# NPFC-2021-SSC PS07-WP18

# **Standardized CPUE of Pacific saury (*Cololabis saira*) caught by the Korean’s stick-held dip net fishery up to 2020**

Kyum Joon PARK and Sung-Il LEE

*National Institute of Fisheries Science, Republic of Korea*

# **Introduction**

Pacific saury (*Cololabis saira*) is widely distributed in the subarctic and subtropical areas of the North Pacific Ocean from inshore waters of Japan and Kuril Islands eastward to Gulf of Alaska and southward to Mexico (Parin 1960). The species migrates seasonally from the subtropical Kuroshio Current in winter to the Subarctic Oyashio Current in summer, for feeding on zooplankton (Shimizu et al., 2009, Taki, 2011). The preferred water temperature for Pacific saury is 13-18 ℃, and the vertical distribution is from near-surface down to around 230m depth (Eschmeyer et al. 1983, Syah et al. 2017). The highest cpue (catch per unit effort) of Pacific saury was found when the SST ranged from 14 to 16 ℃ (Tseng et al. 2013).

After the first exploratory stick-held dip net (SHDN) fishing from Korea had been conducted in the Northwest Pacific Ocean in the 1960s, three commercial fishing vessels commenced saury fishing in the area in 1985 (Jo 2003). Since then, the Korean stick-held dip net fishery has grown rapidly year-by-year, and the largest catch, 50 thousand tons, was made in 1997. Korea’s Pacific saury catch information has been managed by two organizations: Korea Overseas Fisheries Association (KOFA) and National Institute of Fisheries Science (NIFS). KOFA collected total catches and NIFS collects logbook data from fishing vessels as subsamples. The logbook contains daily catch and additional information such as light power (kw) and amount of catch by size (S, M, L and 2L). However, since September 2015, an electronic reporting system (ERS) replaced the traditional logbook and has been collecting the catch data in near real-time. Accordingly, some of the data categories have been changed after the introduction of the ERS (e.g. the catches by fish size were unified into the total catch and light power information was excluded).

# **Method**

The logbook information from 2001 to 2020 was used in the CPUE standardization process. We standardized the CPUE of Pacific saury derived from the Korean fishery according to the standardization protocol (NPFC - 2017 - TWG PSSA02 - Report Annex D) agreed in the 2nd meeting of Technical Working Group on Pacific Saury Stock Assessment (Appendix 1).

## Commercial fishery data sources

Data used in this study were obtained from log books which had been reported to KOFA and NIFS from 2001 to early 2015 and electronic log books reported by ESR from late 2015 to 2020 including information on date, fishing location (longitude and latitude), catch in weight (mt), sea surface temperature (SST) measured using an on-board thermometer, and GRT(gross tonnage) of the fishing vessels. Inter-annual variation of monthly fishing ground of Korea SHDN for Pacific saury from 2001 to 2020 shown in Fig. 1.

## Area definition

Fishing ground of the Korean SHDN fishery was divided into six subareas based on oceanographic characteristics and jurisdictions, reviewed at TWG-PSSA03. Based on this definition, two cases of categories were applied as variables to this analysis. Area2, which the original definition of 6 subareas are applied to its subareas, and Area1, where subareas of Area2 are grouped into, depending on logical character as continental (subarea 1,2 and 3), near continental (subarea 4 and 5) and far area (subarea 7) by concerning logical character (Appendix II).

## Factors considered

The factors of year, month, fishing area, gross registered tonnage (GRT) of fishing vessels and SST what we used in the previous study on CPUE standardization (NPFC-2018-SSC-PS03-WP07), were incorporated as explanatory variables in the CPUE standardization of this study (Table 1). The correlation matrix for the explanatory variables is shown in Fig. 2.

## Statistical methods for CPUE standardization

i. Model specification

Generalized linear model (GLM) was used to standardize the CPUE, because there was no zero catch in Korean SHDN fishery. CPUE used in this study was defined as catch per vessel per day.

The full model used in GLM analysis is given as:

ln(CPUE) = Intercept + Year + Month + Area + Sst + Grt + two-way interactions + ε,

where Year, Month and Area are categorical variables composed of categories of 20 years (2001–2020), 6 months (June – November) and 6 areas (Area2), respectively (Table 1). The other area category (Area1) was made of 6 subareas grouped into 3, namely, continental, near continental and far area. GRT1 and 2 of the vessel tonnage variables has 3 and 4 categories, which are divided at intervals of 100 ton (Grt1) and 50 ton (Grt2), respectively. There are 4 and 7 categories of SST, which are divided at intervals of 4°C (Sst1) and 2°C (Sst2), respectively. The optimal categorizations regarding Area, GRT and SST were determined through model selections. Two-way interactions for all combinations of explanatory variables were incorporated in the full model.

ii. Model selection and diagnosis

We applied Akaike Information Criteria (AIC), Bayesian information criterion (BIC) and the proportions of deviance explained relative to the total explained deviance to measure the predictive ability and determined the selected model with the minimum AIC and BIC through the GLM analyses. For model diagnosis, the percent deviation explained was calculated in addition to Q-Q plot and histogram of residuals.

iii. Calculation of standardized CPUE

Time series of standardized CPUE were estimated using the selected models from the GLM analyses. The formula is given as;

Where, is CPUE indices in ith year, is the observation number in ith year, is the kth fitted CPUE data in ith year.

# **Results and Discussion**

In the selected model from the GLM analysis, where deviance explained was 21.66%, vessel size and sea surface temperature were divided at intervals of 50 ton (Grt2) and those of 2°C (Sst2), respectively (Table 2). Analysis of deviance indicated that all selected explanatory variables were significant at the significant level of 0.01 in the selected model (Table 3). The Q-Q plot, the histogram and the boxplot of residuals for evaluating the assumption on error distribution are shown in Fig. 3. The indicated residuals were distributed normally around 0, even though long tails were observed at the both ends. There were no tendencies in residuals across year. The CPUE were appropriately modeled using the explanatory variables which were selected through the GLM analysis and the changes of the CPUE by categorical variables.

Annual standardized CPUE with 95% confidence intervals using the selected models from the best GLM and nominal CPUE are showed in Table 4. The standardized CPUE and nominal CPUE indicated a similar tendency. Both CPUEs were the lowest in 2020 since 2001 (Fig. 4).

# **References**

Akaike, H. 1974. A new look at the statistical model identification, IEEE Trans. Automat. Contr. 19(6):716-723.

Eschmeyer W. N., Herald E.S and Hamman H. 1983. A field guide to Pacific coast fishes of North America. Houghton Mifflin Company. 336pp.

Jo, H. S. 2003. Catch specification of Pacific Saury, *Cololabis saira*, caught by stick-held dip net fishery (Doctoral dissertation). Pukyong National University, Busan, Korea.

NPFC (2017). NPFC - 2017 - TWG PSSA02 - Report Annex D

NPFC. (2018). NPFC-2018-SSC-PS03-WP07

Parin, N.V. 1960. The range of the saury (*Cololabis saira* Brev.-Scombresocidae, Pices) and effects fo oceanographic features on its distribution. Proc. Acad. Sci. USSR 130(3):649-652.

Shimizu Y., Takahashi K., Ito S.I. Kakehi S., Tatebe H., Yasuda I., Kusaka A., et al. Transport of subarctic large copepods from the Oyashio area to the mixed water region by the coastal Oyashio intrusion, Fisheries Oceanography , 2009, vol. 18 (pg. 312-327)

Syah, A. F. Saitoh, S. I. Alabia, I.D., and Hirawaka, T. 2016. Predicting potential fishing zones for Pacific saury (*Cololabis saira*) with maximum entropy models and remotely sensed data. Fishery Bulletin, 114(3) 330-343.

Taki K. Distribution and population structure of *Thysanoessa inspinata* and its dominance among euphausiids off northeastern Japan, Journal of Plankton Research , 2011, vol. 33 (pg. 891-906)

Tseng, C-T., Su, N-J., Sun, C-L., Punt, A. E., Yeh, S-Z., Liu, D-C., and Su, W-C. 2013. Spatial and temporal variability of the Pacific saury (*Cololabis saira*) distribution in the northwestern Pacific Ocean. – ICES Journal of Marine Science, 70: 991–999.

Table 1 Summary of explanatory variables used in GLM analyses

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Cases | Categorical | Detail | Note |
| Year | *Year* | 20 categories | 20 years from 2001 to 2020 |  |
| Month | *Month* | 6 categories | 6 months from June to November |  |
| Area | *Area1*  *Area2* | 3 categories  6 categories | 1(Ⅰ, Ⅱ, Ⅲ), 2(Ⅳ, Ⅴ), 3(Ⅶ)  Ⅰ, Ⅱ, Ⅲ, Ⅳ, Ⅴ, Ⅶ | See Appendix II. |
| Sea surface temperature | *Sst1*  *Sst2* | 4 categories  7 categories | Sst<10℃;10℃≦Sst＜14℃，14℃≦Sst≤18℃, 18℃<Sst  Sst<10℃;10℃≦Sst＜12℃，12℃≦Sst＜14℃，…, 18℃≦Sst<20℃; 20℃≤Sst | at intervals of 4℃  at intervals of 2℃ |
| Vessel tonnage | *Grt1*  *Grt2* | 3 categories  4 categories | Grt<400tons; 400≦Grt＜500; 500≤Grt  Grt<400tons; 400≦Grt＜450; 450≦Grt＜500; 500≤Grt | at intervals of 100 tons  at intervals of 50 tons |

Table 2 Selected GLM models based on AIC and BIC values

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Best model in GLM analysis | AIC | BIC | % deviance explained |
| 1 | *Intercept*+*Year+ε* | 62,903 | 63,072 | 10.82 |
| 2 | *Intercept*+*Year*+*Month+ε* | 62,426 | 62,635 | 12.68 |
| 3 | *Intercept*+*Year*+*Month*+*Area1+ε* | 62,208 | 62,433 | 13.51 |
| 4 | *Intercept*+*Year*+*Month*+*Area1+Sst2+ε* | 62,195 | 62,429 | 13.57 |
| 5 | *Intercept*+*Year*+*Month*+*Area1+Sst2+Grt2+ε* | 62,079 | 62,337 | 14.02 |
| 6 | *Intercept*+*Year*+*Month*+*Area1+Sst2+Grt2+Year:Month+ε* | 60,981 | 61,979 | 18.65 |
| 7 | *Intercept*+*Year*+*Month*+*Area1+Sst2+Grt2+Year:Month +Year:Area1+ε* | 60,818 | 62,066 | 19.43 |
| 8 | *Intercept*+*Year*+*Month*+*Area1+Sst2+Grt2+Year:Month +Year:Area1+Month:Area1+ε* | 60,767 | 62,072 | 19.66 |
| 9 | *Intercept*+*Year*+*Month*+*Area1+Sst2+Grt2+Year:Month +Year:Area1+Month:Area1+Area1:Sst2+ε* | 60,734 | 62,055 | 19.78 |
| 10 | *Intercept*+*Year*+*Month*+*Area1+Sst2+Grt2+Year:Month +Year:Area1+Month:Area1+Area1:Sst2+Year:Grt2+ε* | 60,325 | 62,032 | 21.51 |
| 11 | *Intercept*+*Year*+*Month*+*Area1+Sst2+Grt2+Year:Month +Year:Area1+Month:Area1+Area1:Sst2+Year:Sst2+ε* | 60,727 | 62,201 | 19.93 |
| 12 | *Intercept*+*Year*+*Month*+*Area1+Sst2+Grt2+Year:Month +Year:Area1+Month:Area1+Area1:Sst2+Year:Grt2+Year:Sst2+ε* | 60,320 | 62,180 | 21.66 |

Table 3 Analysis of deviance for the selected model in GLM

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SS | Df | F | Pr(>F) | Signif. codes |
| Year | 2135.4 | 19 | 143.83 | < 2.2e-16 | \*\*\* |
| Month | 434.7 | 5 | 111.26 | < 2.2e-16 | \*\*\* |
| Area1 | 47.0 | 2 | 30.10 | 8.802e-14 | \*\*\* |
| Sst2 | 7.6 | 1 | 9.78 | 0.0017668 | \*\* |
| Grt2 | 111.2 | 3 | 47.45 | < 2.2e-16 | \*\*\* |
| Year:Month | 616.5 | 92 | 8.58 | < 2.2e-16 | \*\*\* |
| Year:Area1 | 147.3 | 31 | 6.08 | < 2.2e-16 | \*\*\* |
| Month:Area1 | 49.1 | 7 | 8.98 | 4.155e-11 | \*\*\* |
| Area1:Sst2 | 12.3 | 2 | 7.85 | 0.0003891 | \*\*\* |
| Year:Grt2 | 393.7 | 48 | 10.50 | < 2.2e-16 | \*\*\* |
| Year:Sst2 | 33.7 | 19 | 2.27 | 0.0012447 | \*\* |

Residuals 10370.1 19137

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05

Table 4. Nominal and standardized CPUE of Korean stick-held dip net fishery in the Pacific from 2001 to 2020

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | Nominal CPUE  (mt/vessel/day) | Standardized CPUE | SD of Standardized CPUE | 95% CI  Upper Lower | |
| 2001 | 9.44 | 3.82 | 1.58 | 3.88 | 3.75 |
| 2002 | 11.46 | 3.13 | 1.83 | 3.2 | 3.06 |
| 2003 | 17.75 | 5.93 | 2.97 | 6.05 | 5.81 |
| 2004 | 12.21 | 4.78 | 3.22 | 4.93 | 4.63 |
| 2005 | 18.62 | 9.97 | 3.79 | 10.17 | 9.76 |
| 2006 | 16.76 | 8.22 | 5.53 | 8.75 | 7.69 |
| 2007 | 15.73 | 7.15 | 6.01 | 7.6 | 6.7 |
| 2008 | 20.97 | 10.69 | 3.9 | 10.98 | 10.4 |
| 2009 | 11.61 | 4.37 | 2.11 | 4.48 | 4.26 |
| 2010 | 16.58 | 8.02 | 4.48 | 8.43 | 7.6 |
| 2011 | 13.08 | 4.74 | 2.64 | 5.01 | 4.47 |
| 2012 | 11.28 | 3.86 | 1.39 | 3.97 | 3.74 |
| 2013 | 11.69 | 4.67 | 1.43 | 4.74 | 4.6 |
| 2014 | 19.22 | 8.01 | 3.83 | 8.2 | 7.81 |
| 2015 | 7.90 | 3.4 | 1.5 | 3.47 | 3.33 |
| 2016 | 12.20 | 5.47 | 2.39 | 5.58 | 5.35 |
| 2017 | 7.49 | 3.36 | 2.71 | 3.47 | 3.25 |
| 2018 | 9.08 | 5.25 | 2.54 | 5.36 | 5.15 |
| 2019 | 6.40 | 3.37 | 1.38 | 3.44 | 3.3 |
| 2020 | 5.35 | 2.45 | 1.04 | 2.5 | 2.39 |

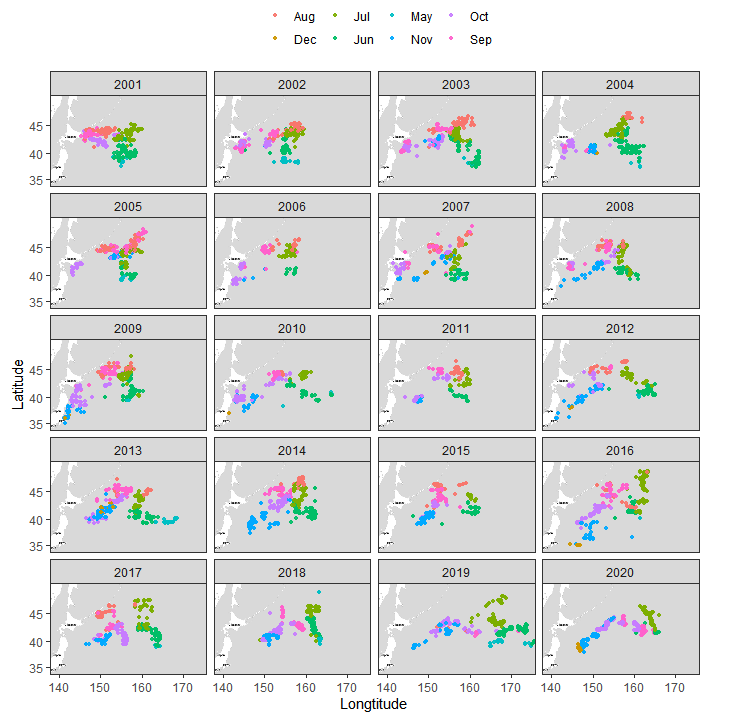


Fig. 1. Distribution of fishing ground of Korean stick-held dip net fishing vessels from 2001 to 2020.

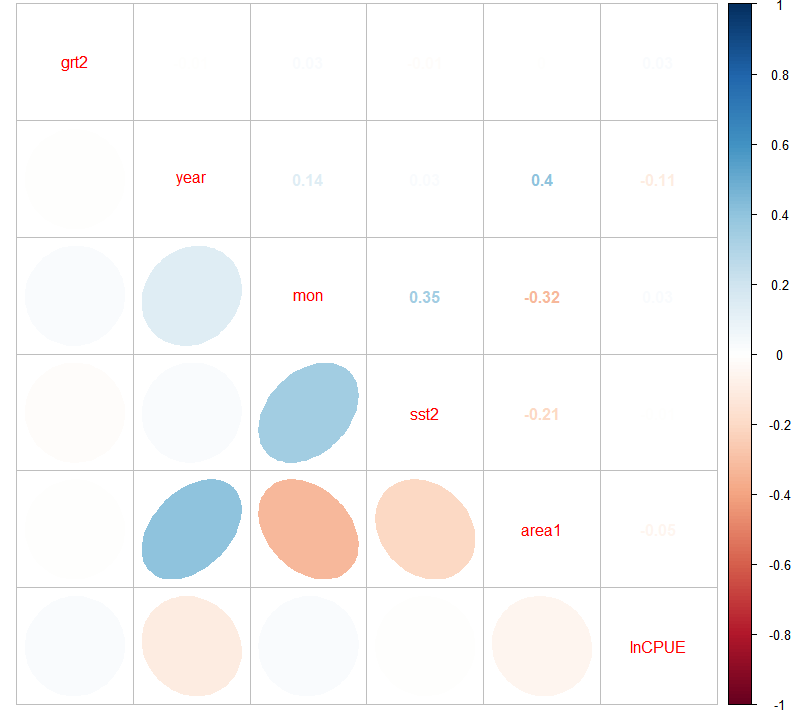


Fig. 2. Correlation matrix of explanatory variables used in the analysis.

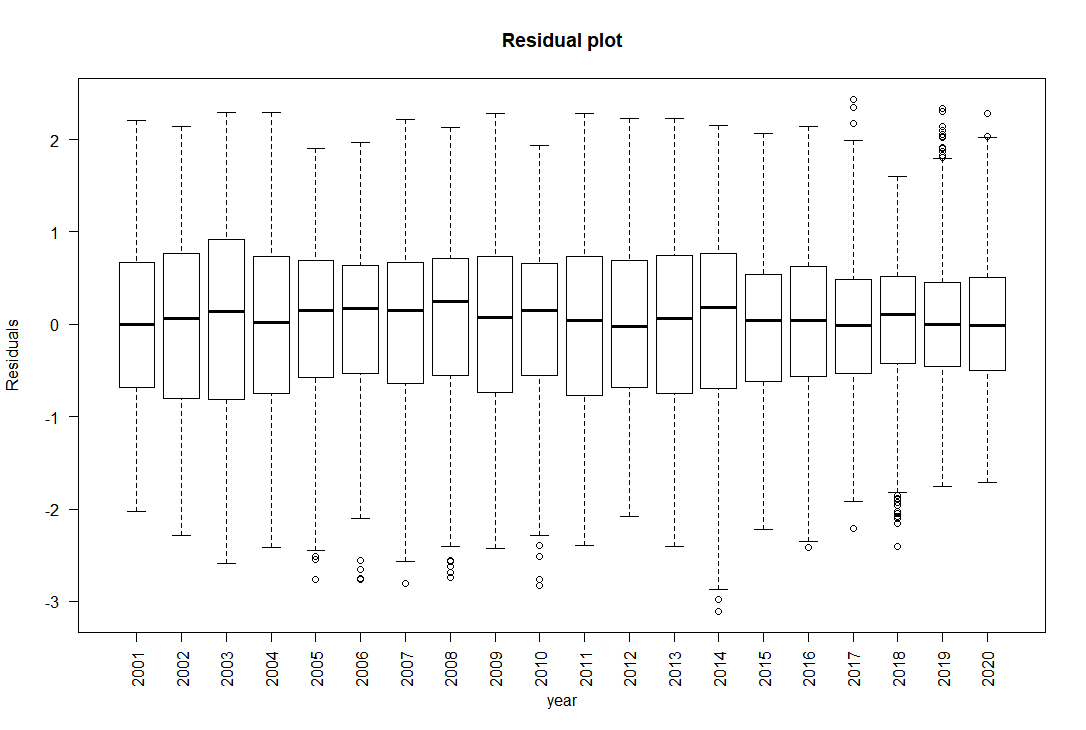
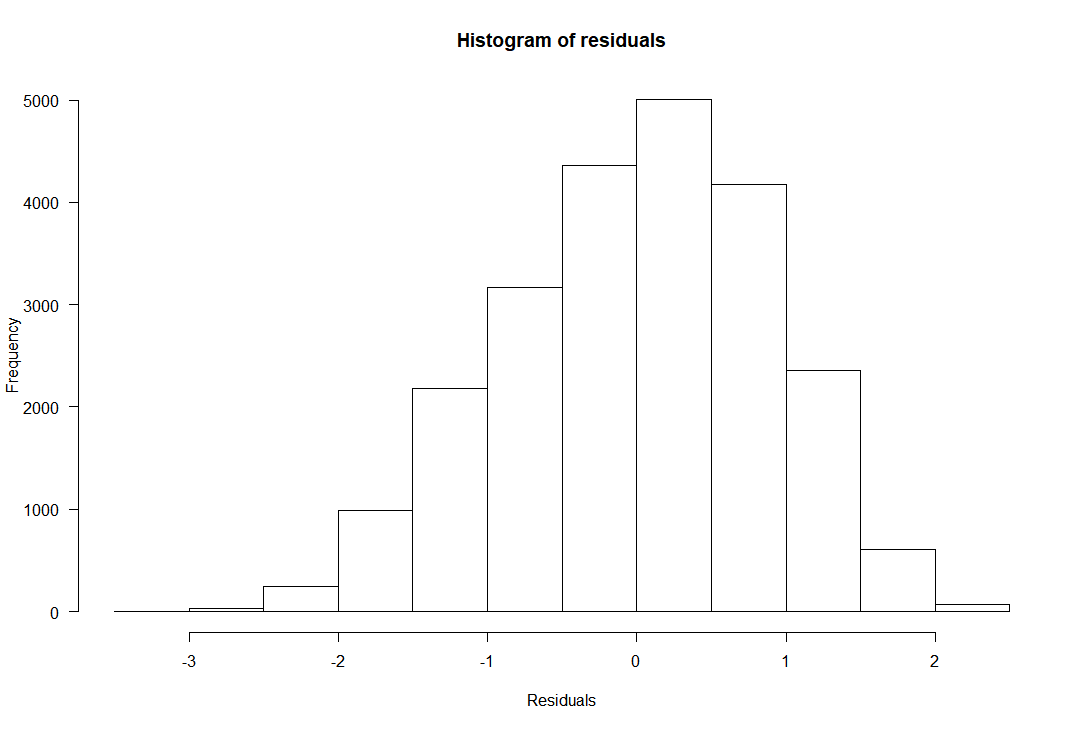
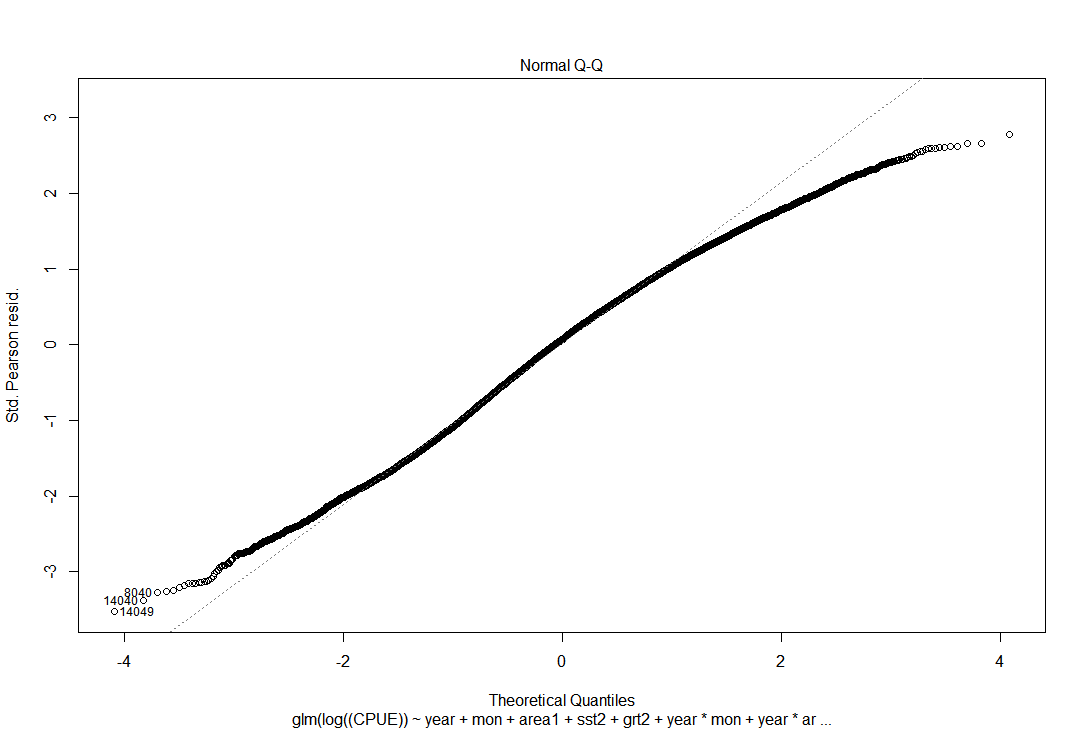


Fig. 3. Q-Q plot, histogram of residuals and residual plots cross years form the best GLM.

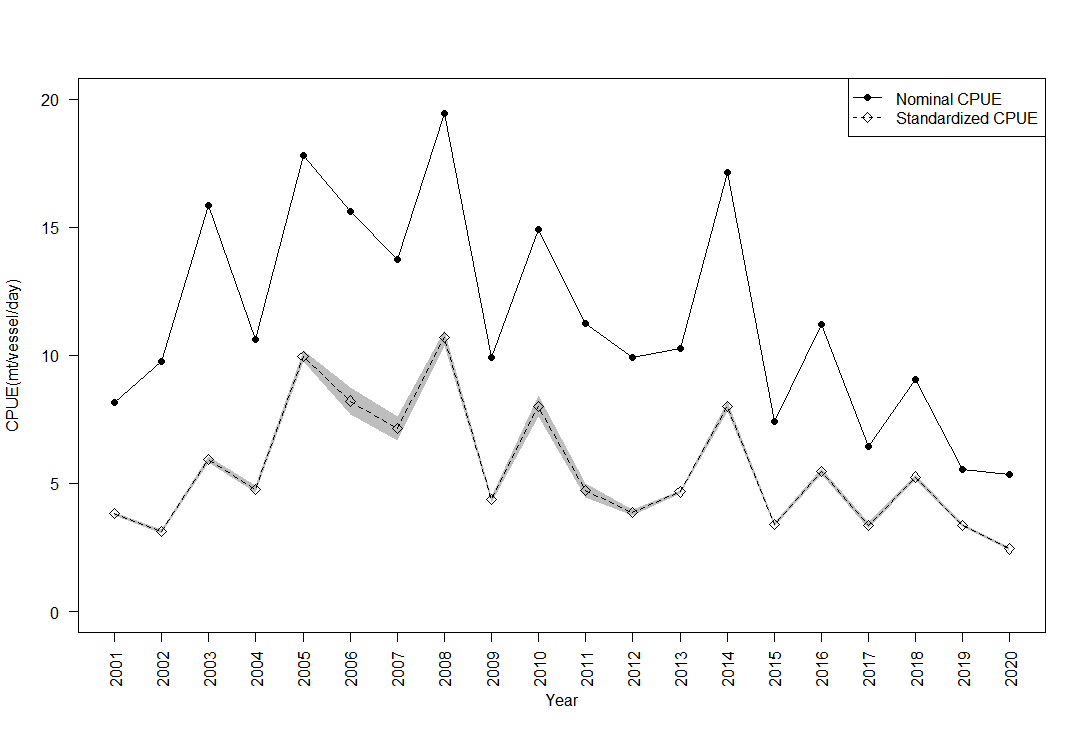


Fig. 4. Annual nominal CPUE and GLM estimated standardized CPUE of Korean stick-held dip net fishing vessels in Northwest Pacific from 2001 to 2020. Gray zone indicates 95% confidence band for the standardized CPUE.

# **Appendix I** Checklist for the CPUE standardization.

|  |  |  |
| --- | --- | --- |
| (1) | Conduct a thorough literature review to identify key factors (i.e., spatial, temporal, environmental, and fisheries variables) that may influence CPUE values; | Yes  (*see* previous working paper, NPFC-2018-SSC-PS03-WP07) |
| (2) | Determine temporal and spatial scales for data grouping for CPUE standardization; | Yes (*see* Table 1) |
| (3) | Plot spatio-temporal distributions of fishing efforts and catch to evaluate spatio-temporal patterns of fishing effort and catch; | Yes (*see* Fig. 1) |
| (4) | Calculate correlation matrix to evaluate correlations between each pair of those variables; | Yes (*see* Fig. 2) |
| (5) | Identify potential explanatory variables based on (1)-(4) as well as interaction terms to develop full model for the CPUE standardization; | Yes (*see* Table. 1) |
| (6) | Fit candidate statistical models to the data (e.g., GLM, GAM, Delta-lognormal GLM, Neural Networks, Regression Trees, Habitat based models, and Statistical habitat based models); | Yes (GLM) |
| (7) | Evaluate the models using methods such as likelihood ratio, AIC/BIC and cross validation; | Yes (AIC and cross validation) |
| (8) | Evaluate if distributional assumptions are satisfied and if there is a significant spatial/temporal pattern of residuals in CPUE standardization modeling; | Yes (*see* Fig. 3) |
| (9) | Extract yearly standardized CPUE and standard error by a method that is able to account for spatial heterogeneity of effort, such as least squares mean or expanded grid. If the model includes area and the size of spatial strata differs or the model includes interactions between time and area, then standardized CPUE should be calculated with area weighting for each time step. Model with interactions between area and season or month requires careful consideration on a case by case basis; | Yes (*see* Method and Fig. 4) |
| (10) | Recommend a time series of yearly standardized CPUE and associated uncertainty; | Yes (*see* Table 3) |
| (11) | Plot nominal and standardized CPUEs over time; | Yes (*see* Fig. 4) |

**Appendix II** Area definition that applied to area categorization. Area1 is defined by combining subareas of area2 with 1(1, 2, 3 of area2), 2(4, 5 of area2) and 3(7 of area2)

