Received: 5 September 2011



# Standardization of albacore CPUE by Japanese longline fishery in the Indian Ocean

Takayuki Matsumoto and Koji Uosaki

National Research Institute of Far Seas Fisheries
7-1, Orido 5-chome, Shimizu, Shizuoka, 424-8633 Japan

## 1. INTRODUCTION

The albacore in the Indian Ocean has been exploited since the early 1950s. The albacore catch has been increasing with fluctuation, and it reached about 48,000 t in 2008 at the historical highest level, though the range of the catch had been from 10,000 t to 30,000 t during the period from the 1960s to the mid 1990s. Japanese longline fishery commenced in this Ocean in 1952. The fishery caught albacore ranging from 9,000 to 18,000 t in the 1960s that corresponds to the beginning of the long history of the fishery. Since then the catch decreased rapidly and reached 400 t in 1977. This drastic change is due to the change of target species of the longline fishery, i.e., yellowfin tuna and albacore to southern bluefin tuna and bigeye tuna, during the 1970s. The catch continued to be a low level ranging from 400 t to 2,500 t until early 1990s. After that the catch slightly increased and was 6,200 t in 2006, which was highest during the past 40 years. However, it is still about one third of the catch at the peak in 1964. Summary of albacore fishery in the Indian Ocean by Japanese longline is reported by Matsumoto and Uosaki (2011).

For the Indian albacore caught by Japanese longline fishery, CPUE standardization using the General Linear Model (GLM) with the assumption that the error structure belongs to log-normal had been carried out for 1960-1991 (Uozumi, 1994) and for 1960-2002 (Uosaki, 2004). However, possibly GLM with negative binominal error structure is better for standardization of albacore CPUE by Japanese longline which includes certain amount of zero catch data. In this document, the standardization of albacore CPUE by Japanese longline was conducted based on two models mentioned above and the results were compared, and discussed which model is more appropriate.

## 2.MATERIALS AND METHODS

### 2.1. Data

The data used here is the logbook data that has been compiled at National Research Institute of Far Seas Fisheries (NRIFSF) based on the logbook mandatory submitted by the fishermen of the longline vessel larger than 20 gross ton (GRT). The data is aggregated by month, 5˚x5˚ block, and number of hooks per basket (HPB), for 1975-2010. The data in 2010 is preliminary. CPUE was defined as the number of fish caught per 1,000 hooks. Observations with less than 5,000 hooks were excluded from this analysis.

### 2.2. Standardization

For the model of standardization of albacore CPUE, generalized linear model were used. Albacore catch by Japanese longline fishery includes a certain proportion of zero-catch data. In that case, the model with negative binominal error structure (NB model), instead of that with log-normal error structure (LN model), may be better.

Therefore, these two models were examined. The former model used is almost the same as that used by Uozumi (1994) or Uosaki (2004). These models include main effects of year, season, subarea and gear configuration. Quarter was used for fishing season categorized into four levels. The subarea was categorized into eight levels (Fig. 1) and the gear configuration was categorized into four levels (4-7, 8-11, 12-15 and 16-21 HPB). Because the information of gear configuration was not available for 1960-1974, each observation was regarded as the 4-7 HPB. The classification of subarea was defined based on the spatial distribution patterns of nominal CPUE of albacore and of species composition of longline catch. This stratification was modified from Uozumi (1994), and the same as that by Uosaki (2004). In order to include observations with no catch of albacore, a constant was added to the CPUE. The constant 0.3 was used as 10% of mean CPUE. The model used was:

| Error structure | Model |
|---|---|
| LN model | $\ln(CPUE+0.3) = \mu + Y_i + Q_j + A_k + G_l + Q{*}A_{jk} + Q{*}G_{jl} + e_{ijkl}$ ----------------- (1) |
| NB model | $Catch = H \cdot \exp(\mu + Y_i + Q_j + A_k + G_l + Q{*}A_{jk} + Q{*}G_{jl} + e_{ijkl})$ --------------- (2) |

where $\mu$: intercept

$Y_i$: effect of year in year $i$ $Q_j$: effect of quarter in quarter $j$

$A_k$: effect of subarea in area $k$ $G_l$: effect of gear in gear $l$

$Q{*}A_{jk}$: interaction term between quarter and area in quarter $j$ and area $k$

$Q{*}G_{jl}$: interaction term between quarter and gear in quarter $j$ and gear $l$

H: number of hooks used. Catch: catch in number

$e_{ijkl}$ : error term

Standardized CPUE for LN model was calculated as follows:

$$\text{Standardized CPUE}_i = \text{EXP}\,(\,\text{LSM}(Y_i) + \text{MSE}/2\,) - C$$

where $\text{LSM}(Y_i)$: least square mean of year effect in year $i$

MSE: Mean square error

C: constant (10% of mean CPUE)

The analyses were conducted using SAS 9.2.

### 2.3. Comparison between two models using information criteria

Shono (2001) introduced how to judge statistically which model is more appropriate between log-normal (LN) and Poisson (PO) models and illustrated the use of information criteria was applicable by considering equivalent response variables for both models. This method was applicable to judge between the LN and the NB models and was applied here for this study.

As an example, it is described using the case of models for Target period. First, instead of formula (1), following formula was considered:

$$\ln(Catch+0.3) = \ln(H) + \mu + Y + Q + A + Y{*}Q + Q{*}A + \varepsilon$$

where ln(Catch+0.3) shows expectation of catch in number, which belongs to normal distribution, and constant=0.3 in this case as used in formula (1). In the actual calculation using SAS, ln(H) should be set as 'offset' on GENMOD procedure. Second, the catch-and-effort data was fit to this model and obtain information criteria as the LN model. Finally, the two information criteria obtained from this LN, and the NB model obtain though model fitting with formula (2) were compared.

Shono (2001) also showed how to calculate maximum log-likelihood in LN and PO models, and pointed out that SAS package dose not calculate appropriately maximum log-likelihood (MLL) and need to add a term shown as follows;

In the case of LN model,

$$MLL = \ln L(\hat{\mu}, \hat{\lambda} \mid C + \text{constant}) = \begin{pmatrix} \text{'loglikelih ood' obtained from SAS output for} \\ \text{GENMOD procedure with log - normal distribution} \end{pmatrix} - \sum_{i=1}^{n} \ln(C_i + \text{constant})$$

In the case of PO model,

$$MLL = \ln L(\hat{\lambda} \mid C) = \begin{pmatrix} \text{'loglikelih ood' obtained from SAS output for} \\ \text{GENMOD procedure with Poisson distribution} \end{pmatrix} - \sum_{i=1}^{n} \ln(\Gamma(C_i + 1))$$

where,

$n$ is number of observations,

$\hat{\mu}$ is maximum likelihood estimator (MLE) of $\mu$ ($\mu=(\mu_1, \ldots, \mu_n)$)

$\hat{\sigma}^2$ is MLE of $\sigma$

Calculation for MLL for NB model is the same as that for PO model. In order to judge better fitting to model among models, well-known information criteria AIC, BIC (Schwarz 1978) and c-AIC (Sugiura 1978) were used. Those are defined as follows;

AIC=-2*(MLL)+2*p
BIC=-2*(MLL)+p*ln(n)
c-AIC=-2*(MLL)+2*n*p/(n-p-1)
where p is number of parameters.

## 3. RESULT AND DISCUSSION

The analysis of variance for the GLM analyses is shown in **Table 1**. This shows all the effects were significant at 0.1 % level. Table 2 shows the results of calculation of information criteria. The smaller the criteria are, the better the model is. Therefore, these results show that LN model is better than NB model for standardization of albacore CPUE. Table 3 shows CPUE indices with CV and confidence intervals. The distribution of standardized residual by LN model indicated not to be largely unbiased as shown in Fig. 2 (distribution of standardized residual), Fig. 3 (QQ-plots) and Fig. 4 (box plots) although QQ-plot for NB model shows slight skew. In the case of LN model, standardized CPUE was high at about 10-12 fish/1000 hooks during 1960-1965, and then rapidly decreased to about 2 fish/1000hooks, 20 % of the level during the 1965-1978 period (Fig. 5). Since then the CPUE became stable at the level in the view of whole time series analyzed. However, the CPUE showed slight increasing trend since 1995. CPUE in 2010 recovered to the level in the early 1970s. Comparing CPUE by LN model with that by NB model, the trend of both indices was very similar. However, the trend was different during early 1960s; CPUE by NB model decreased during early 1960s and CPUE by LN model was comparatively stable during that period.

Uosaki (2004) demonstrated that since late 1960s, Japanese longline fishery has been running without targeting albacore, and that the fishing effort has not deployed in the region where albacore is abundant, though a part of the longline fleet had primarily caught albacore in the 1960s. From this situation the standardized CPUE obtained here may not reflect the abundance of albacore in the Indian Ocean. At least after 1975 Japanese longline has caught albacore only in the geographical margin of the region where albacore abundantly distributed, as pointed out by Uozumi (1994).

The standardized CPUE using the data only from Area 2 and Area 4 (modified model), where albacore is generally abundant, was shown in Fig. 7 just for comparison to that shown above (reference model). This indicated that the CPUE for the modified model showed the similar trend to that for the reference model, and that the standardized CPUE even in the abundant region was as low as in the other region after 1970s. This suggests that the longline fishery operated without targeting albacore even in this region.

**REFERENCE**

Matsumoto, T. and K. Uosaki. 2011. Review of Japanese longline fishery and its albacore catch in the Indian Ocean. IOTC–2011–WPTmT03–13. 20pp.

Schwarz, G. 1978. Estimating the dimension of a model. Ann. Statist., 6: 461-464.

Shono, H. 2001. Comparison of statistical models for CPUE standardization by information criteria—Poisson model vs. log-normal model. IOTC/WPM/2001/05. 12pp.

Sugiura, N. 1978. Further analysis of the data by Akaike's information criterion and the finite correction. Common. Statist. Theor. Meth., 7: 13-26.

Uosaki, K. 2004: Brief review of Japanese longline fishery and its albacore catch in the Indian Ocean. IOTC-WPTMP-2004-08.

Uosaki, K. 2004: Update of the standardized CPUE of the albacore caught by Japanese longline fishery in the Indian Ocean. IOTC-WPTMP-2004-09.

Uozumi, Y. 1994: The CPUE trend for albacore in Indian Ocean waters caught by the Japanese longline fishery. Proceeding of 5th Expert Consultation on Indian Ocean tunas. IPTP Col. Vol. 8. 147-149.

Table 1. Analysis of variance for the GLM analyses.

**LN model**

| Source | DF | SS | Mean Sq. | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 93 | 50668.04 | 544.82 | 479.89 | <.0001 |
| Error | 54352 | 61705.23 | 1.14 | | |
| Corr. Tot. | 54445 | 112373.26 | | | |

R-square= 0.45089 C.V.= 712.7197

| Source | DF | Type III SS | Mean Sq. | F Value | Pr > F |
|---|---|---|---|---|---|
| Y | 50 | 8945.9 | 178.9 | 157.6 | <.0001 |
| Q | 3 | 1153.8 | 384.6 | 338.8 | <.0001 |
| A | 7 | 25431.8 | 3633.1 | 3200.2 | <.0001 |
| G | 3 | 236.8 | 78.9 | 69.5 | <.0001 |
| Q*A | 21 | 2771.4 | 132.0 | 116.2 | <.0001 |
| Q*G | 9 | 57.0 | 6.3 | 5.6 | <.0001 |

**NB model**

| Source | DF | Chi-Square | Pr>Chi |
|---|---|---|---|
| Y | 50 | 4573.79 | <.0001 |
| Q | 3 | 1569.56 | <.0001 |
| A | 7 | 9263.48 | <.0001 |
| G | 3 | 194.27 | <.0001 |
| Q*A | 21 | 3469.37 | <.0001 |
| Q*G | 9 | 98.96 | <.0001 |

Table 2. Results of calculation for several information criteria (AIC, BIC and c-AIC) to judge which models is appropriate. Smaller value is better in each information criteria. Log likelihood (SAS output): value of log-likelihood in output of GENMOD procedure of SAS package. MLL: maximum log-likelihood, n: number of observations, p: number of parameters.

| Period | Log likelihood | Σ(ln(C+const)) | MLL | n | p | AIC | BIC | c-AIC |
|---|---|---|---|---|---|---|---|---|
| LN model | -116105 | 122836 | -238941 | 54552 | 116 | 478113 | 479146 | 478114 |
| NB model | 93051510 | 93318896 | -267385 | 54552 | 116 | 535003 | 536036 | 535003 |

Table 3. Standardized CPUE (number of fish/hooks) with the 95% confidence intervals for each model.

| | LN model | | | | NB model | | | |
|---|---|---|---|---|---|---|---|---|
| Year | Std CPUE | CV | Upper CL | Lower CL | Std CPUE | CV | Upper CL | Lower CL |
| 1960 | 10.500 | 0.053 | 11.689 | 9.429 | 14.880 | 0.0929 | 17.851 | 12.403 |
| 1961 | 10.928 | 0.055 | 12.216 | 9.773 | 16.119 | 0.097 | 19.494 | 13.328 |
| 1962 | 11.036 | 0.045 | 12.090 | 10.071 | 13.108 | 0.0792 | 15.309 | 11.223 |
| 1963 | 9.168 | 0.048 | 10.096 | 8.323 | 10.657 | 0.0831 | 12.542 | 9.055 |
| 1964 | 12.197 | 0.046 | 13.371 | 11.125 | 13.500 | 0.0801 | 15.795 | 11.539 |
| 1965 | 7.709 | 0.044 | 8.436 | 7.042 | 10.828 | 0.0770 | 12.591 | 9.311 |
| 1966 | 6.245 | 0.042 | 6.810 | 5.725 | 8.226 | 0.0737 | 9.504 | 7.120 |
| 1967 | 6.237 | 0.038 | 6.743 | 5.767 | 7.210 | 0.0660 | 8.205 | 6.335 |
| 1968 | 5.259 | 0.039 | 5.704 | 4.848 | 6.006 | 0.0683 | 6.866 | 5.253 |
| 1969 | 4.662 | 0.038 | 5.051 | 4.301 | 6.636 | 0.0670 | 7.567 | 5.819 |
| 1970 | 3.971 | 0.042 | 4.340 | 3.631 | 4.389 | 0.0741 | 5.075 | 3.795 |
| 1971 | 3.108 | 0.042 | 3.404 | 2.836 | 2.870 | 0.0744 | 3.321 | 2.481 |
| 1972 | 2.917 | 0.053 | 3.267 | 2.601 | 2.620 | 0.0927 | 3.142 | 2.185 |
| 1973 | 2.811 | 0.050 | 3.132 | 2.519 | 3.228 | 0.0897 | 3.848 | 2.707 |
| 1974 | 2.889 | 0.045 | 3.186 | 2.618 | 2.952 | 0.0794 | 3.449 | 2.527 |
| 1975 | 1.882 | 0.039 | 2.056 | 1.720 | 1.662 | 0.0706 | 1.908 | 1.447 |
| 1976 | 2.683 | 0.049 | 2.983 | 2.410 | 3.570 | 0.0913 | 4.269 | 2.985 |
| 1977 | 1.669 | 0.050 | 1.872 | 1.485 | 1.643 | 0.0912 | 1.964 | 1.374 |
| 1978 | 1.258 | 0.040 | 1.384 | 1.142 | 0.689 | 0.0735 | 0.796 | 0.597 |
| 1979 | 1.409 | 0.044 | 1.564 | 1.267 | 1.230 | 0.0807 | 1.441 | 1.050 |
| 1980 | 1.592 | 0.041 | 1.751 | 1.445 | 1.113 | 0.0743 | 1.287 | 0.962 |
| 1981 | 1.757 | 0.038 | 1.916 | 1.610 | 1.420 | 0.0675 | 1.621 | 1.244 |
| 1982 | 1.801 | 0.037 | 1.960 | 1.654 | 1.527 | 0.0663 | 1.739 | 1.341 |
| 1983 | 1.804 | 0.035 | 1.954 | 1.664 | 1.397 | 0.0631 | 1.581 | 1.234 |
| 1984 | 1.793 | 0.035 | 1.942 | 1.654 | 1.365 | 0.0631 | 1.544 | 1.206 |
| 1985 | 1.866 | 0.034 | 2.015 | 1.726 | 1.280 | 0.0601 | 1.441 | 1.138 |
| 1986 | 2.217 | 0.033 | 2.387 | 2.058 | 2.231 | 0.0597 | 2.508 | 1.985 |
| 1987 | 2.339 | 0.035 | 2.528 | 2.163 | 2.807 | 0.0636 | 3.180 | 2.478 |
| 1988 | 1.553 | 0.037 | 1.694 | 1.422 | 1.237 | 0.0680 | 1.413 | 1.082 |
| 1989 | 1.485 | 0.042 | 1.637 | 1.344 | 1.064 | 0.0757 | 1.235 | 0.918 |
| 1990 | 1.594 | 0.042 | 1.756 | 1.445 | 1.480 | 0.0748 | 1.714 | 1.278 |
| 1991 | 1.123 | 0.038 | 1.233 | 1.022 | 0.876 | 0.0673 | 1.000 | 0.768 |
| 1992 | 1.371 | 0.039 | 1.505 | 1.247 | 1.383 | 0.0698 | 1.585 | 1.206 |
| 1993 | 1.409 | 0.038 | 1.541 | 1.286 | 1.255 | 0.0673 | 1.432 | 1.100 |
| 1994 | 1.190 | 0.028 | 1.272 | 1.111 | 1.192 | 0.0489 | 1.312 | 1.083 |
| 1995 | 1.087 | 0.025 | 1.156 | 1.020 | 1.048 | 0.0443 | 1.143 | 0.960 |
| 1996 | 1.151 | 0.024 | 1.220 | 1.084 | 1.132 | 0.0422 | 1.229 | 1.042 |
| 1997 | 1.476 | 0.024 | 1.560 | 1.396 | 1.217 | 0.0420 | 1.321 | 1.121 |
| 1998 | 1.715 | 0.024 | 1.814 | 1.621 | 2.147 | 0.0449 | 2.344 | 1.966 |
| 1999 | 1.291 | 0.026 | 1.374 | 1.212 | 1.192 | 0.0464 | 1.305 | 1.088 |
| 2000 | 1.414 | 0.027 | 1.507 | 1.327 | 1.168 | 0.0473 | 1.281 | 1.064 |
| 2001 | 1.433 | 0.027 | 1.526 | 1.345 | 1.496 | 0.0473 | 1.641 | 1.364 |
| 2002 | 1.619 | 0.026 | 1.720 | 1.522 | 1.476 | 0.0468 | 1.618 | 1.347 |
| 2003 | 1.830 | 0.031 | 1.964 | 1.703 | 1.631 | 0.0551 | 1.817 | 1.464 |
| 2004 | 2.081 | 0.029 | 2.219 | 1.950 | 2.415 | 0.0518 | 2.673 | 2.181 |
| 2005 | 1.669 | 0.028 | 1.779 | 1.565 | 1.212 | 0.0489 | 1.334 | 1.101 |
| 2006 | 2.079 | 0.025 | 2.199 | 1.965 | 1.515 | 0.0443 | 1.652 | 1.389 |
| 2007 | 2.357 | 0.025 | 2.493 | 2.227 | 1.792 | 0.0448 | 1.956 | 1.641 |
| 2008 | 2.387 | 0.028 | 2.537 | 2.245 | 1.901 | 0.0485 | 2.091 | 1.729 |
| 2009 | 2.352 | 0.029 | 2.509 | 2.204 | 1.806 | 0.0518 | 1.999 | 1.631 |
| 2010 | 3.467 | 0.037 | 3.751 | 3.202 | 2.710 | 0.0665 | 3.087 | 2.379 |

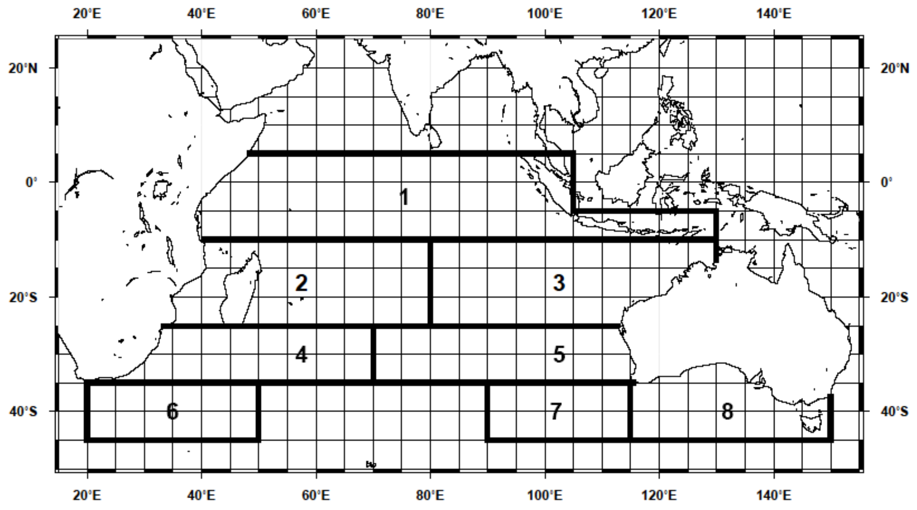


Fig. 1. Subarea used for the GLM analysis.

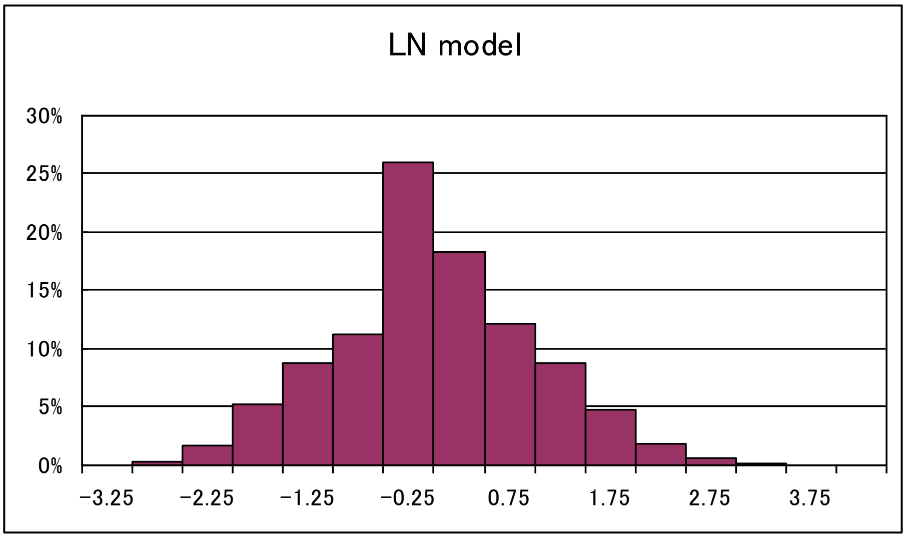


Fig. 2. Distribution of the standardized residual for the GLM analysis (LN model).

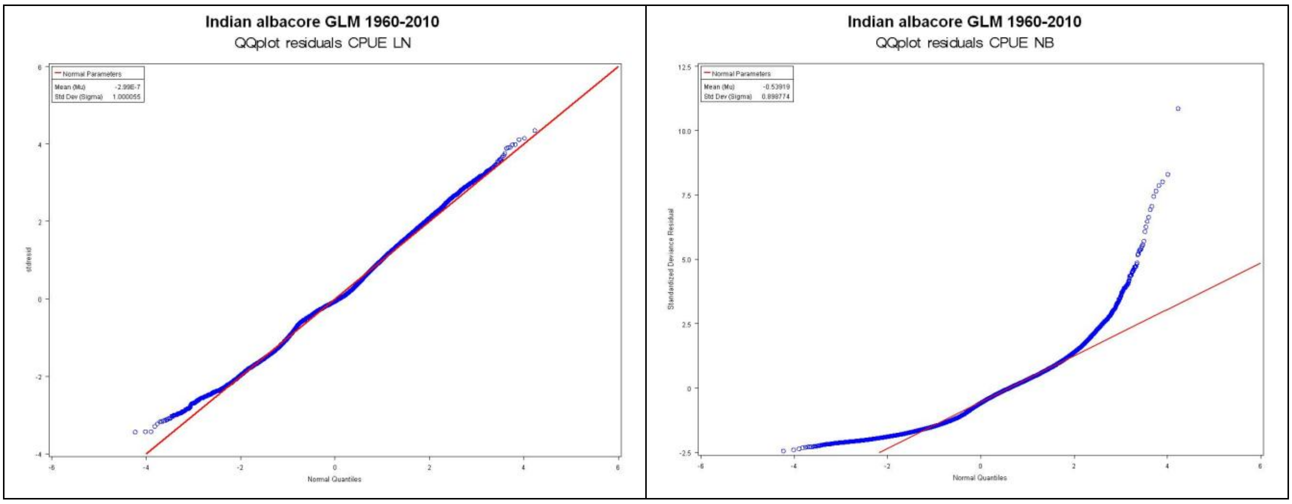


Fig. 3. QQ-plots of standardized residual for the LN and NB models.

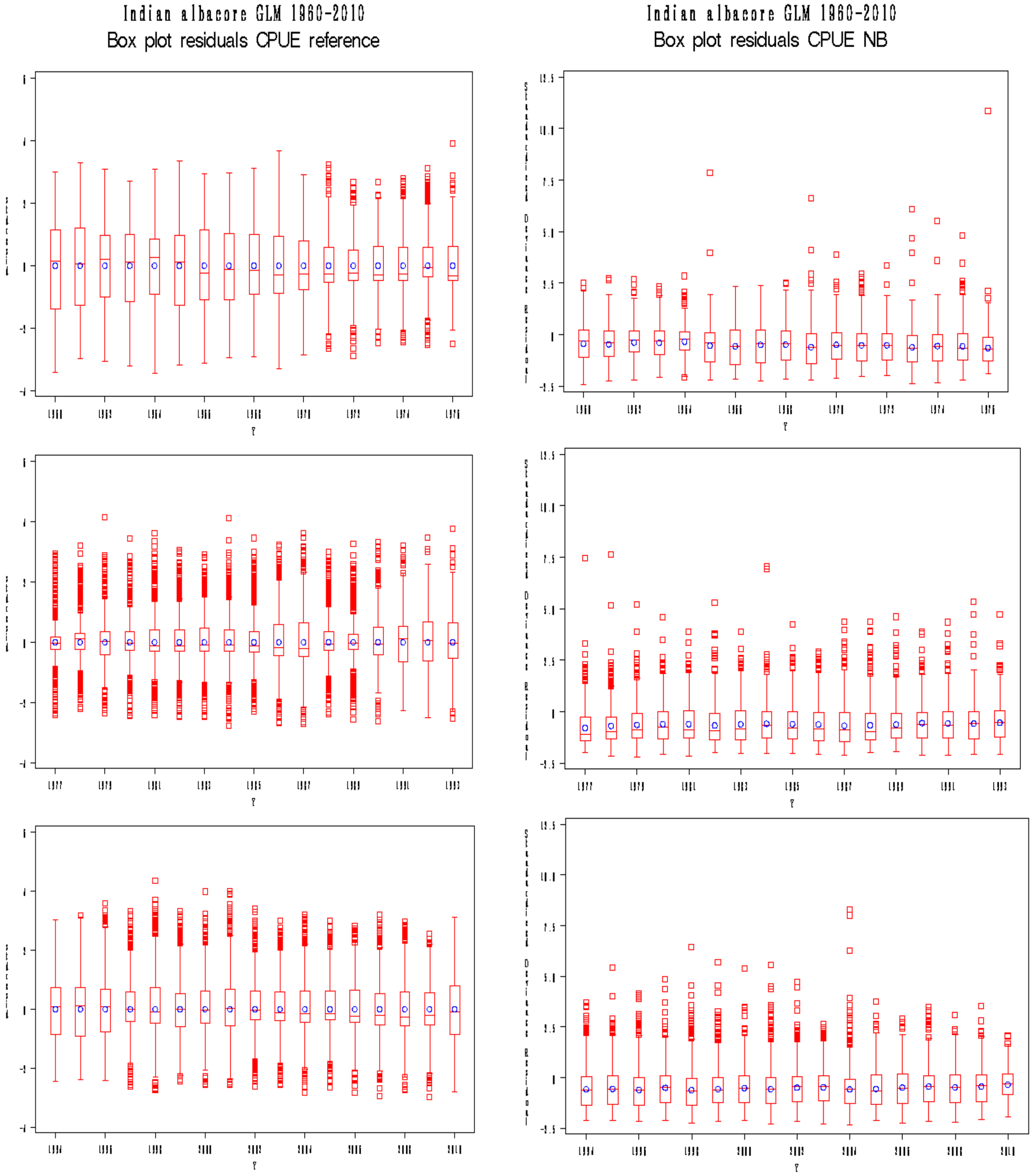


Fig. 4. Box plot of the standardized residual by year for the GLM analysis (left: LN model, right: NB model). Circle: mean, box: 25th and 75th percentile, horizontal line in the box: median, bars: maximum and minimum observation between 1.5 IQR (interquartile range) above 75th percentile and 1.5 IQR below 25th percentile, squares: outliers.

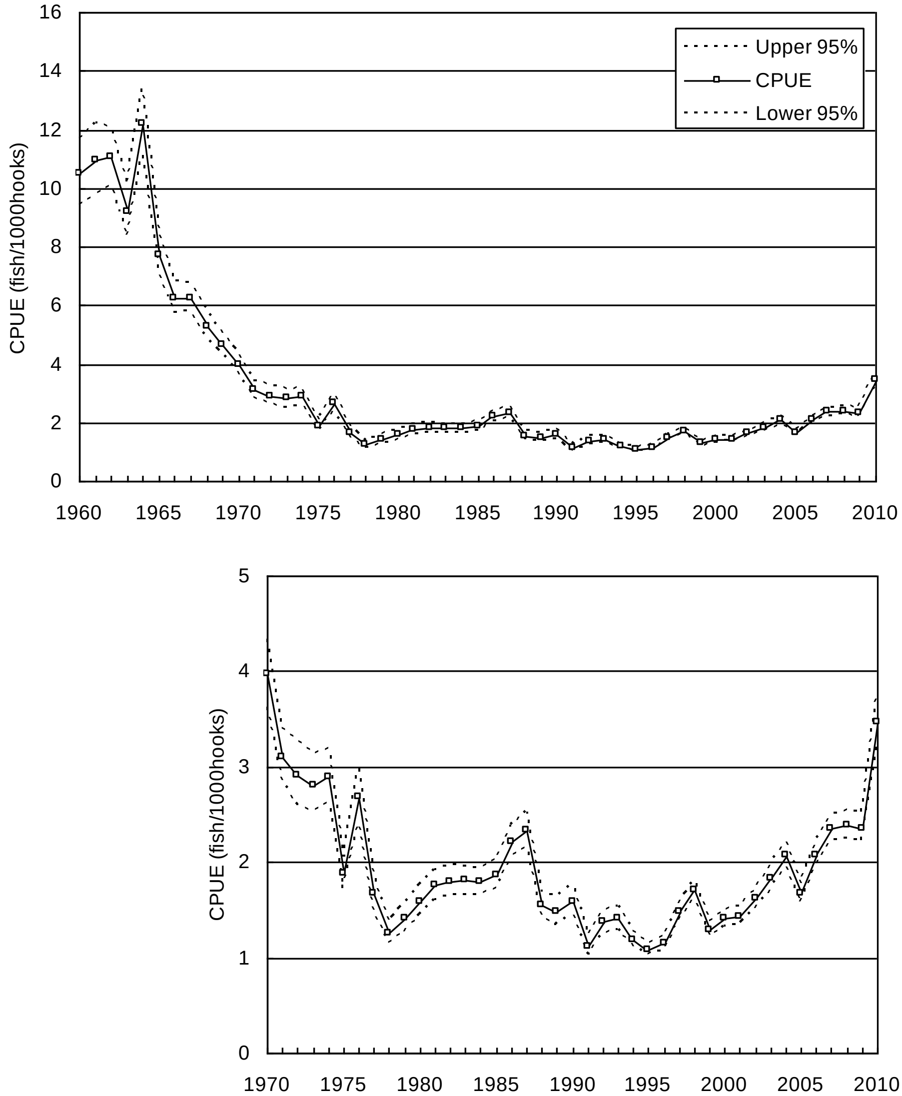


Fig. 5. Standardized CPUE (LN model) for the albacore in the Indian Ocean with confidence intervals. The bottom panel is changed on the scale from top panel and is shown only for 1970-2010.

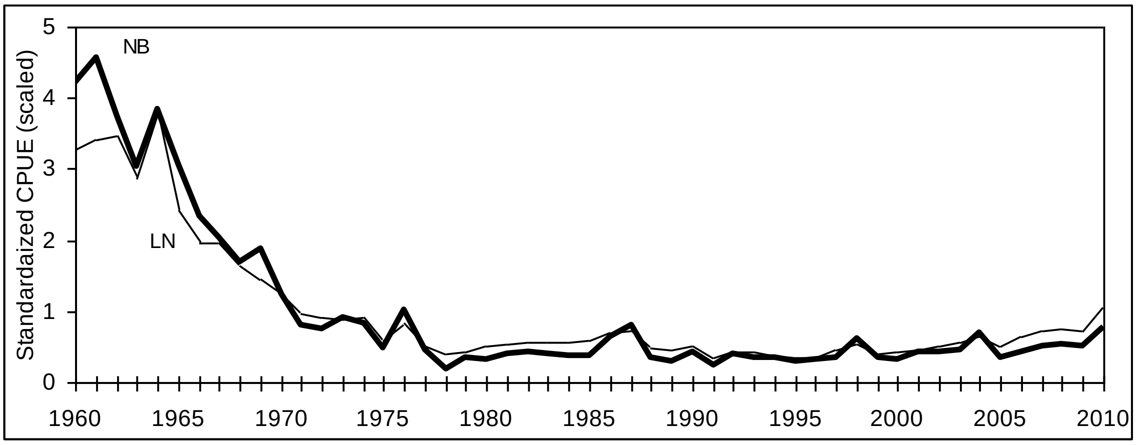


Fig. 6. Comparison of standardized CPUE for albacore in the Indian Ocean by two different models. CPUE indices were scaled by dividing by the average.

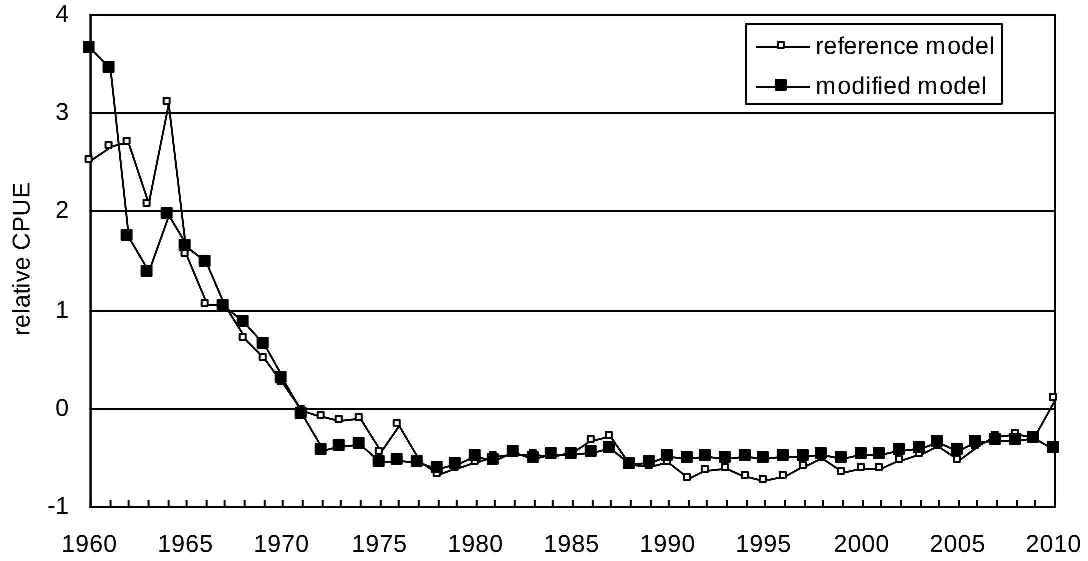


Fig. 7. Standardized CPUEs for the reference and modified models. The CPUE for the modified model were calculated using only from Area 2 and Area 4 where albacore is generally abundant. The CPUE for the reference model is the same as that shown in Fig. 5. Both CPUEs were adjusted with taking difference to mean and dividing standard deviation.