IOTC-2008-WPTT-31

CPUE Standardizations for Yellowfin Tuna Caught by Taiwanese Deep Sea Longline Fishery in the Tropical Indian Ocean Using Generalized Linear Model and Generalized Linear Mixed Model

Shui-Kai Chang[1], Hung-I Liu[2], and Shu-Ting Chang[2]

[1] Institute of Marine Affairs, National Sun Yat-Sen Univeristy, 70, Lienhai Rd., Kaohsiung 80424, Taiwan
[2] Overseas Fisheries Development Council, 19, Sec. 3, Roosevelt Rd, Taipei 100, Taiwan

INTRODUCTION

Yellowfin tuna (YFT) is one of the most important target species for Taiwanese far seas tuna longline fishery operating in the Indian Ocean. The catches of YFT were lower than 20,000 mt before late 1980s and thereafter substantially increased along the increase of bigeye-targeting activities. In this period, the YFT has been bumped up to about 80,000 mt around 1993 and roughly ten years later to about 60,000 mt around 2005. Most of the catches in these two years were made in the waters off Oman.

For stock assessment purpose, the standardizations of CPUE for YFT caught by Taiwanese longline fishery in the Indian Ocean were conducted by generalized linear model (GLM) and generalized linear mixed model (GLMM), on the set by set logbooks data. The logbook data are available only since 1979, and to avoid the effects of complicated regime shifts of high catches in Oman waters and to be in comparable with Japanese standardization CPUE series, the standardizations on Taiwanese series were conducted only for years of 1979-2007 (2006 and 2007 data are still preliminary) and for tropical areas.

MATERIALS AND METHODS

Data analyzing

In a beforehand study that an updated GLM run similar to Run 1 of Liu et al. (2007) was conducted this year with revised 2005 (due to increased data coverage) and additional 2006 data, and focused only on Areas 2 & 5. Fig. 1 shows residual distribution of the run. Obviously, it does not meet the normal distribution assumption because of an additional small mode occurred. We refer the large mode data as Group A data and the small mode data as Group B. A further exploration on the small mode Group B data indicating a very high bigeye tuna (BET) catch composition (Fig. 2), comparing to Group A about 30% only. The high catch composition in Group B was resulted from combination of data with BET>75% and data without any information on YFT or ALB. Excluding the data with incomplete information on YFT or ALB, it might be interpreted that this small mode was from specific BET-targeting fishing activities. It was also noted that most of the Group B data come from years after 1990 when Taiwan started to have more deep set longline activities for bigeye tuna. A preliminary examination on the observer data during 2002-2007 shows that the average bigeye catches composition in BET-targeting vessels operating in the tropical area was 77%. We therefore consider it plausible to use 75% as an *ad hoc* threshold for excluding BET-targeting data for current YFT CPUE standardizations. Based on this analysis, this study applied GLM and GLMM on 1979-2007 logbook data with BET catch composition less than 75% in the tropical area (Area 2 and Area 5 of the new area definition, Fig. 3). The data with incomplete information on YFT (YFT=0) were also removed from the dataset.

Statistical Models

Statistical models of GLM and GLMM were used to model the logarithm of the nominal CPUE (defined as the number of fish per 1,000 hooks) in the report. The main effects considered in this analysis are year, season (Jan.-Mar., Apr.-Jun., Jul.-Sep., and Oct.-Dec.), area (Areas 2 and 5), and target. SST (five categories separated by 15, 20, 25, and 30°C) only considered in one of the GLMM runs. Interactions between the main factors are also included into the model.

The information of number of hooks between floats or hooks per basket (HPB) is usually used as target indicator in the CPUE standardization models. However, this information was only available from 1995 onwards in the logbooks of Taiwanese longline fishery. Alternative indicators were therefore developed in the study: (1) catch ratio of YFT to albacore (ALB), YFT and BET (referred as YFT%), (2) catches of ALB and BET (referred as ALB/BET), and (3) catch ratios of ALB and BET (referred as ALB%/BET%). Four quartile levels were defined into each factor except for ALB due to that tropical area is not ALB fishing ground and therefore only two levels were defined.

Sea surface temperature (SST) has been incorporated into a GLMM run from 1979-2006. This information is kindly provided by Hiroaki Okamoto of National Research Institute of Far Seas Fisheries of Japan. These data were downloaded from NEAR-GOOS Regional Real Time Data Base of Japan Meteorological Agency and JEDEC (Joint Environmental Data Analysis Center) website of Scripps Institution of Oceanography, respectively. The procedure of this data processing is described in Okamoto et al. (2001).

(1) GLM model: The CPUE is predicted as a linear combination of the explanatory variables. At first, the following form was assumed as a full model.

$$\log(CPUE + c) = \mu + Y + S + A + T + Interactions + \varepsilon$$

where *CPUE* is the nominal CPUE of yellowfin tuna,

c is the constant value (i.e. 0.1),

μ is the intercept,

Y is the effect of year,

S is the effect of season,

A is the effect of fishing area,

T is the effect related to the catch ratio of yellowfin or albacore and bigeye tuna,

Interactions is the interactions between main effects,

ε is the error term, $\varepsilon \sim N(0, \sigma^2)$.

(2) GLMM model: This model assumes a delta lognormal error distribution for the positive catch

rates. The model fits separately the proportion of positive sets assuming a binomial error distribution, and the mean catch rate of positive sets (at least one fish was caught) assuming a lognormal error distribution.

Estimated proportion of successful sets is assumed to be the result of r positive sets of a total n number of sets, and each one is an independent Bernoulli-type realization. The estimated proportion (ρ) is a linear function of fixed effects and interactions, by using logit function as a link between linear factor components and binomial errors. The systematic component is defined as:

$$\log\left(\frac{\rho_{Y,S,A,T,SST}}{1-\rho_{Y,S,A,T,SST}}\right)=\alpha_0+\alpha_Y+\alpha_S+\alpha_A+\alpha_T+\alpha_{SST}+\alpha_{Interactions}+\omega$$

with a binomial density:

$$\omega_{Y,S,A,T,SST} \sim Bin(n_{Y,S,A,T,SST},\rho_{Y,S,A,T,SST})$$

For positive observations, which were defined as at least one yellowfin caught, the estimated CPUE rate was assumed to follow a lognormal error distribution (logCPUE) of a linear function of fixed factors and random effect interactions. The systematic component is defined as:

$$\log(\mu_{Y,S,A,ALB,BET,SST}) = \beta_0 + \beta_Y + \beta_S + \beta_A + \beta_T + \beta_{SST} + \beta_{Interactions} + z$$

with a log-normal density:

$$z_{Y,S,A,T,SST} \sim LogNorm(\mu_{Y,S,A,T,SST},\sigma^2_{Y,S,A,T,SST})$$

Standardization runs

This study has conducted a set of standardization runs using logbook data, by both GLMM and GLM model (Table 1). All runs only keep significant factors ($p<0.01$) in the analysis of CPUE by the effective effort. The calculation was done using GLM, GLIMMIX and MIXED procedure of SAS (Ver. 8. 02). The standardized CPUE were then computed from the least square means (LSMaens) of the estimates of the year effects.

RESULTS AND DISCUSSION

Since HPB information was available only since 1995 and hence alternatives were developed to represent target factor for the whole series, we conducted a beforehand GLM analysis on data with HPB since 1995. We performed three GLM runs on the data with same model structure except the indicator of target factor: HPB, YFT% and ALB%/BET%. The result shows there is not obvious difference in standardized CPUE trend (Fig. 4) between HPB case and YFT% case and the trend much more different for HPB and ALB%/BET% case. Therefore, YFT% model structure was used as the base cases and ALB%/BET% model structure was used as a sensitivity run.

Table 1 shows R^2 of all runs. The model of all runs explained less than 50% of the variance, and SenRun_1 was the highest, which explained about 49% of the variance. Table 2-5 shows the ANOVA table of the final models of all runs with information of the highly statistically significant ($p < 0.01$) factors.

Distributions of the standardized residuals for the four runs are shown in Fig. 5. The distribution of the standardized residuals for SenRun_2 and SenRun_3 appear to deviate slightly from normal distribution assumption; on the other hand, base and SenRun_1 appear to meet normal distribution assumption. The normal probability plots are showed in Fig. 6. SenRun_2 show heavy divergences for right tail, however standardized residuals of other of the runs conform roughly to the normal distribution.

Relative standardized CPUEs obtained from all runs are shown in Fig. 7. Relative values are scaled to the average of the estimates and all the series look similar in trend. The four relative CPUEs series were compared with Japanese CPUE series (re-drawn from Okamoto et al., 2008) separately in Fig. 8. The Japanese standardize CPUE was calculated from new Areas 2, 3 and 5 (Japanese tropical fishing ground), and HPB information were used to express the effects of targeting in the GLM model. There are three peaks in all of the Taiwanese series (1986-88, 1992-94, and 2003-05). The first can also be seen in the Japanese series. The other two peaks which shown different patterns with Japanese series are generally corresponding to the two high catch periods described in the Introduction section. Among the four runs, the SenRun3 shows the most similarity to Japanese series. However, due to the different targeting feature and selectivity of the two fleets, it is recommended to apply all the four standardization results to the stock assessment model runs and conduct detail examinations on their performances.

We have used 75% as an *ad hoc* threshold to exclude the BET-targeting data for current YFT CPUE standardizations. Although by doing this there are risks of excluding some informative records of non-BET-targeting activities with incidental high percentage of BET catch or some records from a location with low abundance of bigeye and yellowfin, we consider this practice as a starting point to obtain better (no necessary the best) representative standardized series (comparing to the previous estimated series). The improvements could be noted from the error distributions of Figs. 1 and 5. For future analyses, it would be better to identify and exclude the BET-targeting data on trip (or vessel) basis, rather than on set by set basis to avoid the abovementioned risks. It may also consider not to exclude the BET-targeting data but to include the data with a specific code to retain more information in the model.

Two additional tests on the YFT CPUE standardizations showed that if the standardization was made only on a subset data that have bigeye catch composition > 75%, we would have very similar trend with Japanese series from 1990 onwards; and, that if Japanese series used the same indicator as the base case of this study (catch ratio of YFT to YFT+ALB+BET), then we would also have very similar series for both fleets, especially for the latest five years. These phenomena implied that the two series might not be totally different but there might have some components in the dataset or standardization procedure causing contaminations. This may worth further investigations in the future.

REFERENCES

Okamoto, H., N. Miyabe and T. Matsumoto. 2001. GLM analyses for standardization of Japanese longline cpue for bigeye tuna in the Indian Ocean applying environmental factors. *IOTC Proceedings* 4: 491-522.

Liu, H.I., S.T. Chang and S.K. Chang. 2007. Catch rate standardization runs for Yellowfin tuna caught by Taiwanese deep sea longline fishery in the Indian ocean using generalized linear model and generalized linear mixed model. *IOTC-2007-WPTT-20.*

Table 1. Standardization runs conducted in this report on tropical Indian Ocean yellowfin tuna CPUE of Taiwanese longline fishery.

Runs	MODEL	Target factor	SST	Data period	R^2
base	GLM	YFT%	No	1979~2007	0.2896
SenRun_1	GLMM	YFT%	No	1980~2007	0.4859
SenRun_2	GLMM	ALB% & BET%	No	1980~2007	0.0716
SenRun_3	GLMM	ALB & BET	SST	1980~2006	0.2981

Table 2. ANOVA table of the selected model for base run.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	153	40776.9997	266.5163	593.62	<.0001
Error	222781	100021.4906	0.449		
Corrected Total	222934	140798.4903			

R-Square	Coeff Var	Root MSE	lnyftcpue Mean
0.28961	47.98701	0.670051	1.396317

Source	DF	Type III SS	Mean Square	F-value	P-value
Y	28	1339.737294	47.847761	106.57	<.0001
S	3	130.024265	43.341422	96.54	<.0001
A	1	267.30421	267.30421	595.38	<.0001
YFT%	3	9070.394952	3023.464984	6734.26	<.0001
Y*S	84	2701.412712	32.159675	71.63	<.0001
A*Y	28	465.526948	16.625962	37.03	<.0001
A*S	3	69.998827	23.332942	51.97	<.0001
A*YFT%	3	595.056195	198.352065	441.8	<.0001

Table 3. ANOVA table of the selected model for SenRun_1.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	142	193600.1804	1363.3816	2311.59	<.0001
Error	347281	204827.2496	0.5898		
Corrected Total	347423	398427.43			

R-Square	Coeff Var	Root MSE	lnyftcpue Mean
0.485911	104.1616	0.767986	0.737303

Source	DF	Type III SS	Mean Square	F-value	P-value
Y	27	8723.1294	323.0789	547.77	<.0001
S	3	367.6459	122.5486	207.78	<.0001
A	1	364.2188	364.2188	617.53	<.0001
YFT%	3	159822.251	53274.0837	90325.3	<.0001
Y*A	27	665.6602	24.6541	41.8	<.0001
Y*S	81	4606.1381	56.8659	96.42	<.0001

Table 4. ANOVA table of the selected model for SenRun_2.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	69	28545.1951	413.6985	388.5	<.0001
Error	347354	369882.2349	1.0649		
Corrected Total	347423	398427.43			

R-Square	Coeff Var	Root MSE	lnyftcpue Mean
0.071645	139.9587	1.031919	0.737303

Source	DF	Type III SS	Mean Square	F-value	P-value
Y	27	9946.328231	368.382527	345.95	<.0001
S	3	136.989937	45.663312	42.88	<.0001
A	1	2901.625228	2901.625228	2724.9	<.0001
ALB	1	125.020968	125.020968	117.41	<.0001
BET	3	1043.312817	347.770939	326.59	<.0001
Y*A	27	2275.951763	84.29451	79.16	<.0001
A* ALB	1	65.396791	65.396791	61.41	<.0001
A*BET	3	445.362969	148.454323	139.41	<.0001
ALB*BET	3	66.658416	22.219472	20.87	<.0001

Table 5. ANOVA table of the selected model for SenRun_3.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	143	109100.7598	762.9424	951.72	<.0001
Error	320423	256866.2075	0.8016		
Corrected Total	320566	365966.9673			

R-Square	Coeff Var	Root MSE	lnyftcpue Mean
0.298116	117.5596	0.895348	0.761611

Source	DF	Type III SS	Mean Square	F-value	P-value
Y	26	3928.15917	151.08305	188.47	<.0001
S	3	349.5392	116.51307	145.34	<.0001
A	1	1408.39129	1408.39129	1756.87	<.0001
ALB%	1	657.91277	657.91277	820.7	<.0001
BET%	3	52774.75906	17591.58635	21944.3	<.0001
SST	1	108.44189	108.44189	135.27	<.0001
Y*S	26	1183.19708	45.50758	56.77	<.0001
Y*A	78	6190.6257	79.367	99	<.0001
A*BET%	3	449.80957	149.93652	187.04	<.0001
BET%*ALB%	1	21.56318	21.56318	26.9	<.0001

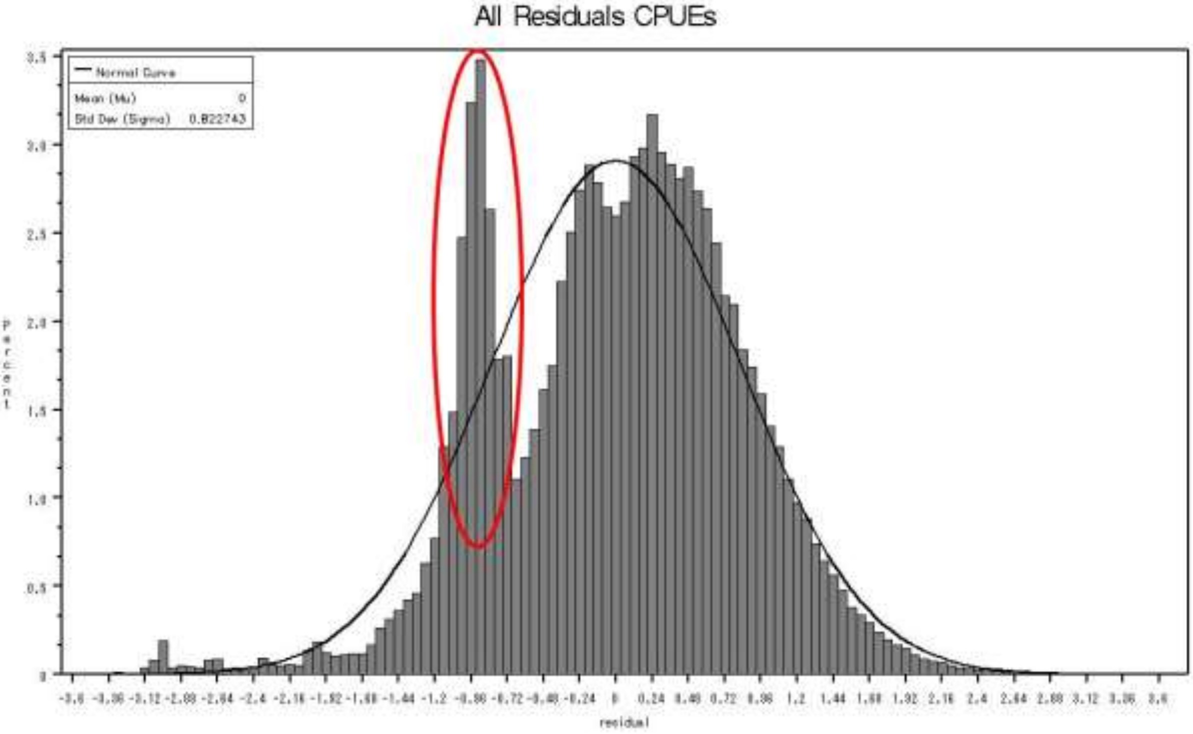


Fig. 1. The residuals distribution result that the beforehand study run carries out appears. The red oval part is Group B.

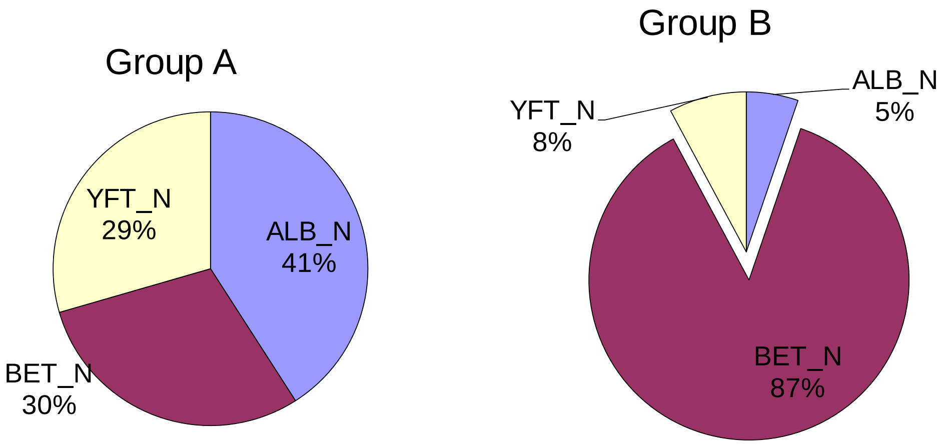


Fig. 2. A further exploration on the small mode Group B data indicating a high BET catch composition (BET/(ALB+YFT+BET)), comparing to Group A. The high catch composition in Group B was resulted from combination of data with BET>75% and data without information on YFT or ALB.

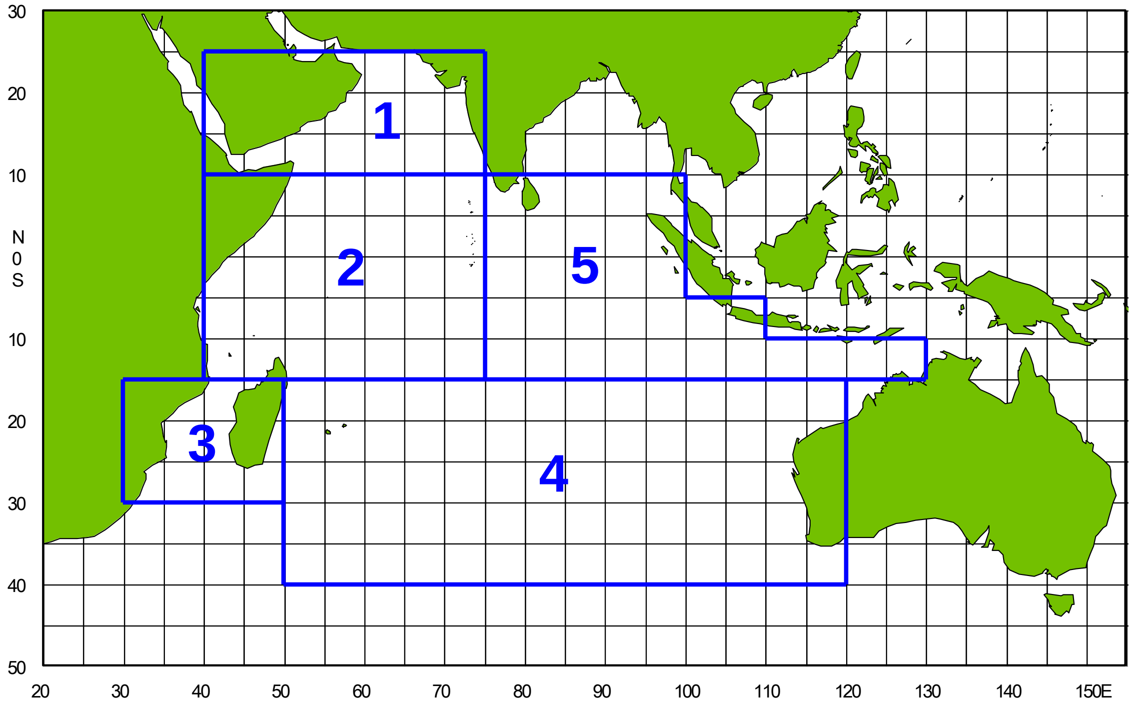


Fig. 3. New area stratification used for the standardization of CPUE for yellowfin tuna in the Indian Ocean in 2008.

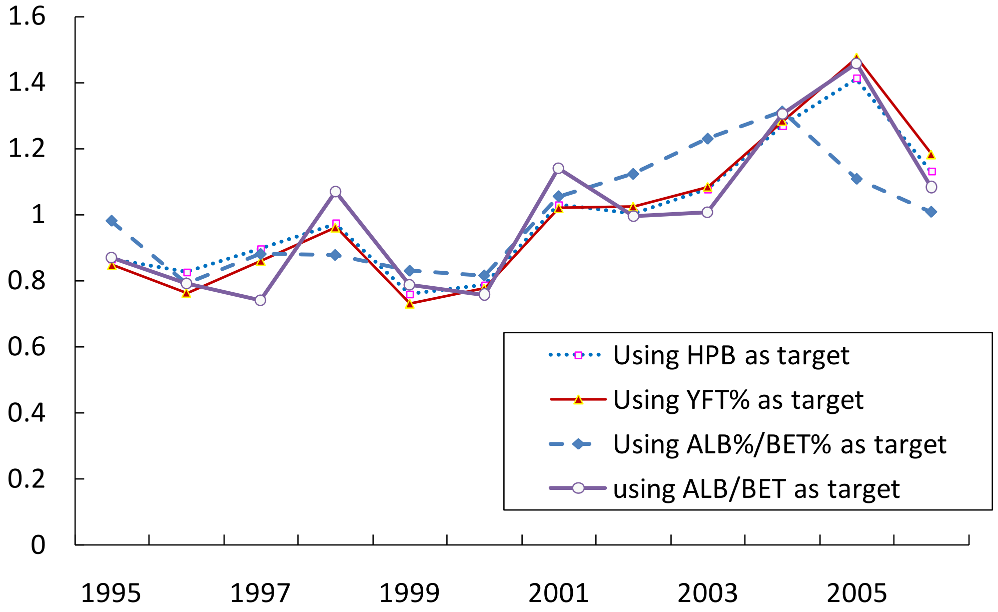
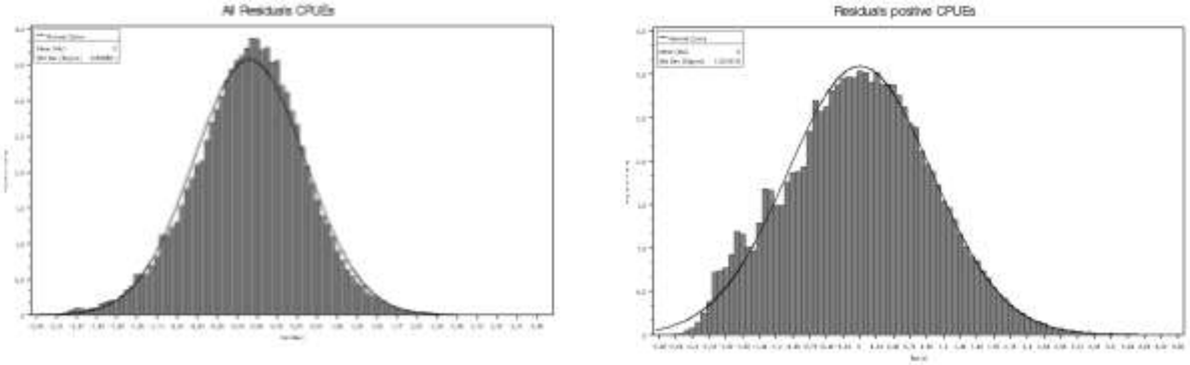


Fig. 4. Standardized CPUE series of NHBF or species composition, there is not obvious difference trend.

base

SenRun 2



SenRun 1

SenRun 3

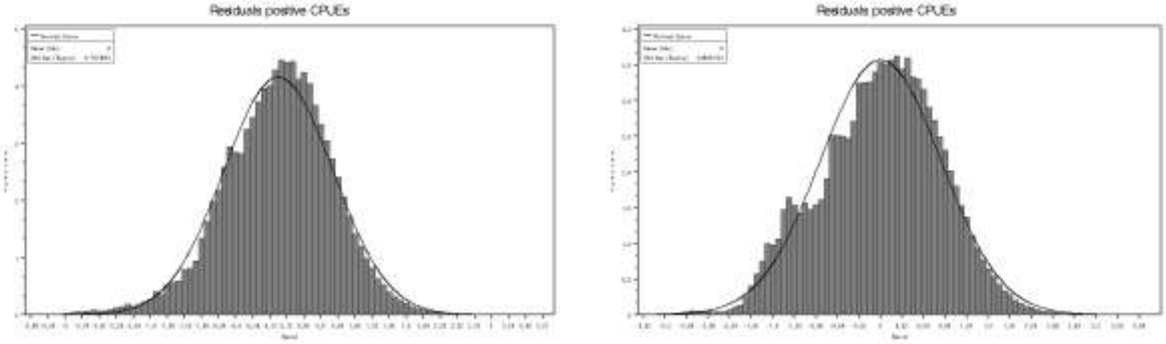


Fig. 5. Distributions of the standardized residuals for the standardization models fitted to the catch and effort data.

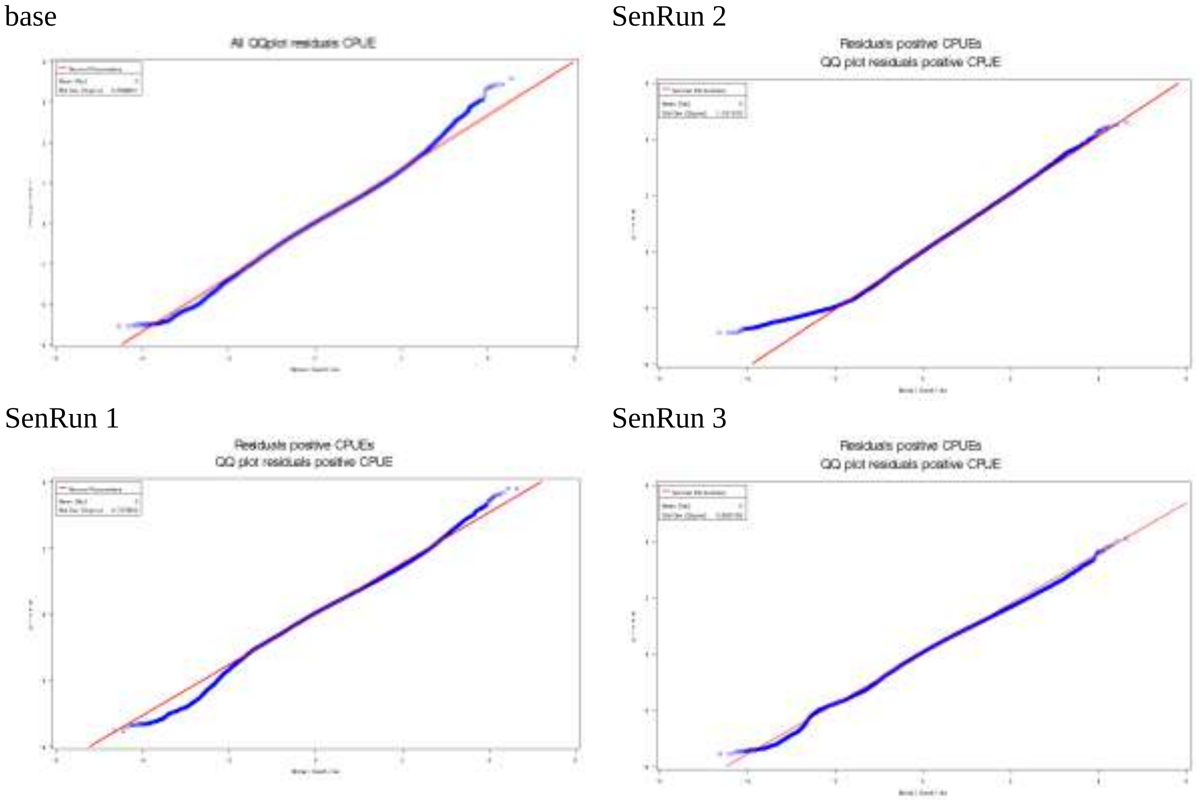


Fig. 6. The normal probability plots for the standardization models fitted to the catch and effort data.

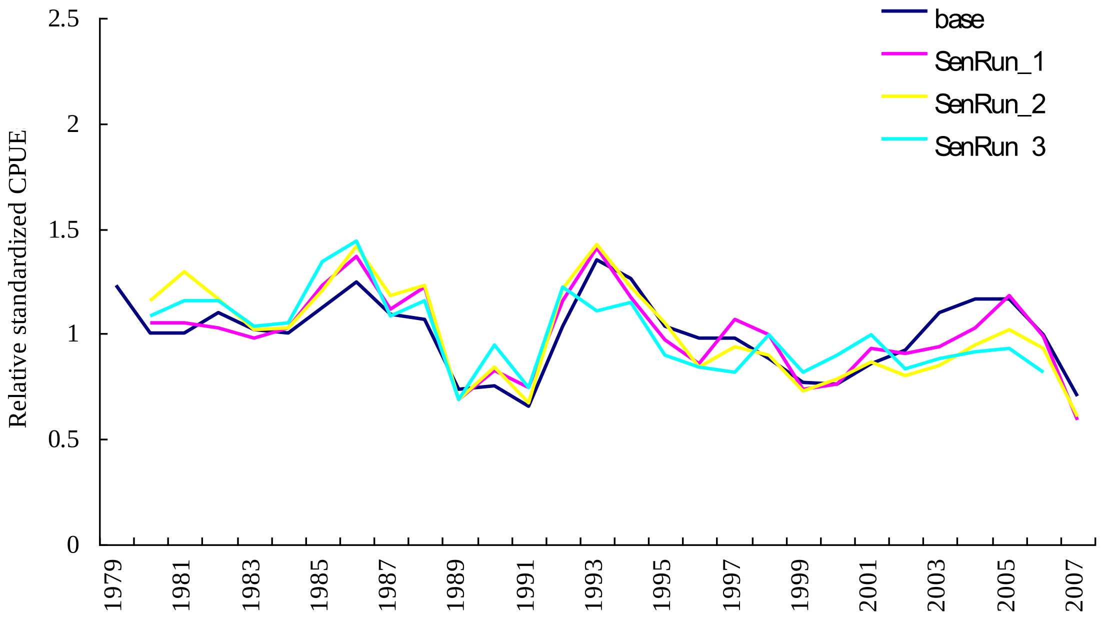


Fig. 7. Trends of relative standardized CPUE of yellowfin tuna caught by Taiwanese longline fishery in the Indian Ocean from four runs.

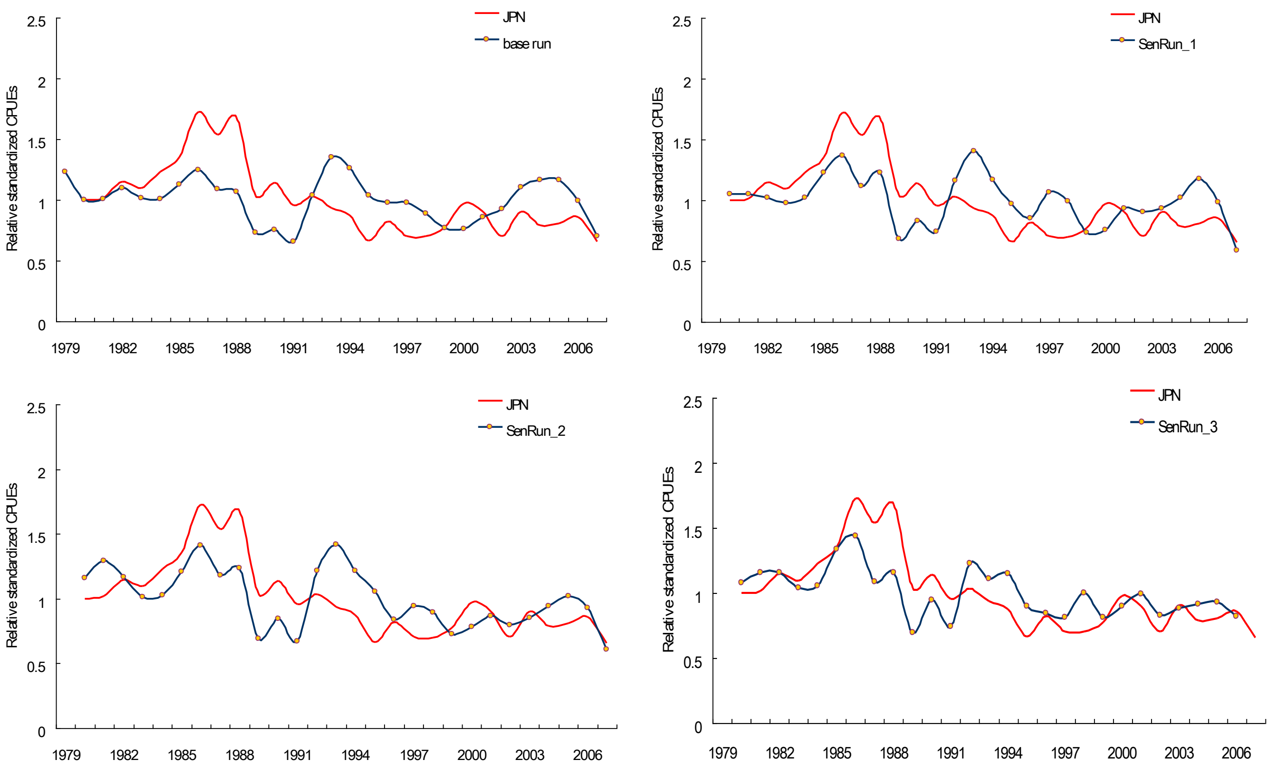


Fig. 8. Relative standardized CPUE of Taiwanese from four runs is compared with Japanese CPUE series separately.