SCRS/2012/054 Collect. Vol. Sci. Pap. ICCAT, 69(3): 1195-1212 (2013)

APPLICATION OF ZERO-INFLATED MODELS TO THE CATCH RATES OF WHITE MARLIN (*TETRAPTURUS ALBIDUS*) BASED ON DATA FROM THE SPANISH SURFACE LONGLINE FISHERY TARGETING SWORDFISH IN THE ATLANTIC OCEAN

J. Ortiz de Urbina[1], B. García-Cortés[2], A. Ramos-Cartelle[2] and J. Mejuto[2]

SUMMARY

*Zero-Inflated-Models were applied to 3,500 fishing trips and 22,184 records of landings of white marlin (*Tetrapturus albidus*) obtained by the Spanish surface longline fishery targeting swordfish during the 1993-2000 and 1988-2010 periods, respectively. The data confirm the low prevalence of this species in the catch of this fishery, with zero catch returns amounting to 80% and 95% for each data set, respectively. The explanatory variables in the final model were based on likelihood ratio tests between nested models and AIC scores. In spite of some data limitations, the results show that it was possible to obtain standardized catch rate indices for both data sets, although there was a considerable year-on-year variability in some cases, as well as several limitations on some of the modelled explanatory variables. In addition, this paper presents the annual values for the standardized indices and discusses the potential effects that the Recovery Plan for this species and the current legislation may have been having on catch rates in recent years.*

RÉSUMÉ

*Des modèles à inflation de zéros ont été appliqués à 3.500 sorties de pêche et 22.184 registres de débarquement de makaire blanc (*Tetrapturus albidus*) obtenus par la pêcherie palangrière de surface espagnole ciblant l'espadon entre les périodes 1993-2010 et 1988-2010, respectivement. Les données confirment la faible prévalence de cette espèce dans la prise de la pêcherie, les registres de captures nulles s'élevant à 80% et à 95% pour chaque jeu de données, respectivement. Les variables explicatives dans le modèle final se sont basées sur les tests du rapport des vraisemblances entre les modèles imbriqués et les scores AIC. En dépit de certaines limitations de données, les résultats font apparaître qu'il a été possible d'obtenir des indices standardisés des taux de capture pour les deux jeux de données, même s'il existait une variabilité interannuelle considérable dans certains cas, ainsi que plusieurs limitations affectant certaines des variables explicatives modélisées. En outre, ce document présente les valeurs annuelles pour les indices standardisés et discute des effets potentiels qu'auraient pu avoir le plan de rétablissement pour cette espèce et la législation actuelle sur les taux de capture au cours de ces dernières années.*

RESUMEN

*Se aplicaron modelos de ceros aumentados a 3.500 observaciones a bordo y a 22.184 registros de desembarcos de aguja blanca (*Tetrapturus albidus*) de la pesquería española de palangre de superficie dirigida al pez espada, durante 1993-2010 y 1988-2010, respectivamente. Los datos confirman la baja prevalencia de esta especie en las capturas de esta pesquería, con registros de captura nula en el 80% y 95% de los casos, según cada conjunto de datos, respectivamente. Las variables explicativas del modelo fueron obtenidas mediante AIC a partir de "likelihood ratio tests". Los resultados muestran que, pese a algunas limitaciones en los datos, fue posible obtener en ambos casos índices estandarizados de tasas de captura aunque con amplia variabilidad interanual en algunos casos y con limitaciones sobre algunas de las variables explicativas modelizadas. Se presentan además valores anuales de los índices estandarizados y se discute el posible efecto que el Plan de Recuperación de esta especie y las medidas de ordenación en vigor podría estar teniendo sobre las tasas de captura de los años más recientes.*

[1] Instituto Español de Oceanografía, P.O. Box 285, 29640 Fuengirola, Málaga. Spain
[2] Instituto Español de Oceanografía, P.O. Box 130, 15080 A Coruña, Spain. http://www.co.ieo.es/tunidos

KEYWORDS

Zero-inflated-models, white marlin, CPUE, surface longline

1. Introduction

White marlin (*Tetrapturus albidus)* (WHM) is an epipelagic istiophorid which can be observed in temperate waters during seasonal migrations, although it prefers warm waters in tropical and subtropical regions. This species -similarly to other billfish- has traditionally been targeted by various artisanal coastal fisheries in several countries and by a large number of different sport, recreational and chartering activities. Additionally, it is caught as low-prevalence by-catch by fisheries targeting tuna and tuna-like species with different fishing gears, such as driftnets, purse seine targeting tropical tunas (with or without FADs) and longlines (surface or deep). Generally, fleets do not carry out a taxonomic identification at the species level and this fish is often recorded together with other billfish species. Records for this species are inadequate or non-existent in many of the artisanal fleets and its prevalence is usually low in commercial landings. Therefore, scientific observations at sea help provide additional information of great interest.

Previous studies carried out on the Spanish surface longline fleet targeting swordfish estimated the prevalence of istiophorids in the catches and the relative importance of WHM in the landings as a whole (1%) and in the total bycatch (1.5%) (Castro *et al.* 2000, Mejuto *et al.* 2009). Despite the low prevalence of WHM in the Spanish surface longline fishery, results showed that approximately 1% of the WHM caught in the Atlantic were released alive, 6% were released and tagged and 11%-16% were dead specimens discarded during the period of that study, with discards showing a downward tendency over time (García-Cortés *et al. in press,* Mejuto *et al.* 2007).

The stock assessments performed by ICCAT in 2006 indicated that the WHM is overexploited and that its biomass will continue to decline unless the regulatory measures adopted by several different countries can reverse this trend (Anon. 2007). ICCAT recommendations focused on reducing commercial landings of longline and purse seine fisheries, as well as on encouraging the release of live specimens which come alongside the vessel in order to maximize their survival. In this sense, the Spanish fleet developed some trials by changing the configuration of surface longlines and including alternative hooks and baits. The results showed that WHM was the only species in which the hook-bait factor was found to be significant at the level established to explain the proportion of positive sets. The results also corroborated the findings of other authors regarding billfish, according to which, the area-time factor was significant and more important than the other variables tested (García-Cortés *et al.* 2010, Mejuto *et al*. 2010, Ortiz and Arocha 2004).

The Generalized Linear Modeling technique (GLM) (Robson 1966, Gavaris 1980, Kimura 1981) was used to estimate standardized catch rates based on data from commercial fleets which have an unbalanced spatial-temporal activity, as it was observed. This was due to the complex migratory behavior of the large pelagic species, usually related with environmental requirements (habitats) and their respective biological processes. The area-time distribution of the fleet and their fishing strategy over time (targeted species, gear configuration, etc.) - among other elements- also seem to be important factors to be considered for such assumption. Consistency in fishing areas over time facilitates this interpretation and increases the reliability of the CPUE information using simpler standardized models (Carruthers *et al.* 2010). However, there are other factors which should also be considered in order to assume such CPUE data as abundance indicators, particularly in the case of by-catch species with a very low prevalence and a high proportion of zero catch records in general. The yearly changes of the predicted CPUE indices should be plausible from a biological point of view.

This document aims to provide a preliminary analysis on the CPUE trends of WHM over time in the Atlantic Ocean within the framework of a research project. It also intends to discuss the applicability of such indices as abundance indicators using two different sets of information.

2. Material and methods

Two different data sets - supplied on a voluntary basis for the purposes of research- were used in the each of the analyses. Data set 1 was made up of records of WHM per fishing operation, which were obtained by scientific observations on board longline vessels targeting swordfish during the 1993-2010 period. Set *2* included records by trip or sub-trip for the 1988-2010 period. In both cases, the fishing effort was measured in thousands of hooks, either directly or by calculating the average number of hooks per haul and the total number of hauls. The nominal CPUE was estimated on the basis of the number of individuals identified as WHM in the catch or from the landings reported for the species. Eight different regions (R) were considered for the analyses (**Figure 1**), which replicates the spatial distribution proposed in previous studies on this species (García-Cortés *et al,* 2010). The variables "year" and "quarter" were used as time variables for the models. Other possible factors were also tested in the different tentative analyses.

Count data on the catch of non-target species are highly skewed and exhibit a high frequency of zero-valued observations (Ortiz *et al*. 2004, Minami *et al.* 2007). The WHM is a regular case of frequent zero catch by set or by trip in surface longline targeting swordfish, particularly in the case of fishing activities in temperate regions. "Zero-inflated models" is a model class capable of dealing with excess zero counts. In essence, they can be interpreted as a two-component mixture model combining a point mass at zero with a count distribution such as Poisson, geometric or negative binomial. The model divides the population into two groups: one group for which the outcome is always zero and one group for which the outcome is drawn from the underlying count data distribution. Each explanatory variable can have an effect on either or both the probability that an individual belongs to the "always zero group" and the magnitude of the count outcome, given that the individual belongs to the "not always zero group". Due to the characteristics of the data sets being analysed, a zero-inflated negative binomial (ZINB) was selected.

Model selection, *i.e.*, explanatory variables in the final model, was based on likelihood ratio tests between nested models and AIC scores. We followed both a forward selection approach (starting with a null model, testing the variables one by one and including them if statistically significant) and a backward elimination approach (starting with a model with all candidate variables and testing one by one for statistical significance, deleting any that was not significant). Standard errors and CVs were estimated using bootstrap techniques (1000 simulations).

Model validation was based on residual analysis. Pearson residuals were plotted against fitted values and against each explanatory variable in the final model. In addition, in order to detect an incorrect specification of the error distribution, "half-normal plots with simulated envelope" (Atkinson 1981, Collet 2003) were generated. The "half-normal quantile plot" could be considered as a diagnosis equalling to the "normal qq-plot" used for lineal models and which allows a graphic assessment of the goodness of fit of the model based on the residuals' behaviour (Pearson type). When the model fit is good, we can expect that the line representing the observed residuals tends to remain between the established confidence limits.

3. Results and discussion

3.1 Data set 1

A total of 3500 fishing sets were modelled for the 1993-2010 period (**Table 1**). The representativeness of the observations is limited due to the small annual coverage (199 observations/year on average), the low prevalence of this species with regard to the main ones, resulting in 80% of the observations not catching WHM at all (**Table 2**), and the fact that the coverage has generally been deployed in specific areas with the aim of obtaining fishing and biological data related to other species.

Table 3 shows the results of the AIC-based model selection. The only explanatory variables that we could use were year, quarter and region. The type of longline and other initially suggested variables, such as the type of gear, were not finally implemented, since the necessary time overlap was not attained, even when reducing the number of categories (levels).

Tables 4, 5 and 6 show the results for count model coefficients (negative binomial; log link), the zero inflation model (binomial; logit link), as well as the estimations obtained from the standardized CPUE index, its confidence intervals 95%, sample size and coefficients of variation. The limitations of this data set become evident when implementing the zero-inflated negative binomial (ZINB) model, even if it initially seemed the most suitable one for this type of data. Thus, the CV (%) values were extremely high, particularly for some of the years analyzed. The potential interactions among the explanatory variables were also tested, but it was not possible to obtain convergence under the ZINB model even when using an alternative delta-lognormal model.

Figure 2 shows the frequency distribution of number of WHM per set and frequency, distribution of number of WHM per fishing operation and for each explanatory variable. **Figure 3** shows the residuals pattern (Pearson) *vs*. fitted (scale response) and *vs*. each explanatory variable. The estimated standardized relative abundance index, its corresponding 95% confidence limits (bootstrap percentile method) and a *loess* fit of the predicted CPUE values are provided in **Figure 4**. The half-normal quantile plot with simulated envelope is provided in **Figure 5**.

3.2 Data set 2

In this case, 22184 fishing trips were modelled for the 1988-2010 period (**Table 1**). The representativeness of the trips is relatively larger (965 trips/year on average), as it includes a greater fishing effort and a better coverage of the commercial activities targeting swordfish by area and quarter, including many trips in temperate regions of both hemispheres. The data confirm the low prevalence of WHM in the commercial records, since 95% of the trips have consistently reported zero catch returns for this species (**Table 2**) every year. Some of these data stem from latitudes in which the presence (and hence, the catch) of WHM is rare or highly unlikely, and that is why the proportion of zero catch returns is higher in data set 2 than in data set 1, which covers more limited regions.

Table 7 shows the results of the model selection based on the likelihood ratio tests and the AIC values. As in data set 1 and due to the same limitations, the only explanatory variables which could be used were year, quarter and region. **Tables 8, 9 and 10** show the results for count model coefficients (negative binomial; log link), the zero inflation model (binomial; logit link), as well as the estimations obtained from the standardized CPUE index, its confidence intervals 95%, sample size and coefficients of variation. This data set has less limitation than the first one, as we can see when implementing the zero-inflated negative binomial (ZINB) model; the CV (%) values obtained are much more moderate and quite consistent in all years except for 2008.

Figure 6 shows the frequency distribution of number of WHM per set and frequency, distribution of number of WHM per fishing operation and for each explanatory variable. **Figure** 7 shows the residuals pattern (Pearson) *vs.* fitted (scale response) and *vs.* each explanatory variable. The estimated standardized relative abundance index, its corresponding 95% confidence limits (bootstrap percentile method) and a *loess fit* of the predicted CPUE values are provided in **Figure 8**.The half-normal quantile plot with simulated envelope is provided in **Figure 9**.

3.3 General comments

Generally, the standardized CPUE results suggest better statistical diagnoses when using data set 2. The volume of information used in each case, the different effort and space/time coverage, as well as other conditioning factors discussed above, account for the different results obtained for each data set. Data set 2 (landing per trip) tends to reduce the variability of the observed CPUE, particularly in these species with quite a low prevalence in this fishery with regard to the operational data, which usually have a smaller coverage due to several factors, including costs. On the contrary, the data per trip are more prone to a higher uncertainty when it comes to the taxonomic identification of the species, the quality of reports and/or the rates of retention of the catch on board, particularly after the implementation of the recent management measures.

Standardized CPUE predictions for the analyzed data sets suggest that increments of CPUE in number of fish between consecutive years ($CPUE_{yr+1}$ *vs.* $CPUE_{yr}$) were high in both data sets, but the magnitude was much higher in the case of the biannual increments of data set 1. The median biannual increment was around 38% and +4% in the case of the data set 2 when the absolute values of the increments or both the negative and positive increments are considered in the calculations, respectively. These values are higher than those reported for other large pelagic species with higher prevalence in the epipelagic system and also with a more frequent catch or by-catch in the surface longline targeting swordfish. A higher variability could be also expected in this case because the CPUE is provided in number of fish versus in biomass, as it was the case of other species reported.

In general, the trends of standardized average CPUE based on data set 2 seem plausible from the biological point of view when it comes to a species of this nature. Nonetheless, the CPUE increases considerably during the 1996-1998 periods, which should be specifically researched and compared with the increases observed by other authors-fleets. In any case, it must be taken into account that the series obtained from both sets could be affected by the recent management measures implemented as a result of the Rebuilding Plan put in place in 2007 (REC-2006-09) and the following years (REC-2010-05, 2011-07) aiming at reducing the landings of WHM and BUM and encouraging live releases, among other actions. In this sense, the trends observed for the 2007-2010 period are similar for both data sets and could be explained by the effects of the management and control measures implemented by CPs rather than by actual abundance drops.

4. Acknowledgments

The authors would like to deeply thank all the members of the team who were involved in the scientific recording of the data, supplied on a voluntary basis during a long period. Without the help of qualified people such as I. González,-González, M. Quintans, E. Alot, M. Marin, O. Soto, B. Costas, A. Carroceda, E. Majuelos, E. Ameal, J.L Torres. J. Fernández-Costa, J.A. Vila and F. González, this paper would not have been possible. We would also like to thank the scientific observers and especially all the skippers of the Spanish longline fleet involved in this scientific collaboration.

5. References

Anon. 2007, Report of the 2006 ICCAT Billfish Stock Assessment. Collect. Vol. Sci. Pap. ICCAT, 60(5): 1431-1546.

Atkinson, A.C. 1985, Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. New York. Oxford University Press.

Carruthers, T.R., Mcallister, M.K. and Ahrens, R.N.M.. 2010, Simulating spatial dynamics to evaluate methods of deriving abundance indices for tropical tunas. Can. J. Fish. Sci. 67:1409-1427.

Castro, J., De la Serna, J. M., Macías D. and Mejuto, J. 2000. Estimaciones científicas preliminares de los desembarcos de especies asociadas realizadas por la flota española de palangre de superficie en 1997 y 1998. Collect. Vol. Sci. Pap. ICCAT, 51(5):1882-1984.

Collet, D. 2003. Modelling binary data, 2nd ed. Chapman and Hall. CRC, New York.

García-Cortés, B., Fernández J., Ramos-Cartelle, A., and Mejuto, J. 2010, Prevalence of istiophorids (Fam. Istiophoridae) on the basis of observations of the Spanish surface longline fleet targeting swordfish (*Xiphias gladius*) in the Atlantic Ocean. Collect. Vol. Sci. Pap. ICCAT, 65(5): 1297-1823.

García-Cortés, B., Ramos-Cartelle, A. and Mejuto, J. 2012, Observations of the white marlin (*Tetrapturus albidus*) carried out on board of the Spanish surface longline fleet targeting swordfish from 1993-2010. Collect. Vol. Sci. Pap. ICCAT, 68(4): 1432-1445 (2012)

Gavaris, S. 1980, Use of a multiplicative model to estimate catch rate and effort from commercial data. Can. J. Fish. Aquat. Sci. 37: 2272-2275.

Kimura, D. K. 1981, Standardized measures of relative abundance based on modeling log (CPUE) and their application to Pacific Ocean. Perch. J. Cons. Int. Explor. Mer. 39: 211-218.

Mejuto, J., García-Cortés, B. and Ramos-Cartelle, A. 2007, Preliminary approach to evaluate the importance of discards and other uses of billfish in the Spanish surface longline fishery carried out in different oceans between 1993-2005. Collect. Vol. Sci. Pap. ICCAT, 60(5): 1547-1554.

Mejuto, J., García-Cortés, B., Ramos-Cartelle, A. and De la Serna, J. M. 2009, Scientific estimations of bycatch landed by the Spanish surface longline fleet targeting swordfish (*Xiphias gladius*) in the Atlantic Ocean with special reference to the years: 2005 and 2006 period. Collect. Vol. Sci. Pap. ICCAT, 64(7): 2455-2468.

Mejuto, J., Ortiz de Urbina, J., Ramos-Cartelle A. and García-Cortés, B. 2010, Changes in billfish catch rates due to the use of different hooks and baits in the configuration of the surface longline gear targeting swordfish (*Xiphias gladius*) in the Atlantic Ocean. Collect. Vol. Sci. Pap. ICCAT, 65(5):1778-1796.

Minami, M., Lennert-Cody, C. E., Gao, W. and Román-Verdesoto, M. 2007, Modeling shark bycatch: The zero-inflated negative binomial regression model with smoothing. Fisheries Research 84: 210-221.

Ortiz, M. and Arocha, F. 2004. Alternative error distribution models for standardization of catch rates of non-target species from a pelagic longline fishery: billfish species in the Venezuelan tuna longline fishery. Fisheries Research 70: 275–297.

Robson, D. S. 1966., Estimation of relative fishing power of individual ships. Res. Bull. Int. Comm. N.W. Atl. Fish, 3: 5-14.

Table 1. Number of observations by explanatory variable (data set 1: left; data set 2: right).

Data set 1		*number observations*	*Data set 2*		*number observations*
Year	1988	-	*Year*	1988	567
	1989	-		1989	647
	1990	-		1990	847
	1991	-		1991	890
	1992	-		1992	886
	1993	160		1993	1158
	1994	214		1994	1151
	1995	195		1995	1374
	1996	242		1996	1585
	1997	294		1997	1761
	1998	171		1998	1543
	1999	267		1999	1127
	2000	246		2000	897
	2001	270		2001	1084
	2002	163		2002	1083
	2003	200		2003	836
	2004	198		2004	807
	2005	164		2005	678
	2006	388		2006	631
	2007	82		2007	580
	2008	101		2008	633
	2009	53		2009	711
	2010	92		2010	708
Quarter	1	579	*Quarter*	1	5059
	2	1123		2	5032
	3	924		3	4988
	4	874		4	7105
Region	R1	455	***Region***	R1	11773
	R2	242		R2	2258
	R3	381		R3	1891
	R4	857		R4	1913
	R5	812		R5	2409
	R6	211		R6	683
	R7	102		R7	164
	R8	440		R8	1093

Table 2. Proportion of zeros by year. (data set 1: left; data set 2: right).

Data set 1		*Data set 2*	
Year	*prop. Zeros*	*Year*	*prop. Zeros*
1988	-	1988	90.48
1989	-	1989	95.05
1990	-	1990	97.28
1991	-	1991	94.04
1992	-	1992	96.28
1993	95.00	1993	96.20
1994	87.85	1994	94.79
1995	77.95	1995	94.54
1996	87.60	1996	91.42
1997	89.12	1997	89.44
1998	90.64	1998	90.86
1999	75.28	1999	95.39
2000	65.04	2000	97.32
2001	81.48	2001	95.20
2002	98.16	2002	99.45
2003	83.00	2003	96.29
2004	59.60	2004	94.92
2005	46.95	2005	93.07
2006	80.15	2006	94.29
2007	78.05	2007	92.76
2008	66.34	2008	97.31
2009	90.57	2009	98.73
2010	85.87	2010	98.45

Table 3. Model selection. (Data set 1; π_i, zero-inflation model; μ_i, count model).

Dropped term	*df*	*AIC*	*Likelihood ratio test*
none	57	4377.337	
YEAR from π_i and μ_i			Does not converge
YEAR from π_i	40	4497.752	χ^2 = 154.42 (df = 17; p < 2.2E-16)
YEAR from μ_i	40	4570.825	χ^2 = 227.49 (df = 17; p < 2.2E-16)
QUARTER from π_i and μ_i	51	4440.719	χ^2 = 75.38 (df = 6; p < 3.203E-14)
QUARTER from π_i	54	4427.824	χ^2 = 56.49 (df = 3; p < 3.308E-12)
QUARTER from μ_i	54	4380.723	χ^2 = 9.39 (df = 3; p = 0.02458
REGION from π_i and μ_i			Does not converge
REGION from π_i	50	4599.911	χ^2 = 236.57 (df = 7; p < 2.2E-16
REGION from μ_i			Does not converge

Table 4. Count model coefficients (negative binomial; log link). Data set 1.

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	-1.6620	0.4690	-3.544	3.940E-04
YEAR1994	1.6206	0.4169	3.888	1.010E-04
YEAR1995	1.7665	0.4073	4.337	1.440E-05
YEAR1996	1.0696	0.4209	2.541	1.104E-02
YEAR1997	0.9672	0.4436	2.180	2.923E-02
YEAR1998	-0.5837	0.4739	-1.232	2.181E-01
YEAR1999	2.5482	0.4266	5.974	2.320E-09
YEAR2000	1.4970	0.4279	3.498	4.680E-04
YEAR2001	2.8113	0.4483	6.271	3.580E-10
YEAR2002	-1.1509	0.7370	-1.562	1.184E-01
YEAR2003	2.5623	0.4639	5.523	3.340E-08
YEAR2004	2.8387	0.3947	7.193	6.350E-13
YEAR2005	2.3834	0.3954	6.028	1.660E-09
YEAR2006	1.8146	0.4201	4.320	1.560E-05
YEAR2007	1.3505	0.4759	2.838	4.544E-03
YEAR2008	1.9139	0.4569	4.189	2.800E-05
YEAR2009	0.5761	0.6437	0.895	3.709E-01
YEAR2010	0.6204	0.5536	1.121	2.625E-01
REGIONR2	-0.1553	0.3831	-0.405	6.852E-01
REGIONR3	-2.0560	0.2659	-7.731	1.070E-14
REGIONR4	-1.9513	0.3722	-5.243	1.580E-07
REGIONR5	-0.6727	0.2013	-3.342	8.320E-04
REGIONR6	-1.2216	0.2222	-5.497	3.870E-08
REGIONR7	-0.8976	0.2886	-3.110	1.872E-03
REGIONR8	-1.1597	0.2374	-4.885	1.030E-06
QUARTER2	-0.0783	0.1578	-0.496	6.198E-01
QUARTER3	0.4091	0.1531	2.672	7.530E-03
QUARTER4	0.0652	0.1324	0.492	6.226E-01
Log(theta)	0.1988	0.1152	1.726	8.435E-02

Table 5. Zero inflation model (binomial; logit link). Data set 1.

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	3.8040	2.1590	1.762	7.804E-02
YEAR1994	1.8770	1.7790	1.055	2.913E-01
YEAR1995	2.0200	1.8910	1.068	2.855E-01
YEAR1996	0.7899	1.8090	0.437	6.623E-01
YEAR1997	3.1530	1.8380	1.715	8.629E-02
YEAR1998	-21.3100	683.9000	-0.031	9.751E-01
YEAR1999	2.2660	1.7800	1.273	2.030E-01
YEAR2000	-28.5600	29340.0000	-0.001	9.992E-01
YEAR2001	6.0650	1.9470	3.115	1.840E-03
YEAR2002	-18.8900	6126.0000	-0.003	9.975E-01
YEAR2003	2.9550	1.9560	1.511	1.308E-01
YEAR2004	2.5020	1.8320	1.366	1.721E-01
YEAR2005	0.7256	2.1770	0.333	7.390E-01
YEAR2006	1.3790	1.8230	0.757	4.492E-01
YEAR2007	-3.8040	2.3170	-1.642	1.006E-01
YEAR2008	-0.3522	2.3500	-0.150	8.809E-01
YEAR2009	2.6510	2.3570	1.125	2.608E-01
YEAR2010	0.1025	5.0060	0.020	9.837E-01
REGIONR2	-4.5980	1.3450	-3.419	6.290E-04
REGIONR3	-10.7500	1.6630	-6.466	1.010E-10
REGIONR4	-4.1090	1.3780	-2.981	2.873E-03
REGIONR5	-6.9050	1.2270	-5.627	1.830E-08
REGIONR6	-20.3300	NA	NA	NA
REGIONR7	-40.3400	312300.0000	0.000	9.999E-01
REGIONR8	-12.0900	1.7820	-6.785	1.160E-11
QUARTER2	2.6990	0.5606	4.815	1.470E-06
QUARTER3	2.9260	0.8944	3.271	1.072E-03
QUARTER4	-0.8650	0.7648	-1.131	2.580E-01

Table 6. Estimated relative abundance index, standard error, CV(%) and confidence intervals (95%) of the Spanish longline. Period: 1993- 2010. Data set 1.

Year	*Index*	*Std error*	*n*	*CV (%)*	*95%LL*	*95%UL*
1993	0.0041	0.0216	160	521.6307	5.812E-14	7.762E-02
1994	0.0033	0.0363	214	1112.9931	7.056E-13	1.113E-01
1995	0.0033	0.0423	195	1293.4393	3.699E-13	9.578E-02
1996	0.0055	0.2965	242	5356.0345	1.037E-11	1.038E+00
1997	0.0005	0.0132	294	2771.4772	3.253E-13	4.653E-02
1998	0.1058	0.1348	171	127.3516	1.538E-03	4.611E-01
1999	0.0056	0.0559	267	1000.2293	1.906E-12	1.610E-01
2000	0.8479	0.4419	246	52.1225	1.139E-01	1.703E+00
2001	0.0002	0.0500	270	30600.8123	2.102E-15	1.868E-01
2002	0.0600	0.0894	163	148.9902	1.005E-16	2.761E-01
2003	0.0029	0.0947	200	3323.1064	7.191E-11	3.412E-01
2004	0.0059	0.0579	198	980.1159	7.511E-12	2.066E-01
2005	0.0219	0.0909	164	414.3391	6.302E-10	2.790E-01
2006	0.0065	0.0308	388	474.2947	6.688E-12	1.026E-01
2007	0.3660	0.1642	82	44.8641	9.795E-10	6.229E-01
2008	0.0395	0.1658	101	419.6564	2.613E-05	4.980E-01
2009	0.0005	0.1202	53	22681.2350	6.935E-14	3.215E-01
2010	0.0070	0.2058	92	2959.1898	1.973E-11	6.560E-01

Table 7. Model selection. (Data set 2; π_i, zero-inflation model; μ_i, count model).

Dropped term	*df*	*AIC*	*Likelihood ratio test*
none	67	14477.64	
YEAR from π_i and μ_i	23	14918.46	$\chi^2 = 528.82$ (df = 44; p < 2.2E-16)
YEAR from π_i	45	14587.70	$\chi^2 = 154.06$ (df = 22; p < 2.2E-16)
YEAR from μ_i	45	14666.08	$\chi^2 = 232.44$ (df = 22; p < 2.2E-16)
QUARTER from π_i and μ_i	61	15097.98	$\chi^2 = 632.34$ (df = 6; p < 2.2E-16)
QUARTER from π_i	64	14504.43	$\chi^2 = 32.79$ (df = 3; p < 3.571E-02)
QUARTER from μ_i	64	14687.17	$\chi^2 = 215.53$ (df = 3; p < 2.2E-16)
REGION from π_i and μ_i	53	15053.12	$\chi^2 = 603.48$ (df = 14; p < 2.2E-16)
REGION from π_i			Does not converge
REGION from μ_i	60	14696.06	$\chi^2 = 232.42$ (df = 7; p < 2.2E-16)

Table 8. Count model coefficients (negative binomial; log link). (Data set 2).

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	-3.2441	0.3156	-10.281	2.000E-16
YEAR1989	0.5521	0.3765	1.466	1.425E-01
YEAR1990	0.1470	0.4215	0.349	7.272E-01
YEAR1991	-0.9711	0.3205	-3.030	2.447E-03
YEAR1992	-1.9624	0.3876	-5.063	4.120E-07
YEAR1993	-1.9211	0.3496	-5.495	3.910E-08
YEAR1994	-2.0845	0.3225	-6.463	1.030E-10
YEAR1995	-1.2344	0.2927	-4.217	2.470E-05
YEAR1996	-0.3884	0.2690	-1.444	1.489E-01
YEAR1997	-0.2754	0.2539	-1.085	2.781E-01
YEAR1998	0.2975	0.2631	1.131	2.582E-01
YEAR1999	-1.2917	0.3259	-3.964	7.380E-05
YEAR2000	-0.7295	0.3798	-1.921	5.474E-02
YEAR2001	-1.0066	0.3279	-3.070	2.143E-03
YEAR2002	-2.6186	0.9772	-2.680	7.367E-03
YEAR2003	-0.5896	0.3742	-1.576	1.151E-01
YEAR2004	-1.5215	0.3652	-4.166	3.100E-05
YEAR2005	-2.0182	0.3418	-5.905	3.520E-09
YEAR2006	-1.8884	0.3883	-4.864	1.150E-06
YEAR2007	-1.7966	0.3412	-5.266	1.400E-07
YEAR2008	-1.0895	0.5957	-1.829	6.742E-02
YEAR2009	-3.3550	0.6819	-4.920	8.650E-07
YEAR2010	-1.9743	0.5965	-3.310	9.330E-04
REGIONR2	-0.0612	0.1647	-0.372	7.101E-01
REGIONR3	0.5185	0.2620	1.979	4.781E-02
REGIONR4	-0.7094	0.3803	-1.866	6.211E-02
REGIONR5	2.7383	0.3181	8.609	2.000E-16
REGIONR6	1.0663	0.6292	1.695	9.012E-02
REGIONR7	3.3498	0.9776	3.426	6.120E-04
REGIONR8	2.7140	0.4138	6.558	5.450E-11
QUARTER2	0.8890	0.2229	3.989	6.630E-05
QUARTER3	3.0544	0.2324	13.144	2.000E-16
QUARTER4	1.7235	0.2229	7.734	1.040E-14
Log(theta)	-1.0992	0.1328	-8.278	2.000E-16

Table 9. Zero inflation model (binomial; logit link). (Data set 2).

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	2.2702	0.2492	9.111	2.000E-16
YEAR1989	0.6154	0.2605	2.363	1.814E-02
YEAR1990	1.1146	0.2852	3.908	9.300E-05
YEAR1991	0.0246	0.2430	0.101	9.194E-01
YEAR1992	0.2011	0.2969	0.678	4.980E-01
YEAR1993	0.1007	0.2680	0.376	7.071E-01
YEAR1994	-0.3553	0.2589	-1.373	1.699E-01
YEAR1995	-0.1023	0.2267	-0.451	6.518E-01
YEAR1996	-0.3590	0.2025	-1.773	7.622E-02
YEAR1997	-0.5604	0.1948	-2.877	4.011E-03
YEAR1998	-0.2307	0.1968	-1.172	2.410E-01
YEAR1999	0.1464	0.2405	0.609	5.426E-01
YEAR2000	0.7589	0.2885	2.631	8.516E-03
YEAR2001	-0.1535	0.2502	-0.613	5.396E-01
YEAR2002	1.6145	0.6504	2.482	1.305E-02
YEAR2003	0.4026	0.2730	1.475	1.403E-01
YEAR2004	-0.3147	0.2841	-1.107	2.681E-01
YEAR2005	-1.0188	0.3056	-3.334	8.550E-04
YEAR2006	-0.7390	0.3094	-2.389	1.692E-02
YEAR2007	-0.9842	0.2946	-3.341	8.350E-04
YEAR2008	0.4677	0.3604	1.298	1.944E-01
YEAR2009	0.4527	0.5259	0.861	3.893E-01
YEAR2010	0.8307	0.4260	1.950	5.116E-02
REGIONR2	-0.7228	0.1116	-6.479	9.210E-11
REGIONR3	0.9188	0.1635	5.619	1.920E-08
REGIONR4	2.1990	0.2554	8.611	2.000E-16
REGIONR5	1.5402	0.1679	9.171	2.000E-16
REGIONR6	1.1551	0.3311	3.489	4.850E-04
REGIONR7	1.7618	0.6083	2.896	3.774E-03
REGIONR8	1.8376	0.2364	7.775	7.550E-15
QUARTER2	-0.2615	0.1750	-1.494	1.352E-01
QUARTER3	-0.8595	0.1691	-5.083	3.710E-07
QUARTER4	-0.5212	0.1672	-3.118	1.820E-03

Table 10. Estimated relative abundance index, standard error, CV(%) and confidence intervals (95%) of the Spanish longline. Period: 1988- 2010. Data set 2.

Year	*Index*	*Std error*	*n*	*CV (%)*	*95%LL*	*95%UL*
1988	0.120798	0.039177	567	32.431860	0.038261	0.187237
1989	0.118485	0.043279	647	36.526840	0.037862	0.202822
1990	0.048987	0.017484	847	35.691650	0.019731	0.083739
1991	0.044731	0.013736	890	30.706900	0.015459	0.067037
1992	0.014122	0.005590	886	39.581250	0.002862	0.023587
1993	0.016141	0.005537	1158	34.305240	0.004053	0.024961
1994	0.020610	0.005848	1151	28.374880	0.007156	0.029505
1995	0.038552	0.010293	1374	26.699290	0.014767	0.054759
1996	0.112739	0.028917	1585	25.649620	0.047566	0.149573
1997	0.150086	0.033738	1761	22.479190	0.072436	0.199939
1998	0.199993	0.042010	1543	21.005830	0.103655	0.259811
1999	0.029046	0.009943	1127	34.232900	0.009247	0.044293
2000	0.028697	0.010349	897	36.063130	0.010386	0.050148
2001	0.050685	0.016601	1084	32.754200	0.016604	0.079946
2002	0.001895	0.001279	1083	67.530160	0.000204	0.003481
2003	0.046219	0.017954	836	38.845500	0.016635	0.085488
2004	0.034928	0.010208	807	29.225960	0.010494	0.047859
2005	0.038144	0.010616	678	27.831090	0.013355	0.051378
2006	0.034718	0.012610	631	36.321980	0.010901	0.058897
2007	0.046339	0.012678	580	27.359990	0.019332	0.068961
2008	0.026377	0.139423	633	528.574120	0.005399	0.422403
2009	0.002776	0.001128	711	40.639260	0.000928	0.005370
2010	0.007716	0.003173	708	41.123490	0.001832	0.013903

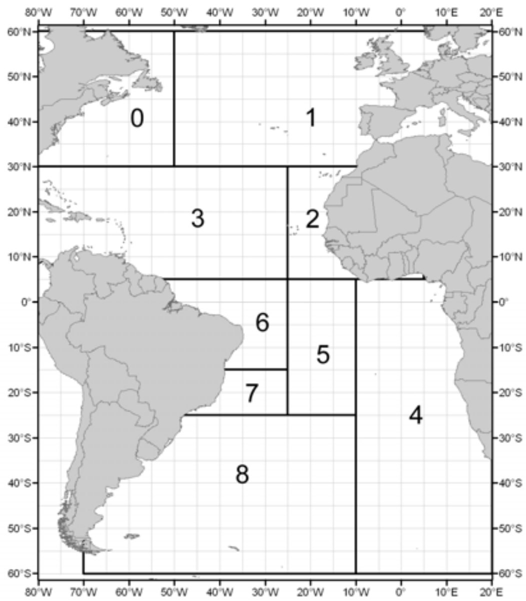

Figure 1. Stratification of geographic areas used for the analysis of white marlin in the Atlantic Ocean (from García-Cortés *et al.* 2010).

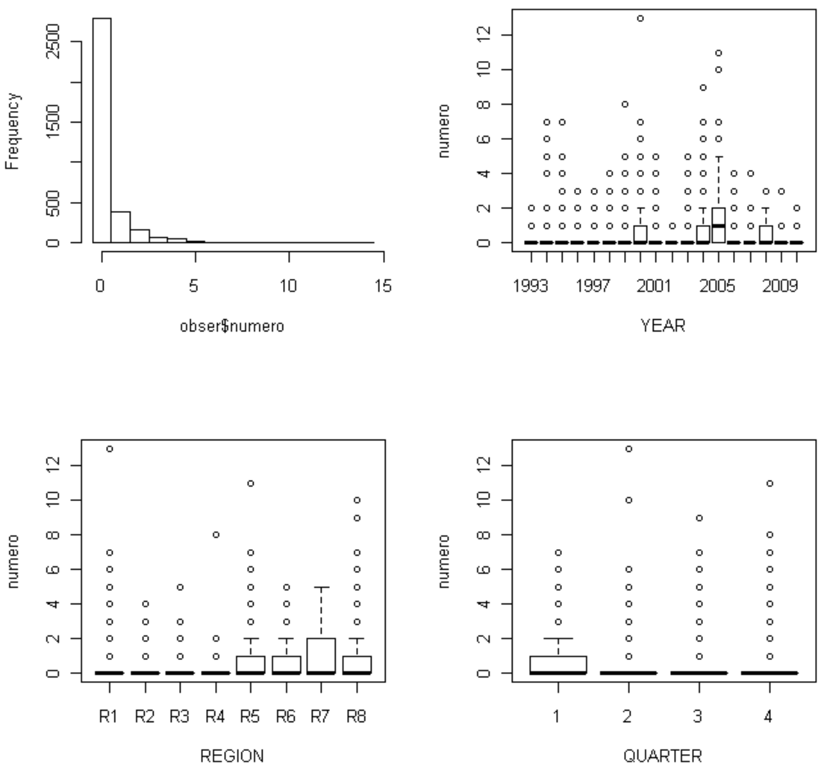

Figure 2. Frequency distribution of number of WHM per fishing set (upper left panel) and frequency distribution of number of WHM per fishing set for each explanatory variable. Data set 1.

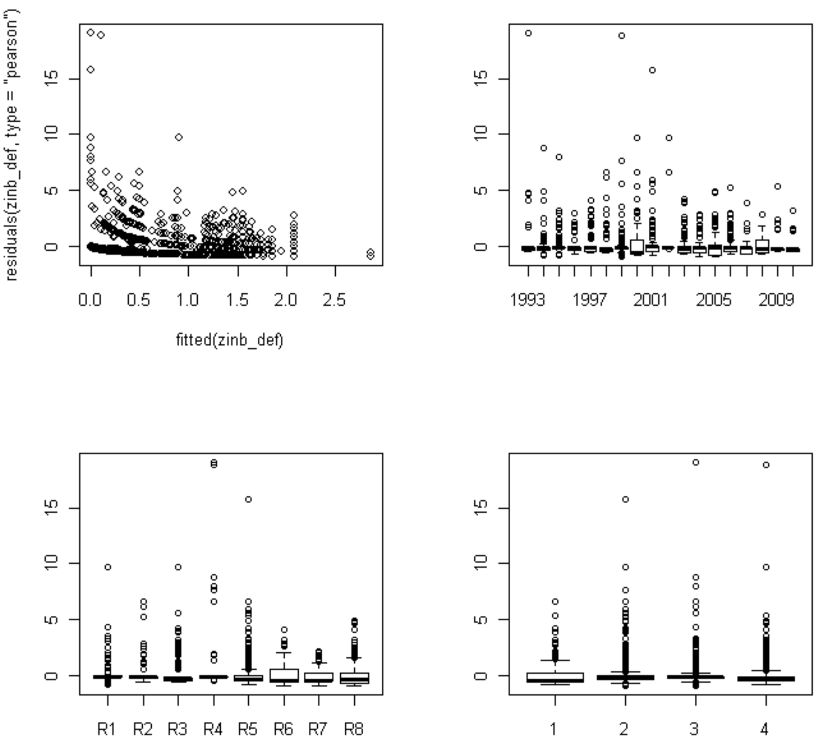

Figure 3. Residuals (Pearson). (upper left panel, residuals *vs* fitted (scale response); upper right panel, residuals *vs* explanatory variable YEAR; lower left panel, residuals *vs* explanatory variable REGION; lower right panel, residuals *vs* explanatory variable QUARTER. Data set 1.

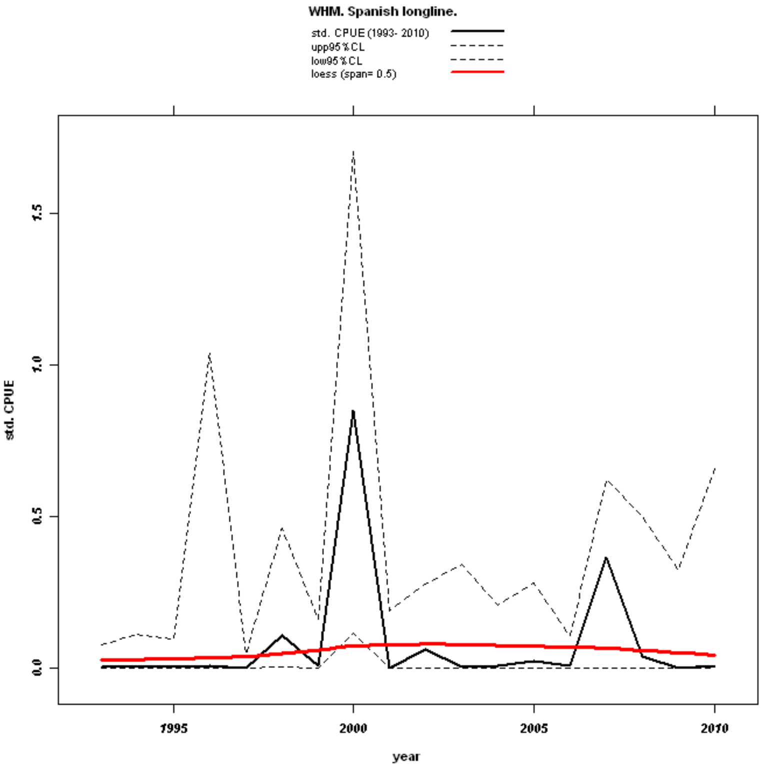

Figure 4. Estimated standardized relative abundance index, corresponding 95% confidence limits (bootstrap percentile method) and loess fit. WHM. Spanish longline. 1993- 2010. Data set 1.

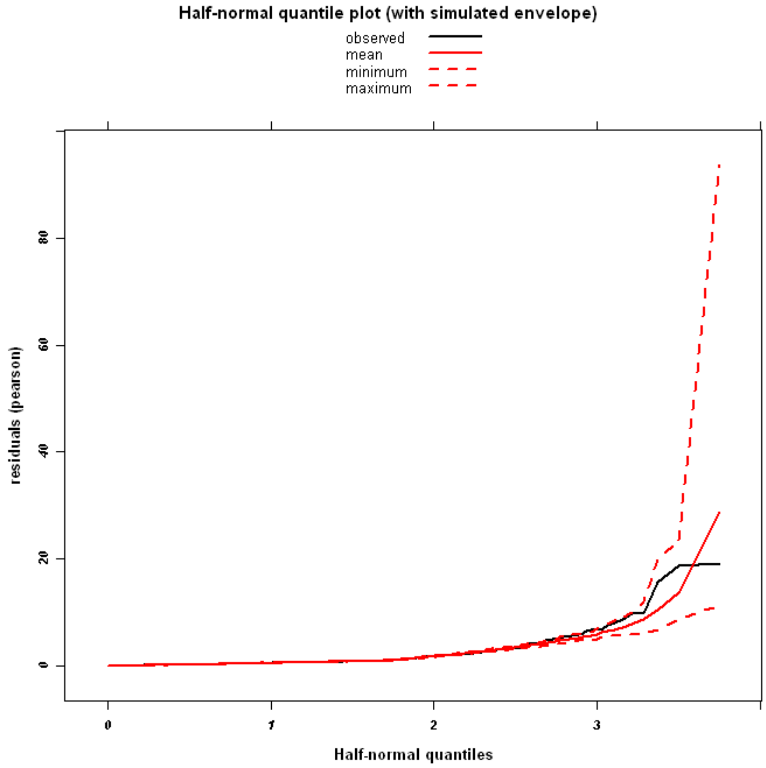

Figure 5. Half-normal quantile plot with simulated envelope. WHM. Spanish longline. 1993- 2010. Data set 1.

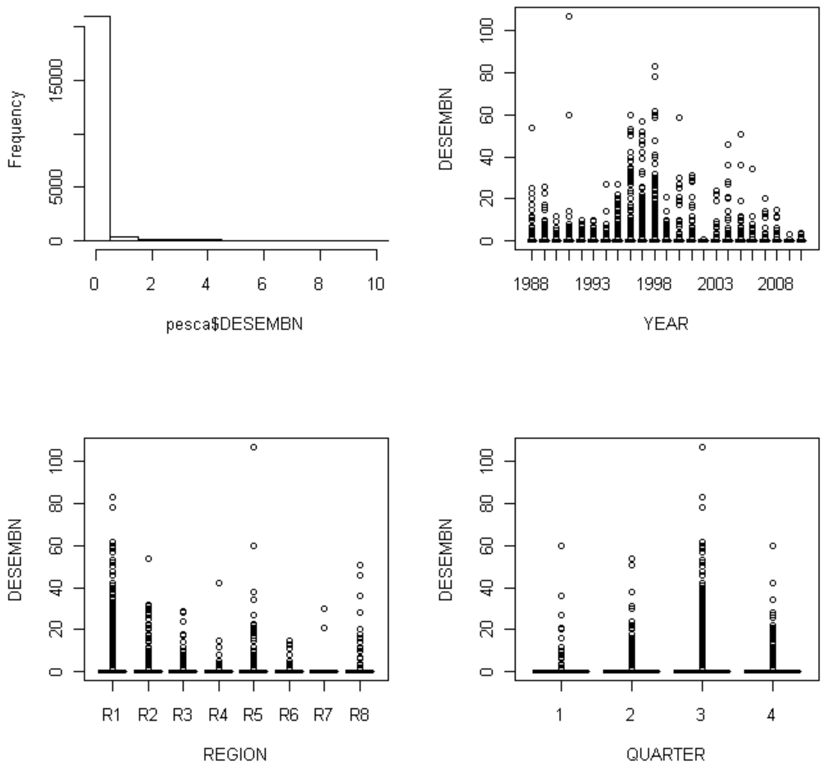

Figure 6. Frequency distribution of number of WHM per trip (upper left panel) and frequency distribution of number of WHM per trip for each explanatory variable. Data set 2.

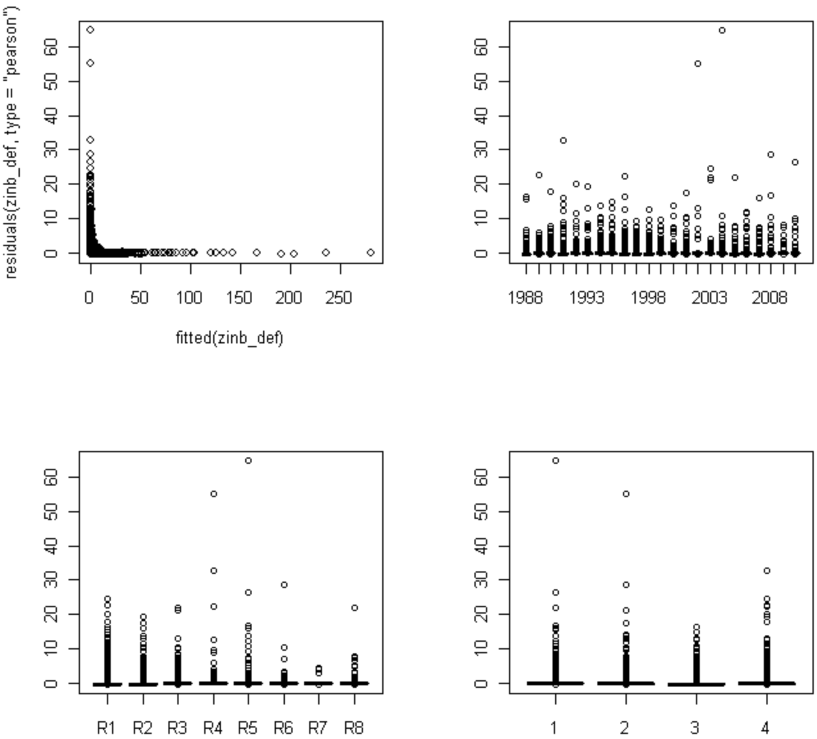

Figure 7. Residuals (Pearson). (upper left panel, residuals *vs.* fitted; upper right panel, residuals *vs.* explanatory variable YEAR; lower left panel, residuals *v.s* explanatory variable REGION; lower right panel, residuals *vs.* explanatory variable QUARTER. Data set 2.

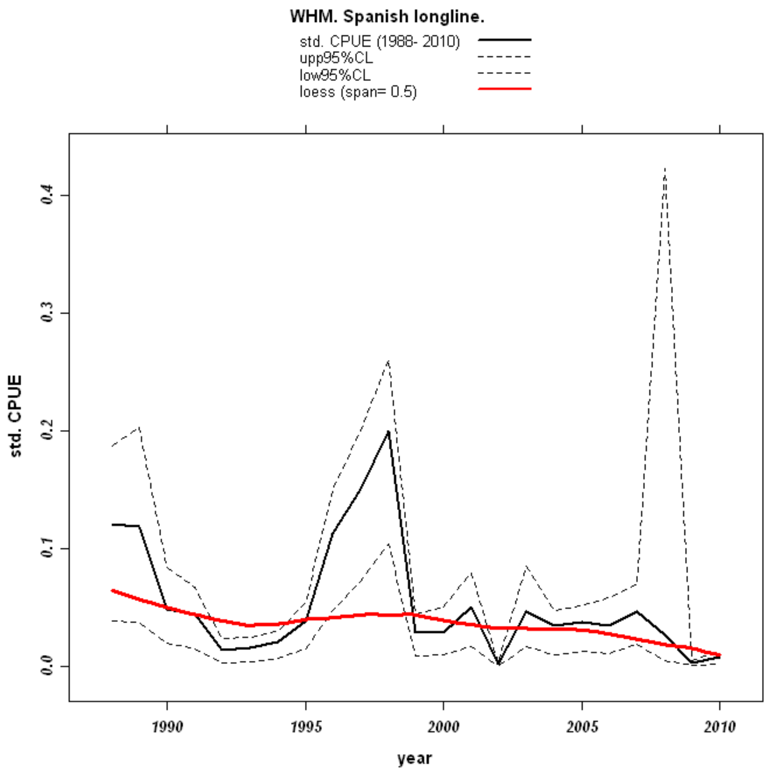

Figure 8. Estimated standardized relative abundance index, corresponding 95% confidence limits (bootstrap percentile method) and loess fit. WHM. Spanish longline. 1988- 2010. Data set 2.

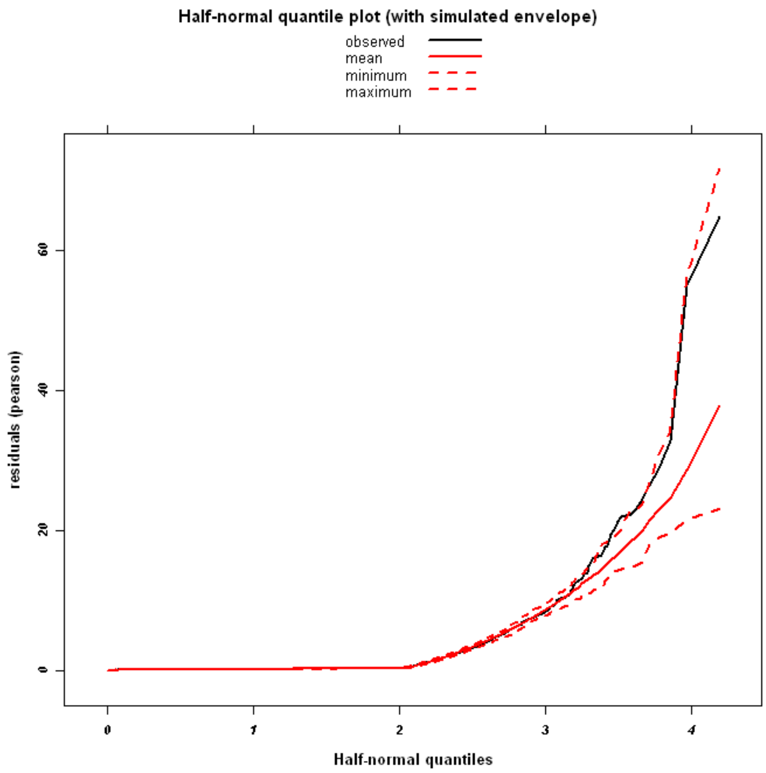

Figure 9. Half-normal quantile plot with simulated envelope. WHM. Spanish longline. 1988- 2010. Data set 2.