Received: 10 October, 2011

STANDARDIZED CPUE OF MAJOR SHARK SPECIES CAUGHT BY THE PORTUGUESE LONGLINE FISHERY IN THE INDIAN OCEAN

Rui Coelho[1,2], Pedro G. Lino[1] & Miguel N. Santos[1*]

SUMMARY

Portuguese longliners targeting swordfish and operating in the Indian Ocean regularly capture elasmobranch fishes as by-catch. Of those, the blue shark (*Prionace glauca*) and the shortfin mako (*Isurus oxyrinchus*) constitute the two main shark species captured. A recent effort by IPIMAR has been recovering historical catch data on elasmobranchs captured since the late 1990's to the present date. This paper reports the CPUE trends of those two major shark species during that time period. Nominal CPUEs for each species were calculated as kg per 1000 hooks deployed. Data from Vessel Monitoring Data (VMS) is available and processed between 2006 and 2010, and for those years the CPUEs were standardized with Generalized Linear Models (GLMs) using year, month (categorized by quarter), location and vessel as explanatory variables. Model validation was carried out with a residual analysis. The results presented in this paper provide the first preliminary trends on elasmobranch catch rates available from the Portuguese longline fishery operating in the Indian Ocean.

KEYWORDS: By-catch, CPUE standardization, generalized linear models, longlines, Prionace glauca, Isurus oxyrinchus.

1: Instituto de Investigação das Pescas e do Mar (IPIMAR), Av. 5 de Outubro s/n, 8700-305 Olhão, PORTUGAL
2: Centro de Ciências do Mar (CCMAR), Univ. Algarve, Campus de Gambelas Ed 7, 8005-139 Faro, PORTUGAL
*: Corresponding author e-mail: mnsantos@ipimar.pt

1. Introduction

Fisheries management is usually based on stock assessment models that require data on the abundance of the species under assessment (Hilborn & Walters 1992). Ideally, data for such models should be fishery-independent with a randomized experimental design. However, when assessing pelagic and migratory species that cover wide geographical areas (e.g. tunas, billfishes and pelagic sharks), fisheries-independent sampling designs would generally be too costly and difficult (even impossible) to implement. Therefore, most stock assessments currently carried out for pelagic species are based on fishery-dependant data, available from the commercial fisheries that interact with those species.

The data usually gathered from the commercial fisheries and analyzed is the Catch per Unit of Effort (CPUE, either in number or biomass), and it is important to standardize those CPUEs to account for effects (consequence of the fishery-dependence) other than the annual effects that are being analyzed. By standardizing the CPUEs, the effects of the covariates considered are removed from the annual CPUE values, and those standardized CPUEs can then be used as annual indexes of abundance.

The objective of this study is to present preliminary CPUE indexes for the two main by-catch shark species (BSH – *Prionace glauca* and SMA - *Isurus oxyrinchus*) captured by the Portuguese pelagic longline fishery targeting swordfish in the Indian Ocean.

2. Material and methods

2.1. Data collection

Data for this study refers to the official fisheries statistics collected by the Portuguese Fisheries authorities. The catch data in the paper refers to the total weight per species captured per fishing set. In a recent effort by the Portuguese Marine Research Institute (*IPIMAR*), this historical catch data from the Portuguese longliners targeting swordfish in the Indian Ocean started to be compiled and analyzed. The current database comprises information from 10,242 fishing sets carried out between 1998 and 2010. Information on effort (number of hooks used per set) is available for 8,085 of those fishing sets carried out between 1999 and 2010 (**Table 1**).

General location (FAO major fishing areas: 47, 51 or 57) is available for the entire time series, while starting in 2005 more detailed information (on the FAO Subarea) started to be collected. The Vessel Monitoring System (VMS) was implemented on these vessels during the 2000's, although these data is currently being filtered, processed and analyzed. The location data between 2006 and 2010 is already integrated in the database, while the information from the previous years is still being processed and not yet available. Therefore, this study presents nominal CPUE information between 1999 and 2010, while the preliminary standardized CPUEs were only estimated and analyzed between 2006 and 2010 (using VMS data integrated with the available catch and effort data) (**Table 1**).

Table 1: Number of fishing sets with catch, effort and location (VMS) information carried out by the Portuguese longline fleet in the Indian Ocean. The percentage of sets per year analyzed for this paper is indicated. NA refers to those years for which VMS information is not yet processed and therefore not used for the analysis.

Year	Sets with catch information	Sets with effort information	Sets with VMS information	% used for analysis
1998	113	0	NA	NA
1999	247	195	NA	NA
2000	327	324	NA	NA
2001	701	443	NA	NA
2002	877	578	NA	NA
2003	866	525	NA	NA
2004	756	495	NA	NA
2005	896	652	NA	NA
2006	2221	1886	1559	70.2
2007	1723	1479	1300	75.4
2008	360	360	242	67.2
2009	525	525	381	72.6
2010	630	623	522	82.9

2.2. Data analysis

The response variable considered for this study was Catch per Unit of Effort (CPUE), measured as biomass of live fish (kg) per 1,000 hooks deployed.

The standardized CPUEs were estimated with Generalized Linear Models (GLMs) using the Delta method approach. This approach was chosen because pelagic sharks are captured as by-catch in this fishery, and there are therefore some sets with zero catches in the database, particularly for SMA. With the Delta method approach two separate models are estimated. The first model assumes a binomial error distribution with a logit link function that is used to model the proportion of fishing sets with positive catches. For this model, the binomial response variable was coded with 1 = set with positive catches for the species of interest and 0 = set with zero catches for the species of interest. The second model uses the nominal CPUE of the positive sets for the species of interest as the response variable, and assumes that those positive catches follow a normal error distribution after a log-transformation.

The initial covariates considered for the models were:

- Year (analyzed between 2006 and 2010);
- Vessel (categorical variable corresponding to the different vessels);
- Quarter of the year (4 categories: 1 = January to March, 2 = April to June, 3 = July to September, 4 = October to December);
- Latitude (4 categories divided by the quartiles);
- Longitude (4 categories divided by the quartiles).

Significance of the explanatory variables was assessed with likelihood ratio tests comparing each univariate model to the null model (considering a significance level of 5%), and by analyzing the deviance explained by each covariate. Goodness of fit and model validation was carried out with a residual analysis.

The final standardized CPUEs were estimated by least square means (LSMeans). The final expected CPUE was calculated as the yearly probability of having a positive set multiplied by the expected catch rate conditional to the set being positive.

All statistical analysis for this paper was carried out with the R Project for Statistical Computing version 2.13.0 (R Development Core Team 2011).

3. Results and Discussion

3.1. Description of the catch and effort

The total effort (in number of sets and hooks deployed) of the Portuguese longline fleet in the Indian Ocean remained relatively constant between 1999 and 2004, followed by an increase during 2006-2007. For the more recent years of 2008 to 2010 the effort was again similar to the initial years of the early 2000's (**Figure 1**). The total BSH and SMA catches also tended to increase initially, with a peak during 2006-2007, followed by a sharp decrease in 2008. During recent years, a slight increase has been observed. Through the entire period the catch variations accompanied the effort trend (**Figure 1**).

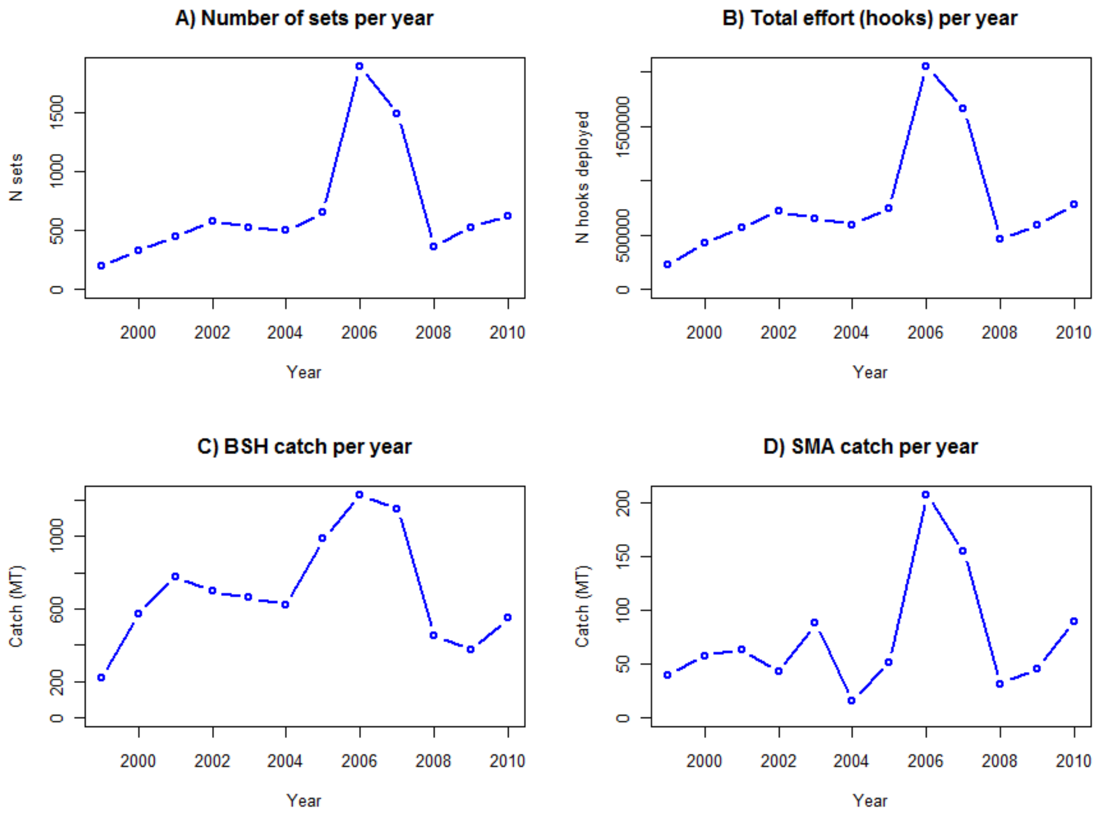

Figure 1: Descriptive plots of the total effort in sets (A) and hooks (B), and the total catch of blue shark (C) and shortfin mako (D) (in metric tons - MT) for the Portuguese longline fleet operating in the Indian Ocean.

3.2. CPUE analysis for blue shark

The percentage of fishing sets with zero catches for BSH in the dataset was 4.0%. The nominal BSH CPUE data was asymmetrical and skewed to the right, but after a log-transformation the data becomes more symmetrical and bell shaped, closer to a normal distribution (**Figure 2**). Therefore, using a Gaussian distribution for modeling the log-transformed blue shark CPUE data seems to be a reasonable procedure for CPUE standardization.

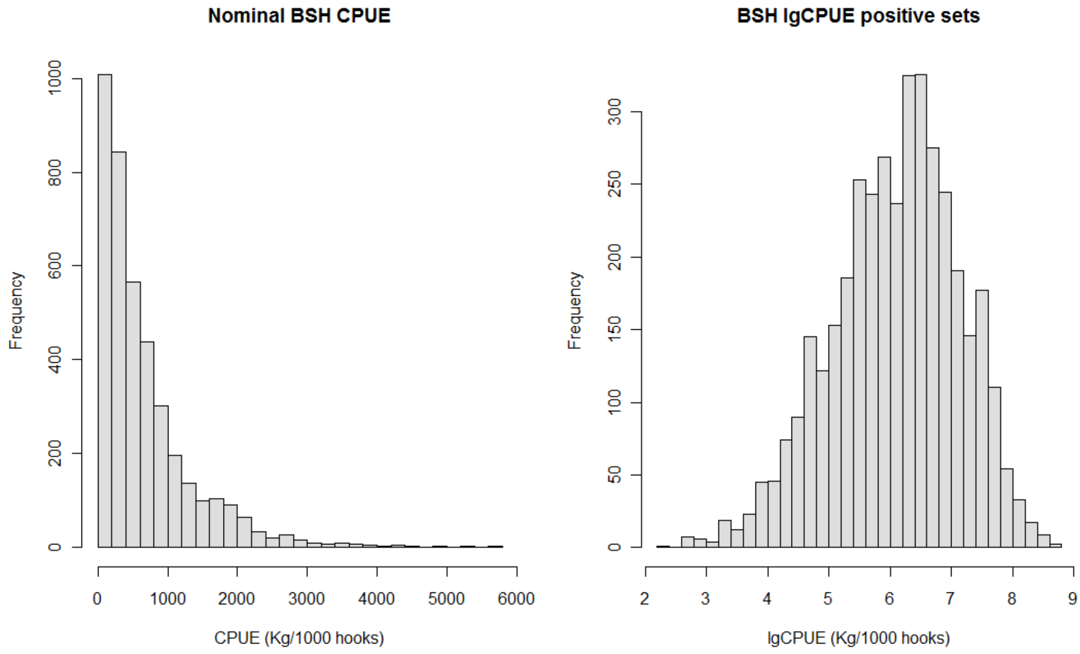

Figure 2: Distribution of the nominal blue shark CPUE, and the log-transformed blue shark CPUE conditional to the set being positive.

The explanatory variables tested for the BSH all contributed significantly for explaining part of the deviance, and therefore the model used was the complete simple effects model (**Table 2**). The covariate that contributed more for explaining part of the deviance was the vessel effect, followed by quarter of the year. The other covariates contributed less, but were also significant. As more VMS data becomes available in the near future, we expect to add additional years to the analysis and also include significant first order interactions in the models.

In terms of residual analysis, the Pearson residuals seemed to be randomly distributed along the predicted values and without any noticeable trends in terms of increasing or decreasing variance. The QQ plot showed a very good fit of the residuals to the expected normal values, and the histogram of the distribution of the residuals also followed a bell shaped normal distribution (**Figure 3**). Some potential outliers were identified, but given the preliminary nature of these models those outliers were not excluded from the final models.

Table 2: Deviance of the parameters used for the blue shark models. For each parameter it is indicated the degrees of freedom (Df), deviance explained (absolute value and percentage), residual deviance left after incorporating each parameter, total cumulative deviance explained by the model (in %), and significance of each parameter.

Model for positive catch rate values						
Parameter	Df	Deviance	Resid. deviance	% deviance	% total deviance	P-value (Chi^2 test)
Null	1		4079.7			
Year	4	115.13	3964.6	2.82	2.82	< 0.001
Vessel	13	870.87	3093.7	21.35	24.17	< 0.001
Quarter	3	399.15	2694.6	9.78	33.95	< 0.001
Latitude	3	155.01	2539.5	3.80	37.75	< 0.001
Longitude	3	90.96	2448.6	2.23	39.98	< 0.001
Model for proportion of positive catches						
Parameter	Df	Deviance	Resid. deviance	% deviance	% total deviance	P-value (Chi^2 test)
Null	1		1337.5			
Year	4	13.743	1323.8	1.03	1.03	0.008
Vessel	13	109.831	1213.9	8.21	9.24	< 0.001
Quarter	3	29.523	1184.4	2.21	11.45	< 0.001
Latitude	3	8.358	1176.1	0.62	12.07	0.039
Longitude	3	34.937	1141.1	2.61	14.68	< 0.001

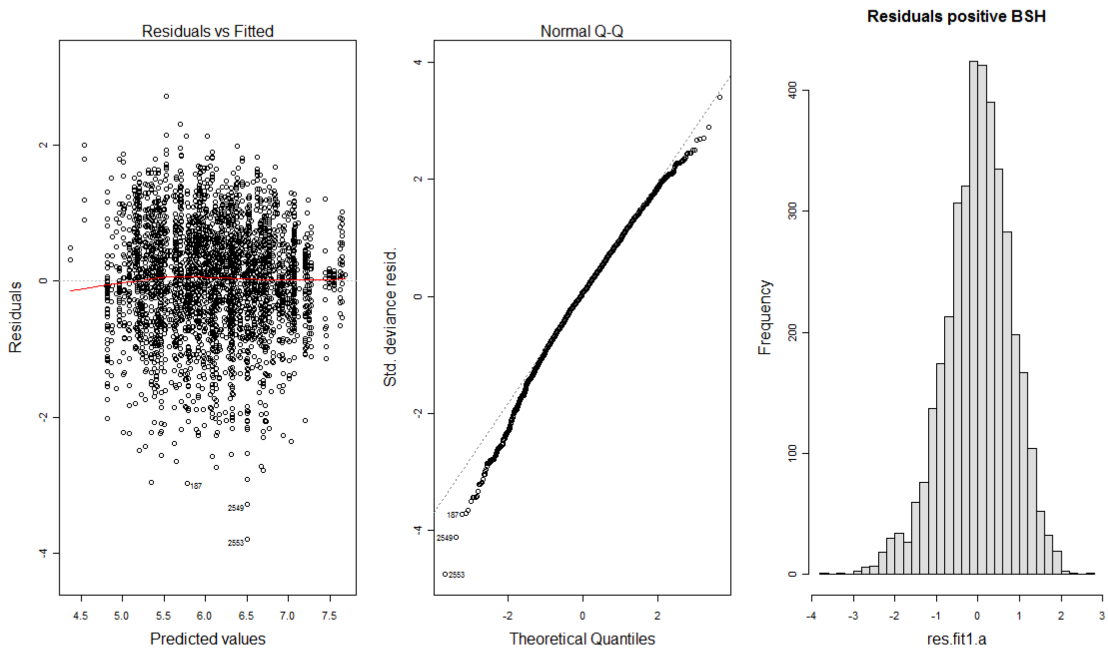

Figure 3: Residual analysis of the final simple effect model used for the blue shark CPUE standardization. Left graphic represents the Pearson residual along the fitted values; the middle graphic represents the QQPlot; and, the graphic on the right side represents the histogram of the frequency distribution of the Pearson residuals.

The nominal CPUEs of the BSH catches between 1999 and 2010 (**Figure 4**) showed some significant variability along the years and a general decreasing trend. The years with the highest nominal CPUEs were 2000, 2001 and 2005. For the standardized series analyzed (between 2006 and 2010), no apparently significant trends are noticeable with the standardized CPUEs remaining relatively stable between those years (**Table 3, Figure 4**). It should be noted that the time series of standardized CPUEs analyzed is still very short (5 years), and should therefore be regarded as a preliminary analysis.

Table 3: Nominal CPUEs and relative index of abundance (kg/1000 hooks) for blue sharks captured by the Portuguese pelagic longline fishery in the Indian Ocean. For the standardized CPUEs it is indicated the standard error (SE), the Coefficient of Variation (CV in %), and the upper and lower limits of the 95% Confidence Intervals (CI).

Year	Nominal CPUE	Index	CV (%)	SE	Lower 95% CI	Upper 95% CI
1999	955.0					
2000	1349.4					
2001	1379.7					
2002	965.5					
2003	994.3					
2004	1024.5					
2005	1429.2					
2006	622.8	567.8	3.2	18.5	531.6	604.0
2007	742.4	790.7	6.3	53.2	686.5	894.9
2008	977.8	522.9	7.0	34.0	456.3	589.6
2009	619.1	721.8	5.9	41.8	639.8	803.8
2010	721.7	606.7	3.4	20.5	566.5	646.8

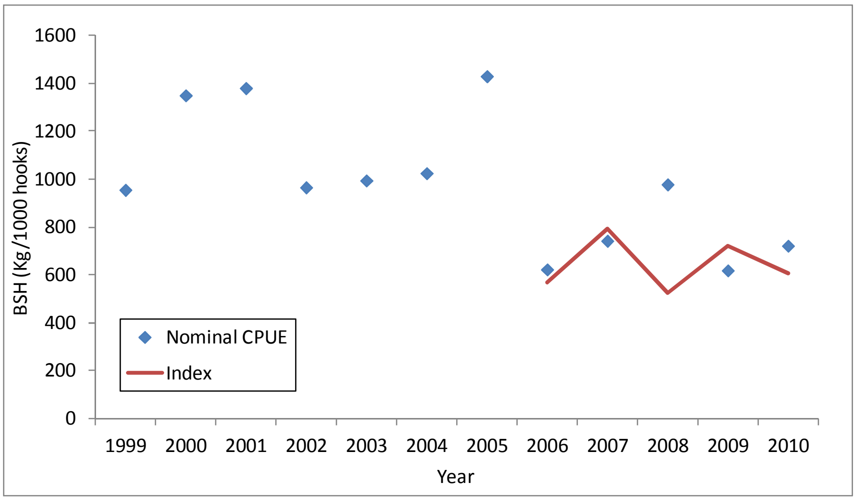

Figure 4: Plot of the annual relative index of abundance for the blue shark captured by the Portuguese pelagic longline fishery in the Indian Ocean. Blue-diamond markers represent nominal CPUEs and the solid red line represents the standardized CPUEs.

3.3. CPUE analysis for the shortfin mako shark

For the shortfin mako the percentage of fishing sets with zero catches was much higher than for the blue shark, specifically 36.7%. The nominal SMA CPUE distribution was also skewed to the right, with a peak of initial zero values (**Figure 5**). With a log-transformation of the positive sets the data becomes more symmetrical and bell shaped, closer to what is expected by a normal distribution (**Figure 5**). Therefore, and like in the procedure also carried out for the blue shark, using a Gaussian distribution for modeling the log-transformed shortfin mako CPUEs also seems a reasonable procedure for this species.

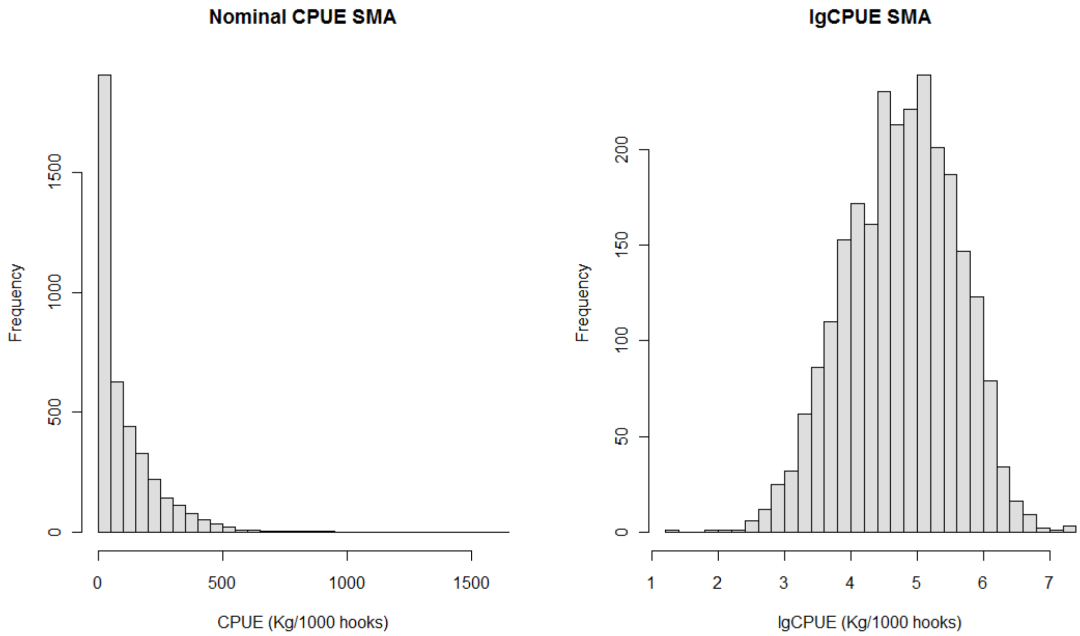

Figure 5: Distribution of nominal CPUEs and log-transformed CPUEs (conditional to the positive fishing sets) for shortfin mako captured in the Indian Ocean by the Portuguese longline fleet.

All the explanatory variables tested for the SMA contributed significantly for explaining part of the deviance, and therefore the models used were the complete simple effects model (**Table 4**). The factor that contributed mostly for explaining the deviance was the vessel effect, followed by year, longitude and latitude. Quarter of the year seemed to be the variable contributing less for the models. Likewise, as more VMS data becomes available we expect to add additional years to the analysis.

Table 4: Deviance of the parameters used for the shortfin mako models. For each parameter it is indicated the degrees of freedom (Df), deviance explained (absolute value and percentage), residual deviance left after incorporating each parameter, total cumulative deviance explained by the model (in %), and significance of each parameter.

Model for positive catch rate values						
Parameter	Df	Deviance	Resid. deviance	% deviance	% total deviance	P-value (Chi^2 test)
Null	1		1731	0	0	
Year	4	31.1	1700	1.80	1.80	< 0.001
Vessel	13	225.7	1474	13.04	14.84	< 0.001
Quarter	3	9.0	1465	0.52	15.36	0.001
Latitude	3	14.1	1451	0.81	16.17	< 0.001
Longitude	3	16.6	1435	0.96	17.13	< 0.001
Model for proportion of positive catches						
Parameter	Df	Deviance	Resid. deviance	% deviance	% total deviance	P-value (Chi^2 test)
Null	1		5251	0		
Year	4	71.0	5180	1.35	1.35	< 0.001
Vessel	13	968.1	4212	18.44	19.79	< 0.001
Quarter	3	9.2	4203	0.18	19.96	0.026
Latitude	3	26.9	4176	0.51	20.48	< 0.001
Longitude	3	36.5	4139	0.69	21.17	< 0.001

In terms of model validation, the Pearson residuals seemed randomly distributed along the predicted values and without any noticeable trends in terms of increasing or decreasing variance. The QQ plot showed a very good fit of the residuals to the expected normal values. Finally, the histogram of the residuals also followed a bell shaped normal distribution (**Figure 6**). Some potential outliers were identified (particularly two that were easily identified with the Persons residuals), but given the preliminary nature of these models they were not removed.

The nominal CPUEs of SMA catches between 1999 and 2010 (**Figure 7**) showed some significant variability along the years. The standardized CPUEs analyzed between 2006 and 2010 also showed some variability (**Table 5, Figure 7**). Like for the blue shark analysis, it should be noted that the time series of standardized CPUEs analyzed for the shortfin mako is still very short, and should therefore also be regarded as preliminary.

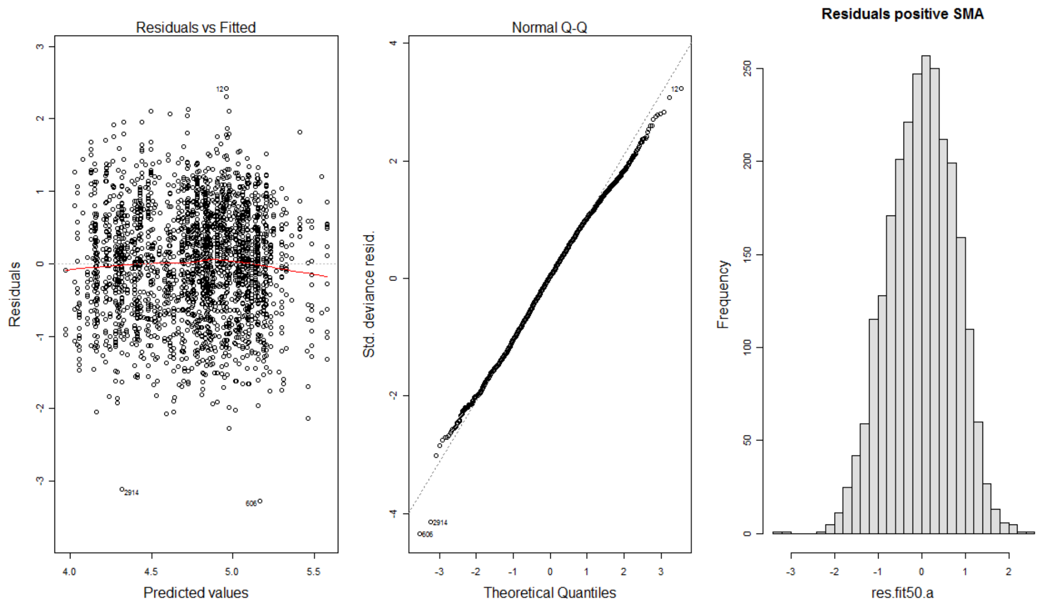

Figure 6: Residual analysis of the final model used for the shortfin mako CPUE standardization. Left graphic represents the Pearson residuals along the fitted values; the middle graphic represents the QQPlot; and, the right side graphic represents the histogram of the frequency distribution of the Pearson residuals.

Table 5: Nominal CPUEs and relative index of abundance (kg/1000 hooks) for shortfin mako captured by the Portuguese pelagic longline fishery in the Indian Ocean. For the standardized CPUEs it is indicated the standard error (SE), the Coefficient of Variation (CV in %), and the upper and lower limits of the 95% Confidence Intervals (CI)..

Year	Nominal CPUE	Index	CV (%)	SE	Lower 95% CI	Upper 95% CI
1999	169.5					
2000	135.6					
2001	106.4					
2002	56.7					
2003	133.3					
2004	23.1					
2005	72.5					
2006	109.0	115.9	5.1	4.6	106.8	125.0
2007	97.4	129.6	7.5	8.6	112.7	146.6
2008	78.9	117.2	7.7	10.7	96.3	138.2
2009	68.3	167.7	7.6	11.5	145.2	190.2
2010	126.8	106.8	4.3	6.1	94.7	118.8

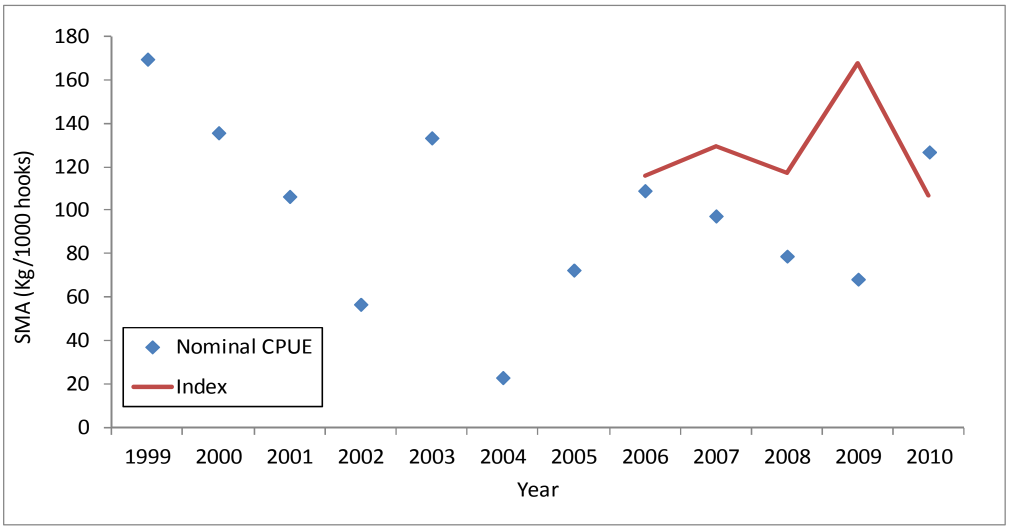

Figure 7: Plot of the annual relative index of abundance for shortfin mako from the Portuguese pelagic longline fishery in the Indian Ocean. Blue-diamond markers represent the nominal CPUE, and the solid red line represents the standardized CPUEs.

3.4. Final considerations

Using GLMs with the Delta method approach is a commonly used procedure to analyze fisheries data with zeros in the response variable, and has been previously applied to the blue and mako sharks (e.g. Cortés 2009; Mejuto et al. 2009). With the Delta method approach, two separate models are estimated. The first model represents the expected probabilities of capturing at least one specimen during each set. The second model estimates the expected mean catch rate conditional to the fact of having captured at least one specimen in the set. In this particular study, we assumed that the first model followed a binomial distribution (binary response variable with a logit link function), while in the second model the log-transformed catch rates (of the positive sets) were assumed to follow a normal distribution.

Other alternatives for dealing with zeros in the response variable are available to standardize CPUEs. An extensive revision on available methodologies for CPUE standardization was carried out by Maunder & Punt (2004). The Delta method has the particularity that it can be used when the response variable is continuous but has a mass of zeros, such as the case of our study where the CPUEs were calculated as biomass (kg) per effort (1000 hooks). Maunder & Punt (2004) also mentioned other possible approaches such as the Zero-Inflated Poisson (ZIP) and the Zero-Inflated Negative Binomial (ZINB), but those models can only be used when the response variable is discrete (e.g. count data). In such cases the effort could be used as an offset variable to the model.

The models and CPUEs standardizations presented in this study should be regarded as preliminary, as this collection of historical data is still being carried out by *IPIMAR*. The historical VMS data from the yearly 2000's is still being analyzed, and as the data becomes filtered and processed to be incorporated in the catch and effort databases we expect to present more complete models that include the complete time series for all years.

In terms of modeling, the Delta method and models chosen seem adequate for this analysis as verified by the residual analysis. However, future work will explore other alternatives for modeling continuous response variables with a mass of zeros, such as tweedie models (e.g. Candy 2004; Shono 2008). Additionally, we also expect to explore alternative modeling options that can account for the lack of independence between the samples, such as mixed models or generalized estimation equations.

4. Acknowledgments

Data was collected within the EU Data Collection Framework. VMS data was collected by the "Sistema integrado de informação e apoio à vigilância, fiscalização e controlo da actividade da pesca - DGPA/SIFICAP-MONICAP". Rui Coelho was partially supported with a grant from FCT (Ref: BDP 40523 / 2007) co-funded by "POCI-2010 - Programa Operacional Ciência e Inovação 2010" and "FSE - Fundo Social Europeu". Rui Coelho would like to acknowledge the EU Commission (DG MARE) for supporting his attendance at the Working Party on Ecosystem and By-catch meeting to present this paper.

5. References

Candy, S.G. 2004. Modeling catch and effort data using generalized linear models, the tweedie distribution, random vessel effects and random stratum-by-year effects. *CCAMLR Science*, 11: 59-80.

Cortés, E. 2009. Standardized catch rates for blue and mako sharks from the U.S. pelagic longline logbook and observer programs. *Collect. Vol. Sci. Pap. ICCAT*, 64 (5): 1595-1613.

Hilborn, R. & Walters, C.J. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman and Hall: New York. 570 p.

Maunder, M.N. & Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. *Fish. Res.*, 70: 141–159.

Mejuto, J., García-Cortés, B., Ramos-Cartelle, A. & De la Serna, J.M. 2009. Standardized catch rates for the blue shark (*Prionace glauca*) and shortfin mako (*Isurus oxyrinchus*) caught by the Spanish surface longline fleet in the Atlantic Ocean during the period 1990-2007. *Collect. Vol. Sci. Pap. ICCAT*, 64 (5): 1509-1521.

R Development Core Team. 2011. R: A language and environment for statistical computing. Version 2.13.0. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Shono, H. 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. *Fish. Res.*, 93: 154–162