CPUE standardization of black marlin (*Makaira indica*) caught by Taiwanese large-scale longline fishery in the Indian Ocean

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

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

ABSTRACT

This paper briefly describes historical patterns of black marlin catches caught by Taiwanese large-scale longline fishery in the Indian Ocean. The cluster analysis was adopted to explore the targeting of fishing operations. In addition, the delta-inverse Gaussian generalized linear models were selected to conduct the CPUE standardizations of black marlin caught by Taiwanese large-scale longline fishery. The results indicate that the targeting effects (clusters) provided the most significant contributions to the explanation of the variance of CPUE for the models with positive catches, while the catch probability might be mainly influenced by the targeting of fishing operations. The standardized CPUE series obtained from different delta model assumptions revealed quite similar trends for all models except for the delta-lognormal model. 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 the northern areas (NW and NE) gradually increased until the mid-2010s, then declined from 2015 to 2022, before rising again in the last year.

1. INTRODUCTION

Black marlin is considered to be a bycatch species of industrial and artisanal fisheries. Gillnet fisheries have increased year by year, accounting for more than 50% of mean annual catch from 2017-2021, followed by line (30.4%) and longline (8.4%), with remaining catches recorded with other gears (4.4%) (IOTC, 2023). Catches have increased steadily since the 1990s, from 2,500 t in 1991 to over 13,000 t since 2004. In recent years, catches have further increased sharply from 13,000 t in 2012 to over 22,000 t in 2016 – the highest catches recorded in the Indian Ocean – due to increases reported by the offshore gillnet fisheries of Iran. Sri Lanka has developed gillnet and longline fisheries since the mid-1990s, and catches have continued to increase from 1,000 t to 3,000 t (IOTC, 2020). The catches were mainly made by Iran (gillnet, 35.8%), India (gillnet and troll, 20.5%) and Sri Lanka (gillnet, line and fresh longline, 17.5%) from 2017 to 2021. Taiwan has made only about 2% of total catches of black marlin in the Indian Ocean (IOTC, 2023).

IOTC conducted a stock assessment for black marlin in the Indian Ocean in 2021, but the results were highly uncertain due to a sharp increase in catches while the opposite trend of CPUE (IOTC, 2021). Therefore, this study conducted CPUE standardization for black marlin to obtain the relative abundance indices for further stock assessment.

2. MATERIALS AND METHODS

2.1. Catch and Effort data

In this study, daily operational catch and effort data (logbook) by 5x5 degree longitude and latitude grid for Taiwanese longline fishery during 1979-2023 were used. These data were provided by Overseas Fisheries Development Council of Taiwan (OFDC). For the area stratification, this study adopted the four areas stratification for black marlin (Fig. 1). It should be noted that the data in 2023 remains preliminary. For conducting the cluster analysis prior to the CPUE standardizations, the data were aggregated by 10-days duration (1st-10th, 11th-20th, and 21st~ in 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

Because black marlin was a bycatch species of Taiwanese longline fishery, a large amount of zero-catches was recorded in the operational catch and effort data sets. In previous studies, ignoring zero observations or replacing them with a constant was the most common approach. Nowadays, 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 that the use of the delta approach is appropriate for a high proportion of zero catches. Therefore, the delta-generalized linear models with different assumptions of error distribution were applied to conduct the CPUE standardization of black 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 (2021), the models were simply conducted with the main effects of year, quarter, longitude, latitude, and fishing targeting (clusters derived from species compositions of data sets, Wang et al., 2021). The models for positive catches and presence/absence data were conducted as follows:

For CPUE of positive catches:

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

Delta model for presence and absence of catch:

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

where *Catch* is the nominal catch in number of positive catch of black marlin (catch in number/1,000 hooks),

PA is the nominal presence and absence of catch,

Hooks is the effort of 1,000 hooks,

μ 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),
ε^{pos}	is the error term assumed based on various distribution,
ε^{del}	is the error term, ε^{del} ~ Binomial distribution.

To examine the appropriateness of the assumption of error distribution, this study applied lognormal, gamma and inverse Gaussian distributions for the error distribution of the model for the positive catches and specified "log" for the model link function.

The stepwise searches ("both" direction, i.e. "backward" and "forward") based on the values of the Akaike information criterion (AIC) were performed for selecting 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 standardized CPUE indices were calculated based on the estimates of the least square means of the interaction between the effects of year and area. The area-specific standardized CPUE trends were estimated based on the exponentiations of the adjust means (least square means) of the year effects (Butterworth, 1996; Maunder and Punt, 2004). The standardized relative abundance index was 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 standardized CPUE
	$CPUE$	is the adjust means (least square means) of the year effect of the model for positive catches,
	PA	is the adjust means (least square means) of the year effect of the model for presence/absence of catches.

3. RESULTS AND DISCUSSION

3.1. Historical fishing trends

Figs. 2 to 3 show the Taiwanese historical nominal catches (numbers) and CPUE distribution of black marlin based on the logbook data of Taiwanese large-scale longline fishery in the Indian Ocean. The catch of black marlin was mainly caught in tropical areas and offshore waters of the northern Indian Ocean. High CPUE also occurred in the tropical and offshore waters of the northern Indian Ocean over the years.

Black marlin catches were mainly caught in northern waters, especially for the

Area NW. Although the catches in the northwestern Indian Ocean increased significantly around 2012, the catches substantially decreased in the following years (Fig. 4 and Fig. 5).

3.2.Cluster analysis

Cluster analysis and CPUE standardizations were separately conducted for only northern areas (NW and NE, Fig. 1) since sparse catches of black marlin were made in the southern areas (Fig. 2).

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

Fig. 9 shows the black marlin catches and efforts by clusters and areas. Black marlin catches were contained in different clusters in different periods when different levels of efforts. Therefore, the data of all clusters were used to conduct further CPUE standardizations. The proportions of zero-catch of black marlin were very high by year for all areas (Fig. 10).

3.3. CPUE standardization

Based on the AIC model selection for the models for positive catches and delta model shows that some effects did not provide significant improvement to AIC and were excluded in different 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 R^2, AIC and BIC. (Table 1). 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 results obtained from the delta-inverse Gaussian models were selected to produce the standardized CPUE series for further stock assessment.

The ANOVA tables for selected models for each area are shown in Table 2. The results indicate that the effects of *G* (Lon and Lat) provided the most significant contributions to the explanation of variance of CPUE for the models for both positive catches and delta models in NW and NE areas. Thus, the catch rates derived from the positive catches of black marlin might be influenced by the targeting of the fishing operation

The area-specific standardized CPUE series are shown in Fig. 12. The CPUE series revealed quite similar trend for all delta model except for the delta-lognormal model in the area NE. The standardized CPUE of positive catches and catch probability obtained from the selected model are shown in Fig. 13. The trends of CPUE of positive catches and catch probability in the northern areas (NW and NE) were generally similar, and the tread of catch probability revealed obvious fluctuations.

The standardized CPUE series with 95% confidence intervals obtained from the selected model are shown in Fig. 14. The CPUE series in the northern areas (NW and NE) gradually increased until the mid-2010s, then declined from 2015 to 2022, before rising again in the last year.

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.

Butterworth, D.S., 1996. A possible alternative approach for generalized linear model analysis of tuna CPUE data. ICCAT Col. Vol. Sci. Pap., 45: 123-124.

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, 2018. Report of the 16th Session of the IOTC Working Party on Billfish. IOTC-2018-WPB16-R[E].

IOTC, 2020. Report of the 18th Session of the IOTC Working Party on Billfish. IOTC-2020-WPB18-R[E].

IOTC, 2021. Report of the 23rd Session of the IOTC Working Party on Tropical Tunas, Data Preparatory Meeting. IOTC-2021-WPTT23(DP)-R[E].

IOTC, 2023. Report of the 21st Session of the IOTC Working Party on Billfish. IOTC-2023-WPB21-R[E].

Kitakado, T., Wang, S.P., Satoh, K., Lee, S.I., Tsai, W.P., Matsumoto, T., Yokoi, H., Okamoto, K., Lee, M.Y., Lim, J.H., Kwon, Y., Su, N.J., Chang, S.T., Chang, F.C., 2021. Report of trilateral collaborative study among Japan, Korea and Taiwan for producing joint abundance indices for the yellowfin tunas in the Indian Ocean using longline fisheries data up to 2019. IOTC–2021-WPTT23(DP)-14.

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., Xu, W.Q., Lin, C.Y., Kitakado, T., 2021. Analysis on targeting of fishing operation for Taiwanese large scale longline fishery in the Indian Ocean. IOTC-2021-WPB19-11.

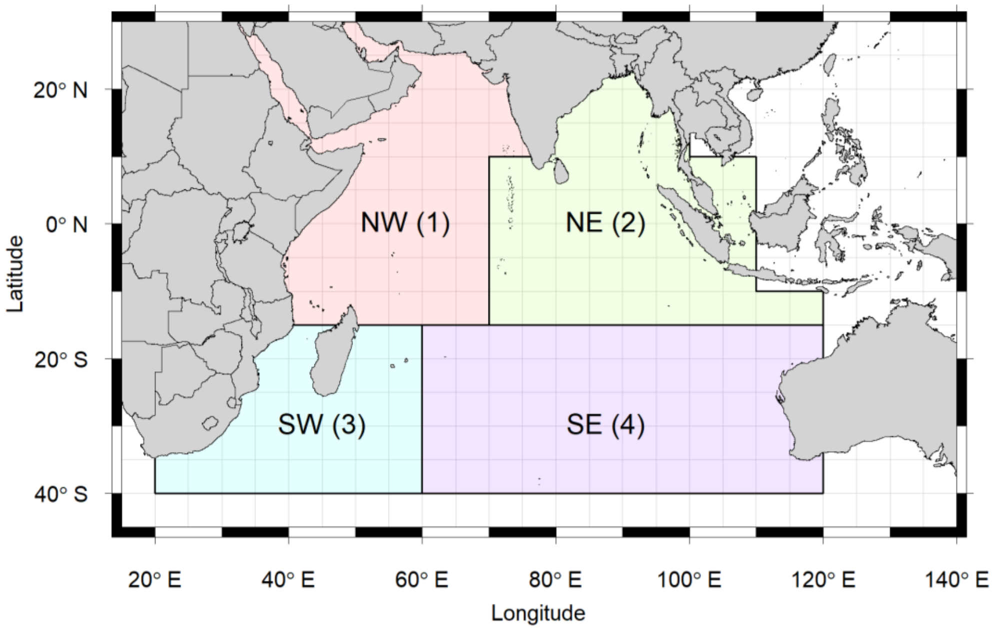


Fig. 1. Area stratification used for black marlin in the Indian Ocean.

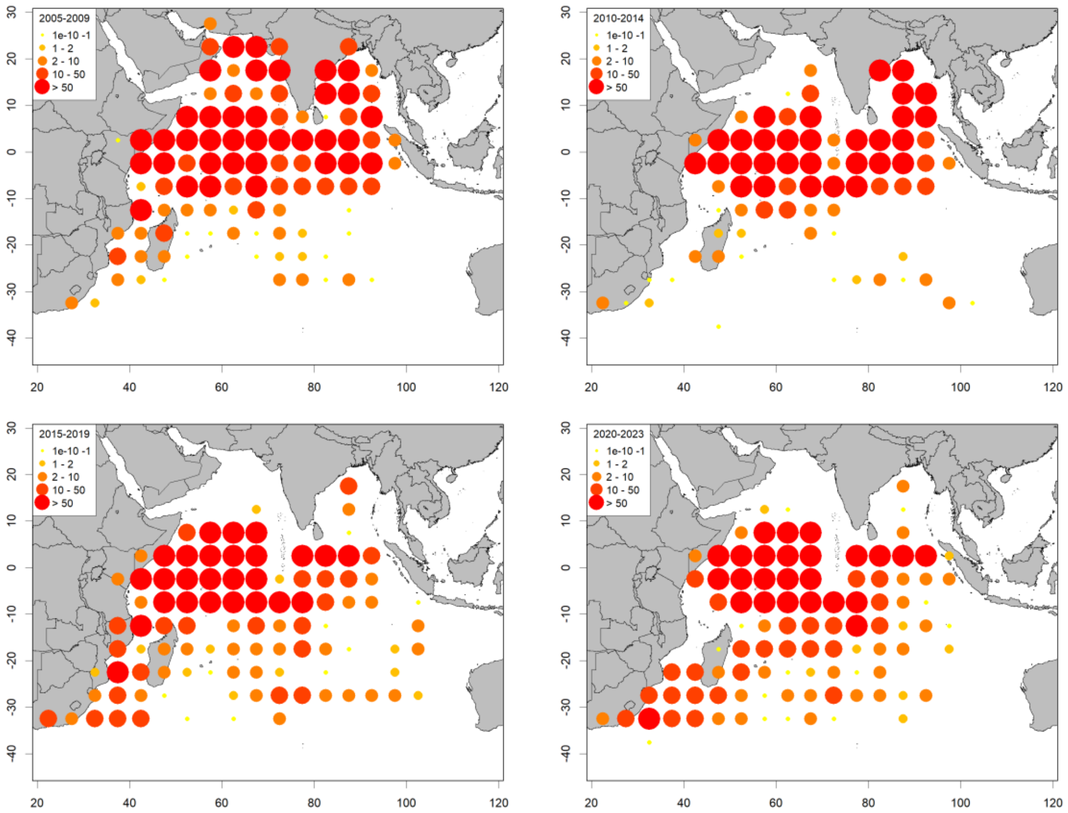


Fig. 2. Black marlin catch distribution of Taiwanese large-scale longline fishery in the Indian Ocean.

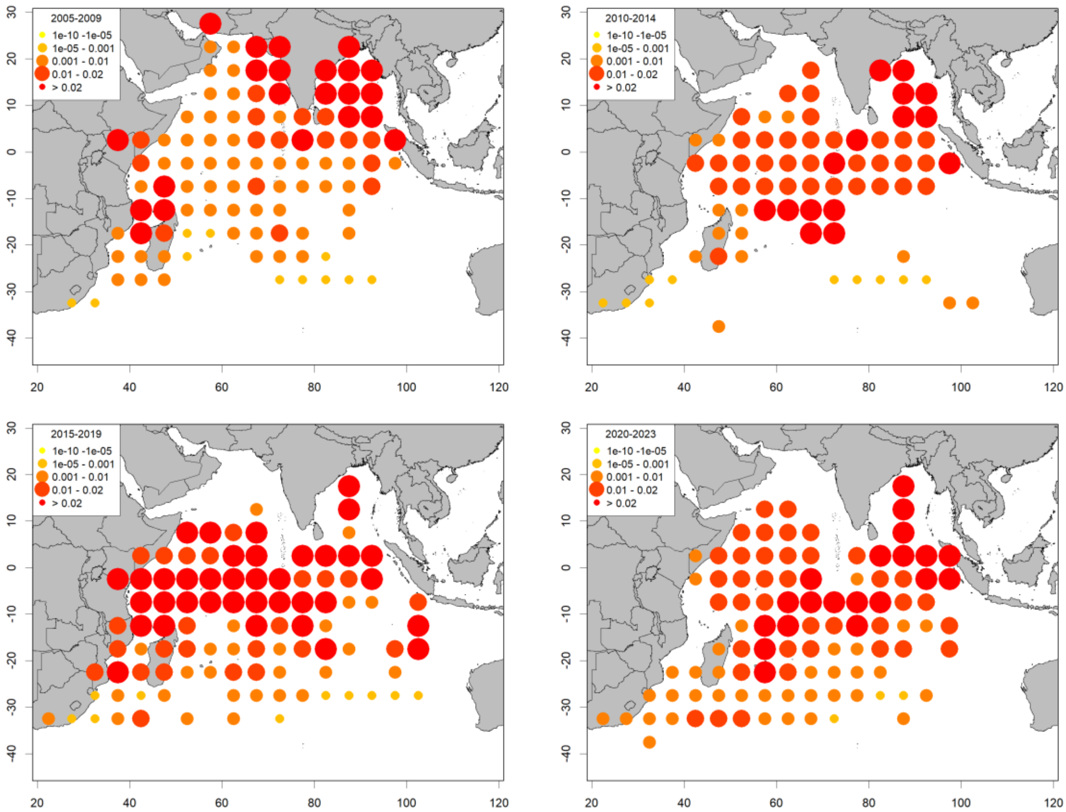


Fig. 3. Black marlin CPUE distribution of Taiwanese large-scale longline fishery in the Indian Ocean.

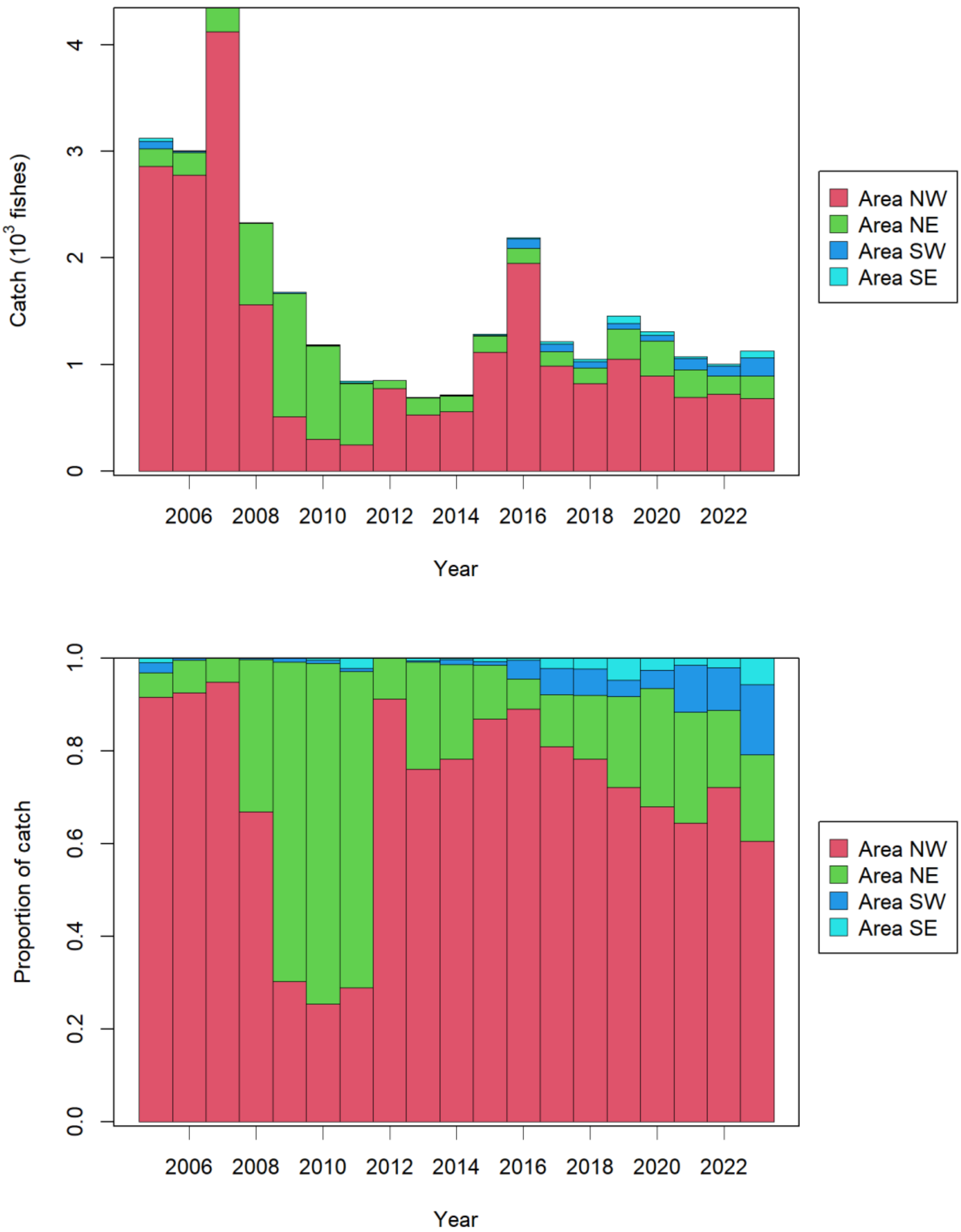


Fig. 4. Annual black marlin catches of Taiwanese large-scale longline fishery in the defined billfish area the Indian Ocean.

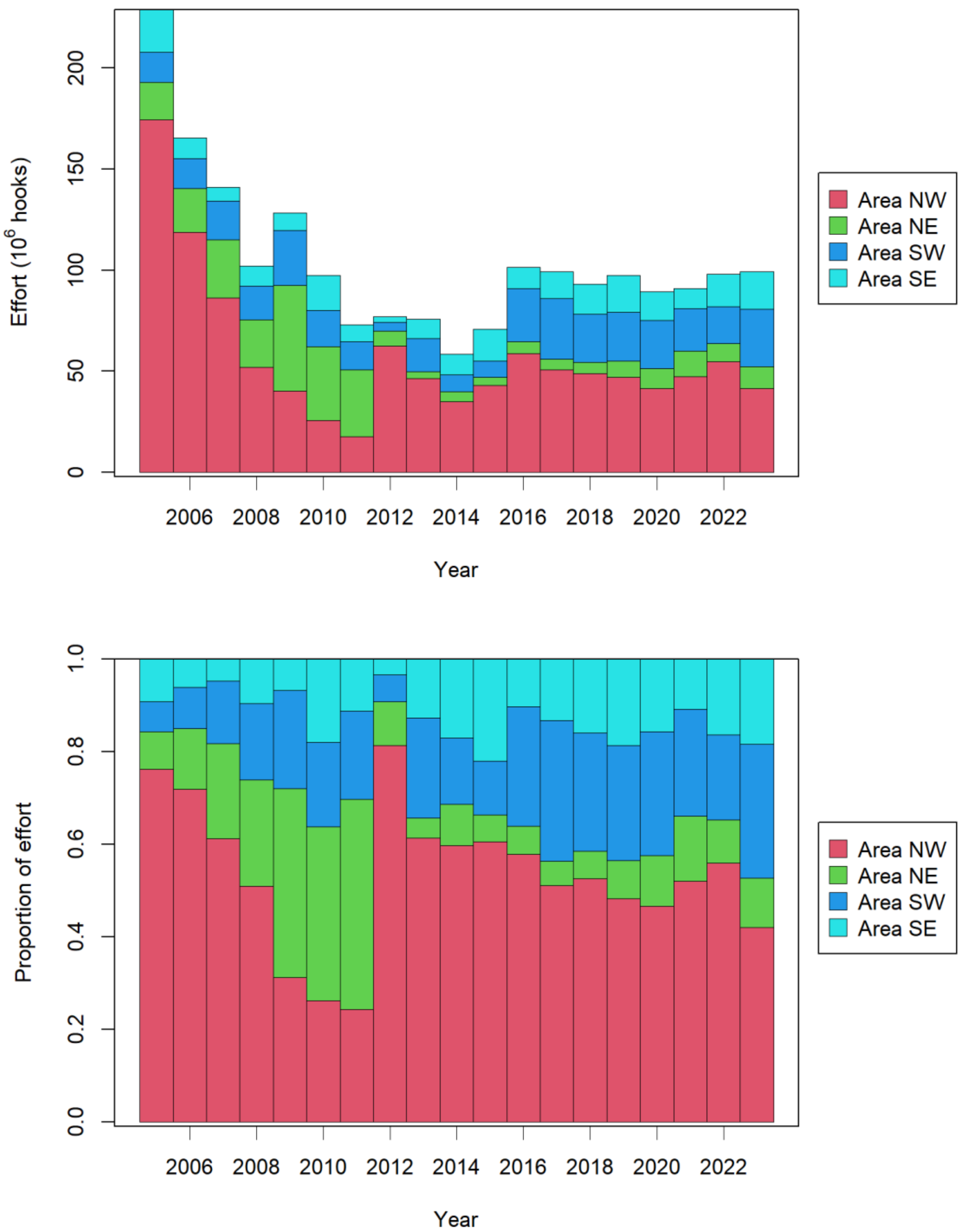
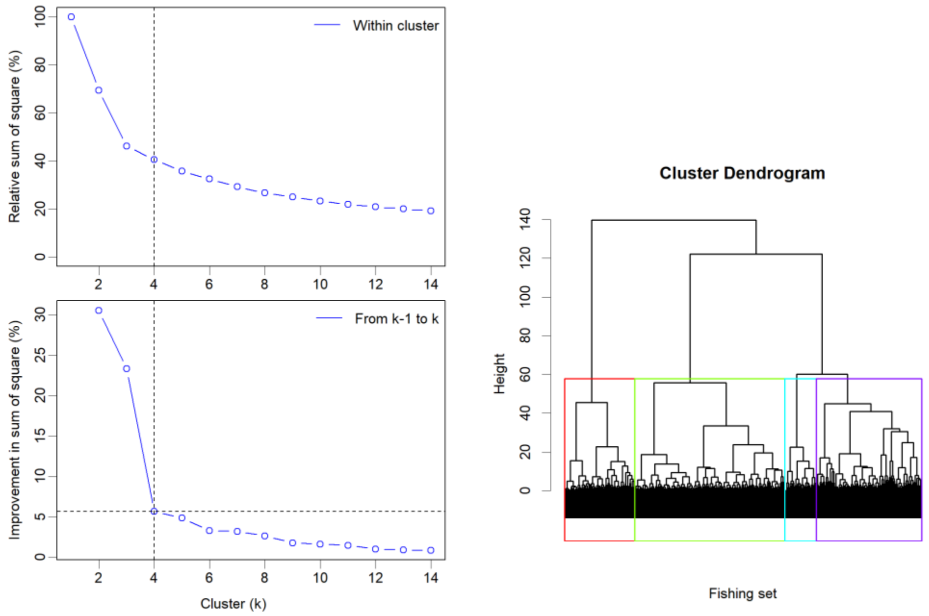


Fig. 5. Annual efforts (number of hooks) of Taiwanese large-scale longline fishery in the Indian Ocean.

Area NW



Area NE

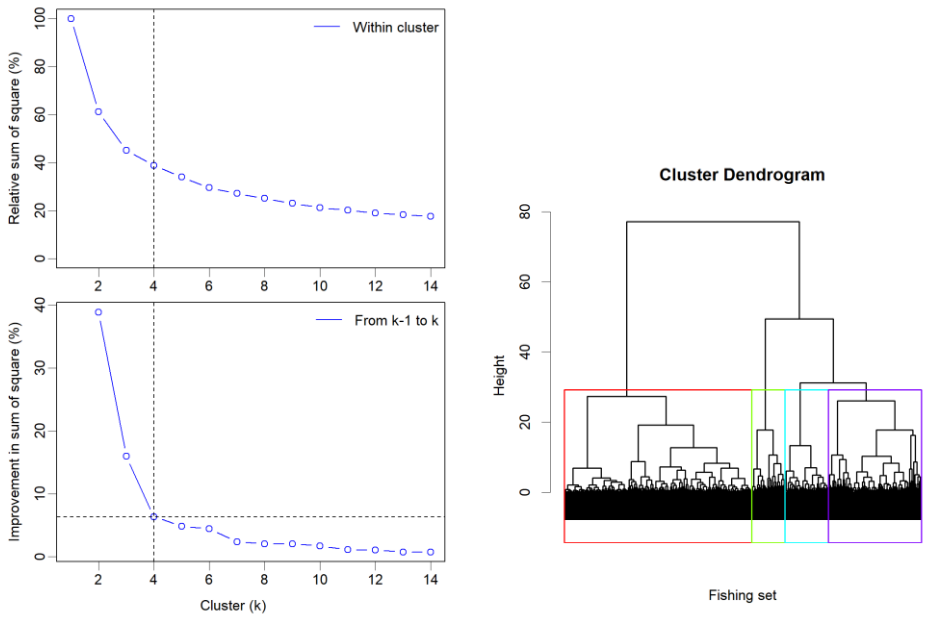


Fig. 6. Sum of squares within clusters for the data of Taiwanese large-scale longline fishery in billfish area of the Indian Ocean.

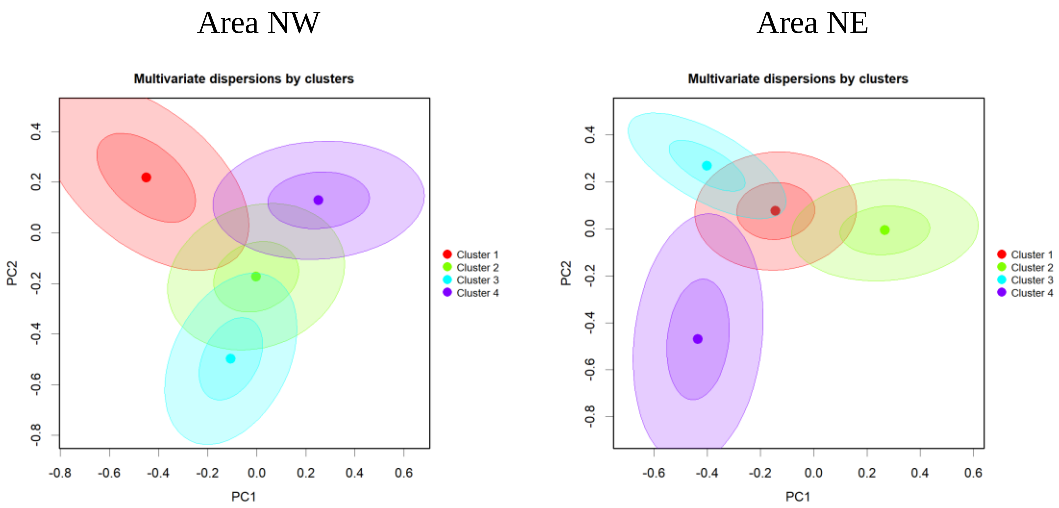


Fig. 7. Multivariate dispersions of the centroids by clusters derived from PCA for the data of Taiwanese large-scale longline fishery in billfish area of the Indian Ocean.

Area NW

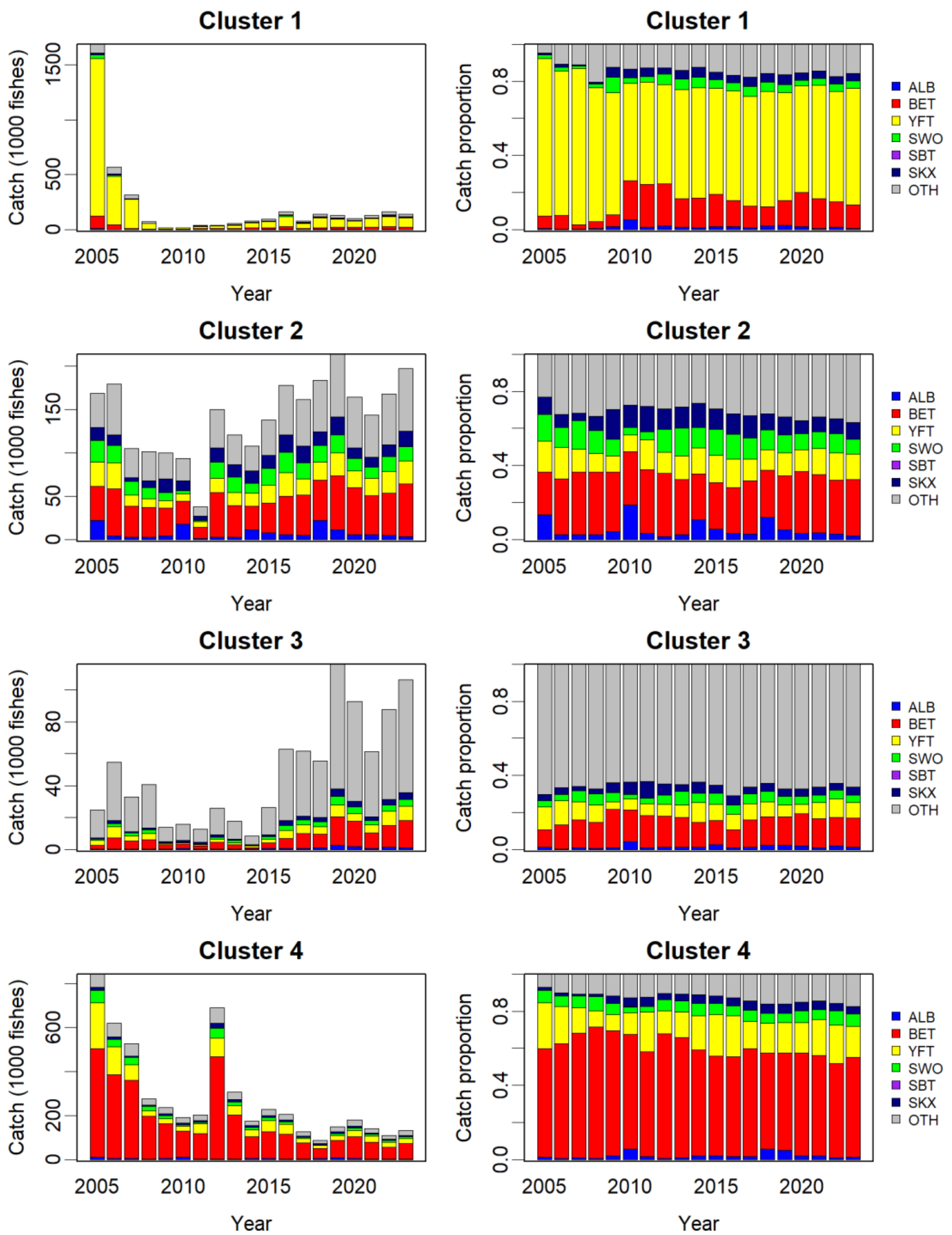


Fig. 8. Annual catches and compositions by species for each cluster of Taiwanese large-scale longline fishery in billfish area of the Indian Ocean.

Area NE

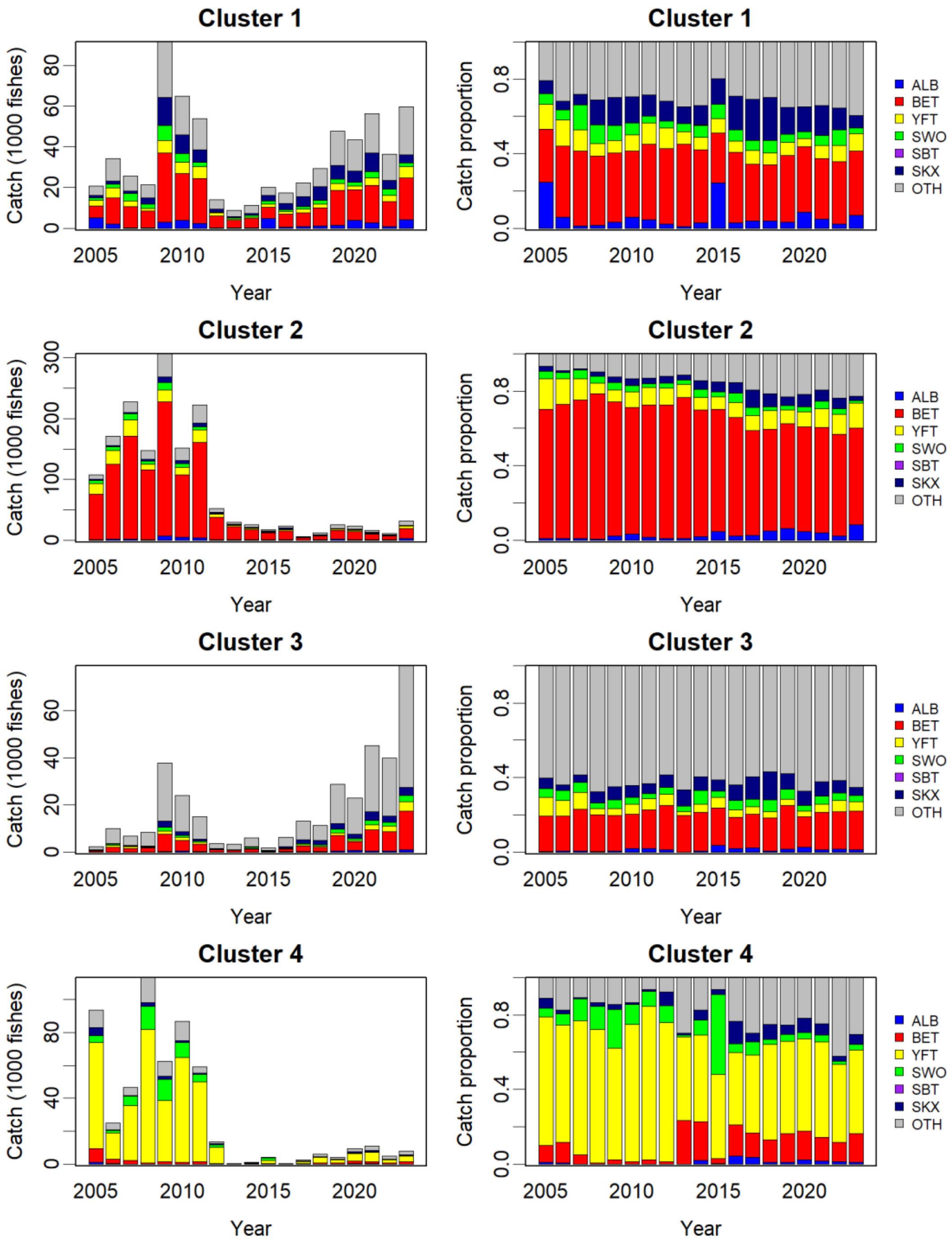


Fig. 8. (continued).

Area NW

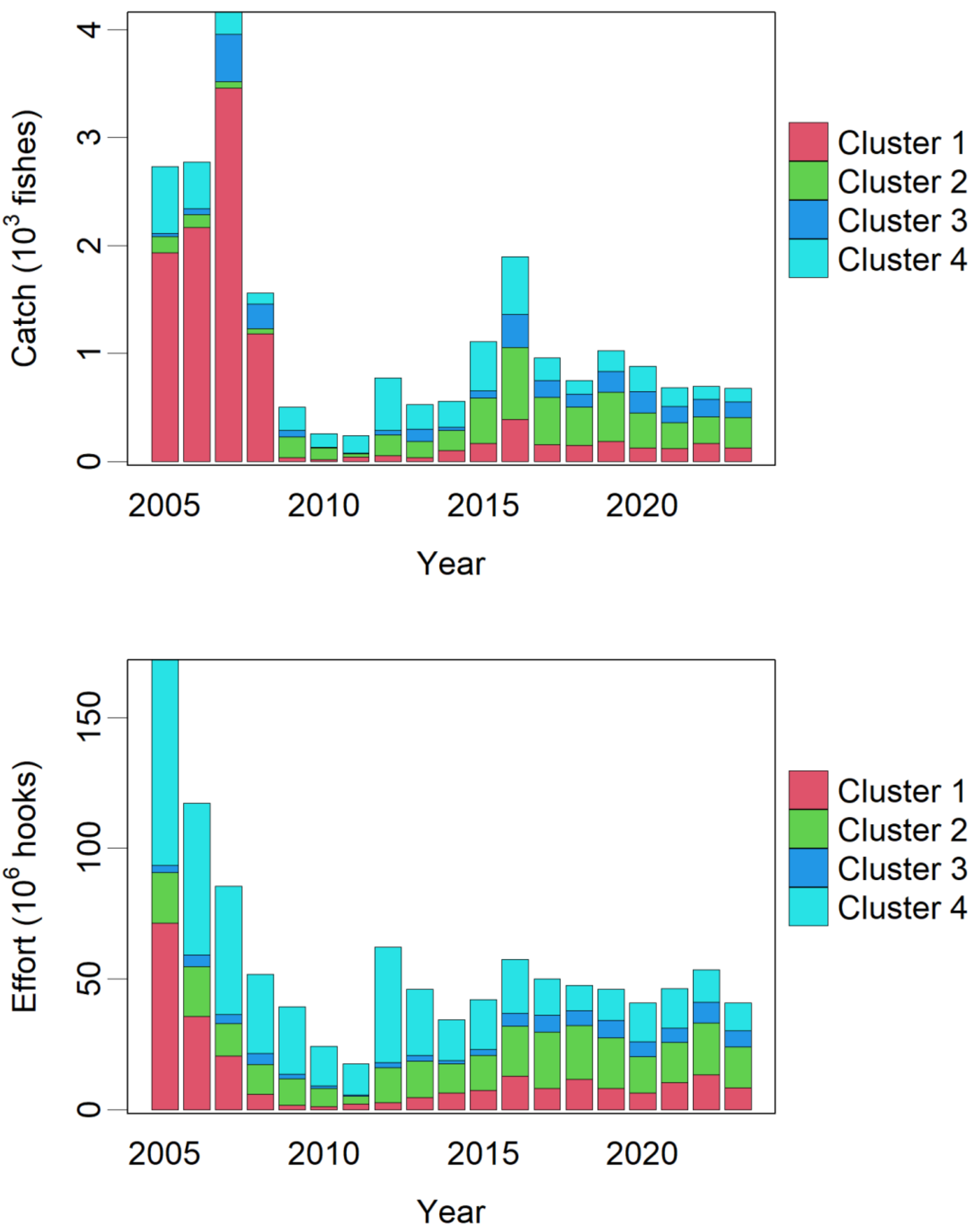
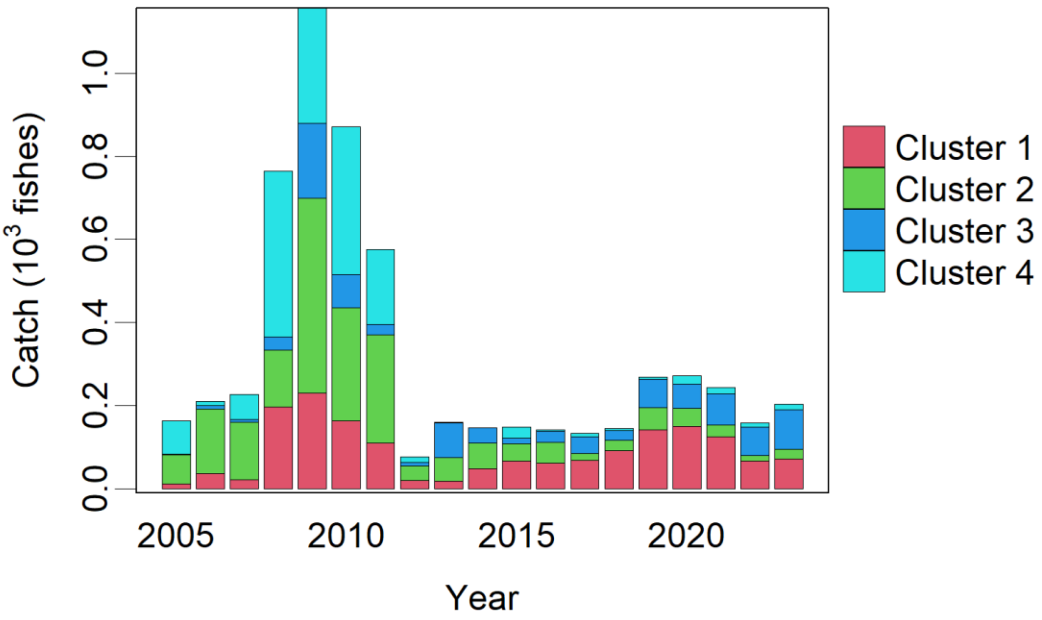


Fig. 9. Annual black marlin catches and efforts for each cluster of Taiwanese large-scale longline fishery in billfish area of the Indian Ocean.

Area NE



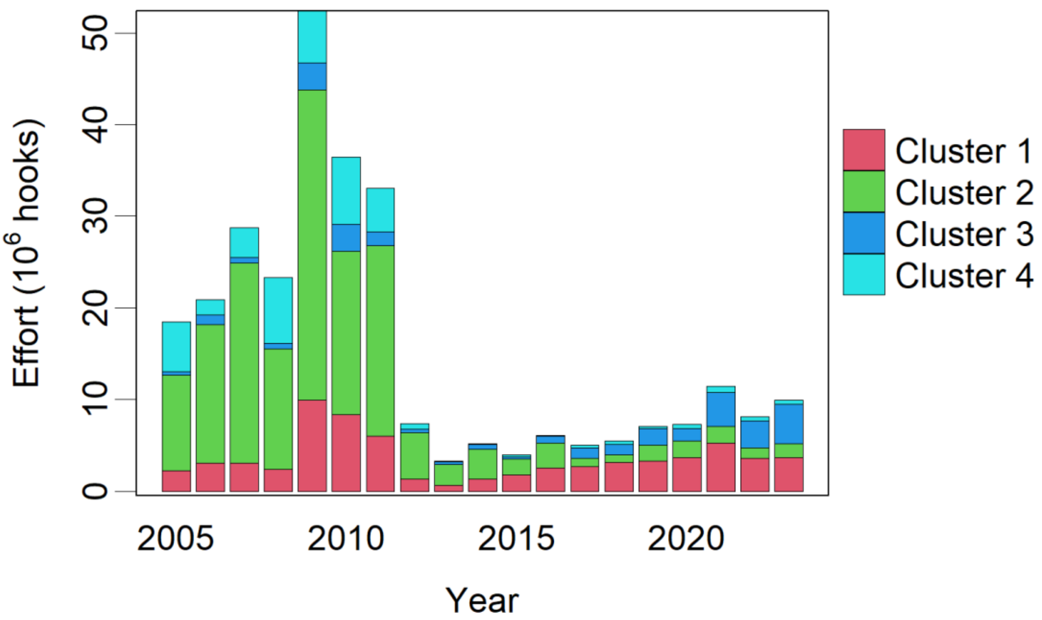
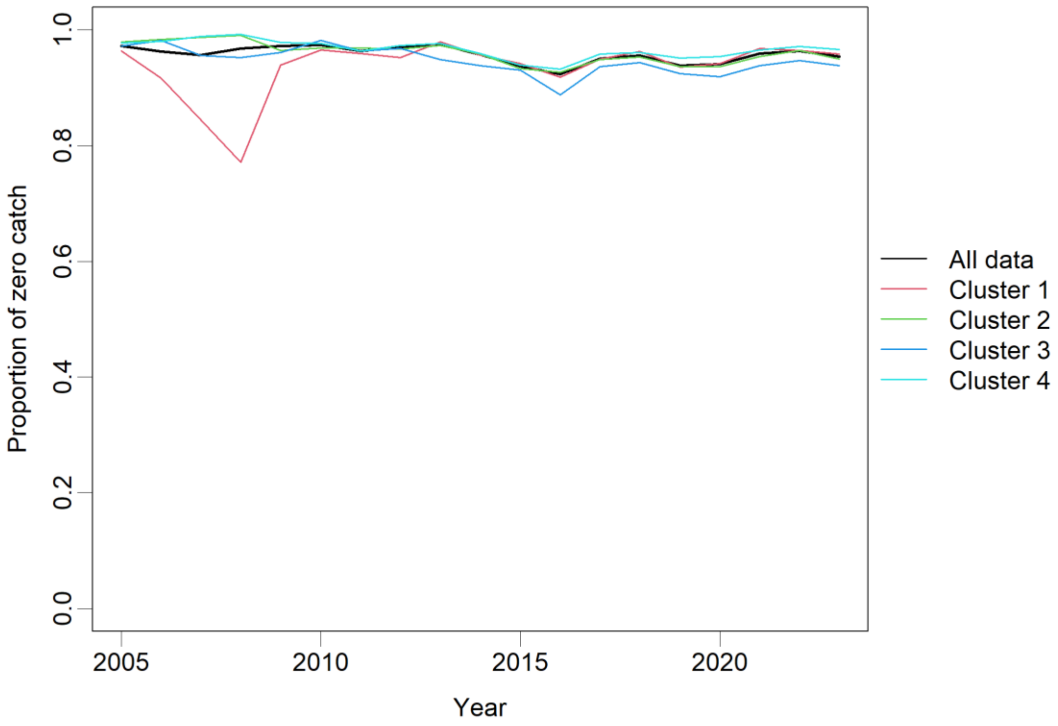


Fig. 9. (continued).

Area NW



Area NE

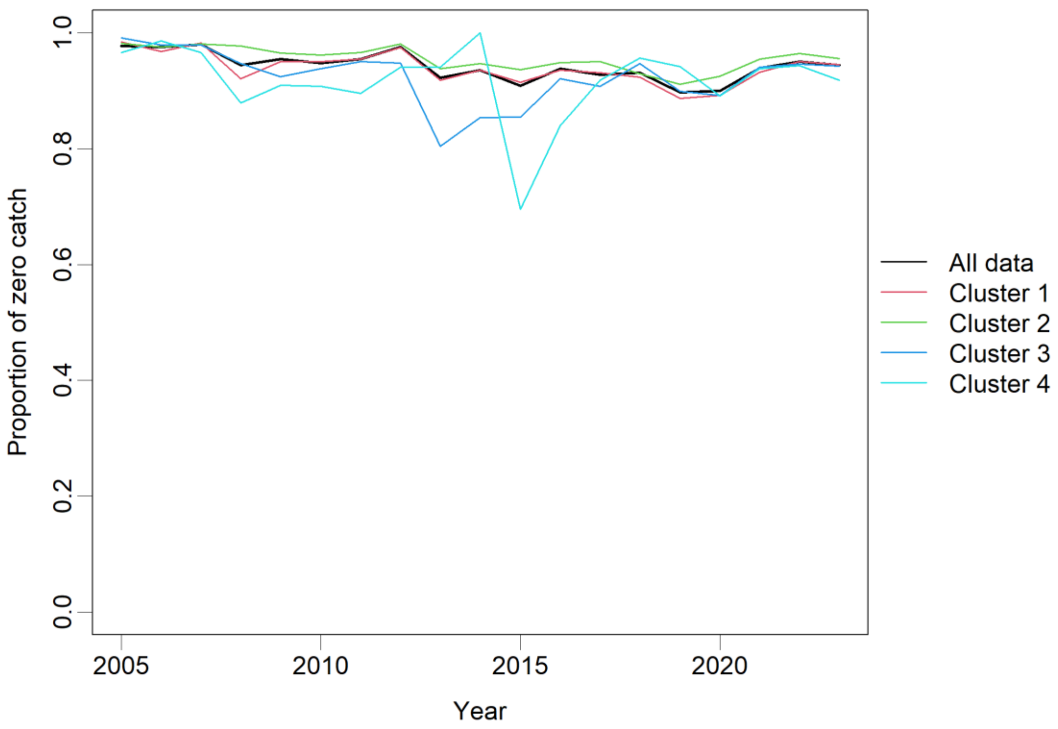


Fig. 10. Annual zero proportion of black marlin catches for each cluster of Taiwanese large-scale longline fishery in billfish area of the Indian Ocean.

Area NW

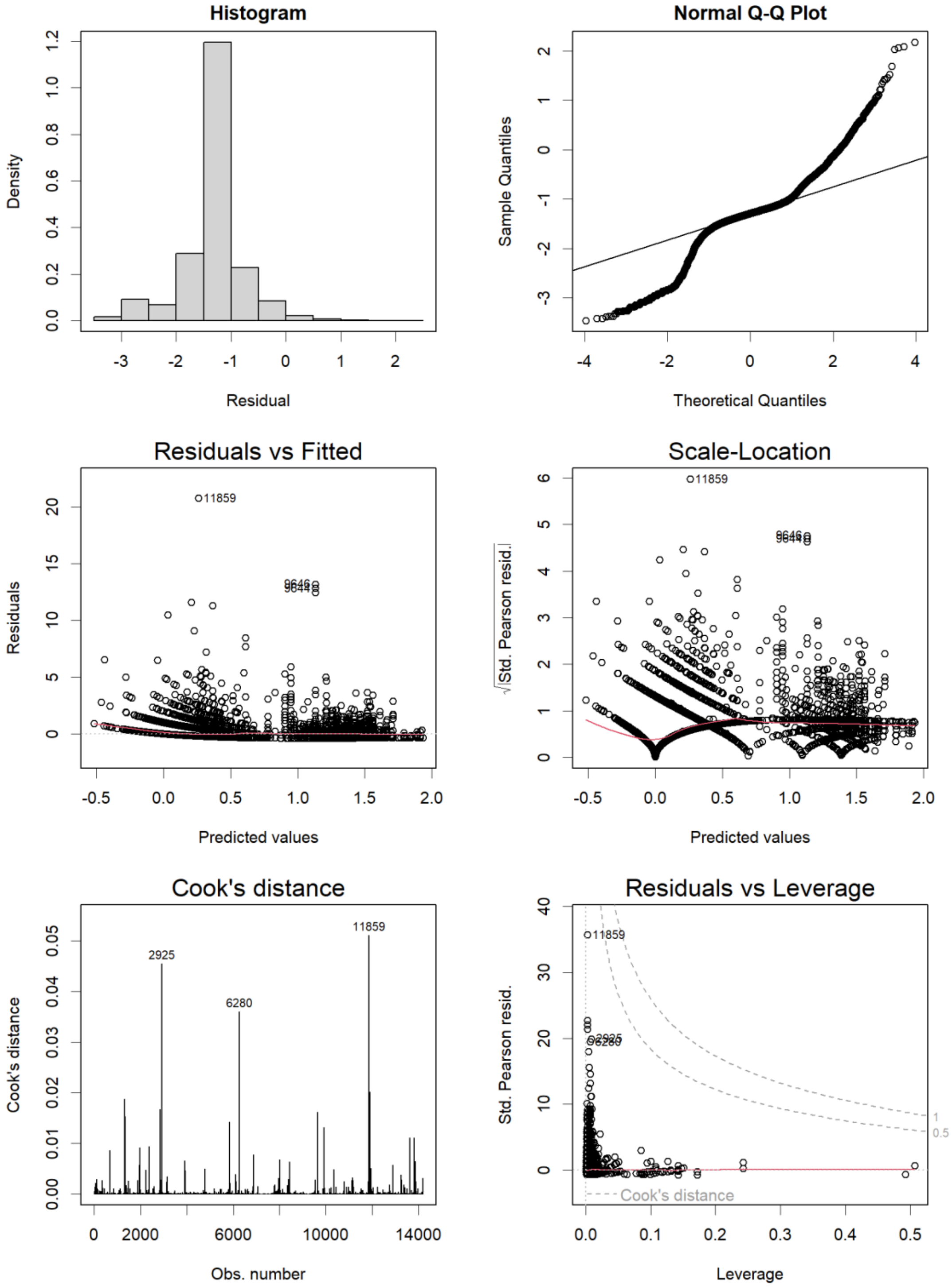


Fig. 11. Diagnostic plots for GLMs with inverse Gaussian error distribution assumption for black marlin caught by Taiwanese large-scale longline fishery in the Indian Ocean from 2005 to 2023.

Area NE

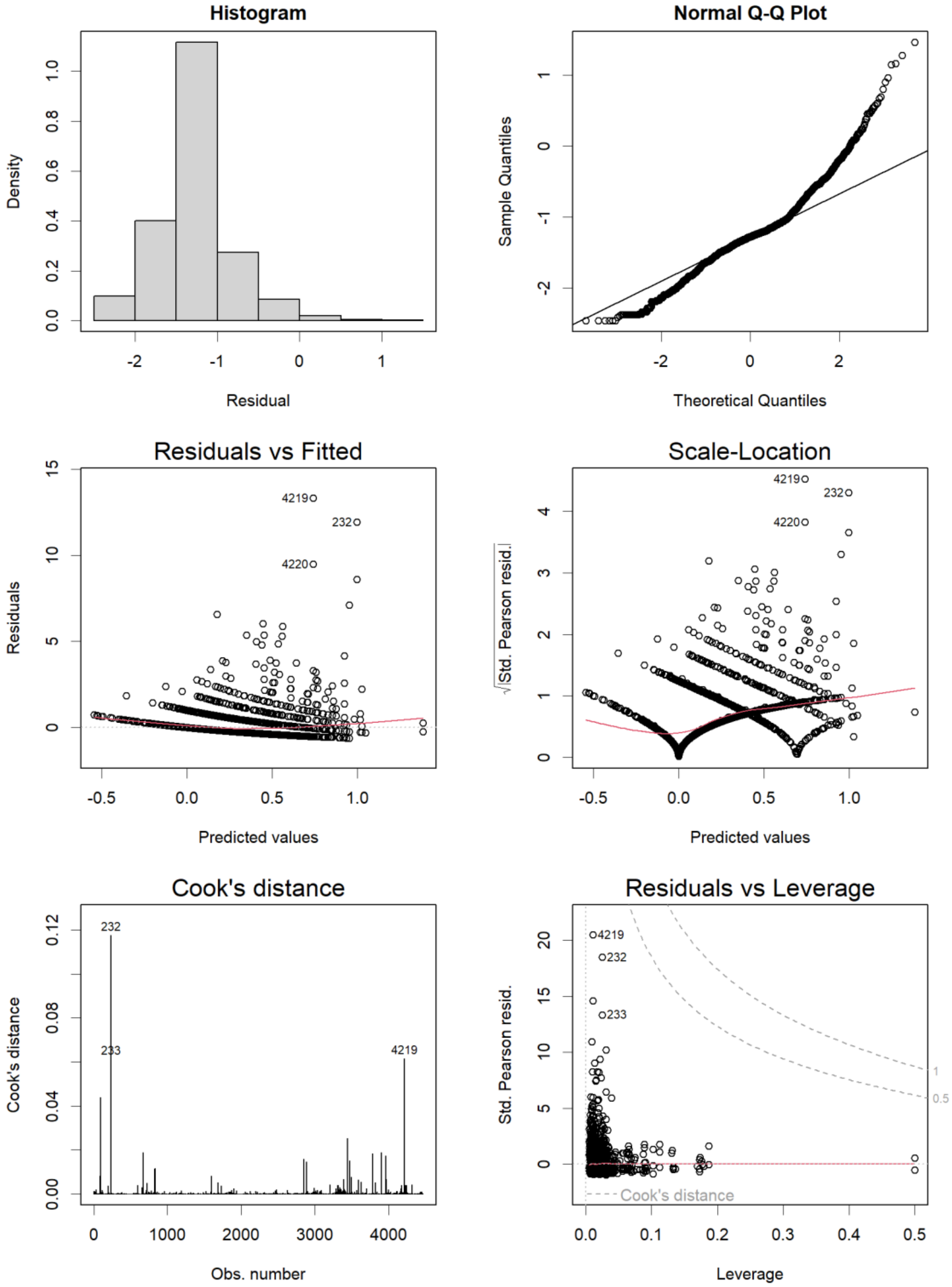


Fig. 11. (continued).

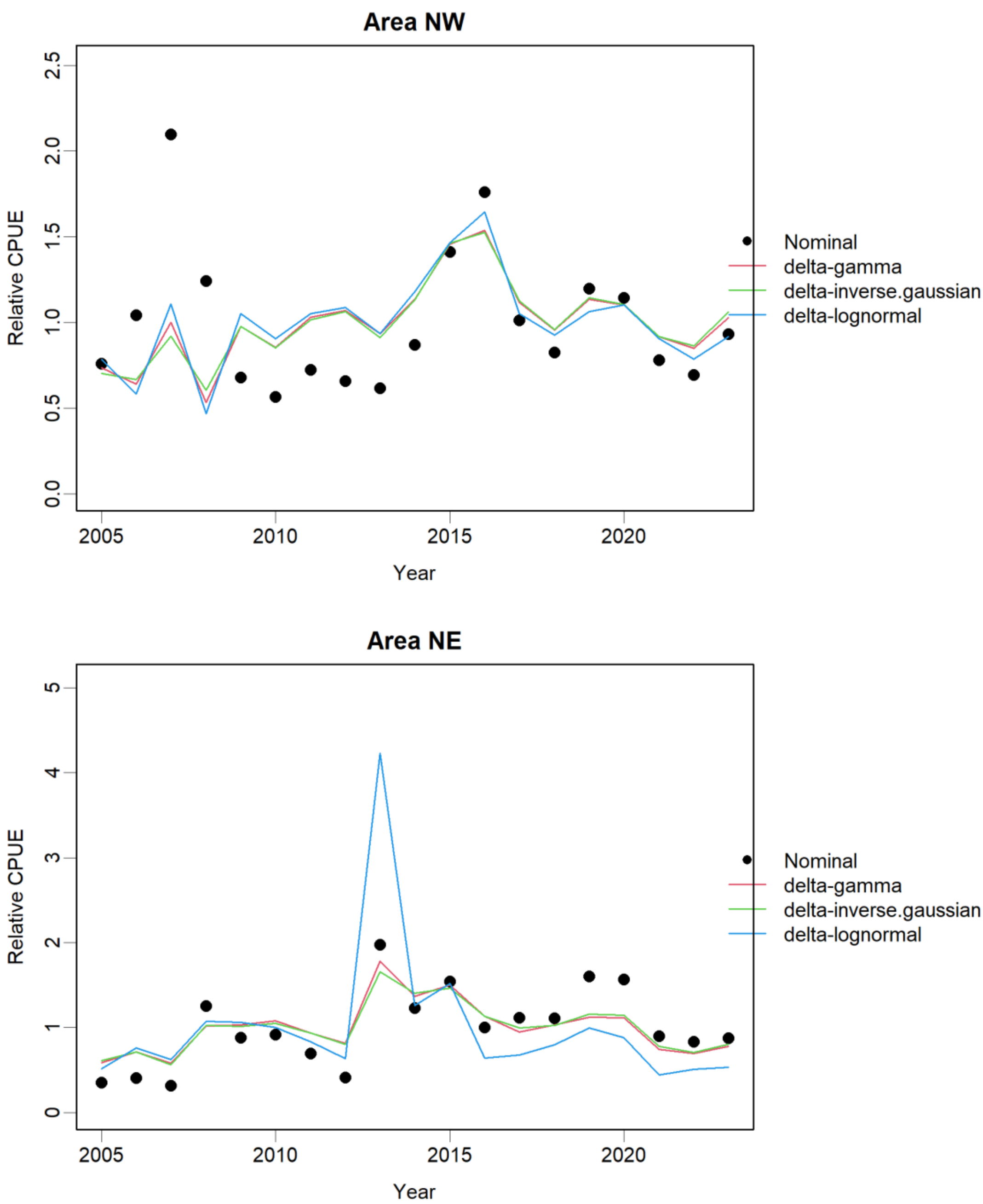
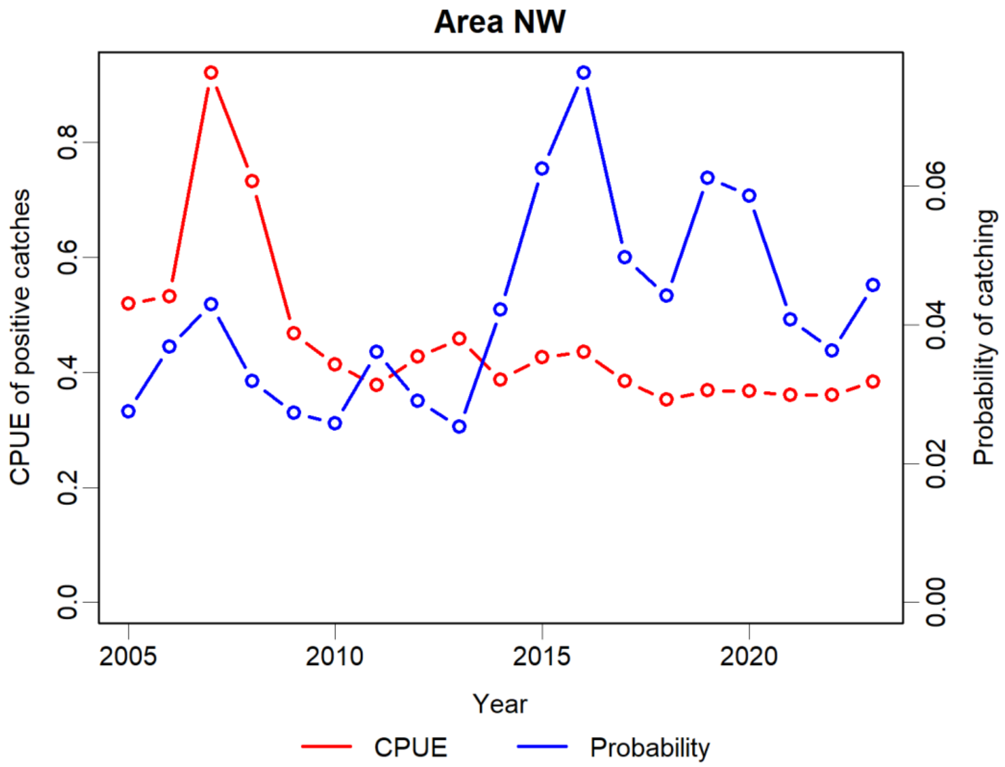


Fig. 12. Standardized CPUE series based on various GLMs for black marlin caught by Taiwanese large-scale longline fishery in the Indian Ocean from 2005 to 2023.



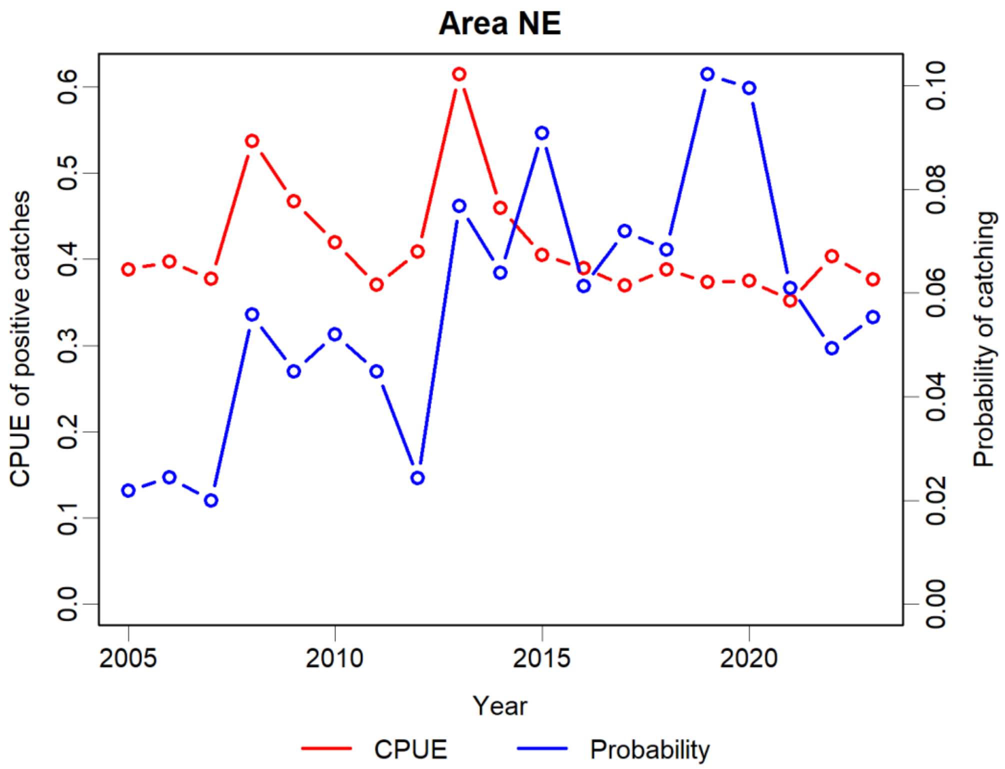


Fig. 13. Standardized CPUE of positive catches and catch probability based on selected model for black marlin caught by Taiwanese large-scale longline fishery in the Indian Ocean from 2005 to 2023.

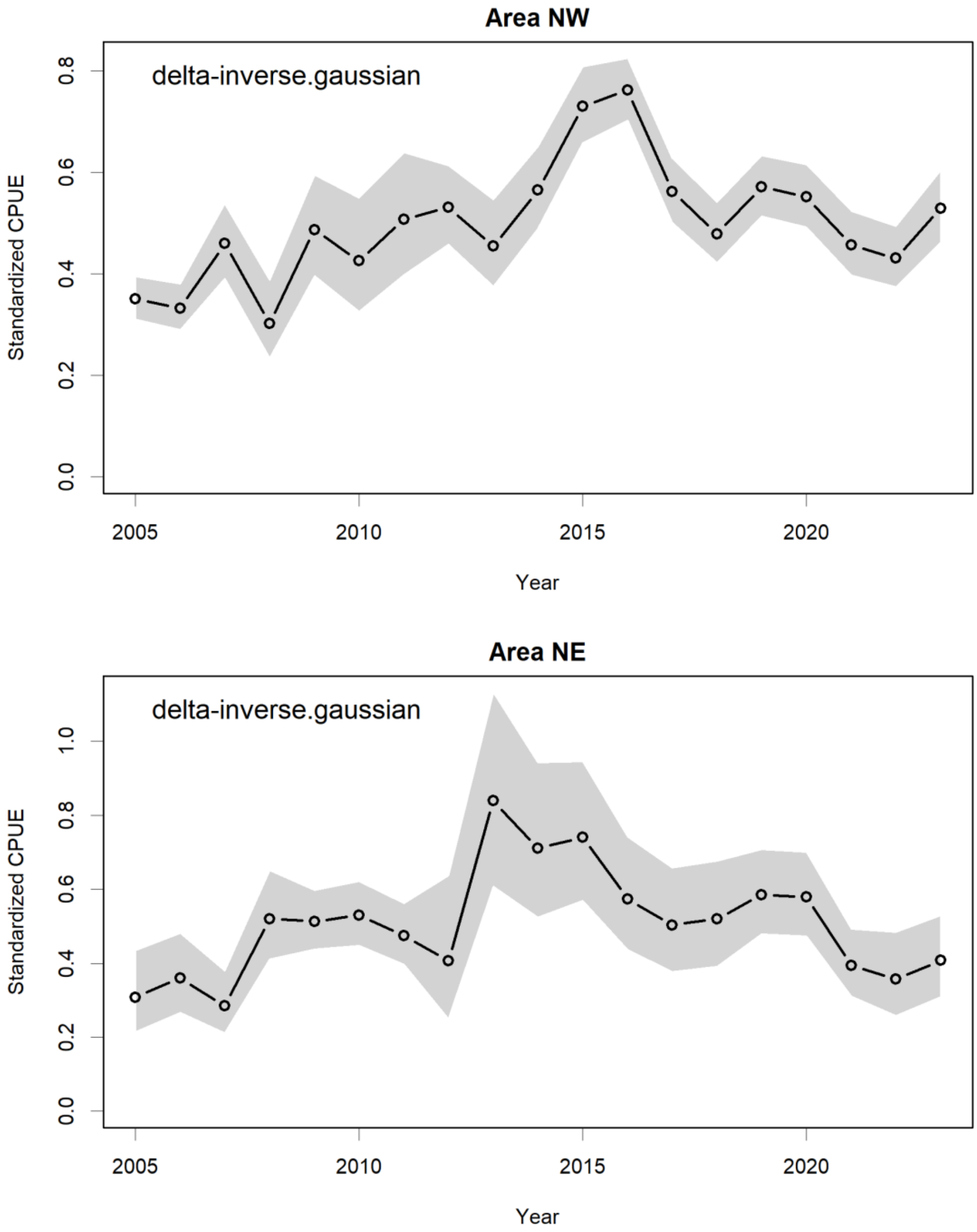


Fig. 14. Standardized CPUE series with 95% confidence intervals based on selected model for black marlin caught by Taiwanese large-scale longline fishery in the Indian Ocean from 2005 to 2023.

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

Area	Model	R2	AIC	BIC
NE	lognormal	0.187	13999	14531
	Gamma	0.188	7308	7801
	inverse Gaussian	0.192	5600	6017
NW	lognormal	0.148	61425	62000
	Gamma	0.332	28688	29263
	inverse Gaussian	0.347	21106	21680

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

NW

Positive catch model with inverse Gaussian:

	Sum Sq	Df	F values	Pr(>F)
Y	66.4	18	10.8780	1.22E-31 ***
Q	2.2	3	2.1726	0.089008942 *
G	238.0	41	17.1231	5.62E-118 ***
T	55.6	3	54.7244	3.68E-35 ***
Q:T	7.8	9	2.5592	0.006152709 **
Residuals	4783.6	14112		
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Binomial model

	LR Chisq	Df	Pr(>Chisq)
Y	2344.8	18	0
Q	101.4	3	7.75E-22 ***
CT	151.7	2	1.17E-33 ***
G	3042.6	47	0
T	183.2	3	1.77E-39 ***
hook	464.1	1	6.02E-103 ***
Q:CT	48.8	6	8.22E-09 ***
Q:T	32.1	9	0.000191 ***
CT:T	28.7	6	6.92E-05 ***
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1		

Table 2. (continued).

NE

Positive catch model with inverse Gaussian:

	Sum Sq	Df	F values	Pr(>F)
Y	27.3	18	5.3893	1.13E-12 ***
CT	3.7	2	6.6036	0.001369 **
G`	36.2	34	3.7824	1.08E-12 ***
T	12.8	3	15.1455	8.35E-10 ***
CT:T	3.4	6	2.0195	0.059652 *
Residuals	1238.9	4406		
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Binomial model

	LR Chisq	Df	Pr(>Chisq)
Y	546.0	18	2.21E-104 ***
Q	269.6	3	3.75E-58 ***
CT	74.8	2	5.69E-17 ***
G	432.2	39	8.64E-68 ***
T	71.1	3	2.47E-15 ***
hook	342.1	1	2.18E-76 ***
Q:CT	46.9	6	1.93E-08 ***
Q:T	53.7	9	2.15E-08 ***
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1		