STANDARDIZATION OF THE CATCH PER UNIT EFFORT FOR ALBACORE (*THUNNUS ALALUNGA*) FOR THE SOUTH AFRICAN TUNA-POLE-LINE (BAITBOAT) FLEET FOR THE TIME SERIES 2003-2015

Henning Winker¹Sven E. Kerwath¹ Wendy M. West¹

SUMMARY

Albacore, Thunnus alalunga, is the main target of the South African tuna pole-line (baitboat) fleet operating along the west and south west coast of South Africa and the South African catch is the second largest in the region with annual landings of around 4000 t. A standardization of the CPUE of the South African baitboat fleet for the time series 2003-2015 was carried out with a Generalized Additive Mixed-Model (GAMM) with a Tweedie distributed error. Explanatory variables of the final model included year, month, geographic position, vessel power, included as a random effect, and targeting, included in form of clustered PCA loadings of the root-root transformed, normalized catch composition. The standardized CPUE mostly trails the nominal CPUE with no overall significant upward or downward trends. The analyses indicate that the CPUE for the South African baitboat fishery for albacore has been stable over the last decade.

RÉSUMÉ

Le germon (Thunnus alalunga) est la principale cible de la flottille de canneurs sud-africains qui opère à l'Ouest et au Sud-Ouest du littoral de l'Afrique du Sud, les captures sud-africaines se trouvant au deuxième rang dans la région, avec des débarquements annuels se chiffrant à environ 4.000 t. Une standardisation de la CPUE de la flottille de canneurs sud-africains a été réalisée pour la série temporelle 2003-2015 à l'aide d'un modèle mixte additif généralisé (GAMM) avec une erreur de distribution Tweedie. Les variables explicatives du modèle final incluaient l'année, le mois, la position géographique, la puissance du navire, incluse comme effet aléatoire, et le ciblage, intégrées sous la forme de chargements de PCA regroupés de la composition de la capture standardisée et transformée en racine quatrième. La CPUE standardisée suit principalement la CPUE nominale sans aucune tendance générale significative à la hausse ou à la baisse. Les analyses indiquent que la CPUE pour la pêcherie de canneurs sud-africains ciblant le germon s'est maintenue stable au cours de la dernière décennie.

RESUMEN

El atún blanco, Thunnus alalunga, es el principal objetivo de la flota atunera de cebo vivo de Sudáfrica que opera en la costa oeste y suroeste de Sudáfrica, y la captura de Sudáfrica es la segunda más grande de la región, con desembarques anuales de aproximadamente 4.000 t. Se realizó una estandarización de la CPUE de la flota de cebo vivo sudafricana para la serie temporal 2003-2015, utilizando un modelo mixto aditivo generalizado (GAMM) con un error de distribución Tweedie. Las variables explicativas del modelo final incluían año, mes, posición geográfica, potencia del buque, incluida como efecto aleatorio, y especie objetivo, incluida en forma de cargas PCA agrupadas de la composición de la captura normalizada transformada en raíz cuarta. La CPUE estandarizada sigue principalmente la CPUE nominal sin tendencias generales significativas ascendentes o descendentes. El análisis indica que la CPUE para la pesquería de cebo vivo sudafricana de atún blanco ha sido estable durante la última década.

KEYWORDS

Albacore, Thunnus alalunga, standardized CPUE, baitboat, tuna pole, catch, effort, GAMM

¹ Department of Agriculture, Forestry and Fisheries, Private Bag X2, Roggebaai 8012, Cape Town, South Africa, Corresponding author:HenningW@daff.gov.za

1. Introduction

Traditionally, albacore, Thunnus alalunga, is the main target of the South African tuna pole (baitboat) fleet which operates in waters up to 1000 km off the South and West coast of South Africa and off Namibia. The South African catch is the second largest in the region with annual landings of around 5000 t. A large part of the catch is comprised of juvenile and sub adult fish below 900 mm FL which are abundant along the Southern African Atlantic coasts from November to May. The fishery started in the late 1970s and originally targeted yellowfin tuna, T. albacares, but switched to albacore when yellowfin moved off the Cape waters in 1980, a pattern that repeated itself in the middle of the first decade of the 21st century, when the yellowfin became abundant again around the Cape. Although tuna occur in mixed shoals, catches of bigeye tuna T. obesus and skipjack Katsuwanus pelamis are caught in low numbers in comparison to albacore and yellowfin in this fishery. The tuna pole fishery, originally managed as part of the linefishery, became a separate sector after an environmental emergency was declared in 2000 due to the collapse of most of the targeted sparid and sciaenid stocks. After the medium term fishing rights allocation in 2002 the tuna pole fishery sector consisted of 191 vessels of more than 10 m length, of which around 130 were active. Rights were re-allocated in 2013 and with the conclusion of the appeals process, the fishery now consists of 163 rights of which 156 have allocated. Reporting of monthly catch statistics has been compulsory since 1985 and includes daily catch (kg) per species per boat. The fishing positions are also recorded and coded according to a 1×1 degree geographic position. Recently the reporting has been improved to fulfill international data obligations and to facilitate analyses and includes information on fishing hours, number of crew, use of life-bait, Sea-Surface-Temperature and target.

The South African baitboat fleet in the South Atlantic is significant in terms of catch with annual landings of approximately 4000t. Catches vary depending on the availability of albacore and yellowfin tuna in inshore waters. Traditionally the South African fleet has been characterized into three different categories (1) Skiboats, (2) Poleboats and (3) Freezer vessels (Leslie *et al.* 2004): (1) Trailerable skiboats, gamefishing and recreational vessels of less than 25 GRT operating mainly out of harbours around Cape Town and are mostly confined to day trips within a range of 50 nm. Fishes are targeted with pole and with rod and line gear; (2) Pole boats, which represent the bulk of the fleet, are mainly older, displacement type vessels converted from other fisheries. These vessels are up to 30m and 230 GRT. Due to their large size and freezing facilities, these vessels can stay out at sea for long periods and reach the farthest fishing grounds. There is considerable overlap between these categories as many of the modern pole vessels also have freezing capacity. Moreover, there are other factors that influence vessel performance significantly, such as navigational gear and more recently the use of live bait and sonar.

Since 1985, catch statistics have been captured into the National Marine Linefish System (NMLS), a database system developed to capture and analyze recreational and commercial linefishing data (Penney 1993). From 1995 the data capturing procedure began to distinguish tuna pole vessels from linefish vessels. Prior to 2002 the tuna pole sector fell within the multi-species linefishery and there has been a lot of overlap in vessels and targeting. The fishery was only formally separated into tuna pole and traditional linefish sectors in the medium term rights allocation in 2002. As a result of the medium term rights allocation, reporting of tuna catches by species has become mandatory for tuna pole fishery since 2003, whereas in previous years, a large number of records only included an aggregated 'Tuna' catch. The "Pole and Linefishery", as the tuna pole fishery is officially termed since 2015, continues to catch a proportion of the traditional linefish other than tuna species, such as snoek *Thyrsites atun* and yellowtail *Seriola lalandi*. Whereas the former is mainly taken during the off-season, the latter is caught opportunistically year around.

A standardization of the CPUE of the South African baitboat fleet was first undertaken by Punt *et al.* (1996), who used Generalised Linear Models (GLMs), assuming a log-normal error distribution. The model was updated and refined by Leslie in (2000) and Leslie *et al.* (2004), who tested different error structures within a GLM and GLMM framework, using an offset to accommodate zero catches. Smith and Glazer (2007) applied a lognormal GLM to standardize the time series 1999-2005. The work on the standardization has witnessed considerable improvement with the introduction of target as explanatory term and a vast improvement of the underlying data structure (Kerwath *et al.* 2012, West *et al.* 2013). In this contribution we build on the experience of the previous work and provide a standardized CPUE for the time series 2003-2015.

2. Materials and Methods

2.1. Catch and effort data preparation

All records classified as tuna pole trips were extracted from the database for the period 2003-2015 (n = 47.965). Each record included the following information: (1) date, (2) unique vessel number, (3) catch position at a 1×1 degree latitude and longitude resolution and (4) mandatory catch reports in kilogram per day per species. The reported species comprised albacore, yellowfin, big-eye, skipjack tunas and the two non-tuna species yellowtail and snoek.

Records were excluded from the analysis for several reasons. Inspection of the dataset revealed that snoek did not co-occur with albacore in the catches and the only n = 300 records with snoek catches were therefore excluded. The analyses was restricted to designated tuna pole vessels that have fished for at least four years with a minimum of a total 100 days of albacore catch over the period of 2003 to 2015. This way the retained vessels accounted for 90.4% of the records.

Skippers have at times only reported the total catch weight at the end of a multi-day trip. As the data structure only allows for reporting catch per day, the data capturer will then divide the total catch by the number of days the vessel was at sea, thereby creating artificial, repetitive catch per day records. Including these data in the analysis would decrease the overall variance therefore only one of these records was included per trip. The percentage catch records removed from the datasets after this rule was applied was 16.6%.

The total number of records retained for the CPUE standardization was n = 34.407, which corresponds to 72% of the extracted tuna pole records.

2.2. Model framework

Albacore CPUE was standardized using Generalized Additive Mixed Models (GAMMs), which included the covariates *year*, *month*, 1×1 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). CPUE was modelled as catch in kilogram 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.* 1996) 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 (**Figure 2**) and correspondingly two clusters were selected as optimal for the CLARA clustering.

The CPUE records were fitted by assuming Tweedie distribution (Candy 2004, 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 (Candy 2004, Dunn and Smyth 2005). Here, we considered the case of 1 , which represents the special case of a Poisson (<math>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 (**Figure 3**). This resulted in a power parameter <math>p = 1.65 with an associated dispersion parameter of $\phi = 20.5$ for the full GAMM.

The full GAMM evaluated for albacore was:

$$CPUE_{i} = \exp(\beta_{0} + Year + s_{1}(Month) + s_{2}(Long, Lat) + FT + \alpha_{V})$$
(1)

where s_I () denotes cyclic cubic smoothing function for Month, s_2 () a thin plate smoothing function for the twodimensional covariate of *Lat* and *Long*, *FT* is the vector of cluster numbers treated as categorical variable, 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 = 102) 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 behavior, 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 February (*Month* = 1), representative of the first 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:

$$E[CPUE_{y}(\mathbf{X}_{0}^{T}\hat{\boldsymbol{\beta}})] = \frac{1}{A} \sum_{a}^{A} \exp(\hat{\boldsymbol{\mu}}_{y,a})$$
⁽²⁾

and

$$\hat{\sigma}_{y}(\mathbf{X}_{0}^{T}\hat{\beta})] = \sqrt{\frac{1}{A}\sum_{a}^{A}\hat{\sigma}_{y,a}^{2}}$$
(3)

where $\hat{\mu}_{y,a}$ is 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 of **X** is transposed.

3. 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). Total deviance explained by the model was 10.1%, (Table 1). The inclusion of the effect of targeting other species of tuna (**Table 1, Figure 4**), contributed to the greatest improvement in deviance explained and accounted for 70% of the total deviance explained. Given the history of the fishery, which started during high availability of yellowfin tuna around the Cape in 1979, this is not surprising. When yellowfin tuna become sporadically available in the inshore waters, a part of the fishery switches targeting. Although the way that targeting was used here has improved the model, 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.

Determining the vessel specifications according to crew size and trip length are both deemed to be poor indicators of vessel type (Leslie *et al.* 2004, Smith and Glazer 2007). It is challenging to obtain this vessel information (gross registered tonnage (GRT), length, use of live bait and sonar information) for the entire fleet, but a classification into vessel type was attempted in the past (Kerwath *et al.* 2012) based on maximum and average number of crew. However, there was no significant improvement in explanatory power by including

vessel type as categorical variable or by using a subset of vessels from each class as indicator vessels. To include *vessel* as a random effect was deemed the most appropriate solution. Although there was notable variation among vessels (**Figure 5**) and inclusion of the random *vessel* effect produced the most parsimonious error structure, it did not have a large effect on the confidence intervals.

As in the previous analyses the time trend signal derived from the analysis is weak with no upward or downward trajectory over the assessment period. The drop in standardized CPUE and nominal CPUE in 2011-2012 is noteworthy and cannot be explained by targeting, weather or effort shifts (**Figure 6**). Overall, however, the analyses presented here indicate that the CPUE for the South African baitboat fishery for albacore has been stable over the last decade. This analysis represents a considerable improvement in the methodology and data quality. Further improvements are possible through more accurate reporting of fishing position, better classification of fishing time and vessel power and possibly the inclusion of environmental parameters such as sea surface temperature, but such factors can only be included here once a reasonably long time series of consistently recorded data have been collected.

References

- Butterworth, D.S. 1996. A possible alternative approach for generalized linear model analysis of tuna CPUE data. ICCAT Col. Vol. Sci. Pap. 45: 123-124.
- Candy, S. G. 2004. Modelling catch and effort data using generalised linear models, the Tweedie distribution, and random vessel effects. CCAMLR Science 11:59–80.
- Carvalho, F.C., Murie, D.J., Hazin, F.H.V., Hazin, H.G., Leite-Mourato, B., Travassos, P., Burgess, G.H., 2010. Catch rates and size composition of blue sharks (*Prionace glauca*) caught by the Brazilian pelagic longline fleet in the southwestern Atlantic Ocean. Aquat. Living Resour. 23, 373-385.
- Dunn, P., and G. Smyth. 2005. Series evaluation of Tweedie exponential dispersion model densities. Statistics and Computing 15(4):267–280.
- He, X., K. A. Bigelow, and C. H. Boggs. 1997. Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery. Fisheries Research 31:147–158.
- Helser, T. E., E. Punt, and R. D. Methot. 2004. A generalized linear mixed model analysis of a multi-vessel fishery resource survey 70:251–264.
- Leslie, R.W. 2000. Updated standardised south Atlantic albacore Thunnus alalunga CPUE for the South African baitboat fishery, 1985 1999. Col. Vol. Sci. Pap. ICCAT. Madrid, Spain.
- Leslie, R.W., Restrepo, V. and Antony, L. L. 2004. Standardised south Atlantic albacore CPUE for the South African baitboat fishery, 1985-2002. Col. Vol. Sci. Pap. ICCAT. Madrid, Spain. 56(4): 1504-1524.
- Ono, K., A. E. Punt, and R. Hilborn. 2015. Think outside the grids: An objective approach to define spatial strata for catch and effort analysis. Fisheries Research 170:89–101.
- Penney, A.J., Krohn, R. G. and Wilke, C. G. 1992. A description of the South African tuna fishery in the southern Atlantic Ocean. Col. Vol. Sci. Pap. ICCAT. Madrid, Spain. 37: 218-229.
- Penney, A.J. 1993. The National Marine Linefish System. In Beckley, L.E. and R.P. van der Elst (eds). Fish, Fishers and Fisheries. ORI Spec. Publ. 2: 68-72.
- Punt, A.E., Penney, A. J. and Leslie, R. W. 1996. Abundance indices and stock assessment of south Atlantic albacore (*Thunnus alalunga*). Col. Vol. Sci. Pap. ICCAT. Madrid, Spain. 43: 361-371.
- R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org.
- Raîche, G., T. A. Walls, D. Magis, M. Riopel, and J.-G. Blais4. 2013. Non-graphical solutions for Cattell's scree test. Methodology 9(1):23–29.
- Smith, C.D. 2005. Annual report of South Africa. NAT/2005/033. ICCAT. Madrid, Spain.
- Smith, C. D. and Glazer, J. 2007. New standardised south Atlantic albacore CPUE for the South African baitboat fishery, 1999-2005. Col. Vol. Sci. Pap. ICCAT. 60 (2): 481-491.
- Struyf, A., M. Hubertt, and P. J. Rousseeuw. 1996. Clustering in an object-oriented environment. Journal of Statistical Software 1(4):1–30.
- Tascheri, R., J. C. Saavedra-Nievas, and R. Roa-Ureta. 2010. Statistical models to standardize catch rates in the multi-species trawl fishery for Patagonian grenadier (Macruronus magellanicus) off Southern Chile. Fisheries Research 105(3):200–214.
- Thorson, J. T., and C. Minto. 2015. Mixed effects: a unifying framework for statistical modelling in fisheries biology. ICES Journal of Marine Science:doi:10.1093/icesjms/fsu213.
- Venables, W. N. and Ripley, B. D. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.

- Kerwath, S.E., Winker, H. and West, W.M. 2012. Standardization of the catch per unit effort for albacore (*Thunnus alalunga*) for the South African tuna-pole (baitboat) fleet for the time series 1999-2010. . Col. Vol. Sci. Pap. ICCAT. 68 (2): 604-614.
- West W.M., Winker H. and Kerwath S.E. 2013. Standardization of the catch per unit effort for albacore (*Thunnus alalunga*) for the South African tuna-pole (baitboat) fleet for the time series 1999-2011. Col. Vol. Sci. Pap. ICCAT. 70 (3): 1247-1255.
- Winker, H., Kerwath, S.E., Attwood, C.G., 2013. Comparison of two approaches to standardize catch-per-uniteffort for targeting behaviour in a multispecies hand-line fishery. Fish. Res. 139, 118-131.
- 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(0):149–159.
- Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC, Boca Raton, Florida.

Parameter	Res. d.f	df	<i>F</i> -value	р
eta_0	34406			
Year	34394	12	44.5	< 0.001
s(Month)	34386	8	38.8	< 0.001
s(Lat,Long)	34380	6	22.2	< 0.001
FT	34379	1	4120.9	< 0.001
% deviance explained		10.1%		

Table 1. Summary statistics of the GAMM fit for albacore tuna pole fishery off South Africa (2003-2015). Res. d.f.: Residual degrees of freedom, d.f.: degrees of freedom.

Table 2. Nominal and standardized albacore CPUE for the period 2003-2015. SE denotes the standard error of the expected log(CPUE).

	_				
Year	ln(CPUE)	SE	CPUE	95% Confidence Interval	nominal CPUE
2003	3.12	0.217	1320.9	1075.1 - 1634.1	948.1
2004	2.97	0.216	923.6	755.0 - 1137.9	721.0
2005	3.12	0.216	1321.4	1079.0 - 1629.8	995.3
2006	3.09	0.216	1228.7	1002.0 - 1517.4	906.4
2007	3.17	0.215	1474.4	1207.5 - 1813.3	1110.3
2008	3.05	0.216	1126.9	918.3 - 1392.5	909.9
2009	3.18	0.214	1502.1	1234.8 - 1840.5	1328.3
2010	3.10	0.214	1272.8	1048.6 - 1556.3	1281.8
2011	3.01	0.214	1032.8	848.9 - 1265.7	779.7
2012	2.92	0.214	830.5	682.5 - 1017.9	758.4
2013	3.09	0.214	1241.4	1020.9 - 1520.4	1114.2
2014	3.16	0.214	1441.3	1186.5 - 1763.8	985.2
2015	3.11	0.214	1284.4	1055.0 - 1575.0	952.5



Figure 1. Mean yearly **a**) albacore catch (tons) and **b**) tuna pole effort (boat days) at the 1×1 degree reporting resolution. Note that one additional area with catches of 10 tonnes further offshore has been omitted.



Figure 2. Illustrating the selected clusters for the two fishing tactics projected over the first two Principal Components (PCs), where only PC1 was determined to be non-trivial. FT 1-YLF: Cluster one is dominated by yellowfin catches. FT 2-ALB: Cluster two dominated is by albacore catches.



Figure 3. Log-likelihood profile for over the grid of power parameters values (1 of the tweedie distribution. The vertical dashed line denote the optimized*p*used in the final standardization GAMM.



Figure 4. Predicted standardized effect sizes for *Month* and Fishing Tactic (*FT*). Grey-shaded areas and error bars denote the 95% confidence intervals, respectively.



Figure 5. Random effects coefficients (black dots) illustrating the deviation from the mean of zero accros the 102 vessels retained for the analysis. Grey shaded areas denote the 95% confidence interval of the mean α_{v} .



Figure 6. Comparison of normalized nominal and standardized CPUE for the albacore tuna pole fishery off South Africa for the time series from 2003 to 2015. 95% confidence intervals denoted by grey shaded areas.