CPUE standardization of blue marlin (*Makaira nigricans*) caught by Taiwanese longline fishery in the Indian Ocean using targeting effect derived from principle component analyses

Sheng-Ping Wang

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

ABSTRACT

In this study, the principle component analysis was conducted based on catch composition of Taiwanese longline fishery in the Indian Ocean. The results indicated that the principle component scores can represent the historical fishing pattern related to characteristics of targeting species. Also, there were appropriate relationships between the principle component scores and the numbers of hooks between float. The delta-lognormal GLM was used to conduct the CPUE standardization of blue marlin (*Makaira nigricans*) caught by the Taiwanese longline fishery in the Indian Ocean for 1979-2015 because blue marlin was bycatch species of Taiwanese longline fishery and large amounts of zero catches was recorded in the operational data sets. CPUE trends were obviously different by areas, while area-aggregated CPUE obviously declined since the mid 1980; gradually increased in the 1990s; declined again from the late 1990s to the early 2000s; substantially increased in recent years.

1. INTRODUCTION

Blue marlin is considered to be a non-target species of industrial and artisanal fisheries. Longline catches account for around 69% of total catches in the Indian Ocean, followed by gillnets (28%), with remaining catches recorded under troll and handlines. The catches were mainly made by Taiwan (longline, 33%), Indonesia (fresh longline, 28%), Pakistan (gillnet, 14%), Iran (gillnet, 7%), and Sri Lanka (7%). Catches reported by drifting longliners were more or less stable until the late 1970s, at around 3,000 t to 4,000 t, and have steadily increased since then to reach values between 8,000 t and to over 10,000 t since the early 1990's. The highest catches reported by longliners have been recorded since 2012, and are likely to be the consequence of higher catch rates by some longline fleets which appear to have resumed operations in the western tropical Indian Ocean (IOTC, 2015).

Fig. 1 shows the historical catches by species and catch of blue marlin caught by Taiwanese fishery, and the annual proportion of blue marlin was generally less 6% of total catches. Fig. 2 shows the nominal CPUE distribution of blue marlin of Taiwanese fleet and high CPUE generally occurred in the northern waters of 15S.

Because blue marlin was bycatch species of Taiwanese lognline fishery, large amount of zero-catches was recorded in the operational catch and effort data sets of Taiwanese longline fishery. The annual proportions of zero-catch were about 75% of total data sets, while the proportions of zero catch decreased in recent years. 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) was applied to conduct the CPUE standardization of blue marlin in the Indian Ocean.

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). These approaches have been applied to conduct the CPUE standardization for striped marlin in the Indian Ocean (Wang, 2015). IOTC (2015) noted that the use of clustering and PCA was a useful approach in dealing with the absence of HBF, and such techniques help examine sets that are used for targeting certain species groups and use all the data in the database of Taiwan. In addition, IOTC (2015) also agreed that the PCA approach should be used instead of the Clustering approach as this gave better results on AIC and BIC values, when modelling the positive sets.

In this paper, we attempted to classify the data sets in relation to species composition of the catches. The results of PCA were also incorporated into CPUE standardization 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 1980-2015 were provided by Oversea Fisheries Development Council of Taiwan (OFDC). It should be noted that the data in 2015 is preliminary.

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. 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).

2.3. CPUE Standardization

A delta-lognormal GLM was applied to standardize the CPUE. The main effects considered in this analysis are year, month, area, and effects related to the fishing configurations (principal component scores). Fishing areas used in this study were defined by four areas based on the IOTC statistics areas for swordfish in the Indian Ocean (Fig. 3) (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 blue 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 (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:

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

2.4. 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. Principle Component Analysis

Based on the results of PCA, the first principle component (PC1) explained about 55% variance of observations, and cumulative proportions of explained variance of the 2nd (PC2) and 3rd (PC3) principle components reached to 87% and 95% (Table 1). 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. 4-6, Table 2).

Fig. 7 shows the annual distributions of principle component scores. 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. 8 shows the distributions of principle component scores grouped by NHBF. 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, more negative principle component scores occurred when NHBF were less than 11 hooks (i.e. targeting in YFT), while principle component score tended to be positive when NHBF were more than 12 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 7 and 13 hooks (i.e. targeting in SWO), while concentrated values for NHBF more than 13 hooks represented ALB operations and YFT and BET operations.

Figs. 9-11 show the spatial distributions of the principle component scores by decades. The patterns indicate the BET fishing operations conducted in the northern Indian Ocean and ALB fishing operations were in the southern Indian Ocean; YFT fishing operations mainly concentrated in Arabian Sea and Bay of Bengal; SWO fishing operations occurred in the southwestern Indian Ocean since 1990s.

3.2. CPUE standardization

Based on the model selections for the lognormal models incorporated *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 model was:

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

The ANOVA tables for selected lognormal models are shown in the Table 3. Except for the effect of year, the results indicate that the main effects of *PC3* was the most explanatory main effect for the models and the secondarily explanatory main effect is the effect of month and *PC2*. In addition, interactions related to *PC3* and *PCi* also provided significant contributions to explanation of variance. The distribution of residuals was close to the assumption of normal distribution (Fig. 12).

Similarly, all of main effects and interactions were statistically significant and remained in the model for the delta model incorporated principle component scores

(PCi). The selected delta model was:

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

The ANOVA tables for selected delta models are shown in the Table 4. Except for the effect of year, the most explanatory main effect for the mode were the effects of month and area. For delta models, the explanatory abilities of PCi are relatively lower than other main effects.

The area-specific standardized CPUE are shown in Fig. 13. The trends of CPUE series in the northern areas (NW and NE) reveal similar trends and CPUEs obviously declined since the mid 1980 although CPUEs obviously fluctuated in early years; gradually increased in the 1990s; declined again from the late 1990s to the early 2000s; increased after the early 2000s but the substantial increase of CPUE can be observed in NW. The CPUE series in the southern areas (SW and SE) generally fluctuated without apparent trends, but high CPUEs occurred between the late 1990s and early 2000s. Fig. 14 shows the area-aggregated standardized CPUE series of blue marlin in the Indian Ocean. The trend of area-aggregated CPUE series is similar to the CPUE series in the northern areas, NW especially.

REFERENCE

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, 2015. Report of the 13th Session of the IOTC Working Party on Billfish. IOTC-2015-WPB13-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., 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. IOTC–2015–WPB13–31 Rev_1.

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.

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 multispecies catch and effort data. Fish. Res. 155: 149-159.

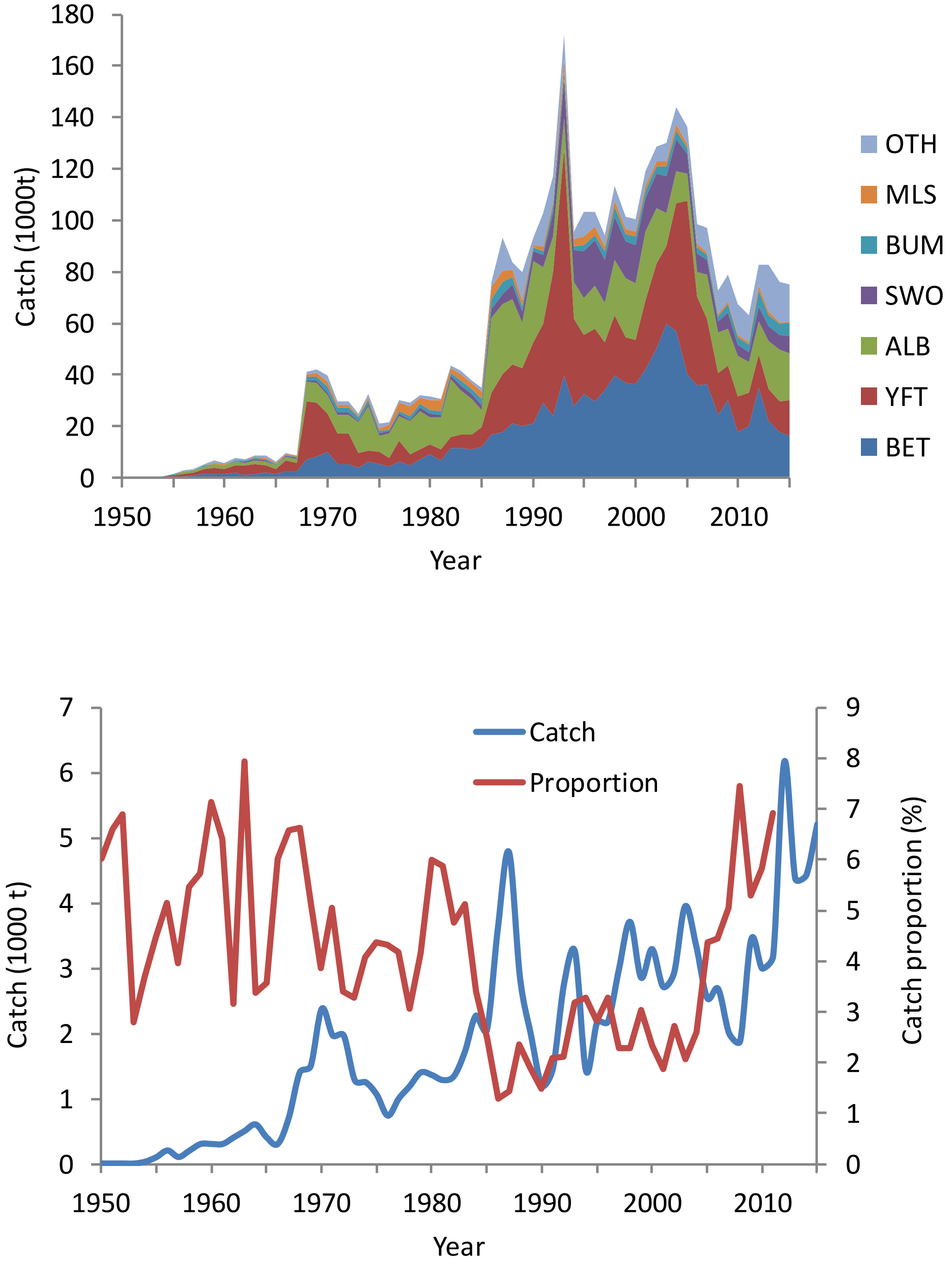

Fig. 1. Annual catches by species (upper panel) and catch of blue marlin (lower panel) caught by Taiwanese longline fishery in the Indian Ocean.

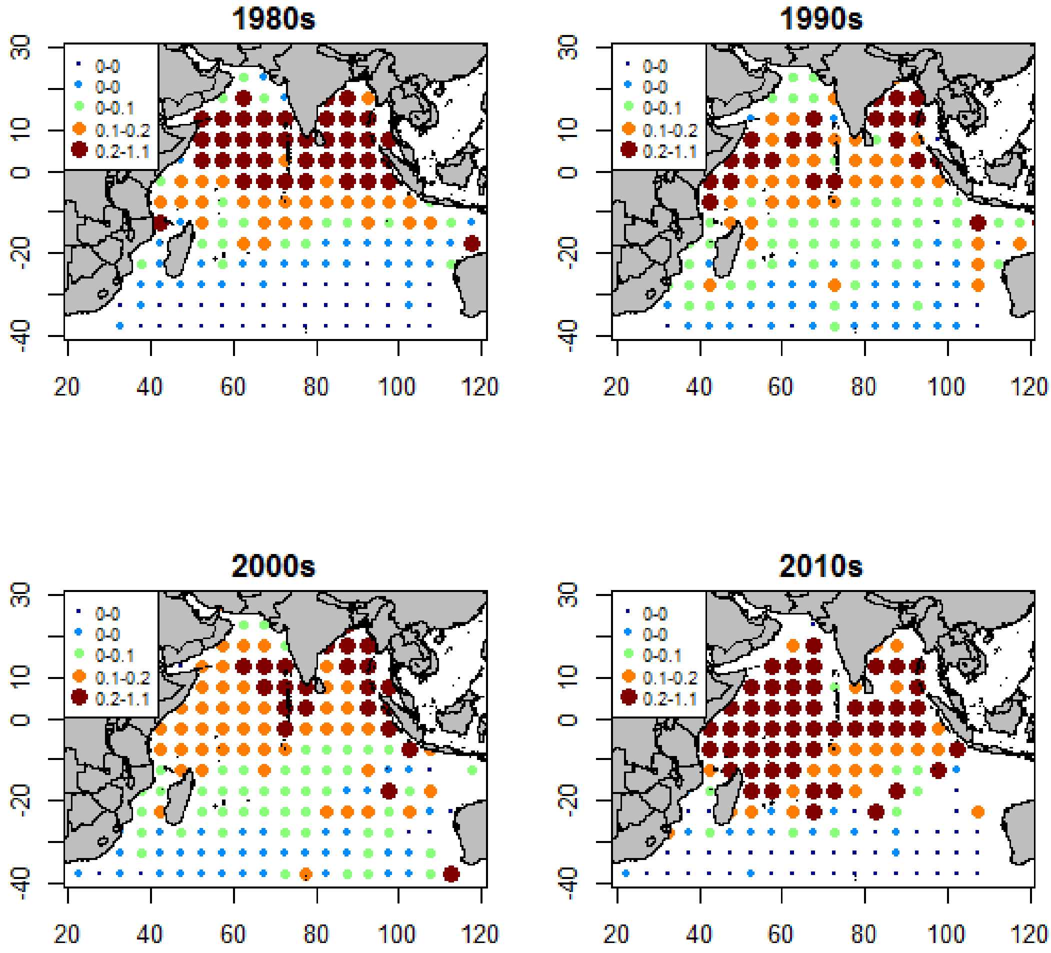

Fig. 2. Nominal CPUE distributions for blue marlin caught by Taiwanese longline fishery in the Indian Ocean.

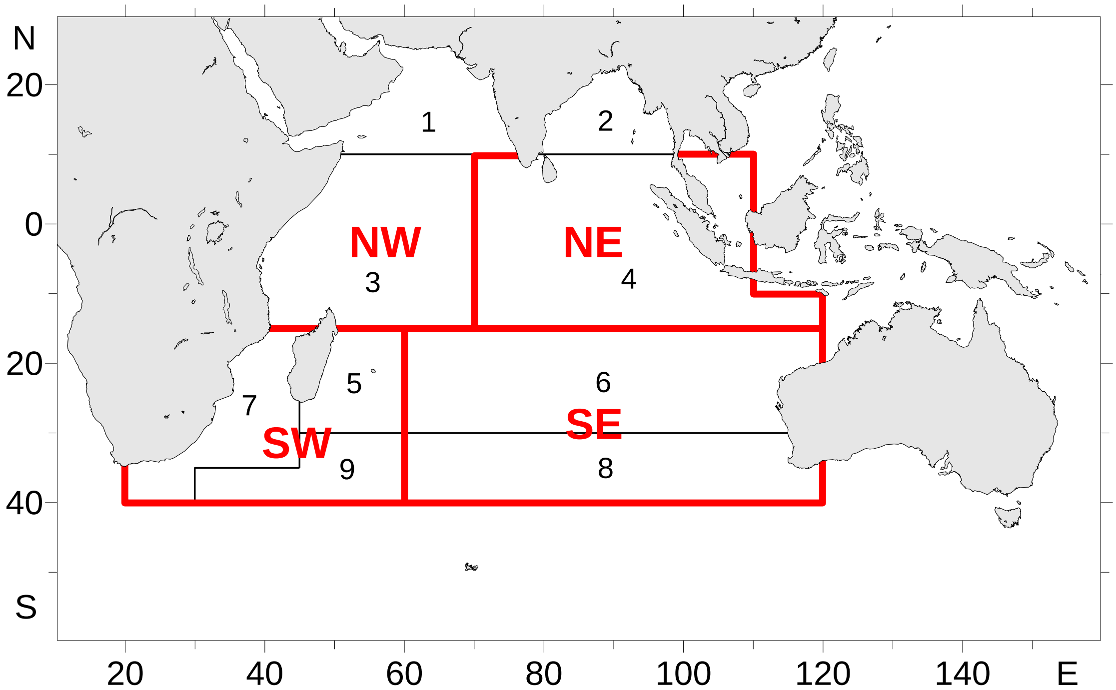

Fig. 3. Area stratification for swordfish in the Indian Ocean.

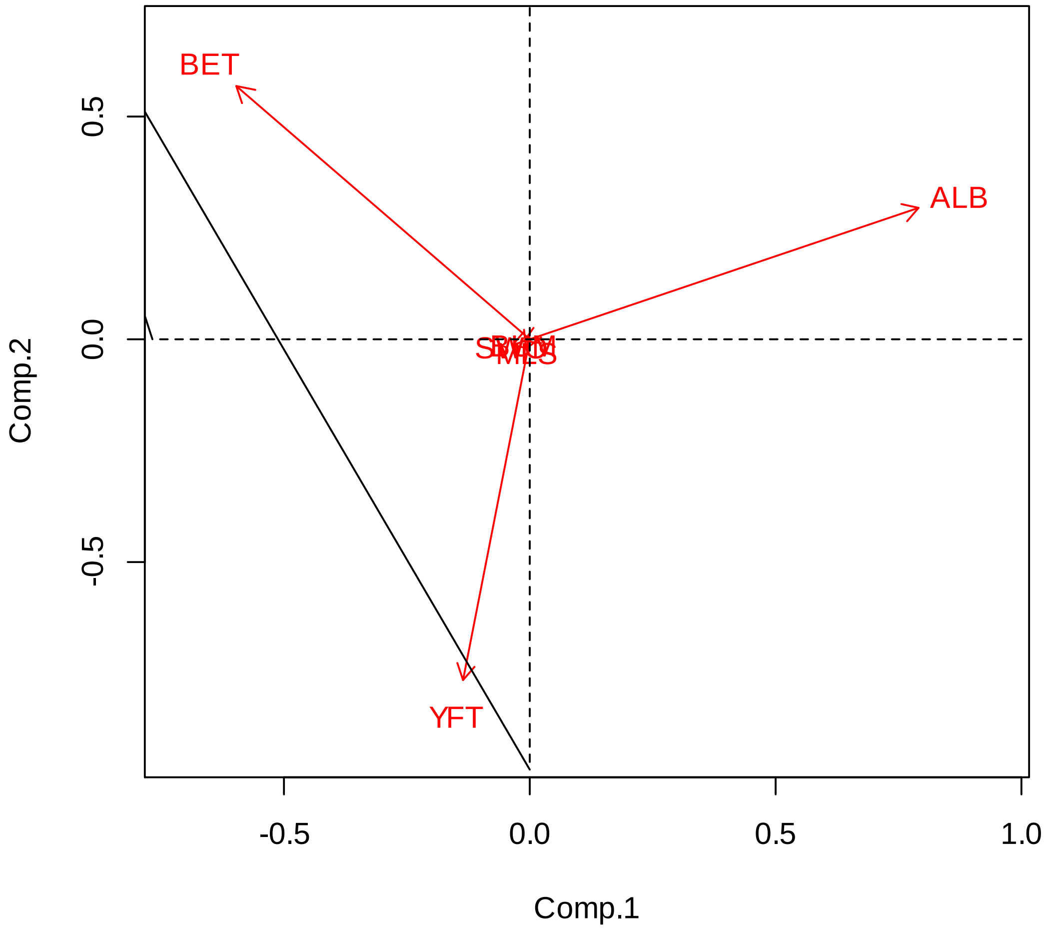

Fig. 4. Relationship between the first and the second principle component for Taiwanese longline fishery in the Indian Ocean.

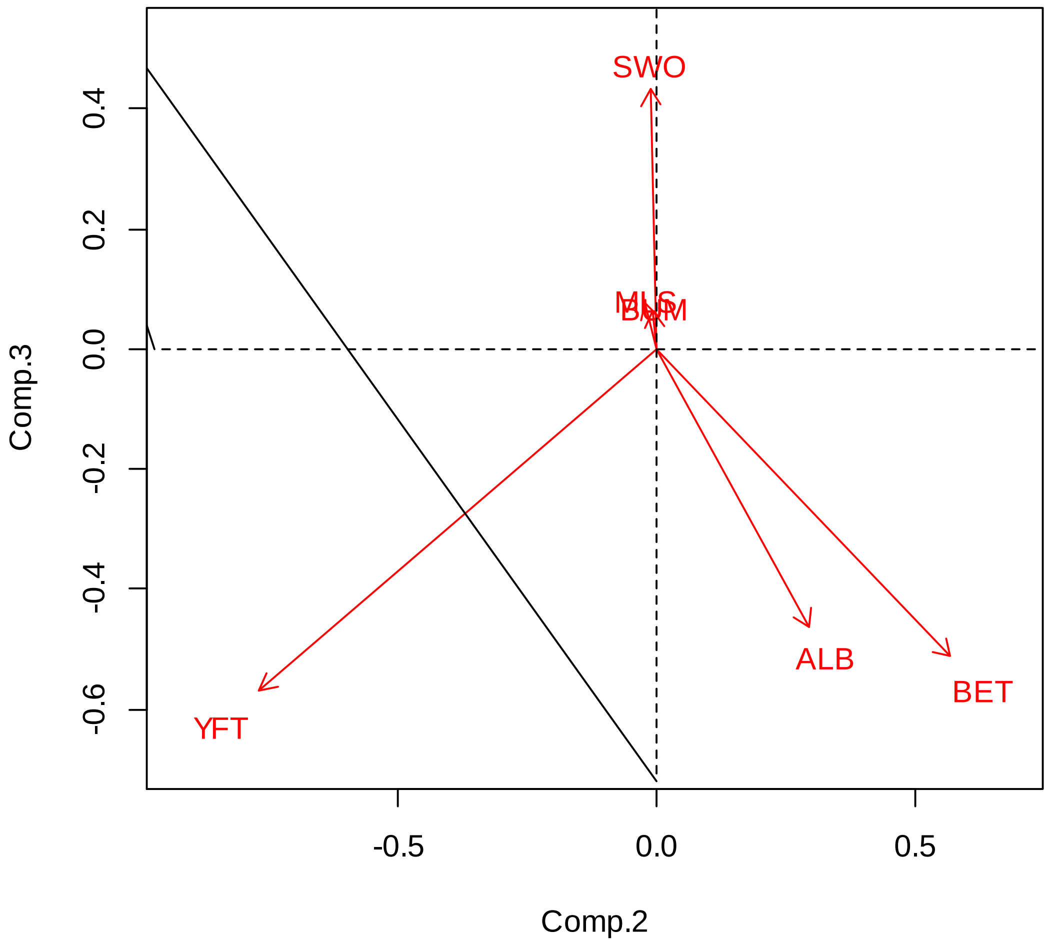

Fig. 5. Relationship between the second and the third principle component for Taiwanese longline fishery in the Indian Ocean.

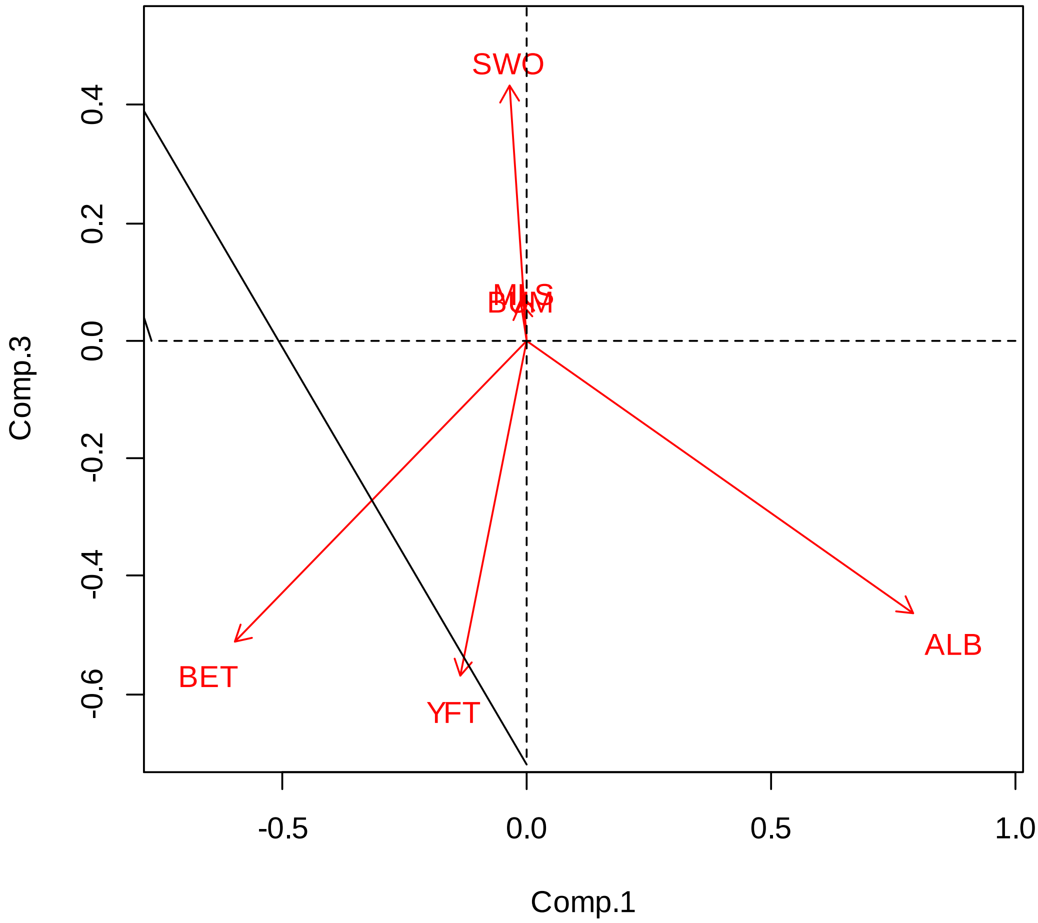

Fig. 6. Relationship between the first and the third principle component for Taiwanese longline fishery in the Indian Ocean.

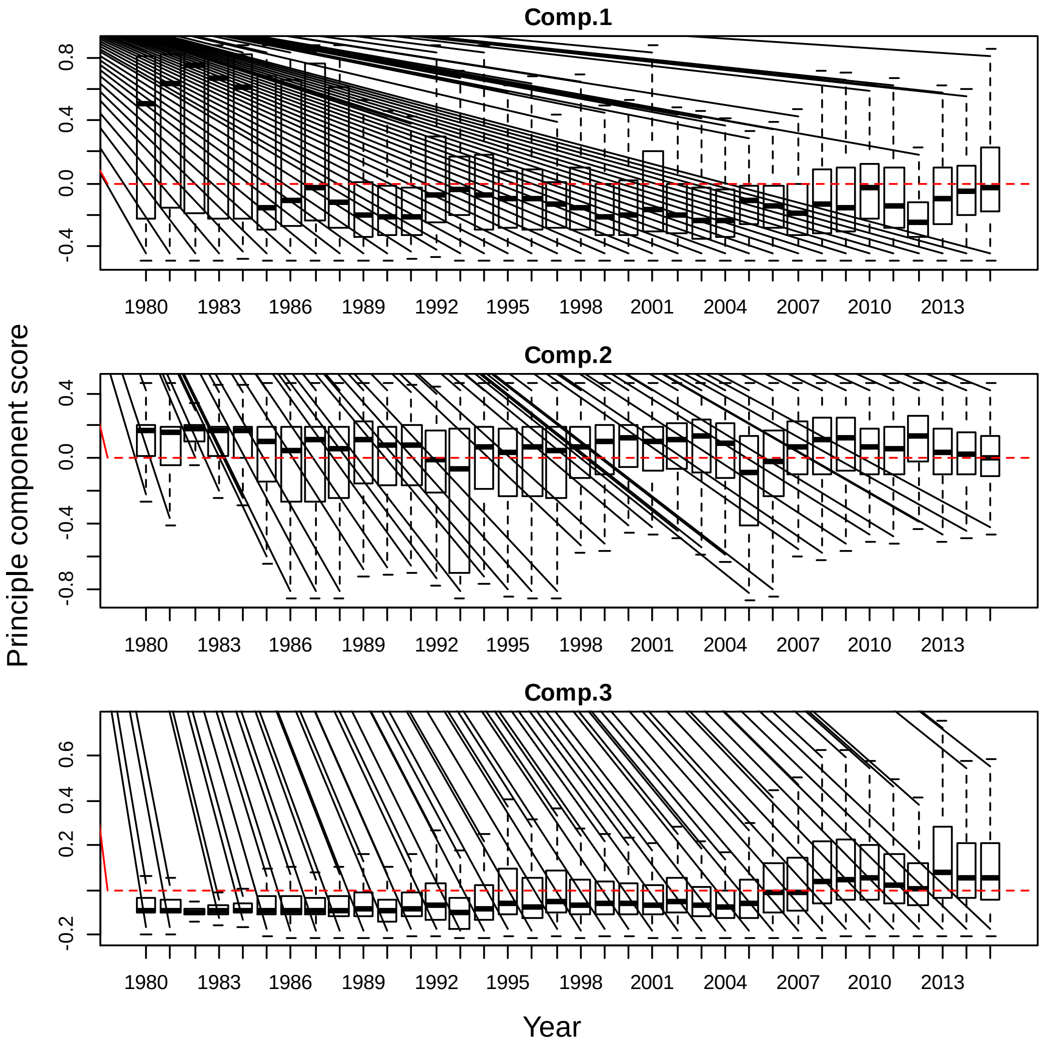

Fig. 7. Principle component scores by year based on the first, the second and the third principle component for Taiwanese longline fishery in the Indian Ocean.

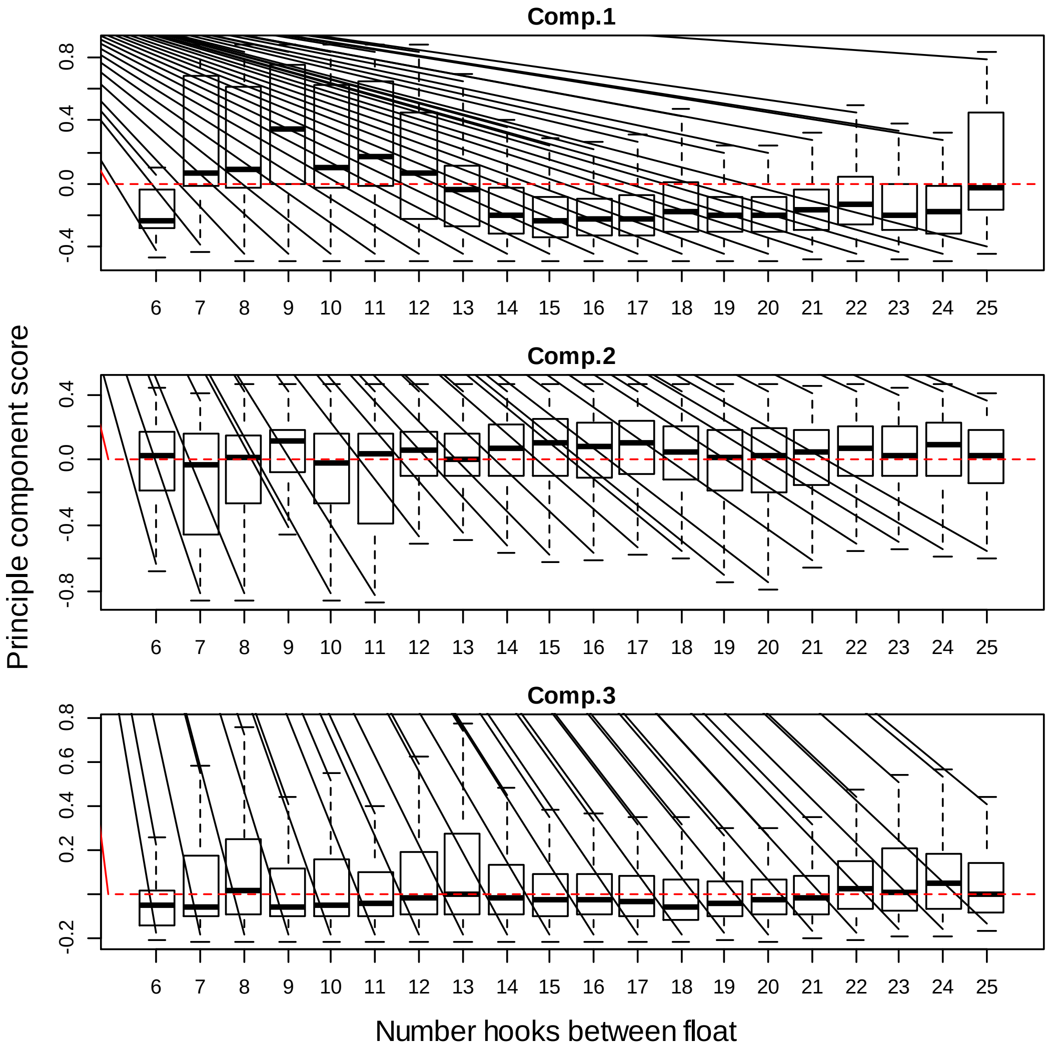

Fig. 8. Principle component scores by number hooks between float based on the first, the second and the third principle component for Taiwanese longline fishery in the Indian Ocean.

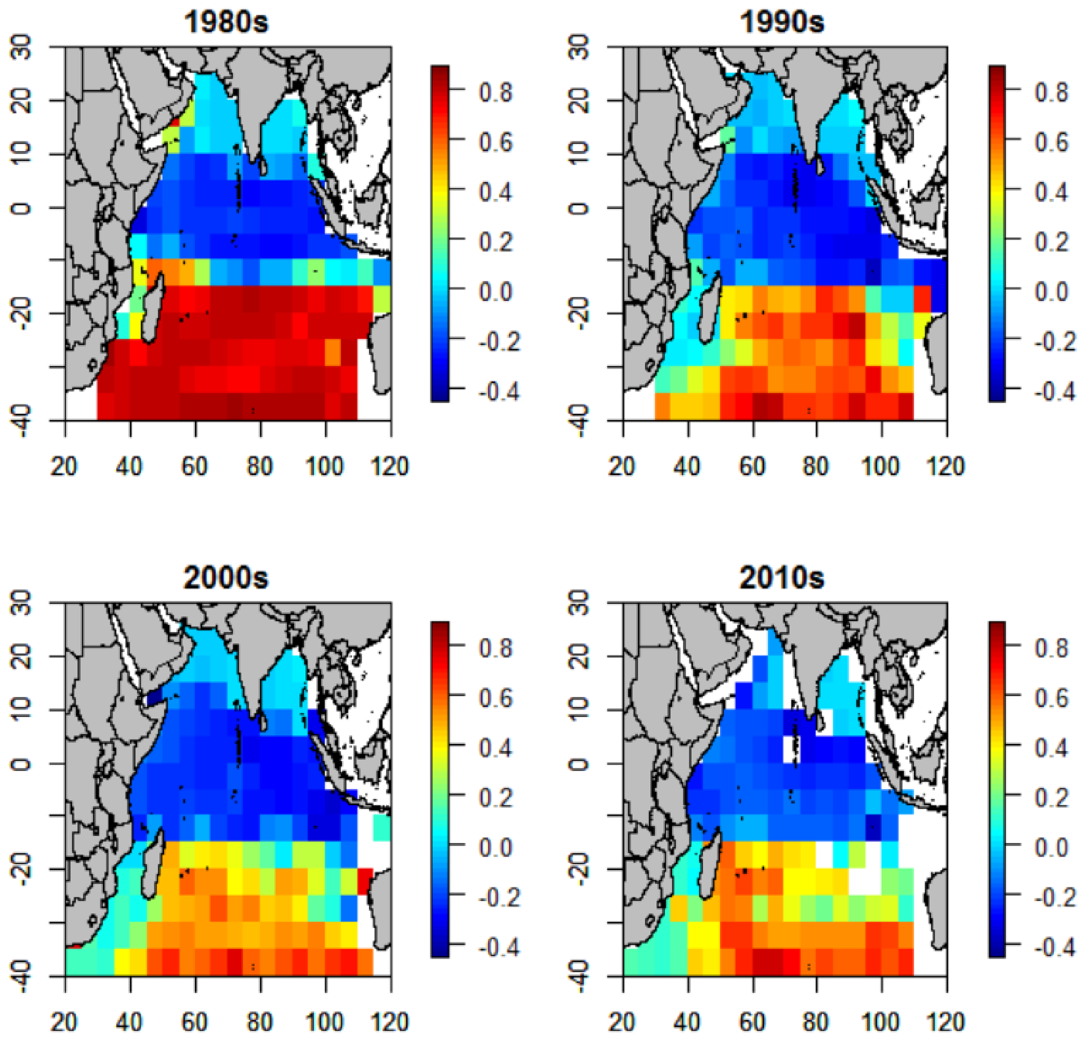

Fig. 9. Distributions of the scores for the first principle component estimated based on catch compositions of Taiwanese longline fishery.

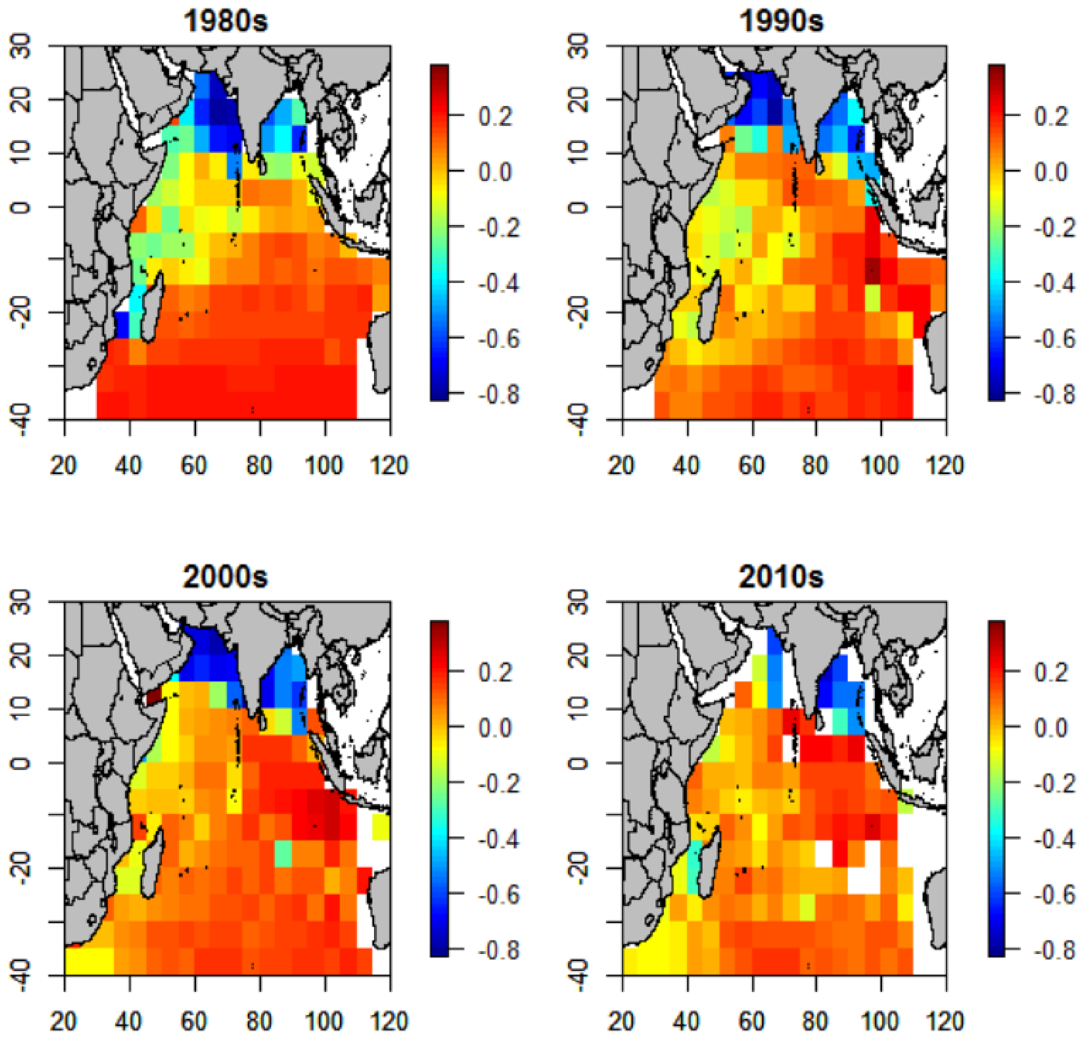

Fig. 10. Distributions of the scores for the second principle component estimated based on catch compositions of Taiwanese longline fishery.

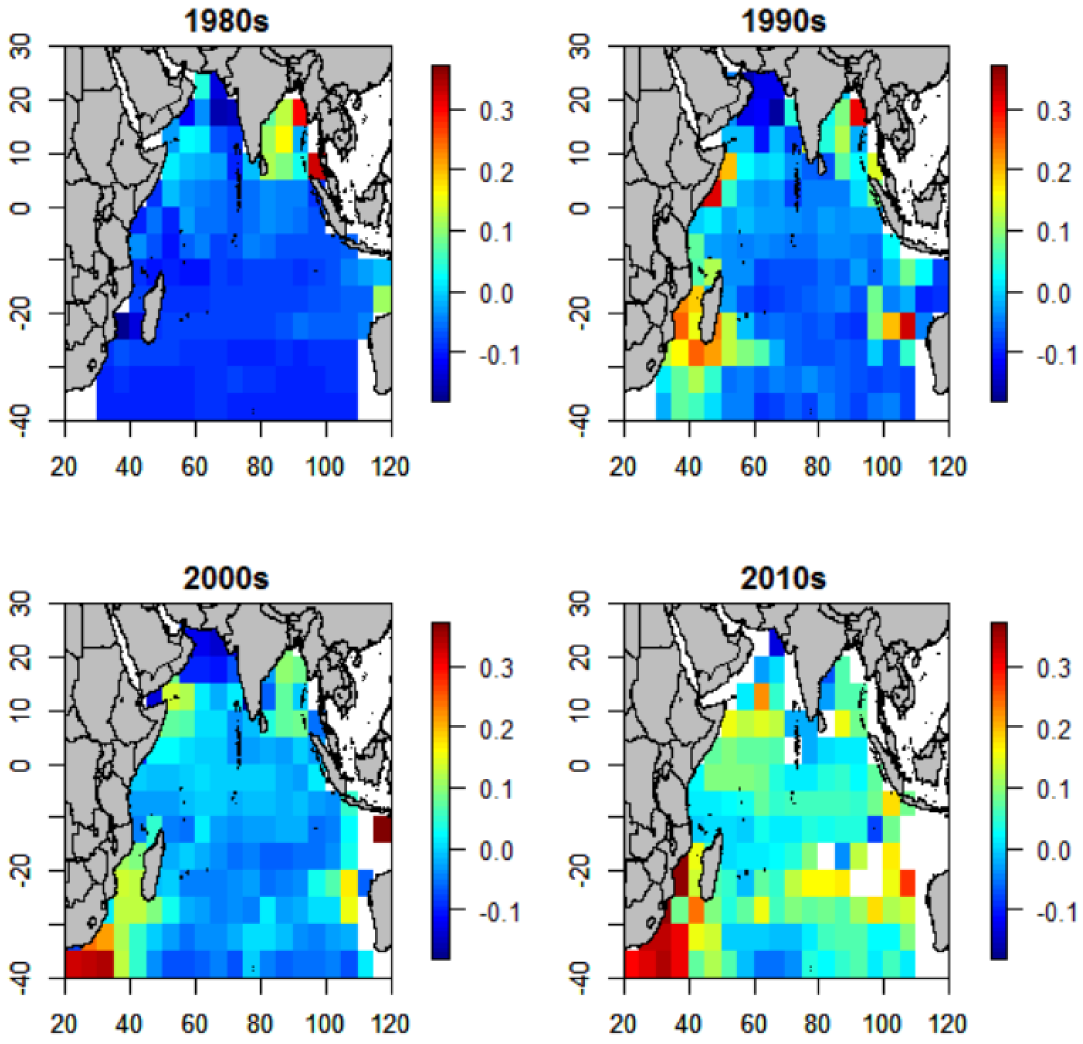

Fig. 11. Distributions of the scores for the third principle component estimated based on catch compositions of Taiwanese longline fishery.

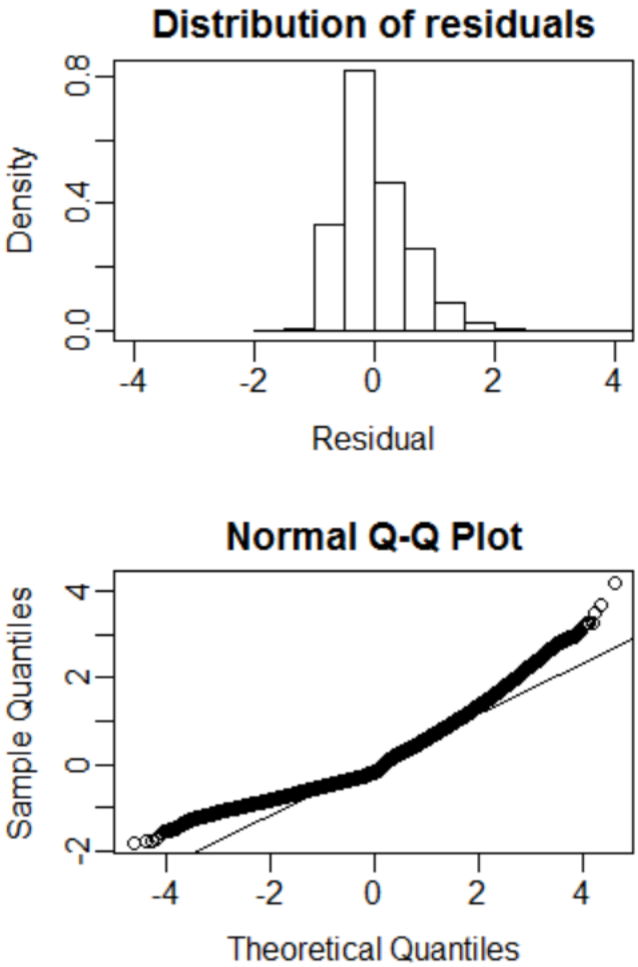

Fig. 12. The distributions and quantile-quantile plots of standardized residuals for lognormal and delta models incorporated principle component scores (*PC_i*) as effects of targeting.

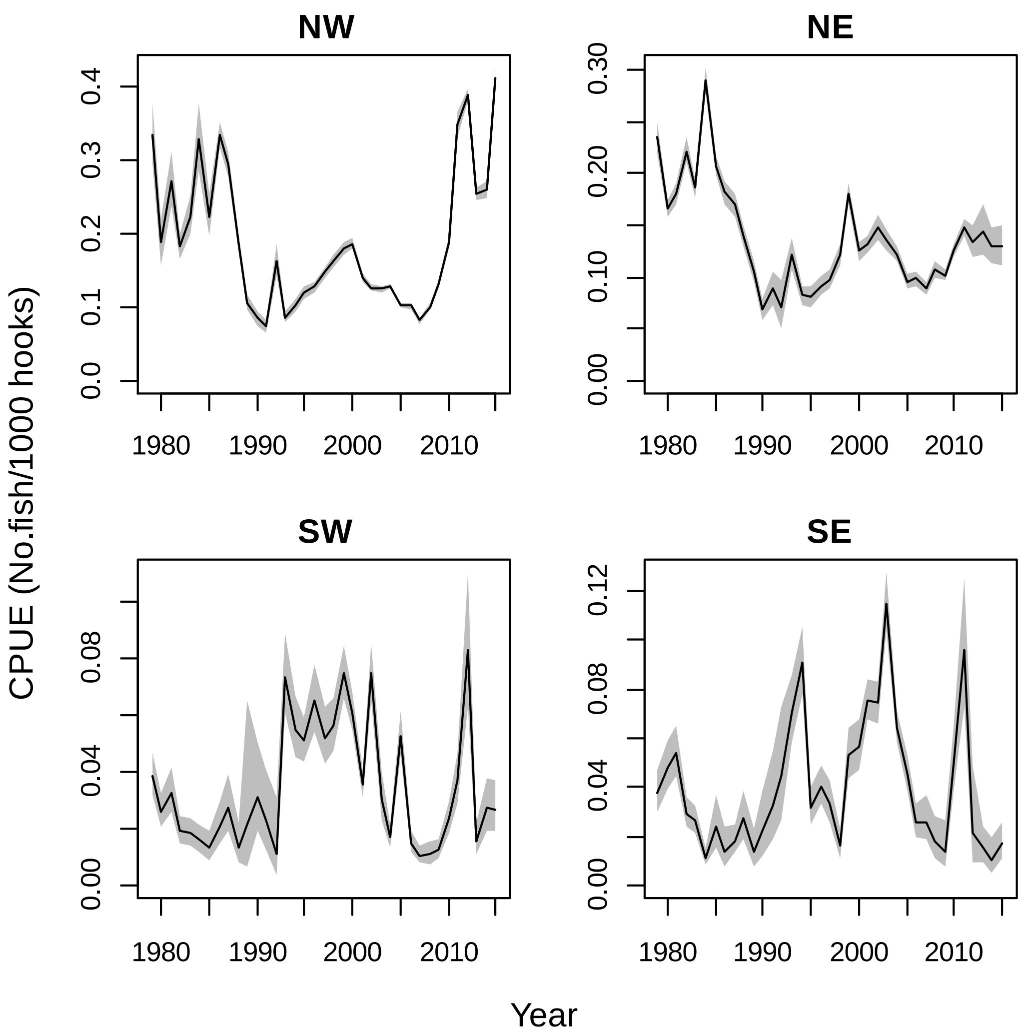

Fig. 13. Area-specific standardized (lines) CPUE with 95% confidence interval (shaded areas) for blue marlin of Taiwanese longline fishery in the Indian Ocean. CPUEs were scaled by the averaged value for each series.

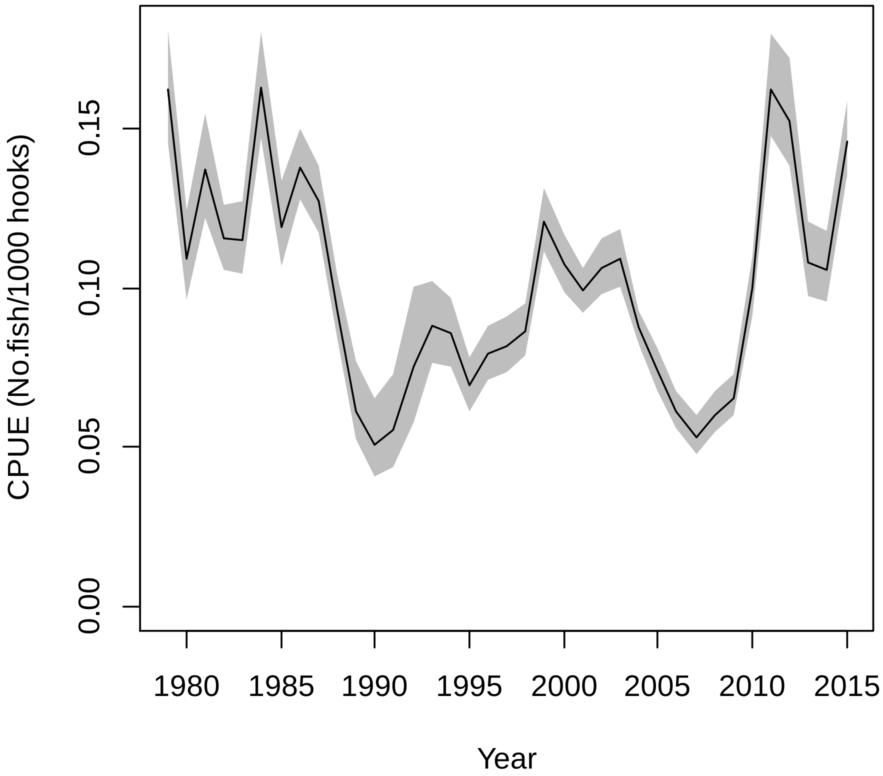

Fig. 14. Area-aggregated standardized (line) CPUE with 95% confidence interval (shaded area) for blue marlin of Taiwanese longline fishery in the Indian Ocean. CPUEs were scaled by the averaged value for each series.

Table 1. 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	0.401	0.304	0.158	0.096	0.057	0.045
Proportion of Variance	0.549	0.316	0.085	0.032	0.011	0.007
Cumulative Proportion	0.549	0.865	0.950	0.982	0.993	1.000

Table 2. 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	0.790	0.296	-0.463	-0.252	-0.084	0.048
BET	-0.596	0.569	-0.513	-0.225	-0.077	0.033
YFT	-0.136	-0.766	-0.570	-0.254	-0.061	0.037
SWO	-0.036	-0.012	0.433	-0.898	-0.070	0.015
MLS	-0.008	-0.021	0.076	0.112	-0.983	-0.124
BUM	-0.014	-0.007	0.064	0.057	-0.113	0.990

Table 3. The ANOVA tables for selected lognormal models incorporated principle component scores (PC_i) as effects of targeting.

Variables	Type III SS	Df	F	Pr(>F)
Y	3292	36	283.093	< 2.2e-16 ***
M	293	11	82.512	< 2.2e-16 ***
A	28	3	28.73	< 2.2e-16 ***
PC1	6	1	19.463	1.03E-05 ***
PC2	18	1	55.168	1.11E-13 ***
PC3	40	1	124.709	< 2.2e-16 ***
Y:A	816	108	23.378	< 2.2e-16 ***
M:A	272	33	25.471	< 2.2e-16 ***
M:PC1	77	11	21.658	< 2.2e-16 ***
M:PC2	47	11	13.288	< 2.2e-16 ***
M:PC3	45	11	12.604	< 2.2e-16 ***
A:PC1	137	3	141.058	< 2.2e-16 ***
A:PC2	16	3	16.808	6.51E-11 ***
A:PC3	43	3	44.442	< 2.2e-16 ***
PC1:PC2	7	1	21.469	3.60E-06 ***
PC1:PC3	334	1	1032.805	< 2.2e-16 ***
PC2:PC3	289	1	895.442	< 2.2e-16 ***
Residuals	78521	243082		

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

Table 4. The ANOVA tables for selected delta models incorporated principle component scores (PC_i) as effects of targeting.

Variables	LR Chisq	Df	Pr(>Chisq)
Y	23173.3	36	< 2.2e-16 ***
M	660.3	11	< 2.2e-16 ***
A	576.7	3	< 2.2e-16 ***
PC1	186.1	1	< 2.2e-16 ***
PC2	65	1	7.40E-16 ***
PC3	159.8	1	< 2.2e-16 ***
Y:A	14073.5	108	< 2.2e-16 ***
M:A	2777.6	33	< 2.2e-16 ***
M:PC1	188	11	< 2.2e-16 ***
M:PC2	85.8	11	1.09E-13 ***
M:PC3	129	11	< 2.2e-16 ***
A:PC1	431	3	< 2.2e-16 ***
A:PC2	683.4	3	< 2.2e-16 ***
A:PC3	619.3	3	< 2.2e-16 ***
PC1:PC2	1113.4	1	< 2.2e-16 ***
PC1:PC3	5498.1	1	< 2.2e-16 ***
PC2:PC3	2994.1	1	< 2.2e-16 ***

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