SCRS/2011/125 Collect. Vol. Sci. Pap. ICCAT, 68(2): 615-631 (2012)

STANDARDIZED CATCH RATES OF ALBACORE (THUNNUS ALALUNGA) CAUGHT BY THE BRAZILIAN FLEET (1978-2010)

Humber A. Andrade[1]

SUMMARY

Standardized catch rates were calculated for the Brazilian fleet using a zero inflated mixture model based on poisson distribution. The database contains information on the longline sets carried out by national and chartered vessels from 1978 to 2010. Models were fit to a pooled dataset with the flag as explanatory factor and to datasets split according to flag. The catch in number of fish was the response variable. Effort was considered as offset. The year and number of hooks per basket were the more important explanatory variables. The standardized catch rates showed symptoms that suggest they may be useless as relative abundance indices. In order to gather appropriate indices, other factors than those considered in this paper should be included in the models.

RÉSUMÉ

Les taux de capture standardisés ont été calculés pour la flottille brésilienne à l'aide d'un modèle mixte modifié en zéro basé sur la distribution de Poisson. La base de données contient des informations sur les opérations palangrières réalisées par des navires nationaux et affrétés de 1978 à 2010. Les modèles ont été ajustés à un jeu de données regroupées, le pavillon étant le facteur explicatif, et à des jeux de données divisés en fonction du pavillon. La prise en nombre de poissons était la variable réponse. L'effort a été considéré comme compensation. L'année et le nombre d'hameçons par panier étaient les variables explicatives les plus importantes. Les taux de capture standardisés font apparaître des symptômes suggérant qu'ils pourraient être inutiles comme indices d'abondance relative. Afin de rassembler des indices appropriés, des facteurs autres que ceux examinés dans le présent document devraient être inclus dans les modèles.

RESUMEN

Se calcularon las tasas de capturas estandarizadas utilizando un modelo mixto de ceros aumentados basado en la distribución Poisson para la flota brasileña. La base de datos contiene información acerca de los lances de palangre llevados a cabo por buques nacionales y fletados entre 1978 y 2010. Los modelos se ajustaron a un conjunto de datos agrupado, siendo el pabellón el factor explicativo, y a un conjunto de datos separado en función del pabellón. La captura en número de los peces era la variable de respuesta. El esfuerzo se consideró como compensación. El año y el número de anzuelos por cesta eran las variables explicativas más importantes. Las tasas de captura estandarizadas mostraban síntomas que sugieren que podrían no ser útiles como índices de abundancia relativa. Para reunir índices apropiados, deberían incluirse en los modelos otros factores que no sean los considerados en este documento.

KEYWORDS

Longlining, catch/effort, catchability, abundance

---

[1] UFRPE – DEPAq, Av. Dom Manoel de Medeiros s/n, Dois Irmãos, Recife-PE – Brazil. Email: humber.andrade@gmail.com

1. Introduction

Estimations of relative abundance indices are input data for several stock assessment models. The simple linear model often used to describe the relationship between abundance (N) and the relative abundance indice (I) in t^{th} year is $I_t = q_t N_t$. In this case N is the abundance in number of fish but biomass could be used as well. If catchability coefficient q_t changes across the years, I is not an acceptable relative abundance indice.

Catchability coefficient (q_t) can change due to several factors related to environment, fishermen behavior, fishing strategy and fishing gears (Cooke and Beddington, 1984; Hilborn and Walters, 1992). For example, if in a given year most of effort is concentrated in a given area, overall q in that year will be different from those of the years in which the effort was concentrated in another area. In that case usefulness of I as relative abundance indice is debatable unless methods used in the analyzes of data allow for removing or, at least, minimizing the effect of factor “area”.

Catch-per-unit-effort (CPUE) of fishing fleets and catch rates of scientific surveys have been often used to estimate relative abundance indices. Scientific surveys may result in less biased estimations because the experimental design may be enough to cope with some of the factors affecting q. In opposition, CPUE of fishing fleet are more biased in the sense they reflect changes in abundance (N) but also changes of fishing strategies and gears aiming at increase catch rate of target species (*e.g.* valuable species). Although catch rates (or CPUE) of fishing fleets are not ideal they are often used because in some circumstances, they are the only information available. That is true for albacore (*Thunnus alalunga*) caught by Brazilian longline fleet, which is the case studied in this work.

There are several approaches to analyze catch rate and CPUE data in order to calculate indices affected mainly by changes of the biomass than by changes of other factors. Those indices are usually denominated “standardized catch rates”. Generalized linear models (GLM) have been often used to “standardize” commercial catch rates (CPUE) and, will be also used in this paper. In the GLM framework a response variable (catch or catch rate) is assumed to follow a probability distribution of the exponential family. One function (monotonous and differentiable) links the response variable to explanatory variables in an additive linear structure. That is the link function. Explanatory variables may be qualitative (factors) or quantitative (covariates). Details about GLM theory and applications may be found in McCullagh and Nelder (1989) and Dobson (2002) and, one review of approaches using GLM is in Maunder and Punt (2004).

If the species of interest is also valuable for fishermen few longline sets will result in catches equal to zero (null catches). Nevertheless, if the species is not the target or, the abundance is not high or, the catchability is low, a large amount of null catches will occur. Albacore was probably the target of some of the longline chartered vessels based in Brazilian ports during some years in 1990s but most of Brazilian national and chartered vessels have been fishing swordfish, yellowfin, blue shark and bigeye tuna (Hsu and Chang, 1993; Arfelli, 1996; Meneses de Lima *et al.*, 2000). Hence, a large amount of null catches is expected for albacore.

When all catches are positive or when null catches are rare, models for positive data only may be used to estimate standardized indices. Also, null catches are allowed in some probability distributions like poisson for counting data. However, if the amount of zeros is large, the data is overdispersed (variance is larger than expected for that distribution of probability). Mixed or hurdle models may be used to model overdispersed counting data (*e.g.* Mullahy, 1986; Ridout *et al.*, 1998). The approach used to cope with zeros in hurdle models is to use Bernoulli or binomial distributions to model the proportion of positive catches, while some probability distribution truncated at zero like (*e.g.* truncated Poisson) is used to model positive catches. In the mixed models zero catch may arise from two distributions. The model for part of zero catches and for all positive catches is a Poisson or any other distribution for counting data. Bernoulli is then used to model the excess of zeros. Models used to cope with overdispersion caused by excess of zeros are often called zero-inflated models, specially the mixed ones.

After fitting the model and checking the fit, estimations of the coefficients and of the variance-covariance matrix are used to calculate values that hopefully are useful as relative abundance indices. Those calculations are based in the idea that variations of q due to changes in fishing grounds, technology and other factors will be represented by coefficients calculated for factors like “area”, “type of gear”, and so on. Coefficients calculated for the factor “year” are then expected to represent the effect of changes of abundance (in annual scale) on the indices (Maunder and Punt, 2004). This may be the case if all important factors affecting q are included in the model and coefficients for “year” arise as filtered signal of changes in biomass across years.

In this paper data from Brazilian longline fishing fleet is analyzed in an attempt to estimate standardized catch rates for albacore using a zero inflated generalized linear model. As mentioned above the factor “year” is assumed to represents the impact of annual changes in abundance on catches and catch rates. Because Brazilian fleet includes national vessels but also vessels chartered from several countries, comments about if “flag” of vessels should be included as factor or if, each database should be analyzed separated, are warranted.

2. Data and analysis

2.1 Database

Database analyzed is the “Banco Nacional de Dados de Atuns e Afins” (BNDA) mantained by the “Subcomitê Científico de Atuns e Afins” (SC-Atuns) that is part of “Comitê Permanente de Gestão de Atuns e Afins” (CPG) of the Ministry of Fisheries and Aquaculture of Brazil. That database contains information about longline sets carried out by Brazilian fleet (national plus chartered vessels) since 1978. It is not a complete database in the sense BNDA does not contain information about all sets but, the sample size is large and I have assumed it is enough to make inferences about catch rates.

Variables available in database are flag of the vessel, date, location and time when the longline was set into the water, number of hooks, number of hooks per basket and time when longline was retrieved. Soak time was calculated based on the difference between the time at the end of longline setting and the time at the end of longline retrieval. Time expended while setting the longline in the water was calculated as the difference between time at start and at end of longline setting operation.

Some of the chartered vessels have been fishing for few years. In this paper only data of national and chartered fleets that fished during a large number of years is considered (**Table 1**). Notice that most of information before 1990 are about Brazilian national fleet and vessels chartered from Japan. After 1990 several vessels were chartered from other countries but, since 2008 only Brazilian national vessels and vessels chartered from Spain were reported. Overall sample sizes are not balanced for factors year and flag in the sense the number of observations area quite different for each of the crossing levels. In opposition, the number of observations is balanced for factors quarter and flag except for vessels chartered from Saint Vincent that fished mainly in the first and fourth quarters (**Table 2**). Nevertheless total number observations in the four levels of factor quarter are similar.

One of explanatory variables considered in the generalized linear models was “area” with three levels as showed in **Figure 1**. Number of longline sets by crossing levels of factors area, flag and quarter are showed in **Table 3**. Motivations for selecting those three areas are the balance of design matrix as well as patterns found when analyzing spatial distribution of catch rate (see results section).

Total number of fishing sets is higher in north than in the other two areas where the number of observations were similar (**Table 3**). In north and east areas most of fishing sets were carried out in the austral summer (first quarter) and spring (fourth quarter) while in the south area most of fishing sets were carried out in austral autumn (second quarter) and winter (third quarter).

Most of fleets have fished mainly in the north area, except the one composed by vessels chartered from Japan that fished mainly in south area (**Table 3**). Among fleets that fished mainly in north area, BRA, BRA-HND and BRA-TAI fished more in the south than in the east area, while the opposite pattern is found for BRA-ESP, BRA-PAN and BRA-PRT.

2.2 Models

In the mixed models used in this paper the response variable (Y) has a zero-inflated Poisson (ZIP) distribution given by

$$(1) \qquad \Pr(Y = y) = \begin{cases} \omega + (1-\omega)\exp(-\lambda) & y = 0 \\ (1-\omega)\exp(-\lambda)\,\lambda^y/y! & y > 0 \end{cases}$$

In the above equation the response variable (Y) is the albacore catch (number of fish) in each of the fishing sets. Catches equal to zero are assumed to arise with probability ω plus probability $(1-\omega)\exp(-\lambda)$. The parameter λ is the mean of Poisson distribution. In this model the excess of zeros with respect to Poisson distribution is represented by the parameter ω. According to Ridout *et al.* (1998) the expectation for those mixed zero inflated Poisson models are

(2) $$E(Y) = (1-\omega)\lambda$$

while the variance is

(3) $$Var(Y) = (1-\omega)\lambda + \left(\frac{\omega}{1-\omega}\right)[(1-\omega)\lambda]^2$$

Regression models based in ZIP distributions have been used since mid 1980s (*e.g.* Mullahy, 1986). Lambert (1992) have used models as

(4) $$\log(\lambda) = X\beta \qquad log\left(\frac{\omega}{1-\omega}\right) = Z\gamma$$

where X and Z are design matrices for explanatory variables while β and γ are vectors of parameters. The two sets of covariates may or may not be the same. Alternative models can be generated by using different link functions for λ or ω other than log and logit. Welsh *et al.* (1996) provided solutions to calculate expections and variance for expectations based on information matrices of Poisson and Bernoulli models. Another alternative is to assume that the parameters estimations follows multivariate normal sample distribution, hence Monte Carlo approach can be used to calculate variance and confidence intervals for the expectations. That numerical approach was the one used in this paper.

The explanatory variables I have considered were:

- $year$ (factor);
- $flag$ (factor);
- hpb (covariate) ⇒ Number of hooks per basket;
- $quarter$ (factor);
- $area$ (factor) ⇒ Three levels as showed in **Figure 1**;
- $soak$ (covariate) ⇒ Soak time as calculated by the difference between time at the end of longline setting and at the end of longline retrieval;
- $dset$ (covariate) ⇒ Time expended when setting the longline as calculated by the difference between time at start and time at end of longline setting operation; and
- $pset$ (factor) ⇒ Period of the day when the longline set started. There are two levels: N (night) – before 9:00 or after 17:00 and D (day).

Three comments are warranted about the structure of the models I have considered: a) Besides main effects I have also considered first order interactions between all possible combinations between two of the explanatory variables; b) I have used two approaches to cope with the multiple fleets. In the first I have fitted models to complete data set including flag of vessels as explanatory variable. In the second approach the dataset was split by flag and then I have fitted eight different models; and c) The logarithm of number of hooks was included in the model as offset in the Poisson model.

In order to estimate the parameters I have used function $zeroinfl(\,)$ of the package pscl developed by Zeileis et al (2008) to run using R software (R Development Core Team, 2011). Estimations are by maximum likelihood using $optim(\,)$ (function of R). Variance-covariance matrix as standard errors are derived numerically using the Hessian matrix returned by $optim(\,)$.

2.3 Selection of variables and models

One alternative to select variables and models is to start with a simple model and, to increase the complexity by adding one term (an explanatory variable or an interaction). Thus some criterion like Akaike Information Criterion (AIC) (Akaike, 1974) or Bayesian Information Criterion (BIC) (Schwarz, 1978) is used to assess if the inclusion of the new term resulted in model improvement. For example, BIC and AIC criteria can point out which is the more parsimonious model based on the tradeoff between bias and variance (Burnham and Anderson, 2002). Nevertheless, because the data is not balanced, design matrix is not orthogonal and the estimations of parameters may be strongly correlated. Therefore the order the parameters enter the model affect the results as calculated based on those criteria. In order to cope with that issue I have started by assessing main effects using the following procedure:

i. Fit several simple models each of them including once at a time the main effect of each explanatory variable;

ii. Calculate the criterion of selection (BIC in this paper) for each of the simple models and select the more parsimonious among them as the “base” model;

iii. Fit several more complex models by adding each at a time the explanatory variables not included in the base model;

iv. Calculate the criterion of selection (BIC) for models fitted in the previous step;

v. If the minimum BIC is the one of the base model, stop the calculations and the base model is selected. Otherwise the new base model is the model with minimum BIC;

vi. Repeat steps *iii*, *iv* and *v* until achieving the stop condition or, until all explanatory variables have been included in the base model. The remaining model is the base model for main effects.

The procedure to cope with interactions is similar to those I have used for main effects:

i. Fit several models by adding each at a time all possible first order interactions between two of the variables included in the base model for main effects;
ii. Calculate BIC for each of the fitted models;
iii. If the minimum BIC is the one of the base model, stop the calculations and the base model is selected. Otherwise the new base model is the model with minimum BIC;
iv. Fit several models by adding each at a time all possible first order interactions between two of the variables included in the base model, except interactions already included in the base model;
v. Calculate BIC for each of the fitted models;
vi. If the minimum BIC is the one of the base model, stop the calculations and the base model is selected. Otherwise the new base model is the model with minimum BIC;
vii. Repeat steps *iv*, *v* and *vi* until achieving the stop condition or, until all possible first order interactions have been included in the base model. The remaining model is the base model for main effects.

3. Results

3.1 Catch, effort and catch rate

Total effort, catch and median of catch rate were calculated for aggregated data (all years and flags) by quarter (**Figure 2**). Although the data is pooled, some seasonal patterns are evident. Effort was concentrated in “north” (A in **Figure 2**) in spring and austral summer. Nevertheless there is a fishing spot in “south” (C in **Figure 2**) close to coast. Effort was concentrated in that fishing ground mainly in autumn and austral winter. Effort in “east” (B in **Figure 2**) was lower than in other areas.

Overall catch distribution reflects spatial distribution of effort but there are exceptions. Though effort was not concentrated in the area close to South American coast between 5^{o} S and 25^{o} S, the catches are somewhat high resulting in high catch rates in that region. That pattern is more evident in the first and fourth quarter. In addition, effort was not concentrated far from coast in the south area but the catches there were relatively high and, consequently the catch rates were also high. That pattern is more evident in the first quarter. Catch rates were also high in some spots in the east area where the effort is low.

Those three fishing areas were selected as they appear in **Figure 1** in an attempt to achieve some balance in the design matrix used in the models but, catch rates distribution was also a motivation. The fishing spot with high catch rates in south coast of South America is inside the “south” area while the large fishing spot close to coast in central coast of South America is inside the “north” area. and, The spreaded small spots with high catch rates in the mid of Atlantic Ocean are in “east” area. The pattern described above is especially evident in the top panel at most right column of **Figure 2**.

Most of the catch rates since 1978 were lower than 10 fish/1000 hooks except in the 1990s when catch rates were very high (**Figure 3**). Variability was also high in 1990s. In opposition, catch rates and its variation were especially low in the last ten years. Finally, most of boxplots indicate positive skewness of catch rate distributions. The pattern described above is consistent in the sense it appears for most of flags. Most of boxes as calculated for different flags in a given year are superposed. Hence catch rates are not quite different for most of fleets (national or chartered vessels).

3.2 Models

Large models with all explanatory variables were selected in all studied cases (pooled or split data). The order the exploratory variables were included in each of the selected models are in **Table 4**. In most of the cases the explanatory variables $year$ and hpb ranked first and second. Interactions were not showed to not clutter the table. Further description of the selected models is in **Table 5**. Results for models fitted to datasets of chartered vessels from Portugal and Saint Vincent are not in that table because the numeric algorithm did not converge.

All interactions and also the main effect hpb appearing in **Table 5** were discarded because the iterative numerical procedure used did not converge when they were included in the models. Notice that all interactions involving year were discarded. Interactions other than those involving $year$ were only dropped from models fitted to pooled dataset and to datasets of vessels chartered from Honduras and Panama.

The numbers of parameters are high in all models. The ratio between the number of estimations (k) and the number of observations (n) are close to or smaller than 0.01 in most of the cases. The model fitted to BRA-HND database is an exception. Although the number of parameters for BRA-HND case is the lowest, the ratio $k/n \cong 0.06$ is the highest. Those ratio calculations are not an objective criterion but they may be used to point for potentially overparameterized models.

Standard diagnostic plots of residuals are in **Figure 4**. All the smooth lines in dispersion diagrams of fitted versus residuals were close to zero (dashed lines in **Figure 4**). That is evidence that the models are not biased in the sense the expectation for residuals is close to zero whatever the fitted value. Exceptions arise in the analyses of BRA-ESP and BRA-HND (third and fourth row of panels in **Figure 4**) in which there were very large fitted values related to quite negative residuals. Nevertheless those large values driving smooth line far from zero were rare.

Distributions of residuals are not normal, but showed positive skewness because right tails of distributions are very heavy (**Figure 4**). That pattern is consistent in the sense it was evident in calculations for the seven databases. Distribution of residuals for database about vessels chartered from Panama (sixth row of panels from top to bottom **– Figure 4**) was the one more close to normal distribution.

3.2 Standardized catch rates

Standardized catch rates showed in **Figure 5** are scaled to the more often fishing scenario for each of the fleets. For example, BRA-JPN fleet has more often fished in second quarter, in “south” area, with effort close to 2300 hooks and so on; hence the standardized catch rates for that fleet are scaled to show the expected number of fish per longline set for that scenario. Solutions for any other scenarios would show similar profiles in the sense the only change would be the scale in the y-axis.

Overall standardized catch rates trend across the years were similar to those showed by nominal catch rates. High values appear in 1990s whatever the model and the database considered (**Figure 5**). The message ones get by looking at nominal or at standardized catch rates are the same. Other remarkable issue is the very large range of the confidence intervals for expectation of standardized catch rates for BRA-HND in 1997 and for BRA-JPN in 1993. As matter of fact those confidence intervals point that the data do not carry information about those two situations. Notice also the skewness of confidence intervals for the years 2009 and 2010. The expectation of standardized catch rates for those years are very close to the upper limit of the confidence interval in the calculations carried out with pooled database and with database of BRA and BRA-ESP fleets.

Standardized catch rates for vessels chartered from China-Taipei (BRA-TAI) are usually higher than those calculated for the other fleets (**Figure 5**). In opposition standardized catch rates as calculated for vessels chartered from Spain (BRA-ESP), Honduras (BRA-HND) and Panama (BRA-PAN) are low. One remarkable issue concerning the calculations is the large variability of the standardized indices. Whatever the database analyzed values for some years were two or three times higher than those calculated for nearby years. That is remarkable because those values should be relative abundance indices and it is very improbable that a fish populations experience such biomass variations.

Standardized indices calculated for pooled data seem to be a kind of "average" index because it show an intermediate time trend pattern if compared to those calculated for the split databases. If one relies in the standardized catch rates for pooled data as relative abundance indices, the conclusion would be that the biomass have increased from early 1980s until the end of 1990s and then, there were a sharp decrease in mid 2000s (**Figure 5**). Biomass did not recover since then though there was a small increasing trend after 2008. Once again it is important to stress that the validity of the above conclusion depends on the reliability of the standardized catch rates as relative abundance indices. Discussion on that issue is warranted in the next section.

Coefficients of correlation as calculated for standardized catch rates for pooled data and for database split by flag were all positive (**Figure 6**). Coefficients are high for BRA-JPN and BRA fleets while it is low for BRA-TAI. Hence the time trend showed by standardized catch rates calculated for pooled data reflects the time trend of values calculated for data of national vessels and of vessels chartered from Japan but not of vessels chartered from China-Taipei. As a matter of fact, standardized catch rates as calculated for vessels chartered from Taipei did not show any consistent time trend (**Figure 5**).

Coefficients of correlation were indeed high for BRA and BRA-JPN that have been fishing for a large amount of years (more than 20). In spite that result points for the conclusion that coefficients of correlations are high when the number of observations are high, that relationship is not so strong, because analyses for BRA-ESP and for BRA-PAN resulted in relatively high coefficients if compared to that calculated for BRA-TAI and, the number of observations are similar for those three fleets.

4. Discussion

Usefulness of standardized catch rates depends on the ability of being able to separate the impact of changes of abundance on catch rates from the impact of factors other than abundance. Ideally the impact of changes in abundance is represented by the coefficients calculated for levels of factor $year$. Nevertheless, if "year" is considered in interactions with other factors, extracting those separated "year" effects may be cumbersome and/or may require some subjective weighting the coefficients (Maunder and Punt, 2004). In this work that difficulty did not arise because iterative algorithm did not converge when considering $year$ in interactions. That convergence failure was probably due to the large number of parameters to be estimated because $year$ has many levels and, consequently, likelihood function may be flat in that situation.

Difficulties concerning convergence avoided several interactions to be included in the models. Models lacking factors and interactions could be biased if those terms are important to explain the variability of the response variable. Nevertheless, even though all those interactions were not included in the models, a couple of simple analyses of residuals pointed out the models are not biased. Further detailed analyses of residuals are required for sound conclusions but, the preliminary results gathered in this paper provide some evidence that those terms not included in models are not so relevant in the sense the models lacking them are not strongly biased.

The strong positive asymmetries of residuals are not of concern because, as a matter of fact, the assumed Poisson distribution may show asymmetry as well. Nevertheless, due to the Central Limit Theorem (CLT) residuals distribution should follow normal distribution approximately. Deviation from normal distribution may be signal that the overdispersion of catch data may be due not only to the large amount of zeros but also, due to a heavy right tail of distribution caused by some large catch observations. That issue should be investigated in the future in order to improve standardization models.

Both nominal catch rates and standardized catch rates scaled for the typical fishing scene for each fleet make evident that vessels chartered from China-Taipei catch albacore more efficiently. Indeed China-Taipei fishermen targeted albacore in the 1990s (Arfelli, 1996). In opposition, although albacore is a valuable tuna species, it was considered as a kind of "by-catch" by fisherman of the other fleets. Hence, there are two kind of standardized catch rates series, those from fleets that do not target albacore and one from the BRA-TAI fleet. Data of the BRA-TAI fleet are not a "random sample" in the sense that those vessels probably fished more often in scenarios in which the density or catchability of albacore were expected to be high. In opposition, other fleets were not aiming at albacore; hence in this sense those samples are expected to represent several situations and not only the ones in which albacore abundance or vulnerability are high. If both standardized series (BRA-TAI and all the others) showed similar time trends, that issue would be not of concern because the conclusion would be the same whatever the "sampler" fleet. Nevertheless, while standardized catch rates as calculated based on BRA-TAI data suggest any consistent time trend, overall, most of the other series suggest an increasing trend until the mid-1990s and then a sharp decreasing trend. The results gathered do not shed light on the discussion about which “sampler” generate more reliable data to calculate standardized indices, the fleet aiming at albacore or the fleets aiming at other species. As a matter of fact, both are biased and the quality of data is lower than those gathered in scientific surveys. Nevertheless, in some circumstances there is only information on commercial catch rate; hence they might be considered in the analyses. It should be stressed that different results may arise depending on the type of commercial data used in the analysis. Comparative studies on the reliability of standardized indices as calculated from “targeting sampler” (*e.g.* BRA-TAI fleet) and “by-catch sampler” (*e.g.* fleet not aiming at albacore) would be very useful.

One could argument that the solution is to put all data together and use $flag$ as factor in order to calculate standardized indices that convey all the available information. Nevertheless that is equivalent to assume that all databases are equally trustful. Maybe one of them (*e.g.* “targeting sampler” data) is less biased than the others and, to equally weight all databases would be a mistake. Moreover, put all together using $flag$ as factor will result in an intermediated standardized catch rates but, more similar to those catch rates of the largest database, which is not necessarily the more reliable one.

Hopefully the standardization calculations should result in values that reflect only the changes of biomass. Hence, time trend of standardized indices should be similar whatever the database used, unless, some important factors were not included in the models and, consequently, their impact on catch rate were not eliminated. Moreover, time trend of standardized catch rates should be biologically meaningful. For example, sharp increasing trends in a small amount of time, are very unlikely for some fish populations. Hence, when peaks show up, one should be concerned about the model because it is probably lacking factors that explain the variability of response variable. Finally, it is important to remember that the motivation to calculate standardized catch rates is the thought that nominal catch rate is biased. Hence, if the calculations were successful, the time trend of standardized catch rate series should be different from that of nominal catch rates. Overall, that was not the case for albacore caught by the Brazilian fleet. The standardized catch rates calculations showed the three symptoms mentioned above: (a) time trend differences related to the flag of the vessel; (b) doubtful biological meaning; and (c) time trend similar to those of nominal catch rate. Therefore, usefulness of those standardized catch rates as relative abundance indices should be carefully considered. The conclusion is that, in order to gather appropriate indices, other factors than those considered in this paper should be included in the models.

References

Akaike, H. 1974, A new look at the statistical model identification. IEEE T. Automat. Contr. 19: 716-723.

Arfelli, C.A. 1996, Estudo da pesca e aspectos da dinâmica populacional de espadarte, Xiphias gladius, L. 1758, no Atântico Sul. 175 f. Tese de Doutorado – Instituto de Biociências, Universidade Estadual Paulista, Rio Claro (SP).

Burnham, K.P. and Anderson, D.R. 2002, Model Selection and Multimodel Inference. 2nd Edition. Springer, New York. 488p.

Cooke, J.G. and Beddington, J.R. 1984, The relationship between catch rates and abundance in fisheries. IMA J. Math. Appl. Med. Biol. 1: 291-405.

Dobson, A.J. 2002, An introduction to Generalized Linear Models. Chapman and Hall/CRC, London. 225 p.

Hilborn, R. and Walters, C.J. 1992, Quantitative Fisheries Stock Assessment. Chapman and Hall, New York. 570p.

Hsu, C.C. and Chang, S.K. 1993, The adjusted longline CPUE of North and South Atlantic albacore stocks by three methods. Collect. Vol. Sci. Pap. ICCAT, 40(2): 333-342.

Lambert, D. 1992, Zero-Inflated Poisson regression, with application to defects in manufacturing. Technometrics. 34(1): 1-14.

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

McCulagh, P. and Nelder, J.A. 1989, Generalized Linear Models. Chapman and Hall, London. 513 p.

Meneses de Lima, J.H., Kotas, J.E. and Lin, C.F. 2000, A historical review of the Brazilian long-line fishery and catch of swordfish. Collect. Vol. Sci. Pap. ICCAT, 51(36): 1329-1358.

Mullahy, J. 1986, Specification and testing of some modified count data models. J. Econometrics, 33: 341–365.

R Development Core Team. 2011, R: A Language and Environment for Statistical Computing. http://www. R-project.org.

Ridout, M., Demétrio, C.G.B. and Hinde, J. 1998, Models for count data with many zeros. In: Proceedings of the XIX International Bio- metric Conference, Cape Town, South Africa, 14–18 December 1998. International Biometric Society, pp. 179–192.

Schwarz, G. 1978, Estimating the dimension of a model. Ann. Stat. 6: 461-464.

Welsh, A.H., Cunningham, R.B., Donnely, C.F. and Lindenmayer, D.B. 1996, Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecol. Model. 88: 297–308.

Zeileis, A., Kleiber, C. and Jackman, S. 2008, Regression Models for Count Data in R. J. Stat. Softw. 27(8). http://www.jstatsoft.org/v27/i08/.

Table 1. Number of longline sets by year and flag as reported in database "Banco Nacional de Dados de Atuns e Afins" (BNDA). BRA stands for national vessels while other columns contain information about vessels chartered from Spain (BRA-ESP), Honduras (BRA-HND), Japan (BRA-JPN), Panama (BRA-PAN), Portugal (BRA-PRT), China-Taipei (BRA-TAI) and Saint Vincent (BRA-VCT).

	Flag								
Year	*BRA*	*BRA-ESP*	*BRA-HND*	*BRA-JPN*	*BRA-PAN*	*BRA-PRT*	*BRA-TAI*	*BRA-VCT*	*Total*
1978	0	0	0	492	0	0	0	0	492
1979	16	0	0	464	0	0	0	0	480
1980	155	0	0	427	0	0	0	0	582
1981	160	0	0	305	0	0	0	0	465
1982	226	0	0	664	0	0	0	0	890
1983	179	0	0	439	0	0	0	0	618
1984	192	0	0	403	0	0	120	0	715
1985	173	0	0	297	0	0	0	0	470
1986	361	0	0	628	0	0	0	0	989
1987	506	0	0	421	0	0	0	0	927
1988	334	0	0	885	0	0	0	0	1219
1989	281	0	0	749	0	0	0	0	1030
1990	261	0	0	29	0	0	0	0	290
1991	135	75	0	289	0	0	491	0	990
1992	178	0	0	58	0	0	980	0	1216
1993	162	0	0	9	0	0	86	0	257
1994	173	0	0	333	0	0	257	112	875
1995	247	0	0	248	0	0	866	167	1528
1996	348	50	0	0	0	0	0	0	398
1997	473	481	3	0	0	0	0	137	1094
1998	465	734	43	24	0	0	124	657	2047
1999	685	1166	272	0	0	171	630	1584	4508
2000	1139	2611	291	34	246	323	782	1738	7164
2001	671	3124	253	13	126	88	887	2005	7167
2002	1396	1013	448	0	396	222	607	1158	5240
2003	501	1457	159	0	119	143	0	0	2379
2004	2384	1332	268	0	1755	36	0	0	5775
2005	2352	1426	355	0	2221	64	0	2	6420
2006	1689	1427	209	0	846	164	0	0	4335
2007	2213	658	168	0	178	0	0	0	3217
2008	181	6	3	0	0	0	0	0	190
2009	285	601	0	0	0	0	0	0	886
2010	146	193	0	0	0	0	0	0	339
Total	18667	16354	2472	7211	5887	1211	5830	7560	65192

Table 2. Number of longline sets by year and flag as reported in database “Banco Nacional de Dados de Atuns e Afins” (BNDA). BRA stands for national vessels while other columns contain information about vessels chartered from Spain (BRA-ESP), Honduras (BRA-HND), Japan (BRA-JPN), Panama (BRA-PAN), Portugal (BRA-PRT), China-Taipei (BRA-TAI) and Saint Vincent (BRA-VCT).

	Flag								
Quarter	*BRA*	*BRA-ESP*	*BRA-HND*	*BRA-JPN*	*BRA-PAN*	*BRA-PRT*	*BRA-TAI*	*BRA-VCT*	*Total*
1	4593	4319	690	1337	1692	269	1723	3280	17903
2	4627	4726	632	2358	1068	345	845	749	15350
3	4557	3624	518	2130	1534	363	1104	788	14618
4	4890	3685	632	1386	1593	234	2158	2743	17321
Total	18667	16354	2472	7211	5887	1211	5830	7560	65192

Table 3. Number of longline sets by area, quarter and flag as reported in database “Banco Nacional de Dados de Atuns e Afins” (BNDA). BRA stands for national vessels while other columns contain information about vessels chartered from Spain (BRA-ESP), Honduras (BRA-HND), Japan (BRA-JPN), Panama (BRA-PAN), Portugal (BRA-PRT), China-Taipei (BRA-TAI) and Saint Vincent (BRA-VCT).

	Area			
Quarter	*North*	*East*	*South*	*Total*
1	11026	5006	1871	17903
2	8397	2022	4931	15350
3	8286	1174	5158	14618
4	11614	3410	2297	17321
Total	39323	11612	14257	65192
Flag				
BRA	12884	1148	4635	18667
BRA-ESP	9862	6090	402	16354
BRA-HND	1375	268	829	2472
BRA-JPN	444	955	5812	7211
BRA-PAN	5004	858	25	5887
BRA-PRT	770	346	95	1211
BRA-TAI	3500	871	1459	5830
BRA-VCT	5484	1076	1000	7560
Total	39323	11612	14257	65192

Table 4. Main effects of the selected models. Dots at right of equations stand for interactions.

Database	*Model*
BRA	$catch \sim year + hpb + dset + quarter + soak + area + pset + \cdots$
BRA-ESP	$catch \sim year + hpb + soak + area + pset + quarter + dset + \cdots$
BRA-HND	$catch \sim year + hpb + soak + dset + quarter + area + pset + \cdots$
BRA-JPN	$catch \sim year + quarter + area + hpb + soak + dset + pset + \cdots$
BRA-PAN	$catch \sim year + soak + quarter + dset + pset + area + \cdots$
BRA-TAI	$catch \sim year + quarter + area + hpb + pset + dset + soak + \cdots$
Pooled	$catch \sim year + hpb + flag + dset + soak + quarter + pset + area + \cdots$

Table 5. Selected zero inflated mixed poisson models when analyzing the pooled data and datasets split base on flag: national vessels (BRA), vessels chartered from Spain (BRA-ESP), Honduras (BRA-HND), Japan (BRA-JPN) and China-Taipei (BRA-TAI). Sample size (n), number of parameter estimations (k) and Bayesian Information Criterion (BIC) are in the three columns at right. Explanatory variables are: year, f lag, hpb – Number of hooks per basket, quarter, area – north, south and east fishing grounds, soak – soak time, dset – Time expended when setting the longline and pset – Period of the day (day or night) when the longline was set.

	Discarded Terms				
Dataset	*Main*	*Interactions*	n	k	*BIC*
BRA	$flag$	$year: hpb, year: dset, year: quarter, year: pset, year: area, year: soak$	10091	144	154703.7
BRA-ESP	$flag$	$year: hpb, year: dset, year: quarter, year: pset, year: area, year: soak$	11956	114	141411.6
BRA-HND	$flag$	$year: hpb, year: dset, year: quarter, year: pset, year: area, year: soak$ $quarter: area, area: pset, soak: pset, dset: pset, hpb: soak$	1273	82	11451
BRA-JPN	$flag$	$year: hpb, year: dset, year: quarter, year: pset, year: area, year: soak$	7189	124	129239.6
BRA-PAN	$flag, hpb$	$year: hpb, year: dset, year: quarter, year: pset, year: area, year: soak$ $soak: area, quarter: area, pset: area, soak: dset, soak: pset, dset: pset$	5707	54	23465
BRA-TAI	$flag$	$year: hpb, year: dset, year: quarter, year: pset, year: area, year: soak$	5822	104	113144.7
Pooled		$year: hpb, year: flag, year: dset, year: soak, year: quarter, year: pset$ $year: area, hpb: soak, dset: soak, soak: quarter$	47289	278	824060.1

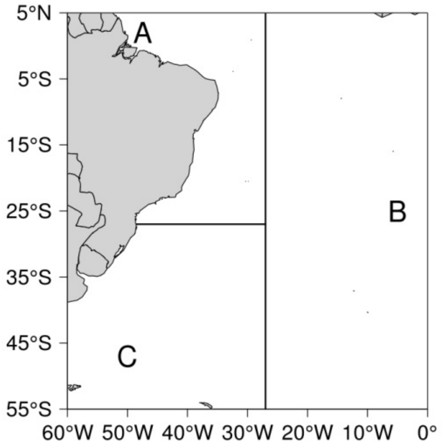

Figure 1. Levels of factor area used as explanatory variable in the generalized linear models. (A) north, (B) east and (C) south areas.

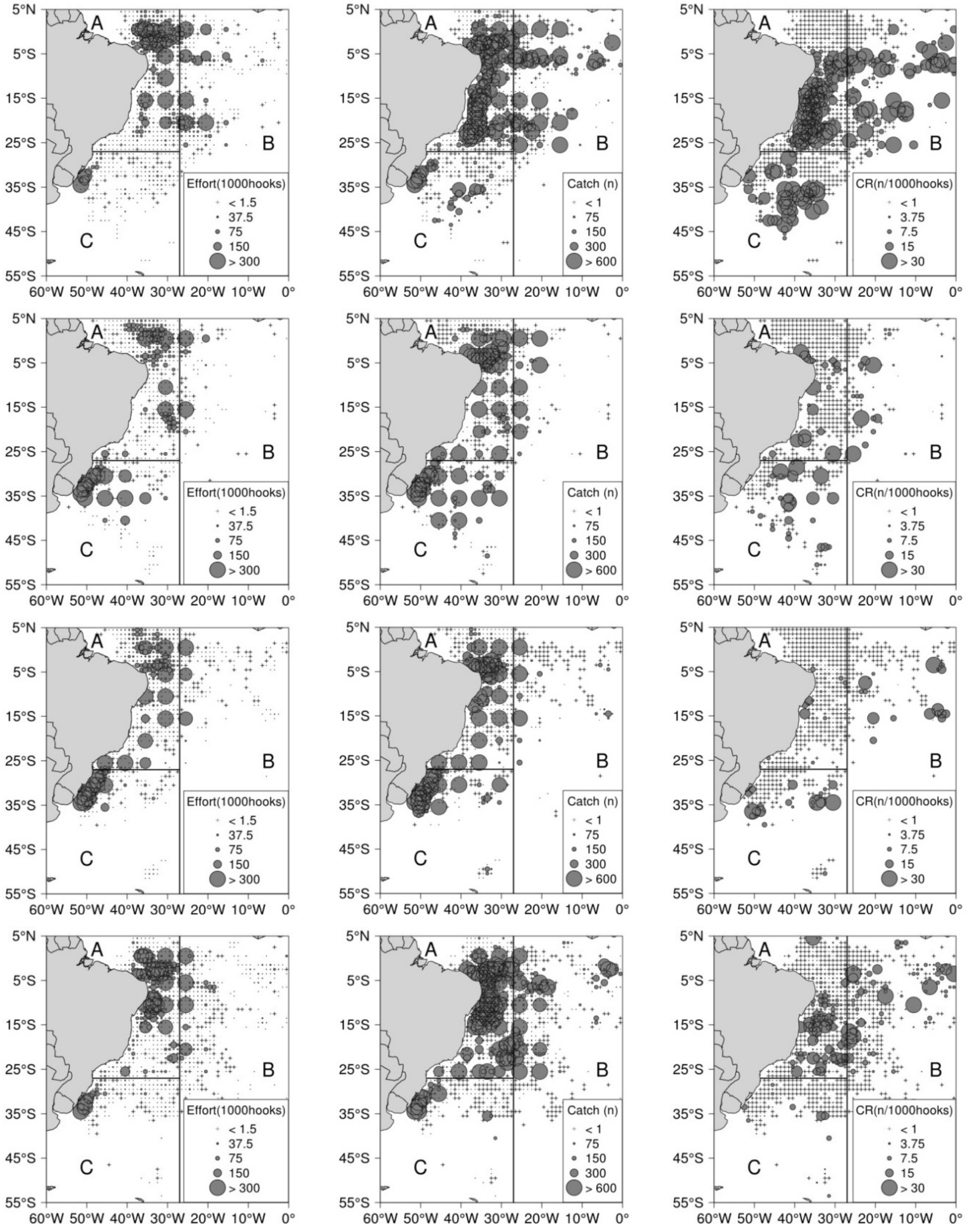

Figure 2. Effort (1000 hooks), catch (n) and median of catch rate (CR) (n/1000 hooks) per quarter as calculated for pooled data (all years and flags). Letters in the panels stand for areas denominated as north (A), east (B) and south (C). Panels from top to bottom stand for first to fourth quarter. Source: "Banco Nacional de Dados de Atuns e Afins" (BNDA).

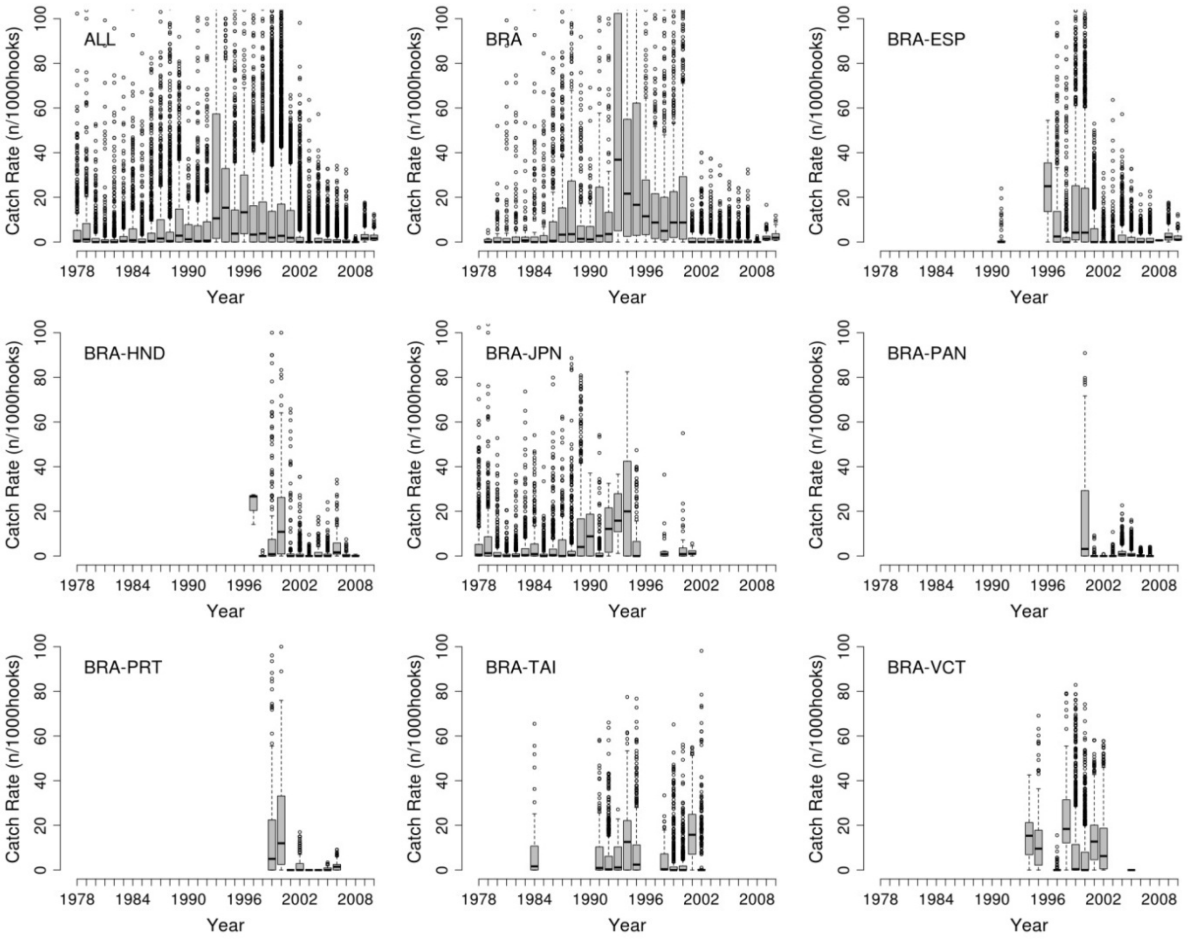

Figure 3. Catch rate (n/1000 hooks) per year. All flags pooled (ALL), Brazilian vessels (BRA), vessels chartered from Spain (BRA-ESP), Honduras (BRA-HND), Japan (JPN), Panama (PAN), Portugal (PRT), China-Taipei (TAI) and Saint Vincent (VCT).

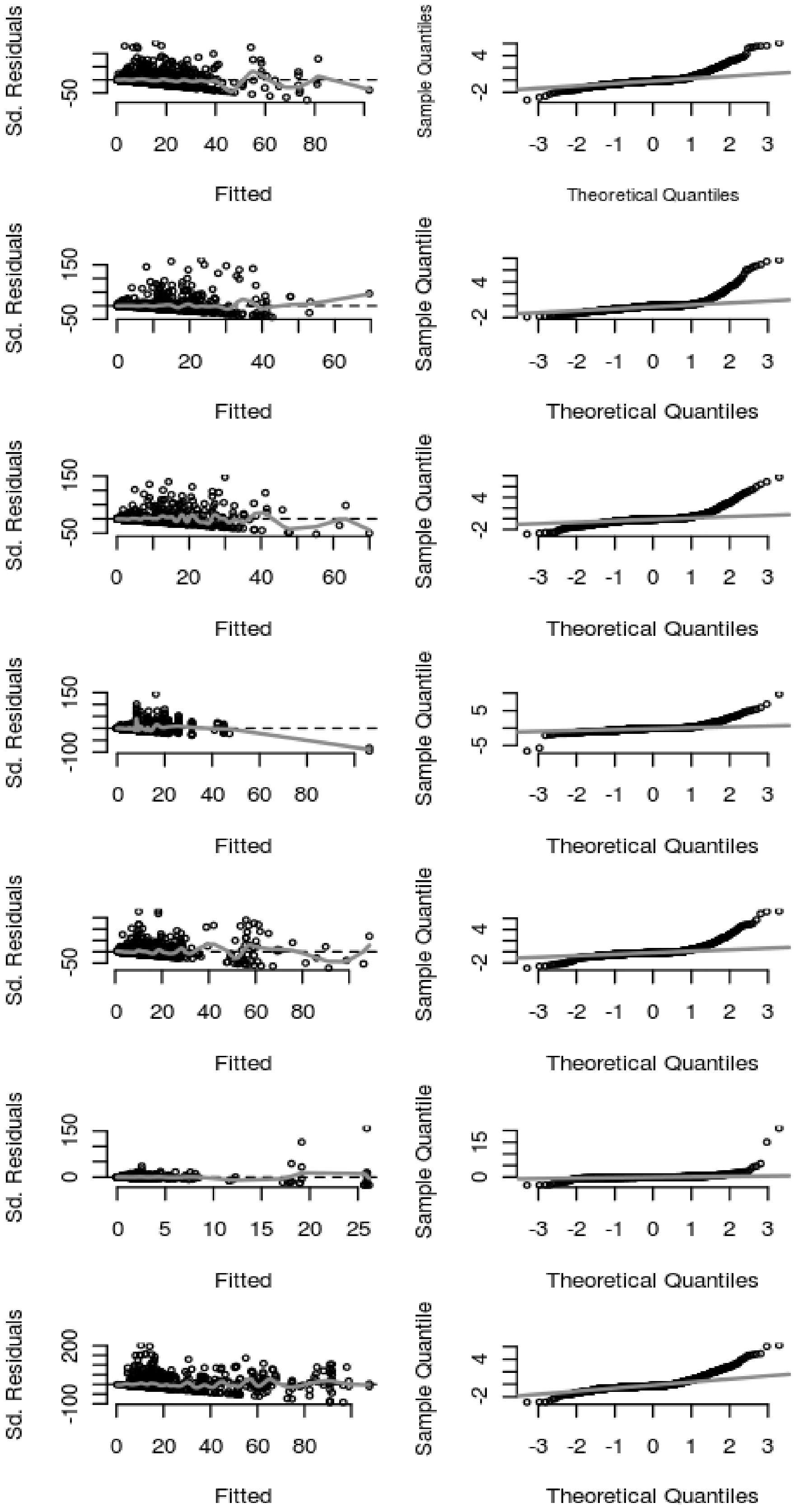

Figure 4. Fitted versus residuals (panels at left) and normal quantiles plots for standardized residuals. Panels from top to bottom show the results for models fitted to seven databases: a) all flags pooled, b) Brazilian national vessels, c) vessels chartered from Spain, d) vessels chartered from Honduras, e) vessels chartered from Japan, f) vessels chartered from Panama and g) vessels chartered from China-Taipei.

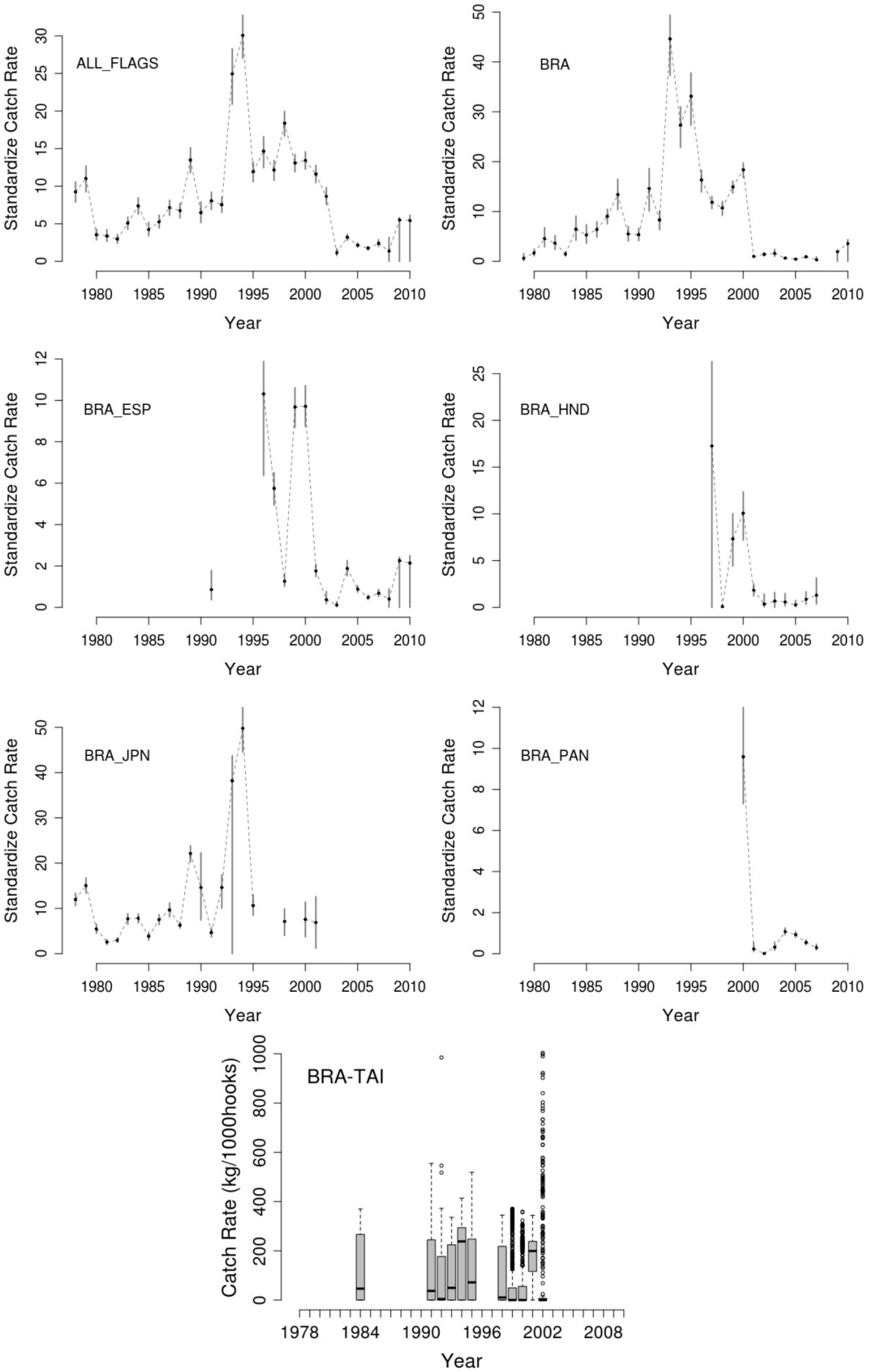

Figure 5. Standardized catch rate as calculated for all flags pooled (ALL.FLAGS), Brazilian vessels (BRA), vessels chartered from Spain (BRA-ESP), Honduras (BRA-HND), Japan (JPN), Panama (PAN) and China-Taipei (TAI). Filled circles stand for the punctual estimation of expectation, vertical gray lines stand for 99% confidence intervals.

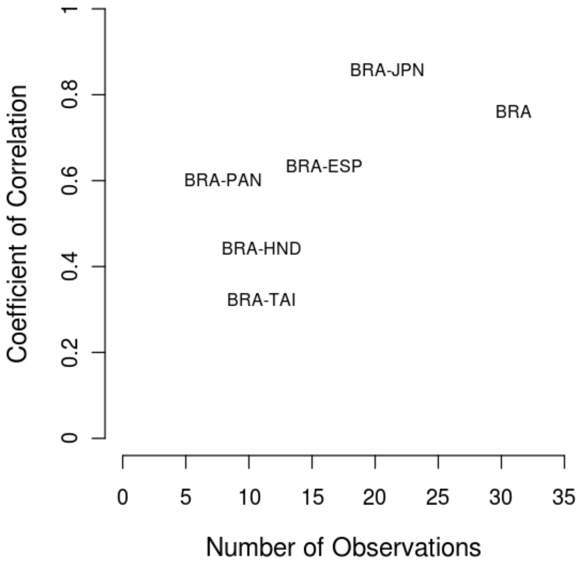

Figure 6. Number of observations and coefficients of correlation between standardized catch rates as calculated for all data pooled and standardized catch rates as calculated for national vessels (BRA), for vessels chartered from Honduras (BRA-HND), Japan (BRA-JPN), from Panama (BRA-PAN), from China-Taipei (BRA-TAI) and from Spain (BRA-ESP).