Standardized catch rates for yellowfin (*Thunnus albacares*) for the European purse seine fleet (1982-2003).

by

Soto M[1]., D. Gaertner[2], J. Dorizo[3], A. Fonteneau[2] , P. Pallarés[1], A. Delgado de Molina[4] and J. Ariz[4]

Abstract

Standardized yellowfin catch rates of the European purse seine fleets using GLM method are presented. Different indexes based in the yellowfin free school catch have been standardized considering time-space strata, as well as catch on FADs as explicative variables.

1. Introduction

In the Indian Ocean, yellowfin is mainly caught by purse seiners. However, relative abundance indexes based on the CPUEs of this fishery have been difficult to obtain due to problems related with changes in catchability of the fleet, as well as the variable use of different fishing modes with different target species, and the increasing fishing power of this fleet (Gaertner&Pallares, 2002).

Nevertheless, the goal of this paper has been to estimate series of CPUEs calculated with various GLM models, in order to estimate biomass trends and variability of adult yellowfin (*Thunnus albacares*) caught on free schools by the PS fishery in the Indian ocean, during the period 1982-2003.

2. Data used

We have used catch and effort data for the main purse seine fleets (France, Spain, European associated and Seychelles), operating in the Indian Ocean for the period 1982-2003 as these fleets have been permanently active in the area and permanently targeting free schools of large yellowfin. Purse seine catch and effort data used in the present GLM analysis were obtained from two different sources:

1) **Catch/Effort IOTC data:** The first data set was the catch and effort data by 1° squares and flags submitted to the IOTC by various fishing countries. Effort was expressed in fishing days without standardization of vessel categories. These data have been cumulated by quarter and by the 10 fishing areas (figure 1) used to analyse the Purse seine data in the area in order to be used in the GLM. All previous exploratory analysis of these areas have been concluding that they are convenient ones to analyze the purse seine data.

2) **Logbook/Vessel data**: The other data set used in the present analysis was obtained from each individual purse seiner on the log book data. The basic data in this analysis was for each boat its yearly catch, its fishing and searching time (e.g. fishing time minus set duration; a parameter that is estimated to be more interesting than fishing time), the age of each vessel each year, its category of size (measured by its carrying capacity), its flag and the estimated distance covered yearly by each boat during its searching time. Distance fished was

[1] IEO. Corazón de María 8, 28002 Madrid Spain.
2 IRD, CRH, Bp 171, 34203 Sète cedex,France
[3] SFA, Seychelles
[4] IEO. PO Box 1373. S/C Tenerife. Spain

calculated as the straight linear distances between the various positions in the log book; the comparison between these log books data and observer or VMS data tend to indicate that these log book distances tend to be underestimated, but at a stable rate of about 30% (unpublished). These distances have been introduced in the GLM as there was a clear indication in the data base that the distances prospected have been permanently increasing during the period. In this data set, only the sub set of catch and effort data in the area south of the Equator was kept and used in the GLM; this technical choice was done in order to eliminate the areas north of the Equator where an effort targeting FADs has been permanently increasing since the early seventies (when the effort south of Equator have been predominantly targeting free schools of yellowfin). The potential advantage of these log book data is that they could partly take into account some changes in the fishing power of purse seine fleets, such as changes in size categories of vessels, their distances prospected, their duration of sets and the increasing age of the purse seine fleet. These detailed factors, such as distance fished, size of the vessels, and their age, have been introduced in the model because of their potential importance and major trends in the IO PS fishery. These trends are well shown by figure 2 to 4.

In both data sets, only the catches on free schools, e.g. the component of the purse seine fisheries targeting yellowfin have been kept and used in the analysis.

3. Methods

3.1. Overall

The GLM method (McCullagh and Nelder, 1989) was used on the 2 data sets in order to evaluate changes in CPUEs that could be representative of abundance fluctuation and trends of the biomass of adult yellowfin.

We have considered different GLM models.

3.2. Logbook/vessel models

To explain the variability of CPUE data, the explicative variables used were the following: year, number of explored squares per day, proportion of days without catches, number of sets per day, mean catch per set, proportion of sets on fads, vessel age and the interaction between vessel category and the country.

The lognormal model have been used in the standardization process. Taking into account the high aggregation level (boat-year), non 0 values of CPUE have been found.

Two alternative definitions of CPUE were considered as a response variable, producing two models with the same explicative variables as mentioned before:

Case 1.

The response variable was defined as:

CPUE=YFT free school catch/ searching time

Case 2

The response variable was defined as:

CPUE=YFT free school catch/ travelled distance

Both response variables were transformed to achieve Normality in the GLM model. Several alternatives for the parameter k in the transformation *y*=log(*x*+k) were investigated, by inspecting histograms and normality Kolmogov-Smirnov tests. Both variables were finally transformed by *y*=log(*x*+0.1).

The model to fit the CPUEs in the case 1 and case 2 is expressed as:

$$\text{Ln}(CPUE+0.1) = year + Nsd + Dnc + Sd + Cs + Sf + Age + Catpais + \varepsilon$$

where, *Nsd* is the number of explored squares per day, *Dnc* is the proportion of days without catches, *Sd* is the number of sets per day, *Cs* is the mean catch per set, *Sf* is the proportion of sets on fads, *Age* is the age of the boat, *Catpais* is a combination of vessel category and country, and ε is the independent error vector identically distributed that follows a $N(0, \sigma^2)$.

The interaction between vessel category and country, *Catpais*, was incorporated in the model as a unique variables coded as:

Catpais Code	country-vessel category
1	France - <1500 m^3
2	France - 1500-2000 m^3
3	France - >2000 m^3
4	Spain - <1500 m^3
5	Spain - 1500-2000 m^3
6	Spain - >2000 m^3

Relative CPUE indices were calculated as the coefficients of the year component, obtained from the contrast treatment option to model the coefficients matrix, that estimates all the values relative to the first year.

3.3. Catch/Effort model: IOTC data

To standardize this data a Delta-Lognormal model (case 3) was used.

The Delta approach (Lo et al, 1992, Soto et al. 2002) splits the model in two components: the probability that cpue was bigger than zero, P{cpue>0}, and the distribution of the positive values of cpue. Both could be modelled independently to obtain, first, an adjustment of the positive cpue probability, and second, the expected cpue conditioned to obtain a cpue value bigger than zero. Then, the Delta-Lognormal method comprises two lineal generalized models using the Binomial and Lognormal distributions, respectively.

The relative CPUE for every year is:

$$CPUE = \mu \cdot p$$

where μ is the standardized cpue for the positive catches, calculated as the coefficients of the year component of the Gaussian model, and p is the standardized proportion of positives, obtained as the year coefficients of the binomial model.

CPUE variable was defined in the Delta model as:

CPUE = catches YFT on free school/fishing days.

For both components of the Delta model (i.e. binomial and lognormal) the explicative variables were: country, area, season and proportion of catches on fads over the total catch.

Country is a categorical variable coded as:

	Country
1	France
2	Seychelles
3	NEI
4	Spain

The zones selected for the analysis are shown in Figure 1.

Each year was divided in four quarters with the following codes:

Season Code	Months
1	January-February
2	March-April-May-June
3	July-August-September-October
4	November-December

Positive CPUE:

The model to adjust positive CPUEs is:

Ln(*CPUE*) = *year* + *zone* + *country* + *proportion of catches on fads* + *fishing season*

For the Gaussian glm to model positives values of CPUE, the proportions of fads was considered as a numerical variable and denoted by *Pfads*.

Proportion of positive CPUE

On the other side, to estimate the proportion of positive CPUEs, a Bernouilli random variable was created with value 0 or 1, depending if the CPUE was nil or positive, respectively. Then, the average of this variable is calculated in each strata defined by each year, season, country, proportion of fads, season and area combination, and the number of observations existing in every strata are calculated to use them as weights in the binomial model.

The probability of CPUE to be positive, is modelled through a Binomial GLM with the logit function as a connection between the explicative variables and the response variable, i.e., the appearance of positive CPUE in each strata is a Binomial random variable with a probability *p* given by the model:

Log(*p/(1-p)*) = *year* + *zone* + *country* + *proportion of catches on fads* + *fishing season*.

In order to apply a Delta-Lognormal method, the proportion of catches on fads was categorized only in two groups for the binomial glm, to avoid unbalance in the analysis. It is denoted by *Pfadc*:

Proportion FADs Code	Proportion FADs
1	[0,0.5]
2	(0.5,1]

Automatic forward and backward selection of variables (Draper *et al*., 1986) was applied to all the models, through the function step.glm of SPLUS, that selects variables based on the value of the Cp statistic.

Variances for CPUE indices were obtained from the analytical variances of the coefficients in the Gaussian models, and from the Delta approximation (Miller, 1986) in the Delta-Lognormal approach.

4. Results

4.1. Logbook/Vessel models

Function step.glm was applied to cases 1 and 2 and no explicative variables were removed by the Cp criteria. However, F statistics in ANOVA table of both models (Tables 1 and 2), show no significance for the *Age* and *Catpais* variables. It was decided to remove them of the models by the simplicity principle.

Figure 5 shows the CPUE index fitted by the GLM for case 1 (CPUE=YFT free school catch/ searching time), with limits defined as the CPUE±Standard deviation CPUE. Diagnostic plots for case 1 are shown in Figure 7. The model fits well, there are no patterns in the residuals. Normal probability plot shows some discrepancy from normal distribution for some extreme observations, due to the proximity to 0 values of some observations.

Nominal CPUE and fitted CPUE of case 1 are compared in Figure 3, and for case 2 (CPUE=YFT free school catch/ travelled distance) in Figure 9. Both fitted CPUEs are compared in Figure 8. Case 1 was selected as the best fit for the CPUE, because the trend in CPUE from case 2 since 1990 is extremely increasing.

4.2. Catch/Effort model

Tables 3 and 4 show ANOVA tables for binomial and Gaussian components of the Delta model applied to task II data. Figure 10 shows the corresponding CPUE index obtained multiplying year coefficients from the two components compares with nominal CPUE. Displacement in trends since year 1991 could be due to the change of data treatment in that period.

Country variable is not significant in the binomial component, i.e. there is no statistically differences between the Spanish and French purse seine fleets to found and set successfully school of tuna.

However, in the Gaussian component, country is significant, i.e., the Spanish fleet obtains higher efficiency in terms of catch by successful fishing days.

4.3. Overall

Table 5 shows the CPUE indices and standard errors for each of the three models. Case 1 was selected as the best fit for the CPUE between the three models.

5. Discussion

The absence of significant effect of the age of the vessels would mean that each purse seiner tend to keep a constant fishing power over time, even for the old vessels. This result tend to be against the empirical guess that modern vessels are systematically more efficient at a given size than the old vessels.

Another possible explanation may result in the apparent independence between the introduction on board of new equipments and the age of the vessel, and this whatever the nationality and class category of the vessel.

In addition, the validity of this statistical result tend to be limited by the fact that the year effect and the age effects cannot be distinguished by the model.

As most of the most important technological factors that are conditioning changes in fishing efficiency of purse seiners have not been identified in the present GLM, its results remain quite doubtful when assuming that any of these GLM indices correspond to fluctuation of the Yellowfin biomass. The ESTHER program (Gaertner, Pallarés, 2002) has shown that the improvement in the design and performances of purse seines, the use of bird radars, of improved navigation radars, of increasingly efficient sonars, of computers, the increasing use of satellite information, etc, are multiple factors that cannot be introduced in the present GLM but that are all producing increased efficiency of the PS fleet.

As a global consequence, the present GLM indices should tend to overestimate the real trend of stock biomass, as only few of the factors contributing to increased fishing powers are handled in the present model.

And as a conclusion, the increasing trends shown at variable degrees by all our GLM indices, probably mask a stability or a decline of the yellowfin stock. Such decline would not be surprising during a 22 years period during which total yellowfin catches have been steadily increasing and multiplied by a factor of 8 between 1982 and 2003.

References

Draper, N. R., y Smith, H., 1981. Applied regression analysis. Ed. Wiley series in probability and mathematical statistics, second ed.

Gaertner D. and Pallares P. (2002) Efficacité des Senneurs Thoniers et Effort Réels (ESTHER). Programme n° 98/061. Union Européenne, DG "Fisheries" (DG XIV), Bruxelles, (Belgique). Rapport Scientifique 187 p.

Lo. N. C. H., Jacobson, I. D. and Squires, J. L. , (1992). Indices of relative abundance from fish spotter data based on delta-lognormal models. Can. J. Fish. Aquat. C. 49(12):2515-2526.

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

Miller, R. G. 1986. Beyond ANOVA. Basic of Applied Statistics. J. Wiley and sons, New York.

Peña, D., 1998. Estadística, modelos y métodos. Vol. 1 y 2. Ed. Alianza Editorial.

Soto, M, Morón, J., Pallarés, P. (2002). A new approach to standardize catch rates for yellowfin (*Thunnus albacares*) for the Spanish purse seine fleet (1984-1995). WPTT-02-27.

Table 1: ANOVA table for Case 1 (response variable CPUE=YFT free school catch/searching time)

	Df	Sum of Sq	Mean Sq	F	Pr(F)
year	21	180.4767	8.59413	45.7978	0
Nsd	1	23.2123	23.21226	123.6973	0
Dnc	1	42.5878	42.58778	226.9487	0
Sd	1	24.9134	24.91343	132.7628	0
Cs	1	49.3055	49.30551	262.7472	0
Sf	1	72.6052	72.6052	386.9104	0
Residuals	850	159.5057	0.18765		

Table 2: ANOVA table for Case 2 (response variable CPUE=YFT free school catch/travelled distance)

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
year	21	234.3749	11.1607	46.3256	0
Nsd	1	179.4372	179.4372	744.8035	0
Dnc	1	35.2316	35.2316	146.2385	0
Sd	1	4.6266	4.6266	19.2042	1.3214E-05
Cs	1	39.6068	39.6068	164.3989	0
Sf	1	75.0277	75.0277	311.4231	0
Residuals	850	204.781	0.2409		

Table 4: ANOVA of binomial case for proportion of positives in Delta model (case 3, IOTC data)

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Year	22	705.908	32.0867	12.5336	0
Country	3	10.348	3.4494	1.3474	0.2572595
Area	9	2529.721	281.0801	109.7942	0
Season	3	598.892	199.6308	77.9789	0
Pfads	1	886.866	886.8662	346.4237	0
Residuals	2109	5399.171	2.5601		

Table 5: ANOVA of gaussian model for positives in Delta model (case 3, IOTC data)

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Year	22	10069.03	457.683	228.509	0
Country	3	1718.49	572.83	285.999	0
Area	7	4876.6	696.658	347.822	0
Season	3	207.8	69.268	34.583	0
Pfadc	1	2662.06	2662.063	1329.096	0
Residuals	11878	23790.59	2.003		

Table 6: Fitted CPUE indices and standard errors for each case.

	Case 1 CPUE=YFT free school/search time		Case 2 CPUE=YFT free school/distance travelled		Case 3 CPUE=YFT free school/fishing days	
year	CPUE	St.Error	CPUE	SE	CPUE	SE
1982					0.0299	0.2299
1983	0.879	0.295	0.8516	0.3343	0.0983	0.2186
1984	0.58	0.262	0.9012	0.2966	0.0981	0.4029
1985	0.705	0.26	1.2186	0.2951	0.0686	0.3025
1986	0.68	0.266	1.2312	0.3019	0.2112	0.4175
1987	0.418	0.268	1.159	0.3038	0.2312	0.5441
1988	0.658	0.266	1.2729	0.3017	0.7637	0.6395
1989	0.158	0.264	0.7669	0.2989	0.2047	0.3252
1990	0.849	0.262	1.4305	0.2969	0.4779	0.4619
1991	1.045	0.262	1.7421	0.2973	2.9058	1.4341
1992	0.83	0.262	1.4966	0.2966	3.2323	1.4009
1993	0.767	0.261	1.4297	0.2961	3.1976	1.3297
1994	0.743	0.262	1.4421	0.297	3.4503	1.4259
1995	0.666	0.262	1.3591	0.2969	3.0878	1.4378
1996	0.892	0.26	1.6803	0.2951	3.2583	1.4487
1997	0.564	0.258	1.3292	0.2925	1.5935	1.2901
1998	0.663	0.26	1.2054	0.2943	1.4737	1.1672
1999	1.099	0.261	1.4998	0.2953	2.194	1.3269
2000	1.164	0.263	1.5814	0.2983	2.336	1.5069
2001	0.952	0.261	1.5842	0.2959	2.4116	1.5476
2002	1.138	0.264	1.8652	0.2994	2.0842	1.4939
2003	1.194	0.267	1.8978	0.3024	3.8009	1.7773

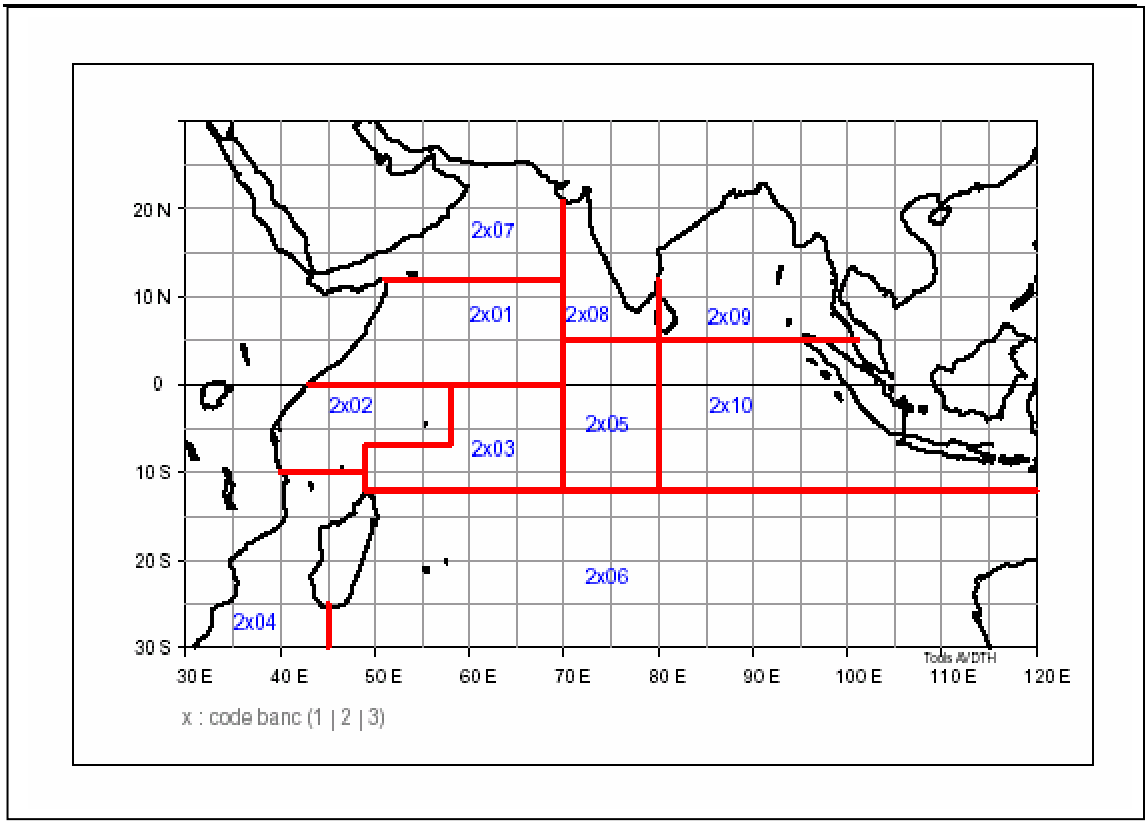

Figure 1: Fishing areas in the Indian ocean.

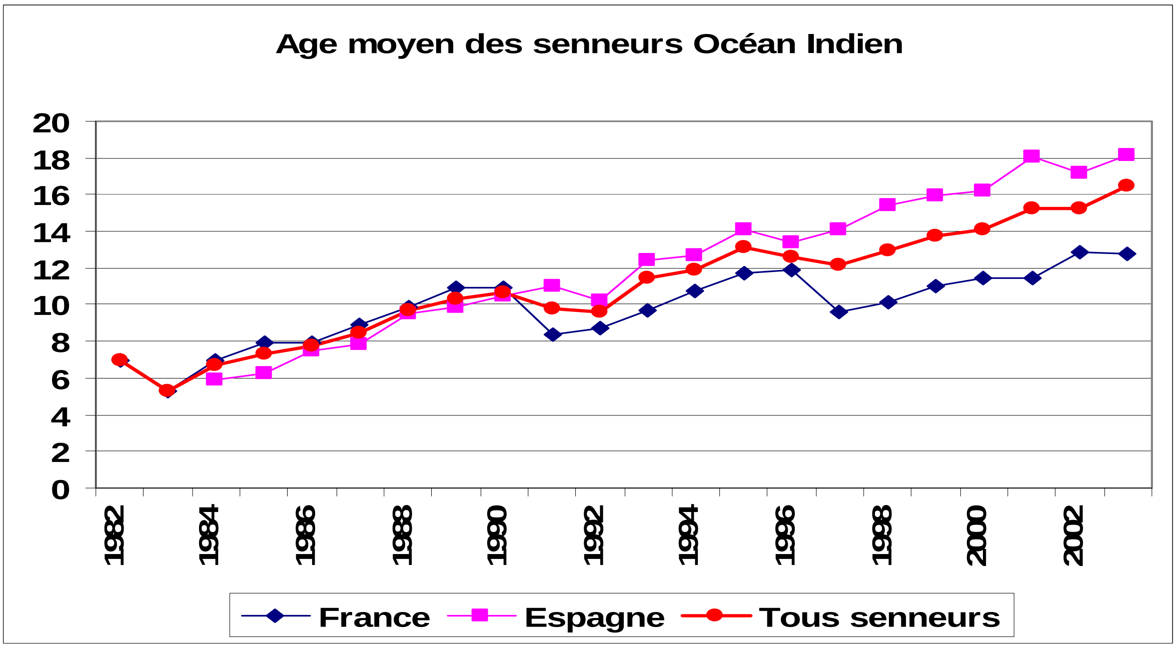

Figure 2: Average age of purse seine vessels in the Indian ocean.

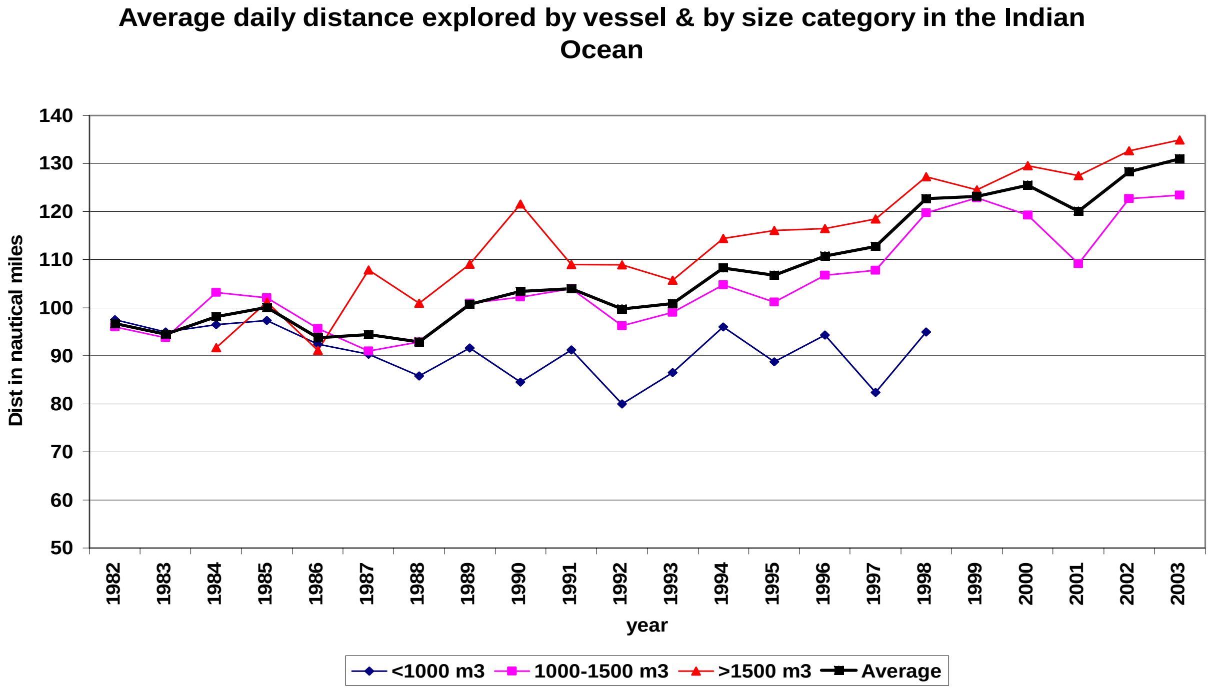

Figure 3: Average daily distance explored by vessel and by size category in the Indian Ocean

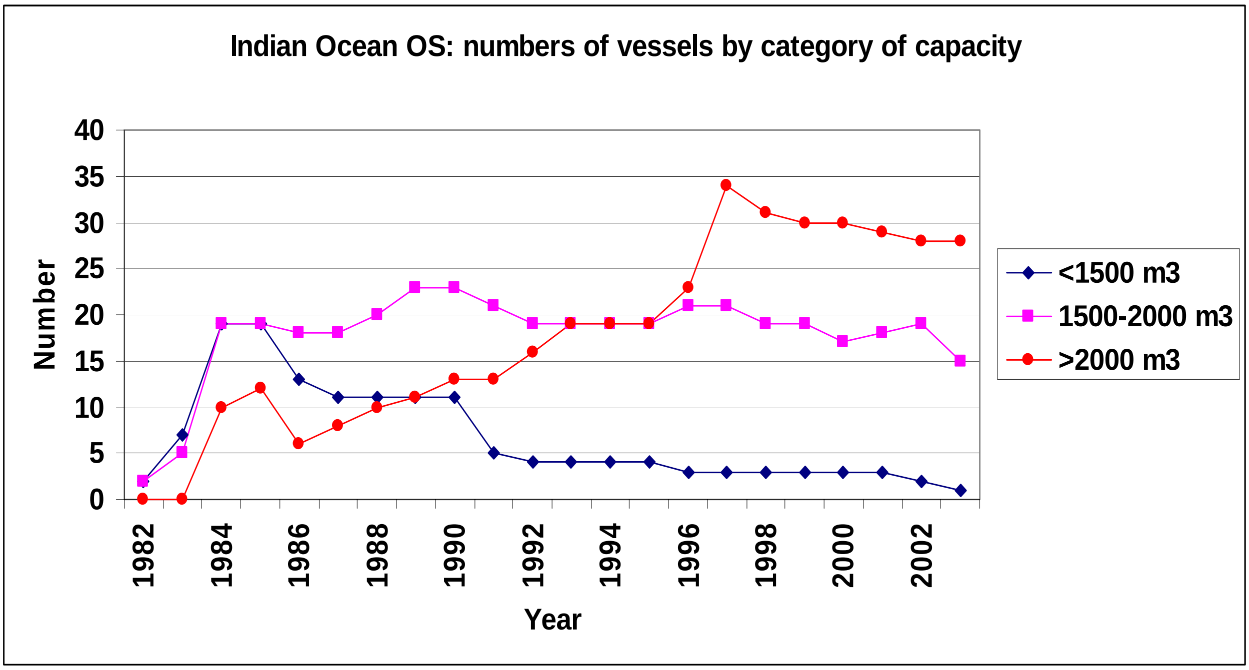

Figure 4: Number of vessels by capacity category in the Indian ocean.

Figure 5: CPUE index obtained from the coefficients of the factor year of case 1. Limits are defined as CPUE ±std. deviation.

Figure 6: Nominal CPUE (catches YFT on free school /searching time) and CPUE fitted in Case 1.

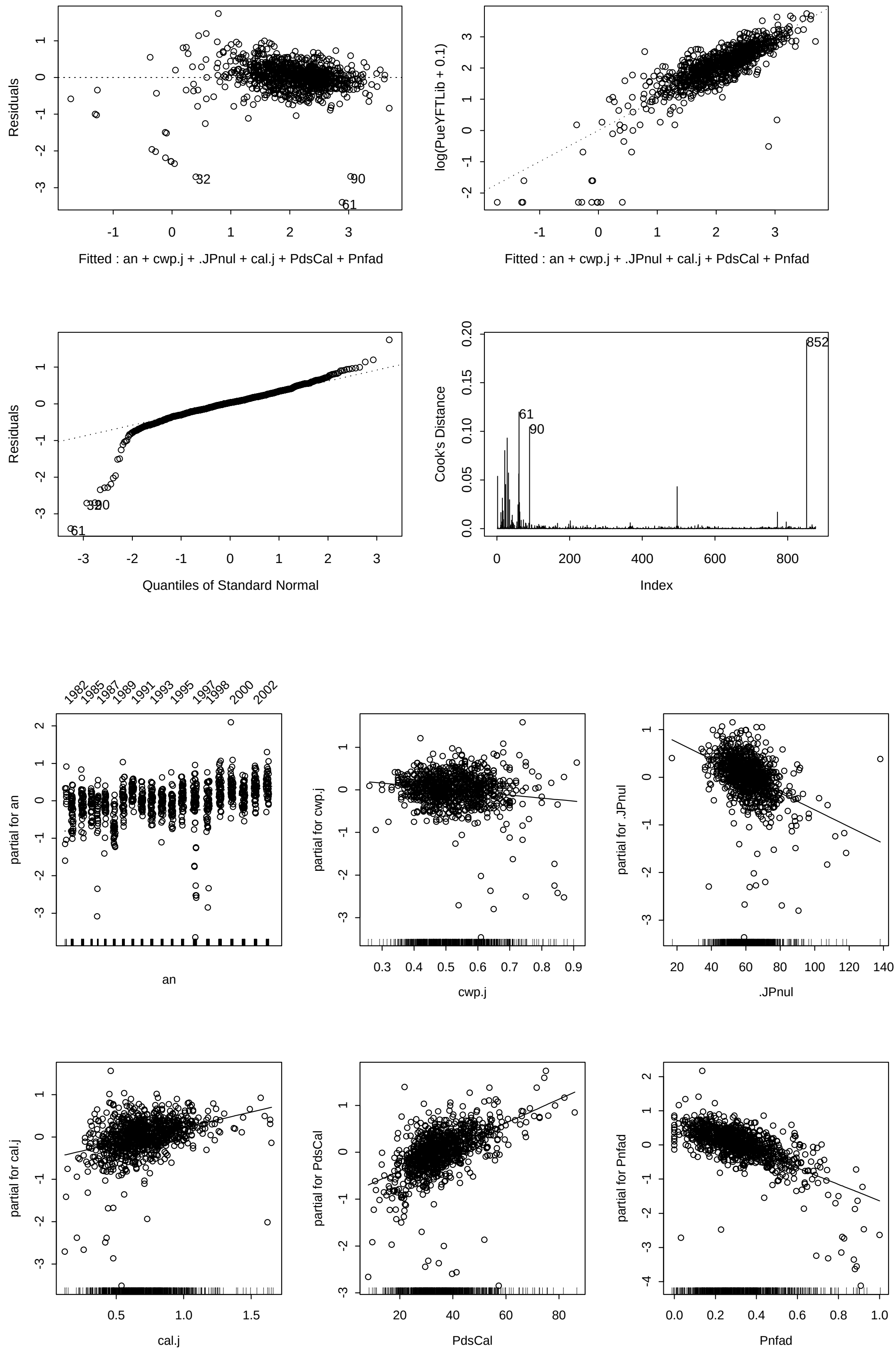

Figure 7: Diagnostic plots for Case 1

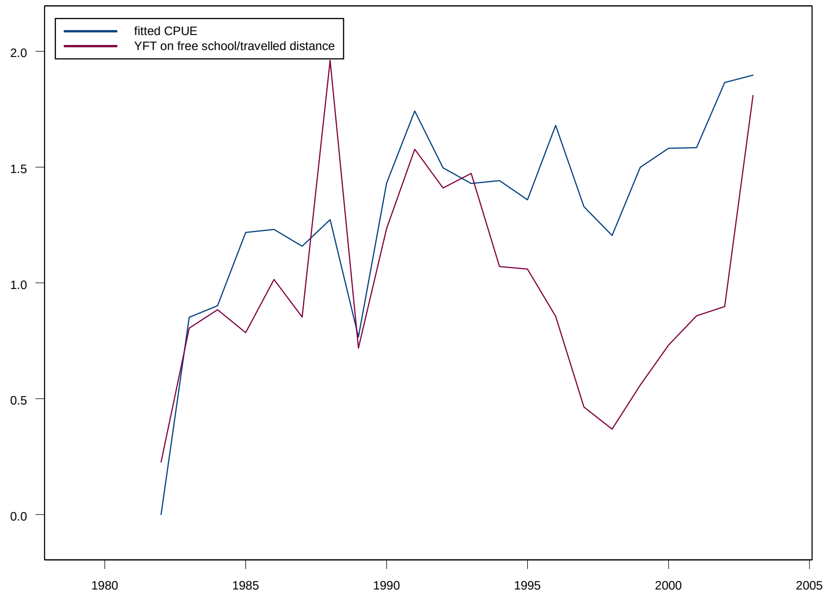

Figure 8: Nominal CPUE (catches YFT on free school/travelled distance) and CPUE fitted in Case 2.

Figure 9: Fitted CPUE of Case 1 (CPUE=YFT on free school/searching time) and Case 2 (CPUE=YFT on free school/travelled distance).

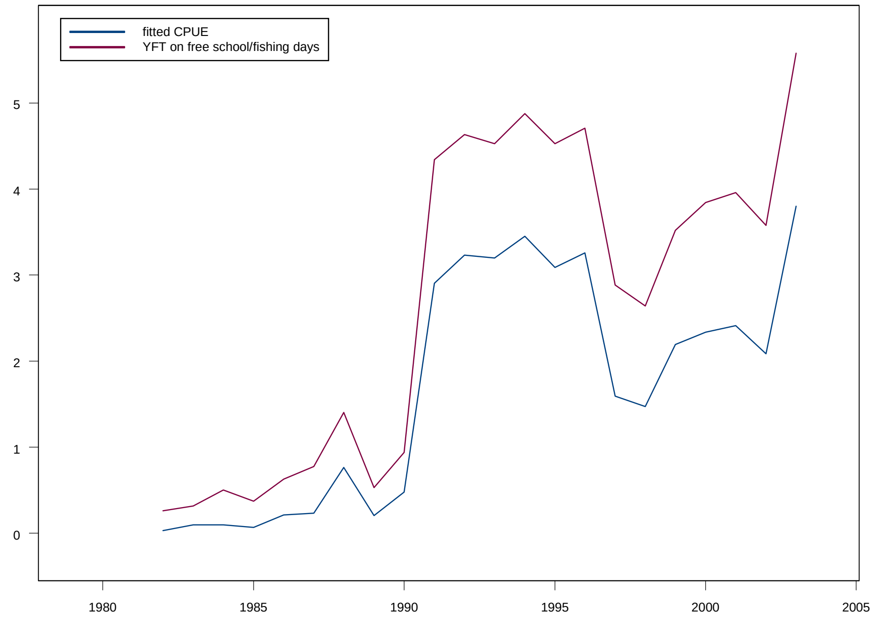

Figure 10: Nominal CPUE (catches YFT on free school/fishing days) and CPUE fitted in Delta model (Case 3, IOTC data).