# CPUE standardization of striped marlin (*Tetrapturus audax*) caught by Taiwanese large-scale longline fishery in the Indian Ocean

Yun-Ju Chen, Sheng-Ping Wang*, Wen-Qi Xu, Chih-Yu Lin

Department of Environmental Biology and Fisheries Science, National Taiwan Ocean University, Keelung, Taiwan.
* Corresponding author: wsp@mail.ntou.edu.tw

**ABSTRACT**

This study aggregated and analyzed catch, effort and length data of striped marlin caught by Taiwanese large longline fisheries in the Indian Ocean and conducted CPUE standardization for striped marlin for 2005-2023. This paper briefly describes historical patterns of fishing operations and striped marlin catches caught by Taiwanese large-scale longline fishery in the Indian Ocean. The groups of data sets derived from cluster analysis based on species compositions were incorporated in the CPUE standardization models as a covariate for explaining the target to obtain the relative abundance indices for further stock assessments. Except for the delta-lognormal models, the standardized CPUE series obtained from different model assumptions revealed similar trends. The Standardized CPUE indices obtained from the delta-inverse Gaussian models should be more appropriate than other models based on statistical diagnostics. The CPUE series in both NW and NE areas generally increased from 2009 to 2013 and then decreased after 2013.

## 1. INTRODUCTION

Striped marlin is largely considered to be the bycatch species of industrial fisheries. Most of the striped marlin were caught by the longline fishery before the mid-1990s. After that, gillnet catches gradually increased while longline catches gradually decreased. In recent years, the proportion of gillnet catches has surpassed that of the longline fishery. Gillnets account for around 50% of total catches in the Indian Ocean between 2014 and 2018, followed by longlines (40%). The remaining catches are mostly recorded under troll and handlines. The catch trends of striped marlin in the Indian Ocean varied, ranging from 2,000 t to 8,000 t per year. In particular, catches reported under longlines highly varied, with lower catch levels between 2009 and 2011 largely due to declining catches reported by Taiwan. In 2021, the lowest catch reached 2,600 tons. (IOTC, 2023).

The striped marlin were mainly caught by Taiwan and Japan. Before the 1970s, Japan was the main country for the striped marlin. Thereafter, Taiwan catches increased significantly and became the most important country for striped marlin in the Indian Ocean. In recent years, catches of striped marlin have increased by Indonesian fisheries and small-scale longline fisheries, which occupied a very important proportion. The distribution of striped marlin catches has changed since the 1980s with most of the catches now taken in the north-west Indian Ocean. In recent years, the catches of striped marlin caught by Taiwan and Japan revealed a decreasing trend, and the reason is still unclear. However, changes in fishing grounds and catches are thought to be related to changes in access agreements to the EEZs of coastal countries in the Indian Ocean instead of changes in the distribution of the species over time. in recent years, catches of striped marlin from the coastal gillnet fisheries of I.R. Iran and Pakistan have steadily increased to an average of 1,600 t annually and contribute around 66% of the total catches of striped marlin in 2021 (IOTC, 2023).

Since the current stock status of striped marlin was pessimistic (IOTC, 2021), this study conducted CPUE standardization for striped marlin in the Indian Ocean to provide relative abundance indices for further stock assessments.

## 2. MATERIALS AND METHODS

### 2.1. Catch and Effort data

In this study, daily operational catch and effort data (logbook) with 5x5 degree longitude and latitude grid for Taiwanese longline fishery during 1979-2023 were provided by Overseas Fisheries Development Council of Taiwan (OFDC). For the area stratification, this study adopted the four areas stratification for swordfish by Wang and Nishida (2011) (Fig. 1). For conducting the cluster analysis prior to the CPUE standardizations, the data were aggregated by 10-days duration (1st-10th, 11th-20th, and 21st~ for each month) (Kitakado et al., 2021).

As the discussions and suggestions from previous IOTC meetings (2021a; 2021b), Taiwanese data before 2005 were recommended not to be used to analyze the targeting of fishing operations and conduct the CPUE standardization for tropical tuna due to the problem of data quality. However, the data problem might not only influence the misreport for the catches of major tropical tunas but also lead to uncertainties in the catch and effort data for other species. Therefore, CPUE standardizations were conducted using the data from 2005 to 2023 as suggested in previous meetings.

### 2.2. Cluster analysis

The details of the procedures of cluster analysis were described by Wang et al. (2021). This study adopted a direct hierarchical clustering with an agglomerative algorithm, which brings a fast and efficient implementation through features of memory-saving routines in the hierarchical clustering of vector data (Müllner, 2013). The trials were conducted using R function “hculst.vector” of package “fastcluster” (Müllner 2021) with Ward's minimum variance linkage methods (“ward.D” for the argument “method” in “hclust.vector” of R function) applied to the squared Euclidean distances between data points calculated based on the species composition.

The number of clusters was selected based on the elbow method, i.e. the change in deviance between/within clusters against different numbers of clusters. The number of clusters was determined when the improvement in the sum of within-cluster variations was less than 10%.

### 2.3. CPUE Standardization

A large amount of zero-catches was recorded in the operational catch and effort data sets because striped marlin was caught as the bycatch species of Taiwanese longline fishery in the Indian Ocean. Historically, ignoring zero observations or replacing them with a constant was the most common approach. An alternative and popular way to deal with zeros was through the delta approach (Hinton and Maunder, 2004; Maunder and Punt, 2004). IOTC (2016) also noted the use of the delta approach to accommodate the high proportion of zero catches. Therefore, the delta-general linear models with different assumptions of error distribution were applied to conduct the CPUE standardization of striped marlin in the Indian Ocean (Pennington, 1983; Lo et. al., 1992; Pennington, 1996; Andrade, 2008; Lauretta et al., 2016; Langley, 2019).

As the approach of Wang (2018), the models were simply conducted with the main effects of year, quarter, longitude, latitude and fishing targeting (clusters), while interactions between main effects were not incorporated into the models. The models for positive catches and delta model were conducted as follows:

For CPUE of positive catches:

$$Catch = \mu + Y + Q + CT + G + T + offset(\log(Hooks)) + \varepsilon^{pos}$$

For delta model:

$$PA = \mu + Y + Q + CT + G + T + \varepsilon^{del}$$

| | | |
|---|---|---|
| where | *Catch* | is the catch in number/1,000 hooks |
| | PA | is the presence/absence of catch, |
| | Hooks | is the effort of 1,000 hooks, |
| | $\mu$ | is the intercept, |
| | $Y$ | is the effect of year, |
| | $Q$ | is the effect of quarter, |
| | $CT$ | is the effect of vessel scale, |
| | $G$ | is the spatial effect of Lon and Lat 5x5 grid, |
| | $T$ | is the effect of targeting (cluster), |
| | $\varepsilon^{pos}$ | is the error term assumed based on various distribution, |
| | $\varepsilon^{del}$ | is the error term, $\varepsilon^{del}$ ~ Binomial distribution. |

To examine the appropriateness of the assumption of error distribution, this study applied normal, gamma, binomial and inverse.gaussian distributions to the error distribution of the model for the positive catches and specified "log" for the model link function. For the model with inverse.gaussian distribution, the index of power variance function was tested using values of 1.1-1.9.

The stepwise searches ("both" direction, i.e. "backward" and "forward") based on the values of the Akaike information criterion (AIC) were performed to select the explanatory variables for each model. Then, the coefficient of determination ($R^2$), and Bayesian information criterion (BIC) were calculated for the models with selected explanatory variables. The AIC and BIC, which were calculated based on the likelihoods with full constants obtained glm() and glm.nb(), were used to compare the models with different error distributions (e.g. Setyadji et al., 2019). In addition, the dispersion statistics for Pearson residuals were calculated to check whether under- or overdispersions resulted from the models with an assumed error distribution.

The standardized CPUE indices were calculated based on the estimates of the least square means of the interaction between the effects of year and area, and calculated by the product of the standardized CPUE of positive catches and the delta model:

$$DL^{index} = e^{\log(CPUE)} \times \left( \frac{e^{PA}}{1 + e^{PA}} \right)$$

where $DL^{index}$ is the standardized CPUE

## 3. RESULTS AND DISCUSSION

### 3.1. Historical fishing trends

Fig. 2 and Fig. 3 show the striped marlin catch in numbers and nominal CPUE distribution based on the logbook data of Taiwanese large-scale longline fishery in the Indian Ocean aggregated by 5 years. Striped marlin were mainly caught in tropical and coastal waters of the northern Indian Ocean. Although the amount of fish caught in the southern Indian Ocean increased significantly from 2005 to 2023 due to the increase in effort, the distribution of high CPUE over the years was still limited to the coastal waters of the northern Indian Ocean.

Striped marlin catches were mainly made with high effort in northern waters, especially for the northwestern fishing area (NW). Although the catches in the northwestern Indian Ocean increased significantly around 2005, the catches substantially decreased in the following years (Fig. 4 and Fig. 5).

### 3.2. Cluster analysis

CPUE standardizations were separately performed for only northern areas (Fig. 1) since the catches and CPUE of striped marlin in the southern areas were substantially lower than those in the northern areas, especially for recent decades (Figs. 2 and 3).

Based on the results from the elbow method, 4 clusters were selected for Areas NW, SW and SE, while 3 clusters were selected for Area SW (Figs. 6 and 7). For each area, the species compositions revealed different patterns by clusters (Fig. 8).

Fig. 9 shows the striped marlin catches and efforts by clusters and areas and striped marlin catches were contained in different clusters in different periods when different levels of efforts were deployed. Therefore, the data of all clusters were used to conduct further CPUE standardizations. The annual trends of the proportions of zero catches of striped marlin roughly stayed the same over the years for NE, , while the NW area showed significant variations. (Fig. 10).

### 3.3. CPUE standardization

Based on the AIC model selections for the models for positive catches and delta models, all of the effects were statistically significant and remained in the models for all areas. For the models for positive catches, the models with inverse Gaussian error distribution would be the optimal models for all areas based on the values of AIC, BIC and $R^2$ (Table 1). In addition, diagnostic plots for residuals also indicated that the models with inverse Gaussian error distribution (Fig. 11) should be more appropriate than other models because there were less increasing or decreasing trends in the range of predicted values (plots for other models by areas were not shown here but the residuals revealed obvious patterns with predicted values). Therefore, the delta-inverse Gaussian models were selected to produce the standardized CPUE series.

The ANOVA tables for selected models are shown in Table 2. Except for the impact of the effect of *Y*, the effects of G (*Lat* and *Lon*) were the most significant variable for both positive catches and delta models in NW and NE areas.

The area-specific standardized CPUE series are shown in Fig. 12 and the CPUE series revealed similar trends for all models. The standardized CPUE of positive catches and catch probability obtained from the selected model are shown in Fig. 13 and CPUE of positive catches and catch probability revealed similar trends.

The standardized CPUE series with 95% confidence intervals obtained from the selected model are shown in Fig. 14. The CPUE series in both NW and NE areas generally increased from 2009 to 2013 and then decreased after 2013.

**REFERENCE**

Andrade, H.A., 2008. Using delta-Gamma generalized linear models to standardize catch rates of yellowfin tuna caught by Brazilian bait-boats. ICCAT SCRS/2008/166.

Fox, J., Monette, G., 1992. Generalized collinearity diagnostics. J. Am. Stat. Assoc., 87: 178–183.

Fox, J., Weisberg, S., 2019. An R companion to applied regression, Third Edition. Thousand Oaks CA: Sage.

Hinton, M.G., Maunder, M.N., 2004. Methods for standardizing CPUE and how to select among them. Col. Vol. Sci. Pap. ICCAT, 56: 169-177.

IOTC, 2016. Report of the 14th Session of the IOTC Working Party on Billfish. IOTC-2016-WPB14-R[E].

IOTC, 2021a. Report of the 19th Session of the IOTC Working Party on Billfish. 13–16 September 2021, Microsoft Teams Online. IOTC–2021–WPB19–R[E].

IOTC, 2021b. Report of the 24rd Session of the IOTC Scientific Committee. 6–10 December 2021, online. IOTC–2021–SC24–R[E]_Rev1.

IOTC, 2022a. Report of the 20th Session of the IOTC Working Party on Billfish. 12–15 September 2022, Online. IOTC–2022–WPB20–R[E].

IOTC, 2022b. Report of the 25th Session of the IOTC Scientific Committee. 5–9 December 2021, Seychelles. IOTC–2022–SC25–R[E].

IOTC, Review of the Statistical Data Available for Indian Ocean Striped Marlin (1950-2021). IOTC-2023-WPB21-INF03-MLS.

Langley, A.D., 2019. An investigation of the performance of CPUE modelling approaches – a simulation study. New Zealand Fisheries Assessment Report 2019/57.

Lauretta, M.V., Walter, J.F., Christman, M.C., 2016. Some considerations for CPUE standardization; variance estimation and distributional considerations. ICCAT Collect. Vol. Sci. Pap. ICCAT, 72(9): 2304-2312.

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

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

Müllner, D., 2013. fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. Journal of Statistical Software, 53(9): 1-18.

Müllner, D., 2021. The fastcluster package: User's manual, Version 1.2.3. https://cran.r-project.org/web/packages/fastcluster/vignettes/fastcluster.pdf

Pennington, M., 1983. Efficient estimation of abundance, for fish and plankton surveys. Biometrics, 39: 281-286.

Pennington, M., 1996. Estimating the mean and variance from highly skewed marine data. Can. J. Fish. Aquat. Sci., 94: 498-505.

Wang, S.P., 2020. CPUE standardization of swordfish (*Xiphias gladius*) caught by Taiwanese large scale longline fishery in the Indian Ocean. IOTC–2020–WPB18–15_Rev1.

Wang, S.P., Nishida, T., 2011. CPUE standardization of swordfish (*Xiphias gladius*) caught by Taiwanese longline fishery in the Indian Ocean. IOTC-2011-WPB09-12.

Wang, S.P., Xu, W.Q., Lin, C.Y., Kitakado, T., 2021. Analysis on fishing strategy for target species for Taiwanese large-scale longline fishery in the Indian Ocean. OTC–2021–WPB19–11.

REVIEW OF THE STATISTICAL DATA AVAILABLE FOR INDIAN OCEAN STRIPED MARLIN (1950-2021)

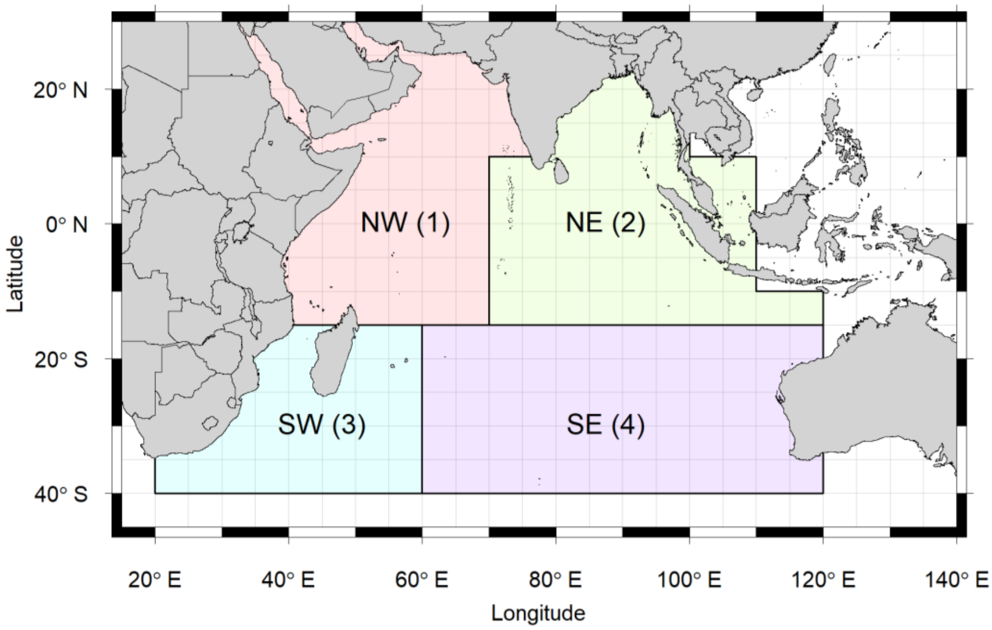



Fig. 1. Area stratification for billfishes in the Indian Ocean.

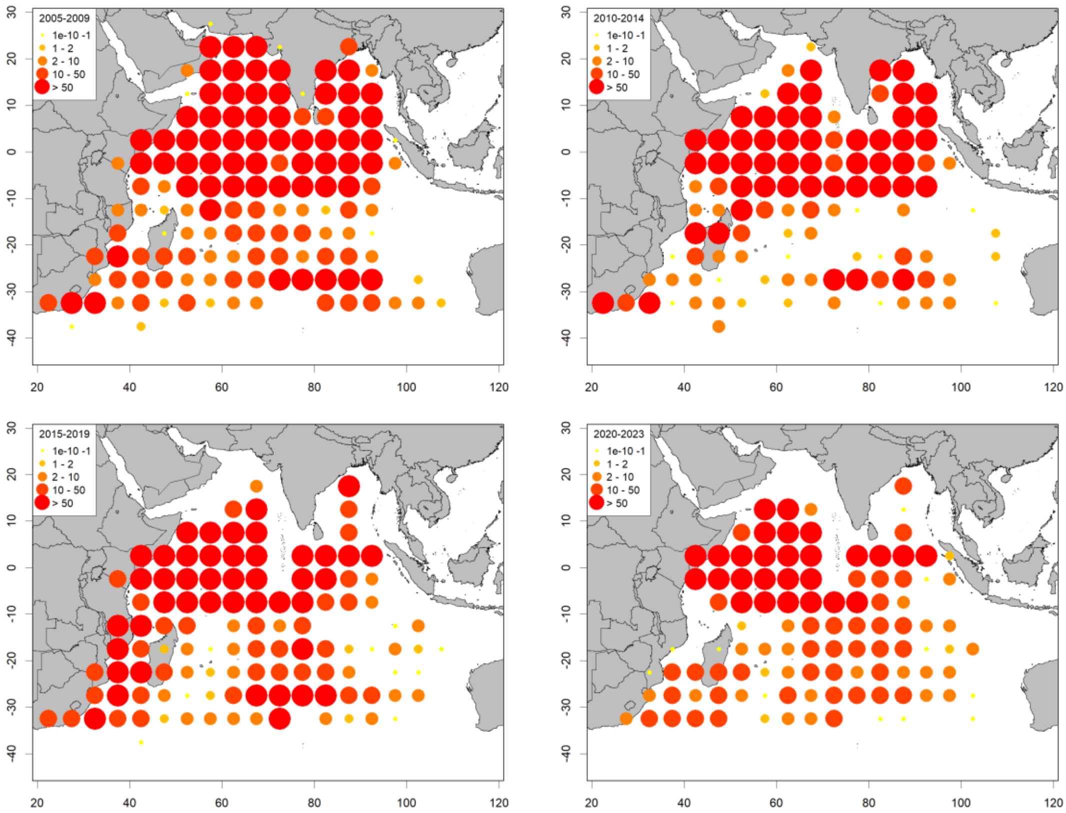



Fig. 2. Striped marlin catch distribution of Taiwanese large-scale longline fishery in the Indian Ocean.

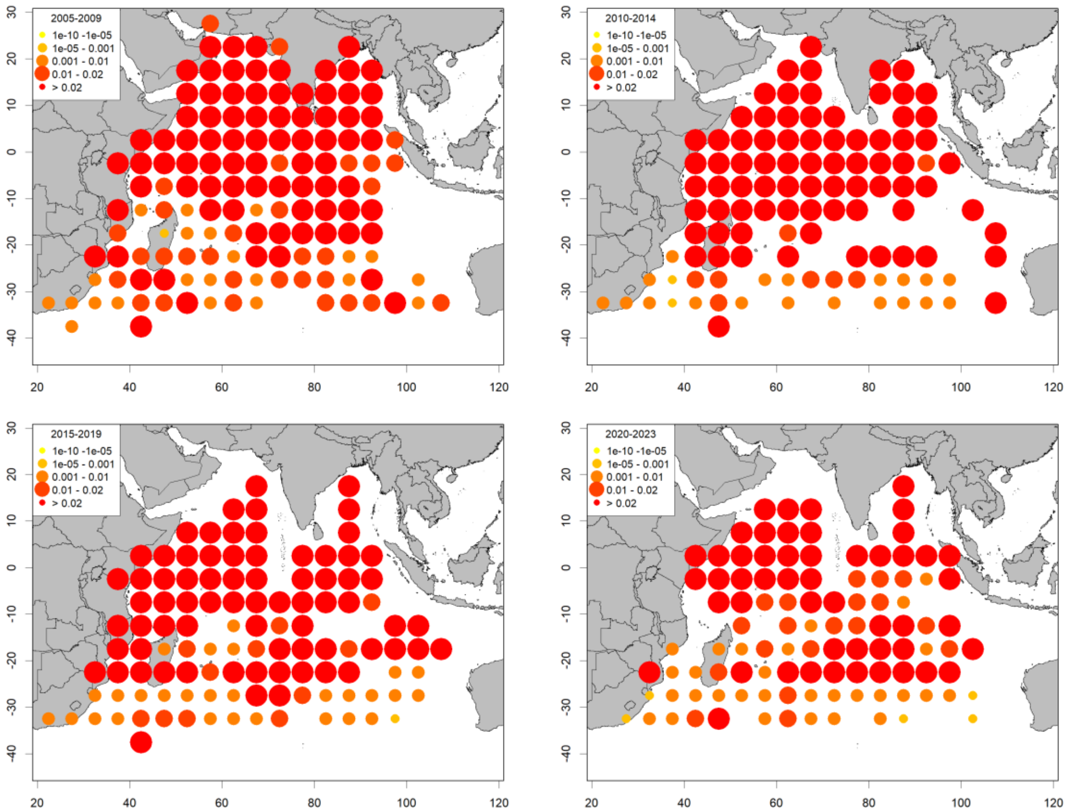



Fig. 3. Striped marlin CPUE distribution of Taiwanese large-scale longline fishery in the Indian Ocean.

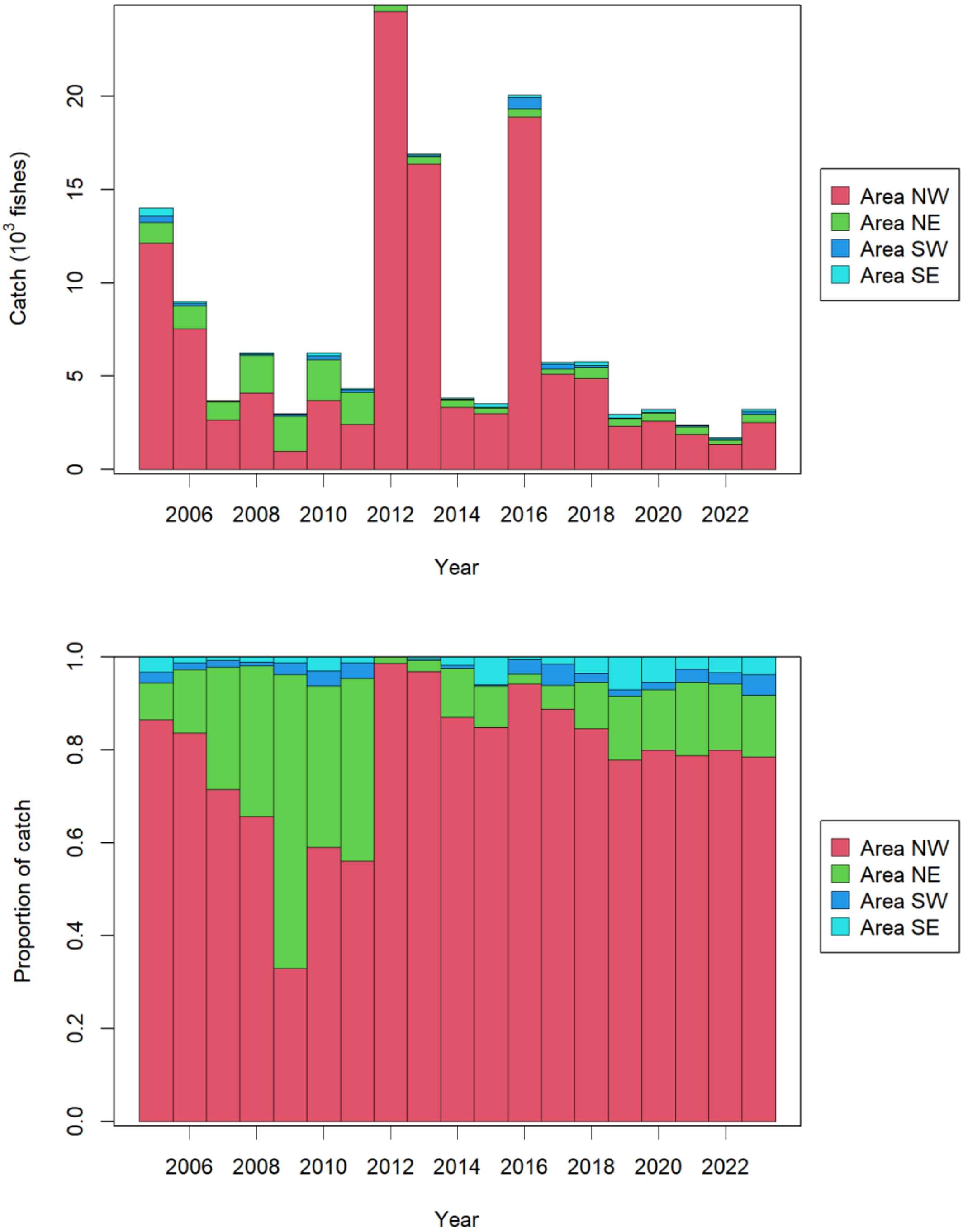



Fig. 4. Annual striped marlin catches of Taiwanese large-scale longline fishery in the Indian Ocean.

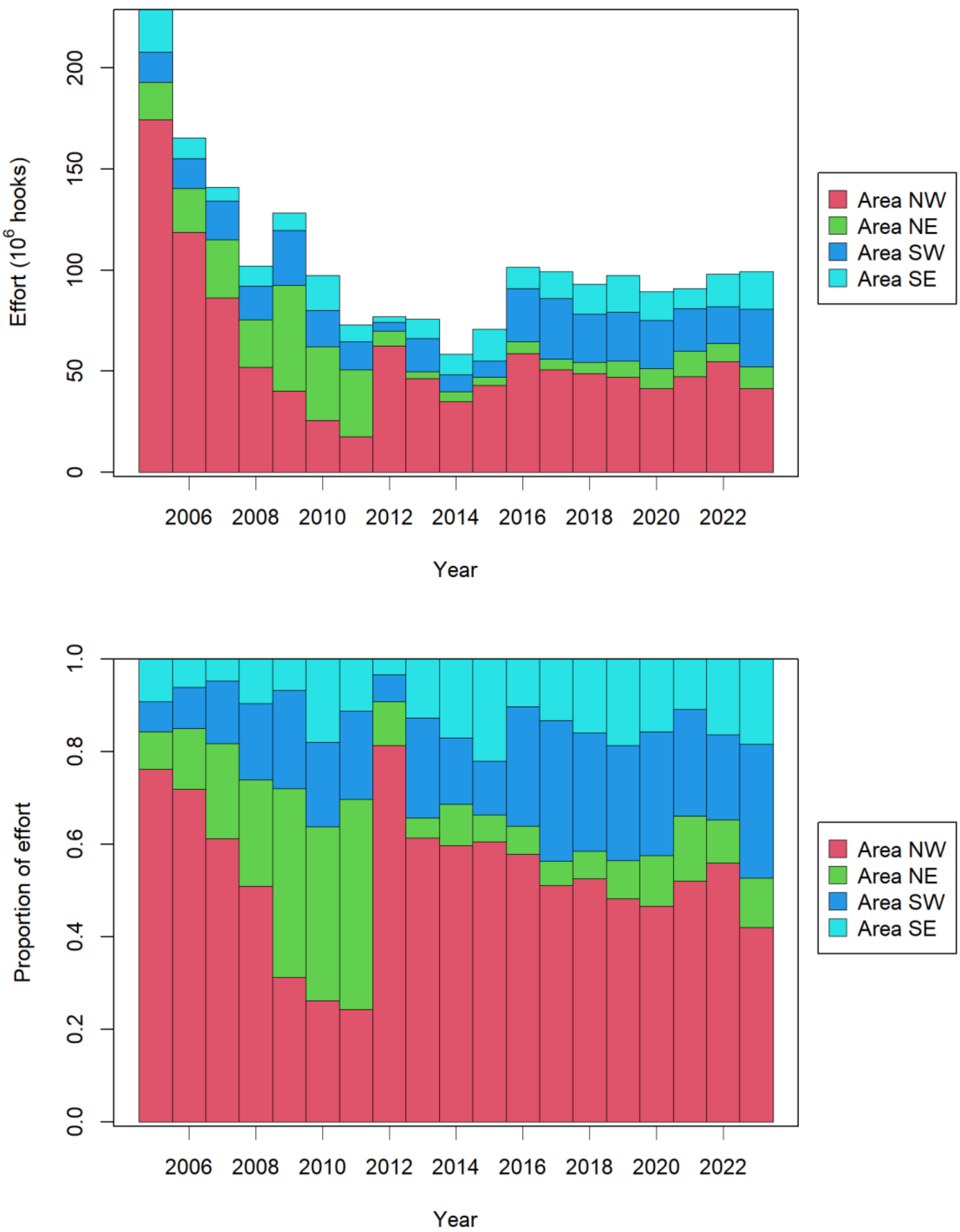
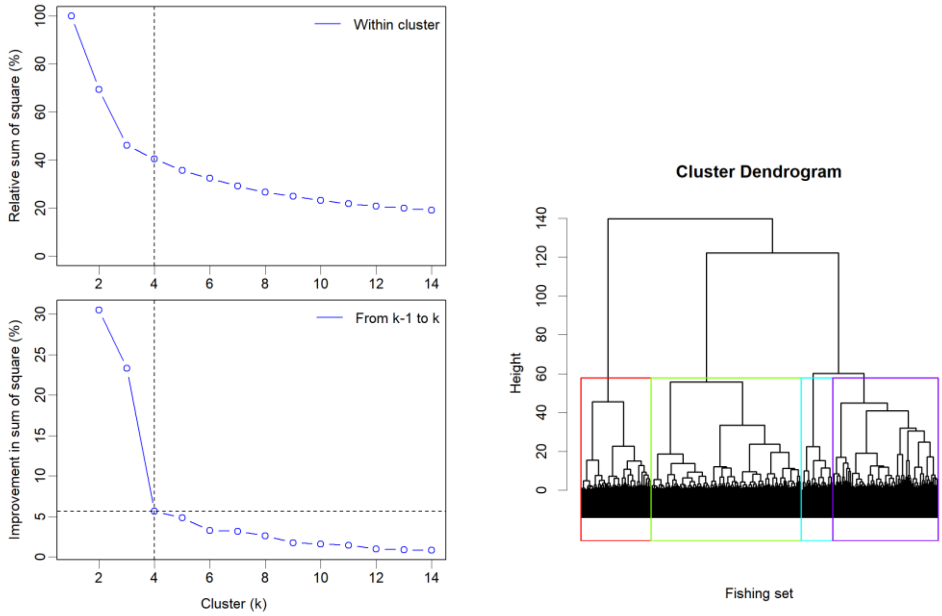



Fig. 5. Annual efforts (number of hooks) of Taiwanese large-scale longline fishery in the Indian Ocean.

Area NW




Area NE

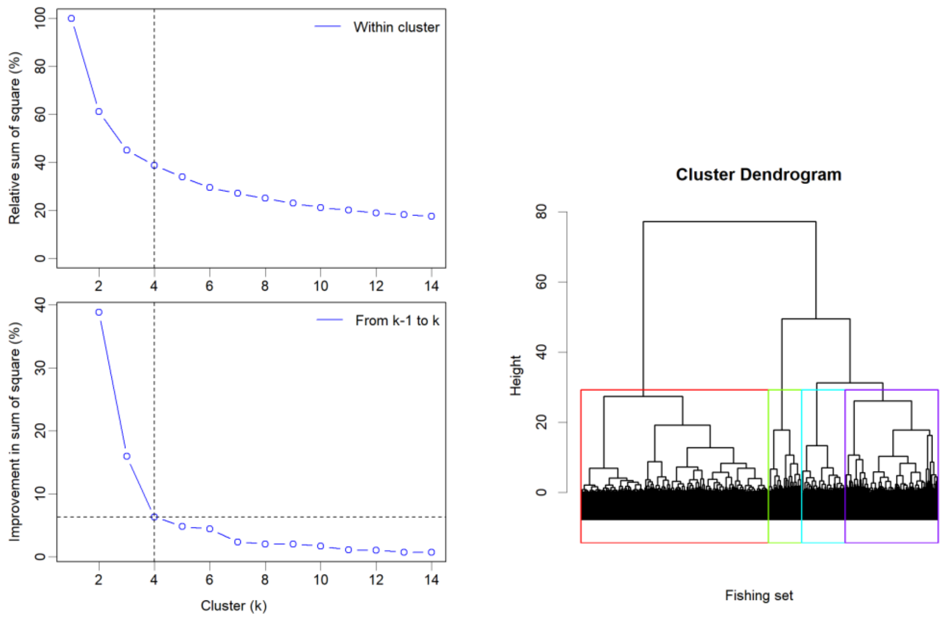



Fig. 6. Sum of squares within clusters for the data of Taiwanese large-scale longline fishery in billfish area of the Indian Ocean.

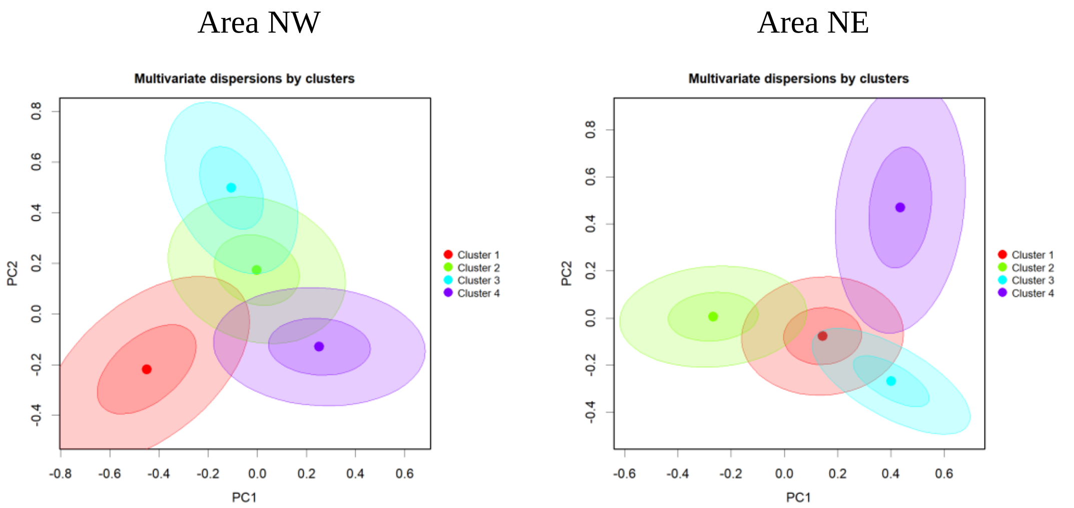



Fig. 7. Multivariate dispersions of the centroids by clusters derived from PCA for the data of Taiwanese large-scale longline fishery in billfish area of the Indian Ocean.

Area NW

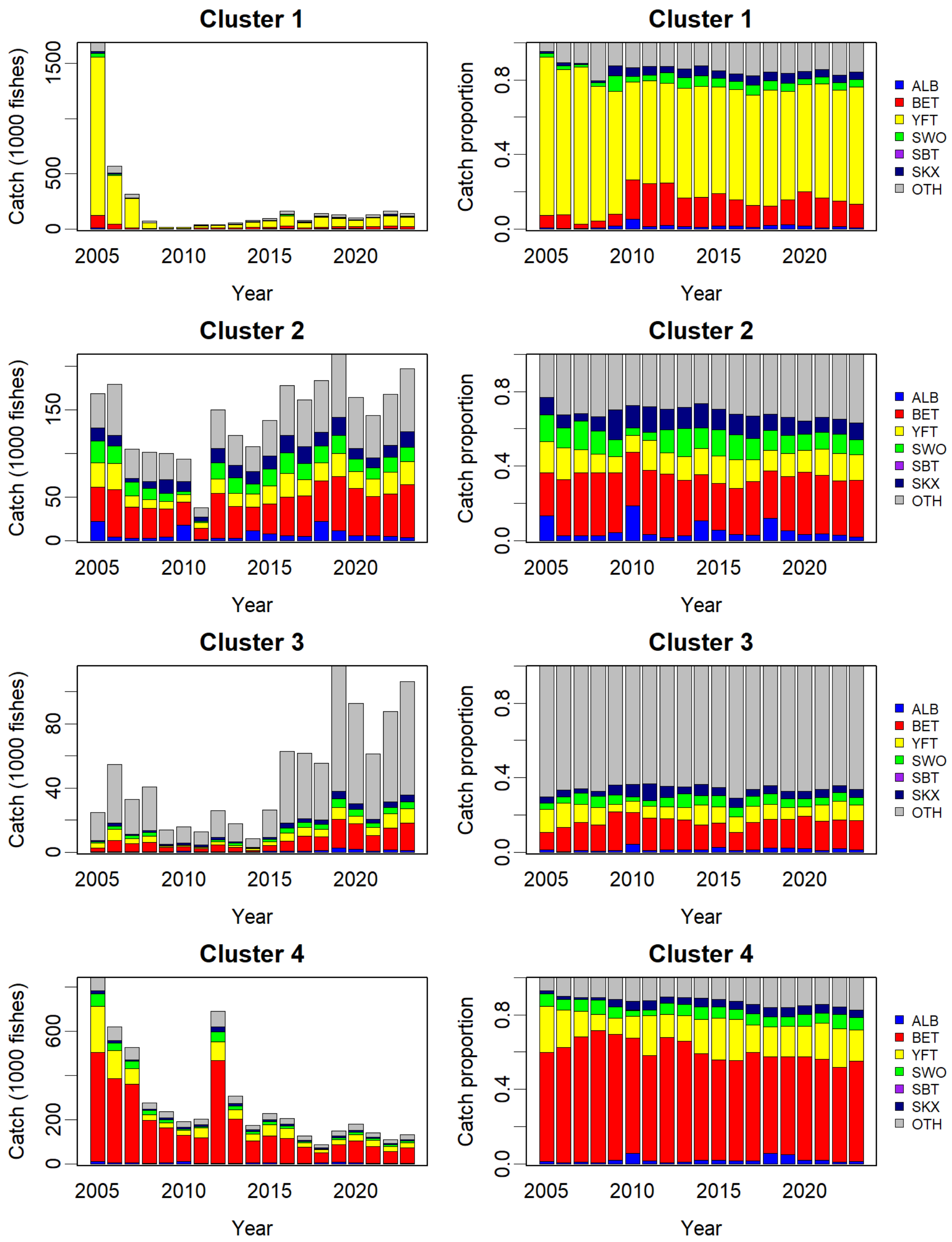



Fig. 8. Annual catches and compositions by species for each cluster of Taiwanese large-scale longline fishery in billfish area of the Indian Ocean.

Area NE

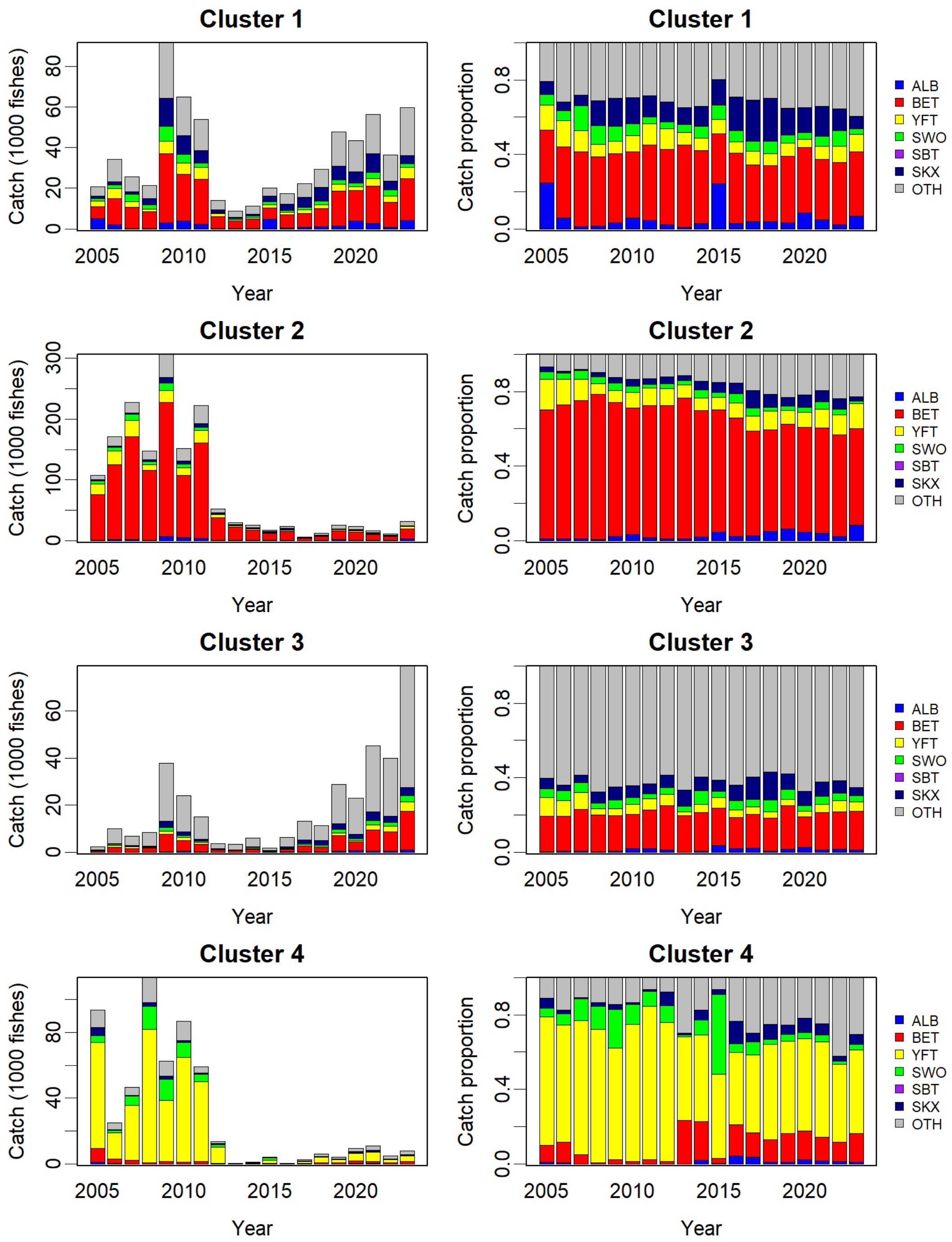
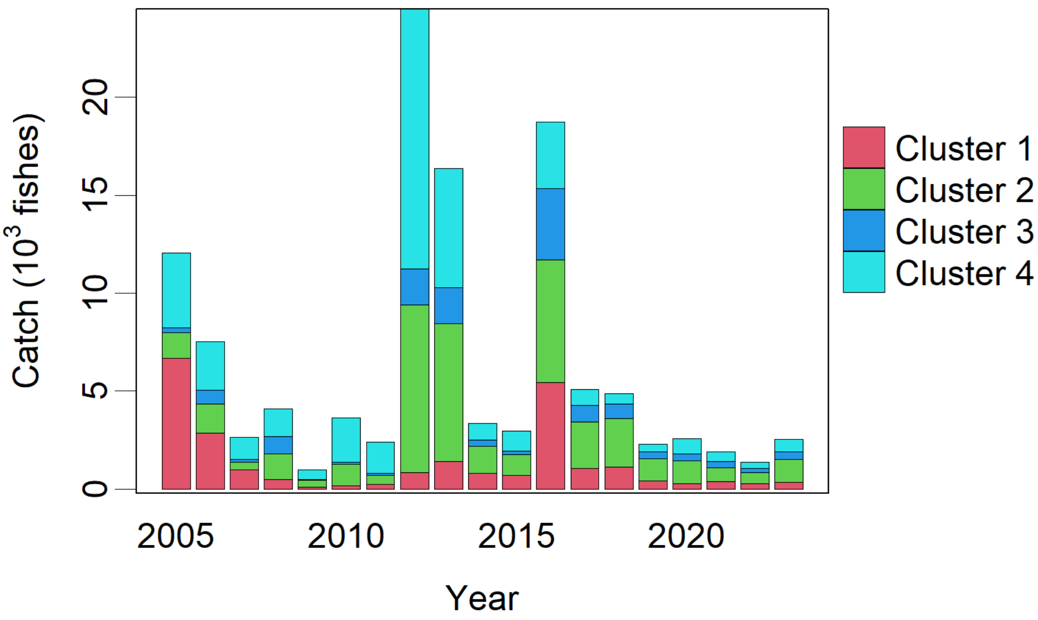



Fig. 8. (continued).

Area NW



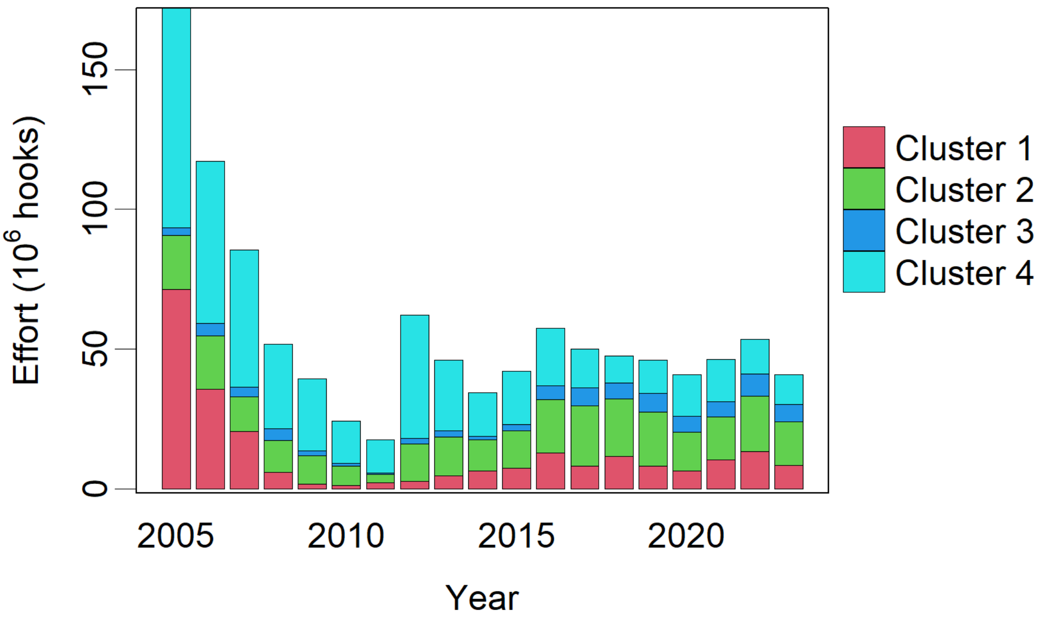
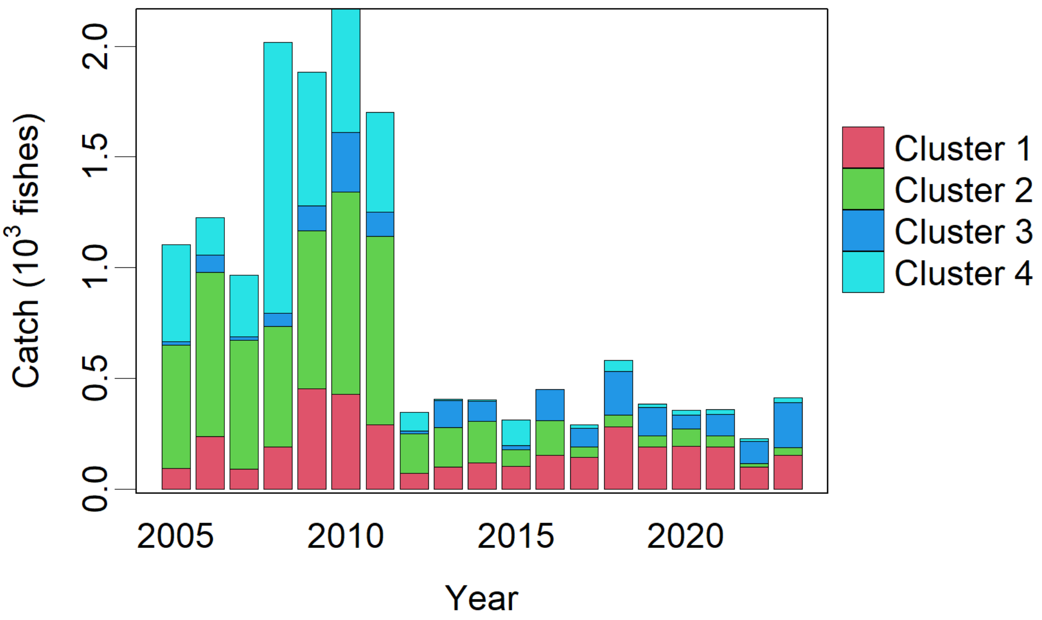




Fig. 9. Annual striped marlin catches and efforts for each cluster of Taiwanese large-scale longline fishery in billfish area of the Indian Ocean.

Area NE



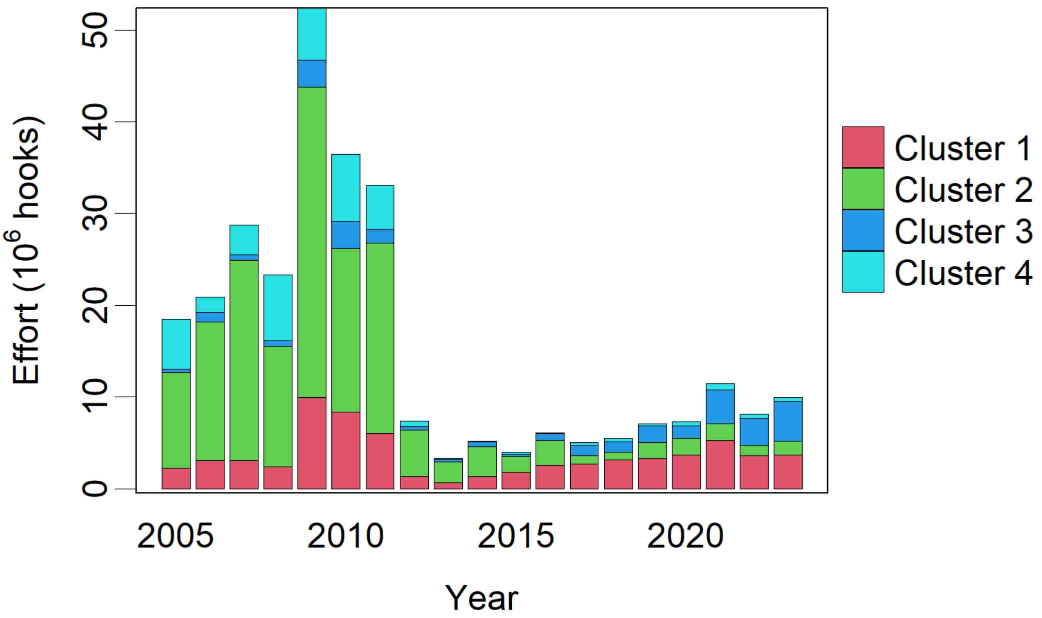
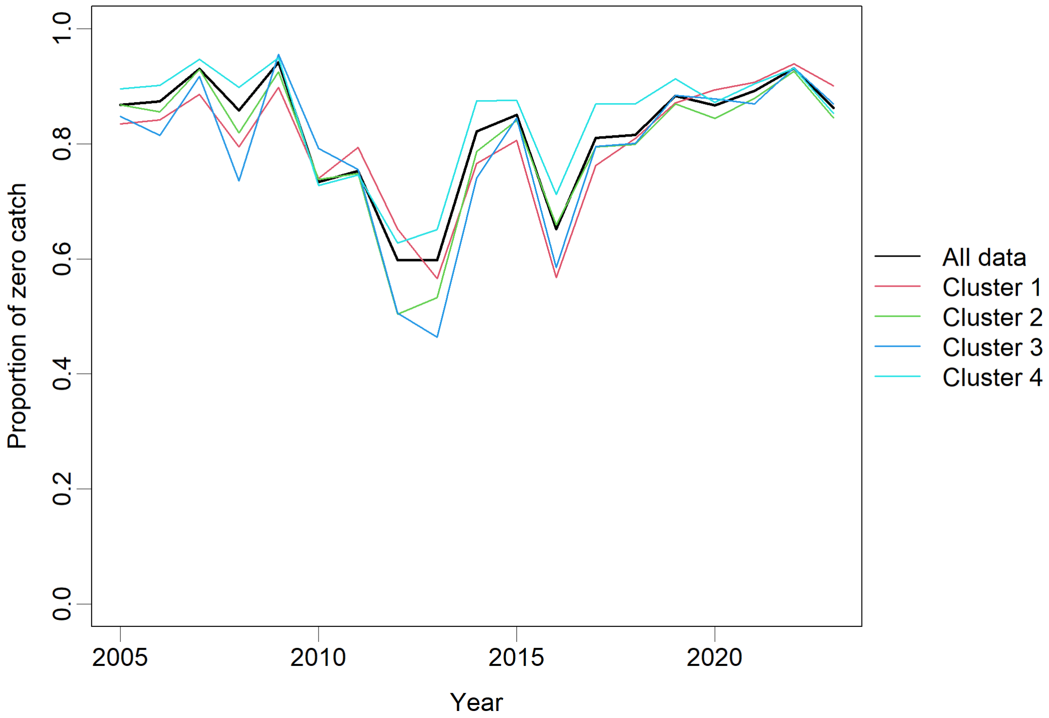




Fig. 9. (continued).

Area NW




Area NE

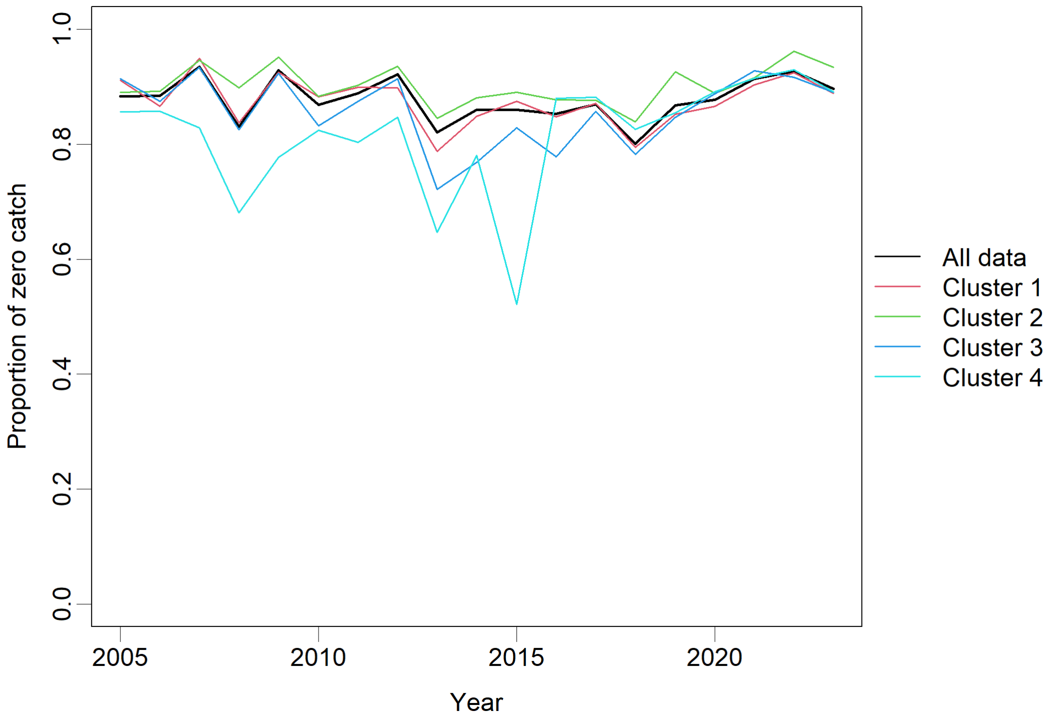



Fig. 10. Annual zero proportion of striped marlin catches for each cluster of Taiwanese

large-scale longline fishery in billfish area of the Indian Ocean.

Area NW

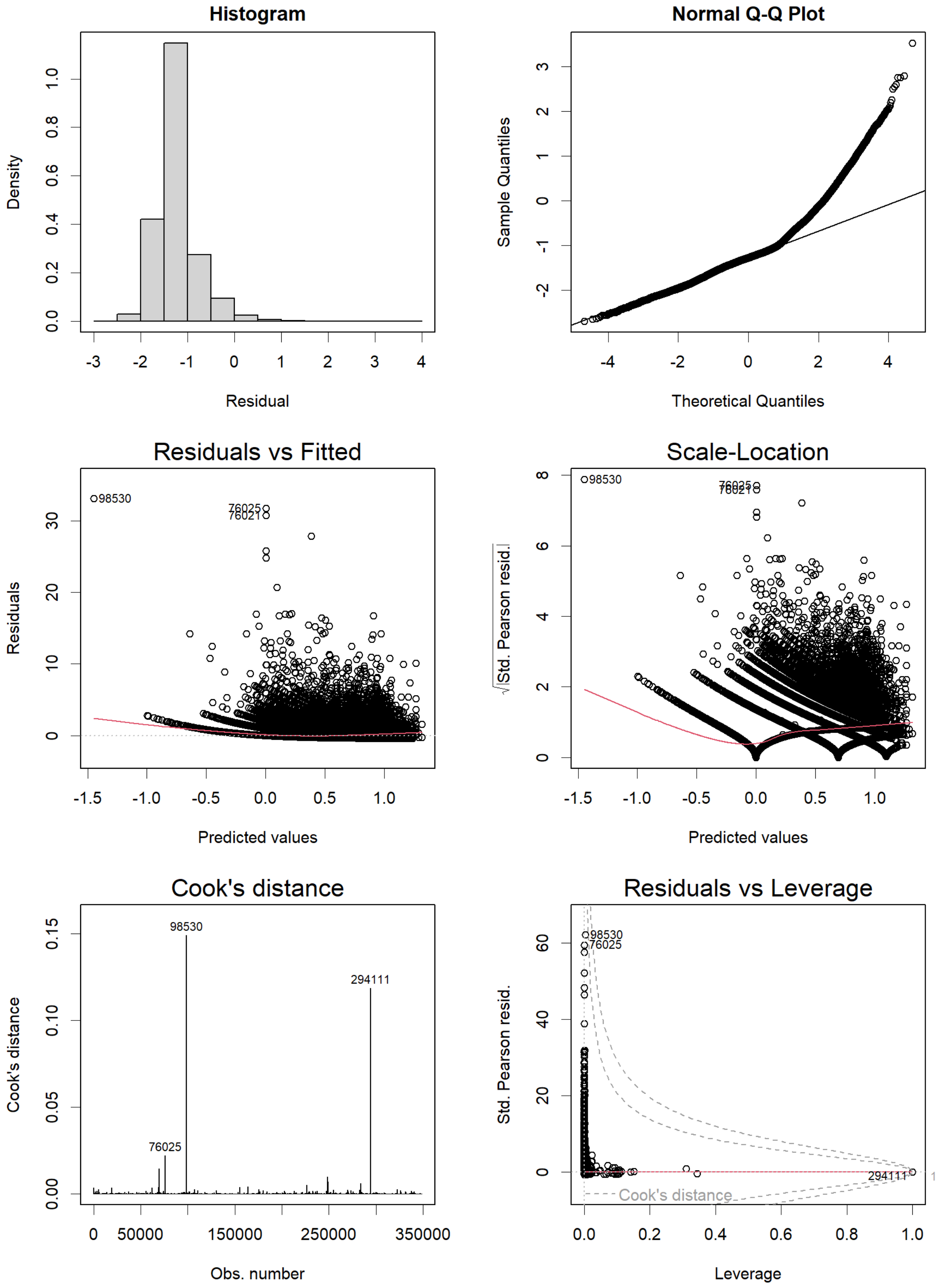



Fig. 11. Diagnostic plots for GLMs with inverse Gaussian error distribution assumption for striped marlin caught by Taiwanese large-scale longline fishery in the Indian Ocean from 2005 to 2023.

Area NE

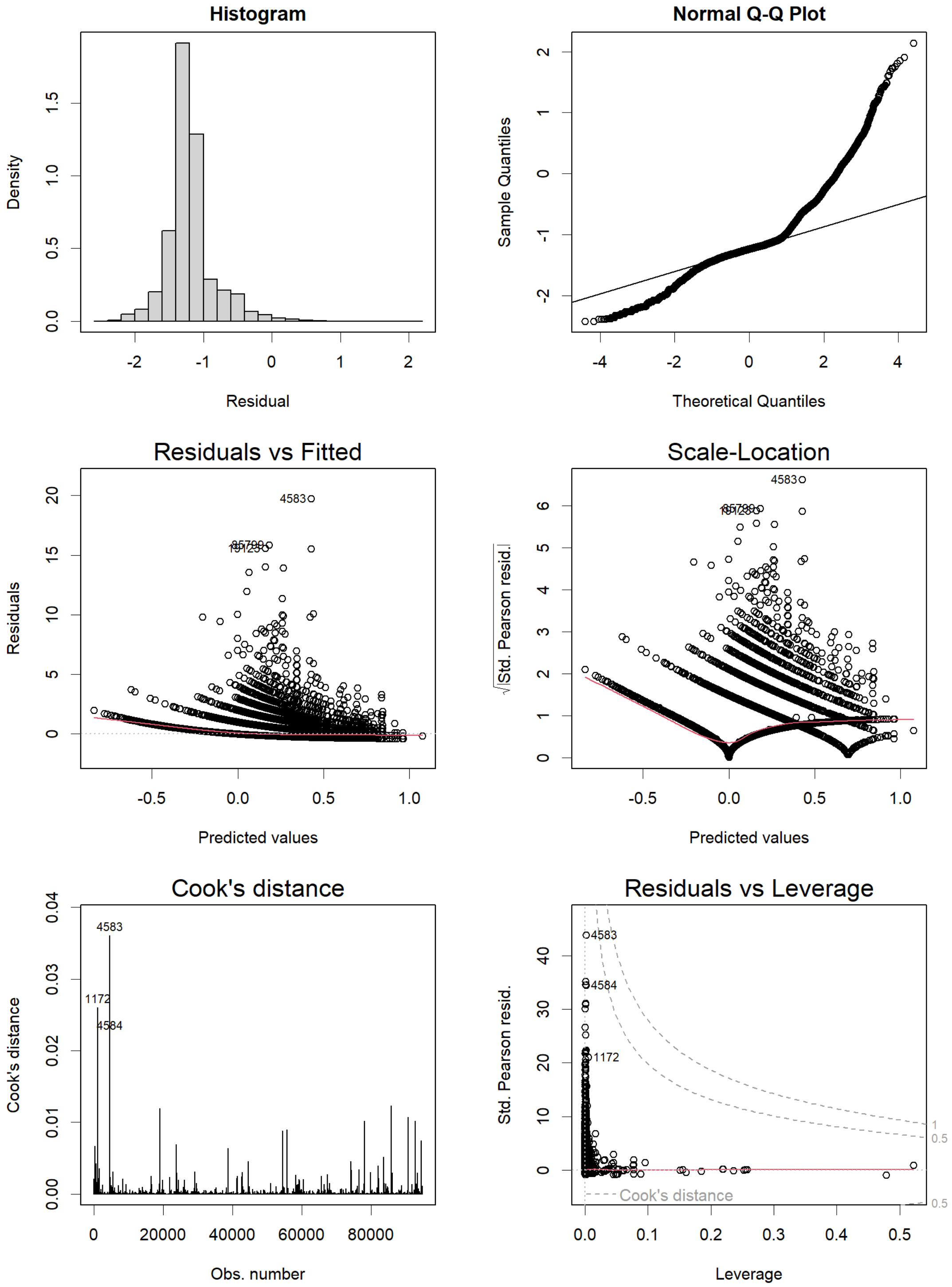



Fig. 11. (continued).

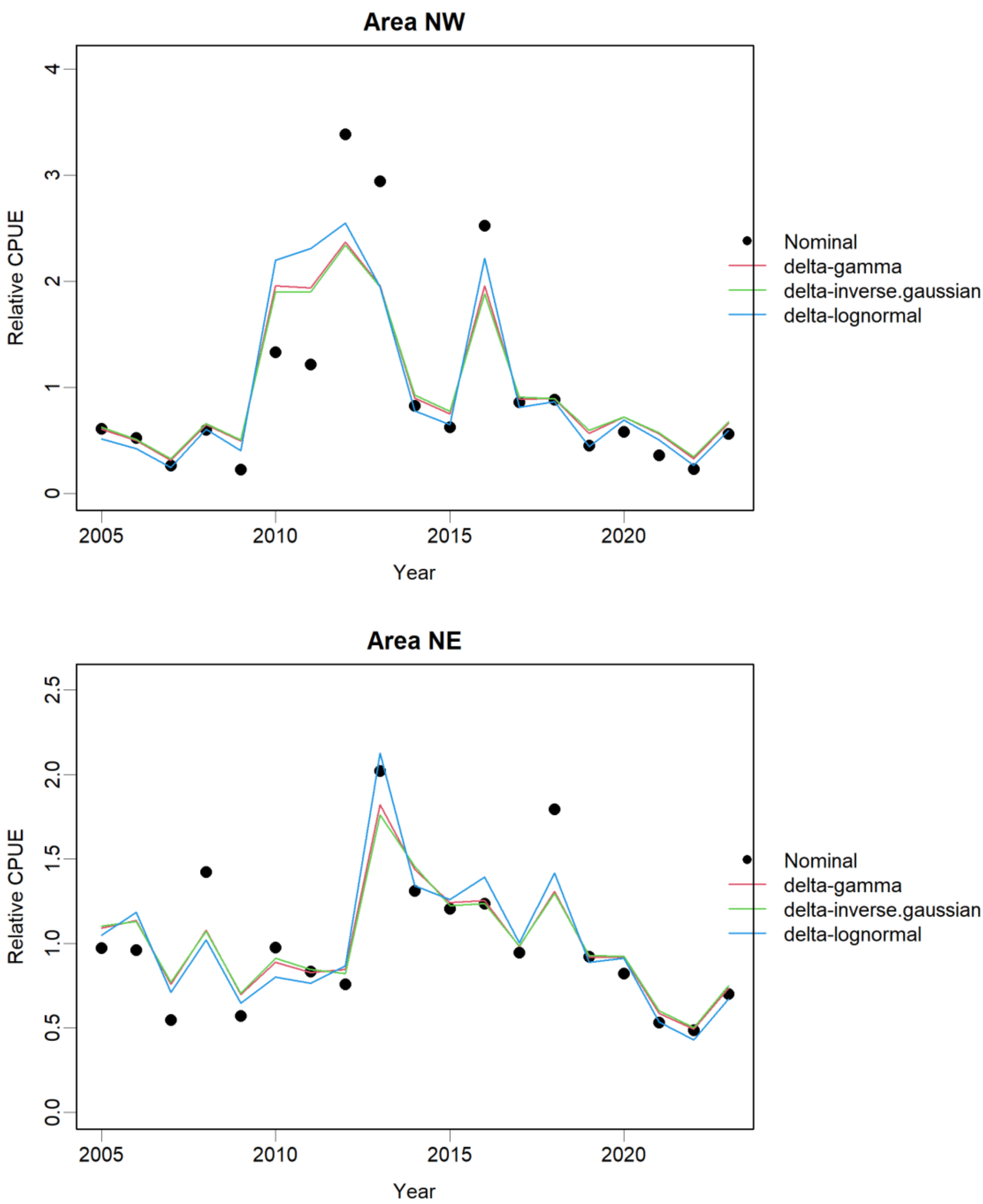



Fig. 12. Standardized CPUE series based on various GLMs for striped marlin caught by Taiwanese large-scale longline fishery in the Indian Ocean from 2005 to 2023.

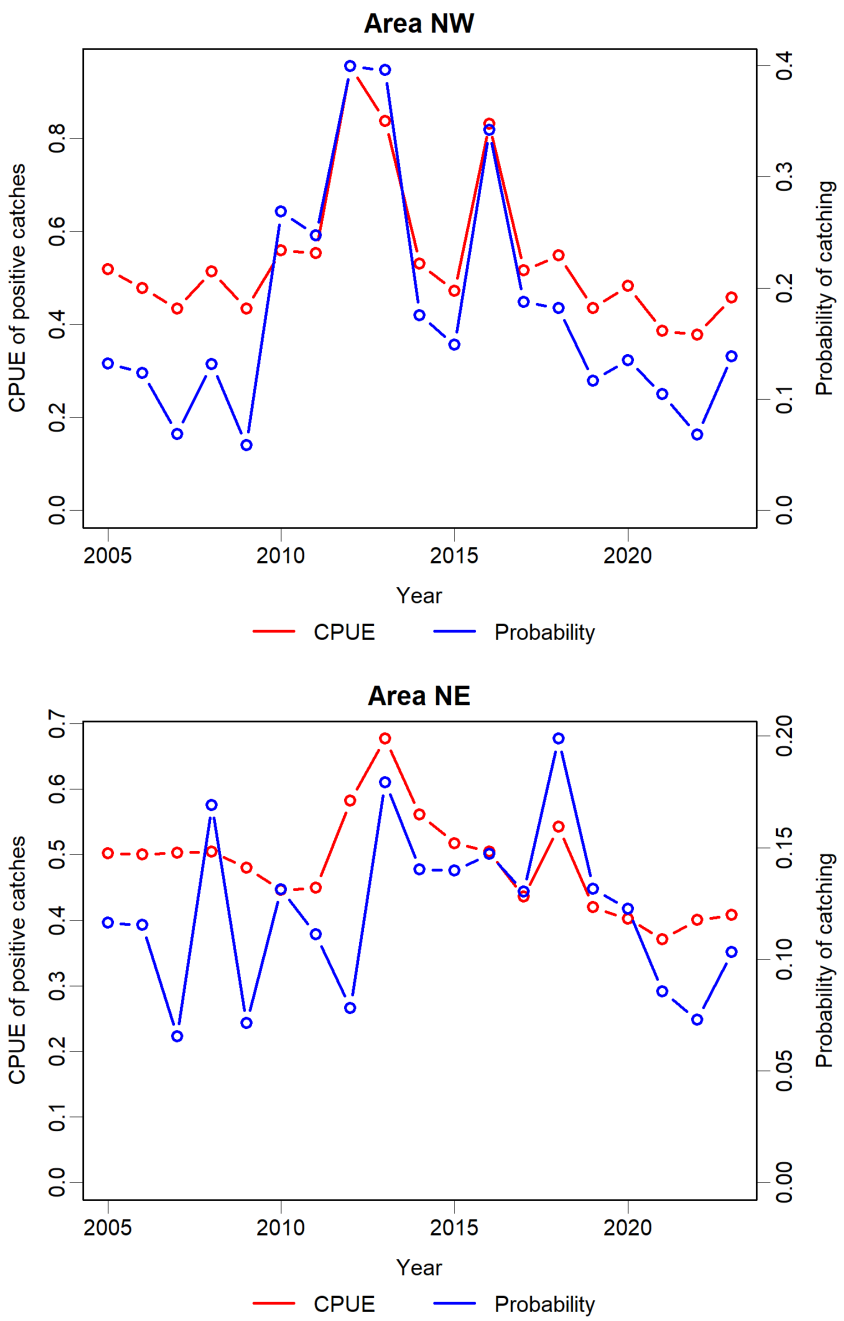



Fig. 13. Standardized CPUE of positive catches and catch probability based on selected model for striped marlin caught by Taiwanese large-scale longline fishery in the Indian Ocean from 2005 to 2023.

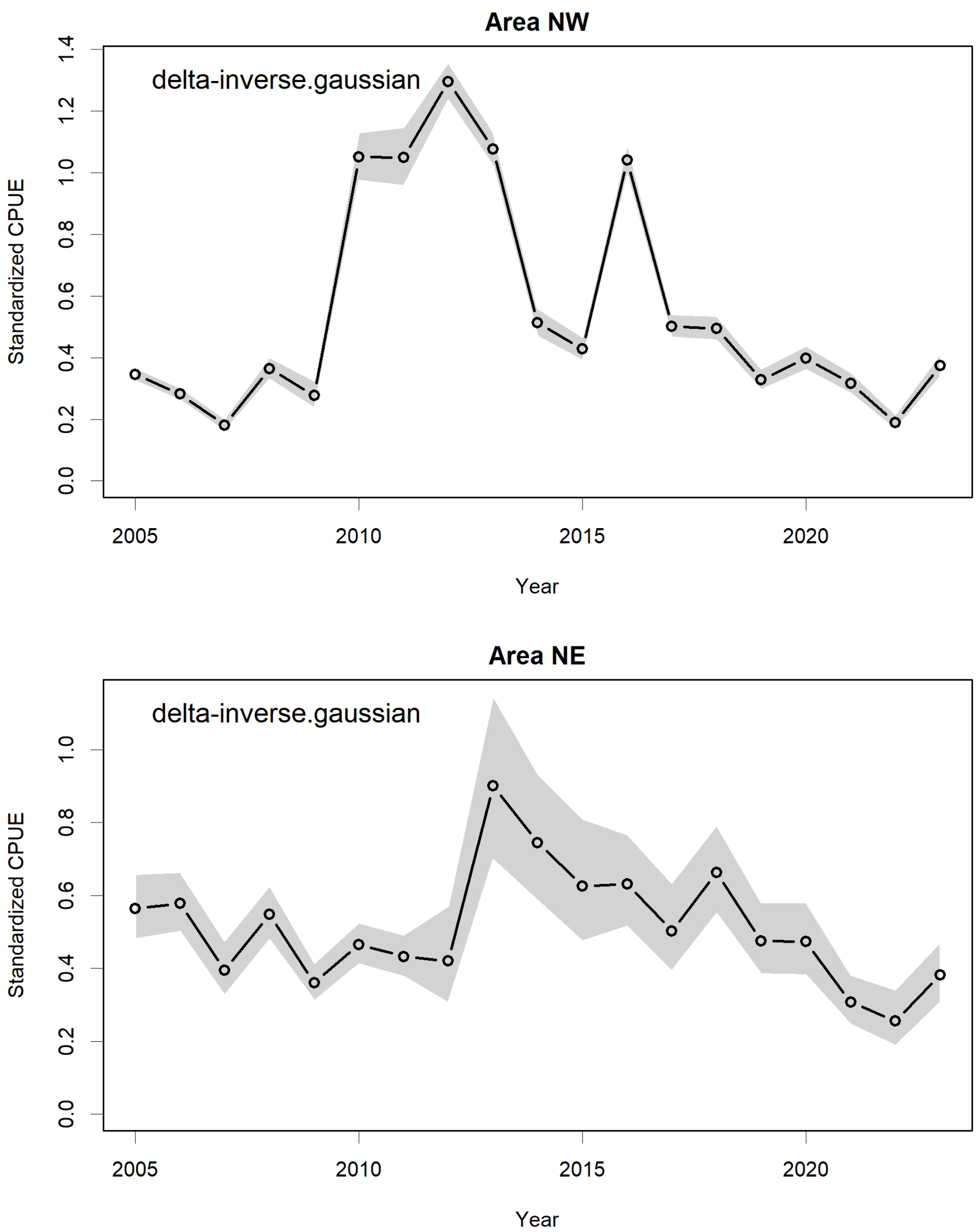



Fig. 14. Standardized CPUE series with 95% confidence intervals based on selected model for striped marlin caught by Taiwanese large-scale longline fishery in the Indian Ocean from 2005 to 2023.

Table 1. Diagnostic statistics for standardized CPUE series based on various models for positive catches of striped marlin caught by Taiwanese large-scale longline fishery in the Indian Ocean from 2005 to 2023.

| Area | Model | $R^2$ | AIC | BIC |
|---|---|---|---|---|
| | lognormal | 0.152 | 962,520 | 963,479 |
| NW | Gamma | 0.273 | 493,177 | 494,139 |
| | inverse Gaussian | 0.291 | 369,212 | 370,174 |
| | lognormal | 0.036 | 185,359 | 186,191 |
| NE | Gamma | 0.077 | 82,470 | 83,302 |
| | inverse Gaussian | 0.091 | 53,443 | 54,275 |

Table 2. ANOVA table for selected standardized CPUE series based on selected GLMs for striped marlin caught by Taiwanese large-scale longline fishery in the Indian Ocean from 2005 to 2023.

Area NW

Positive catch model with inverse Gaussian:

| | Sum Sq | Df | F values | Pr(>F) |
|---|---|---|---|---|
| Y | 1731.8 | 18 | 248.6 | 0 *** |
| Q | 14.5 | 3 | 12.5 | 3.5e-08 *** |
| G | 741.1 | 44 | 43.5 | 0 *** |
| T | 219.8 | 2 | 284.1 | < 2.2e-16 *** |
| Q:T | 56.1 | 6 | 24.2 | 9.8e-29 *** |
| Residuals | 21448.8 | 55429 | | |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Delta model:

| | LR Chisq | Df | Pr(>Chisq) |
|---|---|---|---|
| Y | 23796.9 | 18 | 0 *** |
| Q | 416.3 | 3 | < 2.2e-16 *** |
| CT | 190 | 2 | < 2.2e-16 *** |
| G | 8721.2 | 47 | 0 *** |
| T | 709.5 | 2 | < 2.2e-16 *** |
| hook | 167.4 | 1 | < 2.2e-16 *** |
| Q:CT | 99.4 | 6 | < 2.2e-16 *** |
| Q:T | 204.3 | 6 | < 2.2e-16 *** |
| CT:T | 34.8 | 4 | < 2.2e-16 *** |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 2. (continued).

Area NE

Positive catch model with inverse Gaussian:

| | Sum Sq | Df | F values | Pr(>F) |
|---|---|---|---|---|
| Y | 64.2 | 18 | 9.5 | 9.4e-27 *** |
| Q | 5.7 | 3 | 5 | 1.71E-03 ** |
| G | 50.4 | 34 | 4 | 8.38E-14 *** |
| T | 27.5 | 3 | 24.5 | 9.01E-16 *** |
| Q:CT | 7.3 | 6 | 3.3 | 3.4E-03 ** |
| Q:T | 7.9 | 9 | 2.3 | 1.23E-02 * |
| Residuals | 3852.4 | 10280 | | |

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

Delta model:

| | LR Chisq | Df | Pr(>Chisq) |
|---|---|---|---|
| Y | 1028.1 | 18 | < 2.2e-16 *** |
| Q | 88.7 | 3 | 9.48E-14 *** |
| CT | 56.7 | 2 | 4.77E-13 *** |
| G | 1805.1 | 39 | 0 *** |
| T | 99.2 | 3 | < 2.2e-16 *** |
| hook | 26.9 | 1 | 2.09E-07 *** |
| Q:T | 141.5 | 9 | < 2.2e-16 *** |
| CT:T | 24.8 | 6 | 3.71E-04 *** |

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