Updated standardized CPUE of blue shark by Taiwanese large-scale tuna longline fishery in the Indian Ocean

Wen-Pei Tsai [1,3] and Kwang-Ming Liu [2]

[1] Department of Fisheries Production and Management, National Kaohsiung University of Science and Technology, Kaohsiung 811, Taiwan

[2] Institute of Marine Affairs and Resource Management, National Taiwan Ocean University, Keelung 202, Taiwan

[3] Corresponding author. Email: wptsai@nkust.edu.tw

SUMMARY

The blue shark catch and effort data from observers' records of Taiwanese large longline fishing vessels operating in the Indian Ocean from 2004-2017 were analyzed. Based on the nominal catch per unit effort (CPUE) distribution of the blue shark, four areas, namely, A (north of 10ºS, east to 70ºE), B (north of 10ºS, 70ºE-120ºE), C (south of 10ºS, 20ºE-60ºE), D (south of 10ºS, 60ºE-120ºE) were categorized. To cope with the large percentage of zero shark catch, the CPUE of blue shark, as the number of fish caught per 1,000 hooks, was standardized using a zero inflated negative binomial model. In general, the standardized CPUE series of the blue sharks caught by Taiwanese large-scale longline fishery showed a stable increasing trend (Figure 4). This sable trend suggested that the blue shark stock in the Indian Ocean seems at the level of optimum utilization during the period of 2004-2017. In this report, environmental effects were not included in the model for standardization. The results obtained in this study can be improved if longer time series of observers' data are available and environmental factors were included in the model.

KEYWORDS

Blue sharks, Taiwanese longline fishery, standardized CPUE, by-catch, observer programs, zero inflated negative binomial model

1. Introduction

The Taiwanese longline fishery has operated in the Indian Ocean since the late 1970s. However, the shark by-catch of Taiwanese tuna longline fleets was never reported in the logbook until 1981 because of its low economic value compared with tunas. During the period from 1981 to 2002, only one category “sharks” was recorded in the logbook. The category “sharks” in the logbook has been further separated into four sub-categories namely the blue shark, *Prionace glauca*, mako shark, *Isurus spp*., silky shark, *Carcharihnus falciformis*, and others since 2003. As the Taiwanese longline fishery has widely covered the Indian Ocean, our fishery statistics must be one of the most valuable information that can be used to describe the population status of pelagic sharks.

Blue shark is the major shark by-catch species of Taiwanese large longline fishery. Since FAO and international environmental groups has concerned on the conservation of elasmobranchs in recent years, it is necessary to examine the recent trend of sharks by examining the logbook of tuna fisheries. However, standardization of Taiwanese catch rate on sharks is not straightforward because the logbook data have been confounded with many factors, such as under-reporting, no-recording of sharks and target-shifting effects. Consequently, the observer program for the large longline fishery was conducted to obtain detailed and reliable data for more comprehensive stock assessment and management studies. Recently, the increase of coverage rate of observations enabled us to get a better estimation of shark by-catch. Therefore, it is useful to examine recent trends in relative abundance of the blue sharks using the most recent observer data in the Indian Ocean.

A large proportion of zero values is commonly found in by-catch data obtained from fisheries studies involving counts of abundance or CPUE standardization. The Zero-inflated modeling, which can account for a large proportion of zero values, is an appropriate approach to model zero-heavy data (Lambert 1992; Hall 2000). As sharks are common by-catch species in the tuna longline fishery, the zero inflated negative binomial (ZINB) model was therefore applied to address these excessive zeros of shark catch for CPUE standardization in this study. The CPUEs of blue sharks in the Indian Ocean were standardized using ZINB based on observers’ records data and hopefully the CPUE series can be used in the future blue shark stock assessment.

2. Material and methods

2.1. Source of data

The species-specific catch data including tunas, billfishes, and sharks from observers’ records in 2004-2017 were used to standardize CPUE of blue shark of Taiwanese longline fishery in the Indian Ocean. The summary of these data were shown in **Table 1**. The catch rate of blue sharks might be affected by spatial and temporal factors. We used the following stratification in our models. For temporal factors, we separated the data into 4 quarters: the 1st quarter (January to March), the 2nd quarter (April to June), the 3rd quarter (July to September), and the 4th quarter (October to December). For spatial stratification, based on the nominal CPUE distribution of the blue shark (**Fig. 1**), four areas, namely, A (north of 10ºS, east to 70ºE), B (north of 10ºS, 70ºE-120ºE), C (south of 10ºS, 20ºE-60ºE), D (south of 10ºS, 60ºE-120ºE) were categorized. The areas used in this study are shown in **Figure 2**. For standardization, CPUE was calculated by set of operations based on observers' records during the period of 2004-2017.

2.2. CPUE standardization

A large proportion of sets with zero catch of blue shark (~60%) was found in the observers' records. Hence, to address these excessive zeros, the Zero inflated Negative Binomial model (ZINB, Lambert 1992; Hall 2000) was applied to the standardization of blue shark CPUE. The ZINB is a mixture of two distributions, one distribution is typically a Poisson or negative binomial distribution that can generate both zero and nonzero counts, and the second distribution is a constant distribution that generates only zero counts. The model was fit using glm function of statistical computing language R (R Development Core and Team, 2013) to eliminate the biases by change of targeting species, fishing ground and fishing seasons.

The standardized CPUE series for blue shark was constructed without interaction effects. The main variables chosen as input into the ZINB analyses were year (Y), quarter (Q), area (A), and HPB (number of hooks per basket, HPB). The effect of gear configuration of HPB was used to account for the shift of targeting species. The model is described as:

Catch= Year + Quarter + Area + HPB

For the Zero Inflated Negative Binomial:
(Part 1: count models- Negative Binomial; Part 2: Binomial, link = logit)

The probability distribution of a zero-inflated negative binomial random variable Y is given by

$$\Pr(\mathrm{Y}=y)=\begin{cases}\omega+(1-\omega)(1+k\lambda)^{1/k} & \text{for } y=0\\ (1-\omega)\frac{\Gamma(y+1/k)}{\Gamma(y+1)\Gamma(1/k)}\frac{(k\mu)^y}{(1+k\lambda)^{y+1/k}} & \text{for } y=1,2,\ldots\end{cases}$$

where *k* is the negative binomial dispersion parameter.

The effect of gear configuration of HPB was categorized into the four classes of 1-9, 10-12, 13-14, and >=15, and 4 quarters were categorized: the 1st quarter (Jan-Mar), the 2nd quarter (Apr-Jun), the 3rd quarter (Jul-Sep), and the 4th quarter (Oct-Dec). The area strata used for the analysis were shown in **Figure 2**.

The best model for ZINB models were selected using the stepwise AIC method (Venables and Ripley, 2002). For model diagnostics, the rootograms function in R countreg package (Kleiber and Zeileis, 2016) was used to assess the influence of observations that exert on the model. The distribution of residuals was used to verify the assumption of the ZINB models. These diagnostic plots were used to evaluate the fitness of the models.

3. Results and discussion

The blue shark bycatch data are characterized by many zero values and a long right tail (**Figs. 3**). Overall, 57.39% of the total sets in the Indian Ocean had zero bycatch of blue sharks (**Table 2**). As a result, the following models with many explanatory variables were finally selected. The best models for ZINB models chosen based on AIC were " Catch ~ HPB + Year + Area ". The detail values for nominal and standardized CPUE were listed in **Table 3**. The nominal CPUE of blue shark showed an inter-annual fluctuation, particularly in year 2009 and 2014. However, this variability was reduced in the standardized CPUE series (**Figure 4**). The standardized CPUE series contains the combined effects from two models, one that calculates the probability of a zero observation and the other one estimates the count per year. In general, the standardized CPUE series of the blue sharks caught by Taiwanese large-scale longline fishery showed a stable increasing trend (**Figure 4**). This stable trend suggested that the blue shark stock in the Indian Ocean seems at the level of optimum utilization during the period of 2004-2017.

The diagnostic results from the ZINB model do not indicate severe departure from model assumptions (**Figures 5-7**). The Q-Q normal plots (the upper panel) for ZINB model showed that the error distributions are close to normal (**Figure 5**). There is also no wave-like pattern for the residuals showed that the data is appropriately captured by the model. Additional residual plots for each factor were provided in **Figures 6-7**. The ANOVA tables for each model are given in **Tables 4-5**. Most main effects (except for Quarter factor) tested were significant (mostly $P < 0.01$) and included in the final model. However, the deviance tables were not provided in this study. The standard residuals (e.g., Poisson or deviance) are often not so informative because they mostly capture the modeling of the mean but not of the entire distribution. In case of discrete distributions (as ZINB). To verify that the model solves the problem of excess zeroes, another alternative which checks the marginal distribution of the data is the so-called rootogram. This plot (**Figure 5**) often better at displaying the problems of excess zeros and/or overdispersion than Q-Q plots of randomized quantile residuals (Kleiber & Zeileis, 2016). However, other factors may affect the standardization of CPUE trend. In addition to the temporal and spatial effects, environmental factors are important which may affect the representation of standardized CPUE of pelagic fish i.e., swordfish and blue shark in the North Pacific Ocean (Bigelow *et al.,* 1999), and big-eye tuna in the Indian Ocean (Okamoto *et al.,* 2001). In this report, environmental effects were not included in the model for standardization. The results obtained in this study can be improved if longer time series of observers' data are available and environmental factors were included in the model.

References

Bigelow, K.A., Boggs, C.H., and He, X., 1999. Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. Fish. Oceanogr. 8(3): 178-198.

Cook, R. D., and Weisberg, S., 1982. Residuals and influence in regression, New York, NY: Chapman and Hall.

Efron B., Tibshirani R.J., 1993. An introduction to the bootstrap. London: Chapman and Hall.

Hall, D. B. 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics 56:1030–1039.

Kleiber and Zeileis 2016. Visualizing Count Data Regressions Using Rootograms. The American Statistician, 70(3), 296–303

Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34:1–14.

Okamoto, H., Miyabe, N., and Matsumoto, T., 2001. GLM analyses for standardization of Japanese longline CPUE for bigeye tuna in the Indian Ocean applying environmental factors. IOTC Proceedings 4: 491-522.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Venables, W. N. and Ripley, B. D. 2002. Modern Applied Statistics with S. Fourth edition. Springer.

Walsh, W. A., 2011. Ongoing and Planned Analyses of Catch and Catch Rate Data for Blue Shark Prionace glauca and Shortfin Mako Isurus oxyrinchus in the Hawaii-based Pelagic Longline Fishery: 1995–2010," ISC/11/SHARKWG-2/2. La Jolla, California USA. 28 Nov - 3 Dec 2011.

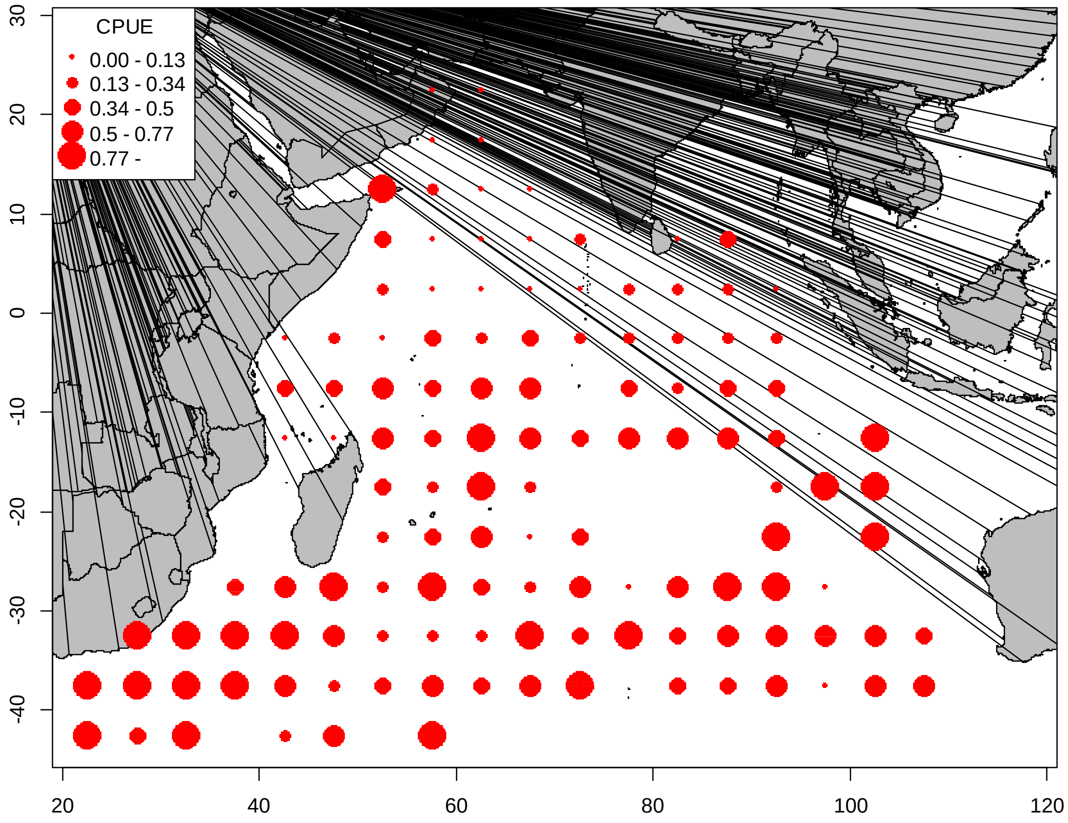

Figure 1. Nominal CPUE distribution of blue shark caught by Taiwanese large-scale tuna longline fishery in the Indian Ocean from 2004 to 2017.

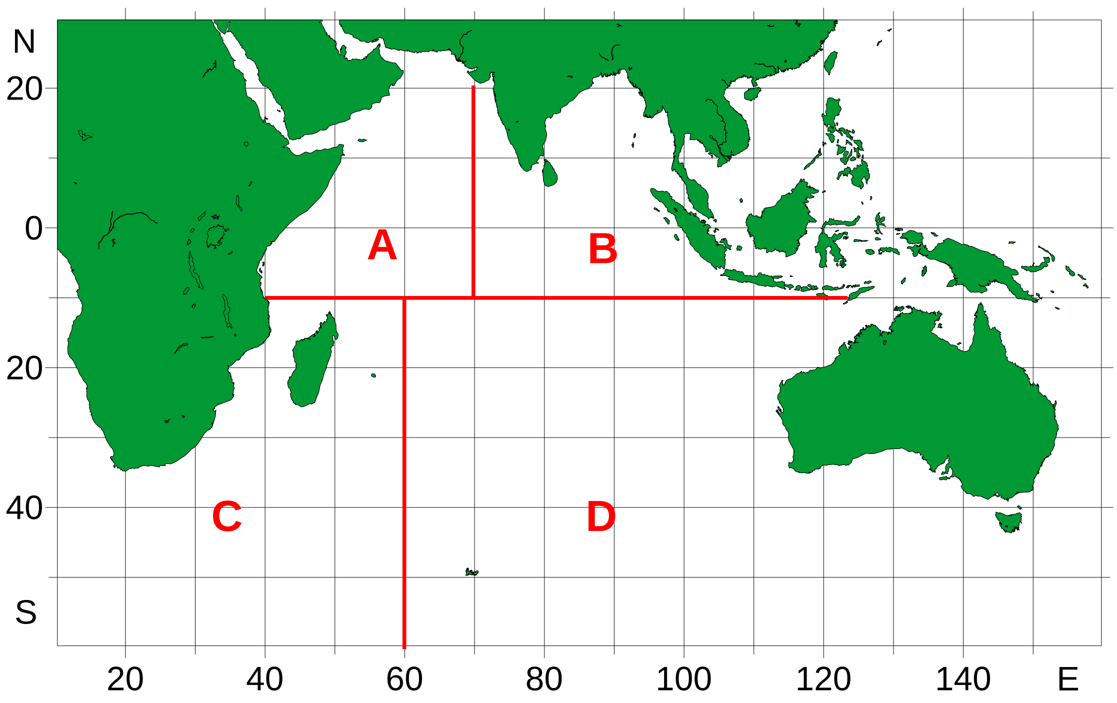

Figure 2. Area stratification used for the estimate of blue shark by-catch of the Taiwanese large-scale tuna longline fishery in the Indian Ocean.

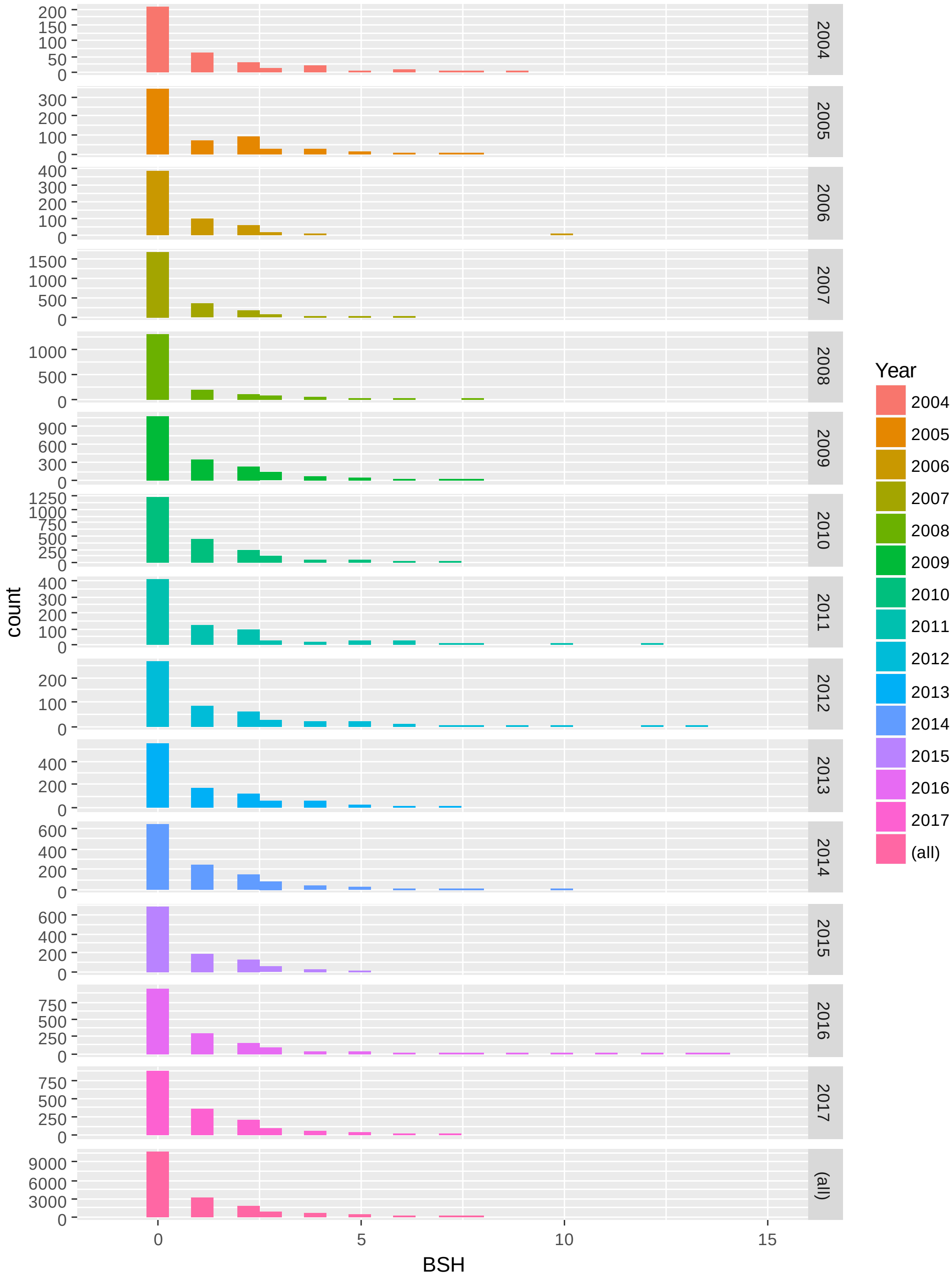

Figure 3. Annual frequency distribution of blue shark bycatch per set in the Indian Ocean, 2004–2017.

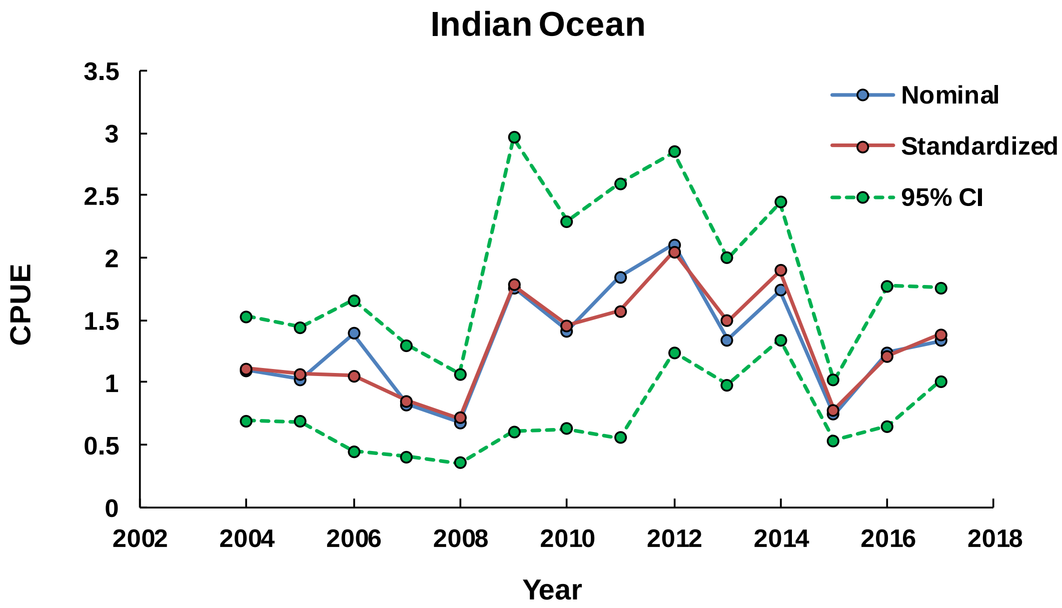

Figure 4. Relative nominal and standardized CPUE with 95% CI of blue shark by Taiwanese longline vessels in the Indian Ocean from 2004 to 2017.

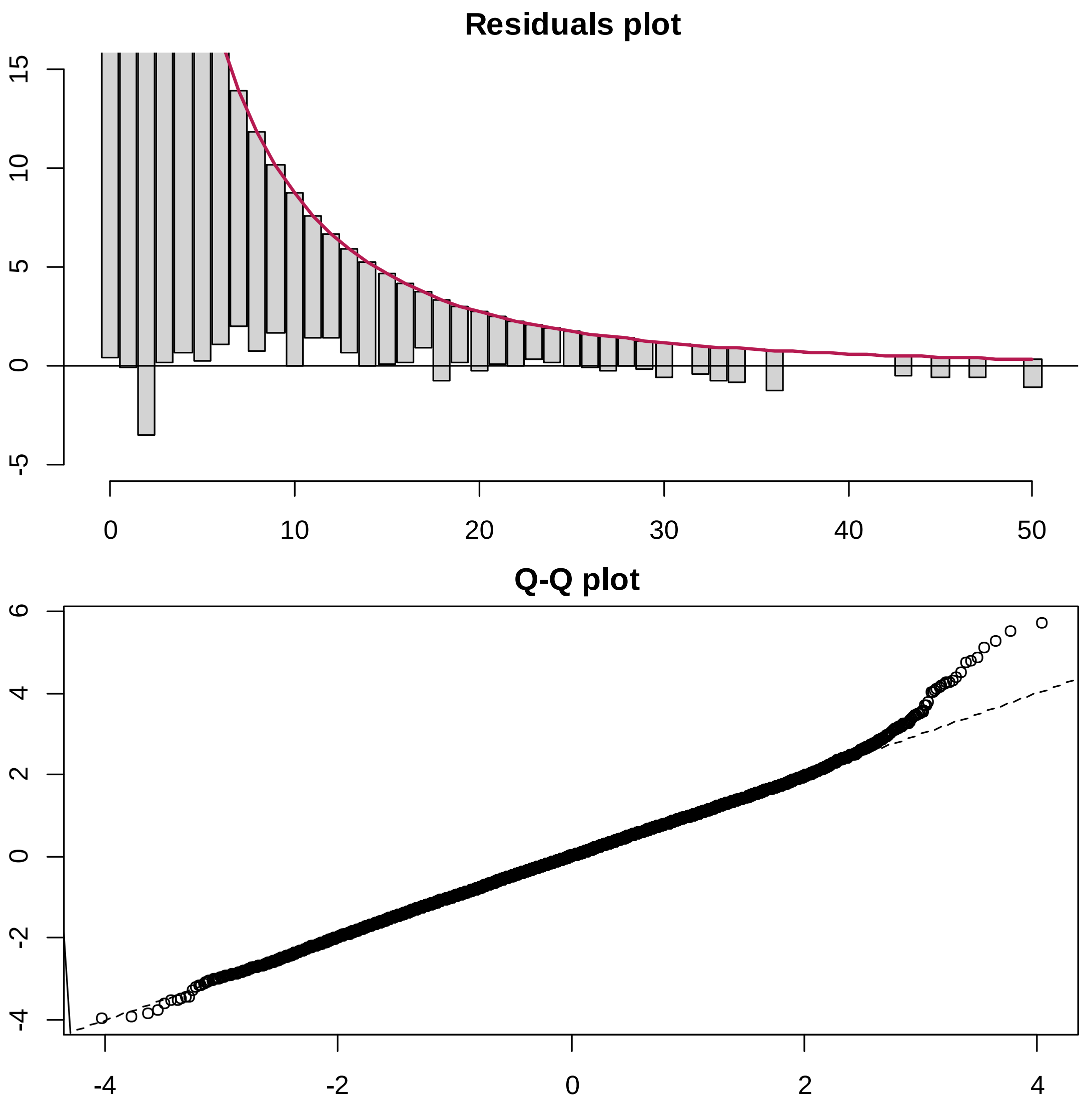

Figure 5. Diagnostic results from the ZINB model fit to the longline blue shark bycatch data.

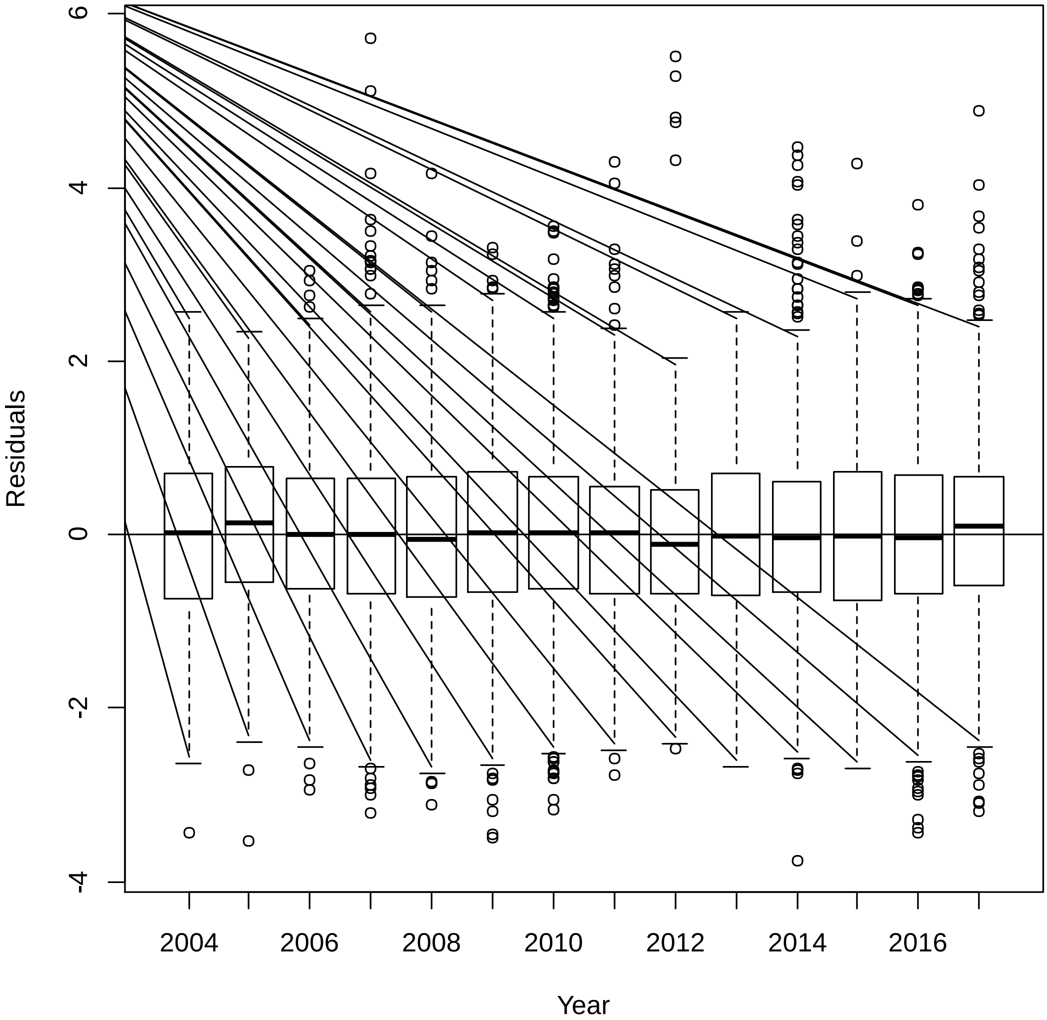

Figure 6. Box plots of the Pearson residuals vs. the covariates for the variables Year.

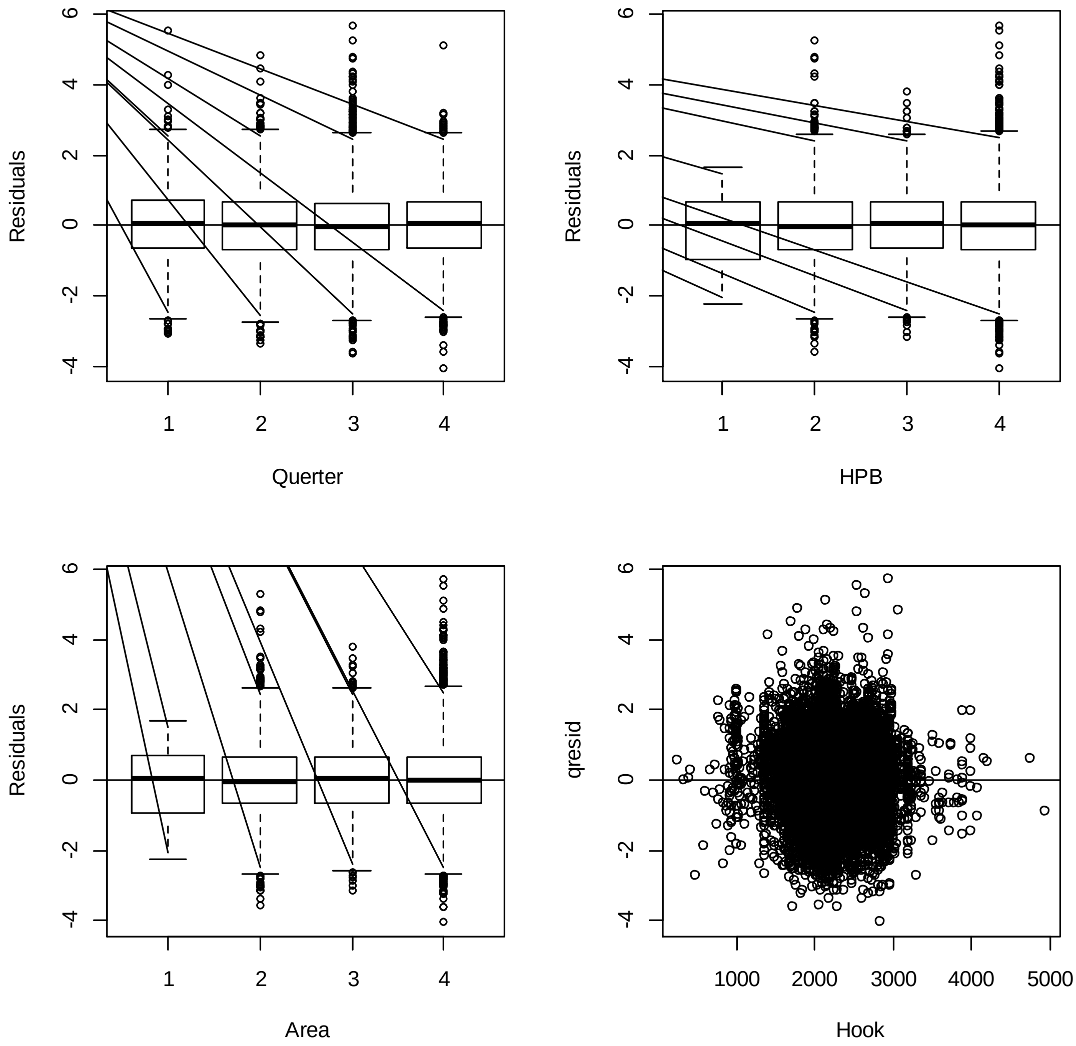

Figure 7. Box plots of the Pearson residuals vs. the covariates for the variables Quarter, HPB, Area and Hook.

Table 1. Summary of information of the observers' data used in this study.

Year	Indian Ocean	
	No. of Hooks	No. of Sets
2004	810,853	349
2005	1,421,228	592
2006	1,419,307	624
2007	5,765,847	2,476
2008	4,248,446	1,781
2009	5,220,475	2,137
2010	5,519,258	2,271
2011	1,876,263	766
2012	1,405,158	507
2013	1,964,276	1,063
2014	2,556,725	1,270
2015	2,151,986	1,089
2016	3,332,972	1,652
2017	3,586,601	1,772
Average	2,948,528	1,311

Table 2. The observed percentage of zero-catch of blue shark for Taiwanese tuna longline vessels in the Indian Ocean from 2004 to 2017.

Year	Percentage of zero-catch
2004	59.03%
2005	58.11%
2006	61.86%
2007	68.09%
2008	73.50%
2009	50.44%
2010	53.81%
2011	53.52%
2012	52.86%
2013	52.02%
2014	49.84%
2015	62.90%
2016	57.51%
2017	49.94%
Average	57.39%

Table 3. Estimated nominal and standardized CPUE values for blue shark of the Taiwanese tuna longline fishery in the Indian Ocean.

Year	Relative values			
	Nominal	Standardized	Lower CI	Upper CI
2004	1.095	1.110	0.691	1.529
2005	1.024	1.065	0.685	1.445
2006	1.388	1.051	0.443	1.658
2007	0.824	0.849	0.400	1.298
2008	0.679	0.709	0.349	1.070
2009	1.758	1.785	0.602	2.967
2010	1.410	1.460	0.627	2.292
2011	1.850	1.573	0.553	2.594
2012	2.105	2.045	1.237	2.854
2013	1.338	1.491	0.982	1.999
2014	1.735	1.894	1.341	2.448
2015	0.739	0.775	0.533	1.018
2016	1.240	1.211	0.648	1.773
2017	1.336	1.384	1.009	1.758

Table 4. Analysis of Deviance Table of count model.

```
Call:
zeroinfl(formula = BSH ~ HPB + Year + Area, data = data, offset = log(Hook),
    dist = "negbin")

Pearson residuals:
    Min      1Q  Median      3Q     Max
-0.6622 -0.5722 -0.4529  0.1891 24.2522

Count model coefficients (negbin with log link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.22810    0.37051 -19.508  < 2e-16 ***
HPB2        -0.35476    0.37619  -0.943 0.345664
HPB3         0.23216    0.37982   0.611 0.541043
HPB4        -0.52370    0.37691  -1.389 0.164698
Year2005    -0.17284    0.14207  -1.217 0.223745
Year2006    -0.26274    0.13610  -1.930 0.053551 .
Year2007    -0.03985    0.12271  -0.325 0.745342
Year2008    -0.47028    0.12945  -3.633 0.000280 ***
Year2009     0.07015    0.12030   0.583 0.559788
Year2010    -0.15895    0.12065  -1.317 0.187694
Year2011     0.20803    0.13205   1.575 0.115157
Year2012     0.49526    0.13797   3.590 0.000331 ***
Year2013     0.32367    0.12793   2.530 0.011405 *
Year2014     0.46804    0.12461   3.756 0.000173 ***
Year2015    -0.39849    0.12920  -3.084 0.002040 **
Year2016    -0.13366    0.12385  -1.079 0.280519
Year2017     0.11943    0.12533   0.953 0.340614
Area2       -0.49402    0.06140  -8.046 8.58e-16 ***
Area3        0.56758    0.05870   9.669  < 2e-16 ***
Area4        0.18643    0.05947   3.135 0.001720 **
Log(theta)  -0.73599    0.02203 -33.415  < 2e-16 ***
```

Table 5. Analysis of Deviance Table of Zero-inflated model.

```
Zero-inflation model coefficients (binomial with logit link):
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     1.90600    0.71600   2.662 0.007768 **
HPB2            1.79470    0.64816   2.769 0.005624 **
HPB3            3.59241    0.72436   4.959 7.07e-07 ***
HPB4           -1.31348    0.62940  -2.087 0.036898 *
Year2005       -0.44156    0.42889  -1.030 0.303221
Year2006      -11.84598   25.79783  -0.459 0.646101
Year2007       -1.74420    0.42850  -4.071 4.69e-05 ***
Year2008       -1.66773    0.45594  -3.658 0.000254 ***
Year2009      -12.16638   20.78389  -0.585 0.558295
Year2010       -3.80486    0.57349  -6.635 3.25e-11 ***
Year2011       -6.04536    0.90043  -6.714 1.90e-11 ***
Year2012       -1.12558    0.55608  -2.024 0.042957 *
Year2013       -1.37401    1.22213  -1.124 0.260897
Year2014       -1.78728    0.50359  -3.549 0.000387 ***
Year2015       -2.36377    0.95694  -2.470 0.013507 *
Year2016       -3.70106    0.69479  -5.327 9.99e-08 ***
Year2017        0.03672    0.51770   0.071 0.943458
Area2          -4.11343    0.51753  -7.948 1.89e-15 ***
Area3          -3.71722    0.36818 -10.096  < 2e-16 ***
Area4          -6.37024    0.64735  -9.841  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 0.479
Number of iterations in BFGS optimization: 116
Log-likelihood: -2.616e+04 on 41 Df
```