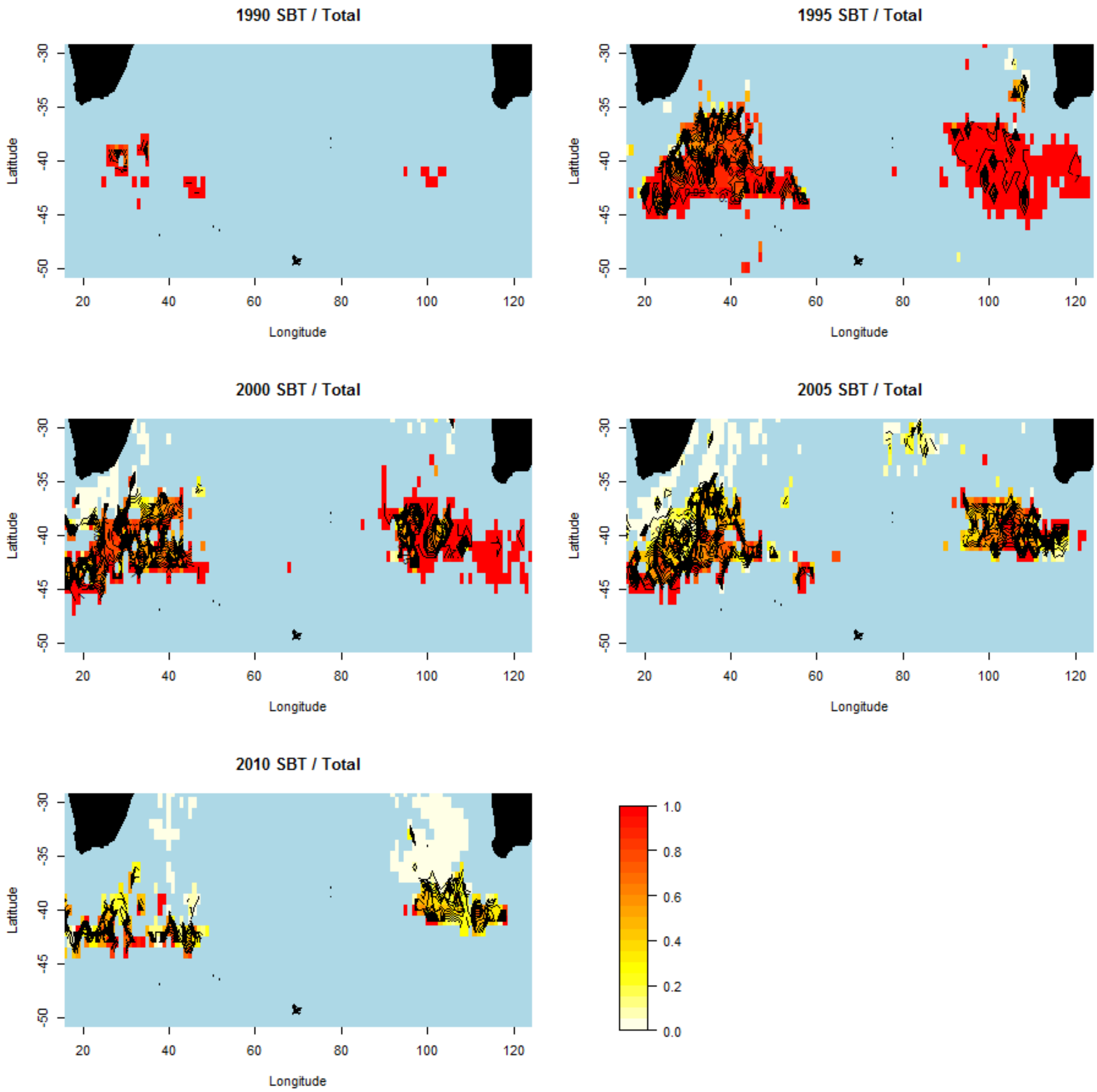

Figure 10: Proportion southern bluefin tuna (SBT) in the total reported catch in numbers by 1 degree square, aggregated over 5 years within the period 1990-2014. Red colour indicates a higher proportion of SBT.

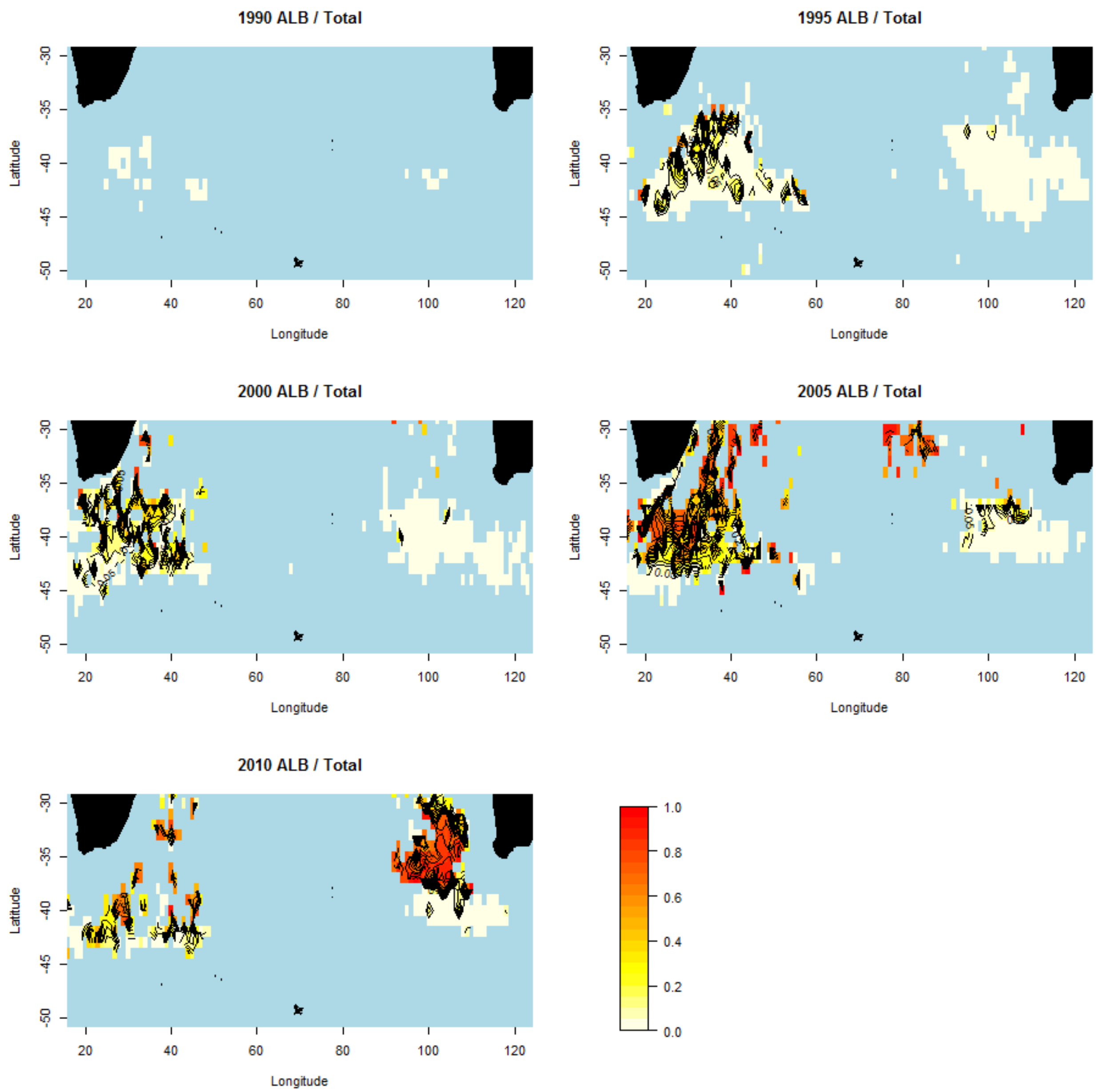

Figure 11: Proportion albacore (ALB) in the total reported catch in numbers by 1 degree square, aggregated over 5 years within the period 1990-2014. Red colour indicates a higher proportion of ALB.

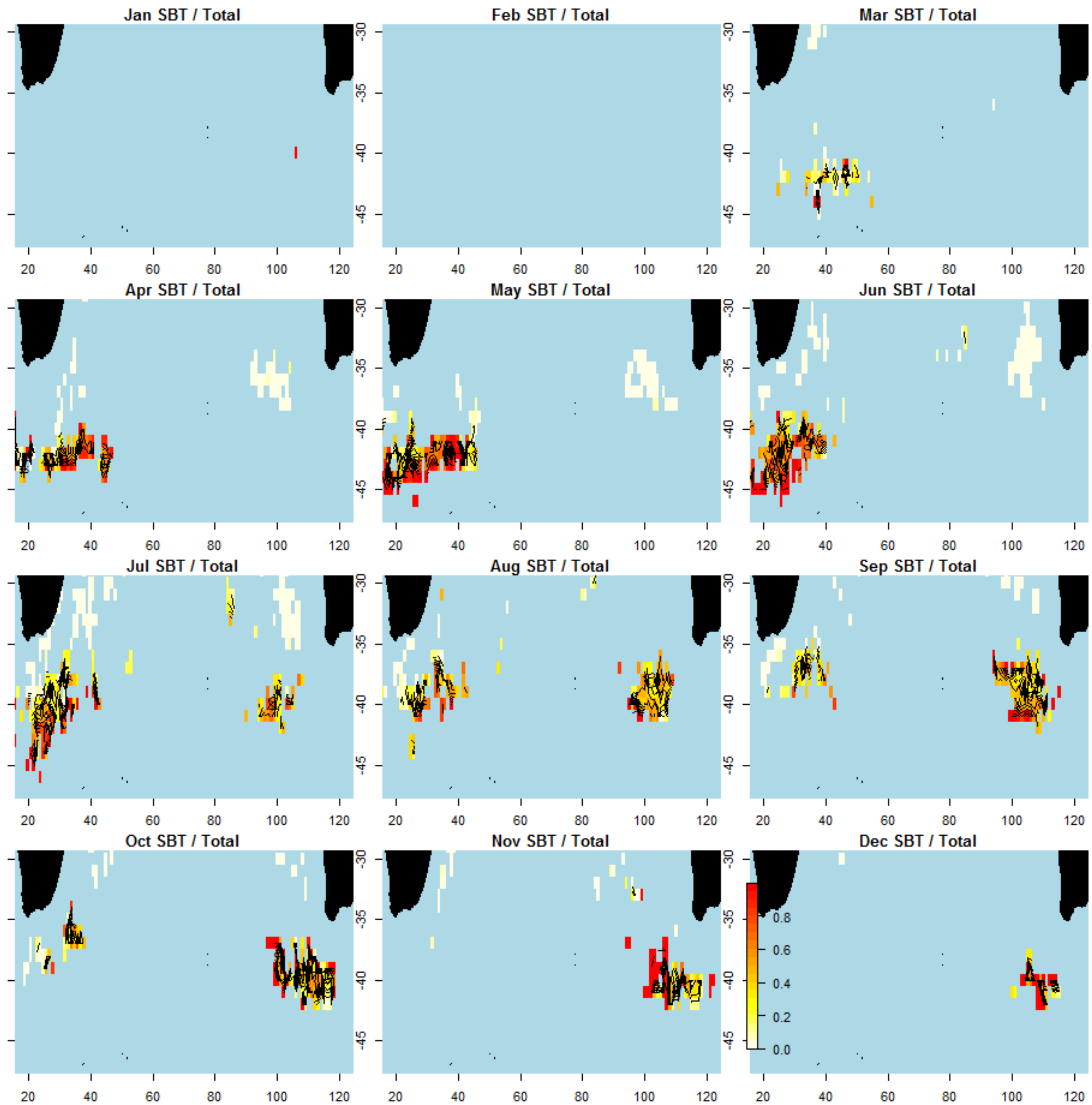

Figure 12: Proportion southern bluefin tuna (SBT) in the total reported catch in numbers by 1 degree square, by month, aggregated over the period 2005-2014. Red colour indicates a higher proportion of SBT.

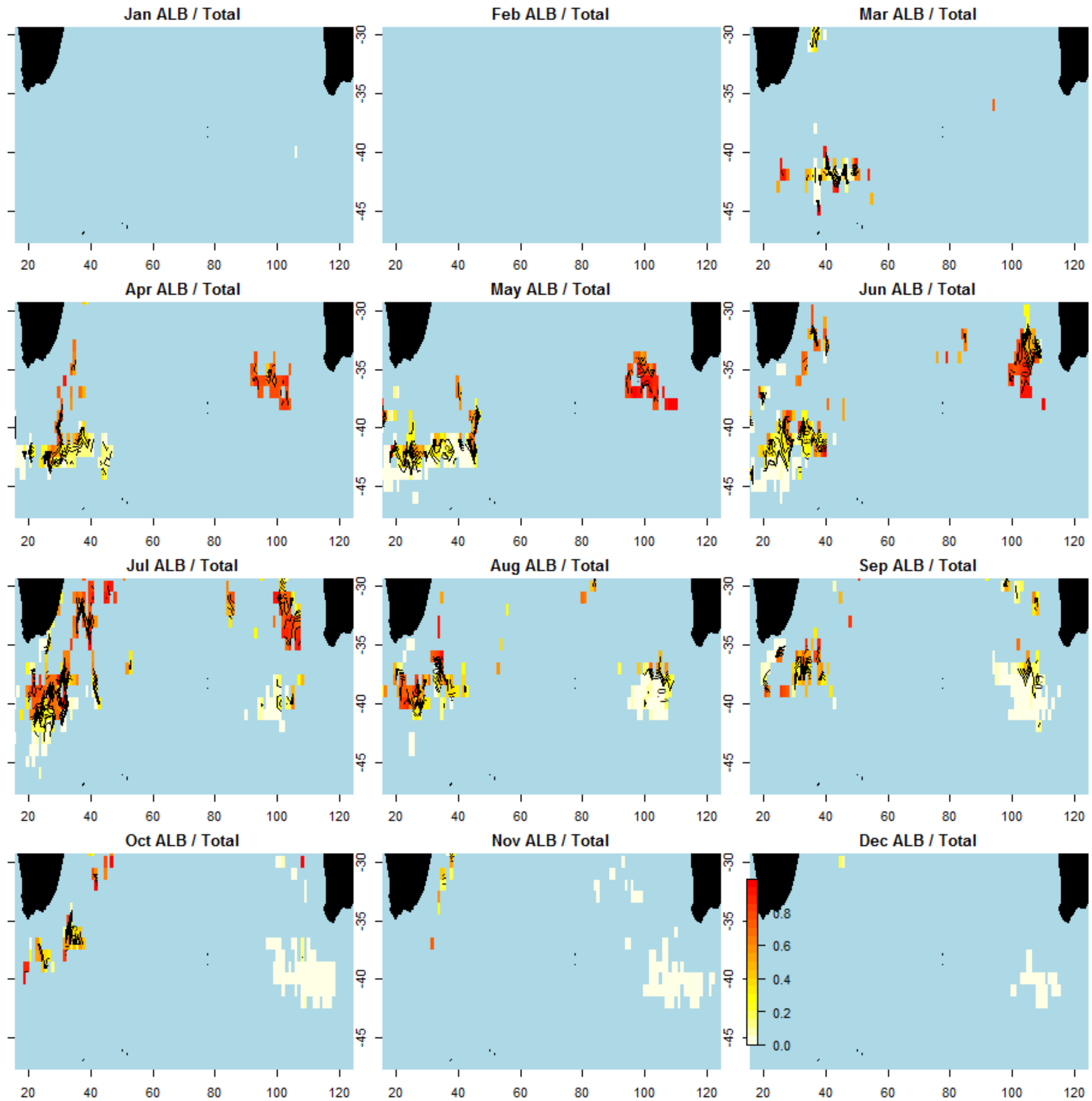

Figure 13: Proportion albacore (ALB) in the total reported catch in numbers by 1 degree square, by month, aggregated over the period 2005-2014. Red colour indicates a higher proportion of ALB

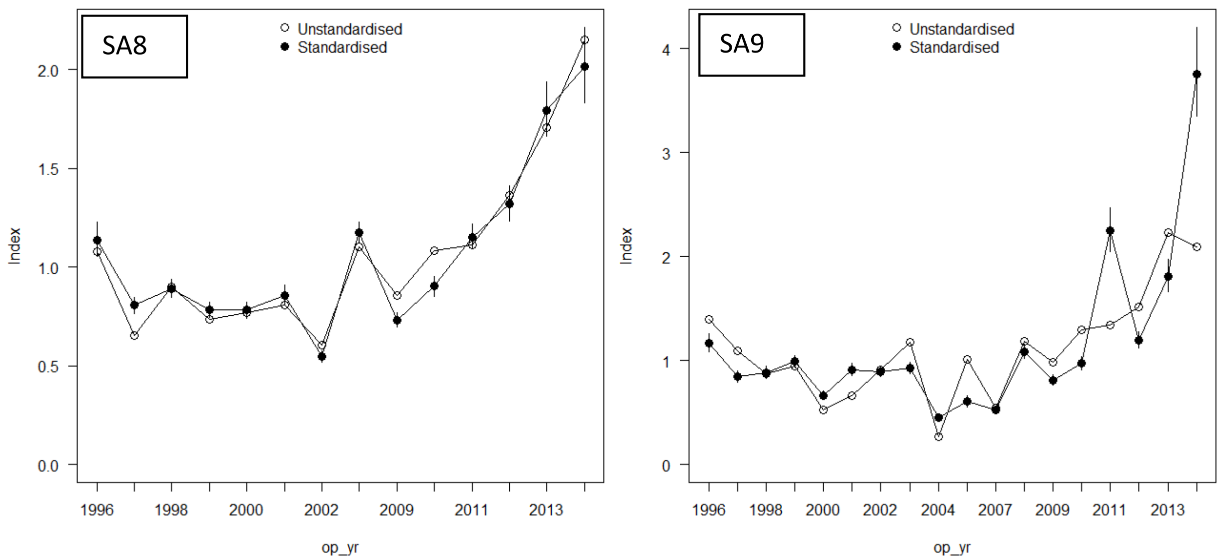

Figure 14: Unstandardized and standardized CPUE indices for statistical areas 8 (left) and 9 (right), based on lognormal GLMs with an added constant.

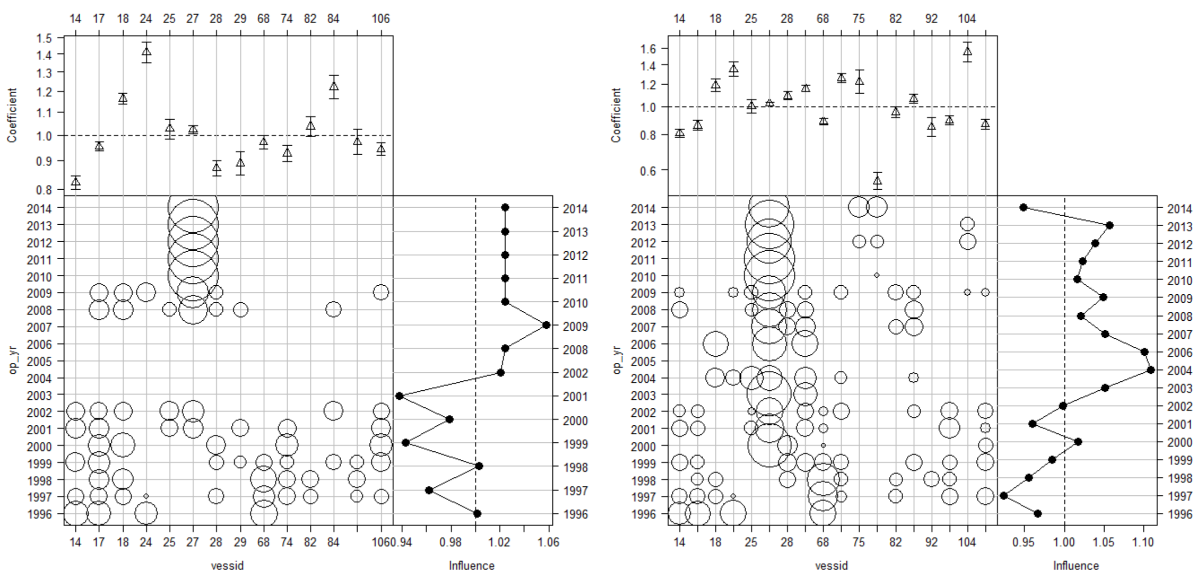

Figure 15: Influence plots for vessel effects for areas 8(left) and 9 (right).

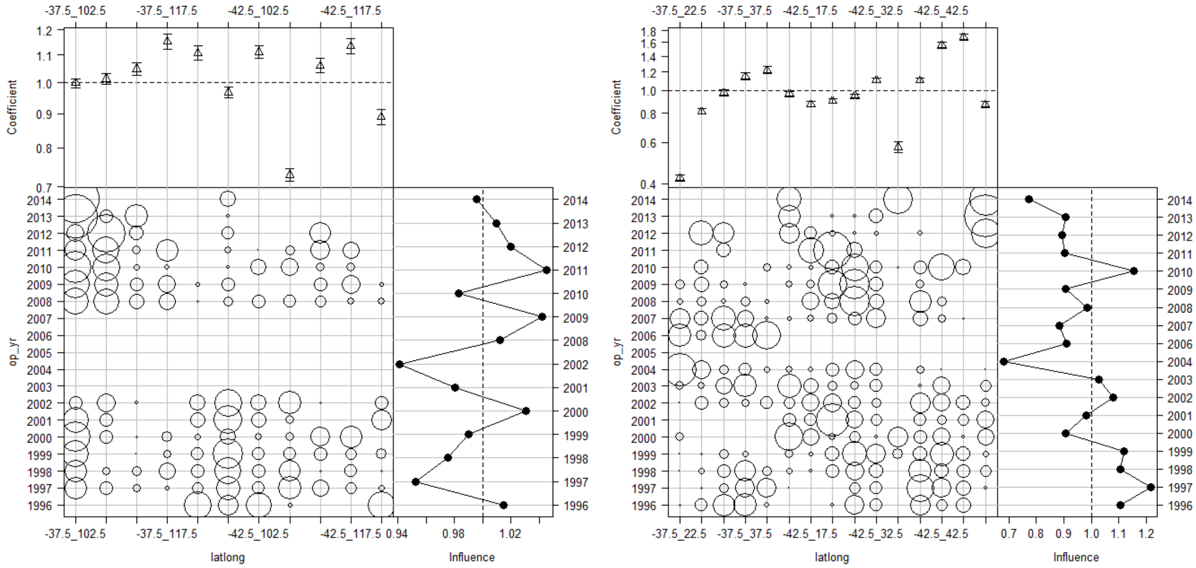

Figure 16: Influence plots for spatial latlong effects for statistical areas 8(left) and 9 (right).

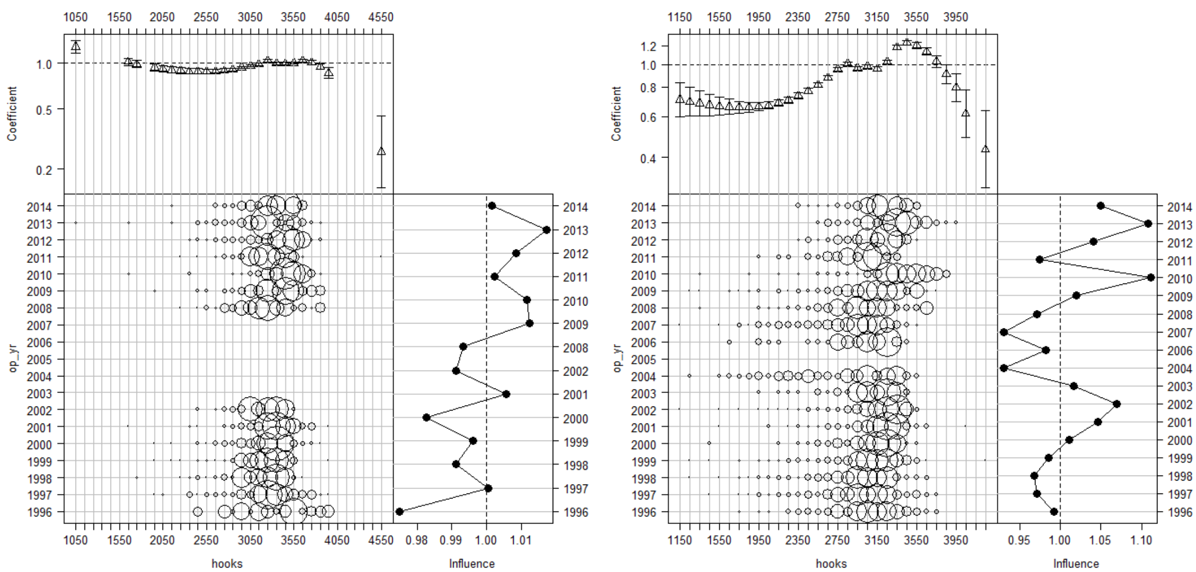

Figure 17: Influence plots for the effects of numbers of hooks for statistical areas 8(left) and 9 (right).

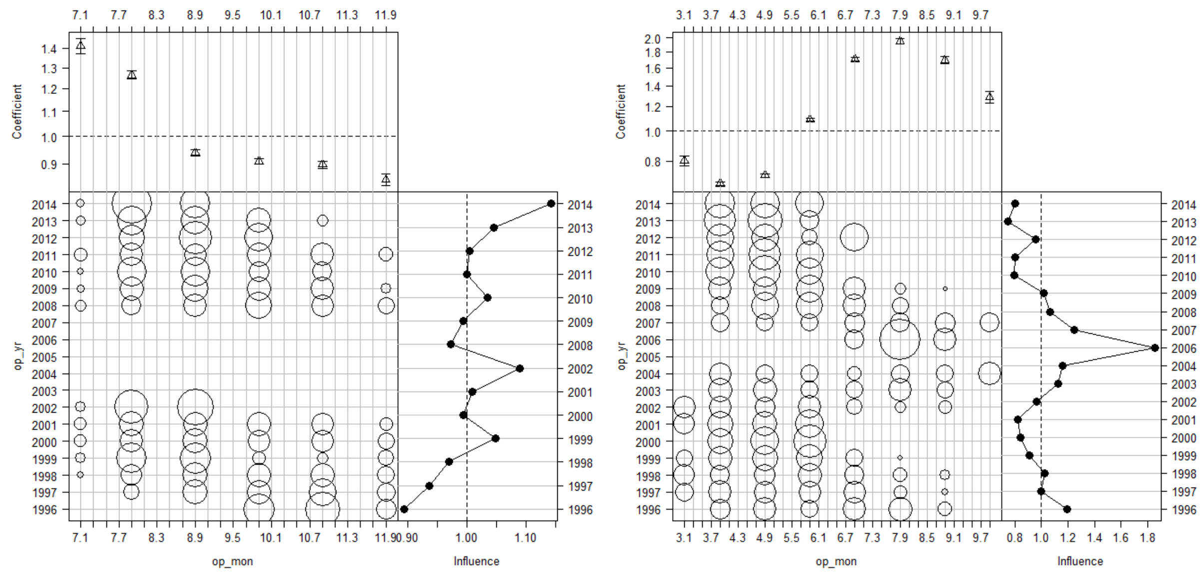

Figure 18: Influence plots for month effects for statistical areas 8(left) and 9 (right).

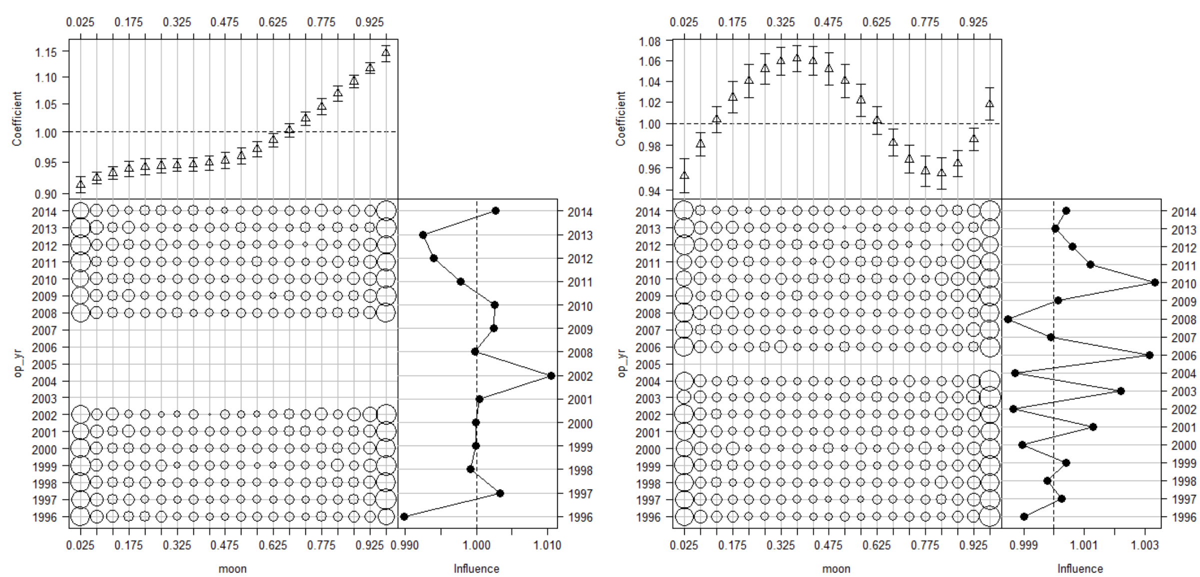

Figure 19: Influence plots for lunar illumination effects for statistical areas 8(left) and 9 (right).

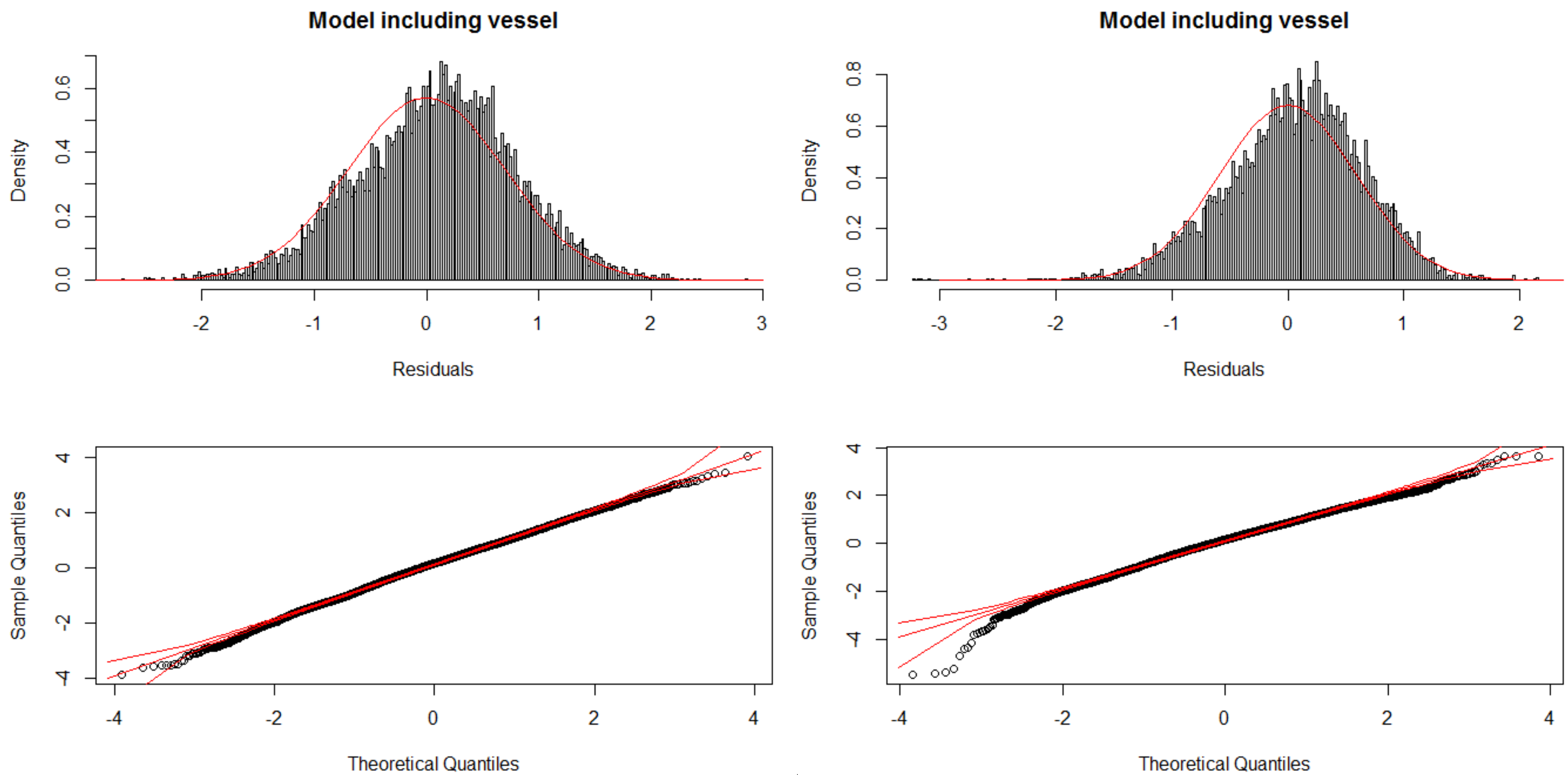

Figure 20: Frequency distributions of the standardized residuals (above) and Q-Q plots of standardized residuals for lognormal constant GLM analyses of statistical areas 8(left) and 9(right).

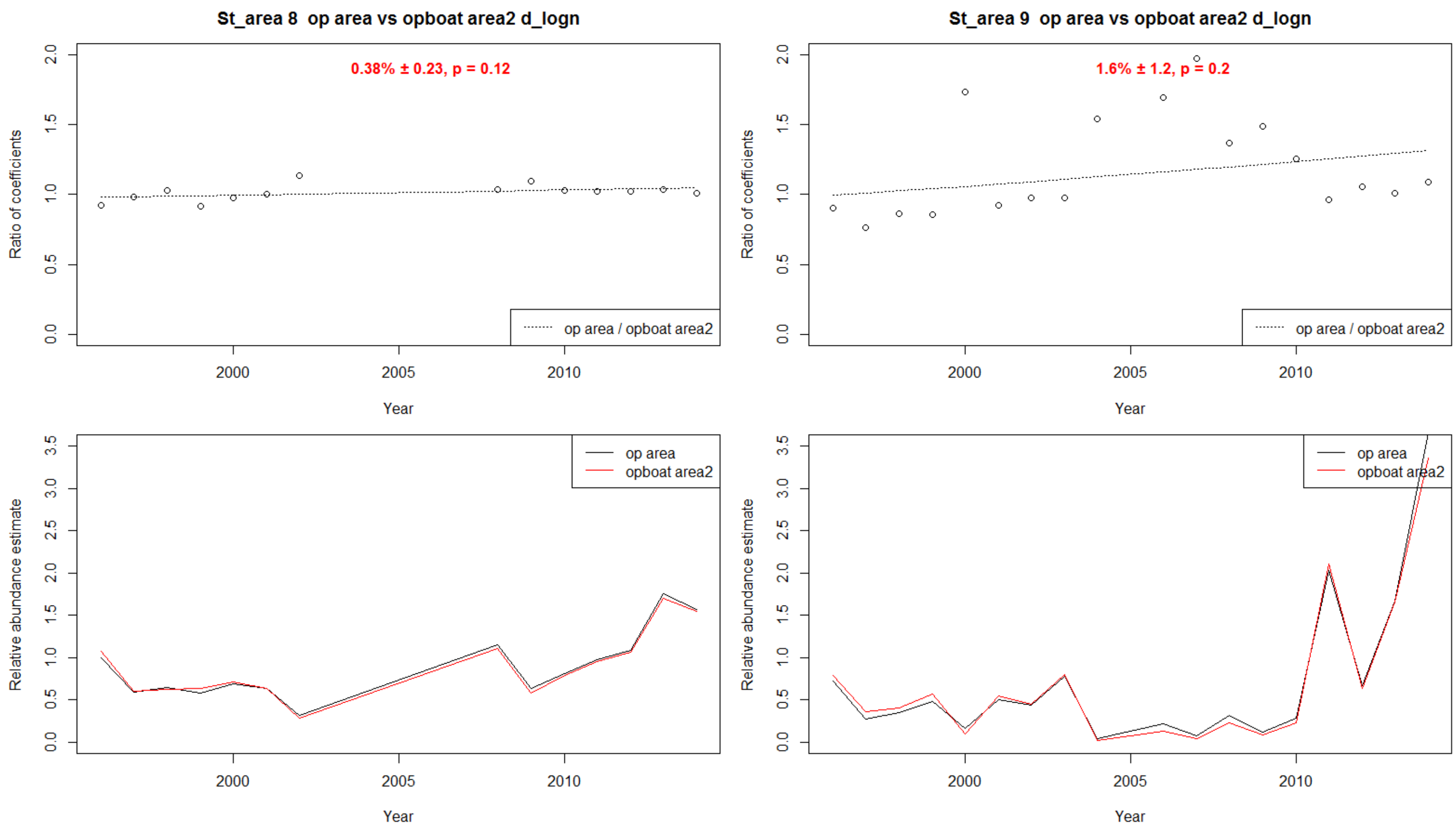

Figure 21: Plots of annual indices of abundance resulting from standardization of SBT CPUE for statistical areas 8 (left) and 9 (right), modelled either with (opboat area) or without (op area) vessel effects. The upper plots show the ratio of the two indices, and the lower plots show the indices.