NPFC-2021-SSC PS07-WP13 (Rev. 1)

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

Chuanxiang Hua1, Siquan Tian1, Yongchuang Shi2, Luoliang Xu1

1College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China

2East China Sea Fisheries Research Institute, Shanghai 200090, China

**Summary**

China’s Pacific saury fishery in the Northwest Pacific Ocean began in 2003. The main fishing gear of Pacific saury fishery is the stick-held dip net. In this paper, catch per unit fishing effort (CPUE) was standardized using generalized linear model (GLM) and generalized additive model (GAM). Four groups of independent variables were considered in the CPUE standardization: spatial variables (Latitude and Longitude), temporal variables (Year and Month), vessel length and environmental variables (SST and SSTG). Log-CPUE was treated as the dependent variable and its error was assumed to follow normal distribution in each model. The model selections of GLM and GAM were based on the Bayesian information criterion (BIC). From the results, Higher Spearman’s correlation and lower mean squared error were observed by GAM. Hence, we prefer to choose the best GAM model to estimated standardized CPUE of Pacific saury.

1. **Background of the Pacific saury fishery**

Pacific saury(*Cololabis saira*) is a highly migratory fish, widely distributed in the Northwest Pacific Ocean (NPO) (Lin, 2003; Sun et al., 2003). At the beginning of the 20th century, the first stick-held net fishing vessel (changed from squid jigging vessel) from China went to the high seas for fishing Pacific saury in the NPO. Now, about 50 Pacific Saury vessels from China operate in the NPO.

**2 Method**

**2.1 The data**

Commercial fishery data of Pacific saury were from 2013 to 2020, which were derived from Pacific Saury Fishery Technical Working Group, Distant-water Fishery Society of China. Distribution of catch (ton) and fishing effort for China Pacific saury fishing fleets in the NPO from 2013 to 2020 were shown in Figure 1.

The Pacific saury is a highly migratory fish, and the distribution of its fishing grounds shows variation during the fishing period (June-November) each year (Tian, 2003); therefore, temporal variables (Year and Month), spatial variables (Longitude and Latitude) were included in the analysis. The formation of the Pacific saury fishing grounds is tightly associated with the marine environment condition (Zhu, 2006). Thus, the Sea surface temperature(SST) and Sea surface temperature gradients (SSTG) were included in the analysis. In addition, the vessel length may affect the catch capacity.

SST data were derived from National Oceanic and Atmospheric Administration (NOAA; [ftp.nodc.noaa.gov](ftp://ftp.nodc.noaa.gov)). The spatial-temporal resolution of the SST data is daily at 0.05°×0.05° grid. SSTG data were calculated by Gradient Magnitude (GM) method (Ortiz, 2004; Howell, 2006). The formula is:



where , , and are SST values of 4 consecutive grids respectively, *i* and *j* is the numbering of row and column, is the longitudinal distance (km) between (*j*-1)th and (*j*+1)th columns, is the latitudinal distance (km) between (*i*-1)th and (i+1)th rows, is SSTG value of the current grid (°C/km).

This study extracted the corresponding oceanographic data from the nearest grid to the grid where the fishery data in the same day. Nominal CPUE was defined as catch per day per vessel, unit: ton/day/v.

Summary of explanatory variables used for CPUE standardization were listed in the table 1. *Year* is a categorical variable of 8 years (2013-2020). *Month* is a categorical variable including the eight calendar months from May to December. *Longitude* and *Latitude* are categorical variables, which divided at intervals of 1°. We attempted two cases (categorical and splined variable) for *Sst* and investigated splined variable for *Sstg*. *Vessellength* is a categorical or continuous variable of 60-76 m vessels, which will affect the catchability (Table1).

Variance Inflation Factor (VIF) and Spearman correlation coefficient among explanatory variables were calculated (Table 2) and correlations among variables were shown in the Figure 2.

**2.2 Full model description and model selection**

Both generalized linear model (GLM) and generalized additive model (GAM) were used to standardize the CPUE.

The full GLM model was:

*log(CPUE) =Year + Month + Longitude\_c + Latitude\_c + Sst + Sstg* *+ Vessellength\_c + interaction + ε*

The full GAM model was:

*log(CPUE)*= *Year* + *Month* + *Longitude\_c* + *Latitude\_c + s(Sst) + s(Sstg) +* *s(Vessel length) + interaction + ε*

whereis the residual, which is assumed to have a normal distribution. *interaction* is an interaction term representing the interactive effect of spatial and temporal factors for the Pacific saury. Full model interaction includes all the possible combination of *Year, Month, Longitude\_c, Latitude\_c.*

The optimal model was selected using the Bayesian information criterion (BIC). Spearman’s correlation between the predicted and observed CPUEs and mean of squared errors between two CPUEs were calculated to evaluate prediction performance.

**2.3 Yearly trend extraction**

The way to calculate the standardization CPUE is the yearly mean of fitted CPUE from the best model. The formula is,

where, is CPUE indices in *i*th year, is the observation number in *i*th year, is the *k*th fitted CPUE data in *i*th year.

The bootstrapped 95% confidence intervals of Standardized CPUE of the optimal GLM and GAM were calculated.

**3 Result and discussion**

In this study we used two models to standardize the CPUEs. VIF and Spearman correlation coefficient among explanatory variables were calculated (Table2). The Maximum VIF<5, indicates that there is no serious multi-collinearity (Tien, 2011). Residuals from both approaches showed an approximately normal distribution around 0, which indicated that the model assumptions were satisfied. The results were shown in Figure 3.

We used same explanatory variables in GLM and GAM analysis (Table 1). The results of the GLM and GAM model selections are shown in Table 3 and Table 6, respectively. The summary of fitting a GLM for the optimal model is shown in Table 4. All explanatory variables are highly significant (*p*<0.01). The summary of fitting a GAM for the best model is shown in Table 7. All explanatory variables are highly significant (*p*<0.01).

Table 9 and Figure 4 shows the annual changes of nominal CPUE and standardized CPUE by GAM and GLM models. There are few differences between fitted CPUEs data by GLM and GAM, which may be related to the assumption of relationships between CPUEs and explanatory variables.

Comparing the results of cross validation tests in GLM and GAM analyses (Table 5 and 8), higher Spearman’s correlation and lower mean squared error (MSE) between observed and predicted of test data were observed by GAM, so we prefer to choose the best GAM model to estimate standardized CPUE.

We standardized CPUE in accordance with the standardization protocol (NPFC - 2017 - TWG PSSA - Report Annex D). The checklist is shown in Appendix 1.

**References**

Howell E A, Kobayashi D R. 2006. El Niño effects in the Palmyra Atoll region: oceanographic changes and bigeye tuna (*Thunnus obesus*) catch rate variability. Fisheries Oceanography, 15(2):477-489.

Lin Longshan. 2003. Fishery survey of Stick-held Net for Cololabis saira in Taiwan. Marine Fisheries (in Chinese), (4): 200-203.

ORTIZ M, AROCHA F. Alternative error distribution models for standardization of catch rates of non-target species from a pelagic longline fishery: billfish species in the Venezuelan tuna longline fishery[J]. Fisheries Research, 2004, 70(1): 275-294.

Sun Manchang, Ye Xuchang, Zhang Jian, et al. 2003. Probe into Pacific saury fisheries in the northwest Pacific Ocean. Marine Fisheries (in Chinese), 25(3): 112-115.

Tian Y J. Variations in the abundance of Pacific saury (*Cololabis saira*)from the Northwestern Pacific in relation to oceanic—climate changes[J]. Fisheries Research, 2003, 60:439-454.

Tien Bui, D., Lofman, O., Revhaug, I., Dick, O. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. Nat. Hazards ,2011,59 (3), 1413-1444.

ZHU Guo-ping, ZHU Qing-cheng. Preliminary study on relationship between Cololabis saira fishing ground and temperature factor in the Northern Pacific Ocean[J]. Marine Sciences, 2006, 30(7):91-96.

**APPENDICES**

Appendix1. Checklist for the CPUE standardization protocol

|  |  |  |
| --- | --- | --- |
| (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* 2.1 The data paragraph 2) |
| (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* table 2 and Fig.2) |
| (5) | Identify potential explanatory variables based on (1)-(4) to develop full model for the CPUE standardization; | Yes |
| (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 and GAM) |
| (7) | Evaluate the models using methods such as likelihood ratio, AIC, BIC or cross validation; | Yes (*see* Table3 and Table6) |
| (8) | Evaluate if distributional assumptions are satisfied and if there is a consistent spatial/temporal distribution 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*2.3 Yearly trend extraction**)** |
| (10) | Recommend a time series of yearly standardized CPUE and associated uncertainty; | Yes (*see* Table 9) |
| (11) | Plot nominal and standardized CPUEs over time.  Overall remarks Recommendations | Yes (*see* Fig. 4) |

**Tables**:

Table 1 Summary of explanatory variables used for GLM and GAM analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Cases | Categorical or continuous | Details | Note |
| Year | *Year* | 8 categories | 8 years from 2013 to2020 |  |
| Month | *Month* | 8 categories | 8 months from May to December |  |
| Longitude | *Longitude\_c* | 25 categories | Longitude<144° ; 144°≦Longitude＜145°，145°≦Longitude＜146，…, Longitude>165° | at intervals of 1° |
| Latitude | *Latitude\_c* | 13 categories | Latitude<38° ; 38°≦Latitude＜39°，39°≦Latitude＜40°，…, Latitude >48° | at intervals of 1° |
| Sea surface temperature | *Sst*  *Sst\_c* | spline  12 categories | Sst<10℃;10℃≦Sst＜11℃，11℃≦Sst＜12℃，…, 19℃≦Sst≤20℃; Sst>20℃ | at intervals of 1℃ |
| Sea surface temperature gradient | *Sstg* | continues（spline） |  |  |
| Vessel length | *Vessellength*  *Vessellength\_c* | continues（spline）  9 categories | Vessellength＜64m，64m≦Vessellength＜76m，…, 76m≦Vessellength | at intervals of 2m |

Table 2 Variance Inflation Factor (VIF) and Spearman correlation coefficient among explanatory variables

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| coefficient/p value | VIF | Year | Month | Longitude | Latitude | SST | SSTG | vessellength |
| Year | 1.42 |  | <0.001 | <0.001 | <0.001 | 0.3238 | <0.001 | <0.001 |
| Month | 3.43 | -0.042 |  | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| Longitude | 4.09 | 0.331 | -0.799 |  | <0.001 | <0.001 | <0.001 | <0.001 |
| Latitude | 1.64 | -0.047 | -0.490 | 0.503 |  | <0.001 | <0.001 | <0.001 |
| SST | 1.38 | 0.006 | 0.464 | -0.399 | -0.440 |  | <0.001 | 0.056 |
| SSTG | 1.29 | -0.170 | 0.339 | -0.423 | -0.377 | 0.205 |  | <0.001 |
| Vessellength | 1.02 | 0.131 | -0.029 | 0.063 | 0.026 | 0.011 | -0.033 |  |

1. Spearman correlation coefficient are under the slope line; *p* values are above the slope line.

Table 3 Result of GLM model selection

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | GLM model | R2 | BIC | Explained deviance（%） |
| 1 | *Ln*(CPUE)~*Intercept+Year+Month+Longitude\_c+Latitude\_c+Sst+Sstg+Vessellength\_c* | 0.4380 | 84741.85 | 46.81% |
| 2 | *Ln*(CPUE)~*Intercept+Year+Month+Longitude\_c+Latitude\_c+Sst+Sstg+Vessellength\_c*+*Year:Month* | 0.4057 | 84616.41 | 41.84% |
| 3 | *Ln*(CPUE)~*Intercept+Year+Month+Longitude\_c+Latitude\_c+Sst+Sstg+Vessellength\_c*+*Year:Month+ Year: Longitude\_c* | 0.3804 | 83679.20 | 37.96% |
| 4 | *Ln*(CPUE)~*Intercept+Year+Month+Longitude\_c+Latitude\_c+Sst+Sstg+Vessellength\_c*+*Year:Month+ Year: Latitude\_c* | **0.3844** | **83186.63** | **38.56%** |
| 5 | *Ln*(CPUE)~*Intercept+Year+Month+Longitude\_c+Latitude\_c+Sst+Sstg+Vessellength\_c*+*Year:Month+ Year: Longitude\_c*+ *Year: Latitude\_c*+ *Month: Longitude\_c*+ *Month: Latitude\_c*+ *Longitude\_c: Latitude\_c* | 0.3506 | 85825.12 | 33.44% |

Table 4 Anova test for best GLM model

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Df | Deviance | Resid. Df | Resid. Dev | *F* | Pr(>*F*) |  |
| NULL |  |  | 28106 | 47980.09 |  |  |  |
| factor(Year) | 7 | 4903.53 | 28099 | 43076.56 | 656.73 | < 2.2E-16 | \*\*\* |
| factor(Month) | 7 | 7943.09 | 28092 | 35133.46 | 1063.82 | < 2.2E-16 | \*\*\* |
| factor(Longitude\_c) | 24 | 1410.63 | 28068 | 33722.84 | 55.10 | < 2.2E-16 | \*\*\* |
| factor(Latitude\_c) | 12 | 529.97 | 28056 | 33192.86 | 41.40 | < 2.2E-16 | \*\*\* |
| Sst | 1 | 134.13 | 28055 | 33058.73 | 125.75 | < 2.2E-16 | \*\*\* |
| Sstg | 1 | 9.43 | 28054 | 33049.30 | 8.84 | 0.0029 | \*\* |
| factor(Vessellength\_c) | 6 | 222.43 | 28048 | 32826.87 | 34.75 | < 2.2E-16 | \*\*\* |
| factor(Year):factor(Month) | 42 | 2321.98 | 28006 | 30504.89 | 51.83 | < 2.2E-16 | \*\*\* |
| factor(Year):factor(Latitude\_c) | 72 | 709.10 | 27934 | 29795.79 | 9.23 | < 2.2E-16 | \*\*\* |

Significant code: \*\*\* 0.001, \*\*0.01, \*0.05

Table 5 The Five-fold cross validation for the best GLM.

|  |  |  |
| --- | --- | --- |
| case | cor\_GLM\_test | MSE\_GLM\_test |
| 1 | 0.5976 | 1.0909 |
| 2 | 0.5983 | 1.0852 |
| 3 | 0.5896 | 1.0849 |
| 4 | 0.5945 | 1.0906 |
| 5 | 0.6007 | 1.0723 |

The spearman’s correlation coefficient is showed in the table.

Table 6 Result of GAM model selection

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | GAM model | R2 | BIC | Explained deviance（%） |
| 1 | *Ln(CPUE)~Intercept+Year+Month+Longitude\_c+Latitude\_c+s(Sst)+s(Sstg)+s(Vessellength)* | 0.3416 | 83734.54 | 34.33% |
| 2 | *Ln(CPUE)~Intercept+Year+Month+Longitude\_c+Latitude\_c+s(Sst)+s(Sstg)+s(Vessellength)+Year:Month* | 0.3504 | 83859.6 | 35.33% |
| 3 | *Ln(CPUE)~Intercept+Year+Month+Longitude\_c+Latitude\_c+s(Sst)+s(Sstg)+s(Vessellength)*+*Year:Month+ Year: Longitude\_c* | 0.4023 | 82662.75 | 40.76% |
| 4 | *Ln(CPUE)~Intercept+Year+Month+Longitude\_c+Latitude\_c+s(Sst)+s(Sstg)+s(Vessellength)*+*Year:Month+ Year: Latitude\_c* | **0.3983** | **82234.32** | **40.22%** |
| 5 | *Ln(CPUE)~Intercept+Year+Month+Longitude\_c+Latitude\_c+s(Sst)+s(Sstg)+s(Vessellength)*+*Year:Month+ Year: Longitude\_c*+ *Year: Latitude\_c*+ *Month: Longitude\_c*+ *Month: Latitude\_c*+ *Longitude\_c: Latitude\_c* | 0.4300 | 84830.02 | 44.28% |

Table 7 Anova test for best GAM model

Parametric Terms:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | df | *F* | P-value |  |
| factor(Year) | 7 | 5.63 | 1.63E-06 | \*\*\* |
| factor(Month) | 7 | 8.99 | 4.09E-11 | \*\*\* |
| factor(Longitude\_c) | 24 | 18.36 | < 2.2E-16 | \*\*\* |
| factor(Latitude\_c) | 12 | 2.87 | 0.00058 | \*\*\* |
| factor(Year):factor(Month) | 44 | 32.28 | < 2.2E-16 | \*\*\* |
| factor(Year):factor(Latitude\_c) | 77 | 8.96 | < 2.2E-16 | \*\*\* |

Approximate significance of smooth terms:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Edf | Ref.df | *F* | p-value |  |
| s(Sst) | 7.94 | 8.58 | 11.67 | < 2.2E-16 | \*\*\* |
| s(Sstg) | 4.82 | 5.83 | 5.93 | 6.85E-06 | \*\*\* |
| s(Vessellength) | 8.85 | 8.99 | 134.41 | < 2.2E-16 | \*\*\* |

Significant code: \*\*\* 0.001, \*\*0.01, \*0.05

Table 8 The cross validation for the best GAM.

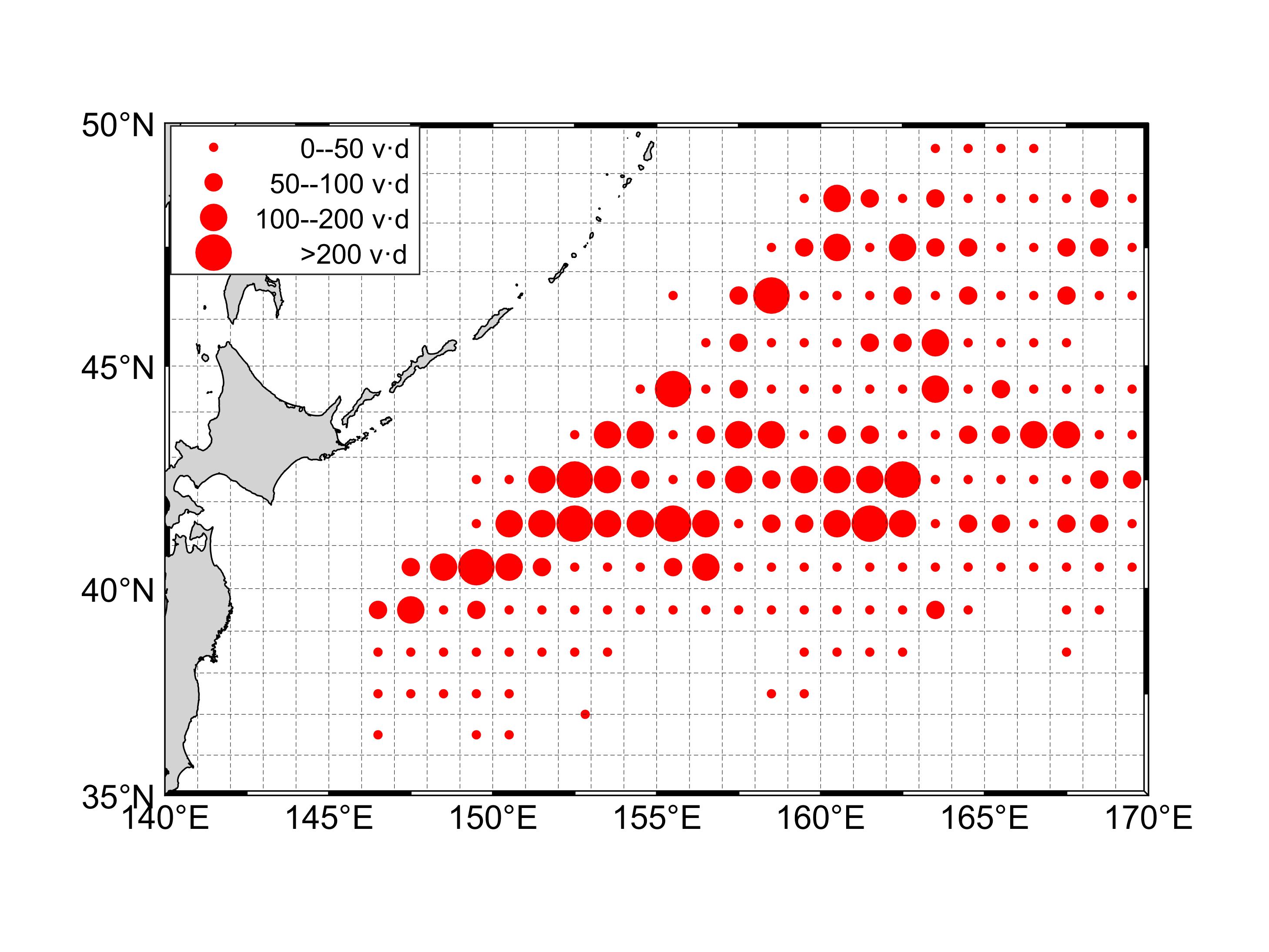
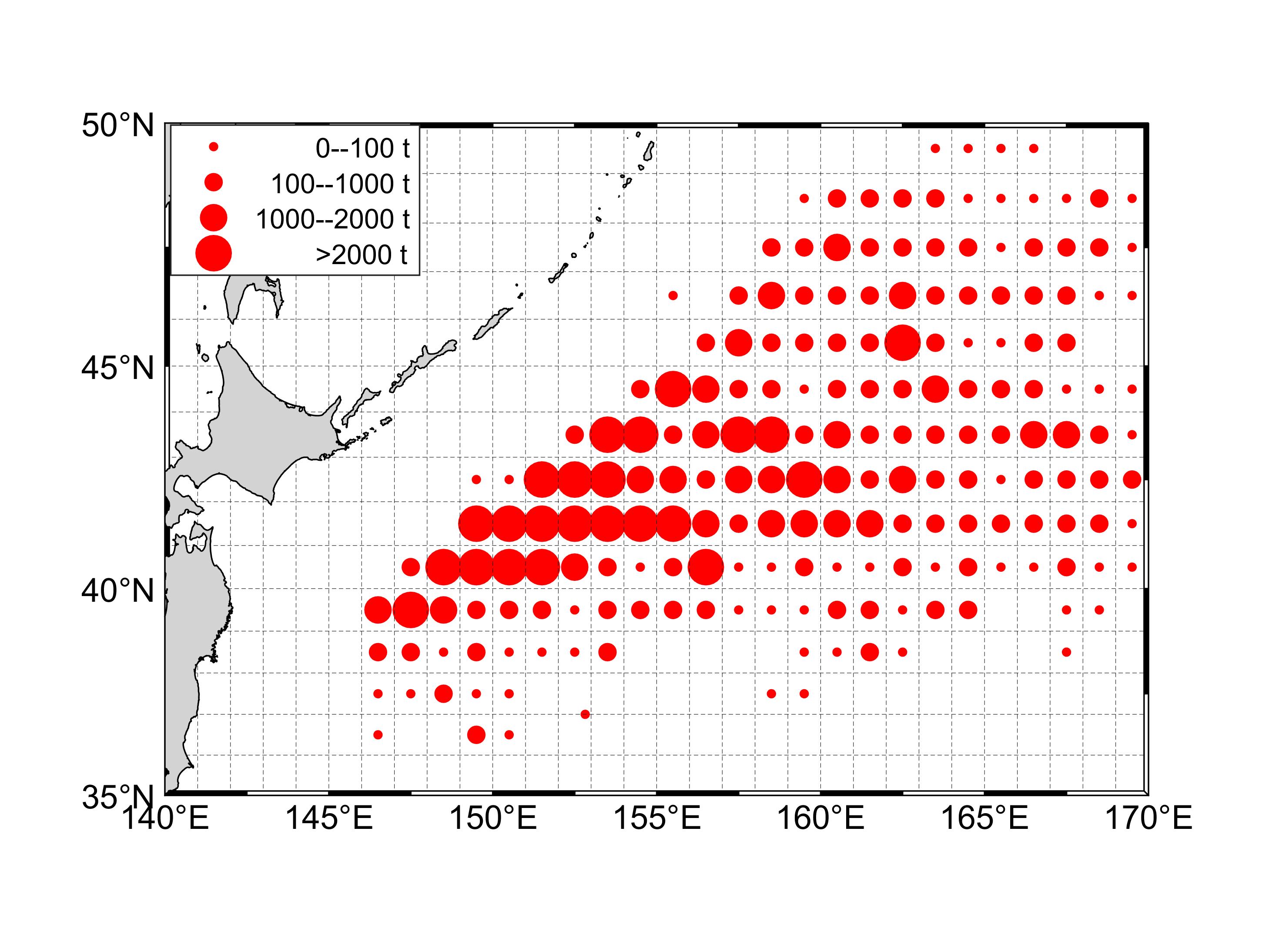
|  |  |  |
| --- | --- | --- |
| case | cor\_GAM\_test | MSE\_GAM\_test |
| 1 | 0.6197 | 1.0342 |
| 2 | 0.6091 | 1.0747 |
| 3 | 0.6120 | 1.0541 |
| 4 | 0.6154 | 1.0746 |
| 5 | 0.6156 | 1.0343 |

The spearman’s correlation coefficient is showed in the table.

Table 9 Nominal and standardized CPUE from 2013 to 2020.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | Nominal CPUE | SD of Nominal CPUE | Standardized CPUE by GLM | SD by GLM | 95% CI by GLM | | Standardized CPUE by GAM | SD by GAM | 95% CI by GAM | |
| 2013 | 20.80 | 19.17 | 13.81 | 4.78 | [13.48 | 14.18] | 13.96 | 5.42 | [13.53 | 14.41] |
| 2014 | 22.11 | 20.62 | 15.85 | 8.27 | [15.49 | 16.20] | 16.24 | 9.29 | [15.82 | 16.60] |
| 2015 | 23.48 | 21.21 | 17.14 | 10.25 | [16.56 | 17.70] | 17.73 | 11.06 | [17.10 | 18.32] |
| 2016 | 15.02 | 18.87 | 9.09 | 6.32 | [8.84 | 9.32] | 9.29 | 6.91 | [9.04 | 9.56] |
| 2017 | 12.12 | 12.82 | 8.40 | 4.16 | [8.22 | 8.53] | 8.50 | 4.50 | [8.34 | 8.68] |
| 2018 | 23.13 | 24.48 | 15.59 | 10.91 | [15.13 | 16.01] | 15.84 | 11.49 | [15.38 | 16.31] |
| 2019 | 10.78 | 12.99 | 6.78 | 3.38 | [6.66 | 6.89] | 6.89 | 3.69 | [6.75 | 7.05] |
| 2020 | 9.51 | 14.09 | 5.84 | 4.08 | [5.53 | 6.07] | 5.95 | 4.37 | [5.48 | 6.16] |

**Figures**:



1. (b)

Fig. 1 Distribution of catch(a) and fishing effort(b) for China Pacific saury fishing fleets in the Northwestern Pacific Ocean from 2013 to 2020.

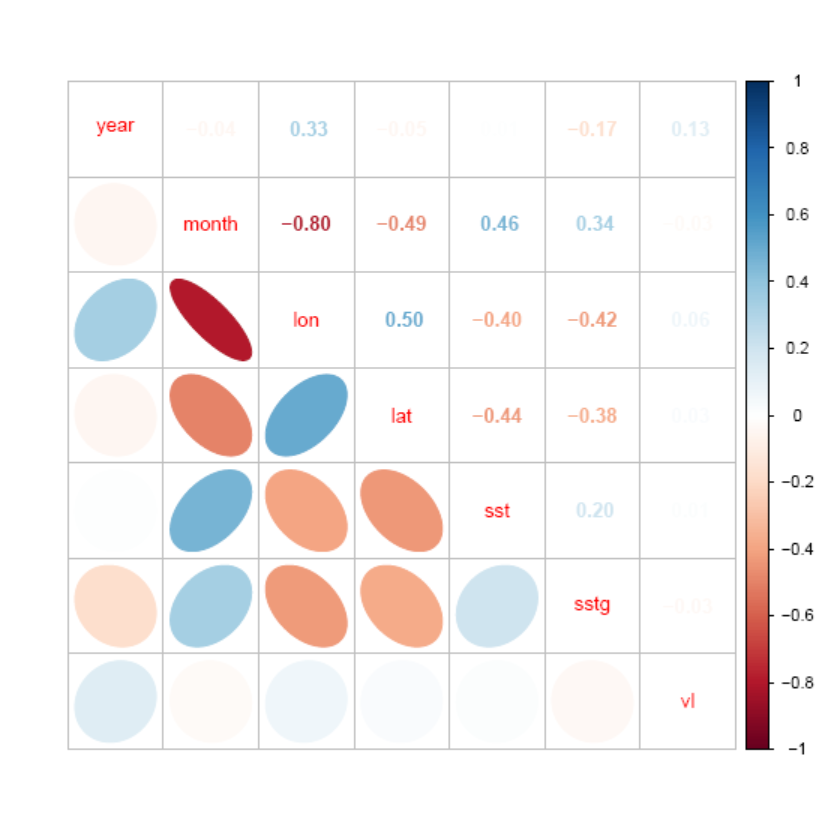


Fig. 2 Correlation matrix of explanatory variables used in the analysis

(a)



(b)

Fig. 3 Normal distribution checks, Q-Q plot and histogram of residuals for the GLM(a) and GAM(b) optimal model.



Fig.4 Annual changes in nominal, GAM and GLM estimated standardized CPUE up to 2020.