STANDARDIZATION OF THE CATCH PER UNIT EFFORT FOR SWORDFISH (*XIPHIAS GLADIUS*) FOR THE SOUTH AFRICAN LONGLINE FISHERY

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

2017-08-25

SUMMARY

Swordfish, *Xiphias gladius* is a target species of the South African pelagic longline fleet operating along the west and east coast of South Africa. A standardization of the CPUE of the South African swordfish-directed longline fleet operating in the IOTC region for the time series 2004-2016 was carried out using a Generalized Additive Mixed Model (GAMM) with a Tweedie distributed error. Explanatory variables of the final model included *year*, *month*, geographic position (*lat, long*) and a targeting factor with two levels, derived by clustering of PCA scores of the root-root transformed, normalized catch composition. Vessel was included as a random effect. Swordfish CPUE had a definitive seasonal trend, with catch rates higher in winter and lower in summer. The standardised CPUE analysis indicates a declining trend over the period 2004-2016.

KEYWORDS

Swordfish, 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 vessels are the only vessels that target swordfish, with negligible bycatch being caught in other fisheries. Pelagic longline fishing by South African vessels began in the 1960s with the main target being southern bluefin tuna (*Thunnus maccoyii*) and albacore (*Thunnus alalunga*) (Welsh, 1968; Nepgen, 1970). This South African Large Pelagic 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. Interest in pelagic longline fishing re-emerged in 1995 when a joint venture with a Japanese vessel confirmed that tuna and swordfish could be profitably exploited within South Africa's waters. Thirty experimental longline permits were subsequently issued in 1997 to target tuna, though substantial catches of swordfish were made during that period (Penney and Griffiths, 1999).

The commercial fishery was formalised in 2005 with the issuing of 10-year long term rights to swordfish- and tuna-directed vessels. On average, 15 South African vessels are active in a year and target swordfish in 20-30m length vessels. Additionally, foreign flagged vessels catch swordfish as bycatch. South Africa's swordfish catches reached a peak in 2002 at 1 187 t, and have been on the decline with average catches of 372 t for the period 2009-2014. The fishery is coastal and swordfish-oriented effort concentrates in the southwest Indian Ocean region (20°- 30°S, 30°- 40°E) and along the South African continental shelf in the southeast Atlantic (30°- 35°S, 15°- 18°E). As such, the fishery straddles two ocean basins, the Indian and Atlantic Ocean. The jurisdictions of the Indian Ocean Tuna Commission (IOTC) and International Commission for the Conservation of Atlantic Tuna (ICCAT) are separated by a management boundary at 20°E. Consequently, all tunas and billfish stocks with the exception of the southern bluefin tuna (*Thunnus maccoyii*) are artificially divided into Atlantic and Indian Ocean stocks along this boundary, regardless of their true stock structure and distribution. Since questions remain about the origin of South African caught swordfish, it remains uncertain if the artificial split in reporting stock indices indeed reflects a biological meaningful separation of stocks.

Here we present standardised catch-per-unit-effort (CPUE) indices that were obtained with a generalised additive mixed model (GAMM) of swordfish catch and effort data from the South African pelagic longline fleet operating in the South Indian Ocean between 2004 and 2016. Catch and effort data were subset to the IOTC area of the South Indian Ocean with the IOTC/ ICCAT transition area removed (20 - 28°E) as suggested by West (2016). The GAMM was 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 swordfish directed longline trips were extracted from the database for the period 2004-2016 (Sets = 4343; hooks = 6 026 929). Each record included the following information: (1) date, (2) unique vessel number, (3) catch position at a 1 x 1 degree latitude and longitude resolution and (4) mandatory catch reports in kilogram per set and (5) hooks per set. Catch and Effort data were subset to only include catches from the IOTC area (Longitude > 29 degrees) with the IOTC/ ICCAT transition area removed.

Model framework

Swordfish CPUE was standardized using Generalized Additive Mixed Models (GAMMs), which included the covariates year, month, 1 x 1 degree latitude (*lat*) and longitude (*long*) coordinates 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' (Raîche *et al.*, 2013). 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.

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 high frequencies of 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 p = 1.3 with an associated dispersion parameter of ϕ = 5 for the full GAMM. The full GAMM evaluated for swordfish 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* = 28) 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 < 0.001) 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 July (*month* = 7), representative of the high catch quarter and *FT* was fixed to the fishing tactic the produced highest average catch rates (*FT* = 2). The expected yearly mean $CPUE_y$ and standard-error of the expected $log(CPUE_y)$ 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

The analysis of deviance for the step-wise regression procedure showed that all of the covariates considered were highly significant ($p < 0.001$) and the inclusion of all considered fixed effects were supported by both the AIC and BIC (Table 1). The inclusion of *month* effect, contributed to the greatest improvement in the deviance explained in the model (47 %), followed by *year* (22%) then *fishing tactic* (19 %) (Table 1). *Fishing tactic* 2 contributed 58% to catches within this cluster, while *fishing tactic* 1 contributed 23% to catches. For *fishing tactic* 1 big eye tuna (*Thunnus obesus*) contributed 41% to catches (Fig. 8.b). The inclusion of targeting, and the justifiable use of the Tweedie distribution (Figs. 5 & 7) have improved the model and our confidence in the swordfish CPUE time-series. However, further analyses could be considered. The amendment of the catch return forms to include the target per catch day, sea surface temperature, use of livebait and hours fished should further improve the standardization of the CPUE data in this fishery in the future. Previous attempts to classify 'catchability' of vessels within the fleet include using *vessel type* as a categorical variable or using a subset of vessels from each class as indicator vessels. This information was challenging to obtain, and neither of these attempts significantly improved the model's explanatory power. As such, including *vessel* as a random effect was deemed the most appropriate solution. Given the notable variation among vessels (Fig.6), it is unsurprising that the inclusion of the random vessel effect produced the most parsimonious error model.

Nominal and standardized CPUE (together with CVs, 95% C.I.) for southern Indian Ocean swordfish caught by domestic South African long-line vessels (> 29 longitude) are presented in Table 2. Swordfish CPUE had a definitive seasonal trend, with higher catch rates in late winter and lower catch rates in summer. The nominal and standardised CPUE time-series were similar, but diverged notably in the last year.

Standardised CPUE rates of approximately 500 kg/1000 hooks were observed in the first year (2004). Thereafter, a consistent decline in CPUE was observed, dropping below 400 kg/1000 hooks in 2007. Standardised CPUE then increased in 2008 to approximately 550 kg/1000 hooks, though a general decline followed. The most recent estimate (2016) was approximately 350 kg/1000 hooks. Overall, the analyses presented here indicate a declining trend in CPUE for the South African swordfish fishery in the IOTC region.

TABLES

Table 1. Results from the GAMM applied to swordfish (*Xiphias gladius*) indicating the deviance explained by parameters selected for the final model.

	DF	AIC	BIC	Deviance	Deviance Explained	% Deviance Explained	P-Value
Null Model	2	57003	57015	116895	0.000	0.00	
Year	14	56686	56775	110086	-6808	21.55	< 0.001
Month	20	55906	56032	95267	-14819	46.91	< 0.001
Latitude/Longitude	28	55701	55880	91365	-3902	12.35	< 0.001
Fishing Tactic	29	55345	55530	85302	-6063	19.19	< 0.001

Table 2. Nominal and standardised CPUE values, including the coefficient of variation (CV) and confidence intervals (LCI, UCI) for swordfish (*Xiphias gladius*) for the period 2004 - 2016.

Year	Nominal	CPUE	CV	LCI	UCI
2004	329	502	0.09	423	596
2005	378	493	0.09	415	585
2006	350	481	0.08	408	568
2007	298	396	0.09	335	468
2008	346	562	0.08	478	660
2009	228	376	0.08	319	444
2010	319	461	0.08	393	540
2011	331	509	0.08	436	596
2012	282	426	0.08	363	501
2013	261	410	0.08	349	482
2014	149	277	0.09	231	332
2015	284	404	0.09	341	479
2016	319	365	0.11	292	456

FIGURES

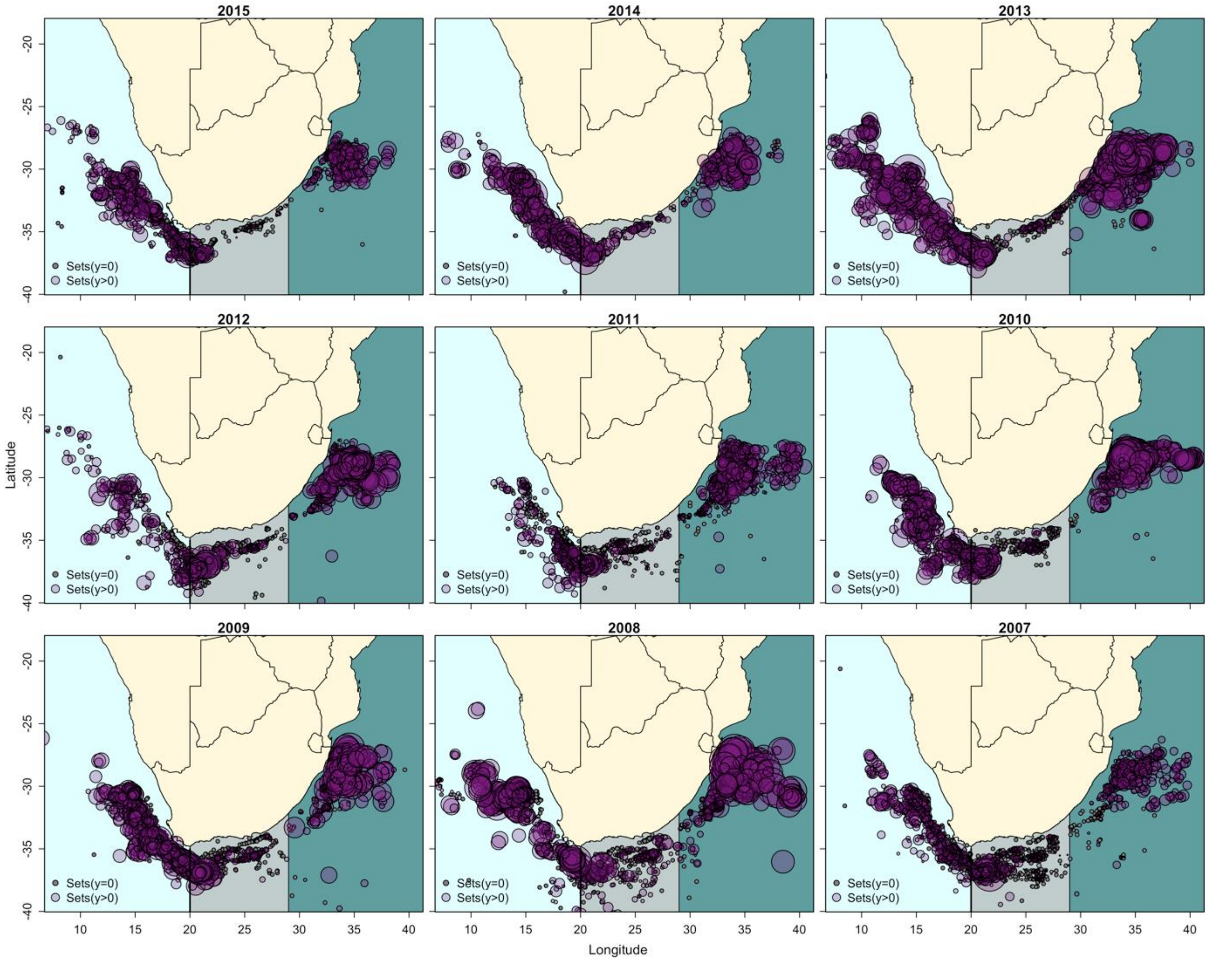

Figure 1. Annual effort for the combined South African longline fleets. Longline sets that did not encounter a swordfish are the smallest circles, and the circle diameter increases proportional to the weight of swordfish caught per set. The black line indicates the ICCAT/IOTC boundary, grey area represents the transition area.

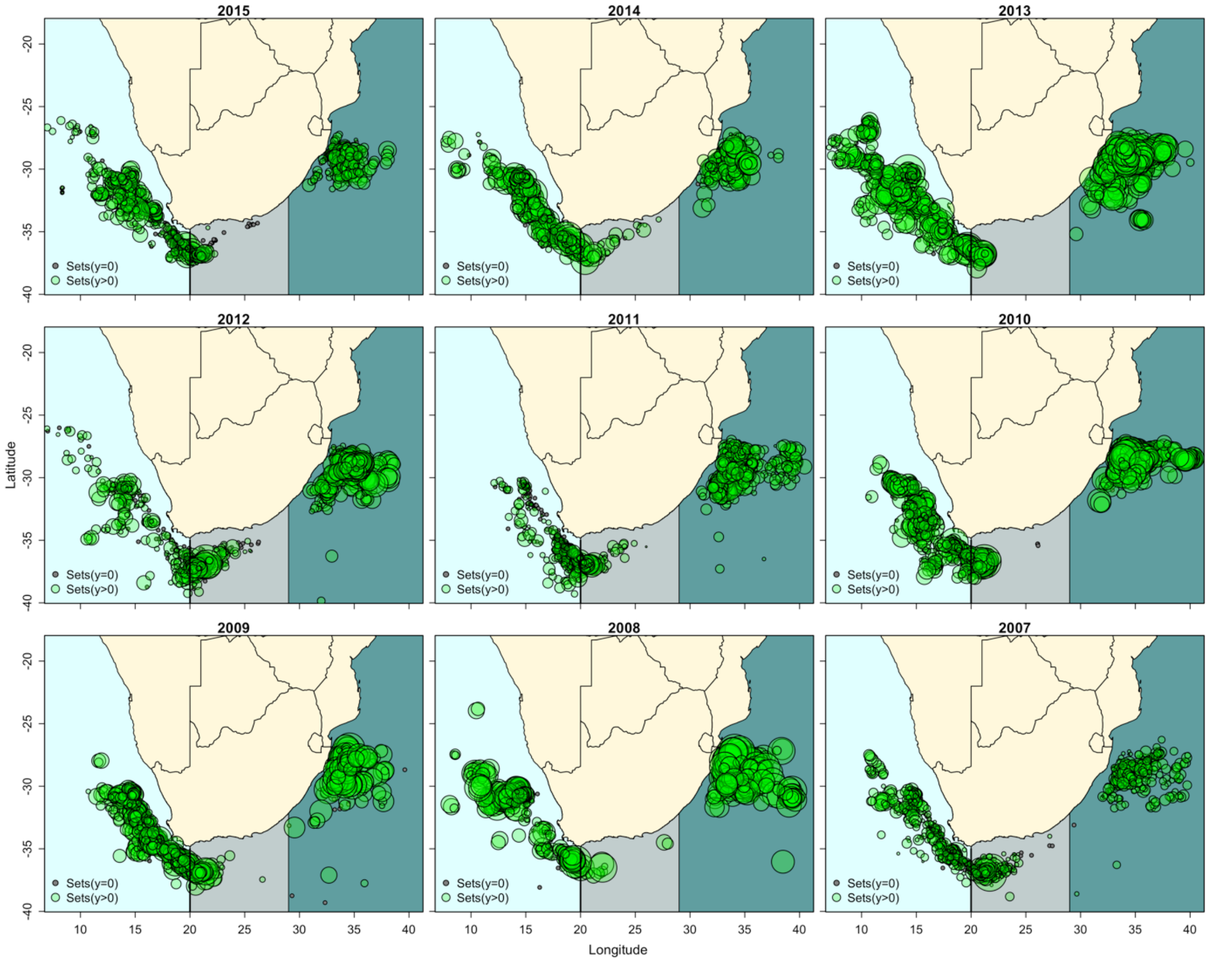

Figure 2. Annual effort for the South African swordfish directed longline fleet. Longline sets that did not encounter a swordfish are the smallest circles, and the circle diameter increases proportional to the weight of swordfish caught per set. The black line indicates the ICCAT/IOTC boundary, grey area represents the transition area.

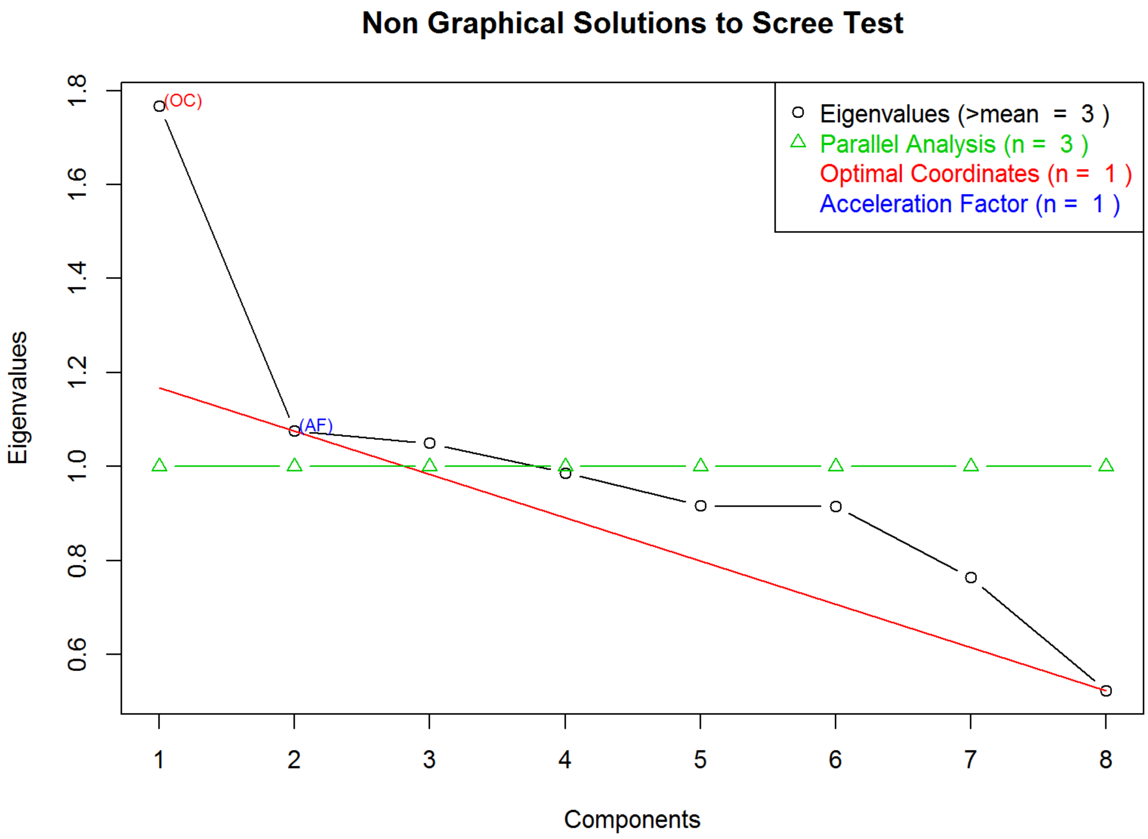

Figure 3. A non-graphical solution to the Scree test to determine the optimal number of clusters in the multivariate analysis to assess the influence of fishing tactic on CPUE estimation.

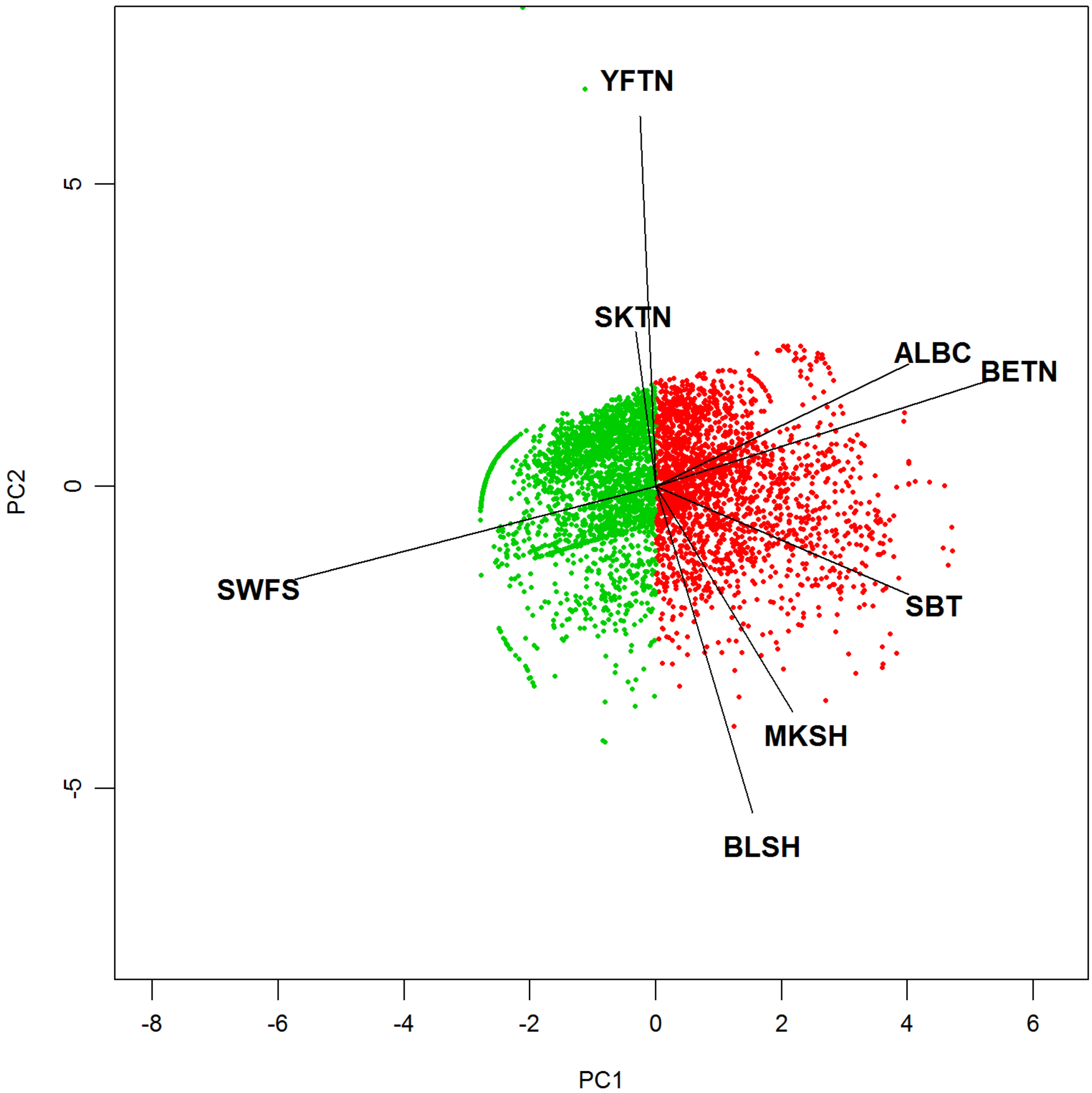

Figure 4. A graphical representation of the two clusters that characterise the different fishing tactics projected over the first two Principal Components (PCs). Cluster one is dominated by bigeye tuna (41%). FT 2: Cluster two is dominated by swordfish (58%).

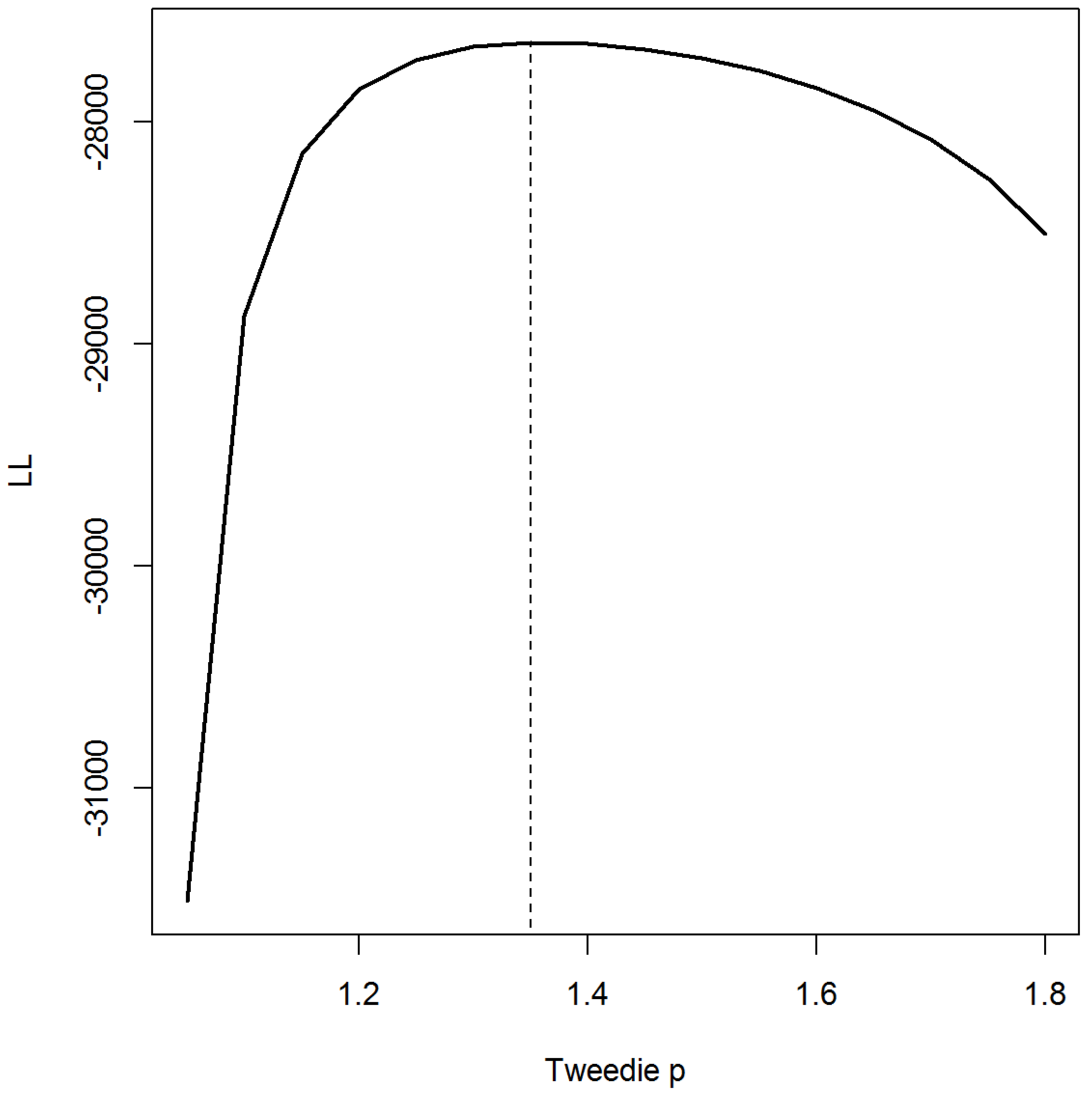

Figure 5. Log-likelihood profile for over the grid of power parameters values ($1 < p < 2$) of the Tweedie distribution. The vertical dashed line denote the optimized p used in the final standardization GAMM.

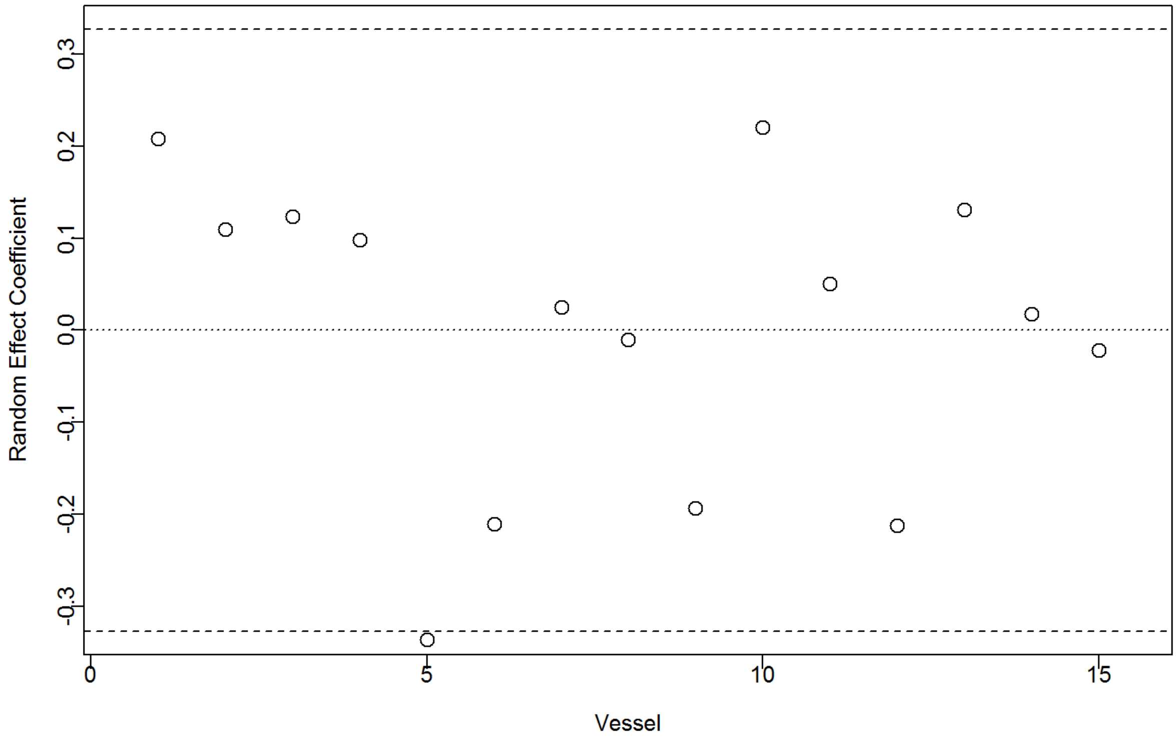

Figure 6. Random effects coeffients (dots) illustrating the deviation from the mean of zero across the 14 vessels retained for the analysis. Dashed lines denote the 95% confidence interval of the mean.

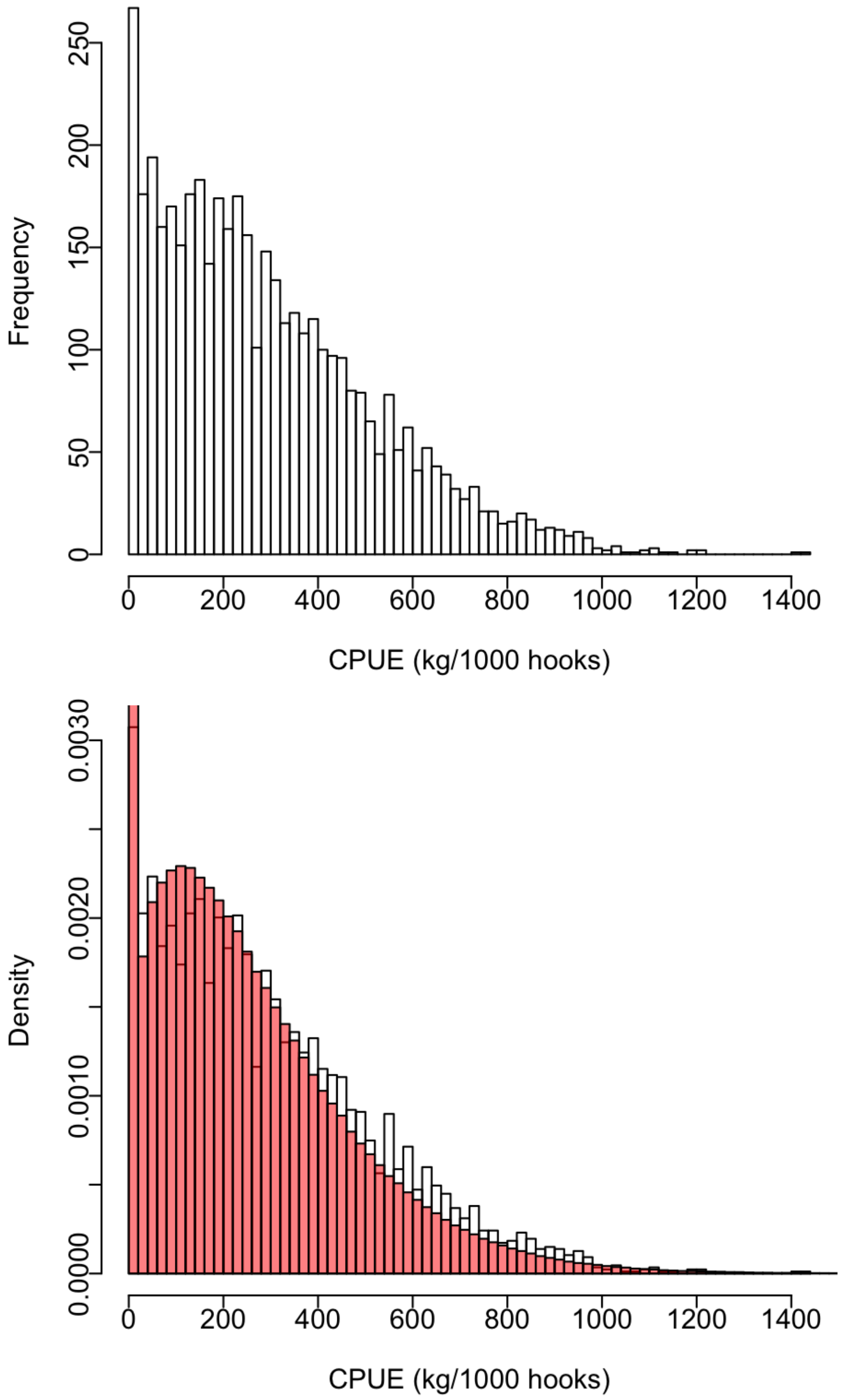

Figure 7. CPUE (kg/1000 hooks) frequency, and density, distributions for the South African swordfish directed longline fishery. The red shaded density denotes the expected density of the response for the Tweedie GAMM, and supports the use of the Tweedie distribution form in the GAMMs.

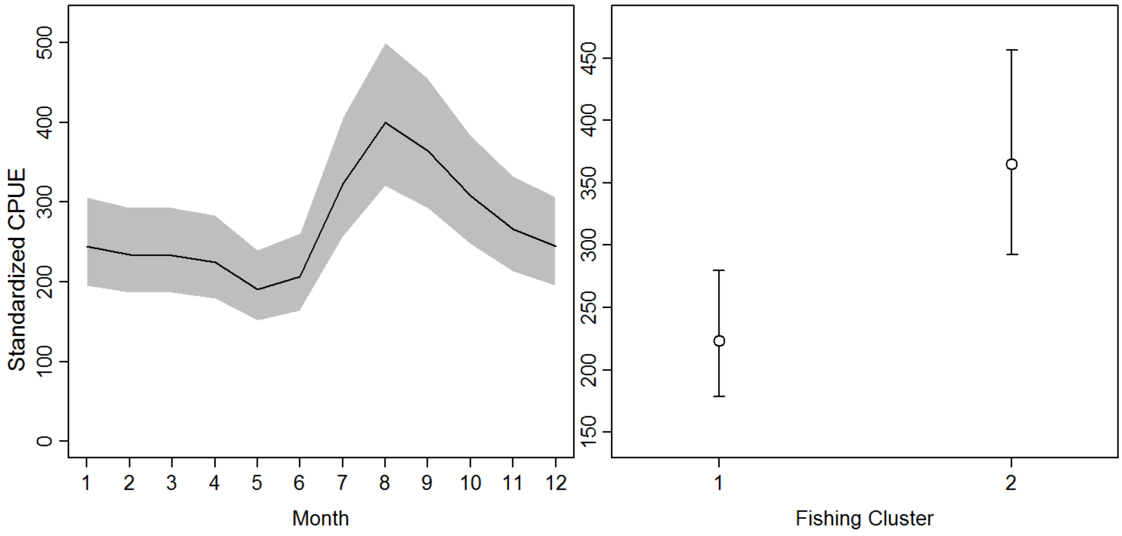

Figure 8. The influence of the fixed effects *Month* and *Fishing Cluster* on the CPUE of swordfish when modelled using the GAMM applied to the South African swordfish directed longline data.

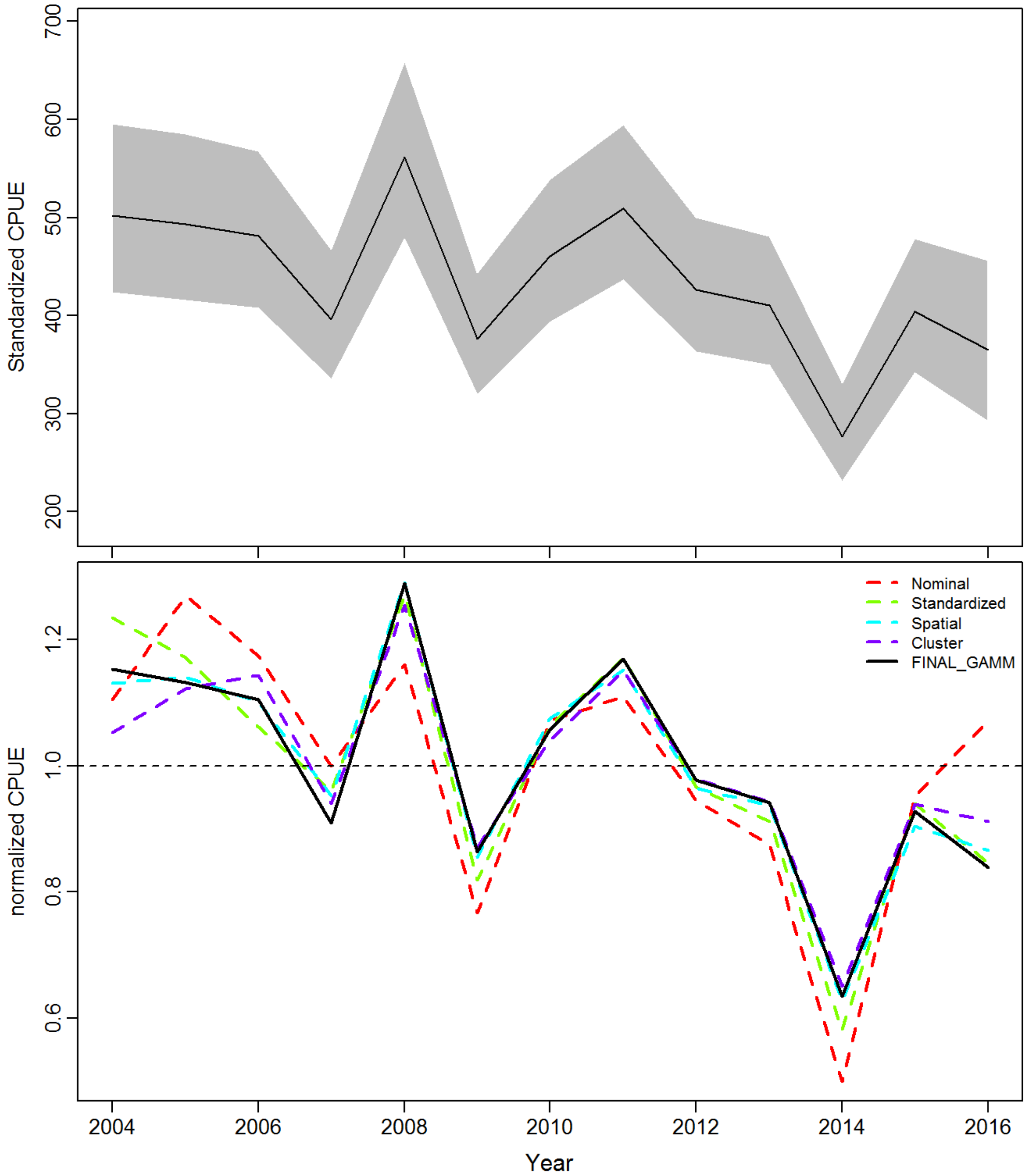

Figure 9. Standardized CPUE for the swordfish directed longline fishery of South Africa for the time period 2004 to 2016 (upper panel). The 95% confidence intervals for the nominal CPUE are denoted by grey shaded areas and comparison of nominal and the various standardized CPUE models (lower panel).

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 (<i>Prionace glauca</i>) caught by the Brazilian pelagic longline fleet in the southwestern Atlantic Ocean. Aquatic Living Resources, 23: 373–385. http://www.alr-journal.org/10.1051/alr/2011005.

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. Elsevier B.V. http://linkinghub.elsevier.com/retrieve/pii/S0165783615001708.

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

Raîche, G., Walls, T. A., Magis, D., Riopel, M., and Blais, J. G. 2013. Non-graphical solutions for Cattell's scree test. Methodology, 9: 23–29.

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.

West, W. 2016. Genetic stock structure and estimation of abundance of swordfish (*Xiphias gladius*) in South Africa. MSc Thesis, University of Cape Town: 1-147.

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. Elsevier B.V. http://linkinghub.elsevier.com/retrieve/pii/S0165783612003311.

Winker, H., Kerwath, S. E., and Attwood, C. G. 2014. Proof of concept for a novel procedure to standardize multi-species catch and effort data. Fisheries Research, 155: 149–159. Elsevier B.V. http://dx.doi.org/10.1016/j.fishres.2014.02.016
http://www.sciencedirect.com/science/article/pii/S0165783614000563.