Received 6 & 29 October 2014

CPUE standardization with targeting analysis for swordfish (*Xiphias gladius*) caught by Taiwanese longline fishery in the Indian Ocean

Sheng-Ping Wang[1] and Tom Nishida[2]

[1] Department of Environmental Biology and Fisheries Science, National Taiwan Ocean University, Keelung, Taiwan.
[2] National Research Institute of Far Seas Fisheries, Fisheries Research Agency, Shimizu, Shizuoka, Japan.

ABSTRACT

In this study, cluster analysis was used to classify the data sets in relation to the species composition of catches. Based on the results of cluster analysis, data sets were grouped into nine clusters and assigned to specific fishing types. The CPUE standardization of swordfish of Taiwanese longline fishery in the Indian Ocean for 1980-2012 using generalized linear model (GLM). Including the effect of cluster related to fishing types substantially improved the performance of GLM. The effect of cluster was the most effective variable for explaining the variance of nominal CPUE. Generally, the trends of standardized CPUE obtained from GLMs with cluster effect were relatively smoother than those obtained from GLMs without cluster effect. In addition, the trends of standardized CPUE were somewhat different when using all data sets or data sets extracted from cluster with higher catch proportions of swordfish, especially for areas SW and SE. Because few data sets were available from swordfish clusters for area SW and SE in early and recent years, the estimates may be highly uncertain. Therefore, we would suggest that the standardized CPUE series obtained from GLM with cluster effect based on all data sets might be more appropriate to be applied to stock assessment as relative abundance indices.

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 of swordfish catch in the Indian Ocean was 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.

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. 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).

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

MATERIAL AND METHODS

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 1980-2012 were provided by Oversea Fisheries Development Council of Taiwan (OFDC). The data of number of hooks between float (NHBF) were available since 1994 but the data were incomplete in 1994. Therefore, the data of NHBF since 1995 were used in the analysis of CPUE standardization.

Cluster analysis

Cluster analysis was conducted based on species composition of the catches. Seven species groups were used in this study, including albacore (ALB), bigeye tuna

(BET), yellowfing tuna (YFT), swordfish (SWO), striped marlin (MLS), blue marlin (BUM) and other species (OTH) (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 21 clusters for taking the mixture of fishing operations into account ($C_2^7 = 21$ ways in which 2 species can be chosen from 7 species groups). Second, a hierarchical cluster analysis with Ward minimum variance method was applied to the squared Euclidean distances calculated from 21 non-hierarchical clusters. Non-hierarchical and hierarchical cluster analyses were conducted using R functions *kmeans* and *hclust* (The R Foundation for Statistical Computing Platform, 2014).

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.

GLM analysis

In this study, General Linear Model (GLM is used to model the logarithm of the nominal CPUE (defined as the number of fish per 1,000 hooks). 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 the GLM. All of effects were treated as category variables. The GLM was conducted as

$$\ln(CPUE + c) = Y + M + A + V + C + \text{interactions} + \varepsilon$$

where *CPUE* is the CPUE of swordfish (catch in number/1000 hooks);
c is the constant value (i.e. 10% of the average nominal CPUE);
Y is the effect of year;
M is the effect of month;
A is the effect of area;
V is the effect of vessel;
C is the effect of targeting (cluster related to fishing type, C/number

of hooks between float, NHBF);
interactions is the interaction between effects.
ε is the error term, $\varepsilon \sim N(0, \sigma^2)$.

In order to investigate the effectiveness of the effect of cluster related to fishing type when fitting GLM to different data sets, four cases with different combination of effect were considered. The working party requested two additional model runs in order to compare the results based on the approach used in this meeting with those based on the approaches used in previous meetings.

Case 1: fit all data sets of 1980-2012 to GLM without the effect of cluster related to fishing type;

Case 2: fit all data sets of 1980-2012 to GLM with the effect of cluster related to fishing type;

Case 3: fit data sets of 1980-2012 extracted from SWO clusters to GLM without the effect of cluster related to fishing type;

Case 4: fit data sets of 1980-2012 extracted from SWO clusters to GLM with the effect of cluster related to fishing type.

Case 5: fit all data sets of 1995-2012 to GLM with the effect of number of hooks between float;

Case 6: fit all data sets of 1995-2012 to GLM with the effect of cluster related to fishing type.

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

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

RESULTS AND DISCUSSION

Cluster analysis

□ For non-hierarchical cluster analysis, Table 1 shows the average proportions SWO catches for 21 K-means clusters. Cluster 4 and 2 consist large average proportions of SWO catches, which are 55.9% and 27.0%. The selection for number of clusters of hierarchical cluster analysis was based on the average proportions of SWO catches obtained from K-means method. The number of clusters was decreased until average proportion of SWO catches of one cluster was still closed to 55.9%. Finally, 9 clusters were chosen and Clusters 4 and 2 contain much more average proportions of SWO catches than other clusters (Table 2 and Fig. 3). Various fishing types were also assigned to each data set based on the results of hierarchical cluster analysis (Table 2).

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.

Figs. 4 and 5 show the historical catches by species and species compositions. It is obvious that Clusters 2 and 4 contained relatively large SWO catches since early-1990s. Large amounts of data sets of Clusters 2 and 4 distributed in the tropical Indian Ocean and the southwestern Indian Ocean (Fig. 6). However, only 9% data sets can be used for GLM if Clusters 2 and 4 were selected (Table 3). Therefore, Clusters 1 and 3, which are classified to be closed to Cluster 2 and 4, were also selected for Cases 3 and 4 when conducting GLM analysis. The data sets extracted from Cluster 1, 2, 3 and 4 were SWO clusters. For data sets of SWO cluster, about 31% of total data sets were used for GLM (Table 3) and these data contains 60% SWO catches and 30% hooks.

CPUEs of SWO obtained from data sets of SWO cluster were obviously higher than those obtained from all data sets (Fig. 7). However, SWO cluster consisted of few data sets and contained few records of SWO catch and hooks before the early-1990s and after the mid-2000s, especially for Area SW and SE (Figs. 8, 9 and 10).

GLM analysis

Table 4 shows the values of R^2 and AIC for four Cases of GLM. The results indicate that including the effect of cluster substantially improved the proportion of explained variances (R^2) and AIC. ANOVA tables for six Cases of GLM were shown in Table5. It is obvious that the effect of cluster was the most effective variable for explaining the variance of nominal CPUE. The effect of cluster was also much more effective against NHBF. Generally, the distributions of standardized residuals were closed to assumption of normal distribution for all Cases and incorporating the effect of cluster make model to fit this assumption much well (Fig. 11). The results of Cases 2 and 4 were selected to present the trends of standardized CPUE for swordfish caught by Taiwanese longline fishery.

The area-specific tends of standardized CPUE based on Cases 2 and 4 were shown in Figs. 12 and 13. For areas NW and NE, the results of Cases 2 and 4 revealed quite similar trends of standardized CPUE. However, the trends of standardized CPUE were somewhat different for areas SW and SE. In addition, the confidence intervals of standardized CPUE for areas SW and SE obtained from Case 4 were much wider than those obtained from Case 2, especially for early and recent years. This may result from few data sets in area SW and SE were available from SWO cluster for early and recent years (Figs. 8, 9 and 10). The estimates may also be highly uncertain for area SW and SE in early and recent years. The area-specific trends of standardized CPUE obtained from the models with cluster or NHBF effects were also shown in Fig. 14. The trend of standardized CPUE obtained from the model with NHBF effect was much more fluctuant than those obtained from the model with cluster effect.

The area-aggregated standardized CPUE also revealed somewhat different patterns among cases (Fig. 15). Similarly, the trend of standardized CPUE obtained from the model based on all data sets (Case 2) was relatively smoother than that obtained from the model based on data of SWO cluster (Case 4). Also, the trend of standardized CPUE obtained from the model with cluster effect (Case 6) was less fluctuant than that obtained from the model with NHBF effect (Case 5).

Based on the results of this study, we would suggest that the standardized CPUE series obtained from GLM with the effect of cluster based on all data sets (Case 2) might be more appropriate to be applied to stock assessment as relative abundance indices.

REFERENCE

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

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.

Nishida, T., Wang, S.P., 2006. Standardization of swordfish (*Xiphias gladius*) CPUE of the Japanese tuna longline fisheries in the Indian Ocean (1975-2004). The fifth session of the IOTC Working Party on Billfish (WPB), Indian Ocean Tuna Commission (IOTC), March 27–31, 2006. Colombo, Sri Lanka. IOTC-2006-WPB-07, 10 pp.

Nishida, T., Kitakado, T., Wang,S.P., 2011. Estimation of the Abundance Index (AI) of swordfish (*Xiphias gladius*) in the Indian Ocean (IO) based on the fine scale catch and effort data of the Japanese tuna longline fisheries (1980-2010). The ninth session of the IOTC Working Party on Billfish (WPB), Indian Ocean Tuna Commission (IOTC), July 4-8, 2011. Victoria, Seychelles. IOTC-2011-WPB09-14.

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.

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.

Wang, S. P., Nishida, T., 2011. CPUE standardization of swordfish (*Xiphias gladius*) caught by Taiwanese longline fishery in the Indian Ocean. IOTC-2011-WPB09-16.

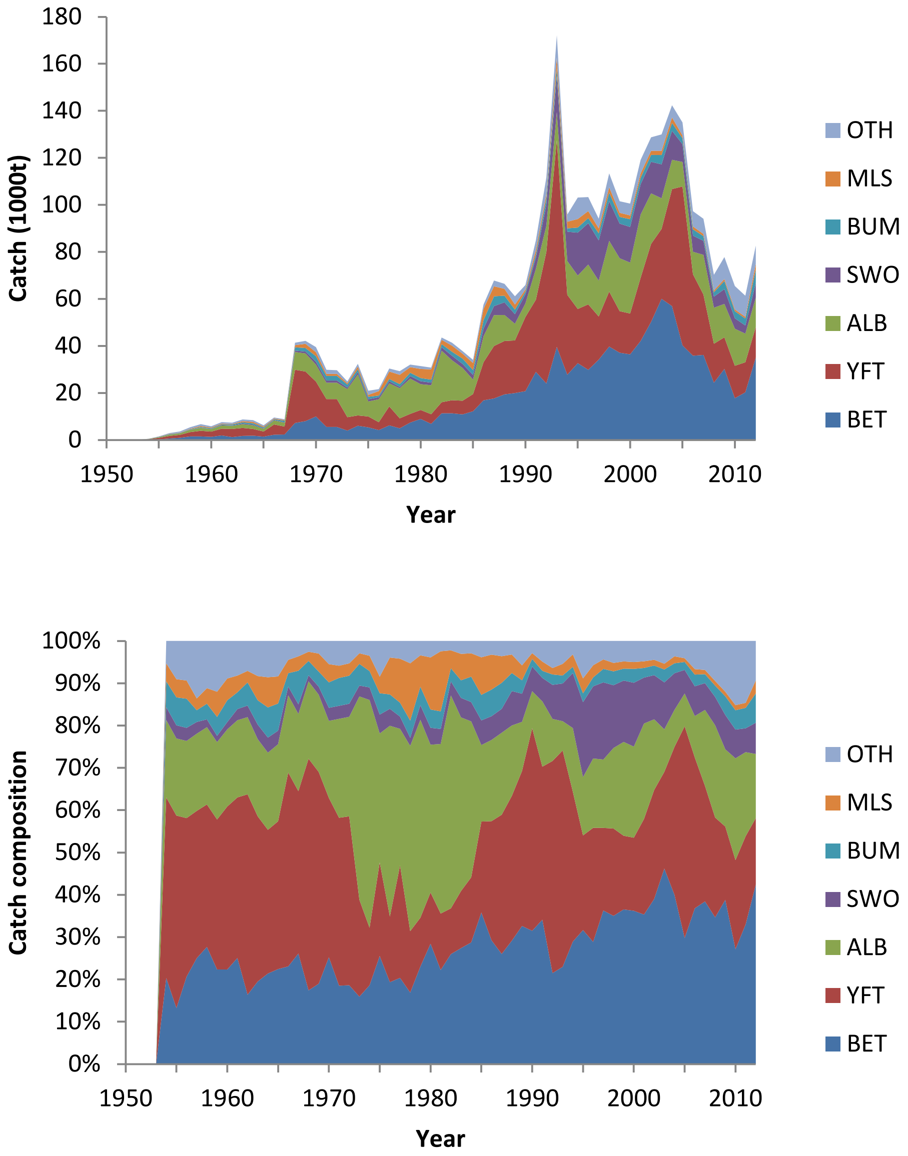

Fig. 1. Annual catches by species and catch compositions of Taiwanese longline fishery in the Indian Ocean.

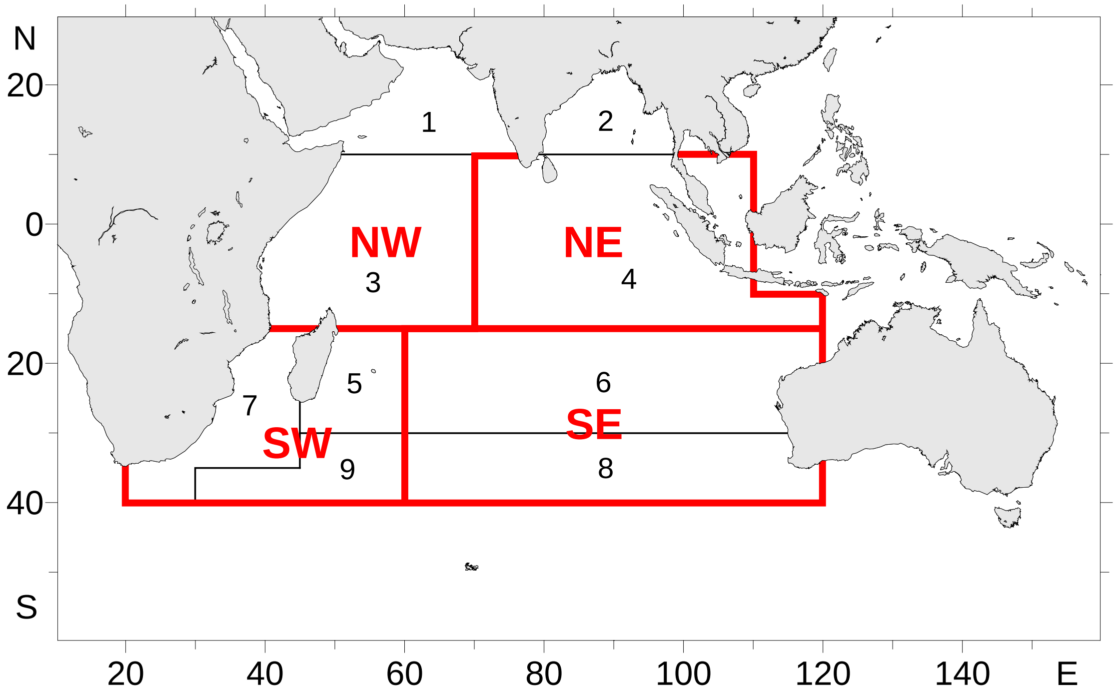

Fig. 2. Area stratification for swordfish in the Indian Ocean.

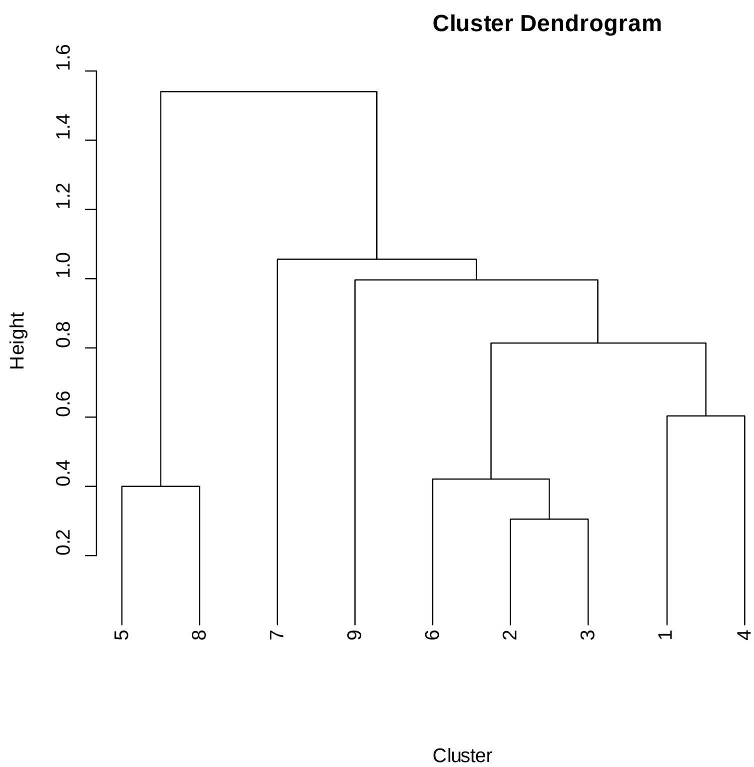

Fig. 3. The dendrogram of hierarchical cluster analysis for classifying the data sets of Taiwanese longline fishery in the Indian Ocean.

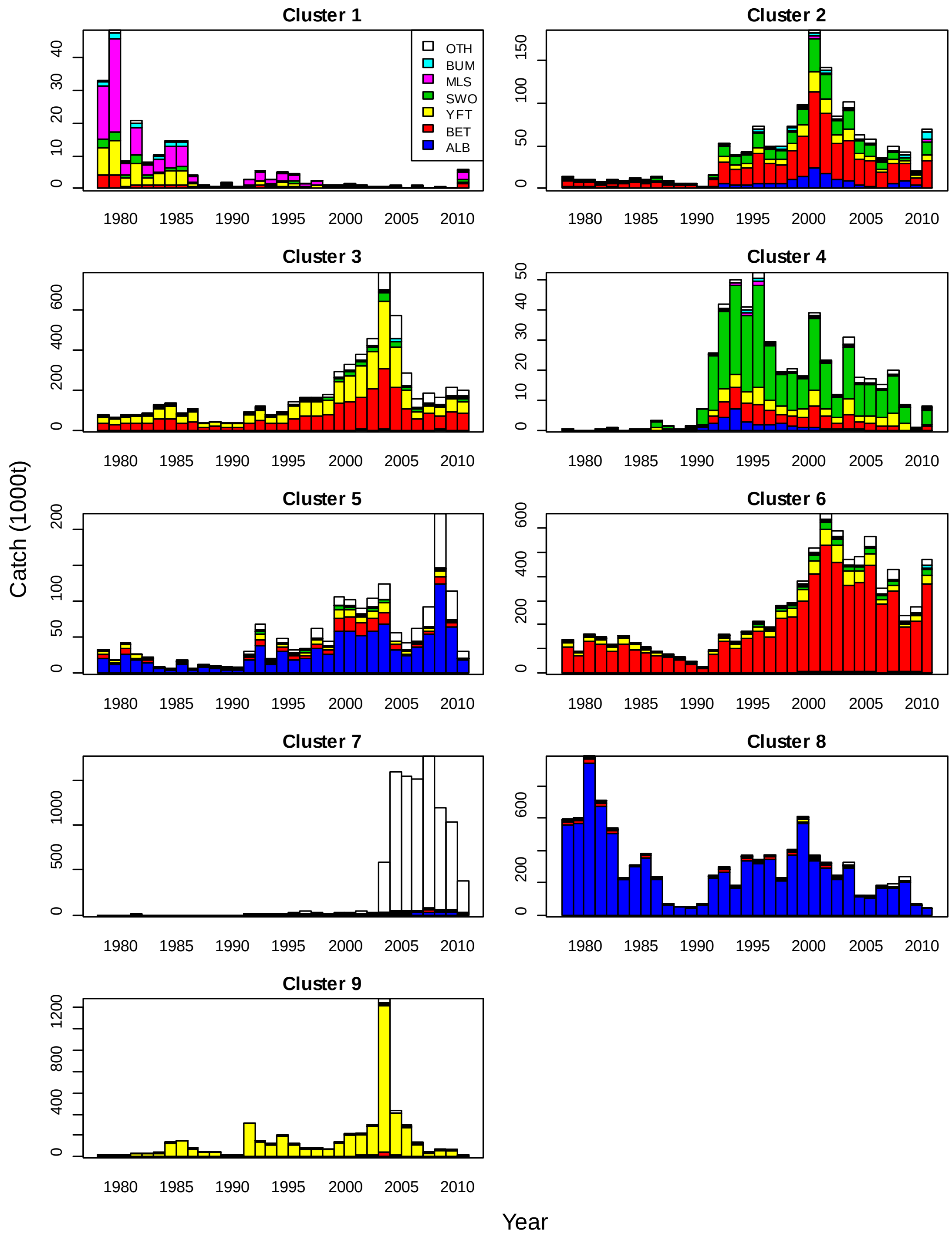

Fig. 4. Annual catches by species of Taiwanese longline fishery in the Indian Ocean for nine clusters.

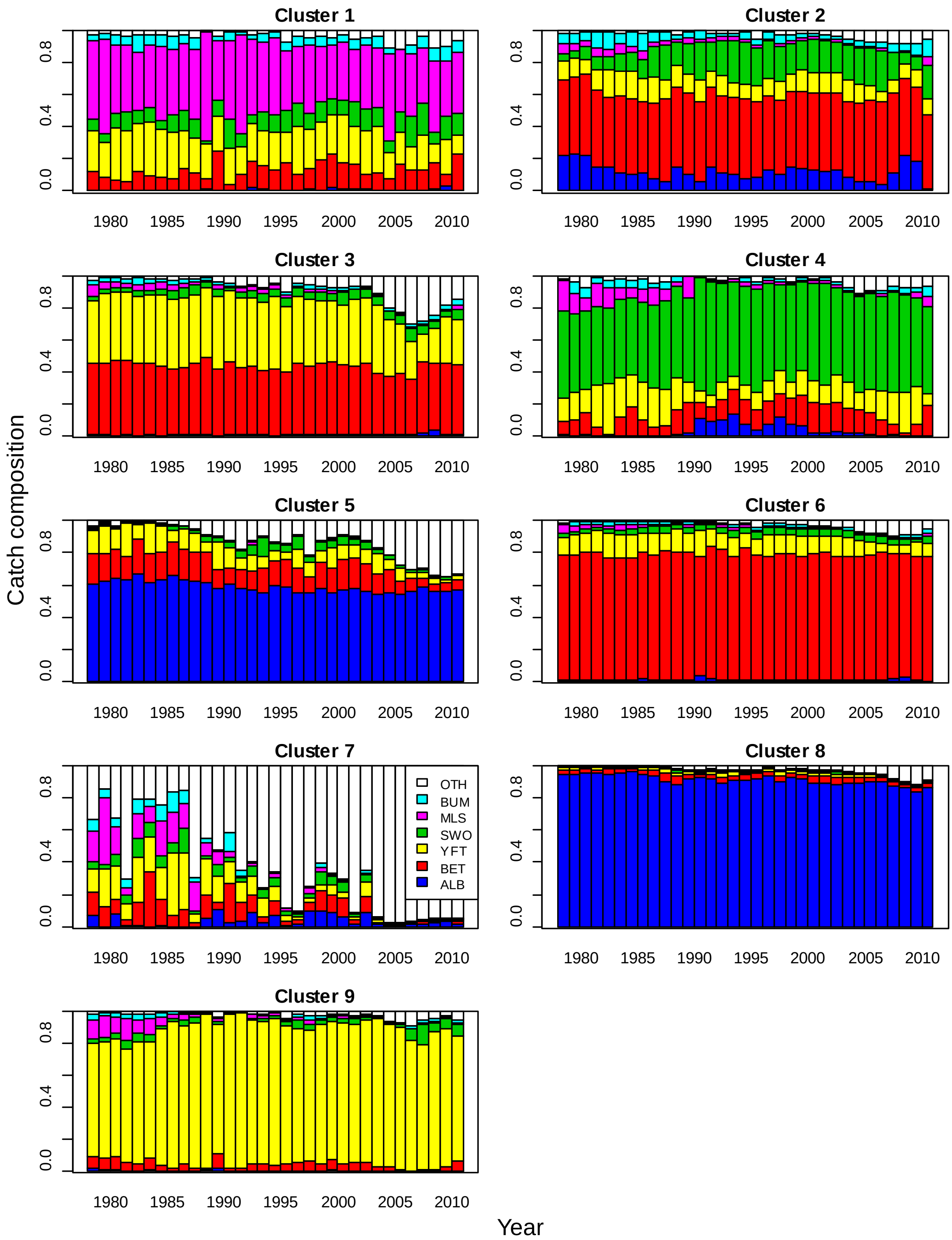

Fig. 5. Annual catch compositions of Taiwanese longline fishery in the Indian Ocean for nine clusters.

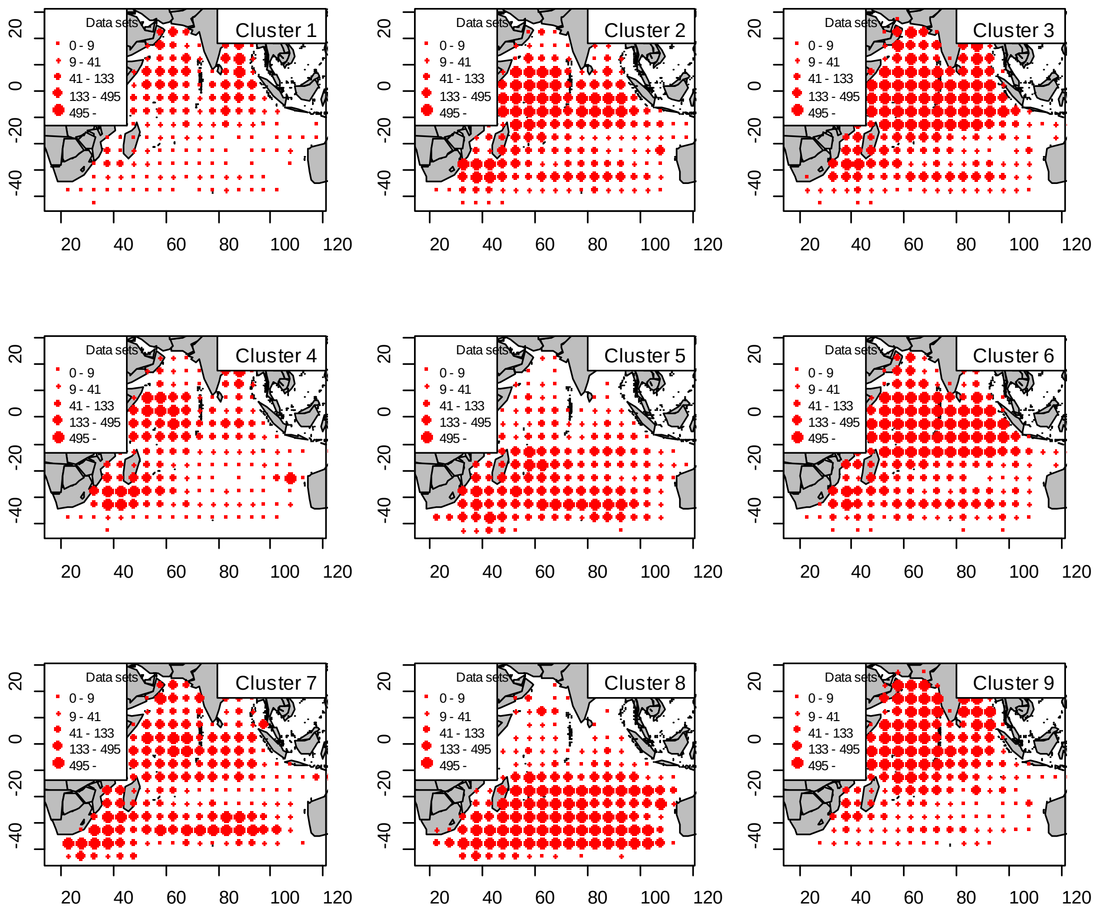

Fig. 6. The distributions for number of data sets of Taiwanese longline fishery in the Indian Ocean for nine clusters.

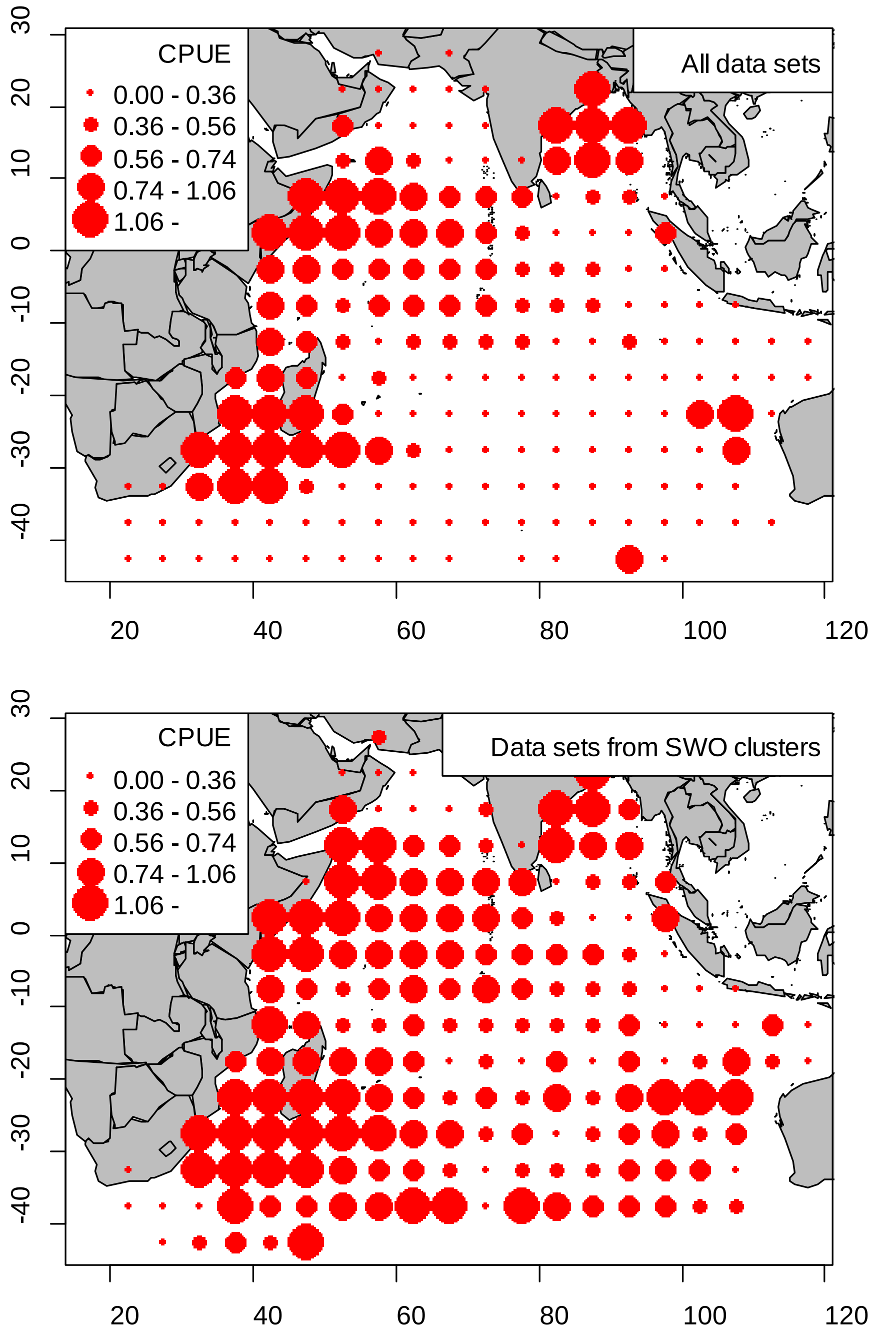

Fig. 7. The distributions of CPUE based on the all data sets and the data sets from SWO clusters.

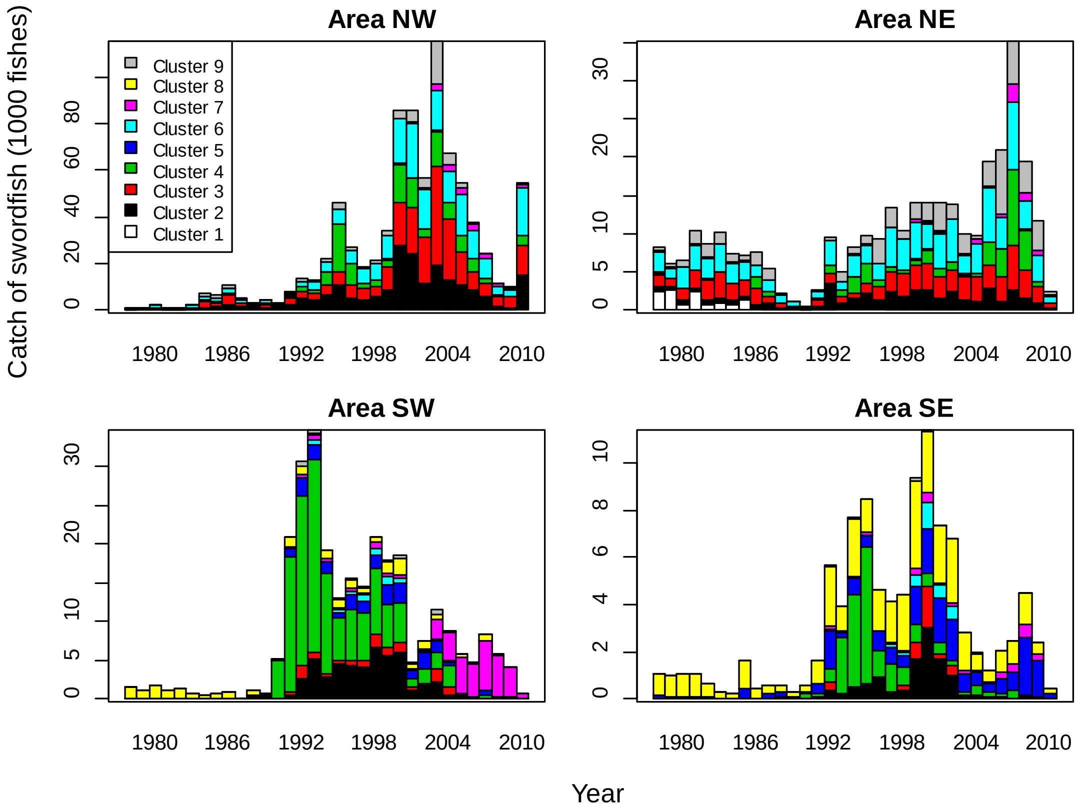

Fig. 8. Annual catches of swordfish by cluster of Taiwanese longline fishery in the Indian Ocean for four areas.

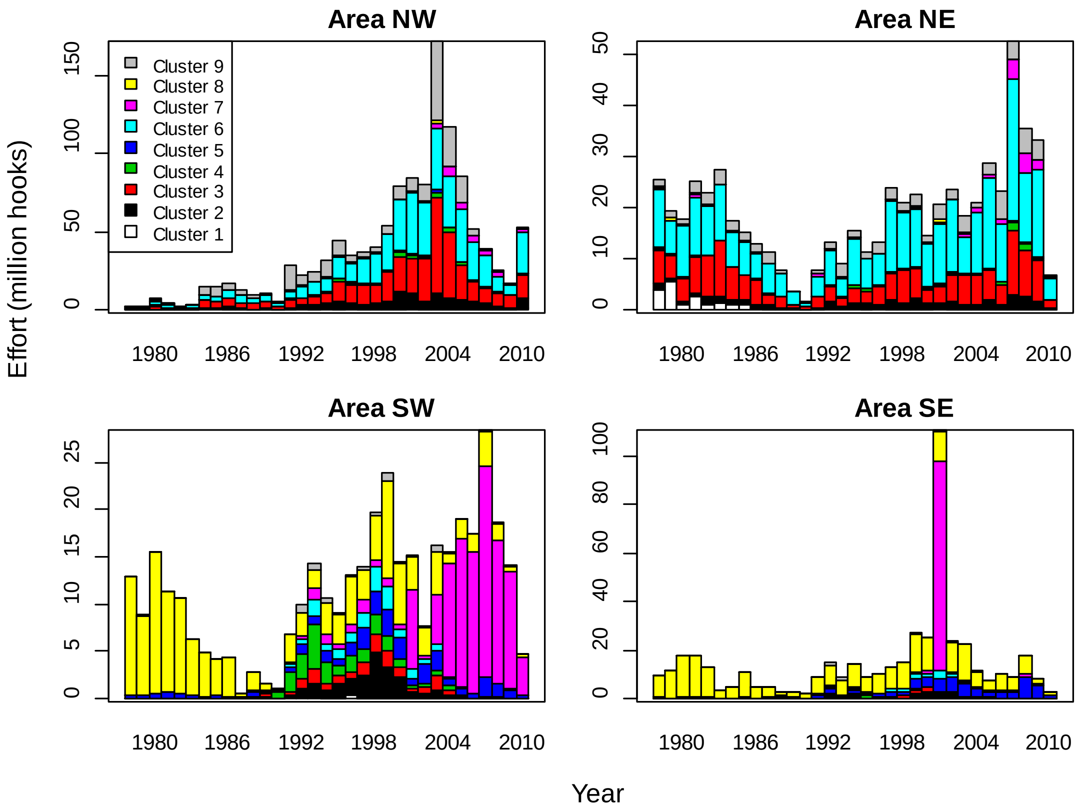

Fig. 9. Annual effort (hooks) by cluster of Taiwanese longline fishery in the Indian Ocean for four areas.

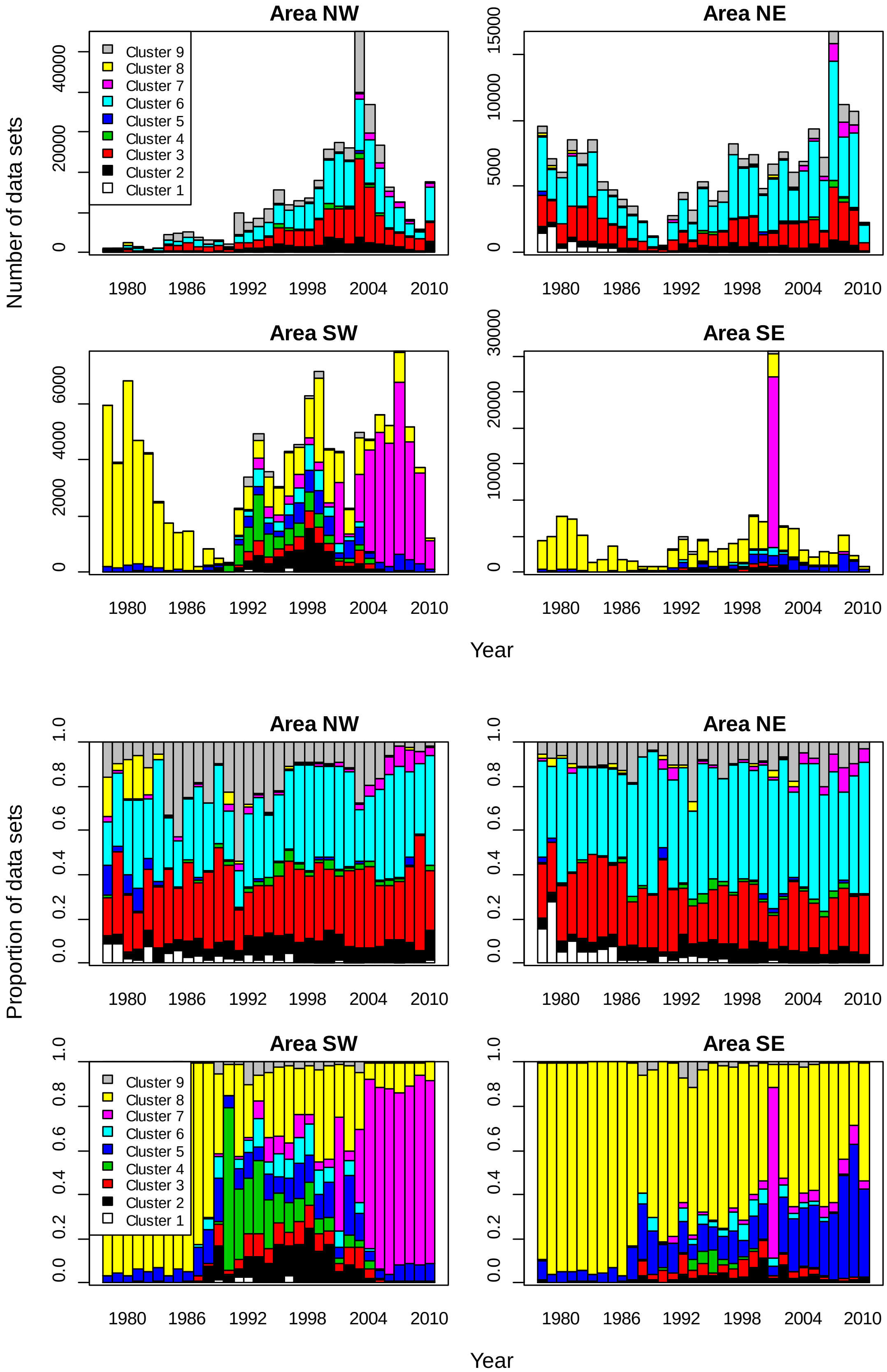

Fig. 10. Annual number of data sets and proportion by cluster of Taiwanese longline fishery in the Indian Ocean for four areas.

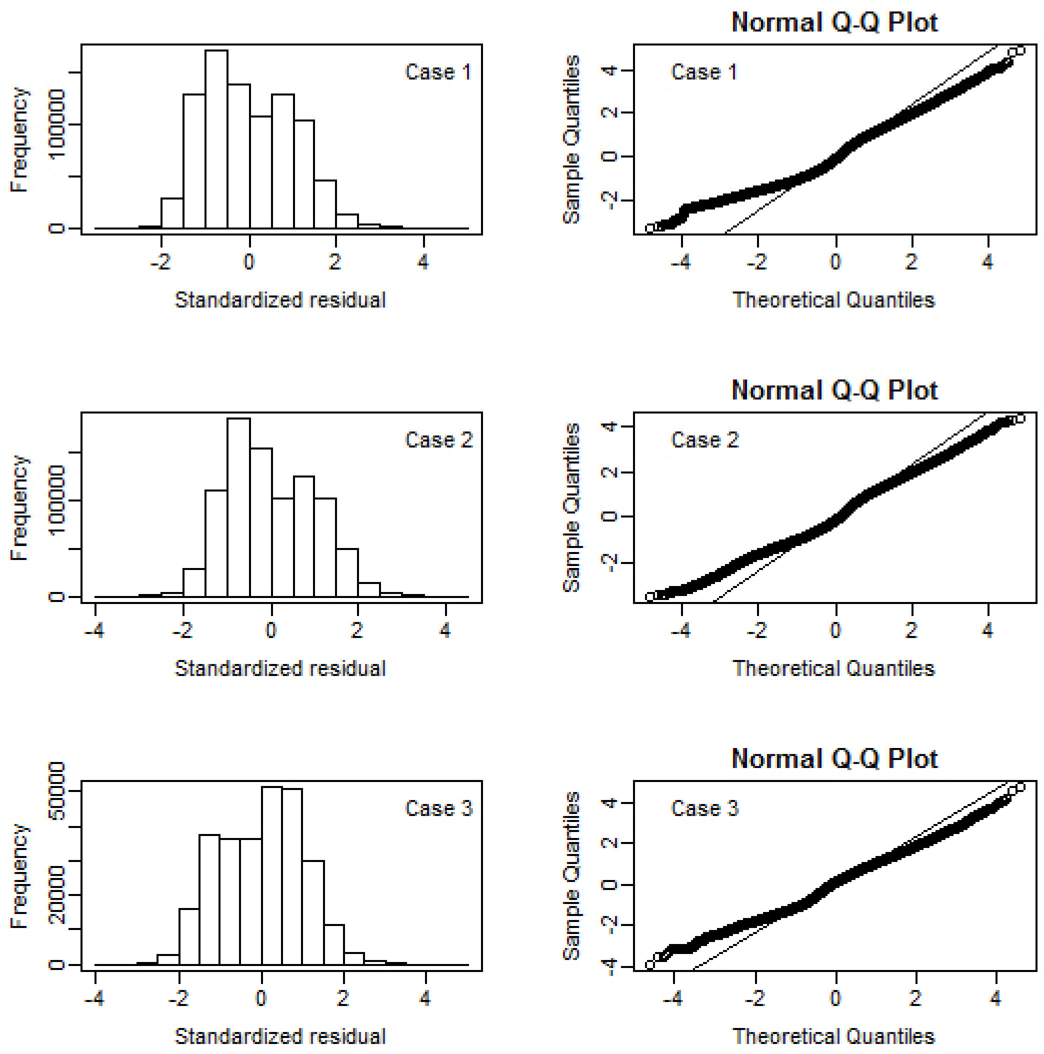

Fig. 11. Distribution of standardized residuals for GLM analyses .

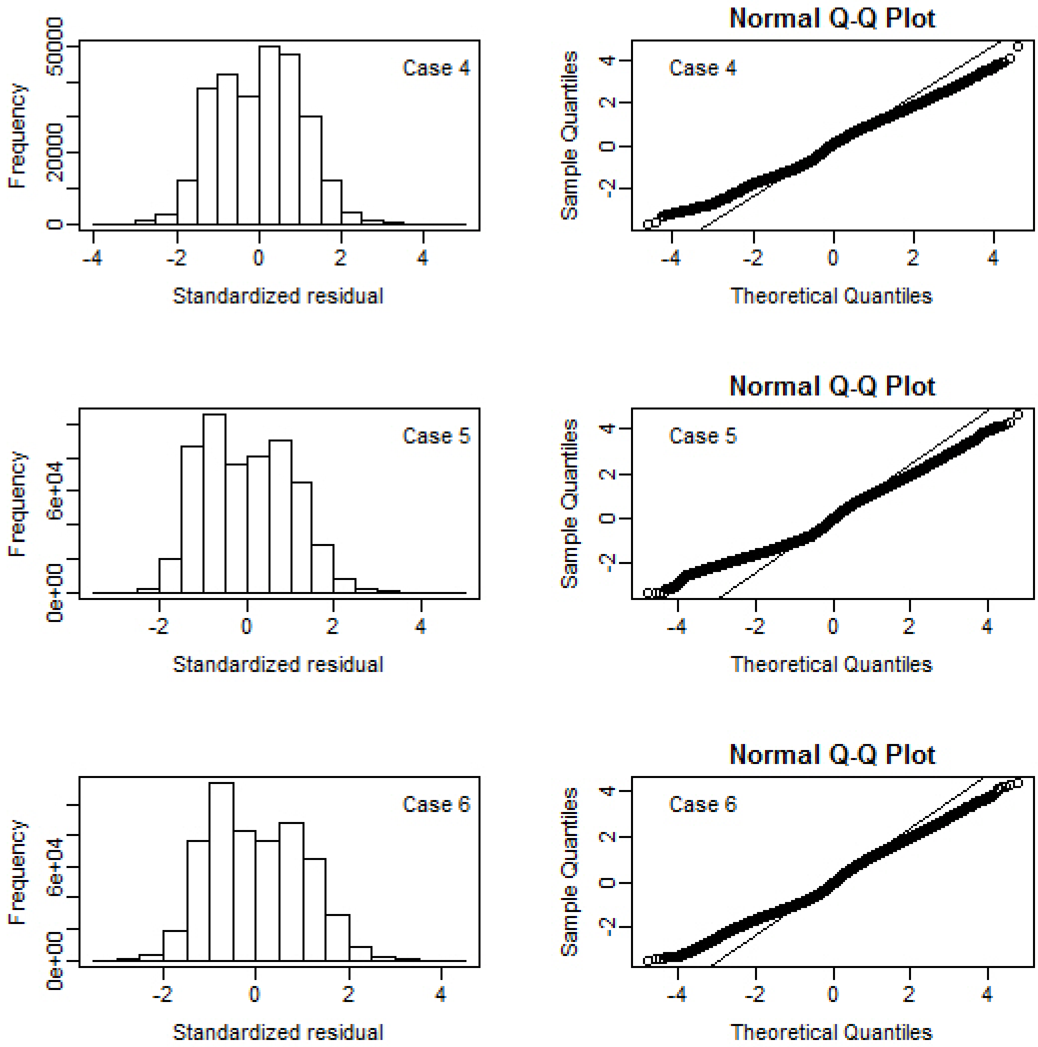

Fig. 11. (continued).

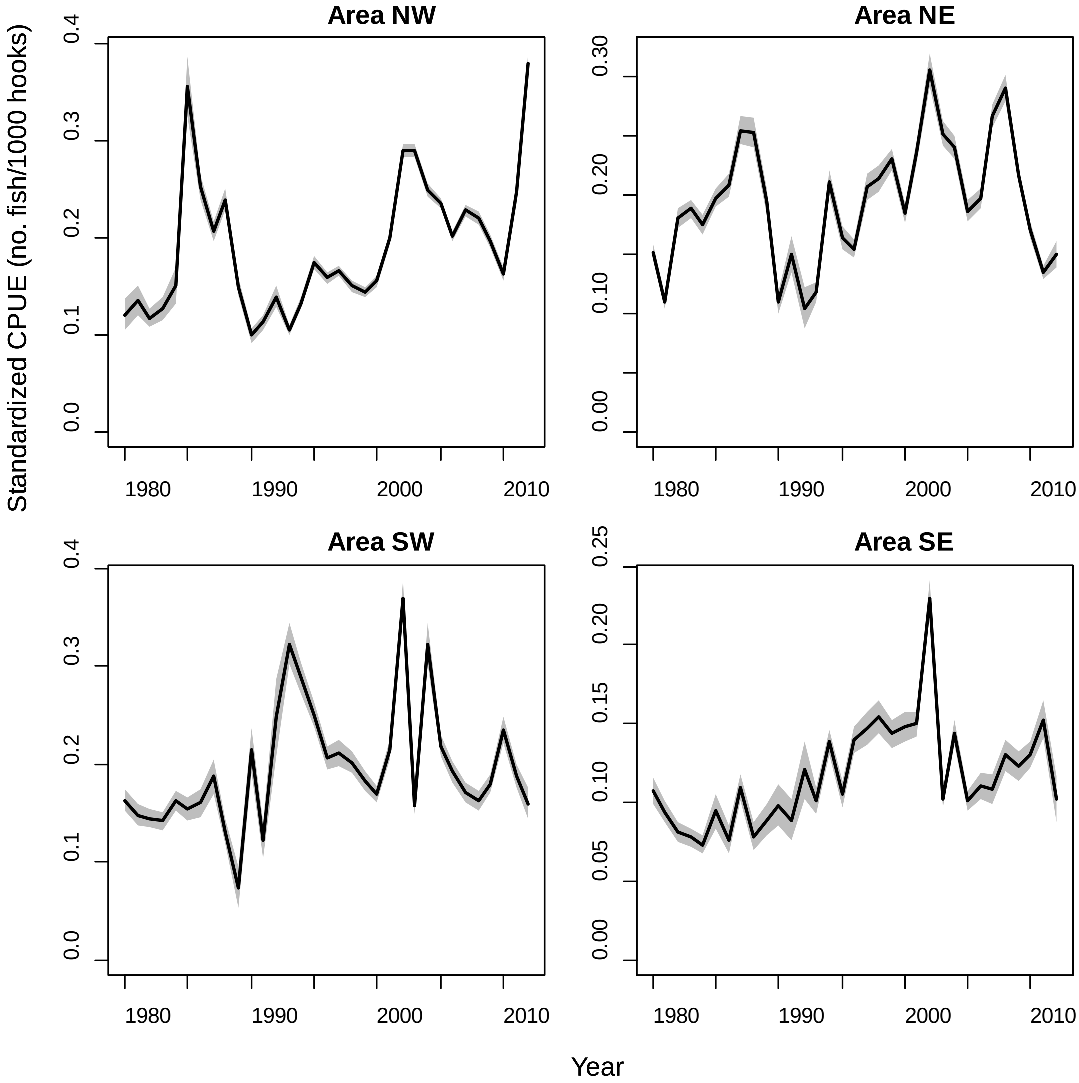

Fig. 12. Area-specific standardized CPUE (lines) with 95% confidence interval (gray shades) for swordfish of Taiwanese longline fishery in the Indian Ocean obtained from Case 2 of GLM analysis.

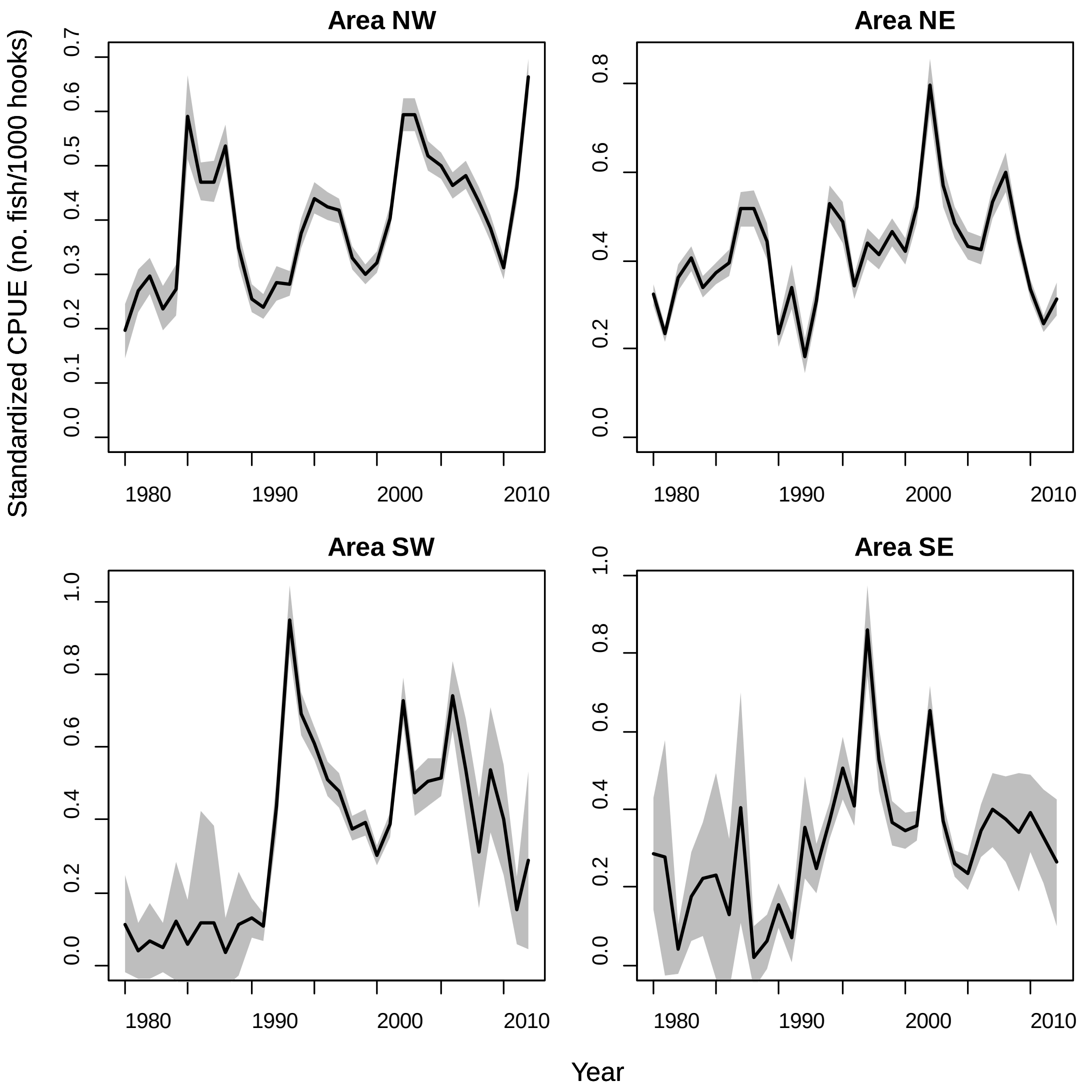

Fig. 13. Area-specific standardized CPUE (lines) with 95% confidence interval (gray shades) for swordfish of Taiwanese longline fishery in the Indian Ocean obtained from Case 4 of GLM analysis.

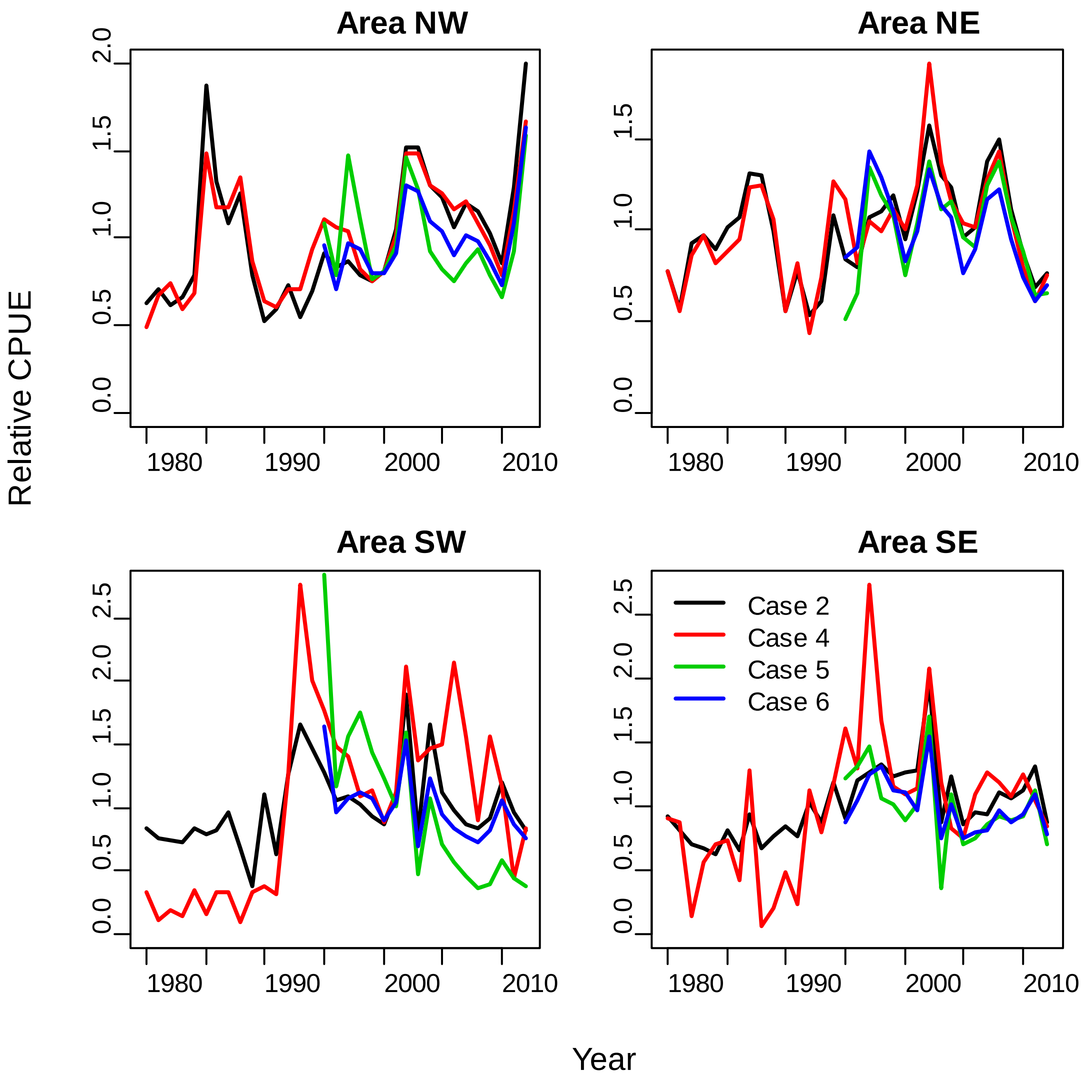

Fig. 14. Area-specific standardized CPUE of swordfish of Taiwanese longline fishery in the Indian Ocean obtained from Cases 2 and 4 of GLM analysis. CPUEs were scaled by the averaged value for each series.

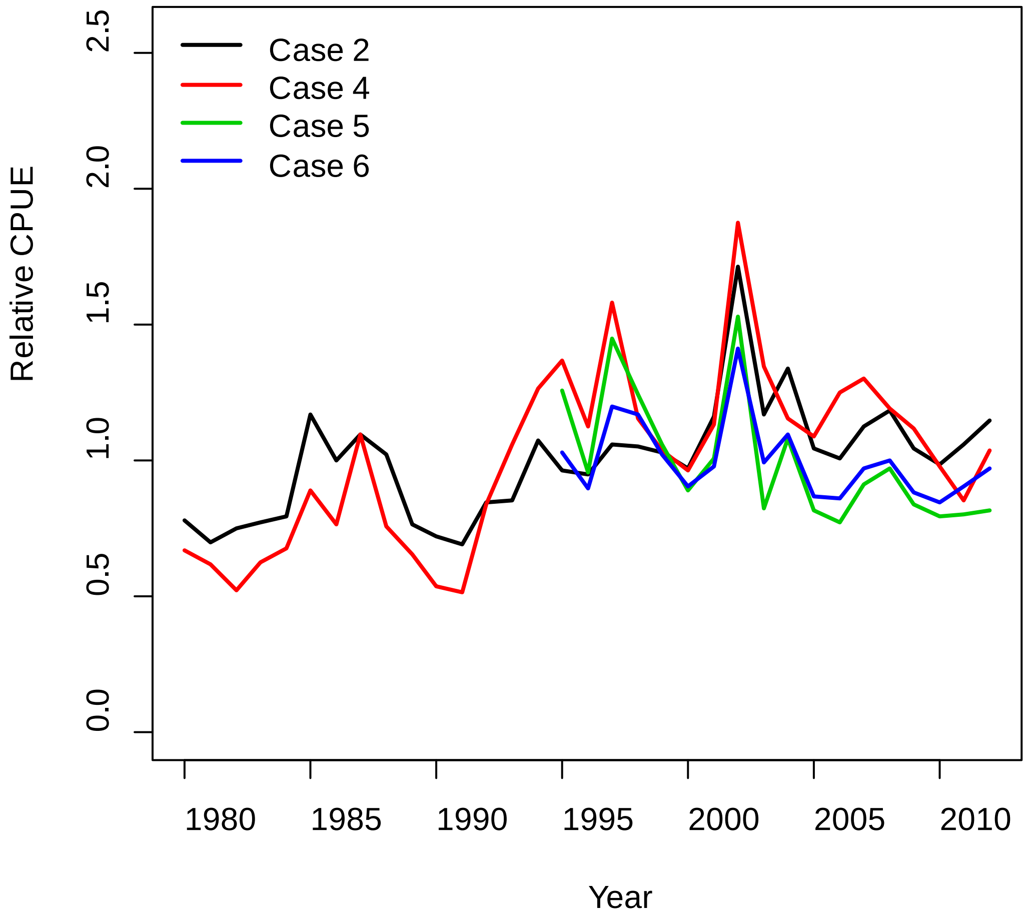

Fig. 15. Area-aggregated standardized CPUE of swordfish of Taiwanese longline fishery in the Indian Ocean obtained from Cases 2 and 4 of GLM analysis. CPUEs were scaled by the averaged value for each series.

Table 1. The average proportions catches of swordfish of Taiwanese longline fishery for 21 non-hierarchical (K-means) clusters.

Cluster	ALB	BET	YFT	SWO	MLS	BUM	OTH
1	0.002	0.080	0.165	0.061	0.346	0.039	0.037
2	0.010	0.483	0.097	0.270	0.017	0.023	0.036
3	0.007	0.317	0.513	0.051	0.016	0.019	0.035
4	0.046	0.128	0.109	0.559	0.018	0.018	0.047
5	0.010	0.121	0.383	0.069	0.019	0.026	0.322
6	0.013	0.396	0.095	0.058	0.011	0.021	0.355
7	0.516	0.057	0.021	0.021	0.003	0.003	0.355
8	0.292	0.436	0.094	0.067	0.009	0.011	0.050
9	0.009	0.633	0.056	0.043	0.014	0.021	0.159
10	0.007	0.278	0.140	0.095	0.045	0.265	0.061
11	0.015	0.047	0.029	0.027	0.006	0.011	0.536
12	0.941	0.018	0.009	0.004	0.001	0.001	0.007
13	0.007	0.496	0.331	0.045	0.017	0.018	0.041
14	0.794	0.059	0.043	0.015	0.004	0.003	0.054
15	0.461	0.093	0.299	0.054	0.008	0.008	0.035
16	0.607	0.226	0.038	0.042	0.006	0.005	0.040
17	0.005	0.843	0.039	0.030	0.007	0.010	0.019
18	0.016	0.011	0.006	0.006	0.001	0.001	0.946
19	0.001	0.010	0.922	0.008	0.007	0.004	0.020
20	0.007	0.677	0.184	0.042	0.013	0.017	0.018
21	0.007	0.081	0.676	0.061	0.034	0.022	0.067

Table 2. The average proportions catches of swordfish and assigned fishing types of Taiwanese longline fishery for 9 hierarchical clusters.

Cluster	ALB	BET	YFT	SWO	MLS	BUM	OTH	Fishing type
1	0.002	0.080	0.165	0.061	0.346	0.039	0.037	MLS+YFT
2	0.074	0.443	0.102	0.198	0.019	0.055	0.043	BET+SWO
3	0.009	0.388	0.334	0.052	0.016	0.020	0.136	BET+YFT
4	0.046	0.128	0.109	0.559	0.018	0.018	0.047	SWO+BET+YFT
5	0.540	0.132	0.090	0.037	0.005	0.005	0.159	ALB+OTH+BET
6	0.007	0.740	0.093	0.037	0.011	0.015	0.049	BET
7	0.015	0.031	0.019	0.018	0.004	0.007	0.722	OTH
8	0.893	0.031	0.020	0.008	0.002	0.002	0.022	ALB
9	0.003	0.041	0.814	0.031	0.019	0.012	0.040	YFT

Table 3. The number of data sets and proportion of data sets of Taiwanese longline fishery for 9 hierarchical clusters.

Cluster	No. data sets	Proportion (%)
1	12,415	1.41
2	57,943	6.58
3	184,557	20.95
4	21,421	2.43
5	34,699	3.94
6	258,188	29.31
7	77,486	8.80
8	144,481	16.40
9	89,678	10.18

Table 4. The numbers of data sets and parameters and values of R2 and AIC for four Cases of GLM analyses.

	No. data sets	No. parameters	R^2	AIC
Case 1	879,369	937	0.20	472,783
Case 2	879,369	1,060	0.33	307,691
Case 3	276,256	823	0.22	89,970
Case 4	276,256	871	0.40	19,938
Case 5	565,600	563	0.20	258,551
Case 6	565,600	636	0.41	158,905

* The values of AIC are only comparable for the cases based on the same data sets.

Table 5. The ANOVA tables for four Cases of GLM analyses.

Case 1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	931	366290.35	393.44	230.06	<.0001
Error	878437	1502239.65	1.71		
Total	879368	1868530.01			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Y	32	14057.00	439.28	256.87	<.0001
V	756	130335.46	172.40	100.81	<.0001
M	11	5951.38	541.03	316.37	<.0001
A	3	6640.04	2213.35	1294.26	<.0001
Y*A	96	42911.51	446.99	261.38	<.0001
M*A	33	10967.41	332.35	194.34	<.0001

Case 2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1051	623775.14	593.51	418.79	<.0001
Error	878317	1244754.87	1.42		
Total	879368	1868530.01			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Y	32	6743.59	210.74	148.70	<.0001
V	756	83463.06	110.40	77.90	<.0001
M	11	904.70	82.25	58.03	<.0001
A	3	1030.70	343.57	242.43	<.0001
C	8	100772.88	12596.61	8888.35	<.0001
Y*A	96	12163.18	126.70	89.40	<.0001
M*A	33	5368.94	162.70	114.80	<.0001
M*C	88	10378.56	117.94	83.22	<.0001
A*C	24	21221.30	884.22	623.92	<.0001

Table 5. (continued).

Case 3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	817	108149.82	132.37	95.87	<.0001
Error	275438	380332.91	1.38		
Total	276255	488482.73			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Y	32	4152.52	129.77	93.98	<.0001
V	642	43500.41	67.76	49.07	<.0001
M	11	700.02	63.64	46.09	<.0001
A	3	462.11	154.04	111.55	<.0001
Y*A	96	6153.02	64.09	46.42	<.0001
M*A	33	4650.32	140.92	102.05	<.0001

Case 4

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	862	193418.26	224.38	209.42	<.0001
Error	275393	295064.47	1.07		
Total	276255	488482.73			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Y	32	1911.09	59.72	55.74	<.0001
V	642	25093.25	39.09	36.48	<.0001
M	11	403.91	36.72	34.27	<.0001
A	3	62.17	20.72	19.34	<.0001
C	3	33810.44	11270.15	10518.80	<.0001
Y*A	96	2842.78	29.61	27.64	<.0001
M*A	33	1619.33	49.07	45.80	<.0001
M*C	33	1760.47	53.35	49.79	<.0001
A*C	9	3083.31	342.59	319.75	<.0001

Case 5. (continued).

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	554	228149.78	411.82	260.98	<.0001
Error	565044	891633.14	1.58		
Total	565598	1119782.92			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Y	17	8110.13	477.07	302.33	<.0001
V	409	71957.64	175.94	111.49	<.0001
M	11	2038.89	185.35	117.46	<.0001
A	3	785.37	261.79	165.9	<.0001
NHBF	2	2449.49	1224.74	776.14	<.0001
Y*A	51	17305.76	339.33	215.04	<.0001
M*A	33	6326.69	191.72	121.5	<.0001
M*NHBF	22	3850.28	175.01	110.91	<.0001
A*NHBF	6	7670.78	1278.46	810.19	<.0001

Case 6

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	644	372414.88	578.28	437.14	<.0001
Error	564954	747368.04	1.32		
Total	565598	1119782.92			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Y	17	3137.39	184.55	139.51	<.0001
V	409	46263.18	113.11	85.50	<.0001
M	11	655.72	59.61	45.06	<.0001
A	3	391.35	130.45	98.61	<.0001
C	8	58207.92	7275.99	5500.10	<.0001
Y*A	51	6170.48	120.99	91.46	<.0001
M*A	33	3405.75	103.20	78.01	<.0001
M*C	88	7111.21	80.81	61.09	<.0001
A*C	24	13441.66	560.07	423.37	<.0001