Standarized CPUE for skipjack tuna from the european purse Seine fleet in the Indian ocean from 1984 to 2013

by

M. Soto[1], F. Fernández[1], A. Delgado de Molina[2], E. Chassot[3]

Abstract

In this document an abundance index is obtained for the tropical tuna skipjack (*Katsuwonus pelamis*) of European purse seine fishery in the Indian Ocean from 1984 to 2013 using generalized linear models. Catch and effort data come from detailed daily logbooks. Catch rates are modeled using the delta lognormal model. The method estimates a cpue from aggregated positive catches, and the proportion of catches, so the final individual abundance indices are calculated multiplying both estimators for each species. Explanatory factors used in the analysis are: year, zone, quarter, holding capacity, country and starting date of the vessel. Year is the most explanatory factor of variability in cpue followed by the fishing area and the quarter. Vessel characteristics have a significant explanatory effect in observed positive catch rates.

Introduction

Since the last two decades, the increasing use of drifting fishing aggregative devices (FADs) by the purse seine fleets operating in the Indian Ocean has changed the length distributions of the tunas tropical landings. In contrast to non-associated school sets which target large fish (mainly yellowfin, *Thunnus albacares*), FADs fishing operations concern skipjack (*Katsuwonnus pelamis*) and juveniles of yellowfin and bigeye tunas (*Thunnus obesus*). With this consideration in mind, the aim of this paper is to develop a standardization procedure of CPUEs for FADs fishing operations. Since, purse seine fishermen may target alternatively associated schools and FADs schools, the presence of a high amount of zero-catch per fishing day may be expected in the data set. As explained in the Method section, in such a situation, delta-lognormal method is an appropriate tool for standardizing CPUEs (Lo et al, 1992, Stefansson, 1996)

Material and Methods

Standarized catch rates of skipjack were estimated using the generalized linear model assuming a delta-lognormal error distribution (Soto *et al.*, 2009). The analysis has been carried out with catch and effort data from logbooks, once the specific composition of catches has been corrected (Anon, 1984, Pallarés y Petit, 2001...) and from detailed fleet data. Catch and effort data are obtained by set, while fleet data contain information

---

[1] Instituto Español de Oceanografía. Corazón de María 8. 28002 Madrid (ESPAÑA).
E_mail:maria.soto@md.ieo.es.
[2] Instituto Español de Oceanografía. Centro Oceanográfico de Canarias. Apdo. de Correos 1373.
38080 Santa Cruz de Tenerife. Islas Canarias (ESPAÑA). E_mail: tunidos@ieo.rcanaria.es.
[3] Institut de Recherche pour le D´eveloppement, UMR 212 EME (IRD/Ifremer/Universit´e de Montpellier 2), Avenue Jean Monnet, BP 171, 34 203 S`ete Cedex, FRANCE.

about age of vessel, physical characteristics (length, holding capacity, GTR, ..) and vessel history. French, Spanish and NEI fleet data have been analyzed together. In this analysis, the NEI fleet was associated to the Spanish purse seine fleet following results of discriminant analysis (Soto et al., 2002). The period considered goes from 1984 to 2013, years where detailed logbooks are available. Fishery started in 1981, but first years haven´t been taken into account as they are less representative.

It was considered a minimum threshold of effort by vessel of 120 fishing days per year. This threshold was selected after to analyze the yields as a function of fishing time of vessels and to observe that there was no correlation between both, neither between fleets nor between the whole of the fleets, and also, that the variability, higher for vessels with short fishing periods, was tending to stabilize from this threshold. Later, a selection of vessels operating in the fishery for more than 15 years was done with the intention of analyzing data from vessels that would contribute to obtain trends more representative of real abundance.

Once the selection of representative vessels was done, there were established categories according to the holding capacity, measured in m^3, trying to balance all the categories with a representative number of observations. This characteristic defines well the vessel capacity as the probability of bias and imprecisions are very little. Vessel categories observed are the following:

Category	Holding capacity
1	< 750 m^3
2	750 - 1249 m^3
3	> 1250 m^3

Considering the possible interaction of the fleet and the category of the vessels as in Soto *et al*. (2003) a mixed variable category-country was defined with the following levels:

Level	Country	Harvest capacity
1	France	< 750 m^3
2	France	750 - 1249 m^3
3	France	> 1550 m^3
4	Spain	< 750 m^3
5	Spain	750 - 1249 m^3
6	Spain	> 1550 m^3

Data of catches and effort were restricted to those obtained from FADs, aggregated by logs per day, because the catches of skipjack of the purse seine fleet during the period considered are obtained almost exclusively from logs. Fishing areas selected for skipjack catches were East Somalia, Maldivas, Chagos and Canal Mozambique, those under fishing on FADs mode.

As it is not possible to allocate effort by set between species, catches were aggregated by day and then by month to avoid the excess of null observations in catches and also in the number of sets; then, a nominal monthly CPUE was defined as:

$$CPUE = \frac{SKJ}{nset}$$

where *SKJ* are the catches of skipjack in tons and *nset* is the nominal effort of the European purse seine fleet measured in number of sets by day aggregated by month. The proportion of catches of of skipjack over total catches is

$$p_{skj} = \frac{SKJ}{total\_catches}$$

The estandarization procedure used was the generalized linear models (GLM) (McCullagh and Nelder, 1989). The *CPUE* index was estimated assuming that *CPUE+k* follows a lognormal distribution based on the observation of the normal QQ-plots and the results of Kolmogorov test, where

$$k = 0.01 \cdot mean\, CPUE$$

The proportion of catches of skipjack, p_{skj}, was modelled independently from the *CPUE* assuming a binomial error distribution.

The independent factors related with abundance considered were: *year*, *fishing area* and *quarter*. Regarding factors related with the vessels, it was considered a combined variable of category of holding capacity and country (*category-fleet*). The age of vessel used in previous studies of *CPUE* (Soto, 2002) was represented by the *operating date* of the vessel in the analysis. Nevertheless, this factor usually causes interactions with the factor year, is not very significant in the models, and masked the abundance effect.

The abundance index was obtained from GLM analysis. By one side, a combined positive *CPUE* was estimated from year LSMeans of the lognormal model. By the other side, estimated proportions of catches were estimated from year LSMeans of the binomial model. The specific index was finally calculated as the product of *year* LSMeans of lognormal model and binomial models. Variance of the indices were calculated using the Delta method (Casella, G., 2002), based on the Taylor development of the function

$$g(\mu, p) = \mu \cdot p,$$

where μ is the estimator of *CPUE* from the lognormal model and *p* the estimator of proportion of catches, assuming that both estimators are independent and there are no covariace terms different from zero.

Analysis and model formulations for the delta model were done using the R statistical software package (R Development Core Team, 2013). In general, model evaluation and diagnosis was carried out through residual analysis (McCullagh and Nelder, 1989). Diagnostic plots are presented for each delta model component: partial residuals for all components, including partial against *year* for each species, and QQ-plots and histograms of Chi-squared residuals for the lognormal component.

A stepwise regression procedure was used to determine the set of systematic factors that significantly explained the observed variability in each model. A Chi-squared test was used to evaluate the statistical significance of an additional factor (McCullagh and Nelder, 1989). Furthermore, the corresponding percentage of deviance explained by each factor relative to the maximum model was estimated to obtain a profile of the most important explanatory factors in the model. A statistically significant variable (*p-value*<0.05) may, in some instances, be omitted from the model if the amount of variation explained by the variable is small in relation to the complexity that it adds (Stefánsson, 1996). The final models included the *year* plus a selection of other explanatory factors that explained more the 5% of the deviance percentage in the models.

3. Results

Positive catches

For the lognormal component, all the factors included are statistically significant. *Year* is the most significant factor to explain the variability observed, followed by by *area* and quarter, explaining 38%, 33% and 25% of the deviance, respectively. Regarding the factors related with the fleet, the combined factor *holding capacity-fleet* explains the 2.4% of the variability, while the *operating date* only explains the 0.1% of the deviance in the model, unless they are statistically significant.

Proportion of positive catches

In the binomial models for the proportion of positives catches, the factor *year* is the most important, explaining almost the 80% of the deviance. *Fishing area* explains the 14% and *starting date of the vessel* the 4.5% of the deviance for SKJ. All factors are significant in the binomial model.

Selected model

The results of deviance analysis are shown in Table 1. For the lognormal model *quarter*, *year* and *fishing area* are the main explanatory factors and selected for the final model. The stepwise regression includes also the *starting* date of the vessel, and harvest capacyti-fleet but it was decided not to include them in the final model as they explain less than 5% of the deviance in the model. For the binomial model, all factors are significant but only Year and area are selected for the final model.

CPUE

Observed and standarized scaled cpue series by specie are shown in Figure 2. There are no clear trends during the period considered in both series, and nominal *CPUEs* are very similar to the standardized *CPUE*s. The nominal *CPUE* is above the estandarize one for almost all the years, and outside the confidence intervals in some cases. The whole period (1982-2013) can be divided in two similar cycles, before and after 1997, where the standardized *CPUE* reaches its minimum value, almost identical than the most recent value considered, 2013. These cycles begin with an increasing average trend in the *CPUE* of 6/7 years and are followed by a decreasing average trend of 10 years.

Fitting diagnoses are show in Figure 2 for lognormal model. The residuals follows a relatively linear expected pattern for aggregated catches in the QQ-plot

4. Discussion

The delta method has been widely used to construct abundance indices for tuna species. No strong trends appear in the series of standarized cpue. The source of variability that comes from the fleet is represented by the factors *harvest capacity-fleet* and *starting date*. The factor *harvest capacity-fleet* represents the effect of vessel class and it is only significant to explain the variability of aggregated catch and not for the proportion of individual catches, i.e. there is no evidence of differences between proportions of individual catches between vessel classes. Also, the *starting date* of the vessel has been removed from the final model because the proportion of explained variability of global catch rates is very little and it is not statistically significant for the proportion of individual catches.

It appears the effect of the vessel is independent of the proportions of catches of skj and the binomial variables are only related with the abundance factors. Fleet factors have been removed from the binomial final model

In general, the standarization procedure showed that vessel characteristics (country, harvest capacity and age of vessel) have a relative minor explanatory effect on the catch rate in the purse seine fishery.

The goal of the standarization procedure is to eliminate the annual variability in the data that is not attributable to the changes in abundance (Maunder y Punt, 2004). This result is in part achieved as it can be seen in Table 3, where the CVs of nominal cpues are higher than the standarized ones.

References

CASELLA, G., Berger, R. L. (2002), Statistical Inference, 2nd ed.

MAUNDER, M. N. and Punt, A. E., 2004. Standarizing catch and effort data : a review of recent approaches. Fisheries Research 70, pp : 141-159.

McCULLAGH, P. and Nelder, J.A., 1989. Generalized linear models, second ed. Ed. Chapman & Hall.

PALLARÉS, P., Ch. Petit. 1998. Tropical tunas: New sampling and data processing strategy for estimating the composition of catches by species and sizes. Col. Vol. Sci. Pap. ICCAT, 48(2): 230-246 (1998).

R Development Core Team. 2013. 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.
SOTO, M., Pallarés, P., Gaertner, D., Delgado de Molina, A. , Fonteneau, A, Ariz, J., 2002. Standardization of tropical purse seine fishing effort by generalized linear model (GLM). IOTC 2002-WPTT-26.

SOTO, M. , E. Chassot and F. Marsac, 2008. Standarized catch rates for yellowfin (*Thunnus albacares*) and skipjack (*Katsuwonus pelamis*) for the European purse seine fleet of the Indian Ocean, 1984-2007. IOTC 2008-WPPT-26.

SOTO, M., Pallarés, P., Delgado de Molina, A., Gaertner, D. 2009. Standardized CPUE for juvenile yellowfin, skipjack and bigeye tuna from the european purse seine fleet in the Atlantic Ocean from 1991 to 2006. *Collective Volume of Scientific Papers ICCAT* **64**, 4 (2009) 1044-1053.

STEFÁNSSON, G. 1996. análisis of groundfish survey abundante data: combining the GLM and delta approaches. ICES journal of Marine Science, 53:577-588.

Table 1: Deviance table for the lognormal model and the proportion of catches. Explanatory factors are emboldened.

Model formulation	**Df**	**Change in Deviance**	**Residual Deviance**	**p-value**	**Percentage of total deviance**
Positive CPUE					
1	1		7646		
Factor					
+year	29	790	6857	1,3E-200	38,3
+quarter	3	523	6334	4,3E-149	25,3
+area	3	699	5635	1,6E-199	33,9
+harvest capacity-fleet	4	50	5585	1,5E-13	2,4
+starting date	1	2	5583	1,1E-01	0,1
Proportion of positive SKJ					
1	1		672		
Factor					
+year	29	100	572	1,1E-09	78,0
+quarter	3	2	570	4,9E-01	1,9
+area	3	18	552	4,5E-04	14,0
+harvest capacity-fleet	4	2	550	7,3E-01	1,6
+starting date	13	6	544	9,5E-01	4,5

Table 2: Relative standarized *CPUE* and standard deviation for YFT_1, *SKJ* and BET_1.

year	CPUE	CV	nominal
1984	0,5	0,2	0,6
1985	0,5	0,1	0,6
1986	0,7	0,1	0,7
1987	0,6	0,1	0,7
1988	0,9	0,1	0,9
1989	0,6	0,1	0,6
1990	0,6	0,1	0,7
1991	0,6	0,1	0,8
1992	0,7	0,1	0,8
1993	0,7	0,1	0,8
1994	0,7	0,1	0,8
1995	0,5	0,1	0,7
1996	0,5	0,1	0,6
1997	0,4	0,1	0,5
1998	0,5	0,1	0,6
1999	0,6	0,1	0,7
2000	0,8	0,1	0,9
2001	0,7	0,1	0,8
2002	1,0	0,1	1,0
2003	0,7	0,1	0,9
2004	0,7	0,1	0,8
2005	0,7	0,1	0,7
2006	0,6	0,1	0,7
2007	0,4	0,1	0,4
2008	0,5	0,1	0,5
2009	0,6	0,1	0,6
2010	0,6	0,1	0,6
2011	0,5	0,1	0,6
2012	0,4	0,1	0,4
2013	0,4	0,2	0,4

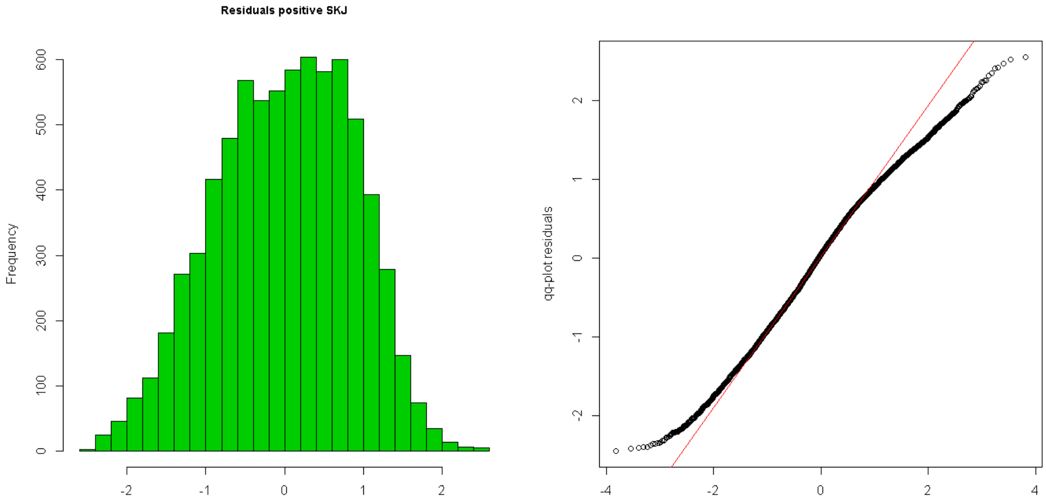

Figure 1:Histogram and Q-Q plot of Chi-squared residuals and partial residuals of lognormal model for the combined cpue.

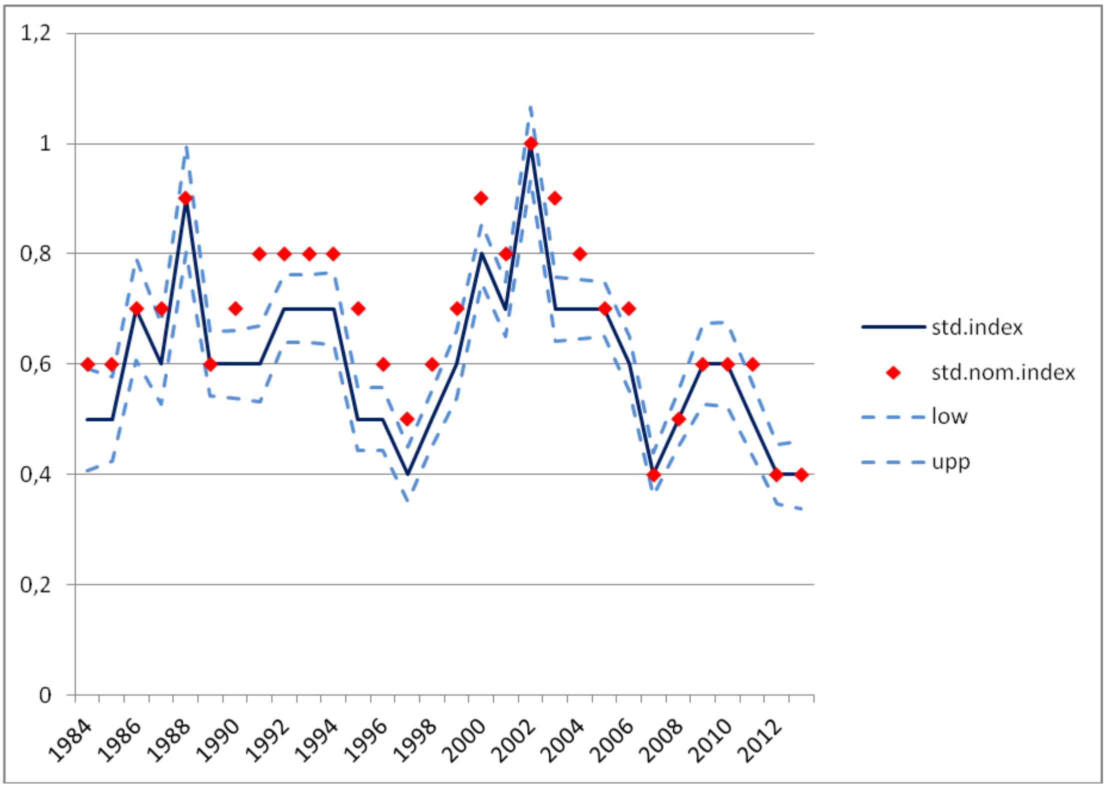

Figure 2: Standarize relative EU Purse Seine CPUE for skipjack in the Indian Ocean. Confidence intervals and nominal values are also plotted.