Received: 17 October
IOTC–2017–WPTT19–26

STANDARDIZATION LONGLINE CATCH PER UNIT EFFORT FOR BIGEYE (*THUNNUS OBESUS*) AND YELLOWFIN TUNA (*THUNNUS ALBACARES*) FROM SOUTH AFRICA

Henning Winker, Denham Parker, Charlene da Silva & Sven E. Kerwath

SUMMARY

Bigeye tuna (*Thunnus obesus*) and yellowfin tuna (*Thunnus albacares*) are important target species for the Japanese flagged vessels, which operate under South African joint-venture agreement in the IOTC region of the South African EEZ. The standardization of catch per unit effort (CPUE) from the joint-venture fleet segment for the period 2004-2016 was carried out using Generalized Additive Mixed Models (GAMMs) with a Tweedie distributed error. Explanatory variables of the final model included *year*, *month*, geographic position and a *targeting* factor with 2 levels, derived by clustering of PCA scores of the root-root transformed, normalized catch composition. Vessels that fished for at least two years were included as a random effect. Standardized bigeye tuna CPUE showed a strong seasonal trend, with catch rates highest between April and July. The standardized CPUE index showed a decline between 2005 and 2008, a slight increase between 2008 and 2010 and a fairly stable trend between 2010 and 2015 and a slight increase again in 2016. Yellowfin tuna showed less pronounced seasonal trend, which peaked in between July and August. The standardised CPUE index for yellowfin tuna showed a sharp decline between 2004 and 2012, followed by a slight increase until 2016. We anticipate that the here presented standardized abundance indices for bigeye tuna and yellowfin tuna could be useful for corroborating other abundance indices for the South-West Indian Ocean.

KEYWORDS

Tropical tunas, standardized cpue, longline, GAMM, targeting, PCA cluster, random effect

AFFILIATIONS

Department of Agriculture, Forestry and Fisheries (DAFF), South Africa

INTRODUCTION

Commercial fishing for large pelagic species in South Africa dates back to the 1960s (Welsh, 1968; Nepgen, 1970). Exploitation of large pelagic species in South Africa can be divided into four sectors, 1) pelagic longline, 2) tuna pole-line 3) commercial linefishing (rod and reel) and 4) recreational line-fishing. Pelagic longline fishing by South African vessels began in the 1960s with the main target being the temperate southern bluefin tuna (*Thunnus maccoyii*) and albacore (*Thunnus alalunga*) (Welsh, 1968; Nepgen, 1970). This South African fishery ceased to exist after the mid 1960's, as a result of a poor market for low quality southern bluefin and albacore (Welsh, 1968). However, foreign vessels, mainly from Japan and Chinese-Taipei, continued to fish in South African waters from the 1970s until 2002 under a series of bilateral agreements.

The commercial longline fishery was only formalised in 2004 with the issuing of 10-year long term rights to the mainly swordfish-directed domestic vessels. On average, 15 domestic South African vessels are active during a year, which target a mix of swordfish from 20-30 m length vessels. Additionally, a varying number (2-12) of Japanese flagged vessels have been fishing under South African Joint-Venture agreement in South Africa’s EEZ. These vessels mostly operate east of the 20°E longitude boundary (Figs. 1 & 2), which separates the Atlantic from the Indian Ocean and thus the jurisdictions of the Indian Ocean Tuna Commission (IOTC) and International Commission for the Conservation of Atlantic Tuna (ICCAT). These joint-venture vessels are considered tuna specialists that mainly target bigeye, yellowfin tuna (*Thunnus albacares*) and Southern Bluefin tuna (*Thunnus maccoyii*). Joint-venture vessels fish under South African fisheries regulations and all catch accrue to South Africa. Observe presence is mandatory for all foreign flagged vessels fishing within South Africa’s, so that observer coverage of this fleet segment 100%.

Here we use Generalised Additive Mixed Models (GAMM) to standardize set-by-set catch-per-unit-effort data of bigeye tuna from Japanese-flagged joint-venture fishing trips for the period 2004-2016. We focus on the Japanese joint-vessels because these are considered to mainly target tuna, whereas skippers of domestic vessels have historical very limited experience in targeting tunas compared to the highly specialised Japanese skippers. As a result domestic vessels usual catch a broader mix of swordfish, sharks and tunas. Catch and effort data were subsetted to ≥25°E of the IOTC area to create a buffer zone to 20°E boundary due to concerns that large yellowfin tuna that seasonally occur in the tuna pole catches off Cape Agulhas at the 20°E boundary could form part of an Atlantic feeding migration. The GAMMs were fitted using a Tweedie distribution and included year, month, latitude, longitude, fishing tactic (targeting) as fixed factors and had a random vessel effect. Targeting was determined by clustering PCA scores of the root-root transformed, normalized catch composition.

MATERIALS AND METHODS

Catch and effort data preparation

All longline records from Japanese flagged Joint-Venture vessels for longline start positions ≥25°E were extracted from the database for the period 2004-2016, excluding 2006, due to the absence of joint-venture fishing operations in this year. The dataset was further subsetted to vessels that fished for at two years. This resulted in a total of sample size of 5709 sets, equating to 1.54 mio hooks. Each record included the following information: (1) date, (2) unique vessel number, (3) start position of the set in latitude and longitude and (4) mandatory catch reports in kilogram per species per set and (5) hooks per set.

Model framework

CPUE was standardized using Generalized Additive Mixed Models (GAMMs), which included the covariates year, month, degree latitude (Lat) and longitude (Long) coordinates of longline start positions and vessel as random effect. In an attempt to account for variation in fishing tactics, we considered an additional factor for targeting derived from a cluster analysis of the catch composition (He *et al.*, 1997; Carvalho *et al.*, 2010; Winker *et al.*, 2013). For the clustering analysis, all CPUE was modelled as catch in metric tons per species per vessel per day. All of the following analysis was conducted within the statistical environment R. The R package 'cluster' was used to perform the CLARA analysis, while all GAMMs were fitted using the 'mgcv' and 'nlme' libraries described in Wood (2006).

Clustering of the catch composition data was conducted by applying a non-hierarchical clustering technique known as CLARA (Struyf *et al.*, 1997) to the catch composition matrix. To obtain the input data matrix for CLARA, we transformed the $CPUE_{i,j}$ matrix of record *i* and species *j* into its Principal Components (PCs) using Principal Component Analysis (PCA). For this purpose, the data matrix comprising the $CPUE_{i,j}$ records for all reported species was extracted from the dataset. The CPUE records were normalized into relative proportions by weight to eliminate the influence of catch volume, fourth-root transformed and PCA-transformed. Subsequently, the identified cluster for each catch composition record was aligned with the original dataset and treated as categorical variable (FT) in the model (Winker *et al.*, 2013). To select the number of meaningful clusters we followed the PCA-based approach outlined and simulation-tested in Winker et al. (2014). This approach is based on the selection of non-trivial PCs through non-graphical solutions for Catell's Scree test in association with the Kaiser-Guttman rule (Eigenvalue > 1), called Optimal Coordinate test, which available in the R package 'nFactors' (Raiche and Magis, 2010). The optimal number of clusters considered is then taken as the number of retained PCs plus one (Winker *et al.*, 2014). The results suggest that only the first PC is non-trivial (Fig. 3) and correspondingly two clusters were selected as optimal for the CLARA clustering (Fig. 4).

The CPUE records were fitted by assuming Tweedie distribution (Tascheri *et al.*, 2010; Winker *et al.*, 2014). The Tweedie distribution belongs to the family of exponential dispersion models and is characterized by a two-parameter power mean-variance function of the form $Var(Y) = \phi\mu p$, where ϕ is the dispersion parameter, μ is the mean and p is the power parameter (Dunn and Smyth, 2005). Here, we considered the case of $1 < p < 2$, which represents the special case of a Poisson ($p = 1$) and gamma ($p = 2$) mixed distribution with an added mass at 0. This makes it possible to accommodate zeros in combination with right-skewed continuous numbers in a natural way when modelling CPUE data (Winker *et al.*, 2014; Ono *et al.*, 2015). As it is not possible to estimate the optimal power parameter p internally within GAMMs, p was optimized by iteratively maximizing the profile log-likelihood of the GAMM for $1 < p < 2$ (Fig. 5). This resulted in a power parameter of p =1.55 for both species with associated dispersion parameters of ϕ = 7.6 and ϕ = 8.4 for the full GAMMs fitted bigeye and yellowfin tuna, respectively. The full GAMM evaluated for both species was:

$$CPUE_i = exp(\beta_0 + Year + s_1(Month) + s_2(Long, Lat) + FT + \alpha_V)$$

where $s_1()$ denotes cyclic cubic smoothing function for *Month*, $s_2()$ a thin plate smoothing function for the two-dimensional covariate of *Lat* and *Long*, *FT* is the vector of cluster numbers treated as categorical variable for “Fishing Tactics”, and α_v is the random effect for Vessel *v* (Helser *et al.*, 2004). The inclusion of individual Vessels as random effects term provides an efficient way to combine CPUE recorded from various vessels (n = 25) into a single, continuous CPUE time-series, despite discontinuity of individual vessels over the time series (Helser *et al.*, 2004). The main reason for treating vessel as a random effect was because of concerns that multiple CPUE records produced by the same vessel may violate the assumption of independence caused by variations in fishing power and skipper skills and behaviour, which can result in overestimated precision and significance levels of the predicted CPUE trends if not accounted for (Thorson and Minto, 2015). The significance of the random-effects structure of the GAMM was supported by both Akaike's Information Criterion (AIC) and the more conservative Bayesian Information Criterion (BIC). Sequential *F*-tests were used to determine the covariates that contributed significantly ($p < 0.05$) to the deviance explained.

Annual CPUE was standardized by fixing all covariates other than *Year* and *Lat* and *Long* to a vector of standardized values X_0. The choices made were that *Month* was fixed to the months where catches were on average highest, and *FT* was fixed to the fishing tactic the produced highest average catch rates (*FT* = 1 for bigeye and *FT* = 2 for yellowfin). The expected yearly mean $CPUE_y$ and standard-error of the expected log(*CPUEy*) for the vector of standardized covariates X_0 were then calculated as average across all *Lat*-*Long* combinations (here forth grid cells) *a*, such that:

$$E[CPUE_y(X_0^T\hat{\beta})] = \frac{1}{A}\sum_{a}^{A} e\,xp(\hat{\mu}_{y,a})$$

and

$$\hat{\sigma}_y\,(X_0^T\,\hat{\beta}) = \sqrt{\frac{1}{A}\sum_{a}^{A}\hat{\sigma}_{y\,,a}^2}$$

where $\hat{\mu}_{y\,,a}$ is the standardized, model-predicted $log(CPUE_{y,a})$ for *Year y* and *Lat* and *Long* for grid cell a, $\hat{\sigma}_{y\,,a}$ is the estimated model standard error associated with $log(CPUE_{y,a})$, *A* is the total number of grid cells and *T* denotes the matrix in which *X* is transposed.

RESULTS AND DISCUSSION

Bigeye tuna and yellowfin tuna were the dominant target species of the Japanese flagged joint-venture vessels, making up for 37% and 43% of the total catch, respectively. The average catch composition of large pelagic species in each of the two fishing clusters showed that cluster 1 comprised a higher proportion the temperate tuna species southern bluefin tuna and albacore as well as swordfish , whereas yellowfin tuna was the dominant species in cluster 2, contributing 55% to of the catches in weight (Table 1). Bigeye tuna was the dominant species in cluster 1 (39%), but almost contributed equally to the catch in cluster 1 (35%).

The justification of using the Tweedie distribution is illustrated in Fig. 6, which shows adequate coverage of observed CPUE when compared to the expected CPUE generated from Monte-Carlo simulations based on the final GAMM fits. The inclusion of a vessel as a random effect and all considered fixed effects were supported by both the AIC and BIC (Tables 2 & 3). Given the notable variation among vessels (Fig. 7), it is unsurprising that the inclusion of the random vessel effect produced the most parsimonious error model.

The analysis of deviance for the step-wise procedure showed that most the covariates considered were highly significant (p < 0.001). The exceptions of covariates with lower significance levels were the *targeting* effect for bigeye tuna (p < 0.01; Table 2) and the month effect for yellowfin tuna (Table 3). For both species, the *year* effect accounted for close to 50% of the total variation in the data explained by the final GAMMs. The *month* effect was the second most important variable for bigeye tuna, explaining 28.3% of the total variation explained, followed by spatial position (21.5%). The inclusion of the *targeting* effect explained the least deviance for bigeye tuna, but accounted for the second most of the deviance explained in the case of yellowfin tuna. In general standardized bigeye tuna CPUE showed a strong seasonal trend, with catch rates highest during the austral winter month June to August (Fig. 8). Standardized CPUE for yellowfin tuna showed less pronounced seasonal trend, which peaked in November, but suggested showed a significant targeting effect as can be inferred from two times higher catch rates for sets within in fishing cluster 2 (Fig. 8).

Nominal and standardized CPUE (together with CVs, 95% C.I.) for big eye and yellowfin tuna caught by joint-venture long-line vessels are presented in Tables 4 and 5, respectively. The standardized CPUE index for bigeye tuna showed a decline between 2005 and 2008, a slight increase between 2008 and 2010 and a fairly stable trend between 2010 and 2015 and a slight increase again in 2016 (Fig. 9). The standardised CPUE index for yellowfin tuna showed a sharp decline between 2004 and 2012, followed by a slight increase until 2016 (Fig. 10). The addition of covariates only slightly influenced the relative abundance trends, when compared to nominal CPUE (Figs. 9 & 10). The most discernible effect was inclusion of the random vessel effect for bigeye tuna, which may be indicative of smaller time-varying vessel (or skipper) efficiency of catching bigeye tuna.

We anticipate that the here presented standardized abundance indices for bigeye tuna and yellowfin tuna based on South African joint-venture vessels could be useful for corroborating other abundance indices for the South-West Indian Ocean.

TABLES

Table 1: Percentage contribution of large pelagic species to the average catch composition of each cluster used as a proxy for fishing tactic. SBT: Southern Bluefin; BET: bigeye tuna; YFT: yellowfin tuna; ALB: albacore; SKJ: skipjack tuna; SWO: swordfish; SMA: shortfin mako and BSH: blue shark

Cluster	SBT	BET	YFT	ALB	SKJ	SWO	SMA	BSH
1	3.4	**39.5**	**24.4**	20.1	0.2	5.6	1	5.7
2	0.1	**35**	**55.1**	1.6	0.1	0.8	2.8	4.4

Table 2: Model statistics for all fixed effects included in the final GAMM for bigeye tuna, summarising the degrees of freedom (df), AIC values, BIC values, residual deviance (Res. dev.), changes in the residual deviance (Dev.), the percentage of the total reduction in deviance explained by each factor (% explained), and corresponding p-values when a *F*-test was applied to test for significance.

Model	df	AIC	BIC	Res. Dev.	Δ Dev.	% explained	P(χ^2)
Null	2	68842.4	68855.7	54108.5	0	0	
+ Year	13	68322.7	68409.2	50272.2	-3836.364	49.27	< 0.001
+ Month	19	68010.3	68135.5	48068.3	-2203.882	28.3	< 0.001
+ s(Lat,Long)	27	67774.1	67956.2	46392.9	-1675.327	21.51	< 0.001
+ FT	28	67765.0	67953.1	46321.6	-71.32146	0.92	< 0.01
Total variation explained (%):				**14.4%**			

Table 3: Model statistics for all fixed effects included in the final GAMM for yellowfin tuna, summarising the degrees of freedom (df), AIC values, BIC values, residual deviance (Res. dev.), changes in the residual deviance (Dev.), the percentage of the total reduction in deviance explained by each factor (% explained), and corresponding p-values when a *F*-test was applied to test for significance.

Model	df	AIC	BIC	Res. Dev.	Δ Dev.	% explained	P(χ^2)
Null	2	70597.2	70610.5	68534.7	0	0	
+ Year	13	69452.1	69538.5	58863.5	-9671.283	47.53	< 0.001
+ Month	19	69315.6	69441.3	57702.7	-1160.732	5.7	< 0.01
+ s(Lat,Long)	28	68862.3	69047.4	54119.0	-3583.738	17.61	< 0.001
+ FT	27	68031.8	68211.9	48188.0	-5931.015	29.15	< 0.001
Total variation explained (%):				**29.69%**			

Table 4: Nominal and standardized CPUE for bigeye tuna caught by Japanese-flagged vessels fishing under South African joint-venture agreement off the south-west coast of South African (≥ 25°E longitude), including CVs and 95% C.I.s for the standardized CPUE produced by the final GAMM.

Year	Nominal	Standardized	CV	lower CI	upper CI
2004	136.9	159.6	0.106	129.7	196.4
2005	267.1	274.9	0.109	222.2	340.1
2005	-	-	-	-	-
2007	165.7	169.5	0.097	140.2	204.8
2008	76.7	78.7	0.098	64.9	95.4
2009	111.2	97.3	0.102	79.6	118.9
2010	175.3	138.1	0.094	114.9	166.1
2011	154.1	128.1	0.093	106.7	153.8
2012	174.4	127.4	0.094	106.1	153.1
2013	150.6	126.9	0.099	104.5	154.1
2014	181.6	134.5	0.111	108.2	167.3
2015	161.0	113.5	0.111	91.3	141.2
2016	216.9	149.8	0.114	119.8	187.5

Table 5: Nominal and standardized CPUE for yellowfin tuna caught by Japanese-flagged vessels fishing under South African joint-venture agreement off the south-west coast of South African (≥ 25°E longitude), including CVs and 95% C.I.s for the standardized CPUE produced by the final GAMM.

Year	Nominal	Standardized	CV	lower CI	upper CI
2004	352.4	385.8	0.105	314.2	473.8
2005	379.3	425.1	0.110	342.5	527.7
2006	-	-	-	-	-
2007	212.8	257.1	0.103	209.9	314.8
2008	160.4	209.3	0.096	173.5	252.4
2009	194.8	237.2	0.104	193.6	290.6
2010	229.6	240.2	0.101	197.0	293.0
2011	114.6	154.4	0.103	126.1	189.1
2012	80.8	101.2	0.106	82.1	124.6
2013	184.6	194.7	0.104	158.8	238.8
2014	174.3	178.4	0.118	141.5	224.8
2015	204.5	207.0	0.118	164.3	260.8
2016	175.3	201.3	0.122	158.3	255.8

FIGURES

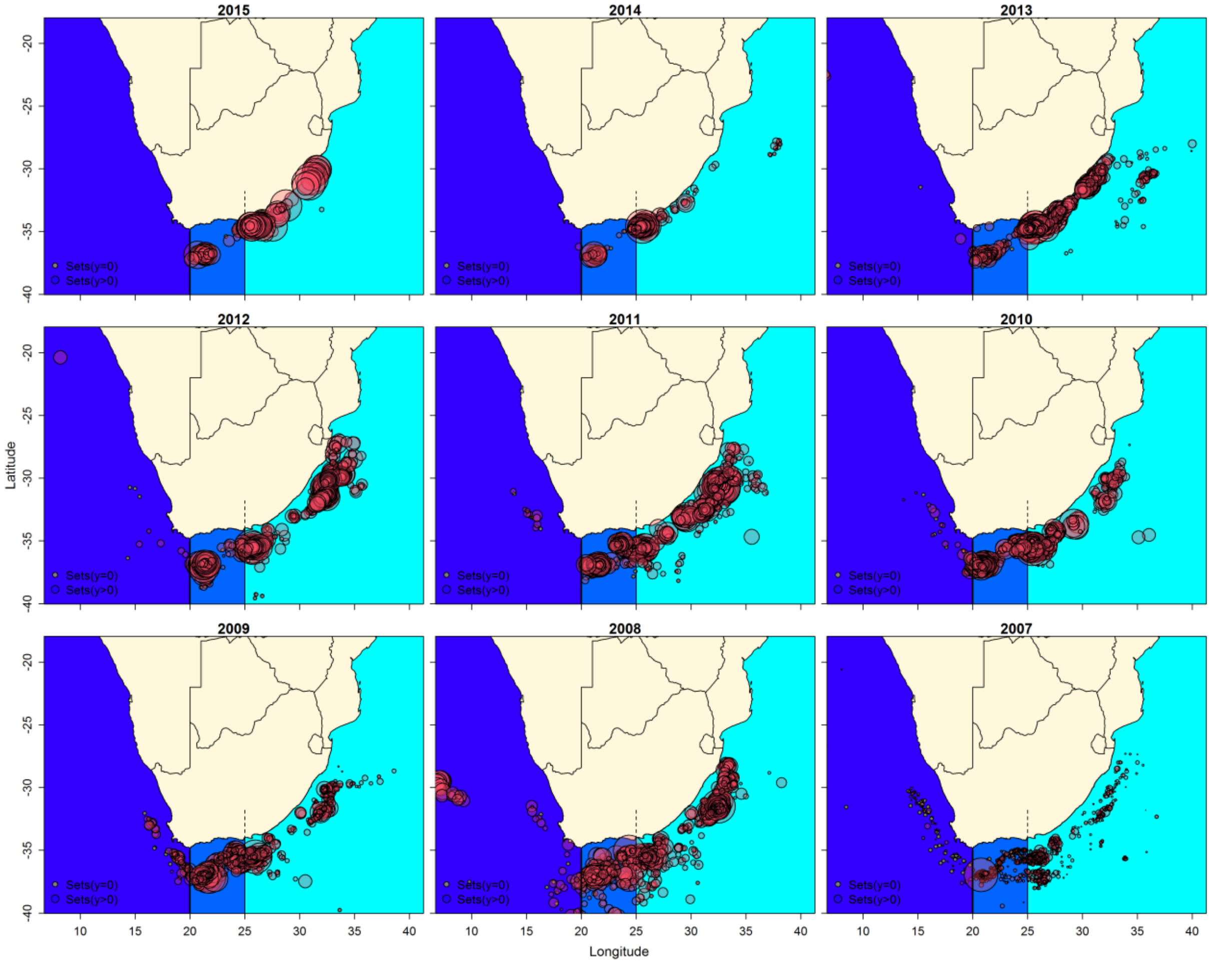

Figure 1: Annual effort for the combined South African longline fleets. Longline sets that did not encounter a bigeye tuna are the smallest circles, and the circle diameter increases proportional to the weight of bigeye tuna caught per set. The black line indicates the ICCAT/IOTC boundary. The blue different shading highlights the Atlantic (ICCAT) region and the considered "buffer" zone between 20°E and 25°E within the IOTC region of South Africa.

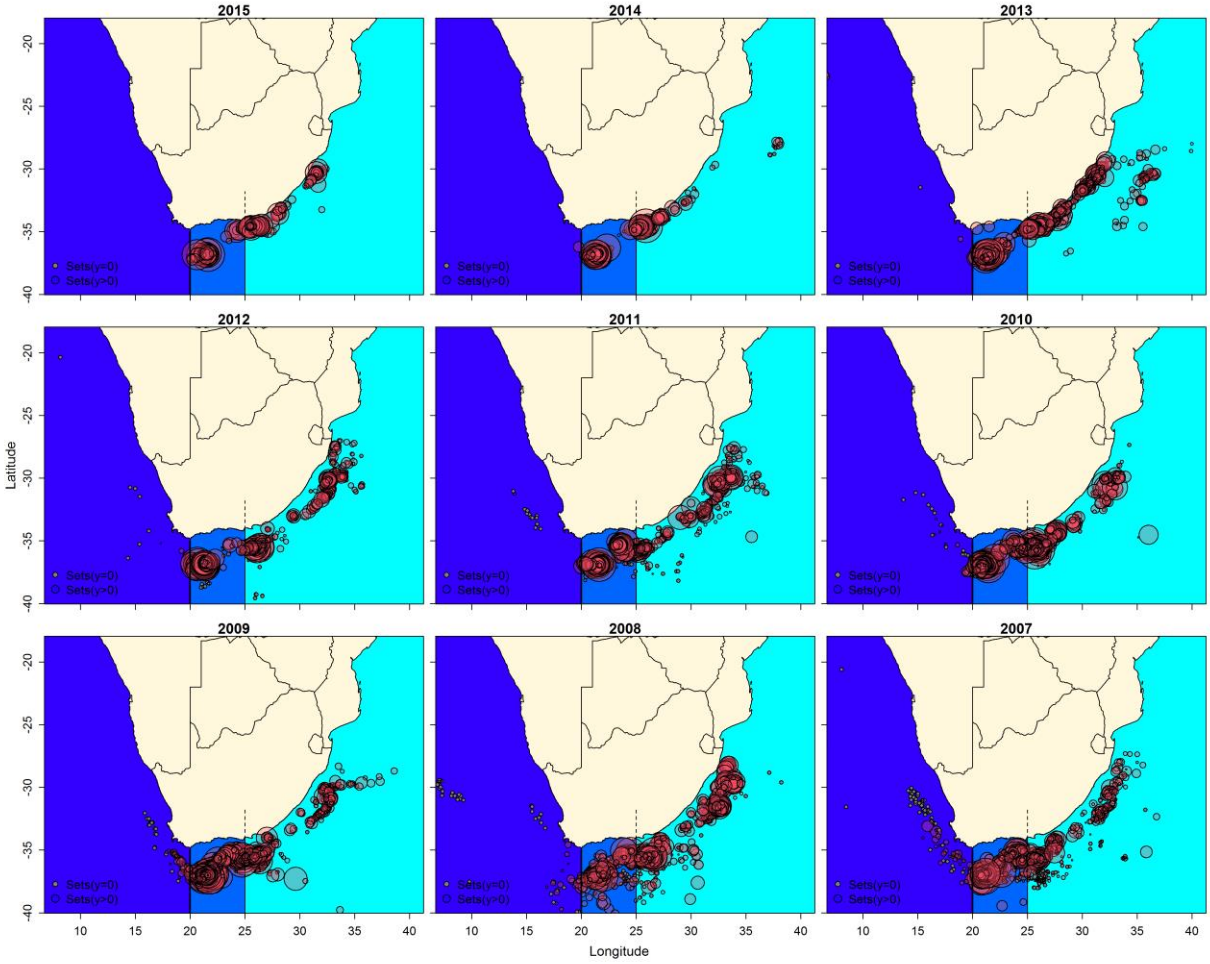

Figure 2: Annual effort for the combined South African longline fleets. Longline sets that did not encounter a yellowfin tuna are the smallest circles, and the circle diameter increases proportional to the weight of yellowfin tuna caught per set. The black line indicates the ICCAT/IOTC boundary. The blue different shading highlights the Atlantic (ICCAT) region and the considered "buffer" zone between 20°E and 25°E within the IOTC region of South Africa.

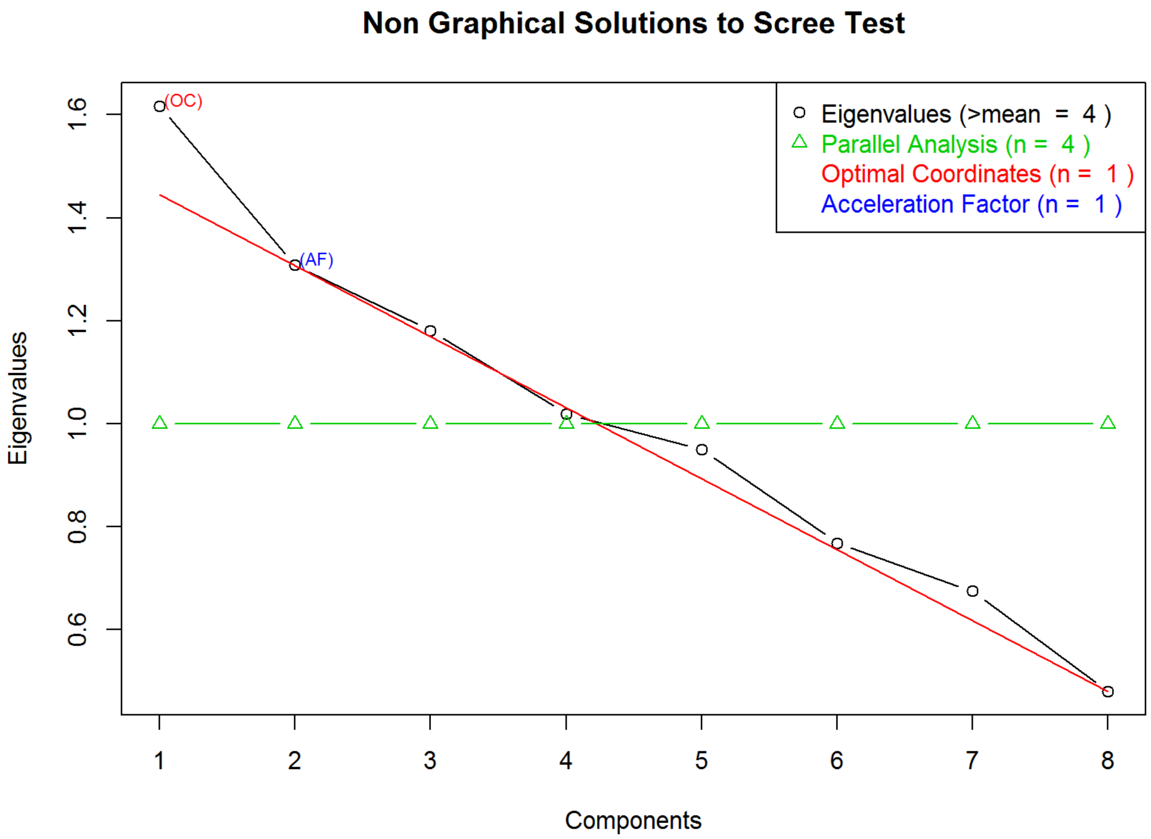

Figure 3: A non-graphical solution to the Scree test to determine the optimal number of clusters in the multivariate analysis to assess the influence of fishing tactic on CPUE estimation.

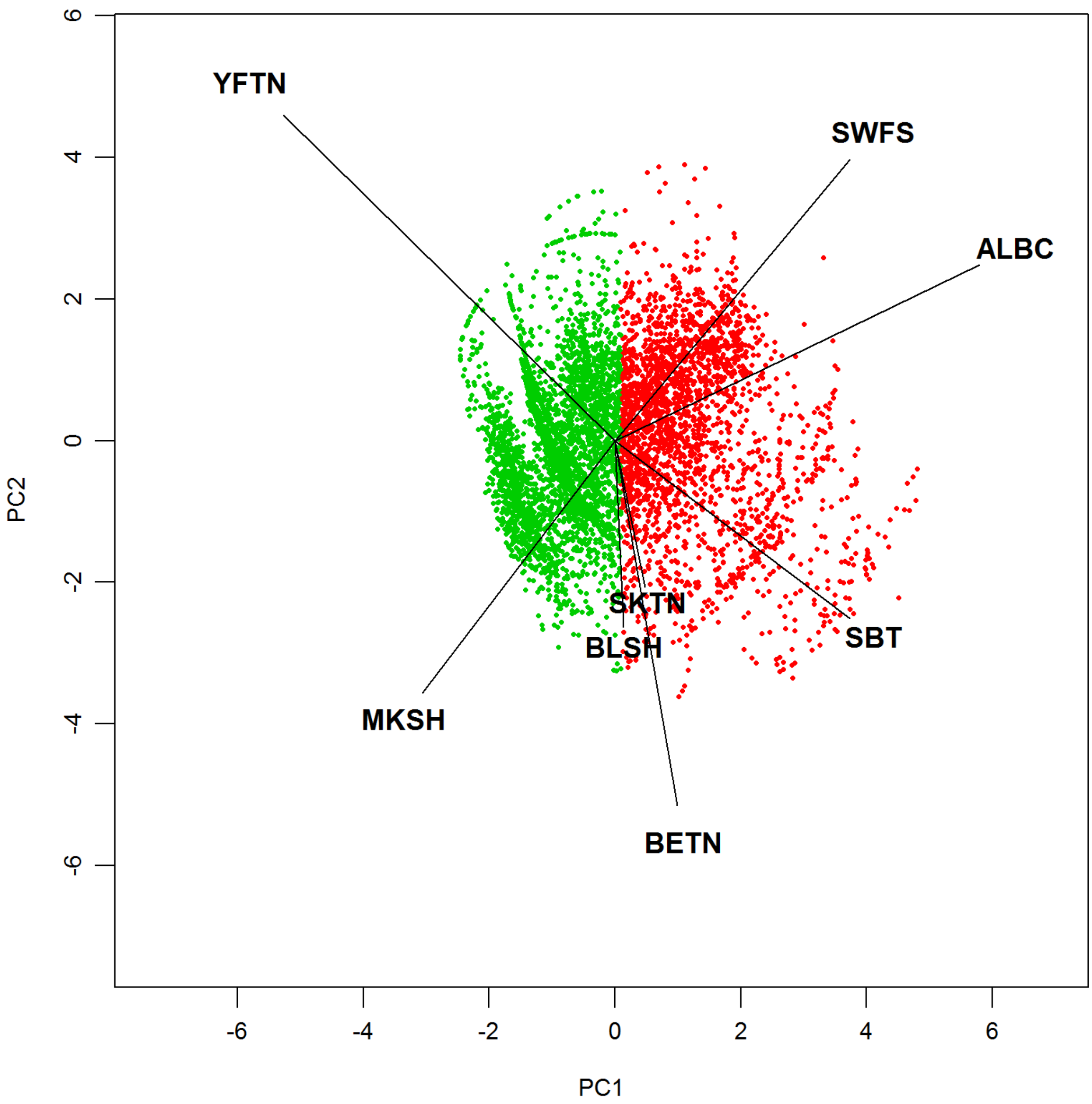

Figure 4: A graphical representation of the two clusters that characterise the different fishing tactics projected over the first two Principal Components (PCs), where only PC1 was determined to be non-trivial. FT 1 (red) and FT 2 (green) compare to Table 1.

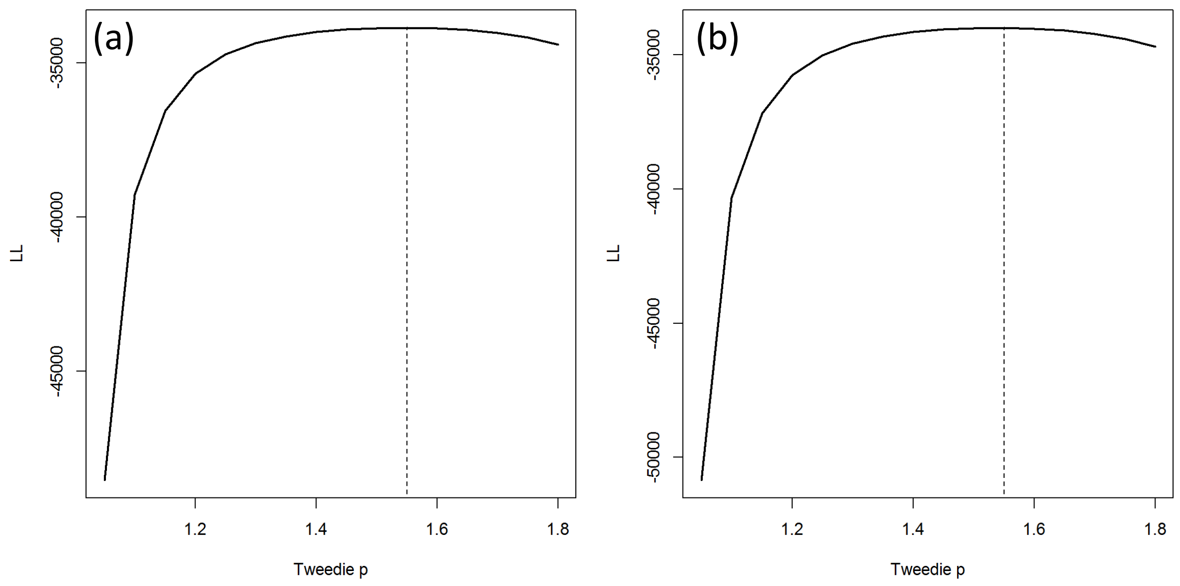

Figure 5: Log-likelihood profile for over the grid of power parameters values ($1 < p < 2$) of the Tweedie distribution for bigeye tuna (a) and yellowfin tuna (b). The vertical dashed line denote the optimized p used in the final standardization GAMM.

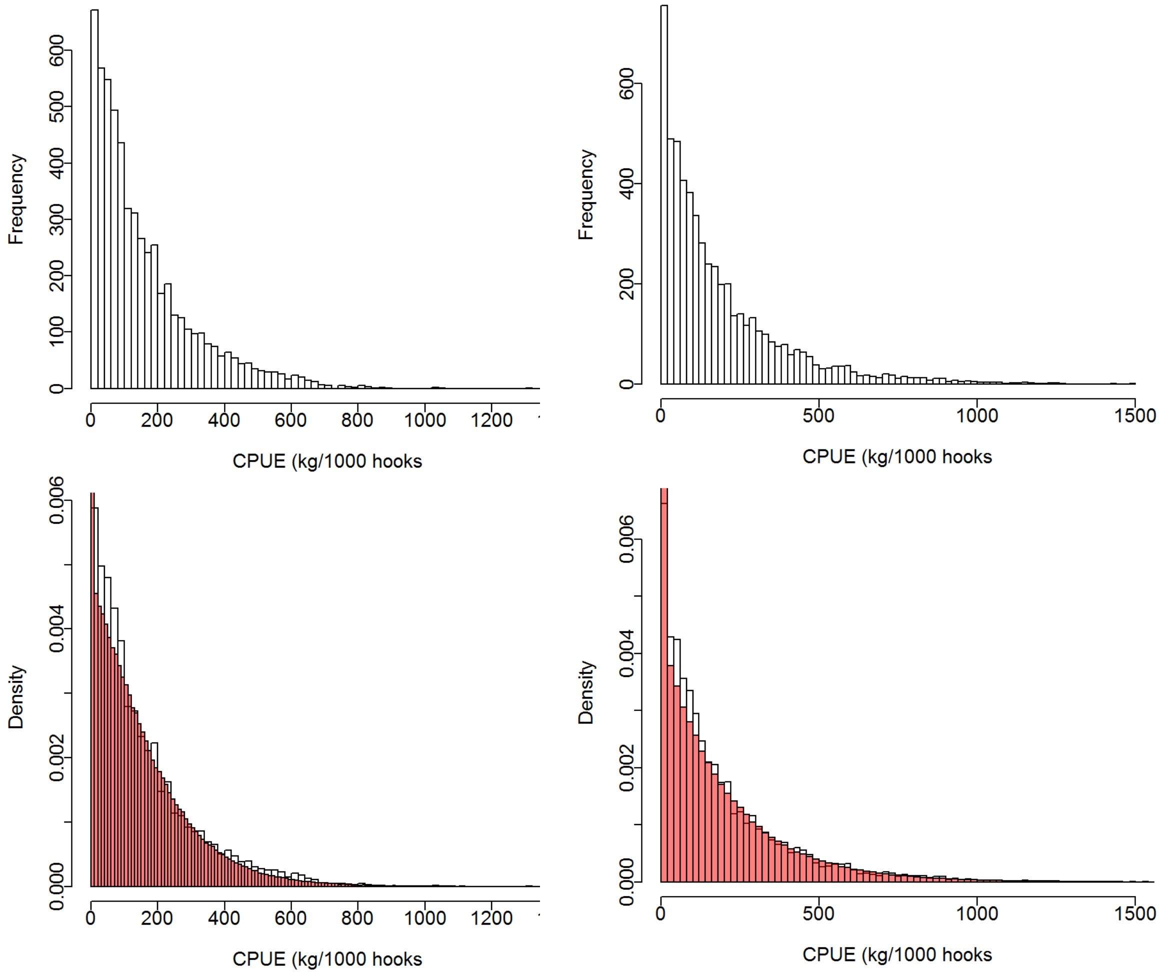

Figure 6: CPUE frequency, and density, distributions for bigeye tuna (left panel) and yellowfin tuna (right panel). The red shaded areas denote the expected distributions generated using Monte-Carlo simulation based on the final GAMM fitted with a Tweedie distribution.

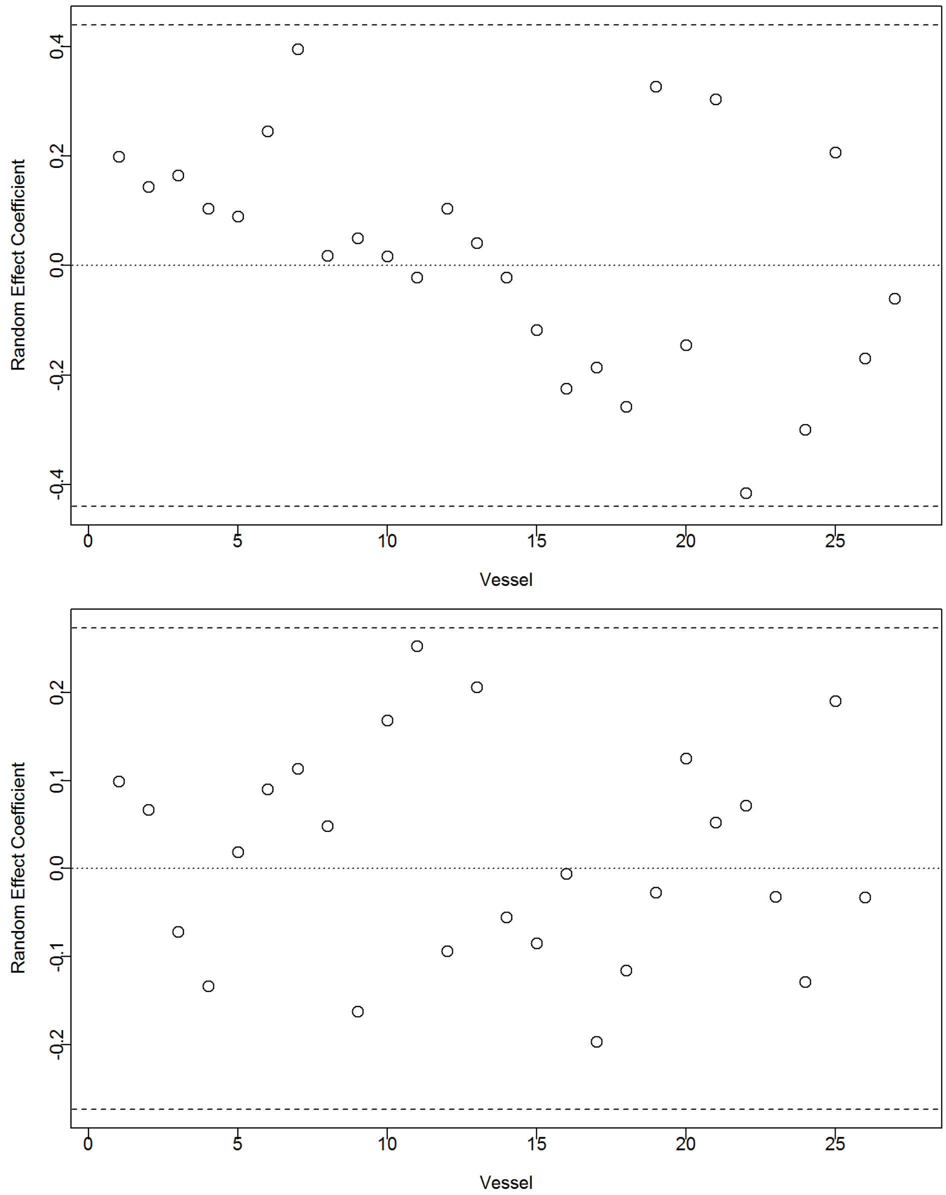

Figure 7: Random effects coefficients (dots) illustrating the deviation from the mean of zero across the 25 vessels retained for the GAMM analysis of bigeye tuna (upper panel) and yellowfin tuna (lower panel). Dashed lines denote the 95% confidence interval.

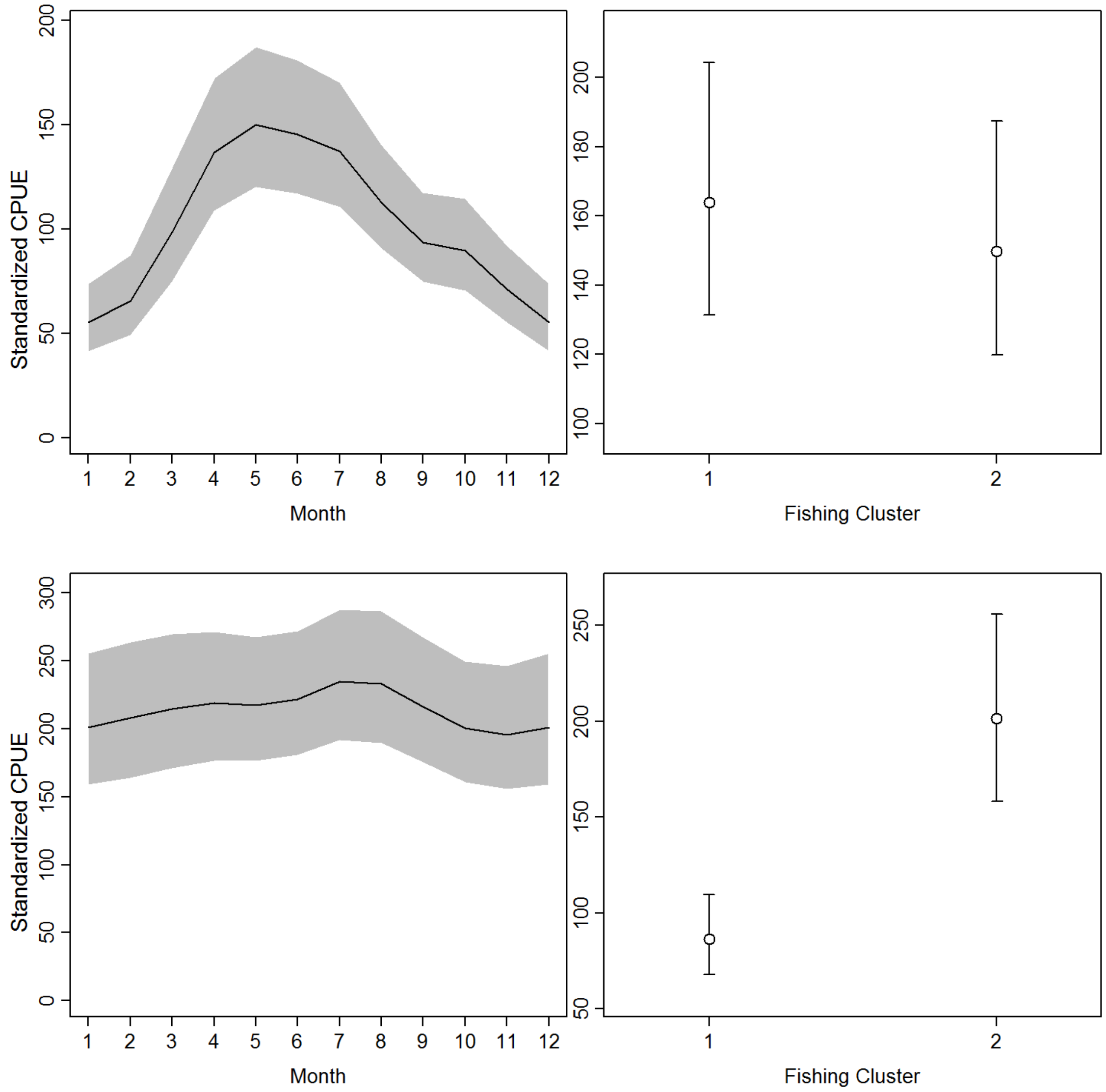

Figure 8: The influence of the fixed effects *Month* and *Fishing Cluster* on the standardized CPUE for bigeye tuna (upper panel) and yellowfin tuna (lower panel) based on the final GAMMs.

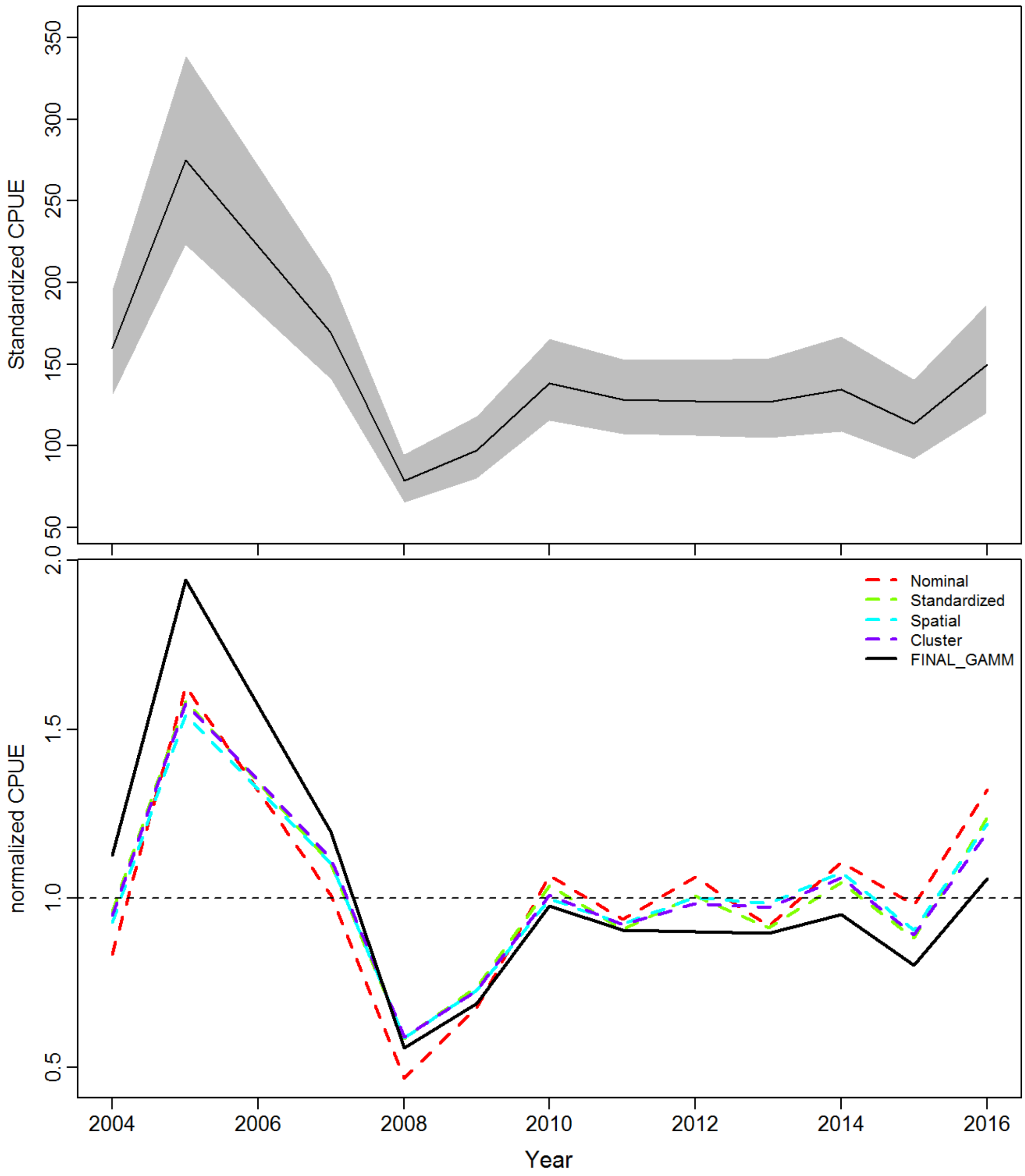

Figure 9: Standardized CPUE for bigeye tuna for the time period 2004 to 2016 (upper panel) based on the final GAMM. The 95% confidence intervals for the nominal CPUE are denoted by grey shaded areas and comparison of nominal and the various standardized CPUE models (lower panel).

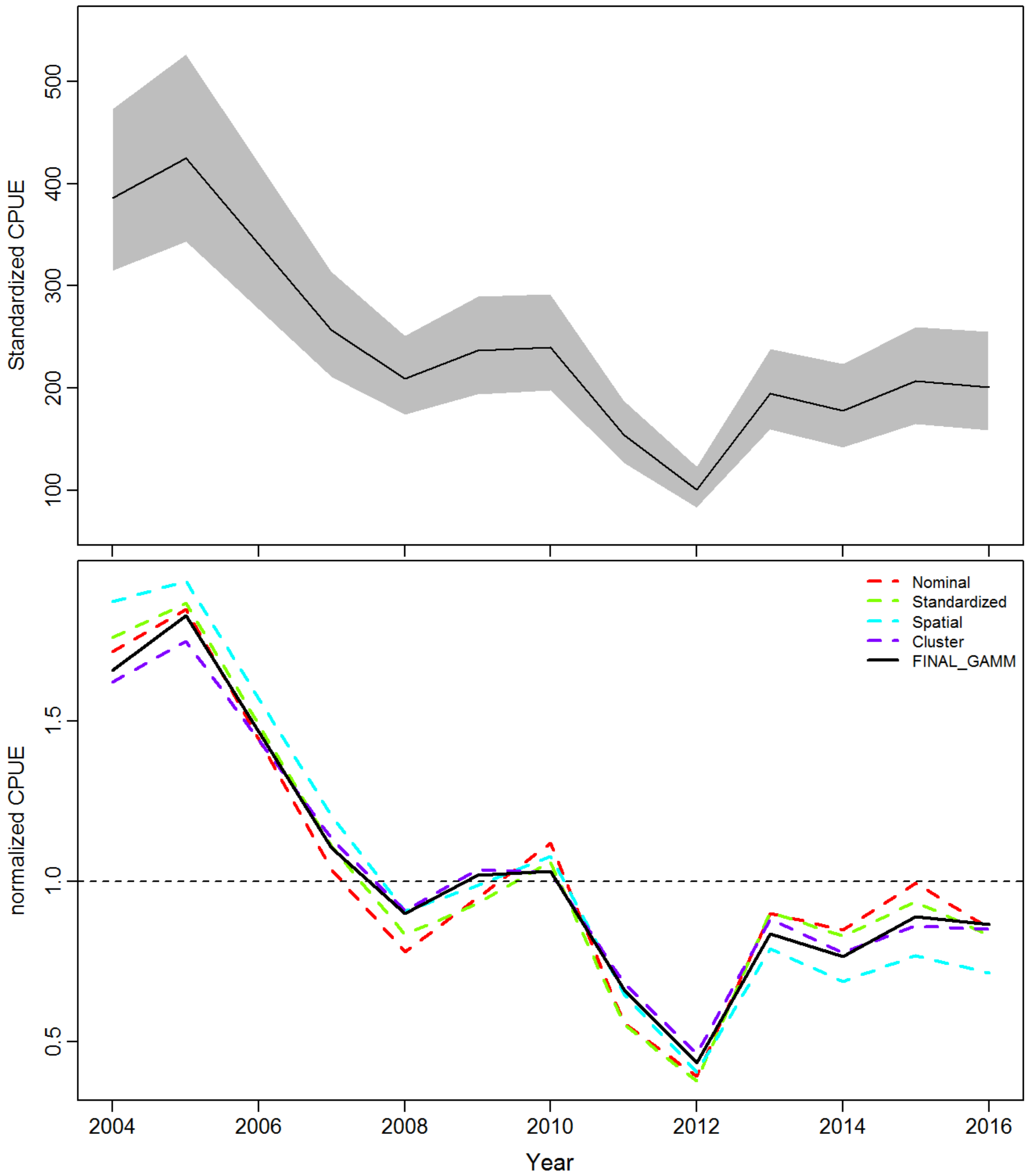

Figure 10: Standardized CPUE for yellowfin tuna for the time period 2004 to 2016 (upper panel) based on the final GAMM. The 95% confidence intervals for the nominal CPUE are denoted by grey shaded areas and comparison of nominal and the various standardized CPUE models (lower panel).

REFERENCES

Carvalho, F. C., Murie, D. J., Hazin, F. H., Hazin, H. G., Leite-Mourato, B., Travassos, P., and Burgess, G. H. *et al.* 2010. Catch rates and size composition of blue sharks (*Prionace glauca*) caught by the Brazilian pelagic longline fleet in the southwestern Atlantic Ocean. Aquatic Living Resources, 23: 373–385.

Dunn, P. K., and Smyth, G. K. 2005. Series evaluation of Tweedie exponential dispersion model densities. Statistics and Computing, 15: 267–280.

He, X., Bigelow, K. A., and Boggs, C. H. 1997. Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery. Fisheries Research, 31: 147–158.

Helser, T., Punt, A. E., and Methot, R. D. 2004. A generalized linear mixed model analysis of a multi-vessel fishery resource survey. Fisheries Research, 70: 251–264.

Nepgen, C. 1970. Exploratory fishing for tuna off the South African west coast. Investigational report (South Africa. Division of Sea Fisheries).

Ono, K., Punt, A. E., and Hilborn, R. 2015. Think outside the grids: An objective approach to define spatial strata for catch and effort analysis. Fisheries Research, 170: 89–101.

Penney, A., and Griffiths, M. 1999. A first description of the developing South African longline fishery. Coll. Vol. Sci. Pap., 49: 162–173.

Raiche, G., Magis, D., 2010. nFactors: Parallel Analysis and Non Graphical Solutions to the Cattell Scree Test. R Package Version 2.3.3.

Struyf, A., Hubert, M., and Rousseeuw, P. J. 1997. Integrating robust clustering techniques in S-PLUS. Computational Statistics & Data Analysis, 26: 17–37.

Tascheri, R., Saavedra-Nievas, J. C., and Roa-Ureta, R. 2010. Statistical models to standardize catch rates in the multi-species trawl fishery for Patagonian grenadier (*Macruronus magellanicus*) off Southern Chile. Fisheries Research, 105: 200–214.

Thorson, J. T., and Minto, C. 2015. Mixed effects: a unifying framework for statistical modelling in fisheries biology. ICES Journal of Marine Science, 72: 602–614.

Welsh, J. 1968. A new approach to research on tuna in South African waters. Fisheries Bulletin of South Africa, 5: 32–34.

Winker, H., Kerwath, S. E., and Attwood, C. G. 2013. Comparison of two approaches to standardize catch-per-unit-effort for targeting behaviour in a multispecies hand-line fishery. Fisheries Research, 139: 118–131.

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