CPUE standardization of swordfish (*Xiphias gladius*) caught by Taiwanese longline fishery in the Indian Ocean

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 general linear models were used to conduct the CPUE standardization of swordfish caught by the Taiwanese longline fishery in the Indian Ocean for 1979-2016 because swordfish was caught as bycatch and large amounts of zero catches was recorded in the operational data sets. The trends of CPUE series were obviously different by areas, while the trend of area-aggregated CPUE series generally revealed an increasing trend before the early 2000s, then substantially decreased and remained at a level lower than that in the early 1980s.

1. INTRODUCTION

Taiwanese longline fishery in the Indian Ocean commenced in the mid-1950s and targeted on yellowfin tuna in the beginning. Following the development of the fishery, two different operation patterns were currently established: the first targets on albacore for canning and the other on tropical tuna species (bigeye tuna and yellowfin tuna) for sashimi market. Since 1990's, however, swordfish has become a seasonal targeting species to some of the fleets. Most catches of swordfish in the Indian Ocean were made by lognline fisheries, especially for Taiwanese longline fishery (seasonal targeting fishery) and Japanese longline fishery (exploited as bycatch), which have the longest period of catch data series. Furthermore, Taiwanese longline fishery made highest proportion of swordfish (about 50-70%) than other fisheries since 1970's although the proportion (about 40-55%) decreased during recent decades (IOTC, 2016).

Fig. 1 shows the historical catches by species and catch proportion of swordfish based on the logbook data of Taiwanese fishery, and the annual proportion of swordfish was generally less 10% of total catches and revealed a decreasing trend after the late 1990s. Fig. 2 shows the nominal CPUE distribution of swordfish of Taiwanese fleet. CPUE in the 1980s was relatively lower than other years but some high CPUE occurred in the offshore area; CPUE substantially increased and occurred in the tropical area and the southwestern Indian Ocean in the 1990s and 2000s; CPUE obviously decreased in the southwestern Indian Ocean in the 2010s.

Because swordfish 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. In recent decades, the annual proportions of zero-catch were about 40-70% 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, a popular way to deal with zeros is through the delta approach (Hinton and Maunder, 2004; Maunder and Punt, 2004). In addition, IOTC (2016) noted the use of the delta-lognormal approach to accommodate the high proportion of zero catches. Therefore, the delta-lognormal GLM (Pennington, 1983; Lo et. al., 1992; Pennington, 1996) was applied to conduct the CPUE standardization of swordfish 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 billfishes in the Indian Ocean (Wang, 2015; 2016). 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 1980-2016 were provided by Oversea Fisheries Development Council of Taiwan (OFDC). It should be noted that the data in 2016 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 the 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, 2017).

2.3. CPUE Standardization

In previous studies, the effect of the fishing areas, which was defined for swordfish in the Indian Ocean (Fig. 3) (Wang and Nishida, 2011), was incorporated into the CPUE standardization models. However, IOTC (2016) considered that the effect of the area may not reflect the influence of the movement of fishing vessels on the CPUE. In addition, IOTC (2016) noted the high number of parameters in the model, especially in the year*area interaction, and the difficulty in interpreting some of these, such as the interactions between the different principal components and other covariates.

In this study, therefore, the models were simply conducted with the main effects considered in this analysis were year, month, 5x5 longitude-latitude grid, and the effects related to the fishing configurations (principal component scores), while interactions between main effects were not incorporated into the models. In addition, CPUE standardizations were also performed by four fishing areas separately (Fig. 3). The lognormal and delta models were conducted as follows:

Lognormal model for CPUE of positive catch:

$$CPUE = \mu + Y + M + A + T + \varepsilon^{\log}$$

Delta model for presence and absence of catch:

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

where	$CPUE$	is the nominal CPUE of positive catch of swordfish (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,
	G	is the effect of 5x5 longitude-latitude grid,
	T	is the effect of targeting (principal component scores (PC_i) derived from the ith principle component),
	ε^{log}	is the error term, ε^{log} ~ normal distribution with 0 mean,
	ε^{del}	is the error term, ε^{del} ~ Binomial distribution.

The models performed by stepwise search ("both" direction, i.e. "backward" and "forward") and selected based on the values of the coefficient of determination (R^2), Akaike information criterion (AIC) and Bayesian information criterion (BIC). The standardized CPUE were 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 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 standardized probability of positive catches:

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

where *CPUE* is the adjust means (least square means) of the year effect of the lognormal model,

$\tilde{P}$ is the adjust means (least square means) of the year effect of the delta model.

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 BET). For PC2, more negative principle component scores occurred when NHBF were less than 11 hooks (i.e. targeting YFT), while principle component score tended to be positive when NHBF were more than 12 hooks (i.e. targeting 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 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 the effects were statistically significant and remained in the models. The selected lognormal model was:

$$\log(CPUE) = \mu + Y + M + G + PC1 + PC2 + PC3$$

The ANOVA tables for selected lognormal models are shown in the Table 3. The results indicated that the main effects of *PC3* (the effect used to identify the targeting SWO or other species) was the most explanatory effect for all of four areas. The effects of *PC1* and *PC2* (the effect used to identify the targeting ALB and YFT or ALB) provided significant contributions to explanation of variance for tropical areas in the northern Indian Ocean (areas NW and NE). The effects of *PC1* (the effect used to identify the targeting ALB) were also significant in the southern Indian Ocean (areas SW and SE). The distribution of residuals was close to the assumption of normal distribution (Fig. 12).

Similarly, all of main effects were statistically significant and remained in the model for the delta model incorporated principle component scores (*PCi*). The selected delta model was:

$$PA = \mu + Y + M + G + PC1 + PC2 + PC3$$

The ANOVA tables for selected delta models are shown in the Table 4. The most explanatory main effect for the mode were the effects of *PC3* and the secondary explanatory effect was the effect of 5x5 grid. The results indicated that both of the targeting and spatial effects influenced catch probability of swordfish in the Indian Ocean.

The area-specific standardized CPUE series are shown in Fig. 13. The CPUE series in the area NW fluctuated without significant trend. The CPUE series in the area NE fluctuated from 1980 to the early 2000s and substantially decreased. In the area SW, CPUE substantially increased from the late 1980s to early 1990s, gradually decreased until the early 2000s, and substantially decreased again and remained at a low level thereafter. In the area SE, the CPUE series gradually increased before the mid-2000s, then substantially decreased until the late 2000s, and revealed an increasing trend in recent years. Fig. 14 shows the area-aggregated standardized CPUE series of swordfish in the Indian Ocean. The trend of area-aggregated CPUE series generally revealed an increasing trend before the early 2000s, then substantially decreased and remained at a level lower than that in the early 1980s.

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.

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.

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.

Hoyle, 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].

IOTC, 2016. Report of the 14th Session of the IOTC Working Party on Billfish. IOTC-2016-WPB14-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., 2016. CPUE standardization of blue marlin (*Makaira nigricans*) caught by Taiwanese longline fishery in the Indian Ocean using targeting effect derived from principle component analyses. IOTC–2016–WPB14–23.

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.

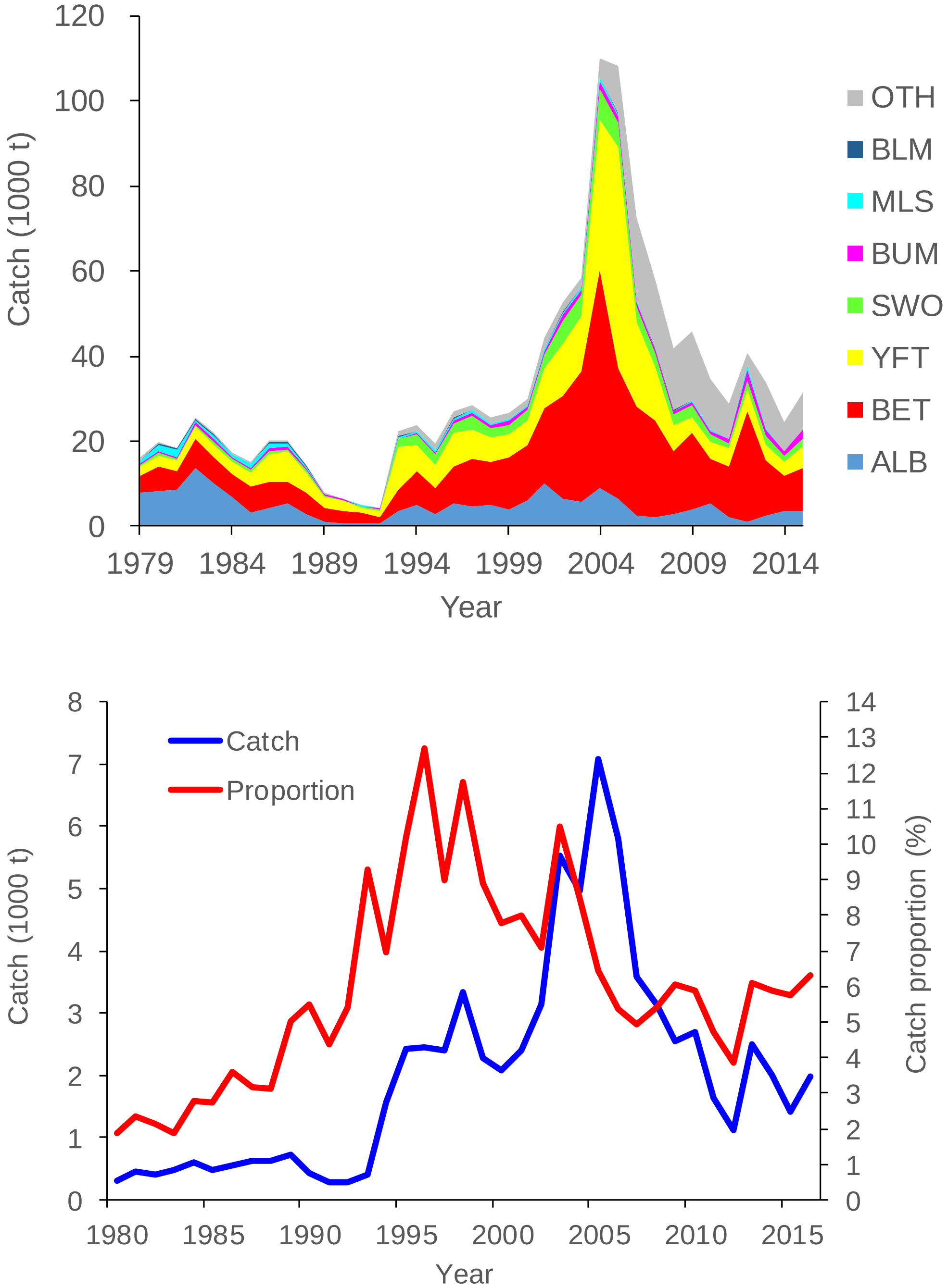


Fig. 1. Annual catches by species (upper panel) and catch of swordfish (lower panel) caught by Taiwanese longline fishery in the Indian Ocean.

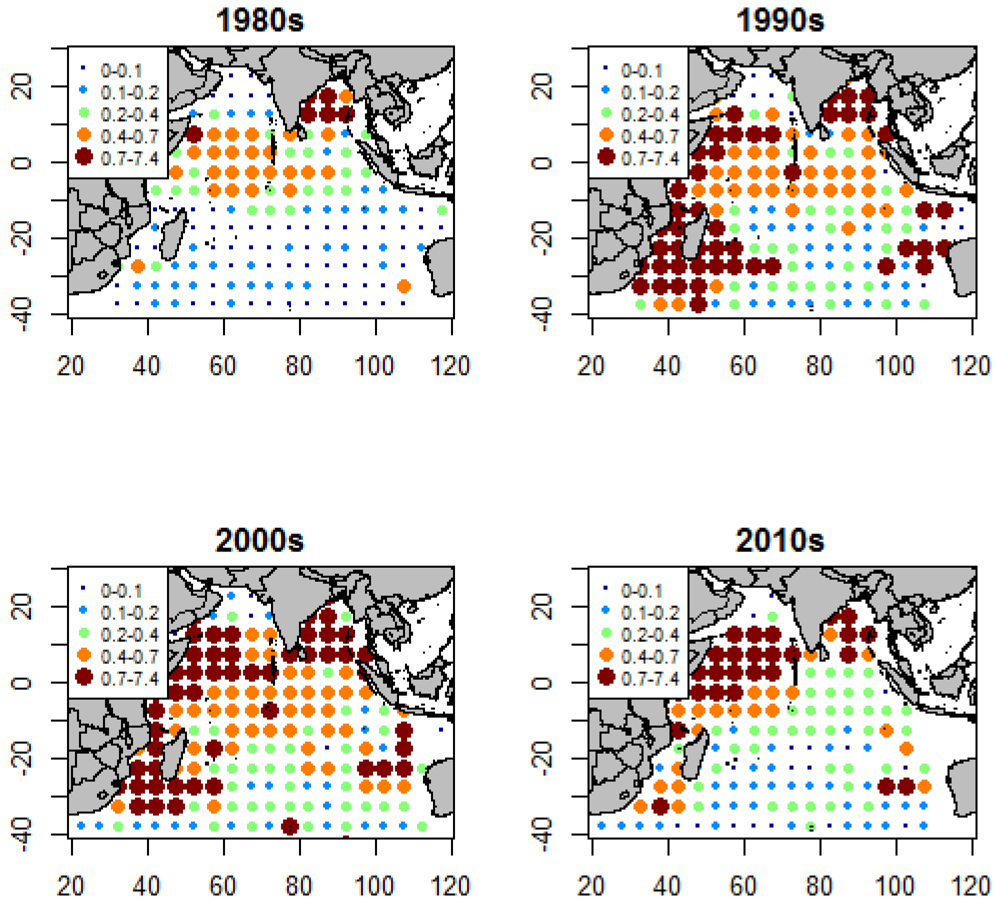


Fig. 2. Nominal CPUE distributions for swordfish caught by Taiwanese longline fishery in the Indian Ocean.

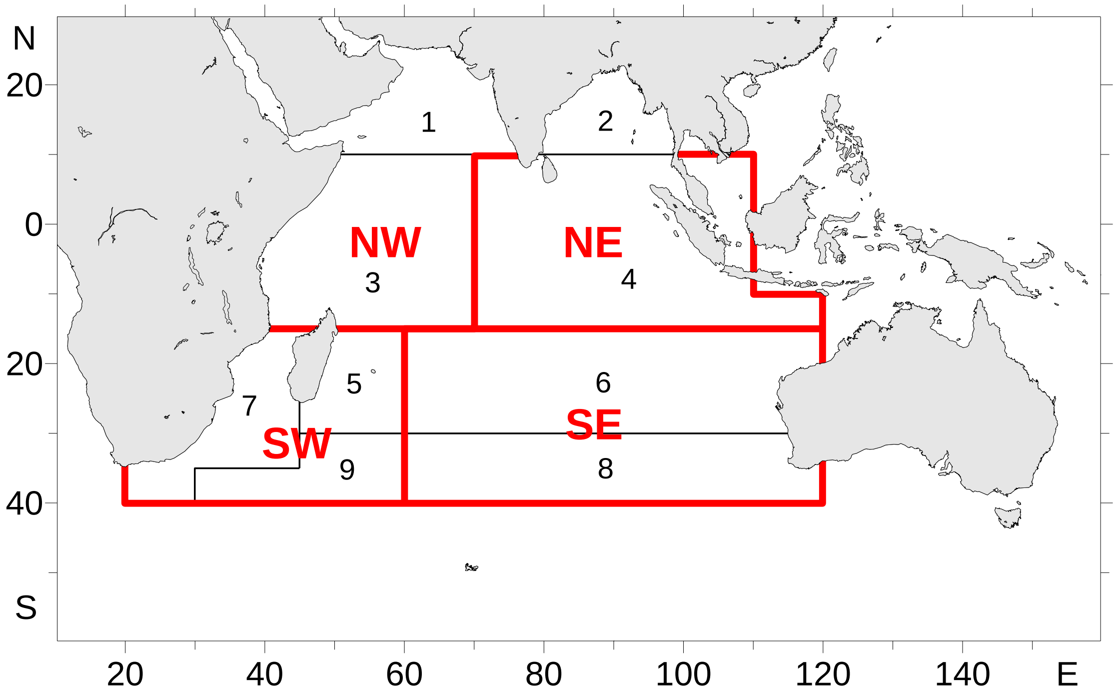


Fig. 3. Area stratification for swordfish in the Indian Ocean.

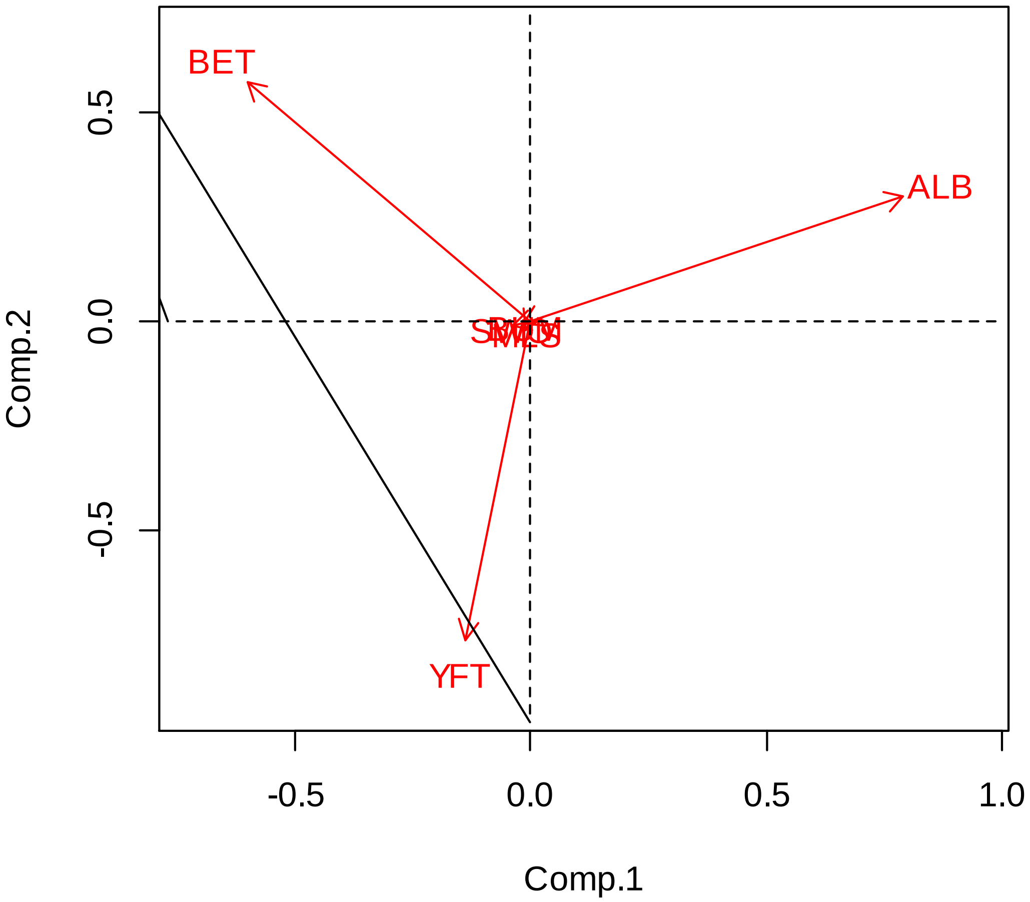


Fig. 4. Relationship between the first and the second principle component for Taiwanese longline fishery in the Indian Ocean.

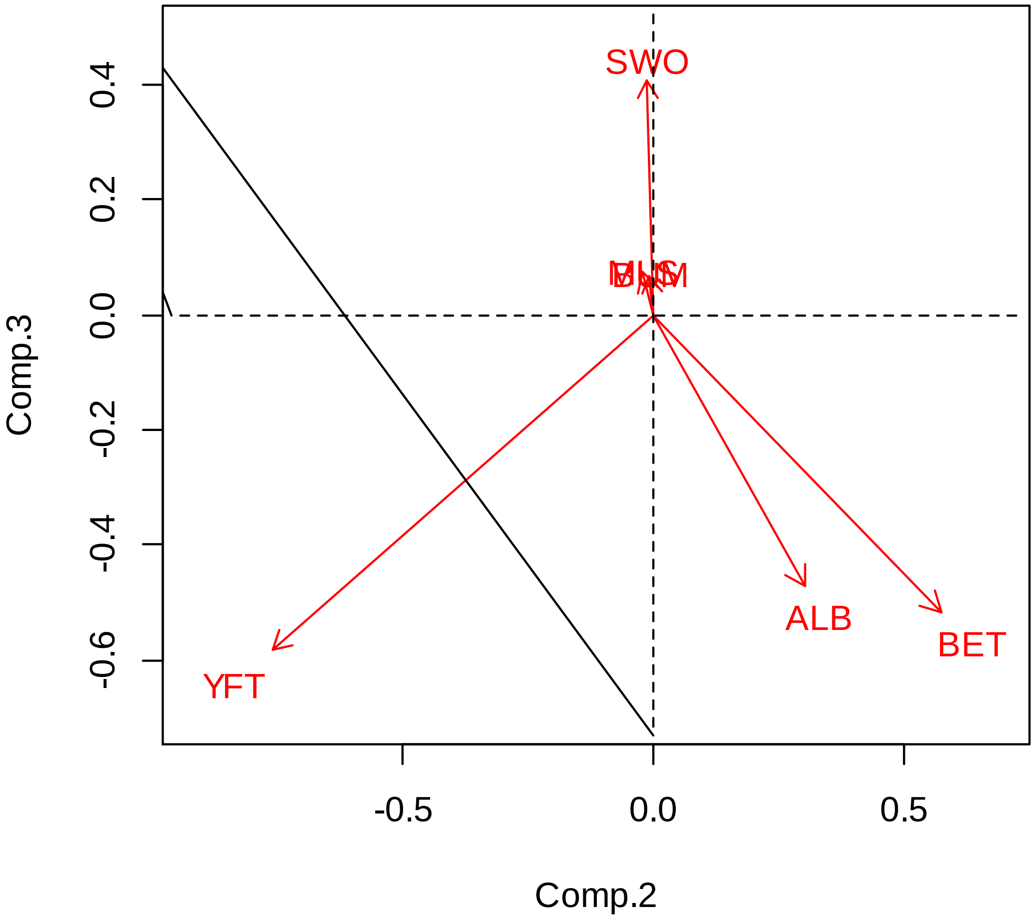


Fig. 5. Relationship between the second and the third principle component for Taiwanese longline fishery in the Indian Ocean.

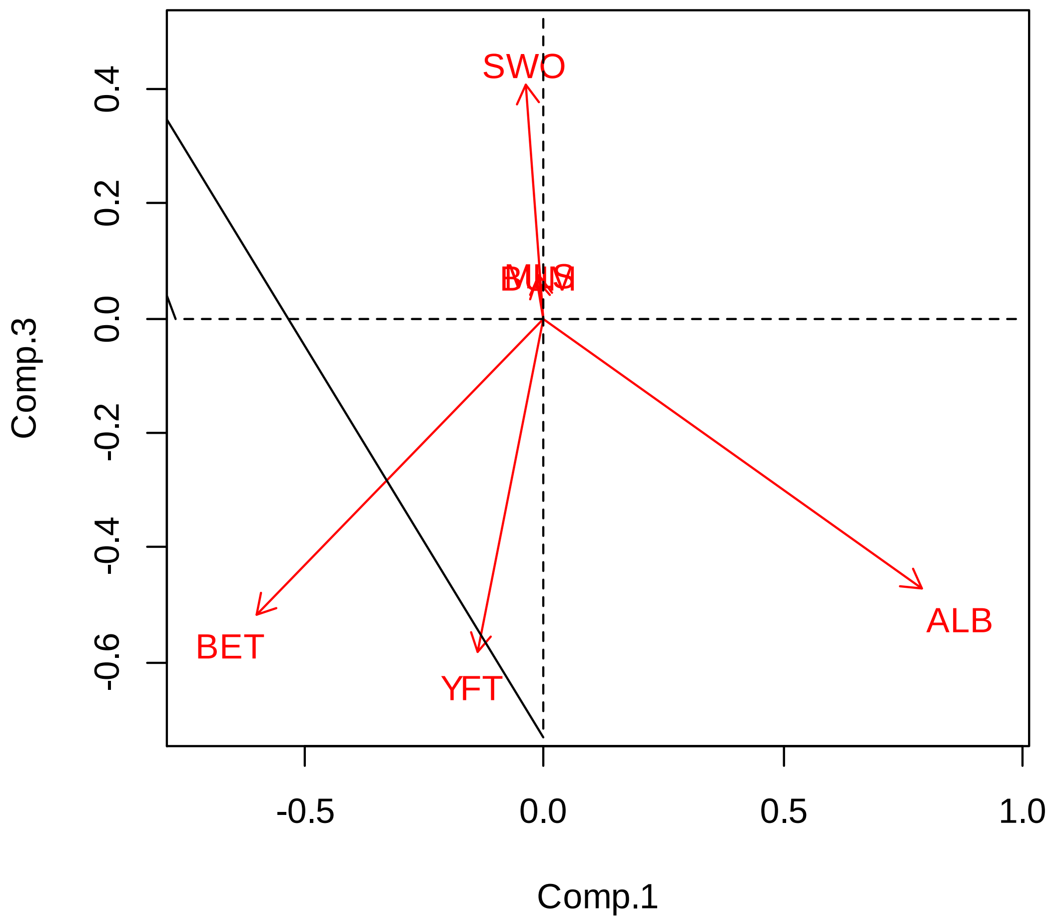


Fig. 6. Relationship between the first and the third principle component for Taiwanese longline fishery in the Indian Ocean.

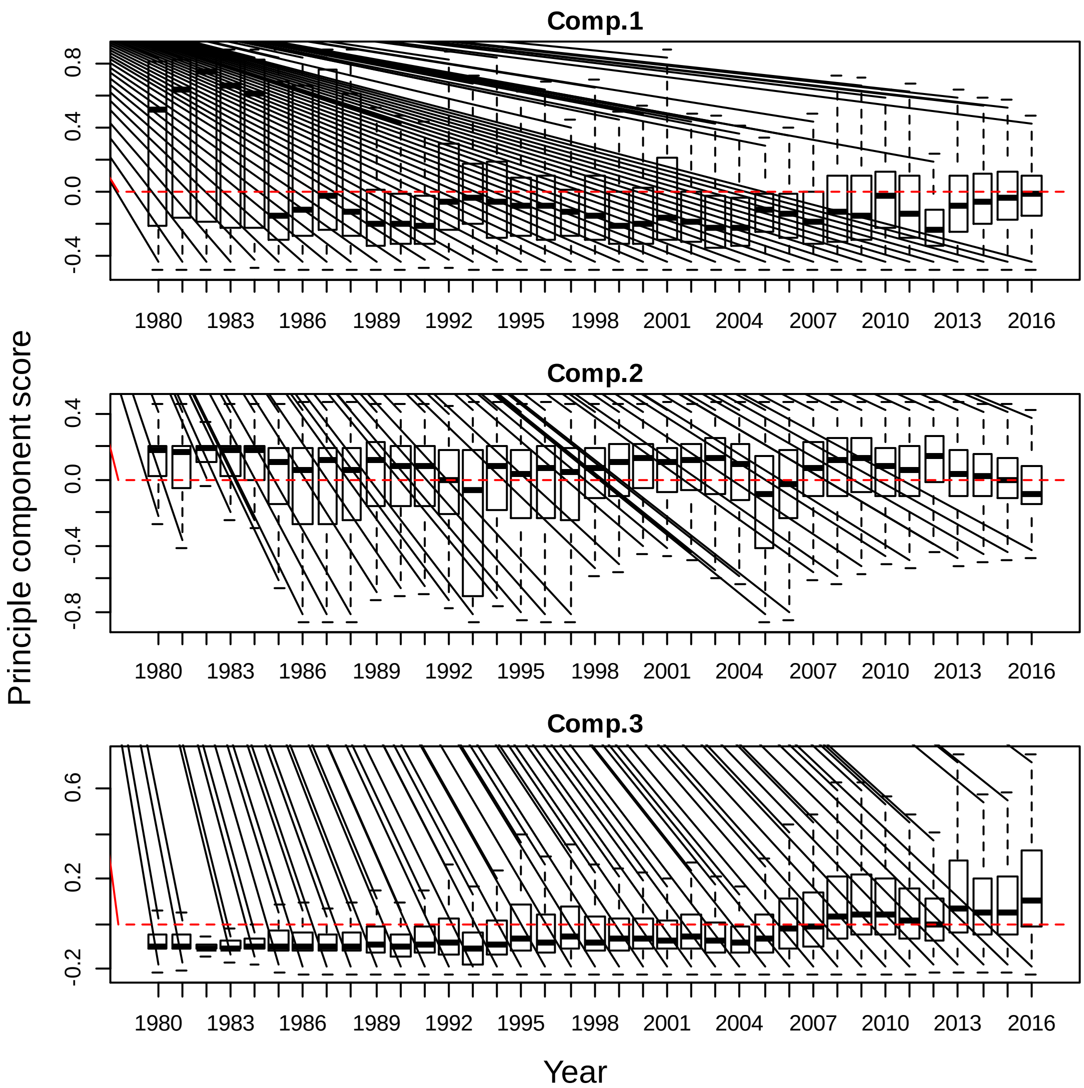


Fig. 7. Principle component scores by year based on the first, the second and the third principle component for Taiwanese longline fishery in the Indian Ocean.

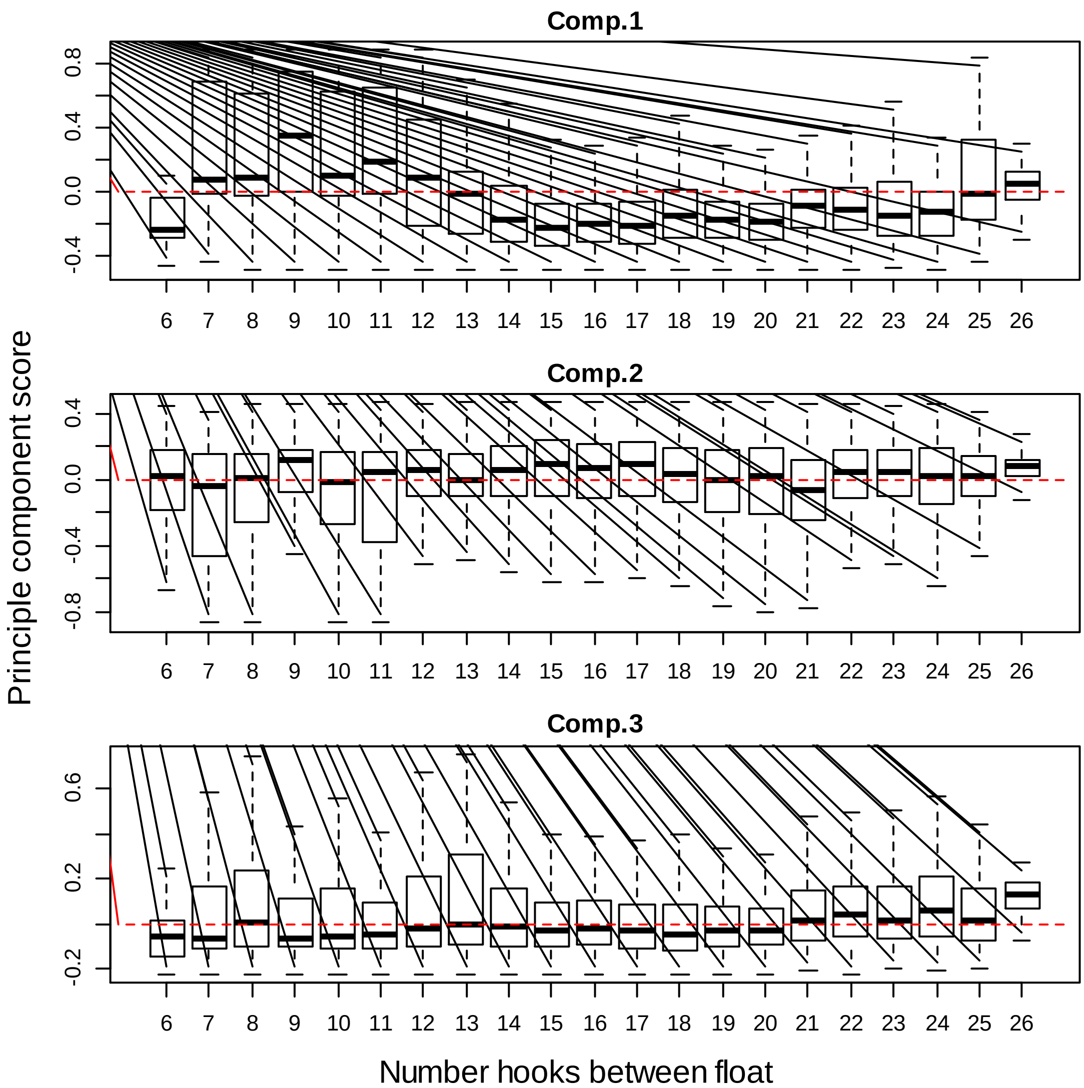


Fig. 8. Principle component scores by number hooks between float based on the first, the second and the third principle component for Taiwanese longline fishery in the Indian Ocean.

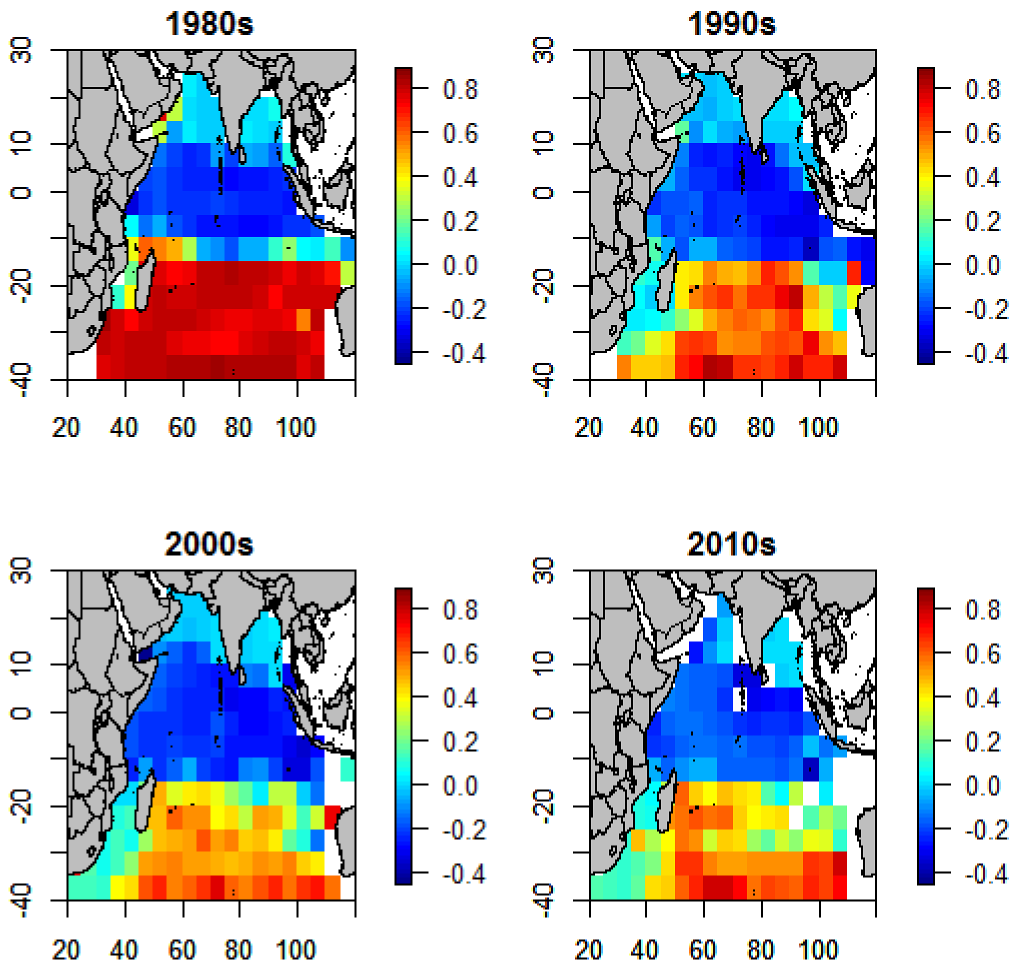


Fig. 9. Distributions of the scores for the first principle component estimated based on catch compositions of Taiwanese longline fishery.

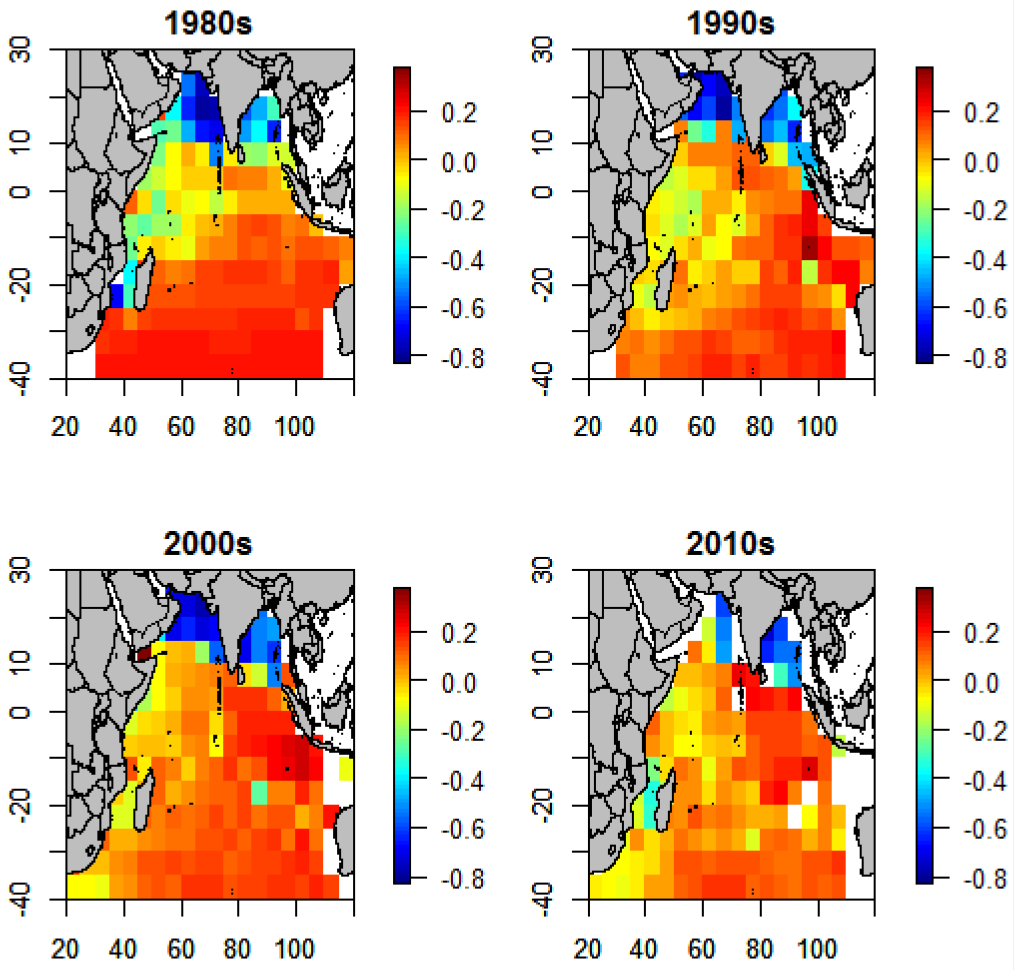


Fig. 10. Distributions of the scores for the second principle component estimated based on catch compositions of Taiwanese longline fishery.

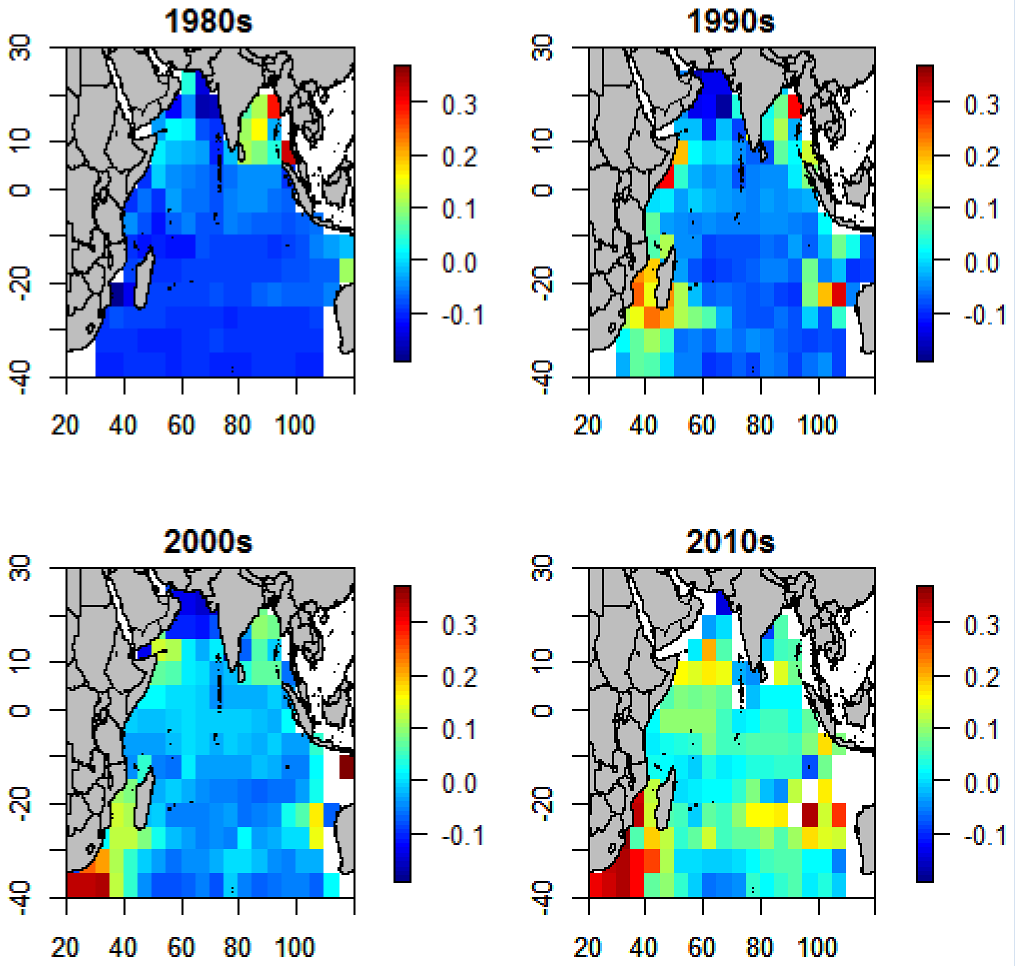
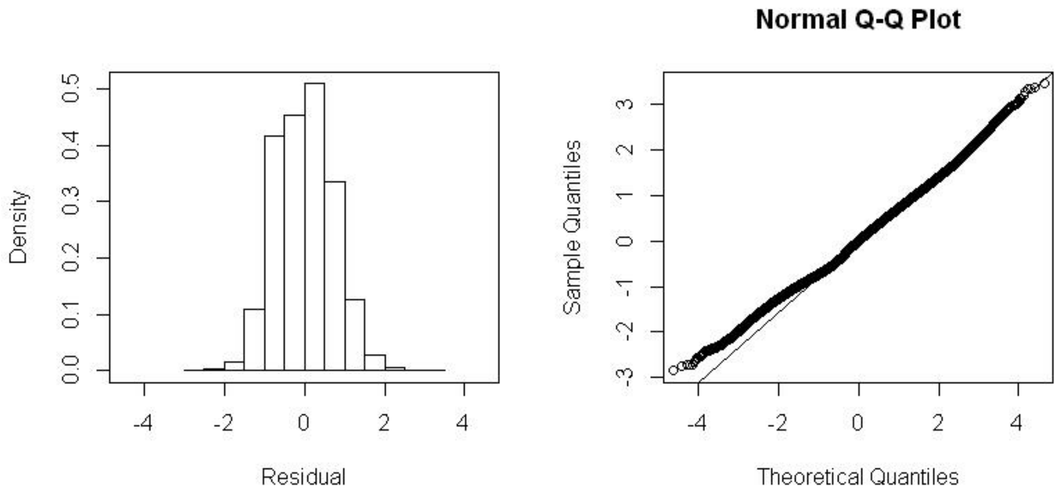


Fig. 11. Distributions of the scores for the third principle component estimated based on catch compositions of Taiwanese longline fishery.

Area NE



Area NE

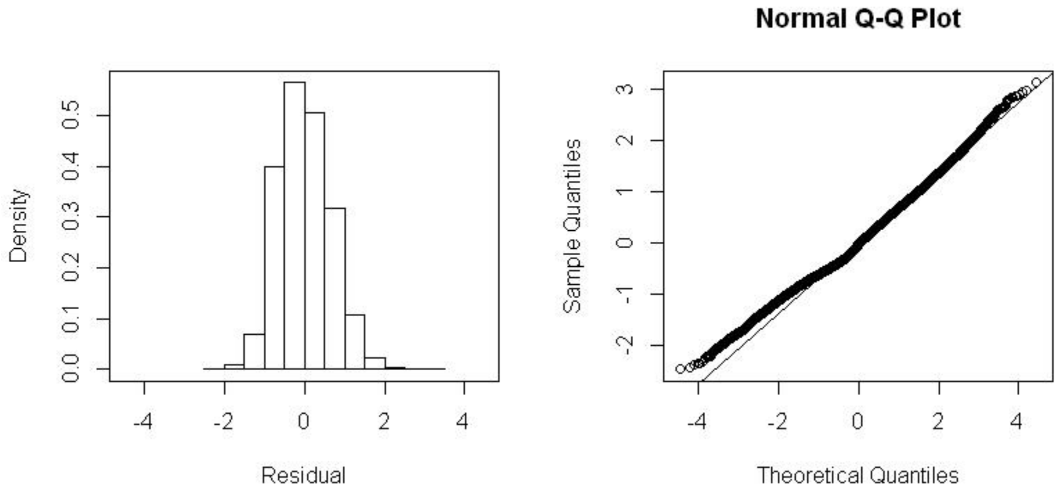
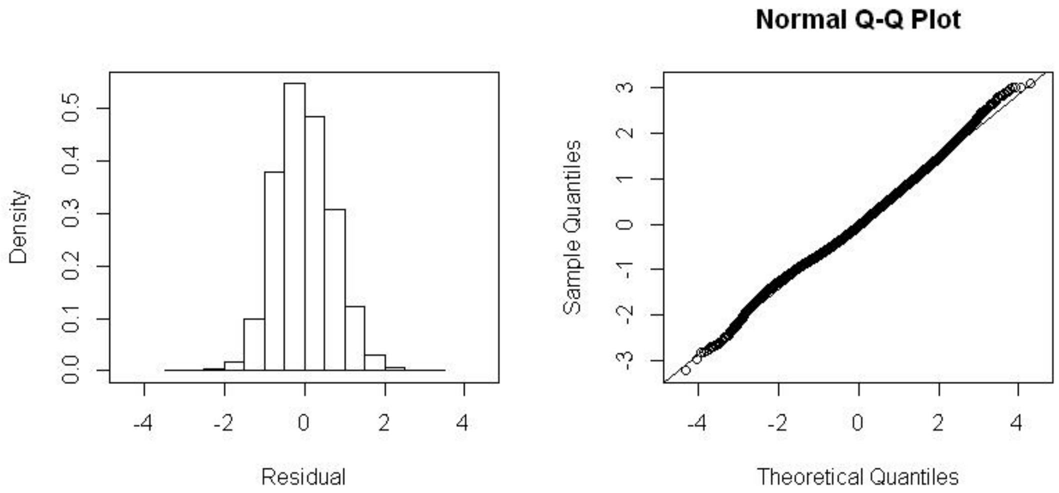


Fig. 12. The distributions and quantile-quantile plots of standardized residuals for lognormal and delta models incorporated principle component scores (PC_i) as effects of targeting.

Area SW



Area SE

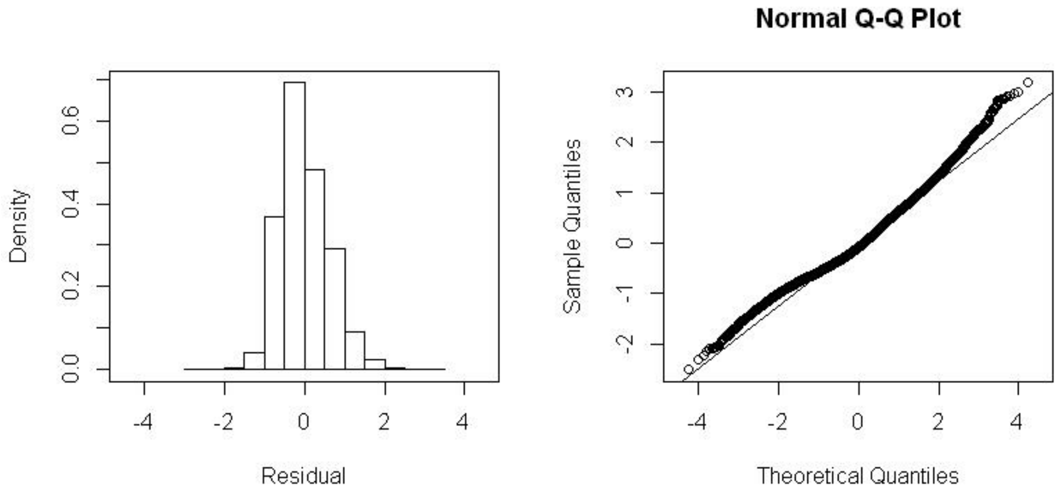


Fig. 12. (Continued).

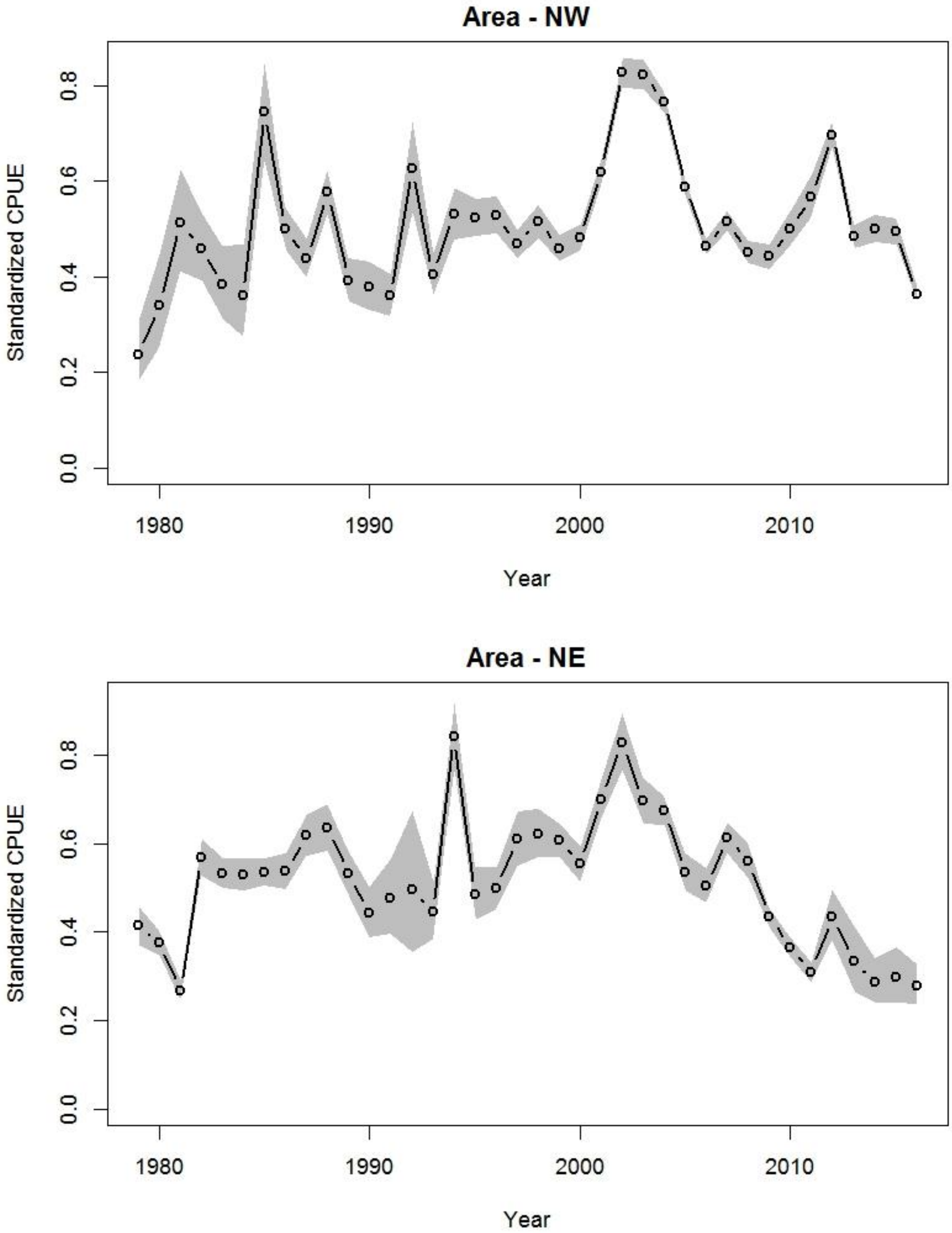
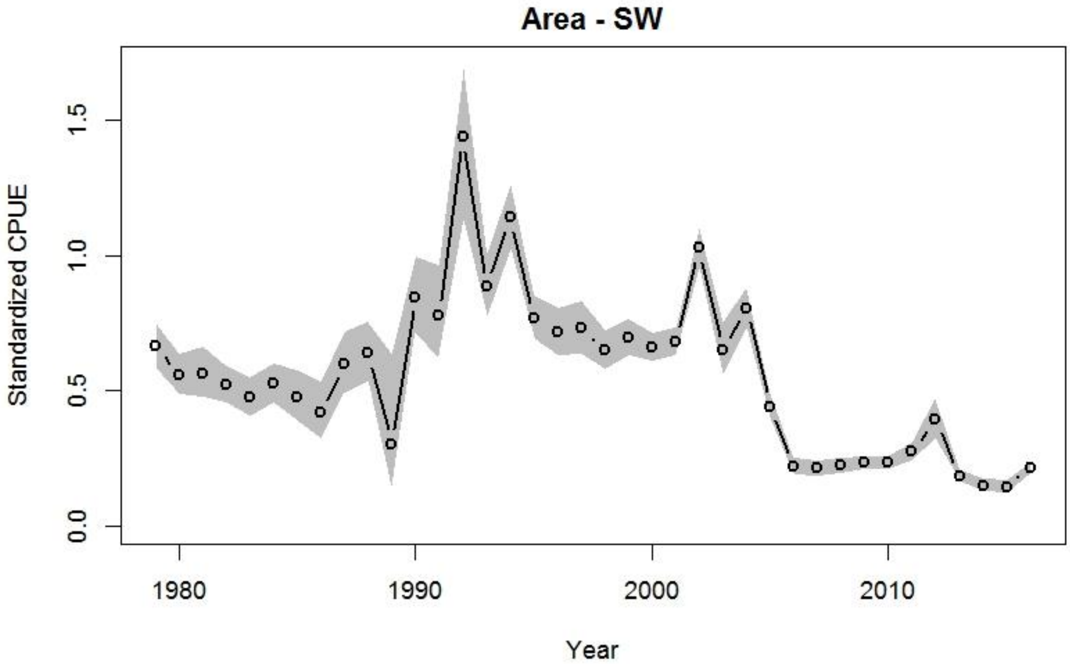


Fig. 13. Area-specific standardized (lines) CPUE with 95% confidence interval (shaded areas) for swordfish of Taiwanese longline fishery in the Indian Ocean. CPUEs were scaled by the averaged value for each series.



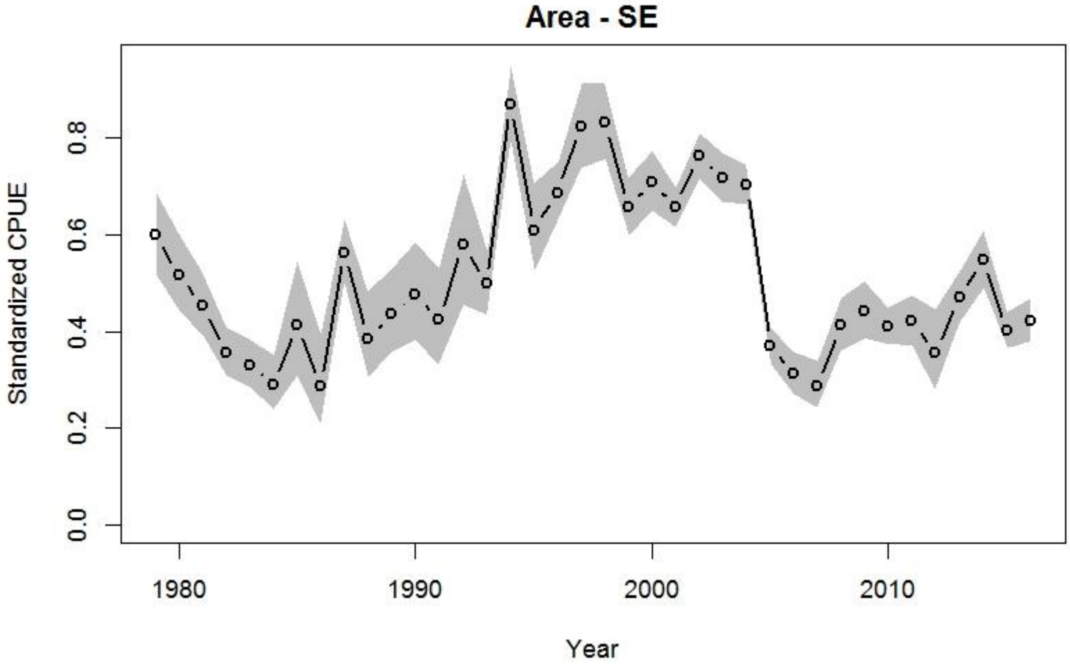


Fig. 13. (Continued).

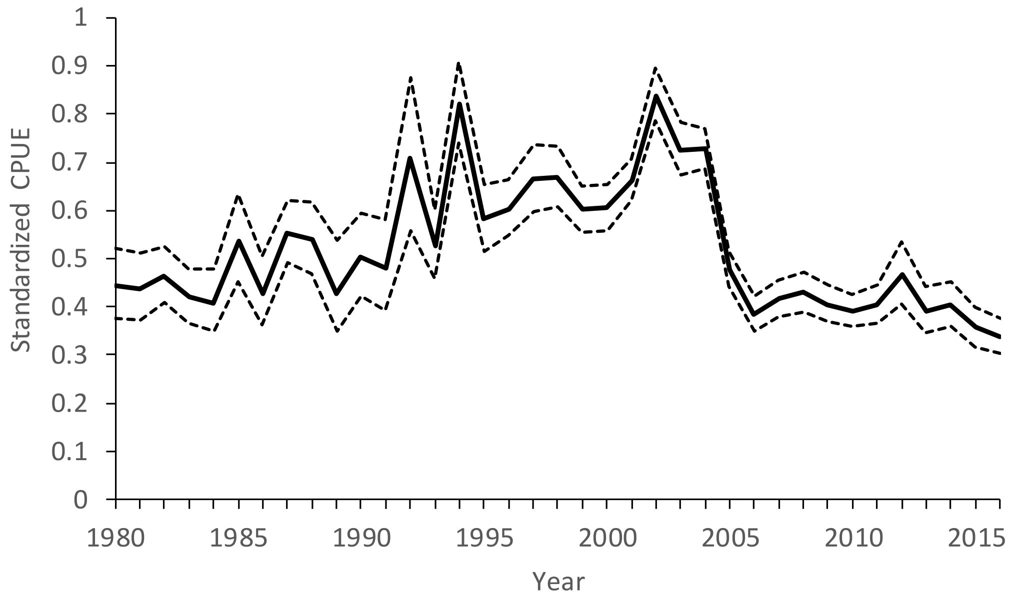


Fig. 14. Area-aggregated standardized (line) CPUE with 95% confidence interval (shaded area) for swordfish of Taiwanese longline fishery in the Indian Ocean. CPUEs were scaled by the averaged value for each series.

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	0.009	0.008	0.018	0.808	0.038	0.022
BET	0.776	0.091	0.469	0.055	0.097	0.040
YFT	0.067	0.702	0.217	0.031	0.077	0.043
SWO	0.030	0.042	0.092	0.013	0.599	0.026
MLS	0.010	0.021	0.017	0.003	0.018	0.062
BUM	0.014	0.020	0.028	0.002	0.023	0.027

Table 3. The ANOVA tables for selected lognormal model.

Area NW

	SS	Df	F	Pr(>F)
Y	6289	37	346.31	< 2.2e-16 ***
M	2554	11	473.15	< 2.2e-16 ***
G	3922	43	185.84	< 2.2e-16 ***
PC1	2585	1	5267.35	< 2.2e-16 ***
PC2	2428	1	4946.79	< 2.2e-16 ***
PC3	26041	1	53058.56	< 2.2e-16 ***
Residuals	137739	280648		

Area NE

	SS	Df	F	Pr(>F)
Y	2671	37	174.026	< 2.2e-16 ***
M	287	11	62.975	< 2.2e-16 ***
G	2280	42	130.891	< 2.2e-16 ***
PC1	462	1	1113.141	< 2.2e-16 ***
PC2	367	1	883.605	< 2.2e-16 ***
PC3	6652	1	16035.466	< 2.2e-16 ***
Residuals	48189	116173		

Area SW

	SS	Df	F	Pr(>F)
Y	5246.9	37	290.654	< 2.2e-16 ***
M	228.6	11	42.602	< 2.2e-16 ***
G	3529.7	30	241.152	< 2.2e-16 ***
PC1	100.5	1	205.979	< 2.2e-16 ***
PC2	35	1	71.648	< 2.2e-16 ***
PC3	9753.4	1	19990.732	< 2.2e-16 ***
Residuals	28844.4	59120		

Area SE

	SS	Df	F	Pr(>F)
Y	1359.7	37	100.511	< 2.2e-16 ***
M	105.6	11	26.265	< 2.2e-16 ***
G	635.3	53	32.786	< 2.2e-16 ***
PC1	179.2	1	490.081	< 2.2e-16 ***
PC2	11.8	1	32.362	1.29E-08 ***
PC3	1880.9	1	5144.36	< 2.2e-16 ***
Residuals	16750.6	45814		

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

Table 4. The ANOVA tables for selected delta models.

Area NW

Variable	LR Chisq	Df	Pr(>Chisq)
Y	8014	37	< 2.2e-16 ***
M	4006	11	< 2.2e-16 ***
G	21269	49	< 2.2e-16 ***
PC1	9543	1	< 2.2e-16 ***
PC2	8381	1	< 2.2e-16 ***
PC3	57424	1	< 2.2e-16 ***

Area NE

	LR Chisq	Df	Pr(>Chisq)
Y	4072.3	37	< 2.2e-16 ***
M	1115.4	11	< 2.2e-16 ***
G	5517.6	42	< 2.2e-16 ***
PC1	3575.5	1	< 2.2e-16 ***
PC2	1296.2	1	< 2.2e-16 ***
PC3	23294.4	1	< 2.2e-16 ***

Area SW

	LR Chisq	Df	Pr(>Chisq)
Y	3596.5	37	< 2.2e-16 ***
M	1568.6	11	< 2.2e-16 ***
G	8565.7	32	< 2.2e-16 ***
PC1	202.9	1	< 2.2e-16 ***
PC2	85.4	1	< 2.2e-16 ***
PC3	9543.7	1	< 2.2e-16 ***

Area SE

	LR Chisq	Df	Pr(>Chisq)
Y	4829.3	37	< 2.2e-16 ***
M	433.4	11	< 2.2e-16 ***
G	984	53	< 2.2e-16 ***
PC1	38.3	1	5.92E-10 ***
PC2	209.8	1	< 2.2e-16 ***
PC3	6241.9	1	< 2.2e-16 ***

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