Received: 6 October 2015

IOTC–2015–WPTT17–24

Standardization of distant water tuna longline Hooking Rate for Yellow fin tuna (*Thunnus albacares*) from Fishery Survey of India Fleet (1981-2012)

AUTHORS : D.K Gulati* and Premchand**

*** Zonal Director, Fishery Survey of India, Cochin Base, Kochi**

**** Director General, Fishery Survey of India, Mumbai**

Abstract

Generalized linear model (GLM) is commonly used to evaluate impacts of environmental as well as fisheries operational variables on fisheries catch per unit fishing effort (CPUE) and to arrive at standardized CPUE which could be used as a relative index in fisheries stock abundance. GLM analysis is an effective way of standardization of CPUE data with catch rates in which there is a high proportion of zeros in the catch data. This paper describes a method for the analysis of yellow fin survey data, incorporating zero and non-zero values into a single model. The database contains information on the long line sets carried out by survey vessels of FSI from 1981 to 2012. The catch in number of fish per 100 hooks was the response variable. The Standardized hooking rates for yellow fin tuna were derived by means of GLM approach. Ten variables, Year, Quarter, Latitude, Longitude, duration (soaking time), catch rate of sailfish, skipjack and marlin, gear and Vessel Type were used to build GLM model. Eight variables were significant in GLM analysis which account for 26% of the variance in nominal CPUE. The purpose of this study is to obtain standardized hooking rate which can be used for yellow fin stock assessment.

Introduction

Yellow fin tuna (*Thunnus albacares*) is an important target species for the Indian long line fishery in the Indian Ocean. Estimations of relative abundance indices are input data for several stock assessment models. The catch and effort data for Indian Ocean yellow fin tuna for the Indian long line fishery was little used to study stock dynamics. The time-series of catch per unit effort (CPUE) for tunas seems to be appropriate for use as an abundance index for assessing the stock (Hsu, 1996; Uosaki, 1996). Moreover, the Generalized Linear Model (GLM) has been used to assess the tuna stock status as presented in many international expert forums in recent years. Catchability coefficient (q) can change due to several factors related to environment, fishermen behavior, fishing strategy and fishing gears (Cooke and Beddington, 1984; Hilborn and Walters, 1992). Catch-per-unit-effort (CPUE) of fishing fleets and catch rates of scientific surveys have been often used to estimate relative abundance indices. If, catch ability coefficient (q) changes across the years, *I* (abundance indices) is not an acceptable relative abundance Indices and scientific surveys may result in less biased estimations because the experimental design may be enough to cope with some of the factors affecting q.

The data of Catch per fishing unit (CPUE), is often used as a relative index of fisheries stock abundance (Hilborn and Walters, 2001; Nishida and Chen, 2004), and it is often influenced by many factors such as fishing capacity (e.g. characteristics of fleet, fishing gear and equipment), environmental factors (e.g. sea water temperature at varied depth, sea surface temperature, dissolved oxygen concentration, marine currents and lunar phases), and spatial and temporal factors (e.g. latitude, longitude, year and month). The CPUE needs to be standardized so that the impact of these factors can be removed or minimized and to got the better performance of CPUE reflecting the changes in population abundance only (Maunder and Punt, 2004).

There have many hypotheses about sub-stocks of Indian Ocean yellow fin tuna, i.e., two sub-stocks (Kikawa *et al*. 1970; Morita and Koto, 1970; Huang *et al*., 1973), three sub stocks (Kurogane and Hiyama, 1958), and four sub-stocks (Nishida, 1992). However, this study considers a single stock as well as western and eastern two major sub-stock hypotheses (Nishida, 1992).

MATERIALS AND METHODS

The standardized catch per unit of effort for the oceanic yellowfin (*Thunnus albacares*) was obtained by means of a General Linear Model (GLM) based on 6112 set records for the 1981-2012 periods. Sampling distribution is shown in (Figure 1). The number of zero catch observations was considerably high, a trait common with survey data in general. Modeling a species abundance based on a data set with a large proportion of zeros can invalidate the assumptions of the analysis and jeopardize the validity of model based inference if not properly accounted for (Lambert, 1992; Maunder and Punt, 2004). Furthermore, the use of 'nominal' CPUE (total catch divided by total effort) as an index of abundance assumes that all other factors affecting the catch rate have remained constant over time (Maunder and Punt, 2004). In reality, factors such as, season, gear type, location, boat size, fishing depth, and others are not constant. Therefore, the nominal catch data need to be standardized to account for the impact on catch rates of factors other than abundance. Standardization techniques can also account for the large proportion of zero observation catches, temporal changes in a species distribution, and the effect of changes in fishing efficiency. In the present study catch rates were modeled with an approach, incorporating zero and non-zero values into a single mode. The main objective of the analysis is to estimate a year effect. The year effect is used to represent the annual relative levels of abundance and is used as the relative index of abundance to include in the stock assessment. The year effect is included in the GLM as a categorical variable.

The lognormal error model is historically the most common method of CPUE standardization (Ortiz and Arocha, 2004). This model assumes that the natural logarithm of the catch rate is distributed normally. The natural logarithm of zero is undefined therefore these values must be dealt with; historically ignoring zero observations or replacing them by a constant (c) was the most common approach (Maunder and Punt 2004) ,the latter method is employed in this study.

In this study the data was analyzed using a lognormal error distribution for only the positive hauls (hauls in which the CPUE was non-zero) and by adding a constant equal to 0.1 to every observation from the FSI Observer Long line database. This value was chosen because it resulted in the most normal-like distribution of residuals (Butterworth 1996). The model provides an analysis technique where many problems usually associated with zero values are alleviated.

A primary goal of catch and effort analysis is to account for and explain as much variability in the data as possible. Fishery independent data (i.e. survey data) are often preferred for catch and effort analysis because standardized fishing and sampling procedures produce data that are designed to be compared from one year to the next.

The catch and effort data (number of hooks) used in the development of relative indices of abundance were collected through the surveys conducted by FSI during 1982-2012. This study is focused on standardizing longline vessels data from the FSI fleet operating in Indian EEZ of Indian Ocean and Bay of Bengal.

The data was provided by Fishery Survey of India and subsequently filtered to remove data

records with missing or faulty information. Records were excluded from the analysis if information on the operational variables was not recorded.

Among the operational variables recorded certain variables were recorded irregularly and thus dropped from the analysis. The final variables used for analysis were; catch (recorded in number of fish caught in operation) as a response variable, whereas, years, quarters, gear, vessel type, latitude(5*5 degrees) and longitude(5*5degrees), soaking time(5 hours strata), catch rate of sailfish, skipjack and marlin.

The selected final model was as follows:

glm(formula = log(hr.yftf. + 0.1) ~ factor(latstrat) + factor(year) + factor(q) + factor(lonstrat) + factor(vessel) + factor(soaking time) + factor(sailo) + factor(skpm) + factor(marl), data = datanj1, na.action = na.exclude)

Results and Discussion

In this paper, an attempt is made to evaluate the spatial distribution of yellowfin tuna caught in Indian and Bay of Bengal oceans around India with the aim of improving understanding of existing patterns of distribution that is a prerequisite for rational and sustainable management, which has to be based on sound scientific findings. The main objectives of this study are:

i) to quantify some of the explanatory variables accounting for explained variability such as year, quarter(fishing seasons), space (over the latitudes and longitudes), soaking time, vessels, gears and proportion of other species caught in the long line operation on the response variable, that is on the expected hooking rate.

ii) to understand the variability in the survey hooking rate obtained from the FSI's vessels over the past 30 years.

The GLM technique has been used to estimate the effective extents of various factors because GLM allows identification of the factors that influence catch rates as well as computation of standardised catch rates, represented by the year effect factor after taking into account the effects of other factors, which are used in many stock assessment methods.

The analysis presented here is done in R, a free statistical software package. R is available from the Comprehensive R Archive Network (http://cran.r-project.org).

One of the main purposes of GLM analysis of catch rates is to provide year effects. For a main effect model, year effects can be derived from the coefficients by setting "options (contrasts=c ("contr. treatment", "contr.poly")) in the R package. However, the year effects could also be extracted with special care, if the model contains interaction term with year. The common approach of extracting the year effect alone from a log-linear model can be replaced by an integral of the fitted model over the entire scale under consideration. This approach, however, yields catch rate indices that are equivalent to the year effects when the model contains no interaction terms (Stefansson 1996). Therefore, only the main effect model was used to estimate the variation of the catch rates over year in this study. Model drop1 was used to compute all the single terms in the scope argument that can be added to or dropped from the model. Model was fitted and computed a table of the changes in fit.

In this paper, GLM method is used to standardize the nominal yellow fin tuna catch rate of long line fishery operating in the Indian Ocean and to evaluate impacts of a list of temporal, spatial and fisheries operational variables on the catch rate. The frequency distribution of the standardized residuals for all variables combined effects is close to that of the normal distribution as assumed in the model (Figure2). The histograms of standardized residuals (%) by year from GLM fitting show likely normality as assumed in the model (Figure3). The results of using the GLM analysis of variance (ANOVA) to examine the logged catch rate for differences among variables are shown in (Table 1).

All the variables except catch rates of sailfish, skipjack and marlin were significant in GLM analysis which account for 26 % of the variance in nominal CPUE. The CPUE variability is mainly attributed to the year and, secondly, to the factor quarter and followed by latitudes (Table1). The other factors such as longitudes, vessels and soaking period were also significant, although less important. The standardized CPUE of 0.16 fish per 100 hooks in 1982, had been gradually increased to its maximum of 0. 83 fish per hundred hooks in 1986, decreased again to 0.25 fish per hundred hooks in the year 1993. The standardized CPUE had been stable during 1999 - 2005 within a range between 0.22 and 0.29 fish per hundred hooks. The hooking rate decreased to a very low in past recent years. A standardized CPUE plot indicated that the standardized trend was less variable than the nominal (Figure4).

The nominal CPUE in 1986 and 1997 (4.11 and 1.5 fish per hundred hooks, respectively) are high compared to other years, which need to be further checked. These GLM results represent an appropriate beginning to this study, but analytical questions remain. At present, the intention is to include available variables in the GLMs and to seek other significant explanatory variables. The lognormal gaussian analyses revealed significant effects of several explanatory variables, but did not explain high percentages of the probability of negative catch for this species. Attempts to find other significant covariates are necessary. Standardized CPUE plot is more stable than the nominal CPUE for the species. This indicated that standardization removed some variability attributable to the explanatory variables.

Deviance analysis results are reported in Table 1. Based on its statistical significance as well as the percentage of the deviance explained by each factor, final model for catch in number of fish included eight factors, *year ,quarter, latitudes, longitudes, gears, soaking time, vessels and catch rate* of skipjack, marlin and sailfish

Diagnostic plots (residuals *vs.* fitted values and cumulative normalized residual plots) are shown in (Figure 5**)**. In general, residual patterns are not far from expected under the normal error distribution assumption, which suggests a reasonably good fit.

The data used in this study are highly unbalanced (Figure 6). The GLM in the R-package can take the unbalanced data into account, which can be a great advantage, but the precision would be improved by a more balanced design. Probably, the precision of the estimates of the coefficients would have increased considerably, had the other associated information collected in the surveys on wind direction, wind velocity, current speed, current velocity etc. been included in the GLM analysis. The precision would definitely have increased considerably had the survey been done in accordance with the planned survey scheme. The models can accommodate temporal and spatial variability as well as the variability of other categories such as gear type, vessel horse power, length of the vessel, skipper's skill and environmental factors. The environmental factors such as temperature, salinity and oxygen are basically responsible for the spatial and temporal distribution of the fish. Most of the FSI's long liners are well equipped with the oceanographic equipments. The environmental data available with FSI, may also be tried. The inclusion of these factors in the GLM analysis would definitely affect the estimates of the coefficients .This should be a concern in the future computational strategy.

References

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.

Hilborn, R., and Walters C. J. 2001. Quantitative fisheries stock assessment choice, dynamics and SWG-10-JM-06 uncertainty. Boston: Kluwer Academic Publishers.

Hilborn, R., Walters, C.J., 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Chapman and Hall, New York.

Hsu, C. C. (1996). Standardization catch per unit effort of Taiwanese longline fishery as abundance index of albacore stock in the Atlantic. ICCAT/SCRS/96/8O, 21 pp

Huang, H. Y. (1995). National report on Taiwanese longline fisheries in the Indian Ocean in 1994. FAO IPTP Collective Volume, 9: 47-48.

Kikawa, S., T. Koto and C. Shingu. (1970). The status of the tuna fisheries of the Indian Ocean as of 1968. Bull. Far Seas Fish. Res. Lab. S-series 2.

Kurogane, K. and Y. Hiyama. (1958). Morphometric comparison of the yellowfin tuna from six grounds in the Indian Ocean. Bull. Jap. Soc. Sci. Fish. 24: 487-494.

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

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

Morita, Y. and T. Koto. (1970). Some consideration on the population structure of yellowfin tuna in the Indian Ocean based on the longline fishery data. Bull. Far Seas Fish. Res. Lab. 4:125-140.

Nishida, T. (1992). Considerations of stock structure of yellowfin tuna (*Thunnus albacares*) in the Indian Ocean based on fishery data. Fish. Oceangr. 1(2): 143-152.

Nishida, T., and Chen, D. 2004. Incorporating spatial autocorrelation into the general linear model with an application to the yellowfin tuna (*Thunnus albacares*) longline CPUE data. Fish. Res. 70: 265-274.

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

Stefánsson ,G .1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches.*ICES Journal of Marine Science* 53: 577–588.

Uosaki, K. (1996). Updated standardized CPUE for albacore caught by Japanese longline fishery in the Atlantic Ocean, 1959-1995. ICCAT/SCRS/96/7I, 5p.

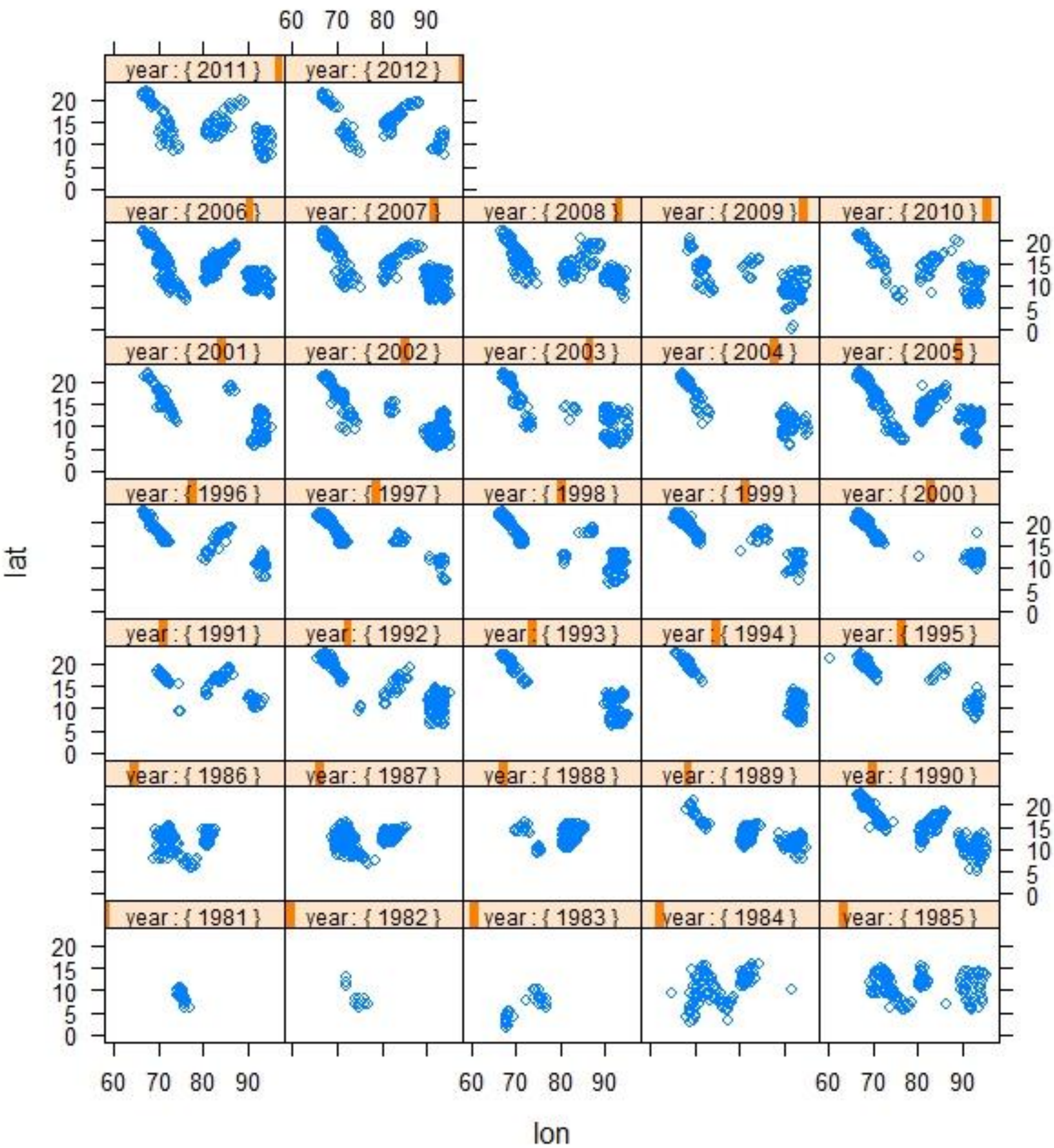
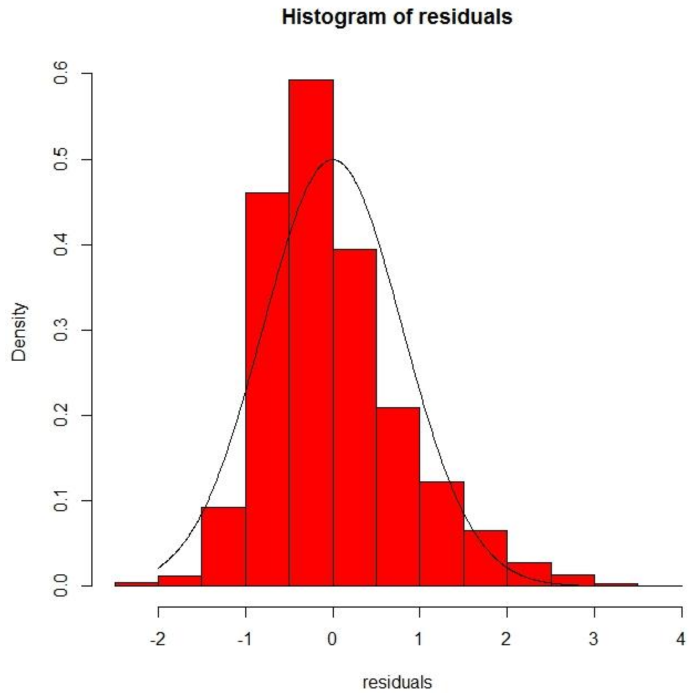
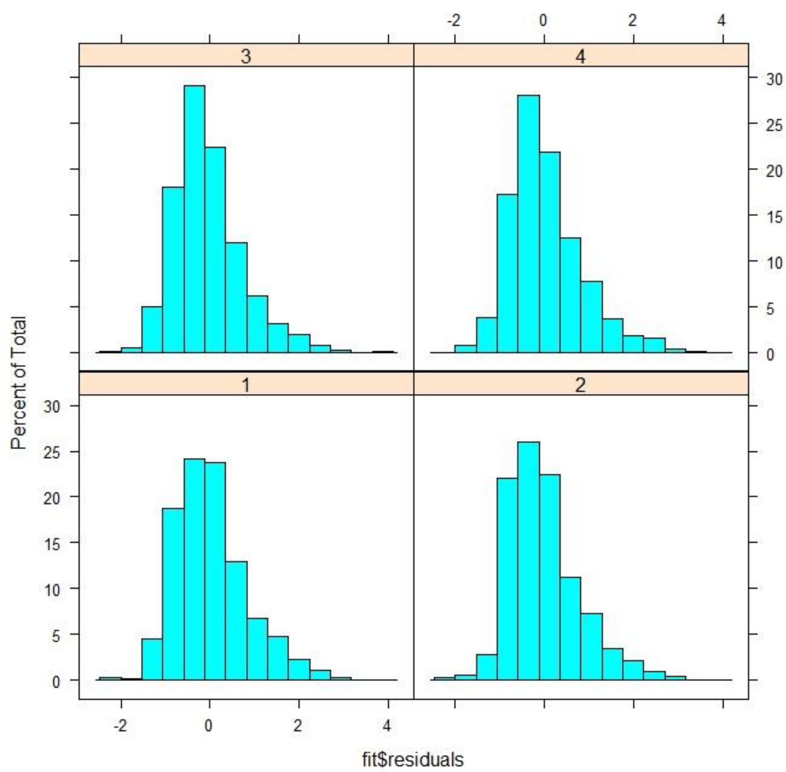


Figure 1. Longline stations sampled during the study period


Histogram of residuals
Density
residuals
0.0
0.1
0.2
0.3
0.4
0.5
0.6
-2
-1
0
1
2
3
4


3
4
1
2
Percent of Total
fit$residuals
0
5
10
15
20
25
30
-2
0
2
4

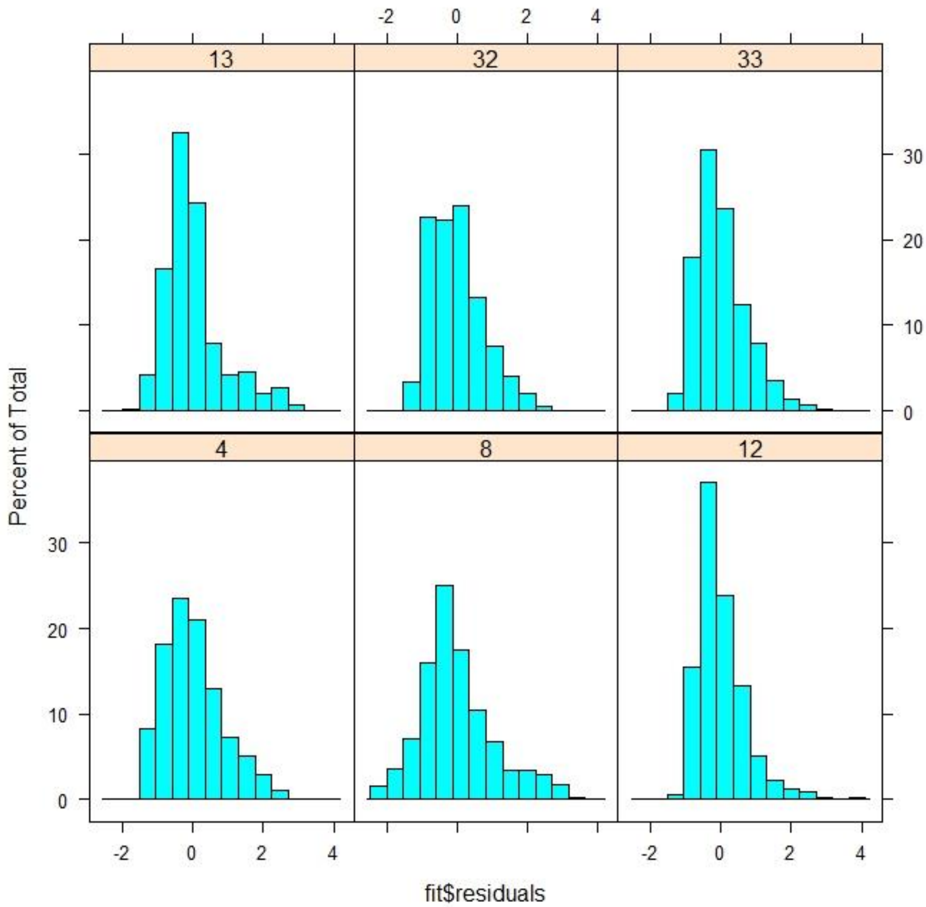


Figure 2. The frequency distribution of the standardized residuals by quarters for all variables combined effects

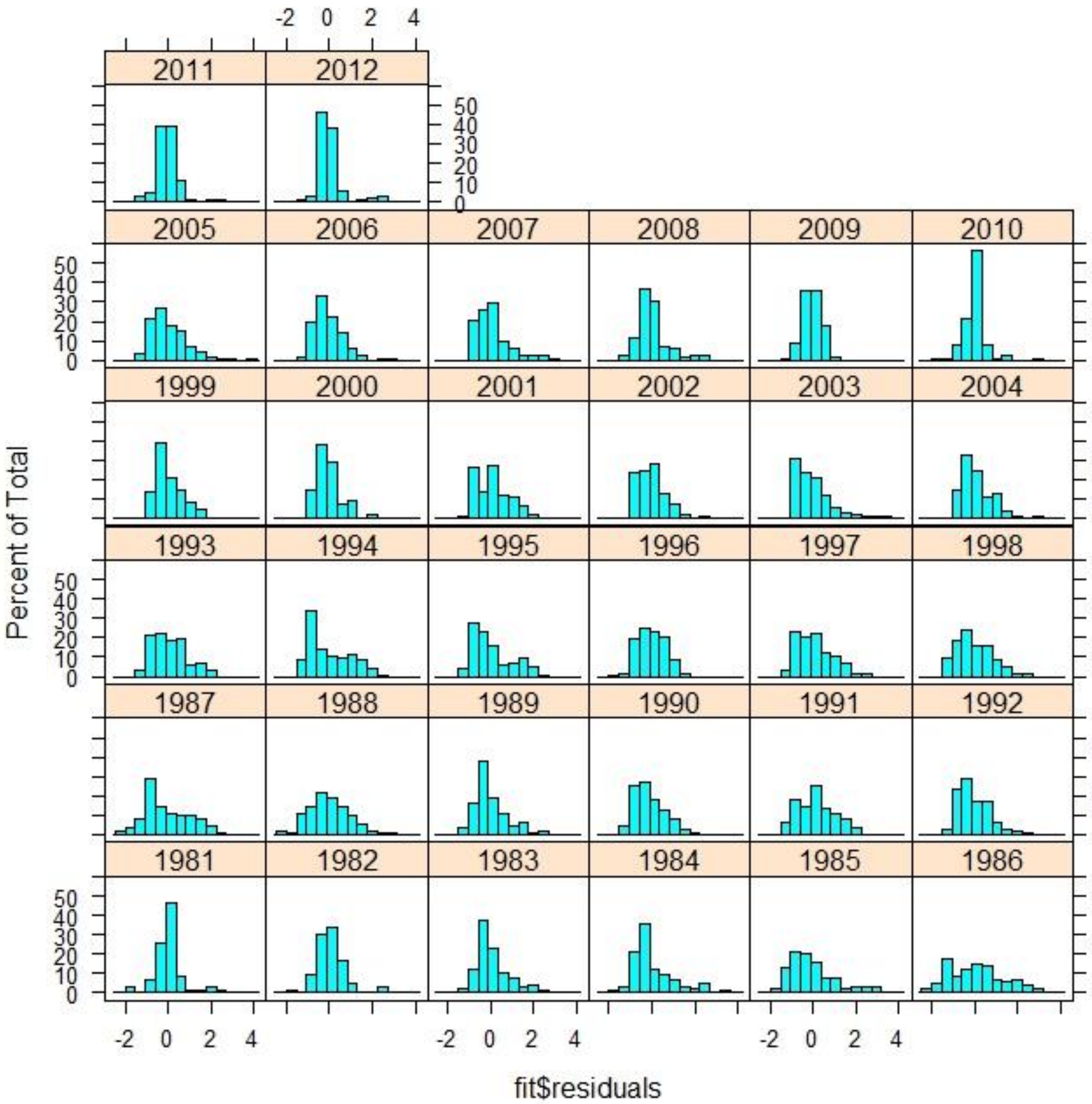


Figure 3. Frequency distribution of standardized residuals (%) by year from GLM fitting

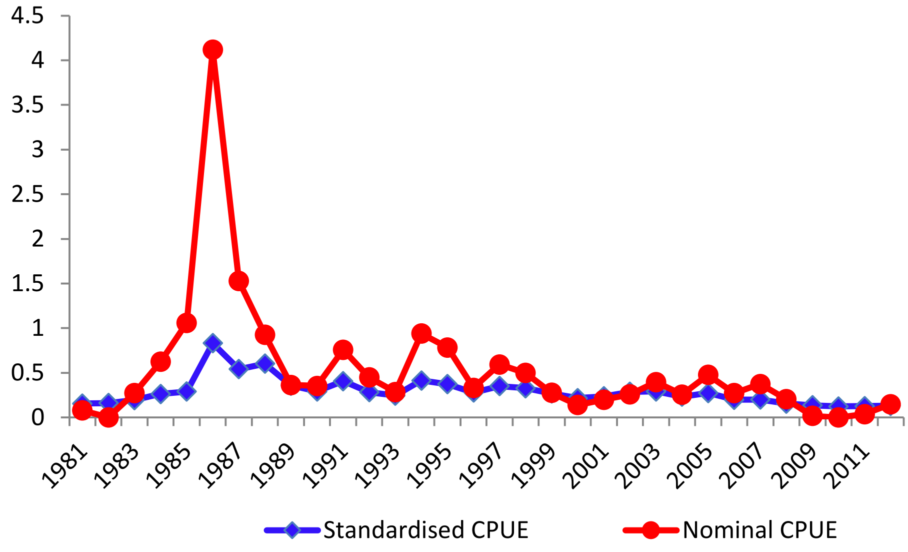


Figure 4. Standardized and nominal CPUE (Hooking Rate) for distant water tuna longline for yellow fin during the period 1981-2012

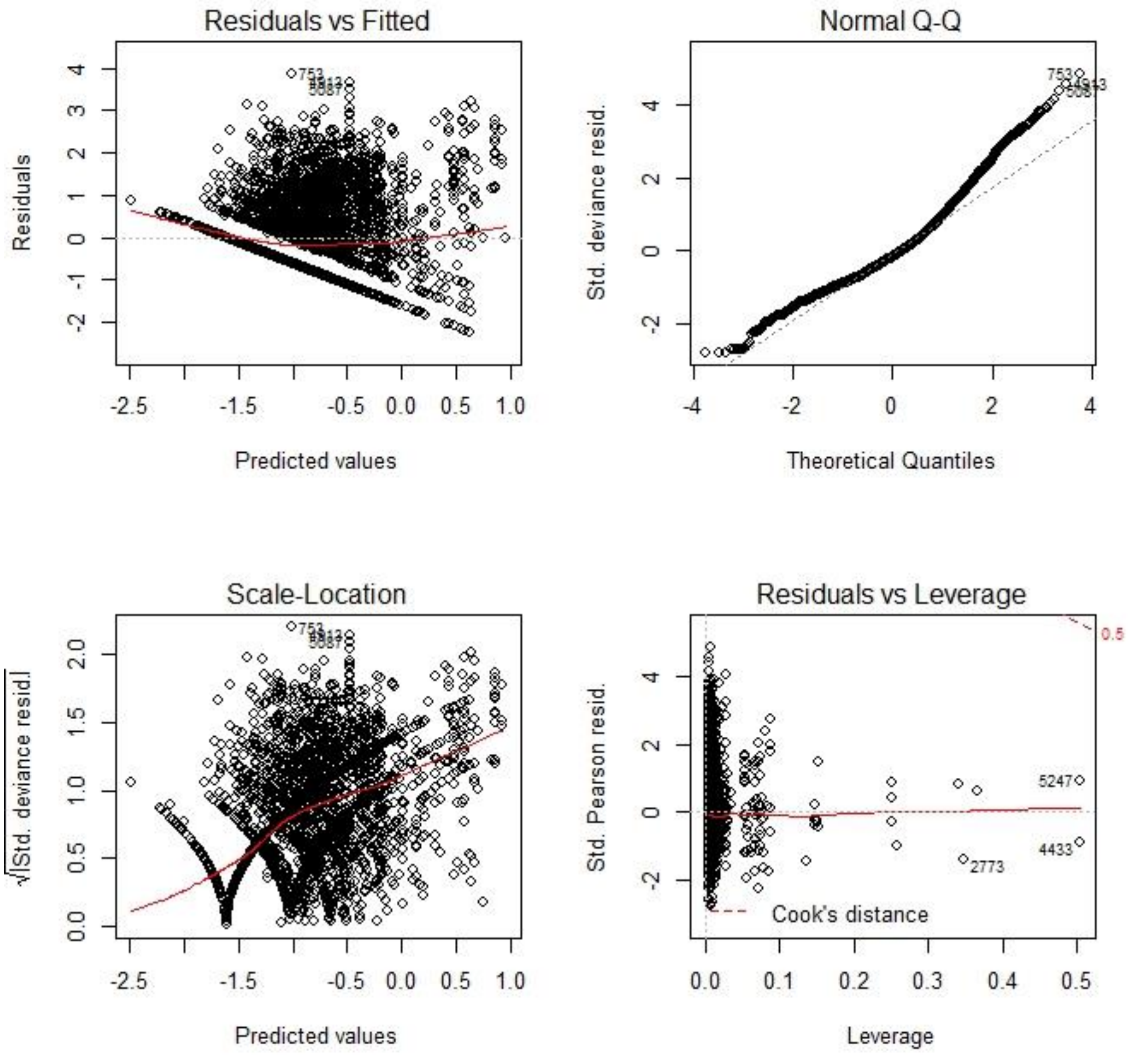


Figure 5. Distributions of the standardized residuals and the normal probability plots for the standardization model fitted to the catch (in number) and effort (hooks) data of FSI's fleet for the time series 1981-2012

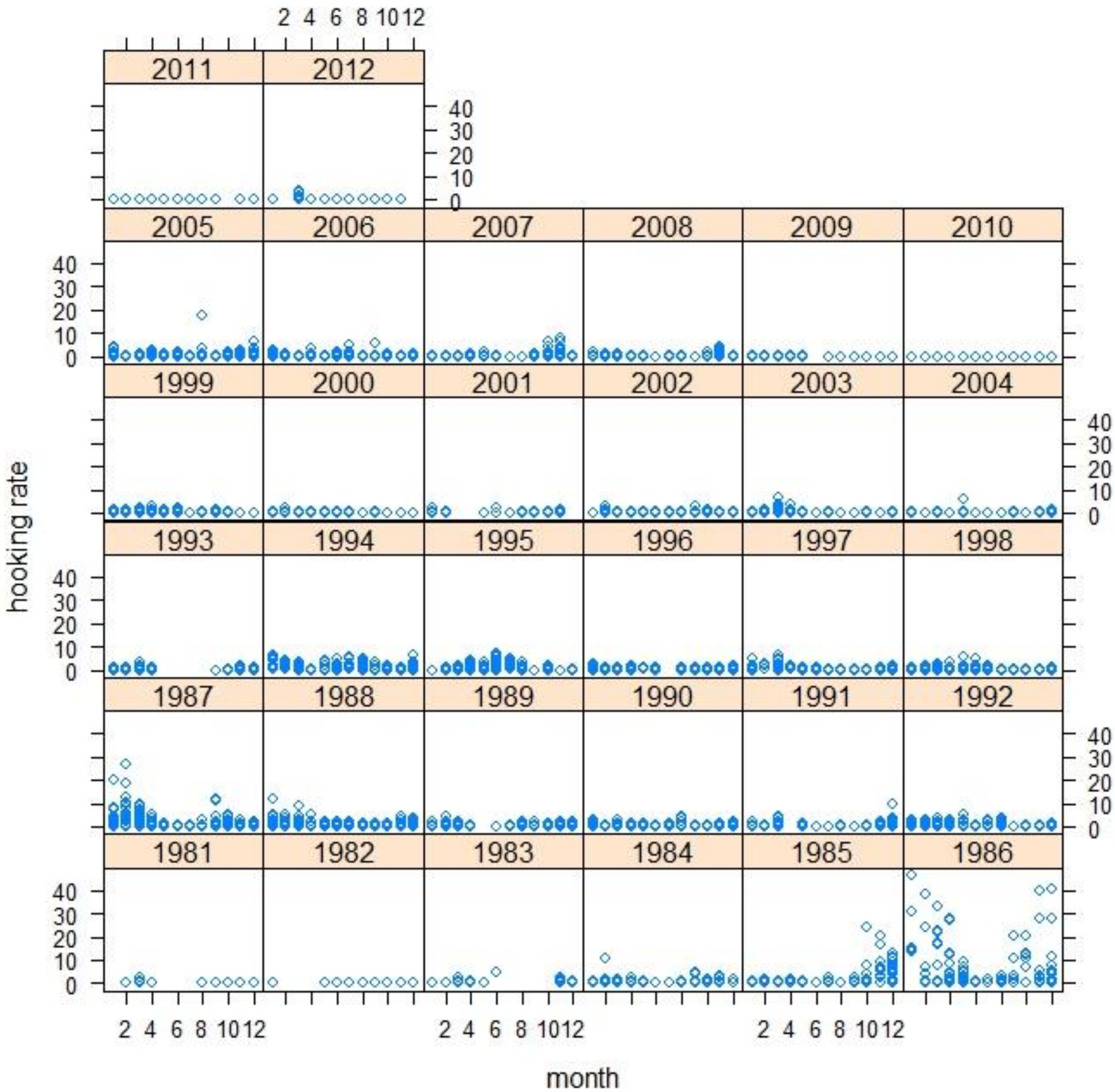


Figure 6. Month-wise and year-wise presentation of Hooking Rate of Yellow fin tuna

Table 1. Results of GLM analysis of variance (ANOVA)

	Df	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)
NULL			6109	5263.9		
factor(latstrat)	4	134.93	6105	5128.9	52.3323	< 2e-16 ***
factor(year)	31	696.54	6074	4432.4	34.8579	< 2e-16 ***
factor(q)	3	251.33	6071	4181.1	129.9699	< 2e-16 ***
factor(lonstrat)	7	89.67	6064	4091.4	19.8734	< 2e-16 ***

factor(vessel)	5	177.53	6059	3913.9	55.0828	< 2e-16 ***
Factor(soaking time)	3	5.52	6056	3908.4	2.8563	0.03569 *
factor(sailo)	3	2.79	6053	3905.6	1.4403	0.22893
factor(skpm)	3	5.92	6050	3899.6	3.0639	0.02692 *
factor(marl)	3	1.83	6047	3897.8	0.9484	0.41617

Significance codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Acknowledgements

The authors wish to express their sincere thanks to the Fishery survey of India, Ministry of Agriculture (Dept. of A.H; D & Fy) for utilizing the data and according permission to present the paper in the IOTC Sixteenth Working Party on Tropical Tunas (WPTT16) scheduled to be held in Bali, Indonesia from 15 to 19 November 2014.