STANDARDIZATION OF TROPICAL PURSE SEINE FISHING EFFORT BY GENERALIZED LINEAR MODEL (GLM)

by

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

ABSTRACT

This document explores different models to standardize fishing effort of the Spanish and French purse seine fleets fishing tropical tuna in the Indian Ocean. Models are based in GLM and include year, country, boat category and age as explanatory variables. Regarding the characteristics of the boats, data available are not enough accurate to explain all the variance due to vessel efficiency, so the resulting effort cannot be considered as a standardized effective effort. On the other hand, the characteristics of the purse seine fleets make difficult to understand the age effect because the negative effect of age can be compensated by positive effects of improvement in the technical equipment of the boats. So a simple model based in the size of the wells category and country is proposed for standardization. Regarding other purse seine fleets it is propose to consider their boats as French for years before 1986 and as Spanish for the more recent years on the standardization purpose. These proposals are based in the results of anova and discriminant analysis conducted.

INTRODUCTION

The procedure used to standardize the nominal fishing effort of the tropical purse seine fleets has been widely discuss by IRD and IEO experts.

Methods used in the Indian Ocean have been similar to those applied in the Atlantic. Basically the procedure consisted in a two step standardization, first by fleet (France and Spain) and secondly between fleets. Regarding the statistical point of view the procedure was not consistent so the development of a new method valid statistically was recommended.

This document has two parts, in the first part different standardization models are presented based in generalized linear model (GLM). All models take into account the physics characteristics of the boats as well as the age of each vessel. Other characteristics related with technical equipment or changes in the use of floating objects are not included in the analyses because this information was missing for many boats. So the resulting standardized fishing effort must be consider as a measure of the nominal fishing effort as they do not take into account the potential increase of vessel's efficiency. In a second part, data from other purse seine fleets are analyze in order to establish a general criterion to assign those data to the French or Spanish fleets.

STANDARDIZATION OF FISHING EFFORT

Material and Methods

Detailed catch and effort data obtained from log book were used in the analyses as well as the boat's main characteristics. Species composition was corrected following the method proposed by Pianet, 2000. Catch and effort data are recorded by set. Boat's data include age, physical characteristics (length, size of wells, GT, etc.) and other information related with the situation of the boat along it life . Spanish and French data have been analyze together to obtain an overall standardized effort. The period considered was 1985 to 1999 because before 1985 the Spanish fleet fished in a exploratory way. In the second part of the analysis data from all the other purse seine fleets fishing in the western Indian Ocean were also considered.

It was first verified that all the characteristics of the boats (length, power, size of wells...) were highly correlated; the size of wells in m^3 was defined as representative of boat's category. This variable is generally well known and unbiased. Based on the distribution of this variable among fleets four levels were considered:

category 1	< 1250 m^3
category 2	1250 – 1499 m^3
category 3	1500-1750 m^3
category 4	>1750 m^3

[1] Instituto Español de Oceanografía. Corazón de María 8. 28002 Madrid (ESPAÑA).
E_mail: pilar.pallares@md.ieo.es.
[2] IRD (UR 109).Centre Halieutique Méditerranéen et Tropical. Avenue Jean Monnet BP 171, 34203 Sete Cedex. E-mail: gaertner@ird.fr.
[3] 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.
[4] IRD. BP 570 – Victoria, Seychelles.

Boats with size of wells higher than 2000 t. were eliminated because they are only present in the Spanish fleet and could unbalance the analysis. Figure 1 shows the distribution of the fleets by size of wells category.

The yearly catch rate -total catch (all species and fishing modes)/total effort (fishing days)- by boat was considered as the response variable.

A minimum of 120 fishing days was established as threshold. A 14% of data were excluded from the analysis applying this level of minimal effort. This threshold was defined after verify that catch rate and fishing days were not correlated and that variability of the catch rate by boat was stabilized after 120 fishing days.

After verify that the coefficients corresponding to age had a continuous negative trend, the age of the boat was considered as a continuous variable in order to reduce the number of parameters to include in the model. This assumption allowed the development of a more generalized model allowing to obtain annual indexes by boat as a function of their category (generally stable through the boat life) and age (in number of years).Taking into account the mean age of boats considered and the distribution of ages through the period analyzed (Figure 2) a maximum of 15 years was established in order to maintain the age effect constant for boats 15 years old and over.

Generalized linear models were applied using the S-Plus statistical software. This software include the contrast treatment option to make the coefficients relatives to the first level of the different factors. This option was used to allow an easier understanding of the results.

Due to the asymmetric distribution of response variable (Figure 3) a log transformation was applied. Year, country, age and boat category were included as explanatory variables.

In order to compare fitted model objects we calculate the Akaike information criterion according to the formula -2*log-likelihood + 2*npar, where npar represents the number of parameters in the fitted model.

RESULTS AND DISCUSSION

Firsts analyses showed year, boat category, country and age as highly significant. A strong interaction between country and boat category was also found. Based on these results a new model was developed taking into account both the statistical and the practical point of views. Regarding the practical aspect, the model should be flexible enough to allow it to build the historical data base of standardization factors only once, and to obtain annual standardized effort in the future for each incoming year.

In order to simplify the model a new variable was created by mixing country and category. This variable was defined as a factor with eight levels: 1 =French boats between 1250 and 1499 m^3; 2 = French boats <1250 m^3; 3 = French boats between 1500 and 1750 m^3; 4 = French boats >1750 m^3; 5 = Spanish boats <1250 m^3; 6 = Spanish boats between 1250 and 1499 m^3; 7 = Spanish boats between 1500 and 1750 m^3; 8 = Spanish boats > 1750 m^3. This variable was called CATPAIS. The French boat between 1250 and 1499 m^3, which was well represented in the fishery during the period has been selected as being the reference vessel..

After some preliminary analyses based on a stepwise model selection two models were formulated:

Model 1

Ln (CPUE) = Year + Catpais +Age+Catpais*Age+ ε,

were CPUE is the observations vector and ε is the independent component error normally distributed $N(0, \sigma^2)$. Year and catpais were considered as fix factors.

Table 1 shows the the anova table of the model.

Including the age-catpais interaction we tried to explore differences in the age coefficient depending of the catpais variable. These differences can be explained because the replacement of the equipment of the boats is not uniform by fleet and/or boat size. This is an important factor because the change and improvement of equipment of the boats is frequent and we can presume that bigger boats have a more systematic replacement than the smaller ones. In fact, the age of the boat as an independent variable has not a significant effect if we include the interaction term.

From this model the Akaike statistic was –170.0494, the explained variance a 55% and all the factors were significant. Distribution of the residuals and partial residuals as well as normal fit of the model (Figure 4) shows a good fit of data to the model.

This model provides predictions of effort regarding category and/or age of the boats. Figure 5 shows the country-category coefficients in original scale, these coefficients represent the effort corresponding to country-category variable relative to the first level in the first year (1985) and age equal to 0. We can see an increase in efficiency by increasing the size of wells. Considering only the catpais effect we can conclude that for all the categories, the French fleet appears to be more efficient than the Spanish fleet. However if we analyze the interaction age:catpais coefficients (Figure 6) we can see there are probably significant improvement in the technical equipment of the Spanish boats over time which appear to be much more important than those of the French fleet. Regarding the effort standardization the model can predict, by boat and year, the relative effort taking as reference a French purse seiner between 1250 and 1500 m^3 aged 0 in 1985.

So the effort corresponding to boat i in year j can be predicted by:

$\ln (cpue_{ij})\hat{} = \alpha+\beta_1+\beta_2*n$

i = boat, j = year

Boat of reference =cat 1(1250-1500 m3), year 1985, age 0

α = ln(cpue boat of reference) = 2.7279

β_1 = coefficient corresponding to category

β_2 = coefficient corresponding to interaction age:category

n = age of the boat

α, β_1 y β_2 in logarithmic scale.

For each boat, the standardization factor can be obtained by dividing the effort predicted by the effort corresponding to the boat of reference both retransformed to the original scale ($e^{(\ln(cpueij))} / e^{\alpha}$).

Model 2

As an alternative to the previous model, a simpler one was formulated including only year and catpais as explanatory variables:

Ln (CPUE) = Year + Catpais + ε,

Table 2 shows the results of this model.

Figure 7 shows the distribution of the residuals, partial residuals and normal fit of the model.

The variance explained by the model was 50% and the Akaike statistic –134.5725.

Looking the coefficients corresponding to the catpais variable (Figure 8) and comparing with those obtained by model 1 we can reinforce the idea that the efficiency of the Spanish fleet is a result of both the size of boats and their capacity to develop technical improvements. On the contrary the efficiency of the French fleet is mainly explained by the size of the boats.

Regarding the statistical point of view, the explained variance as well as the Akaike statistic indicates that Model 1 is better than Model 2. It means that the inclusion of an interaction term between age and CATPAIS improves the model.

In order to analyze the consequence of applying one or another model in the overall fishing effort we have compared the Spanish and French nominal fishing efforts with the standardized efforts obtained by the two models (Figures 9-10).

We can see that standardized effort from model 2 is very close to nominal effort for both the Spanish and French fleets. Looking the coefficients of the variable captais we can see that the standard category is in the middle, closed to Spanish boats between 1250 and 1750 higher than French an Spanish boats smaller than 1250 m^3 and lower than French an Spanish boats bigger than 1750 m^3. Standardized effort from model 1 is slightly lower because of the negative values of the interaction age-catpais coefficient.

It is clear that the variables related with the characteristics of the boats included in the models are not enough accurate to explain well the variance due to vessel efficiency. We can assume that the year effect includes change in catchability as well as in abundance (even though this is difficult to understand because we are using total catch) . Figure 11 shows the year coefficients, the increasing trend at the beginning of the fishery as well as in the nineties may be partially due to increasing catchability, on the other hand the dramatically decrease in 1998 is probably related with the negative effect in catchability of "el Niño".

In conclusion, information of variables related to catchability is not sufficient to obtain accurate effective fishing effort for the Spanish and French purse seine fleets. However it is possible to standardize effort by fleet and category.

Comparing the two models proposed, both include variables such as catpais which are relatively constant along the boat's life. This is important because the same model can be applied during a more or less large period, depending of changes in the fleet structures. In the case of model 1 the fact to include the temporal component (age of boat) as a continuos variable also allows to keep the model through a period. However in this case the period could be shorter because the age structures of the fleets have been continuously changing.

The characteristics of the purse seine fleets make difficult to understand the age effect. During the life of the boats their equipment changes several times. In some cases the changes can affect even their physical characteristics. So the negative effect of age can be compensated by other positive effects. Available information does not allow us to track the equipment of the boats so we can only partially take into account these positive effects in our models.

Taking this fact into account and considering that from the statistical point of view differences are not so significant to refuse model 2 we propose this model to standardize the French and Spanish purse seine effort. To these fleets this model can provide standardized effort in fishing days of French purse seine (between 1250-1499 m^3), aged 0 in 1985.

This result shows the difficulty to standardize the purse seine effort using GLM or similar methods. The problem is general for all the tropical tuna purse seine fisheries. In one side we have the difficulty to obtain accurate information on the technical equipment of the boats and their changes, in other side we have the problem to isolate and understand many different effects specially because they are highly correlated. In addition we must handle with unbalanced analysis due to the fact that changes in the fleets affect all the boats and are made in a short period of time. So it is difficult to have overlapping between equipment as has been showed in Soto et all (2000, 2002). The development of integrated statistical models could be an alternative approach to this problem.

CLASSIFICATION OF OTHER PS FLEETS

METHODS

The effort applied by other purse seine fleets is important, so it must be included in the standarization process. A possible alternative would be to add these fleets to the model as a third level of the fleet variable. Another way is to assume that they are similar to the Spanish or French fleets and to apply to them the correspondent standardization process.

Although two components can be identified as a function of their skipper and ownership nationality, (Spain or France) we treated all these fleets as a unit.

To assign the other fleets to the Spanish or French fleets, there are several statistical options. Regarding the variables we considered the age of the ship, the well capacity and the estimated CPUE as the most representatives. This selection was made taking into account the expert criteria rather than the statistical criteria. Nevertheless, correlation between these variables were high enough to apply multivariant techniques.

For the classification problem, discriminant analysis is a good option when normality of the variables is assumed. Also, logistic regression can be used to classify the fleet when the dependent variable is qualitative, the regression trees when the variables are very heterogeneous and neural networks in a global classification context.

A first approach to explore the influence of each variable on the CPUE and to establish differences between the levels of each factor was the analysis of variance,.

ANOVA

An anova was made taking into account two periods: 1982-1986 and 1987-1999. These periods were definied based on the results of a regression tree, in which the principal grouping variable was the year.

A variable called PAIS (country) corresponding to the fleet was created with three values: 1 to the French fleet, 4 to the Spanish fleet and 2 to other fleets. For each period, the model was the following:

CPUE ~ PAIS + ANTIGÜEDAD + CATEGORIA + CATEGORIA : PAIS

Where CATEGORIA is the category of the ships, depending on the volume of well as previously defined.

With this model we explore if the variable PAIS has a significative effect on the CPUE (p-value very small or F-ratio very large). Once we know if the fleets are different we compare the equality of means for each pair of fleets using the statistic

t0 = (yi. – yj.)/(1/ni + 1/nj)1/2

where n is the number of total observations and k the number of levels of each factor. This statstic follows a t-Student distribution with n-k degrees of freedom. Also, we can obtain confidence intervals for the difference of means. If differences are significant the corresponding intervals do not include the zero.

For the first period, 1982-1986, the fit of the model is poor, because the number of observations is very small. The variable PAIS is less significant, and its interaction with the cathegory of the ship does not contribute to improve the model. The means difference between France and other fleets in the multiple contrasts is smaller than the difference between Spain and other fleets. So other fleets could be assign to the French fleet in that period. Tables 3 and 4 shows the Type III anova and multiple contrasts for the first period.

During the second period, 1987-1999, the variable PAIS is very significant but confidence intervals in the multiple contrast of differences in the means include the zero. However, the absolute value of the estimator of the difference in means between Spain and NEI is smaller than the difference between France and other fleets as well as the confidence intervals. So, we could assign other fleets to the Spanish fleet. Tables 5 and 6 shows the Type III anova and multiple contrasts for the latest period.

The prediction power of the anova is not very high in this case, so another approach was made to improve previous results. The discriminant analysis is an adequate technique to deal with this classification problem.

DISCRIMINANT ANALYSIS

Taking a group of elements (ships) classified into two populations (Spanish fleet and French fleet), we select variables related with the characteristics of the boats such as: carrying capacity, age and CPUE. We tryied then to find a discriminant function to split groups and to classify the elements into the groups. Also, we try to evaluate statistical differences between groups and to establish a classification rule for new elements (other fleets).

Figure 12 shows average CPUE by boat and age for the period 1982-1999. We can see a decreasing trend until the age of 21 years follow by a dramatical increase. Boats older than 21 were eliminated because they were only Spanish and make difficult to find a classification criterium for the other fleets.

The high CPUE values for old boats could be explain by the improvement in their technical equipment through their life. Because the information available concerning these technical changes was too scarce, they do not allow to separate the improvement effect and the age effect. It was then decided to eliminate the age in the analysis, because this variable should be biased. So only the carrying capacity and CPUE were included.

As in the analysis of variance models, two periods were considered, 1982-1986 and 1987-1999. The overall period was also considered.

Carrying capacity and mean CPUE were scaled by their means and standard deviation, and it was assumed that they followed Normal distribution with identical variance.

The linear discriminant function was the following:

PAIS = A*CC + B*CPUE

Where A is the vector of constants and B a 2x2 matrix of coefficients.

If we apply the discriminant function to the elements of the Spanish and French fleets, the resulting values provide the centroides vector (average by group). This vector was used to classify boats from another fleets by comparing their discriminant function values and assigning to the group corresponding to the closer centroide. The Fisher criterium was used for classification.

Table 7 shows the discriminant function predictions.

For all the years (1982-1999) 14 boats were classified as French and 15 as Spanish. By periods, from 1982 to 1986 8 ships were classified in the French fleet and 1 in the Spanish fleet, mainly due to the fact that the Spanish fleet started fishing at the end of the period. From 1987 to 1999 8 ships were classified as French and 14 as Spanish.

It seems that other fleets were closer to the French fleet during the first period and to the Spanish fleet during the later period.

Differences in classification can be explained by the fact that the Spanish fleet started fishing at the end of the first period.

CONCLUSIONS

- The permanent improvement in the technical equipment of the purse seiners make difficult to obtain standardized effective fishing effort. Information available is not sufficient detailed to get such an effort.
- Physical characteristics of the boat are not enough accurate to explain all the variance due to vessel efficiency.
- Although the considerations above, we can standardize effort by fleet and some physical characteristics of the boats. The resulting effort would be closer to nominal than to effective effort.
- Considering the high correlation between physical characteristics of boats we propose the size of wells as a measure of their carrying capacity.
- The models shows that the efficiency of the Spanish fleet is a result of both the size of boats and their capacity to develop technical improvements. On the contrary the efficiency of the French fleet is mainly explained by the size of the boats.
- The characteristics of the purse seine fleets make difficult to understand the age effect, because the negative effect of age can be compensated by positive effects of improvement in the technical equipment of boats.
- The GLM models could not be the best approach to get PS effective effort due to the lack of information related with changes in the characteristics of boats and the special dynamics of these changes. Alternative approach as the integrated statistical models could be a better option.
- Taking into account the misunderstanding of the age effect the model proposed to standardize effort only include the size of the wells and the fleet.
- Unit of reference for standardization is a French purse seiner with a size of well between 1250 and 1500 m^3.
- Regarding other PS fleets we propose to consider these fleets as Spanish for the standardization since 1985 and as French for the previous years.

BIBLIOGRAPHY

ALLEN, R. AND R. PUNSLY: 1984. Catch rates as indices of abundance of yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean. Int.-Amer. Trop. Tuna Comm. Bull. 18: 301-379.

CHAMBERS, J.M. AND T.J. HASTIE. 1991: Statistical Models in S. Ed.Champman

HILBORN, R., AND C.J. WALTERS 1992: Quantitative Fisheries Stock Assessment: Choice, Dynamics, & Uncertainty. Chapman and Hall, New York.

MCCULLAGH, P. AND NELDER, J.A., 1989. Generalized linear models, second ed. Ed. Chapman & Hall.

PALLARÉS, P., D. GAERTNER, M. SOTO, A. DELGADO DE MOLINA AND J. ARIZ, 2001: Estandarización del esfuerzo de las flotas de cerco tropical por medio de modelos lineales generalizados. ICCAT. Col. Doc. Cient. (in press).

PIANET, R., P. PALLARÉS AND CH. PETIT, 2000: New sampling and data processing strategy for estimating the composition of catches by species and sizes in the european purse seine tropical tuna fisheries. IOTC. WPDCS/00/10.

PUNSLY, R. 1987 Estimation of the relative annual abundance of yellowfin tuna, *Thunnus albacares*, in the eastern Pacific Ocean during 1970-1985. Int.-Amer. Trop. Tuna Comm. Bull. 19: 3.

PUNSLY, R.G. AND R.B. DERISO. 1992 Estimation of the abundance of yellowfin tuna, *Thunnus albacares*, by age groups and regions within the eastern Pacific Ocean. Int.-Amer. Trop. Tuna Comm. Bull. 20: 99-131.

ROBSON D.S., 1966: Estimation of the relative fishing power of individuals ships. ICNAF Res. Bul., 3: 5-14.

SABATIER R., J.D. LEBRETON AND D. CHESSEL. 1989: Principal componen analysis with instrumental vaiables as a tool for modelling composition data. Multiway data analysis, p.: 341-352. Ed. R. Coppi & S. Bolasco.

VENABLES W.N. AND B.D. RIPLEY, 1999: Modern Applied Statistics with S-Plus.Third edition. Ed. Springer. 501 p.

TERRANCE J. QUINN II AND R.B. DERISO 1999: Quantitative Fish Dynamics.Oxford Univ. Press. New York.

Table 1. - Anova table from model 1.

ANOVA type III table					
	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
AN	14	9.78681	0.6990578	17.88006	0
ANT	1	0.00633	0.0063322	0.16196	0.687541
CATPAIS	7	1.66140	0.2373435	6.07062	0
ANT:CATPAIS	7	1.99239	0.2846265	7.28000	0
Residuals	468	18.29743	0.0390971		

Table 2.- Anova table from model 2.

ANOVA type III table					
	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
AN	14	9.91545	0.708246	16.61535	0
CATPAIS	7	7.27948	1.039925	24.39649	0
Residuals	476	20.28999	0.042626		

Table 3.- Anova model for the period 1982-1986

ANOVA type III table					
	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
FLEET	2	611.953	305.9763	18.84917	0
CATEGORY	3	811.235	270.4117	16.65827	0
AGE	1	99.211		6.11174	0.014
CAT:FLEET	5	1150.795		14.17856	0
Residuals	503	8165.139	16.2329		

*Table 4.- Multiple comparisons between different levels of variable PAIS (fleet) for the period 1982-1986. Critical point is 2.3784 and response variable is CPUESC. Intervals excluding 0 are flagged by '****' .*

95 % simultaneous confidence intervals for specified linear combinations, by the Tukey method				
FLEET	Estimate	Std. Error	Lower Bound	Upper Bound
1-2	-0.323	1.020	-2.760	2.11
1-4	1.460	1.000	-0.914	3.84
2-4	1.790	0.925	-0.412	3.99

Table 5.- Anova model for the period 1987-1999

ANOVA type III table					
	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
FLEET	2	611.953	305.9763	18.84917	0
CATEGORY	3	811.235	270.4117	16.65827	0
AGE	1	99.211	99.211	6.11174	0.014
CAT:FLEET	5	1150.795		14.17856	0
Residuals	503	8165.139	16.2329		

*Table 6.- Multiple comparisons between different levels of variable PAIS (fleet) for the period 1982-1986. Critical point is 2.3507 and response variable is CPUESC. Intervals excluding 0 are flagged by '****' .*

95 % simultaneous confidence intervals for specified linear combinations, by the Tukey method				
FLEET	Estimate	Std. Error	Lower Bound	Upper Bound
1-2	-0.753	0.570	-2.090	0.586
1-4	-0.283	0.463	-1.370	0.805
2-4	0.471	0.553	-0.829	1.770

Table 7. - Classification of boats from other fleets as French or Spanish fleet by discriminant analysis. Number 1 corresponds to the French fleet and number 4 to Spanish fleet.

BOAT		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
PERIOD	82-99	4	4	4	4	1	1	1	1	4	1	1	1	4	1	1	1	1	4	4	4	4	4	4	4	4	4	1	1	1
	82-86	1	1	1	1	1	1	1	1	4																				
	87-99								1	4	1	1	4	4	1	1	1	1	4	4	4	4	4	4	4	4	4	4	4	1

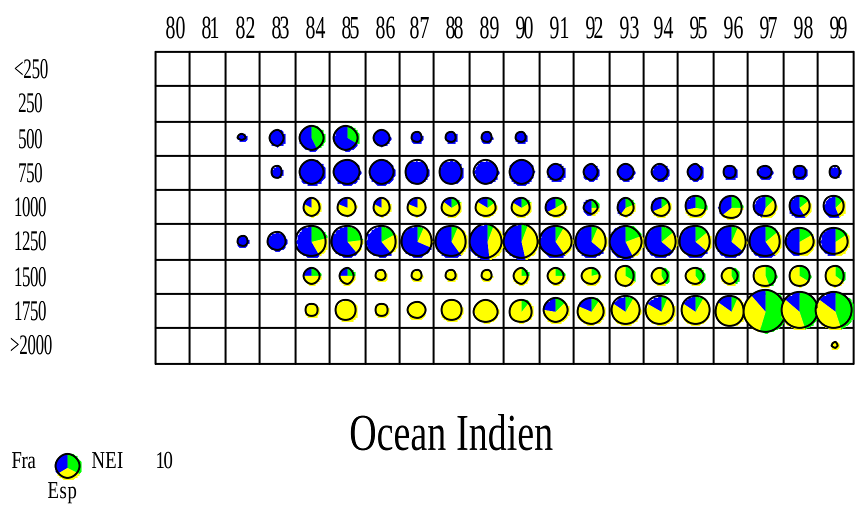

Figure1: Distribution of boats by fleet and well's size.

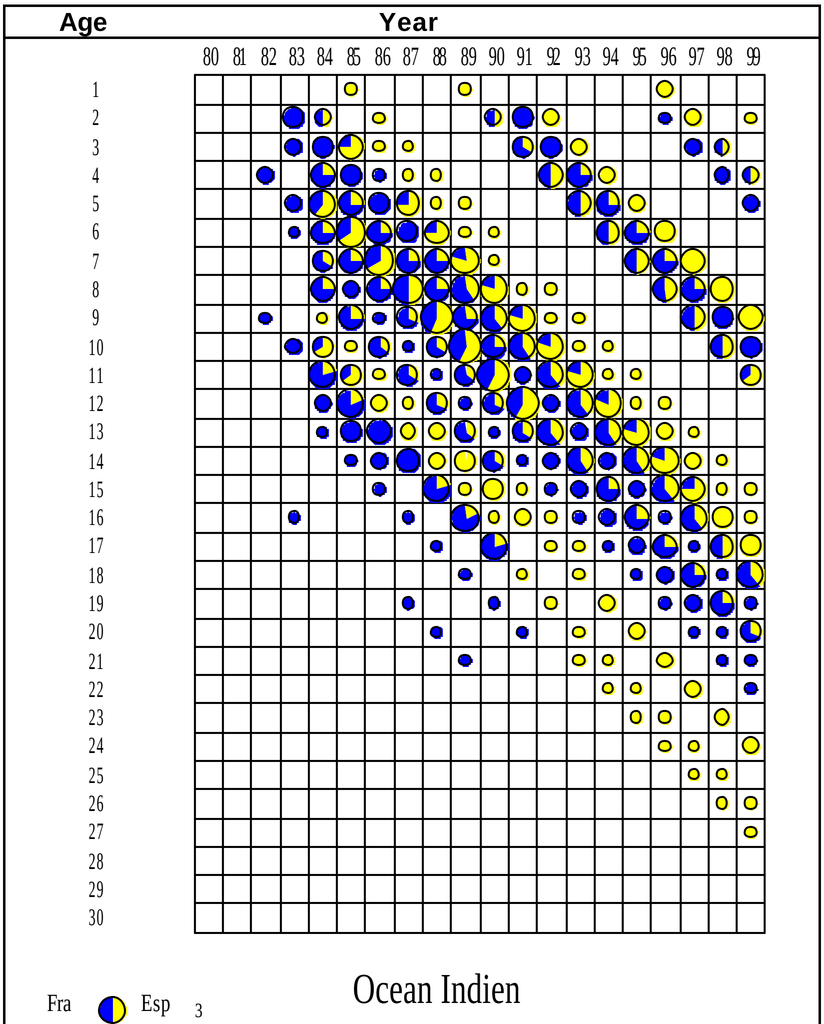

Figure 2: Distribution of boats by fleet and age of boats.

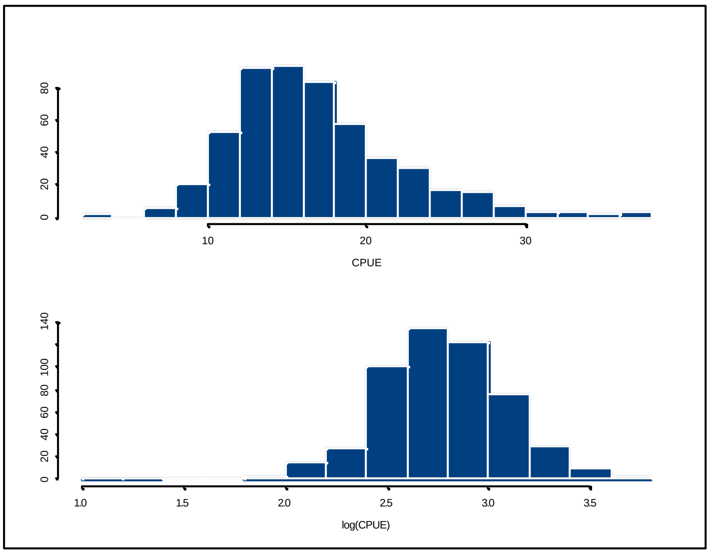

Figure 3: Distribution of cpue and log (cpue).

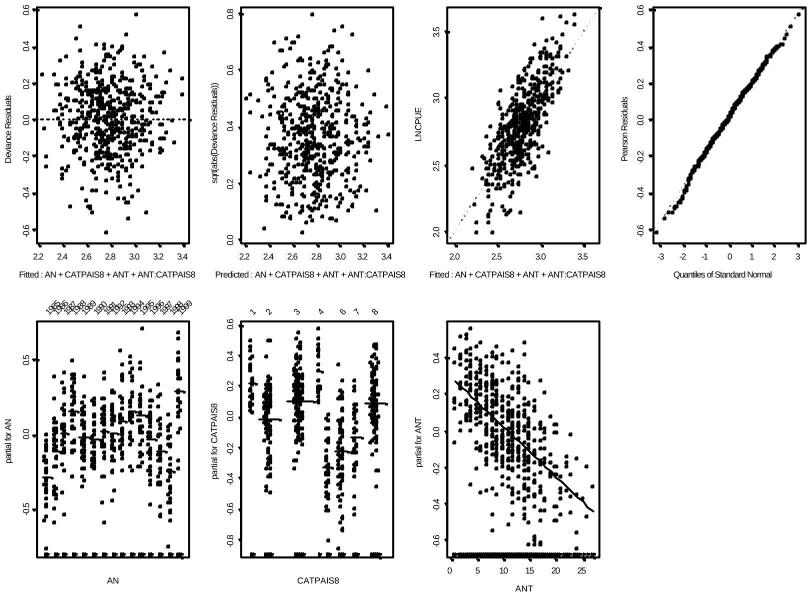

Figure 4: Distribution of residuals, partial residuals and normal fit of model 1.

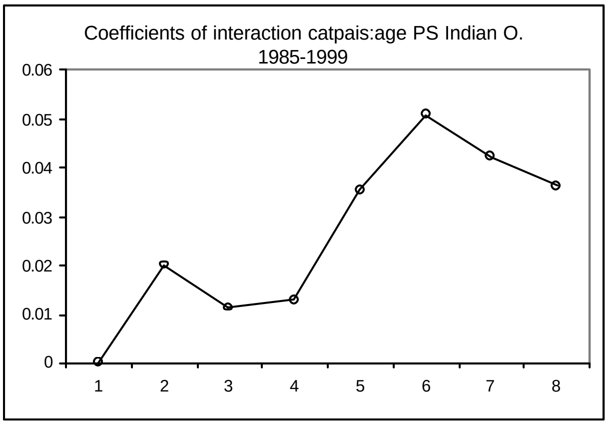

Figure 5: Coefficients of variable catpais estimated by model 1 in log scale.

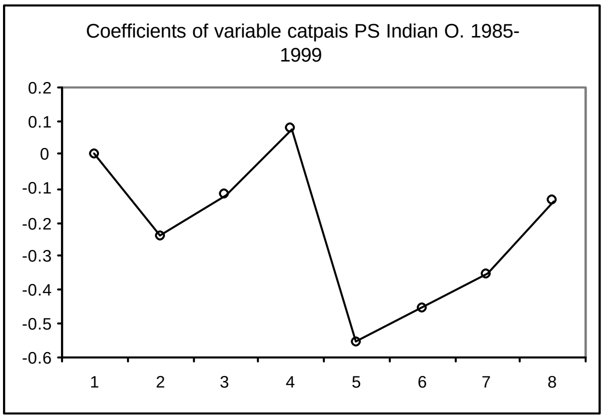

Figure 6: Coefficients of interaction catpais: age estimated by model 1 in log scale.

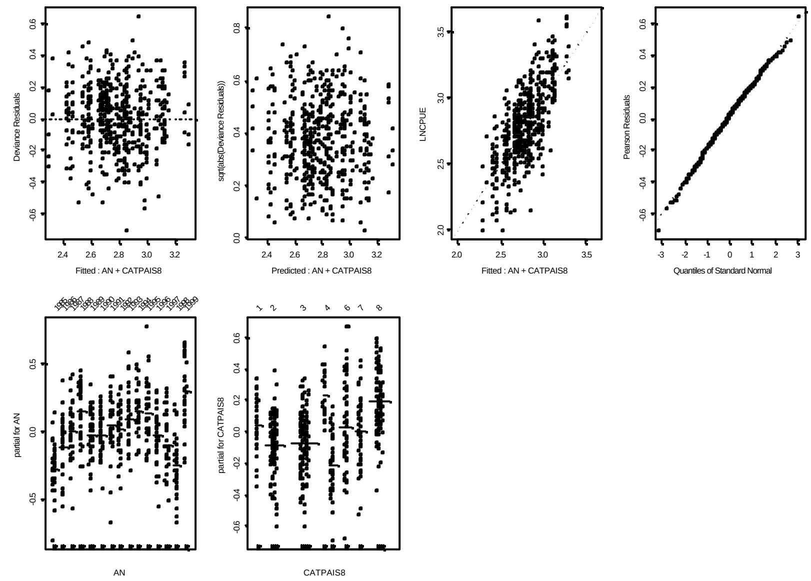

Figure 7: Distribution of residuals, partial residuals and normal fit of model 2.

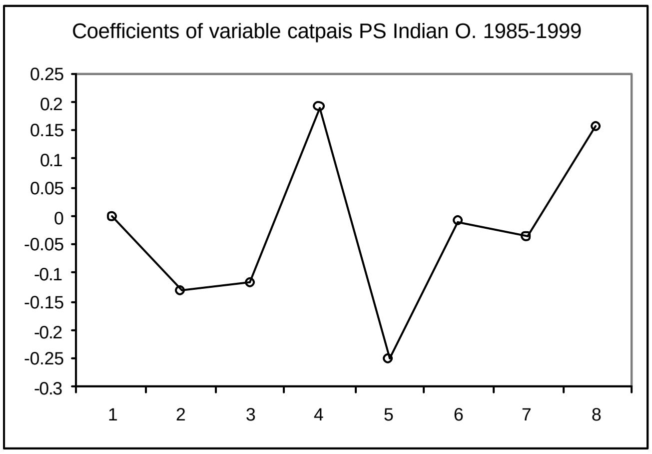

Figure 8: Coefficients of variable catpais estimated by model 2 in log scale.

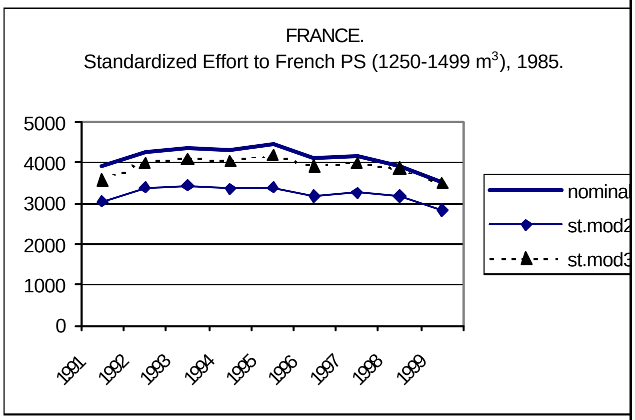

Figure 9: Nominal and standardized fishing effort of the French PS fleet.

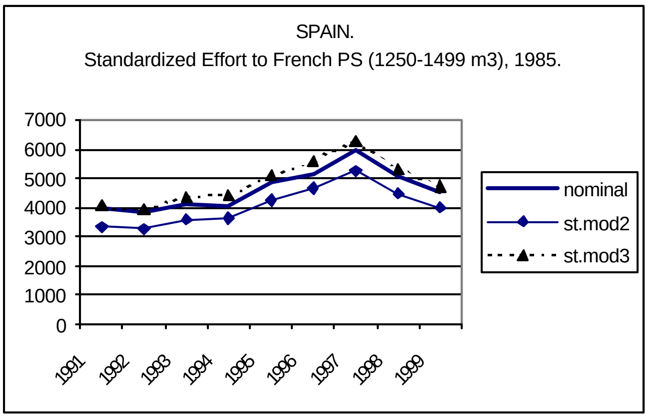

Figure 10: Nominal and standardized fishing effort of the Spanish PS fleet.

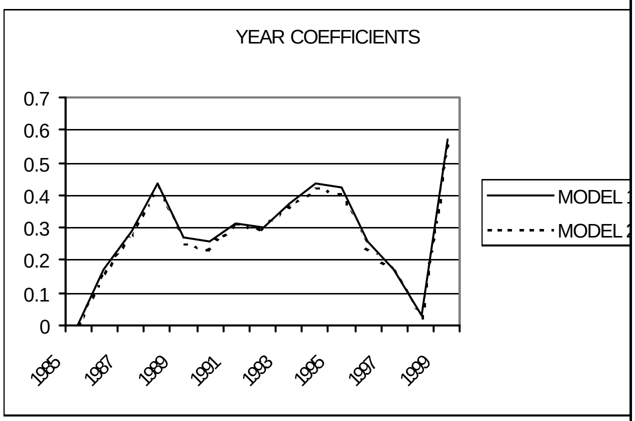

Figure 11: Coefficients of variable year estimated by model 1 and model 2 in log scale.

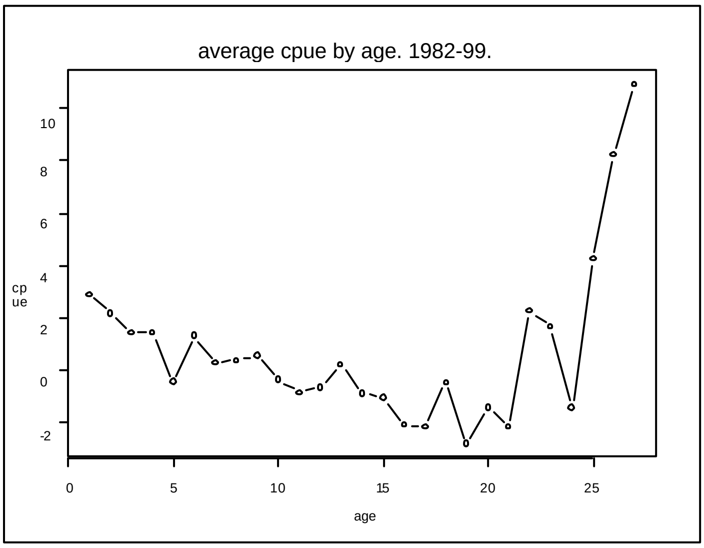

Figure 12. - Average purse seiner CPUE by age in the Indian Ocean for the period 1982-1999.