Received: 26 August 2015

CPUE standardization of striped marlin (*Kajikia audax*) caught by Taiwanese longline fishery in the Indian Ocean using targeting effect derived from cluster and principle component analyses

Sheng-Ping Wang

Department of Environmental Biology and Fisheries Science, National Taiwan Ocean University, Keelung, Taiwan.

ABSTRACT

In this study, cluster analysis and principle component analysis were conducted based on catch composition of Taiwanese longline fishery in the Indian Ocean. Both of clusters and principle component scores can represent the historical fishing pattern related to character of targeting species. Also, there were appropriate relationships between numbers of hooks between float and clusters and principle component scores. Therefore, clusters and principle component scores can be adopted as proxy factors related to fishing characters when information of number of hooks between float is not available. In addition, this study provided a CPUE standardization of striped marlin (*Kajikia audax*) caught by the Taiwanese longline fishery in the Indian Ocean for time periods of 1980-2013. Since striped marlin is caught by Taiwanese longline fleet as bycatch species and large amount of zero catches are recorded in the operational data sets, the delta-lognormal GLM model is adopted to perform the CPUE standardization analysis. The trends of CPUE series in the northern Indian Ocean substantially decreased since 1980 although the CPUE obviously fluctuated in early years, while the CPUE series in the southern Indian Ocean revealed increasing trends with fluctuations before the mid-1990 and sharply decreased thereafter. The trend of area-aggregated CPUE series is similar to the CPUE series in the northern Indian Ocean, which reveals a continuously decreasing trend since 1980.

1. INTRODUCTION

Striped marlin are caught almost exclusively under drifting longlines (72%) with remaining catches recorded under gillnets and troll lines. The catches under drifting longlines have been recorded under Taiwan, Japan, Republic of Korea fleets and, recently, Indonesia and several NEI fleets. In recent years, the fleets of Taiwan (longline) and to a lesser extent Indonesia (longline) are attributed with the highest catches of striped marlin (IOTC, 2014).

Since striped marlin are bycatch species of Taiwanese lognline fleet, large amount of zero-catches are recorded from Taiwanese longline fleet and the proportions of zero-catch reached about 70-80% of total operation sets since 1998 (Wang and Nishida, 2013). Historically, ignoring zero observations or replacing them by a constant was the most common approach. Currently, the most popular way to deal with zeros is through the delta approach (Maunder and Punt, 2004). Therefore, the delta-lognormal GLM (Pennington, 1983; Lo et. al., 1992; Pennington, 1996) is applied to standardize the CPUE in this study.

The characters of fishing operation, such as number of hooks between float (NHBF), material of line, bait and etc., are known to be informative to describe the change in target species. Wang and Nishida (2011) also indicated that the model performance for CPUE standardization was significantly improved when including the effect of NHBF treated as categorical variable. However, NHBF data were available since 1995 and obstructed the incorporation of the effect of NHBF when conducting the CPUE standardization with data before 1995. He et al. (1997) and Hoyle et al. (2014) suggested alternative approaches to account for targeting in multispecies CPUE based on species composition, such as cluster analysis and principle component analysis (PCA).

In this paper, we attempted to classify the data sets in relation to species composition of the catches. The results of cluster analysis and PCA were also incorporated into CPUE standardization of striped marlin in the Indian Ocean as an effect related to fishing operation.

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 1980-2013 were provided by Oversea Fisheries Development Council of Taiwan (OFDC). The data of number of hooks between float (NHBF) were available since 1994 and the collection of NHBF data were more complete since 1995. Therefore, the data of NHBF may not be applicable to conduct the long-term CPUE standardization for fishes caught by Taiwanese longline fishery in the Indian Ocean.

2.2. Cluster Analysis

Cluster analysis can group catch records into categories by automatically identifying similar species-composition groups in the data and this approach has been applied to south Pacific albacore tuna using data from vessels in multiple fleets (e.g. Bigelow and Hoyle, 2012).

Cluster analysis was conducted based on species composition of the catches. Six main species groups were used in this study, including albacore (ALB), bigeye tuna (BET), yellowfing tuna (YFT), swordfish (SWO), striped marlin (MLS) and blue marlin (BUM) (Fig. 1). He et al. (1997) suggested a cluster analysis with two steps to classify the data sets because the large number of data sets precluded direct hierarchical cluster analysis. First, a non-hierarchical cluster analysis (K-means method) was used to group all data sets into 15 clusters for taking the mixture of fishing operations into account ($C_2^6 = 15$ ways in which 2 species can be chosen from 6 species groups). Second, a hierarchical cluster analysis with Ward minimum variance method was applied to the squared Euclidean distances calculated from 15 non-hierarchical clusters. Non-hierarchical and hierarchical cluster analyses were conducted using R functions *kmeans* and *hclust* (The R Foundation for Statistical Computing Platform, 2015).

He et al. (1997) indicated that the choice for the number of clusters to produce was largely subjective. At least two clusters (tuna sets and swordfish sets) were expected. More than two clusters were produced to allow other possible categories to emerge.

2.3. Principle Component Analysis

Hoyle et al. (2014) indicated that a new method to account for targeting in multispecies CPUE based on species composition has recently been developed, which uses scores from PCA as predictor variables in a CPUE standardization model. Previous studies also suggested that PCA may be more effective than the cluster analysis approach (Ortega-García and Gómez-Muňoz, 1992; Pech and Laloë, 1997; MacNeil et al., 2009; Winker et al. 2013; Winker et al. 2014).

In this study, the PCA was performed based on the linear regression models constructed of the catch compositions of six main species groups (ALB, BET, YFT, SWO, MLS and BUM).

$$PC_i = \beta_{1,i}ALB + \beta_{2,i}BET + \beta_{3,i}YFT + \beta_{4,i}SWO + \beta_{5,i}MLS + \beta_{6,i}BUM$$

where PC_i is the ith principle component, and $\beta_{x,i}$ (x = 1, 2, …, 6) is weighting for each composition, respectively.

The principal component scores, derived from the PCA of the catch composition data, were used as continuous nonlinear predictor variables for targeted effects in the CPUE standardization model. In this study, PCA was conducted using R functions *princomp* (The R Foundation for Statistical Computing Platform, 2015).

2.4. CPUE Standardization

The delta-lognormal GLM was applied to standardize the CPUE in this study and the main effects considered in this analysis are year, month, area, vessel scale and NHBF. The main effects considered in this analysis are year, month, area vessel, and cluster related to fishing type. Fishing areas used in this study were defined by four areas based on the IOTC statistics areas for swordfish in the Indian Ocean (Fig. 2) (Wang and Nishida, 2011). Hinton and Maunder (2004) indicated that interactions with the year effect would invalidate the year effect as an index of abundance. For the interaction associated with year effect, therefore, the interaction between year and area effect was only considered in models. The lognormal and delta models are conducted as follows:

Lognormal model for CPUE of positive catch:

$$\log(CPUE) = \mu + Y + M + A + T + \text{interactions} + \varepsilon^{\log}$$

Delta model for presence and absence of catch:

$$PA = \mu + Y + M + A + T + \text{interactions} + \varepsilon^{del}$$

where	$CPUE$	is the nominal CPUE of positive catch of striped marlin (catch in number/1,000 hooks),
	PA	is the nominal presence and absence of catch,
	μ	is the intercept,
	Y	is the effect of year,
	M	is the effect of month,
	A	is the effect of fishing area,
	T	is the effect of targeting (cluster related (CA) to fishing type or principal component scores (PC_i) derived from the ith principle component),
	ε^{log}	is the error term, $\varepsilon^{log} \sim N(0, \sigma^2)$,

ε^{del} is the error term, $\varepsilon^{del} \sim Bin(n, p)$.

The model selection is based on the values of the coefficient of determination (R^2), Akaike information criterion (AIC) and Bayesian information criterion (BIC). The standardized CPUE are calculated based on the estimates of least square means of the interaction between the effects of year and area.

The area-specific standardized CPUE trends are estimated based on the exponentiations of the adjust means (least square means) of the interaction between year and area effects (Butterworth, 1996; Maunder and Punt, 2004).

The standardized relative abundance index is calculated by the product of the standardized CPUE of positive catches and the standardized probability of positive catches:

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

2.5. Adjustment by area size

The estimation of annual nominal and standardized CPUE is calculated from the weighted average of the area indices (Punt et al., 2000).

$$U_y = \sum_a S_a U_{y,a}$$

Where U_y is CPUE for year y,

$U_{y,a}$ is CPUE for year y and area a,

S_a is the relative size of the area a to the four new areas.

The relative sizes of nine IOTC statistics areas for swordfish in the Indian Ocean (Nishida and Wang et al., 2006) were used to be aggregated into four areas used in this study.

Area	NW	NE	SW	SE
Relative area size	0.2478	0.2577	0.1638	0.3307

3. RESULTS AND DISCUSSION

3.1. Cluster Analysis

・ For non-hierarchical cluster analysis (K-means method), Table 1 shows the average proportions of catches for 15 K-means clusters. Cluster 14 consisted of large average proportion of MLS catches, which was 42.9%. The selection for number of clusters of hierarchical cluster analysis was based on the average proportions of MLS catches obtained from K-means method.

For hierarchical cluster analysis, he et al. (1997) suggested considering the number of clusters until the smallest cluster contained less than 10% of the total number of data sets. In case of this study, selecting 5 clusters can achieve criterion of He et al. (1997). However, most clusters consisted of ALB, BET and YFT when 5 clusters were selected and average catch proportions of other species substantially decreased for every cluster (results not shown). This may reduce the effectiveness of cluster effect related to fishing type for species other than ALB, BET and YFT when conducting CPUE standardization by GLM. Finally, 6 clusters were chosen and Clusters 4 and 2 contain much more average proportions of SWO catches than other clusters (Table 2 and Fig. 3).

Figs. 4 and 5 show the historical catches by species and species compositions. It is obvious that data sets assigned to Clusters 1, 2 and 3 belonged to the fishing operations targeting in BET, ALB and YFT, respectively. Data sets assigned to Clusters 4 and 6 were fishing operations targeting in multispecies and Clusters 4 and 6 consisted of relatively large catches of BET and SWO, respectively. Data sets assigned to Cluster 5 consisted catches of species other than main tunas and billfishes.

Based on the historical composition of clusters (Fig. 6), high proportion of ALB cluster (Cluster 2) occurred before mid-1980s and the proportions of BET clusters (Clusters 1 and 4) and YFT cluster (Cluster 3) obviously increased thereafter. The proportions of BET, YFT and SWO-mixed Cluster (Cluster 6) increased since early-1990s. In addition, relatively high proportion of the cluster of other species (Cluster 5) also occurred after 2005.

Fig. 7 shows the frequencies of NHBF by clusters based on the data after 1995. NHBF mainly concentrated on 15 and 10 hooks for BET (Cluster 1) and ALB clusters (Cluster 2), respectively. Multiple modes were observed for YFT and multispecies (Clusters 3, 4, 5 and 6) clusters and this may be resulted from changing in fishing operation for different target species.

3.2. Principle Component Analysis

Based on the results of PCA, the first principle component (PC1) explained about 54% variance of observations, and cumulative proportions of explained variance of the 2nd (PC2) and 3rd (PC3) principle components reached to 85% and 95% (Table 3). According to the weightings of variables for principle components, PC1 can be the targeting indices for ALB (positive direction) and BET (negative direction), PC2 can be the targeting indices for YFT (negative direction) and BET (positive direction), and PC3 can be targeting indices for tunas (YFT, BET and ALB) (negative direction) and SWO (positive direction) (Figs. 8-10, Table 4).

Fig. 11 shows the historical distributions of principle component scores for PC1, PC2 and PC3. For PC1, most of principle component scores were positive before late-1980s (i.e. targeting in ALB), while most of principle component scores were negative thereafter (i.e. targeting in BET). For PC2, more negative principle component scores occurred during late-1980s and mid-1990s (i.e. targeting in YFT), while negative values decreased thereafter due to high proportions of catches occurred for both BET and YFT. For PC3, positive principle component scores gradually increased since early 1990s and this represented the SWO catch trend, especially for years after early 2000s.

Fig. 12 shows the distributions of principle component scores by NHBF for PC1, PC2 and PC3. For PC1, most of principle component scores were positive when NHBF were less than 12 hooks (i.e. targeting in ALB), while most of principle component scores were negative when NHBF were more than 13 hooks (i.e. targeting in BET). For PC2, principle component scores concentrated on negative values when NHBF were less than 13 hooks (i.e. targeting in YFT), while principle component score tended to be positive when NHBF were more than 14 hooks (i.e. targeting in BET) and this indicated that NHBF for BET operations would be more than that for YFT operations. For PC3, most positive principle component scores occurred when NHBF were between 10 and 13 hooks (i.e. targeting in SWO), while negative values occurred for NHBF less than 10 hooks and more than 13 hooks represented ALB operations and YFT and BET operations.

3.3. CPUE standardization

Based on the model selections for the lognormal models incorporated *CA* (clusters) or *PCi* (principle component scores) as effects of targeting, all of main effects and interactions were statistically significant and remained in the models. The selected lognormal models were:

$$\log(CPUE) = \mu + Y + M + A + CA + Y \times A + M \times A + M \times CA + A \times CA$$

and

$$\begin{aligned}\log(CPUE) = \mu &+ Y + M + A + PC1 + PC2 + PC3 \\ &+ Y \times A + M \times A + M \times PC1 + M \times PC2 + M \times PC3 \\ &+ A \times PC1 + A \times PC2 + A \times PC3 + PC1 \times PC2 \\ &+ PC1 \times PC3 + PC2 \times PC3\end{aligned}$$

The ANOVA tables for selected lognormal models are shown in the Table 5. Except for the effect of year, the results indicate that the main effects of *CA* or *PCi* are the most explanatory effects for the models and the secondarily explanatory main effect is the effect of area. The distributions of residuals adequately fit to the assumption of normal distribution (Fig. 13).

Similarly, all of main effects and interactions were statistically significant and remained in the model for the delta model incorporated clusters (*CA*). For the delta model incorporated principle component scores (*PCi*), the interactions between *PC*1 and *PC*2 was excluded based on the AIC stepwise procedure. The selected delta models were:

$$PA = \mu + Y + M + A + CA + Y \times A + M \times A + M \times CA + A \times CA$$

and

$$\begin{aligned} PA = \mu + Y + M + A + PC1 + PC2 + PC3 + Y \times A + M \times A \\ + M \times PC1 + M \times PC2 + M \times PC3 + A \times PC1 + A \times PC2 \\ + A \times PC3 + PC1 \times PC3 + PC2 \times PC3 \end{aligned}$$

The ANOVA tables for selected delta models are shown in the Table 6. Except for the effect of year, the most explanatory main effect for the mode is the effect of area. For delta models, the explanatory powers of *CA* and PC_i are relatively lower than other main effects.

Table 7 shows the values of R^2, AIC and BIC statistics for lognormal and delta models incorporated *CA* or *PCi* as effects of targeting. The results indicate that including the effect of *PCi* can substantially improve the values of AIC and BIC for both lognormal and delta models although the differences in R^2 are not significant.

The area-specific nominal and standardized CPUE are shown in Fig. 14. Generally, trends of standardized CPUEs are close to nominal CPUEs. In addition, the trends of standardized CPUE series derived from the models with *CA* effects are quite similar to the standardized CPUE series derived from the models with PC_i effects. The trends of CPUE series in the northern areas (NW and NE) reveal similar trends and they substantially decreased since 1980 although the CPUE obviously fluctuated in early years. Also, the CPUE series in the southern areas (SW and SE) are similar but they revealed increasing trends with fluctuations before the mid-1990 and sharply decreased thereafter. In recent years, CPUEs seem to slightly increase for all four areas.

Fig. 15 shows the area-aggregated nominal and standardized CPUE series of striped marlin in the Indian Ocean. The trend of area-aggregated CPUE series is similar to the CPUE series in the northern areas and it reveals a continuously decreasing trend since 1980 although the obvious fluctuation occurred in early years.

REFERENCE

Bigelow, K.A., Hoyle, S.D., 2012. Standardized CPUE for South Pacific albacore. WCPFC-SC8-2012/SA-IP-14.

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.

He, X., Bigelow, K.A., Boggs, C.H., 1997. Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery. Fish. Res. 31: 147-158.

Hoyle1, S.D., Langley, A.D., Campbell, R.A., 2014. Recommended approaches for standardizing CPUE data from pelagic fisheries. WCPFC-SC10-2014/ SA-IP-10.

IOTC, 2014. Report of the Twelfth Session of the IOTC Working Party on Billfish. IOTC-2014-WPB-R[E].

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.

Ortega-García, S., Gómez-Muňoz, V., 1992. Standardization of fishing effort using principle component analysis of vessel characteristics: the mexican tuna purse-seiners. Sci. Mar. 56: 17-20.

Pech, N., Laloë, F., 1997. Use of principal component analysis with instrumental variables (PCAIV) to analyse fisheries catch data. ICES J. Mar. Sci. 54: 32-47.

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.

Punt, A. E., Walker, T.I., Taylor, B.L., Pribac, F., 2000. Standardization of catch and effort data in a spatially-structured shark fishery. Fish. Res. 45: 129-145.

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., Nishida, T., 2013. CPUE standardization of striped marlin (*Kajikia audax*) caught by Taiwanese longline fishery in the Indian Ocean. IOTC-2013-WPB11-26 Rev_2.

Winker, H., Kerwath, S.E., Attwood, C.G., 2013. Comparison of two approaches to standardize catch-per-unit-effort for targeting behaviour in a multispecies hand-line fishery. Fish. Res. 139: 118-131.

Winker, H., Kerwath, S.E., Attwood, C.G., 2014. Proof of concept for a novel procedure to standardize multispeciescatch and effort data. Fish. Res. 155: 149-159.

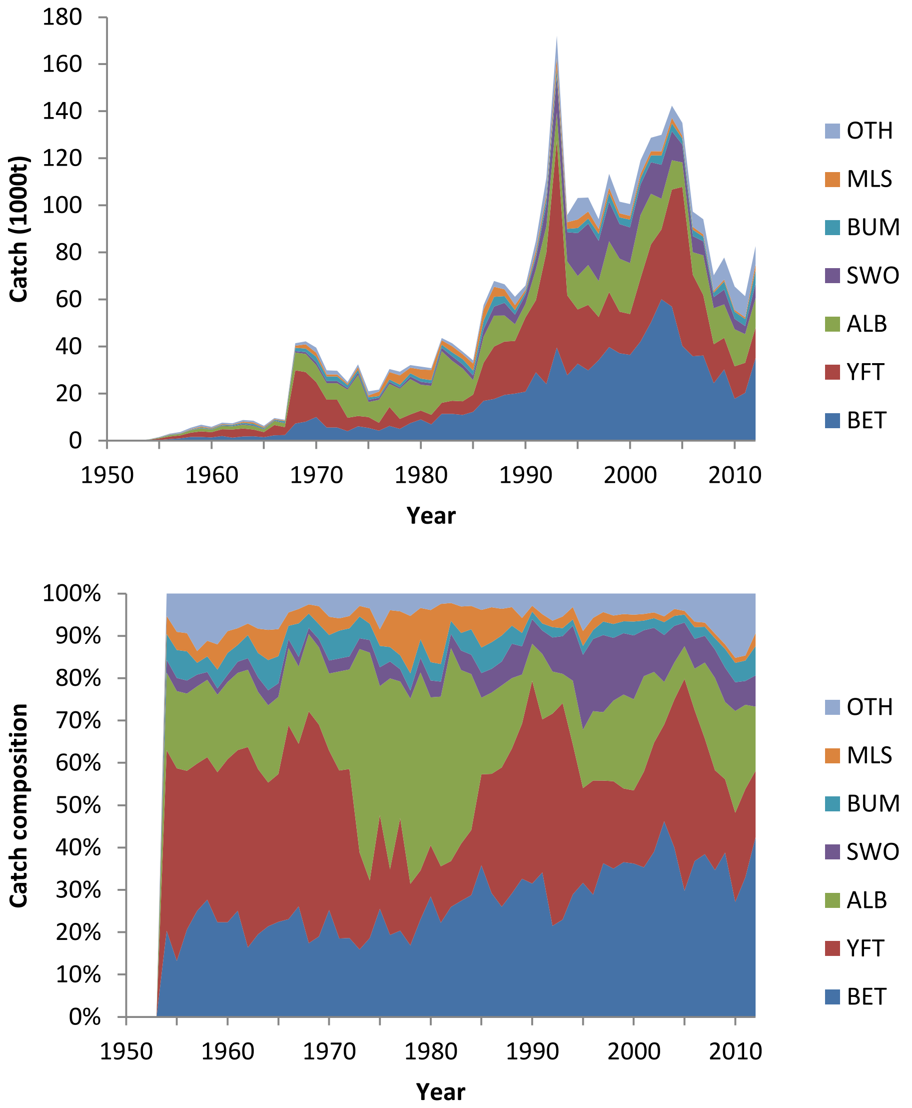


Fig. 1. Annual catches by species and catch compositions of Taiwanese longline fishery in the Indian Ocean.

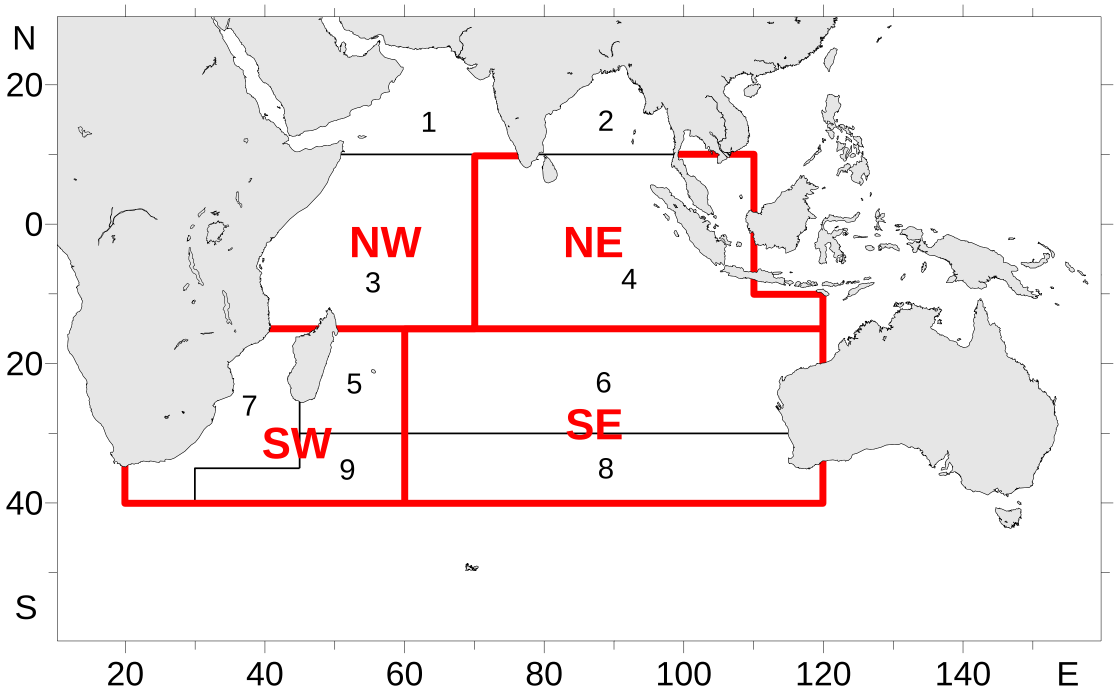


Fig. 2. Area stratification for swordfish in the Indian Ocean.

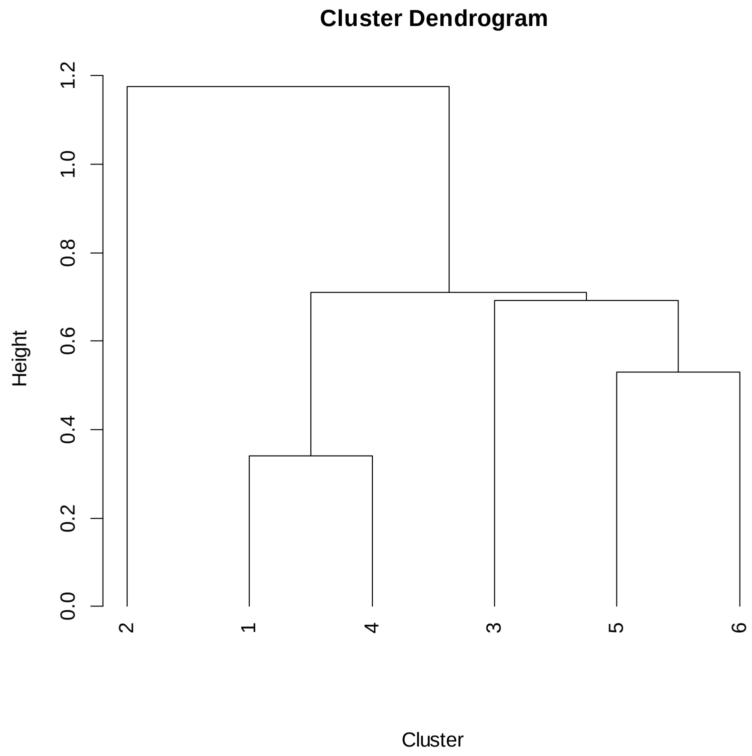


Fig. 3. The dendrogram of hierarchical cluster analysis for classifying the data sets of Taiwanese longline fishery in the Indian Ocean.

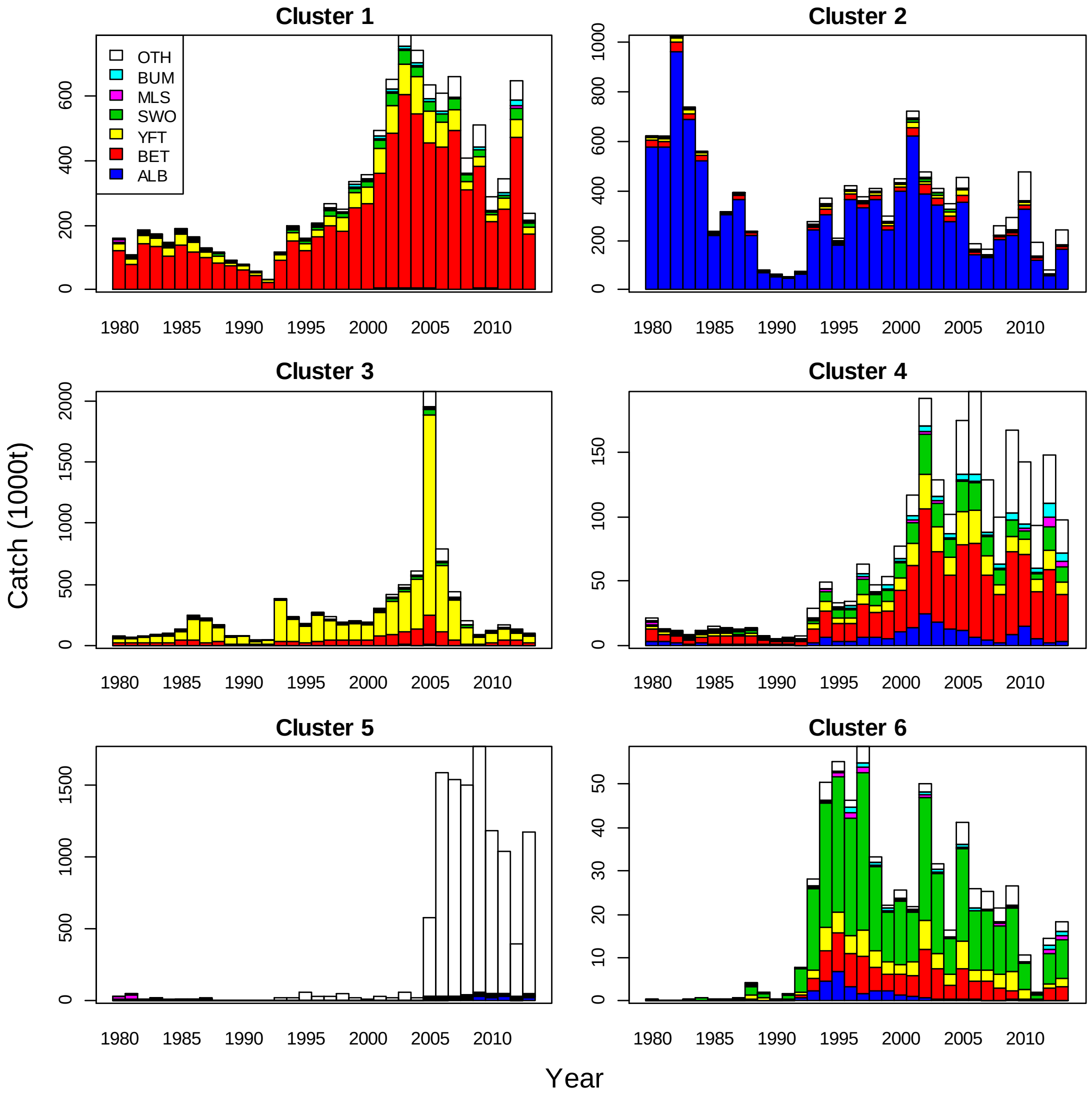


Fig. 4. Annual catches by species of Taiwanese longline fishery in the Indian Ocean for nine clusters.

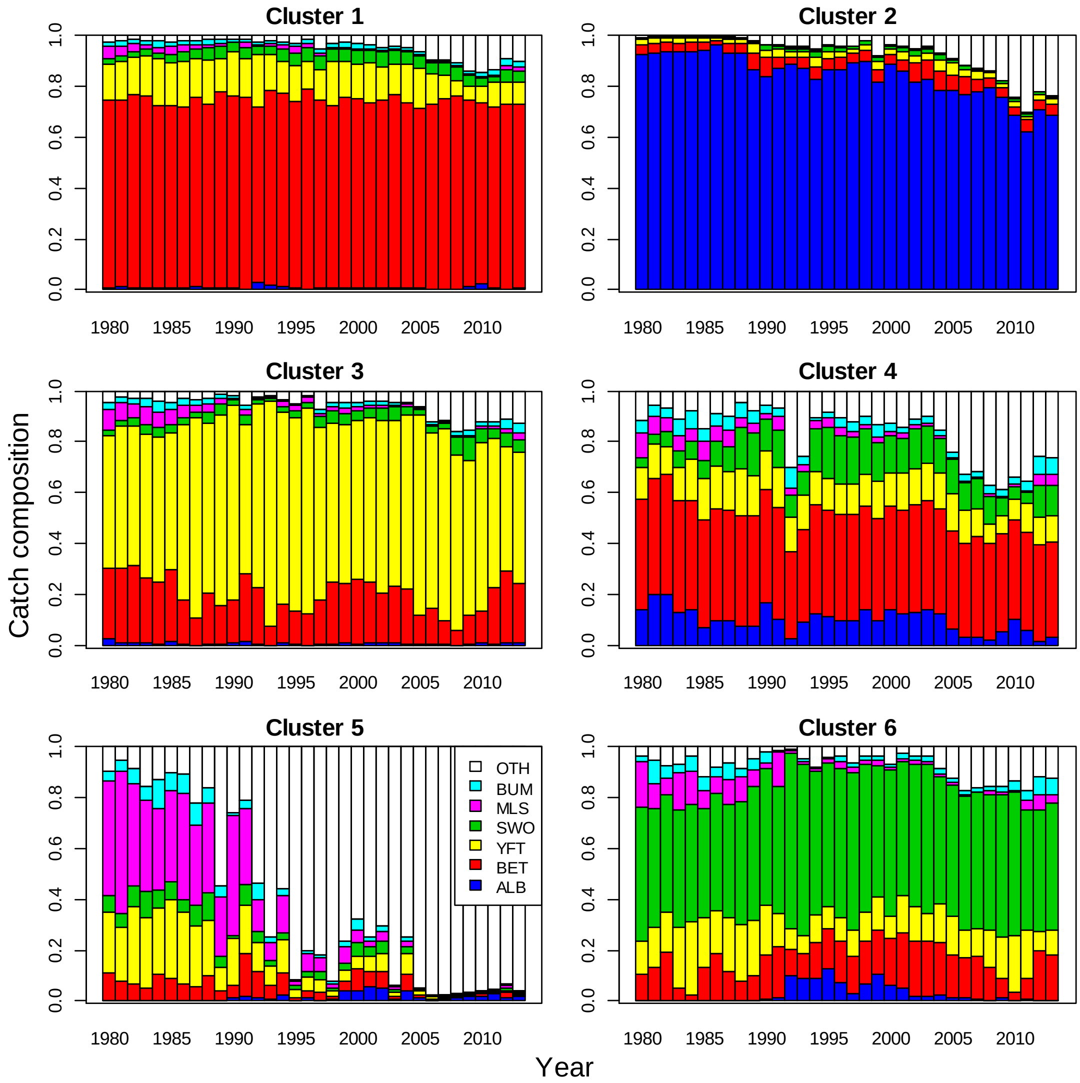


Fig. 5. Annual catch compositions of Taiwanese longline fishery in the Indian Ocean for nine clusters.

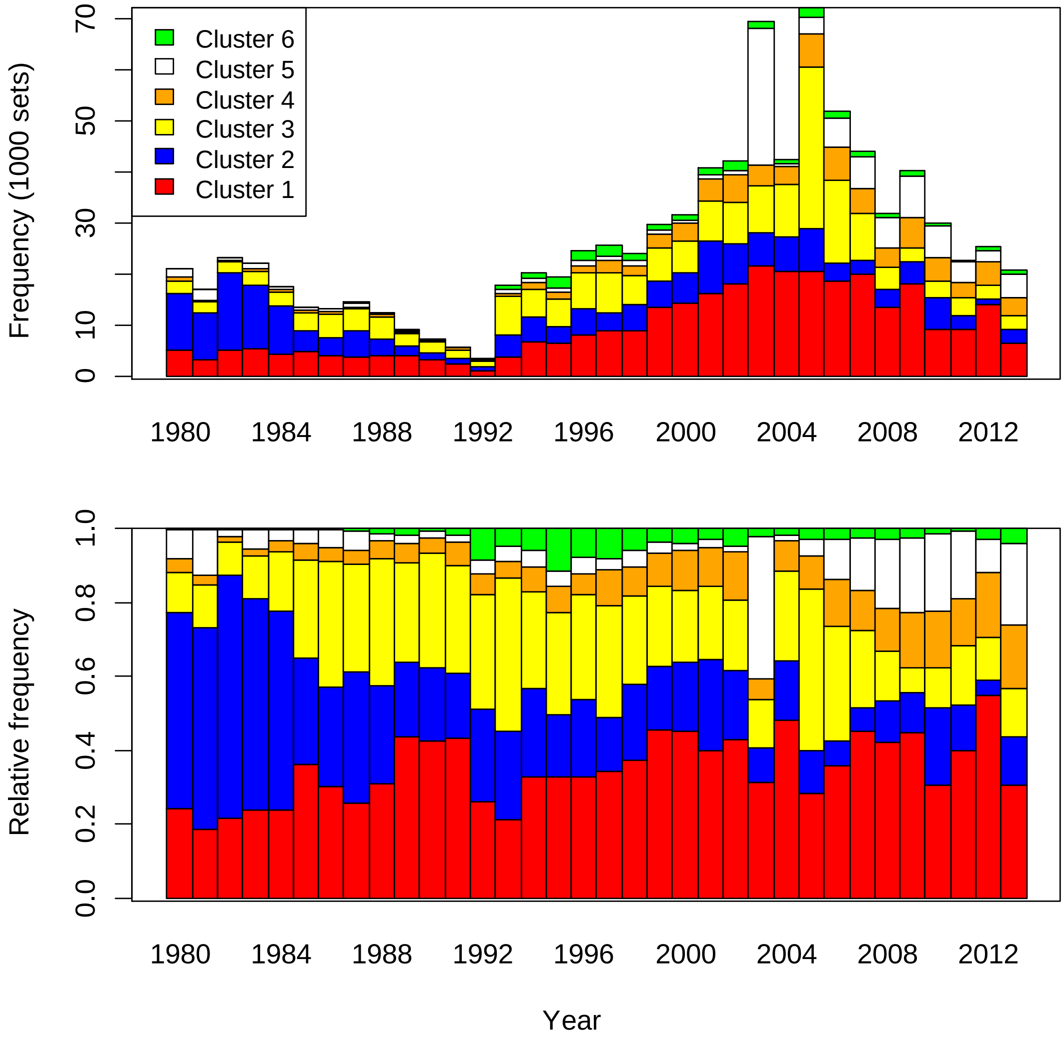


Fig. 6. Annual cluster compositions of Taiwanese longline fishery in the Indian Ocean.

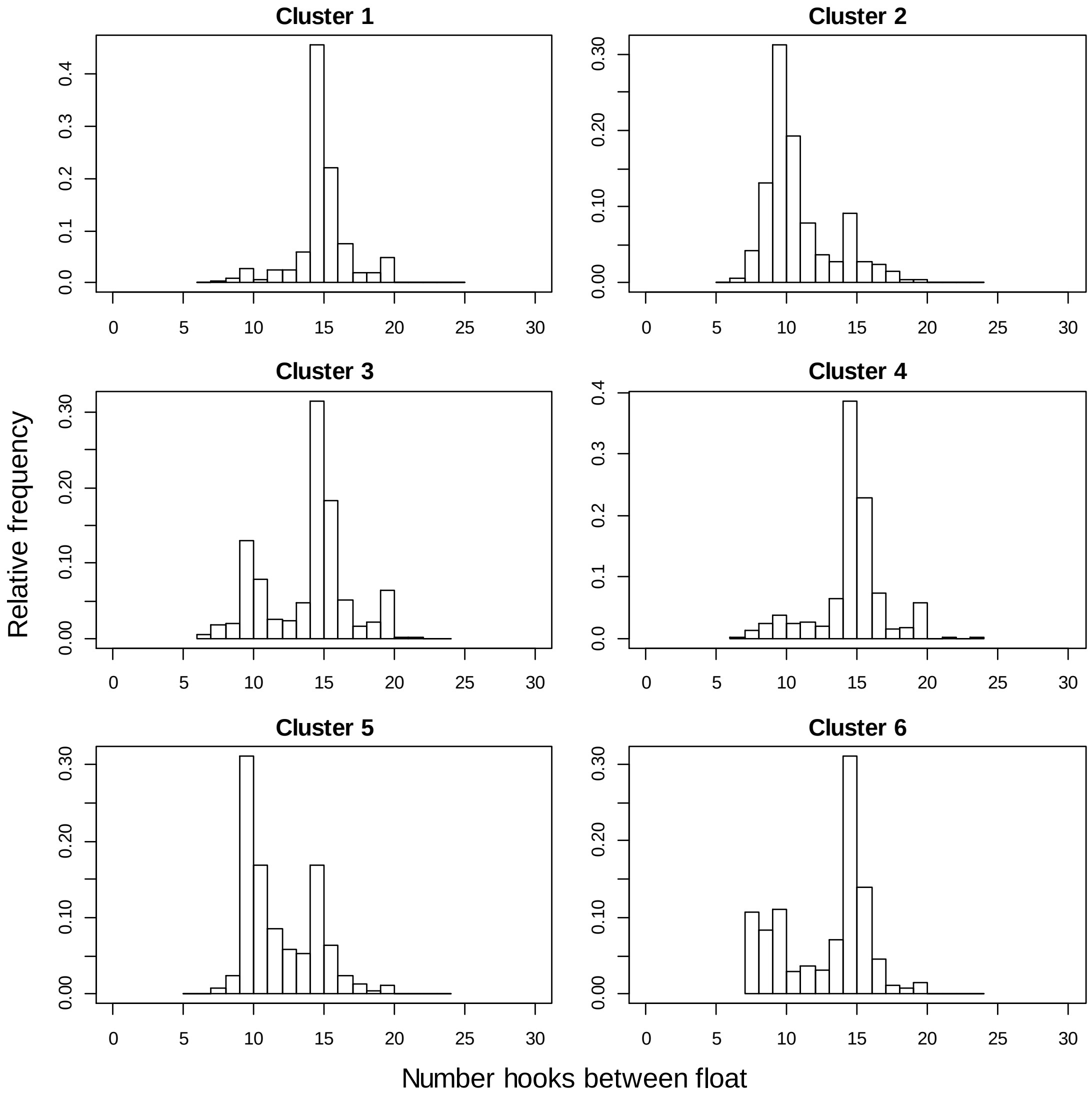


Fig. 7. The distributions of number hooks between float of Taiwanese longline fishery in the Indian Ocean for nine clusters.

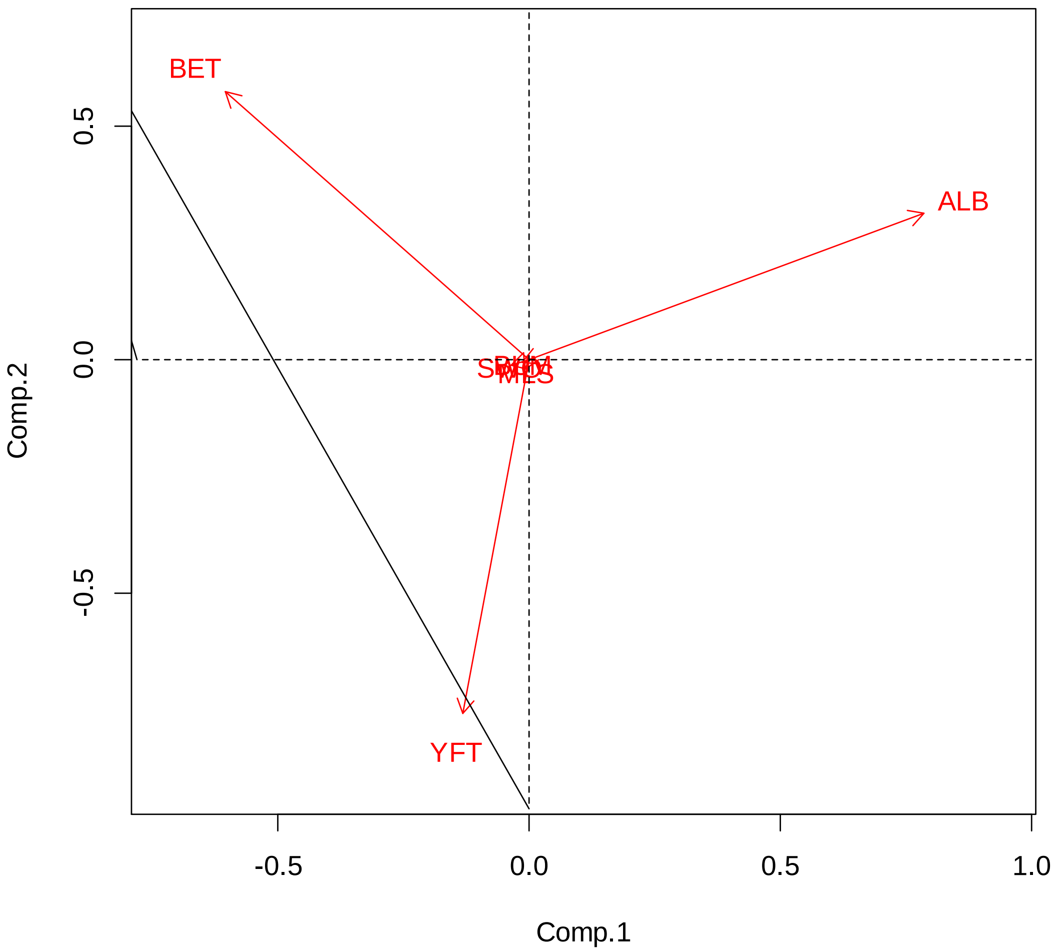


Fig. 8. Relationship between the first and the second principle component for Taiwanese longline fishery in the Indian Ocean.

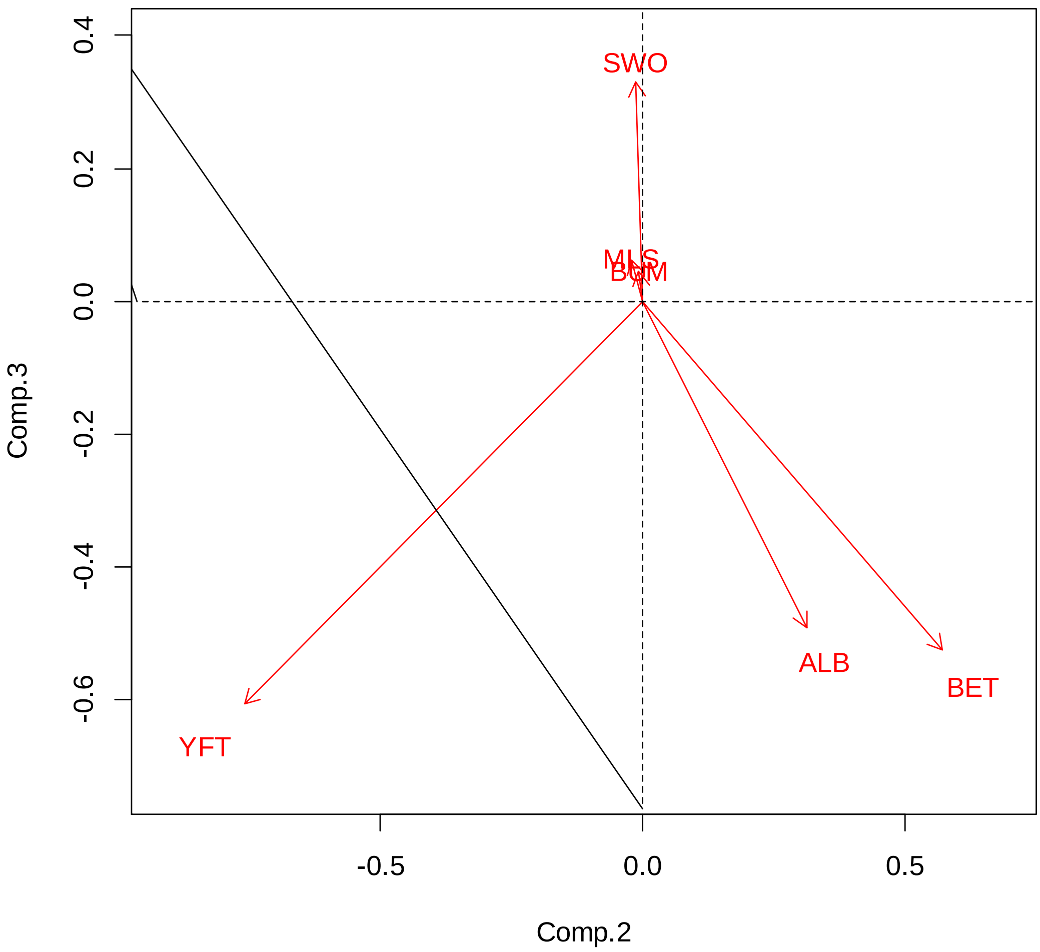


Fig. 9. Relationship between the second and the third principle component for Taiwanese longline fishery in the Indian Ocean.

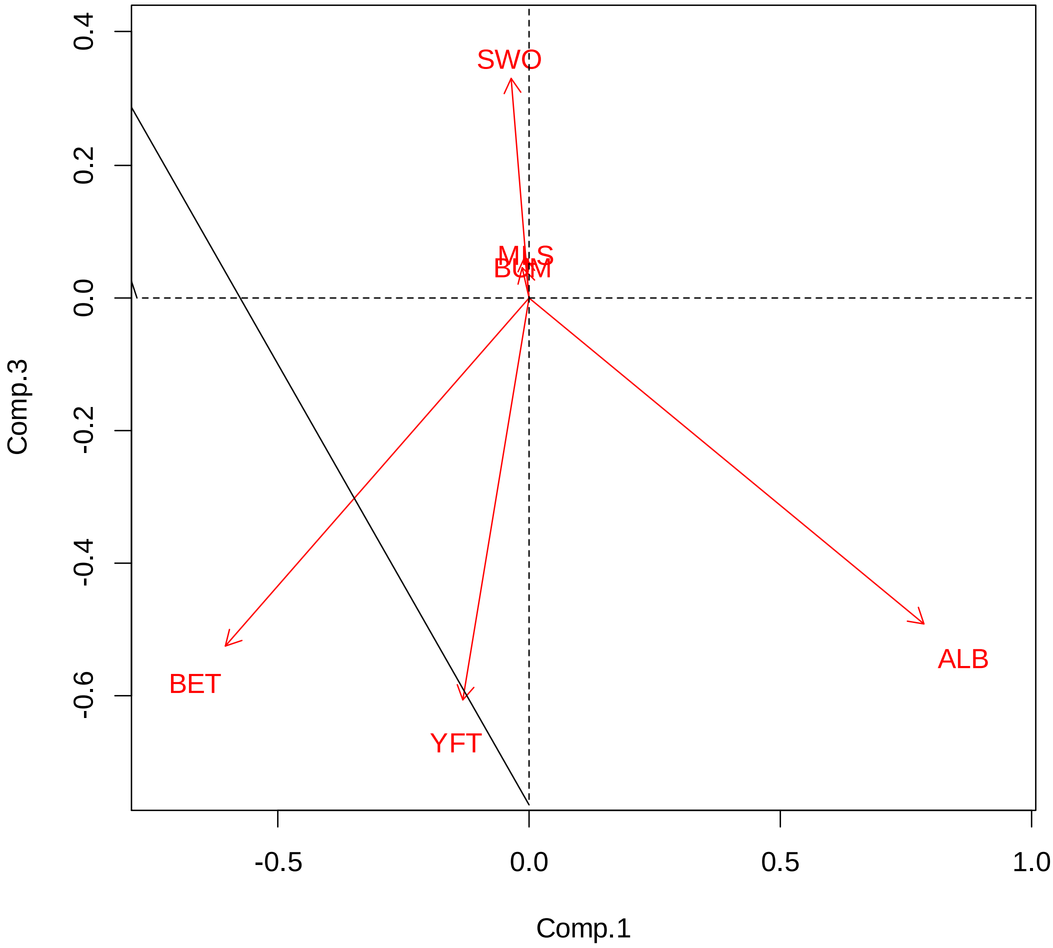


Fig. 10. Relationship between the first and the third principle component for Taiwanese longline fishery in the Indian Ocean.

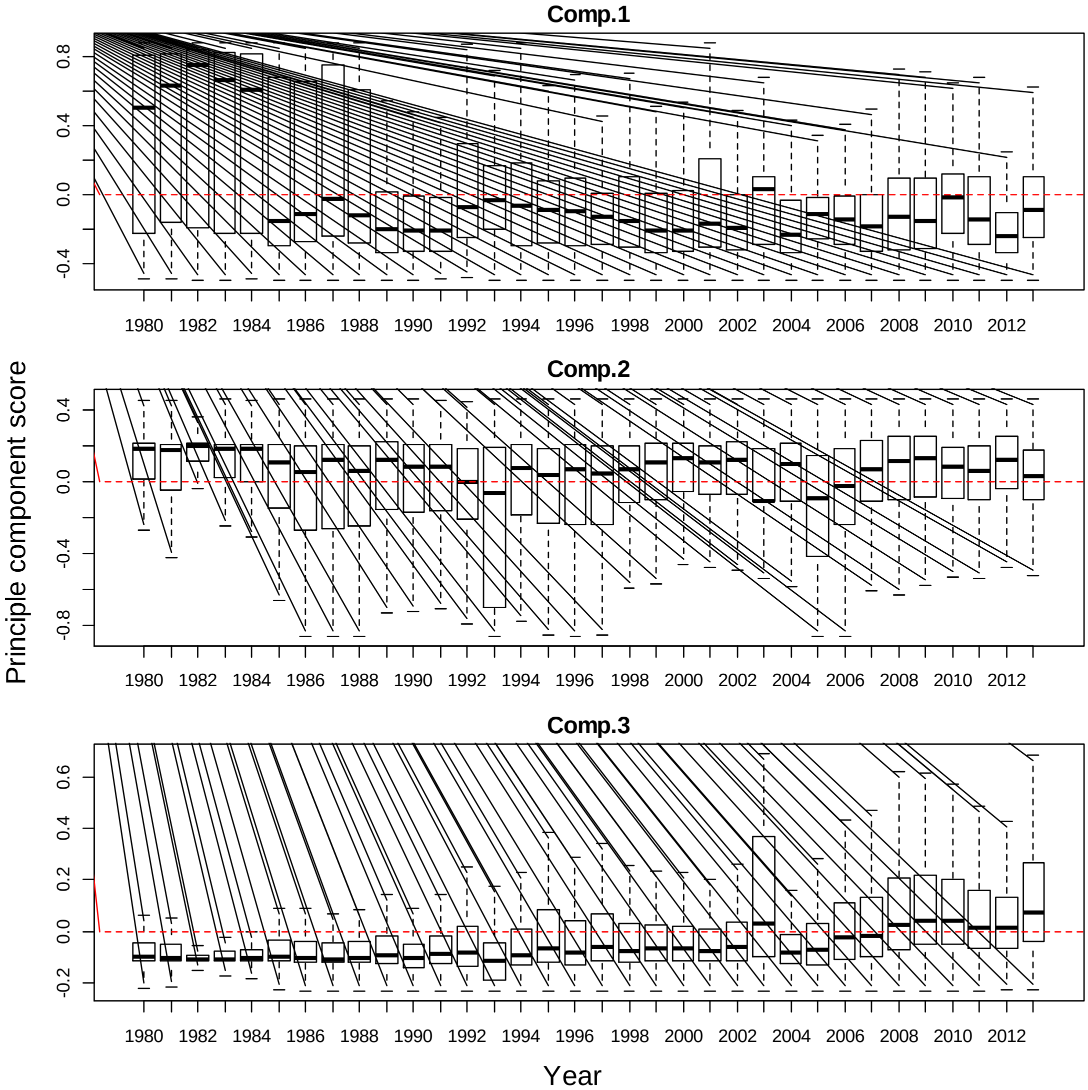


Fig. 11. Principle component scores by year based on the first, the second and the third principle component for Taiwanese longline fishery in the Indian Ocean.

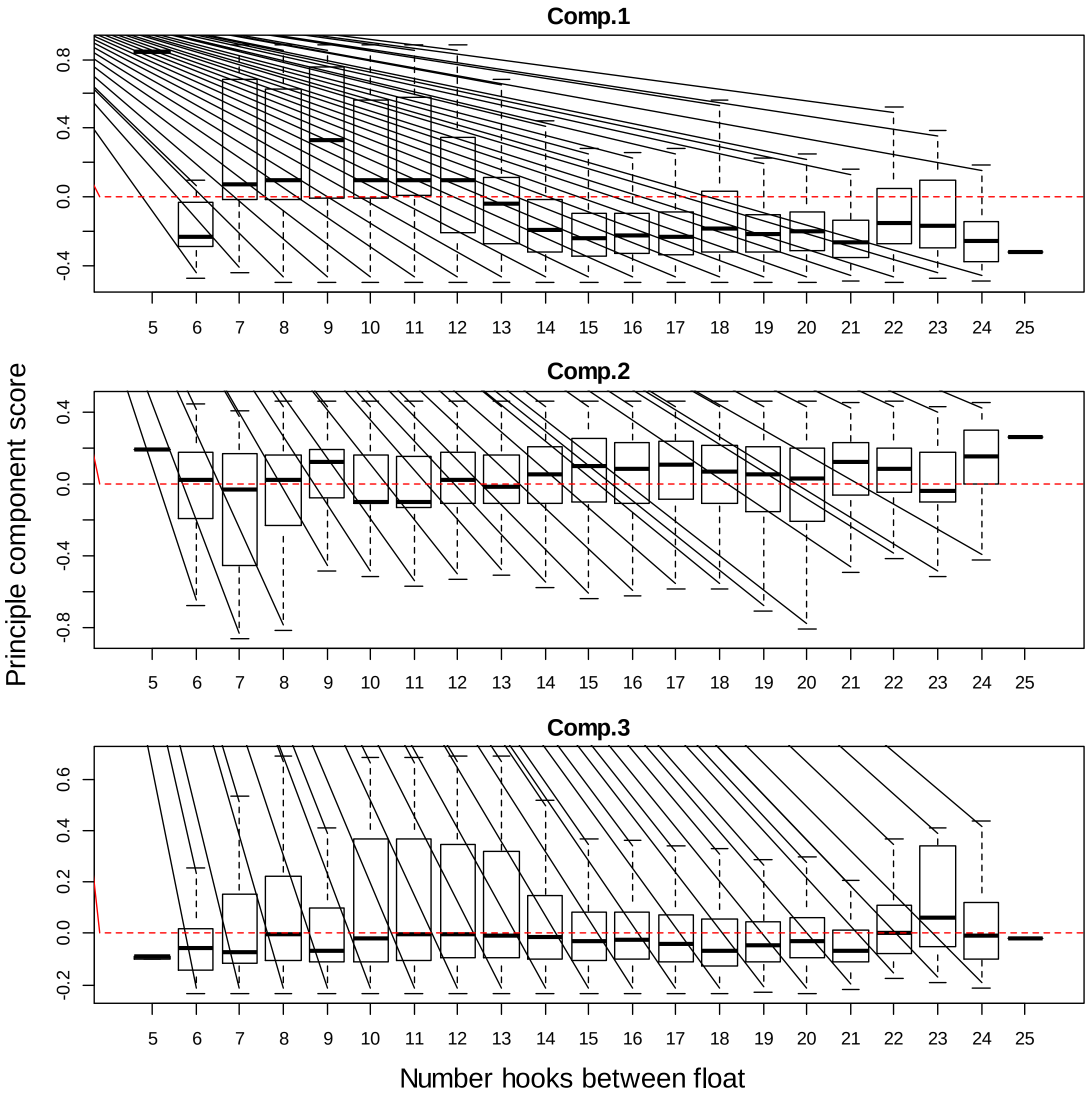
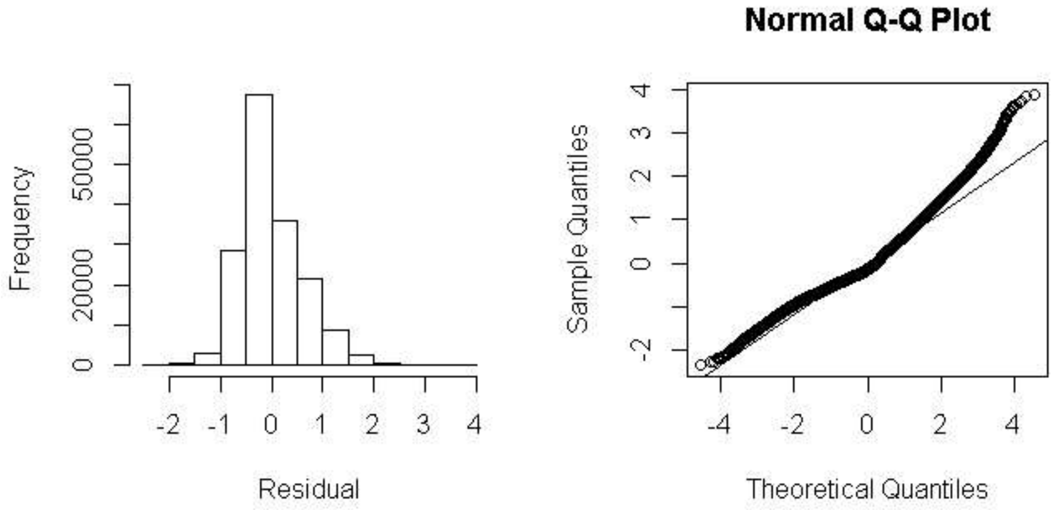


Fig. 12. Principle component scores by number hooks between float based on the first, the second and the third principle component for Taiwanese longline fishery in the Indian Ocean.

Model incorporated *CA* effect



Model incorporated PC_i effect

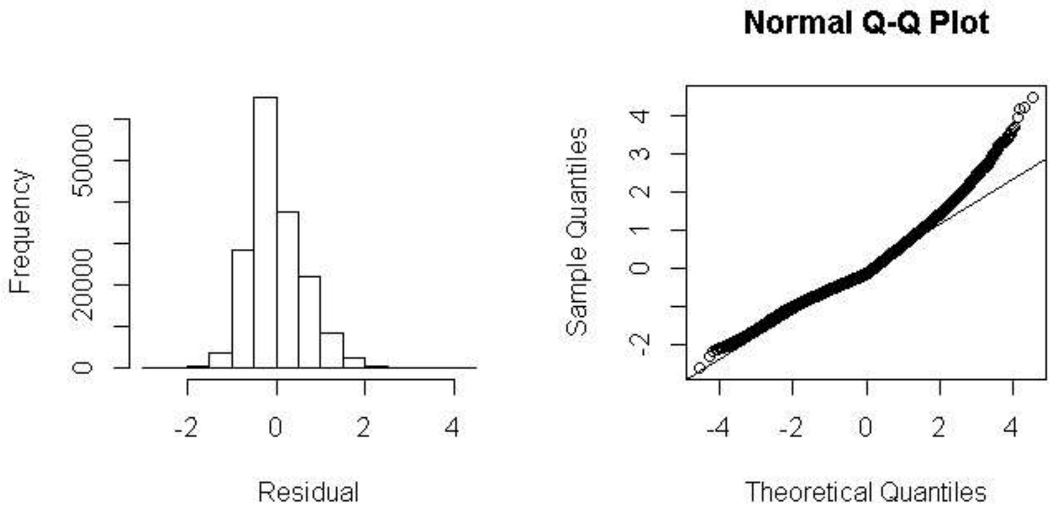


Fig. 13. The distributions and quantile-quantile plots of standardized residuals for lognormal and delta models incorporated clusters (*CA*) or principle component scores (PC_i) as effects of targeting.

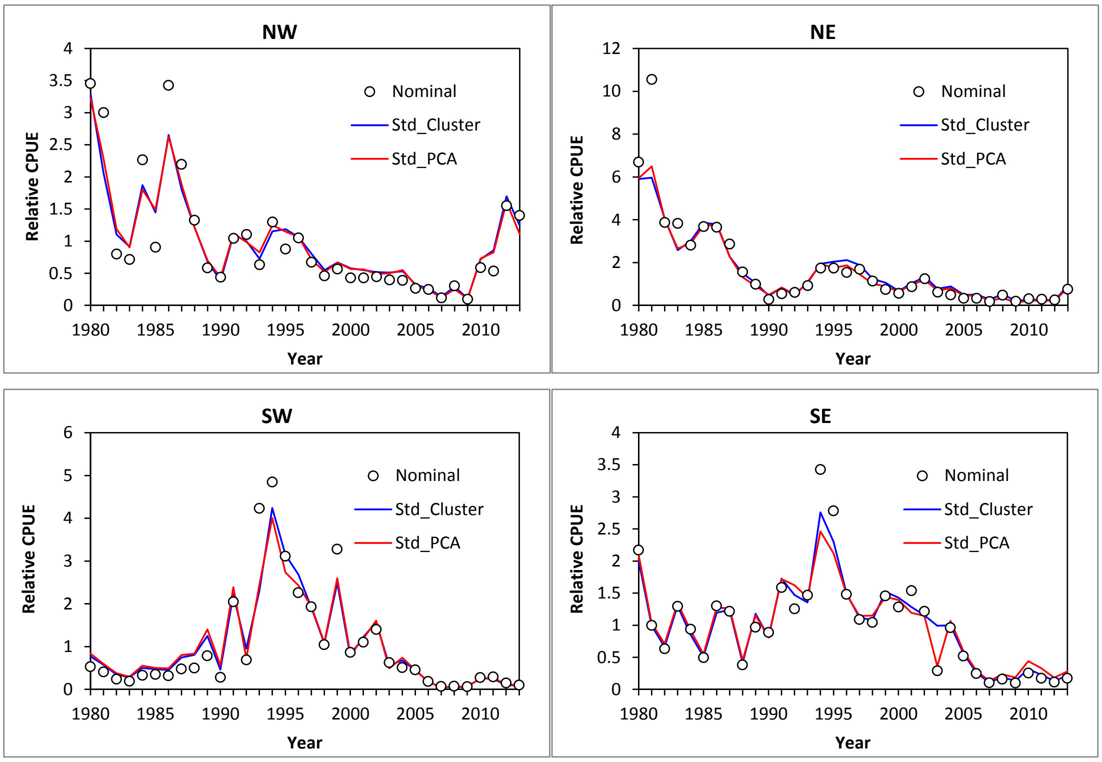


Fig. 14. Area-specific standardized CPUE of striped marlin of Taiwanese longline fishery in the Indian Ocean. CPUEs were scaled by the averaged value for each series.

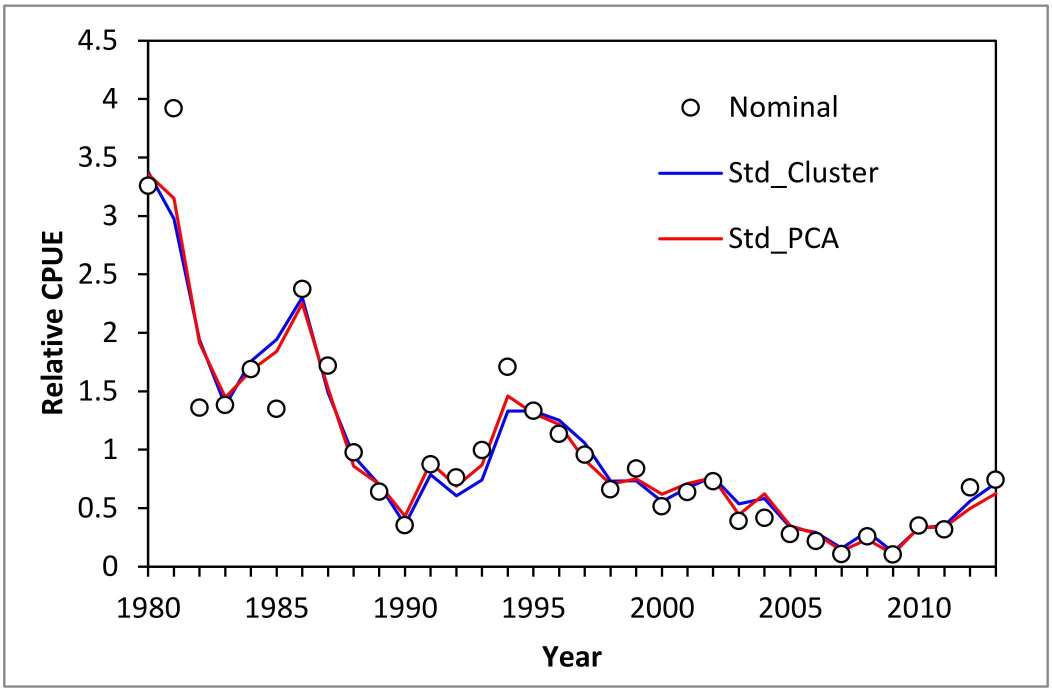


Fig. 15. Area-aggregated standardized CPUE of striped marlin of Taiwanese longline fishery in the Indian Ocean. CPUEs were scaled by the averaged value for each series.

Table 1. The average proportions of catches by species for Taiwanese longline fishery in the Indian Ocean based on 15 non-hierarchical (K-means) clusters.

Cluster	ALB	BET	YFT	SWO	MLS	BUM
1	0.011	0.627	0.057	0.092	0.016	0.026
2	0.938	0.019	0.009	0.004	0.001	0.001
3	0.006	0.611	0.242	0.031	0.012	0.014
4	0.001	0.017	0.894	0.013	0.010	0.006
5	0.447	0.074	0.129	0.058	0.008	0.007
6	0.285	0.419	0.090	0.057	0.007	0.011
7	0.776	0.057	0.045	0.015	0.004	0.003
8	0.012	0.020	0.019	0.014	0.004	0.013
9	0.004	0.834	0.050	0.023	0.007	0.009
10	0.614	0.223	0.029	0.023	0.003	0.003
11	0.009	0.371	0.120	0.142	0.024	0.048
12	0.036	0.149	0.102	0.530	0.018	0.022
13	0.008	0.395	0.410	0.045	0.016	0.019
14	0.002	0.094	0.199	0.069	0.429	0.058
15	0.010	0.132	0.578	0.065	0.030	0.027

Table 2. The average proportions of catches by species for Taiwanese longline fishery in the Indian Ocean based on 6 hierarchical clusters.

Cluster	ALB	BET	YFT	SWO	MLS	BUM
1	0.007	0.705	0.112	0.046	0.011	0.015
2	0.821	0.050	0.032	0.014	0.003	0.002
3	0.006	0.200	0.610	0.041	0.018	0.017
4	0.060	0.380	0.115	0.127	0.021	0.041
5	0.011	0.028	0.040	0.020	0.053	0.018
6	0.036	0.149	0.102	0.530	0.018	0.022

Table 3. Summary of principle component analysis based on the catch composition for Taiwanese longline fishery in the Indian Ocean.

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	0.396	0.304	0.168	0.102	0.058	0.043
Proportion of Variance	0.536	0.315	0.096	0.035	0.011	0.006
Cumulative Proportion	0.536	0.851	0.947	0.982	0.994	1.000

Table 4. Principle component loadings based on the catch composition for Taiwanese longline fishery in the Indian Ocean.

Species	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
ALB	0.784	0.313	-0.491	-0.204	-0.060	-0.029
BET	-0.605	0.572	-0.525	-0.166	-0.051	-0.014
YFT	-0.132	-0.758	-0.607	-0.197	-0.035	-0.019
SWO	-0.036	-0.012	0.330	-0.943	-0.029	0.004
MLS	-0.008	-0.021	0.062	0.054	-0.984	0.154
BUM	-0.013	-0.006	0.044	0.017	-0.150	-0.987

Table 5. The ANOVA tables for selected lognormal models incorporated clusters (*CA*) or principle component scores (PC_i) as effects of targeting.

Model incorporated *CA* effect

Variables	Type III SS	Df	F	Pr(>F)
Y	2933	33	236.245	< 2.2e-16 ***
M	111	11	26.706	< 2.2e-16 ***
A	95	3	83.931	< 2.2e-16 ***
CA	778	5	413.814	< 2.2e-16 ***
Y:A	1173	99	31.498	< 2.2e-16 ***
M:A	1246	33	100.342	< 2.2e-16 ***
M:CA	808	55	39.034	< 2.2e-16 ***
A:CA	628	15	111.261	< 2.2e-16 ***
Residuals	63365	168415		

Model incorporated PC_i effect

Variables	Type III SS	Df	F	Pr(>F)
Y	3261	33	260.737	< 2.2e-16 ***
M	231	11	55.383	< 2.2e-16 ***
A	143	3	125.722	< 2.2e-16 ***
PC1	41	1	107.24	< 2.2e-16 ***
PC2	95	1	251.595	< 2.2e-16 ***
PC3	89	1	234.094	< 2.2e-16 ***
Y:A	1268	99	33.786	< 2.2e-16 ***
M:A	965	33	77.149	< 2.2e-16 ***
M:PC1	50	11	12.098	< 2.2e-16 ***
M:PC2	224	11	53.667	< 2.2e-16 ***
M:PC3	129	11	30.937	< 2.2e-16 ***
A:PC1	159	3	139.576	< 2.2e-16 ***
A:PC2	153	3	134.841	< 2.2e-16 ***
A:PC3	462	3	406.059	< 2.2e-16 ***
PC1:PC2	71	1	188.387	< 2.2e-16 ***
PC1:PC3	177	1	466.061	< 2.2e-16 ***
PC2:PC3	333	1	878.036	< 2.2e-16 ***
Residuals	63836	168442		

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

Table 6. The ANOVA tables for selected delta models incorporated clusters (*CA*) or principle component scores (PC_i) as effects of targeting.

Model incorporated *CA* effect

Variables	LR Chisq	Df	Pr(>Chisq)
Y	28855.9	33	< 2.2e-16 ***
M	991.3	11	< 2.2e-16 ***
A	1678.7	3	< 2.2e-16 ***
CA	985.5	5	< 2.2e-16 ***
Y:A	18322.4	99	< 2.2e-16 ***
M:A	1845.7	33	< 2.2e-16 ***
M:CA	1437.4	55	< 2.2e-16 ***
A:CA	3649.1	15	< 2.2e-16 ***

Model incorporated PC_i effect

Variables	LR Chisq	Df	Pr(>Chisq)
Y	30091.6	33	< 2.2e-16 ***
M	1329.8	11	< 2.2e-16 ***
A	1529	3	< 2.2e-16 ***
PC1	320.7	1	< 2.2e-16 ***
PC2	1489.7	1	< 2.2e-16 ***
PC3	415	1	< 2.2e-16 ***
Y:A	18949.4	99	< 2.2e-16 ***
M:A	1473.6	33	< 2.2e-16 ***
M:PC1	320.6	11	< 2.2e-16 ***
M:PC2	863.3	11	< 2.2e-16 ***
M:PC3	391.8	11	< 2.2e-16 ***
A:PC1	229.1	3	< 2.2e-16 ***
A:PC2	438.5	3	< 2.2e-16 ***
A:PC3	1347.4	3	< 2.2e-16 ***
PC1:PC3	2136	1	< 2.2e-16 ***
PC2:PC3	1617.5	1	< 2.2e-16 ***

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

Table 7. The values of R^2, AIC and BIC statistics for lognormal and delta models incorporated clusters (*CA*) or principle component scores (PC_i) as effects of targeting.

	Log-normal model		delta-model	
Statistics	Cluster	PCA	Cluster	PCA
R^2	0.351	0.346	0.176	0.188
AIC	314,045	315,240	717,327	707,137
BIC	316,614	317,538	720,314	709,796