North Pacific Fisheries Commission

NPFC-2022-TWG CMSA05-WP09

Standardized CPUE of Chub mackerel (*Scomber japonicas*) caught by the China's lighting purse seine fishery up to 2020

Yongchuang Shi[1], Heng Zhang[1], Qiuyun Ma[2], Guoqing Zhao[1], Chao Yang[1],

[1]Key and Open Laboratory of Oceanic Laboratory of Remote Sensing Information Technology in Fishing Resource, East China Sea Fisheries Research Institute, Chinese Academy of Fishery Science. Shanghai 200090, China

[2]College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China

Summary

Chub mackerel (*Scomber japonicas*) is key economic and ecological specie in the Northwest Pacific Ocean. Most of the Chub mackerel catch is harvested by the lighting purse seine fishery in China. In this paper, we standardized catch per unit fishing effort (CPUE) 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 Chl-a). 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 BIC. From the results, Higher Spearman's correlation and lower mean squared error were observed by GAM. Besides, the standardized CPUE trend of GAM model is similar with that of nominal CPUE. Therefore, we prefer to choose the best GAM model to estimated standardized CPUE of Chub mackerel fishery.

1. Background of the Chub mackerel fishery

Chub mackerel (*Scomber japonicas*) is a highly migratory fish, widely distributed in the high seas of the Northwest Pacific Ocean (Yukami., 2009). The annual catches of Chub mackerel recorded in 2020 were about 61,166 tons in China, which accounted for 16.70% of the global production. Now, about 50 Chub mackerel vessels from China operate in the Northwest Pacific Ocean. The main fishing area of China is shown in Figure 1.

2 METHOD

2.1 The data

Full-commercial fishery data were from 2014 to 2020, which were derived from Technical Group for Chub mackerel Fishery, Distant-water Fishery Society of China. Distribution of catch (ton) and fishing effort for China Chub mackerel fishing fleets in the Northwestern Pacific Ocean from 2014 to 2020 was shown in Figure 2. The catch of Chub mackerel in region 146—155°E and 39—44°N is higher than other regions (Fig.2a).

The Chub mackerel is a highly migratory fish, and the distribution of its fishing grounds shows large variation during the fishing period (April–November) each year (Yatsu, 2002), therefore, temporal variables (year and month), spatial variables (longitude and latitude) were included in the analysis. The distribution of the Chub mackerel fishing grounds is tightly associated with the marine environment (Zhang, 2009). Thus, Sea surface temperature (SST) and Chlorophyll-a concentration (Chla) were included in the analysis. In addition, the vessel length may affect the quantity of the catch, which was included in this study.

SST data were derived from National Oceanic and Atmospheric Administration (NOAA; ftp.nodc.noaa.gov). The spatial-temporal resolution of the SST data is daily at 0.1°×0.1° grid. The monthly Chla data from the Moderate Resolution Imaging Spectroradiometer (MODIS) on board the satellite Aqua platform provided by Asia-Pacific Data Research Center were used for this study. The Chla data was from website: http://apdrc.soest.hawaii.edu/data/data.php.

This study extracted the corresponding oceanographic data from the nearest grid to the grid where the fishery data existed at the same date. Nominal CPUE were 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 7 years (2014—2020). *Month* is a categorical variable including the nine calendar months from April 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 *Chla*. *Vessellength* is a categorical or continuous variable of 44—61 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 3.

2.2 Full model description and model selection

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

The full GLM model was:

log(CPUE) =Year + Month + Longitude_c + Latitude_c + SST + Chla +Vessellength_c + interaction+ε

The full GAM model was:

log(CPUE)= Year+ Month+ longitude_c + latitude_c + s(SST) + s(Chla) + s(vessel length) +interaction+ε

where ε is 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 Chub mackerel. 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,

$$\overline{CPUE_i} = \frac{1}{n_i} \times \sum_{k=1}^{n_i} CPUE_k^{fitted}$$

where, $\overline{CPUE_i}$ is CPUE indices in *i*th year, n_i is the observation number in *i*th year, $CPUE_k^{fitted}$ 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. Variance Inflation Factor (VIF) and Spearman correlation coefficient among explanatory variables were calculated (Table1). The

Maximum VIF<5, indicates 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 4 and Figure 5.

We used same explanatory variables in GLM and GAM analysis (Table 1).The result of the best GLM and GAM models 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) except for Chla. The summary of fitting a GAM for the best model is shown in Table 7. All explanatory variables are highly significant (p<0.01) except for Chla.

Table 9 and Figure 6 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. The checklist is shown in Appendix 1.

REFERENCES

Yukami R, Ohshimo S, Yoda M, Hiyama Y, 2009. Estimation of the spawning grounds of chub mackerel *Scomber japonicus* and spotted mackerel *Scomber australasicus* in the East China Sea based on catch statistics and biometric data. Fish. Sci. 75(1), 167-174.

Yatsu A, Mitani T, Watanabc C, *et al*., 2002. Current stock status and management of Chub mackerel *Scomber japonicus* along the Pacific coast of Japan. Fisheries science, 68(supl):93-96.

Zhang G W, Chen X J, Li G, 2009. Bio-economic model and its application of Chub mackerel in the East China Sea and Yellow Sea. Journal of Shanghai Ocean University, 18(4):447-452.

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

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.2)
(4)	Calculate correlation matrix to evaluate correlations between each pair of those variables;	Yes (*see* table 1 and Fig.3)
(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.5 and Fig.6)
(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. 6)

Tables:

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

Variables	Cases	Categorical or continuous	Details	Note
Year	*Year*	7 categories	7 years from 2014 to2020	
Month	*Month*	9 categories	9 months from April to December	
Longitude	*Longitude_c*	25 categories	Longitude<144° ; 144°≤Longitude＜145°; 145°≤Longitude＜146;…, Longitude>165°	at intervals of 1°
Latitude	*Latitude_c*	18 categories	Latitude<33°; 33°≤Latitude＜34°; 34°≤Latitude＜35; …, 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℃
Chlorophyll-a concentration	*Chla*	continues（spline）		
Vessel length	*Vessellength* *Vessellength_c*	spline 10 categories	Vessellength＜44m; 44m≤Vessellength＜46m; …, Vessellength≥60m	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	Vessellength	SST	Chla
Year	1.09		0.1926	0.2354	<0.001	<0.001	<0.001	<0.001
Month	1.31	-0.042		<0.001	<0.001	0.6481	<0.001	<0.001
Longitude	2.92	0.331	-0.799		<0.001	<0.001	<0.001	<0.001
Latitude	3.06	-0.047	-0.490	0.503		<0.001	<0.001	<0.001
Vessellength	1.01	0.006	0.464	-0.399	-0.440		0.2350	0.4574
SST	1.28	-0.170	0.339	-0.423	-0.377	0.205		<0.001
Chla	1.91	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 The best GLM model

GLM model	R^2	BIC	Explained deviance (%)
Ln(CPUE)~*Intercept+Year+Month+Longitude_c+Latitude_c+SST +Chla +Year:Month*	**0.3753**	**99421.86**	**60.59%**

Table 4 Anova test for best GLM model

	Df	Deviance	Resid. Df	Resid. Dev	*F*	Pr(>*F*)	
NULL			33416	46828.89			
factor(Year)	6	5450.20	33410	41378.69	794.2872	< 2.2E-16	***
factor(Month)	8	1729.77	33402	39648.91	189.0669	< 2.2E-16	***
factor(Longitude_c)	24	326.02	33378	39322.89	11.87804	< 2.2E-16	***
factor(Latitude_c)	17	112.07	33361	39210.83	5.764192	2.2E-13	***
SST	1	3.81	33360	39207.02	3.3343	0.0006	***
Chla	1	0.80	33359	39206.21	0.702344	0.0082	**
factor(Year):factor(Month)	44	1106.32	33315	38099.89	21.98598	< 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	0.5743	1.2492
2	0.5819	1.3579
3	0.5173	1.1074
4	0.5528	1.2497
5	0.5649	1.0975

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

Table 6 The best GAM model

GAM model	R^2	BIC	Explained deviance (%)
Ln(CPUE)~Intercept+Year+Month+Longitude_c+Latitude_c+s(SST) +s(Chla)+ Year:Month	**0.3725**	**99408.76**	**48.72%**

Table 7 Anova test for best GAM model

Parametric Terms:

	df	*F*	P-value	
factor(Year)	6	90.06098	< 2.2E-16	***
factor(Month)	8	7.798984	1.62E-10	***
factor(Longitude_c)	24	5.701314	< 2.2E-16	***
factor(Latitude_c)	17	5.180736	1.42E-11	***
factor(Year):factor(Month)	46	21.45609	< 2.2E-16	***

Approximate significance of smooth terms:

	Edf	Ref.df	*F*	p-value	
s(SST)	3.74	4.73	4.09	0.00013	***
s(Chla)	4.56	5.56	2.06	0.0359	**

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	0.6891	1.0578
2	0.6147	1.1247
3	0.6589	0.9957
4	0.6765	1.0629
5	0.6367	0.9874

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

Table 9 Nominal and standardized CPUE from 2014 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
2014	22.59	13.68	19.24	11.14	[18.43 22.34]	19.77	8.91	[18.54 21.12]
2015	18.61	11.91	15.49	6.81	[14.75 16.51]	16.38	5.48	[14.67 16.75]
2016	16.41	10.28	14.58	3.70	[13.83 14.92]	14.97	3.67	[14.64 15.32]
2017	15.47	9.99	13.31	5.65	[12.48 14.53]	14.15	5.55	[13.65 14.52]
2018	16.48	10.05	14.08	4.05	[13.59 15.12]	14.72	3.72	[14.07 15.07]
2019	19.24	12.43	17.57	7.69	[16.42 18.24]	18.03	6.93	[17.38 18.57]
2020	14.98	8.97	12.38	9.59	[11.05 13.17]	13.05	8.61	[12.68 13.49]

Figures:

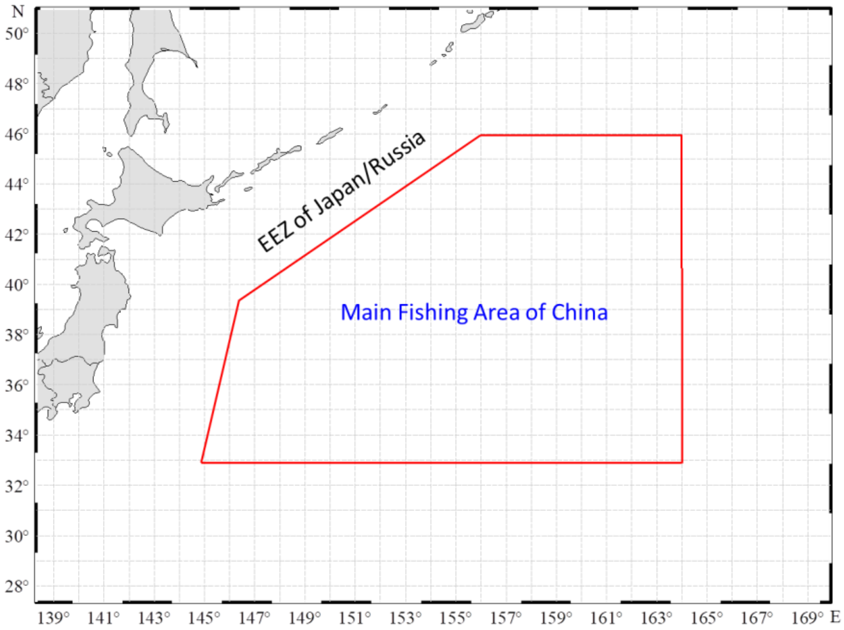
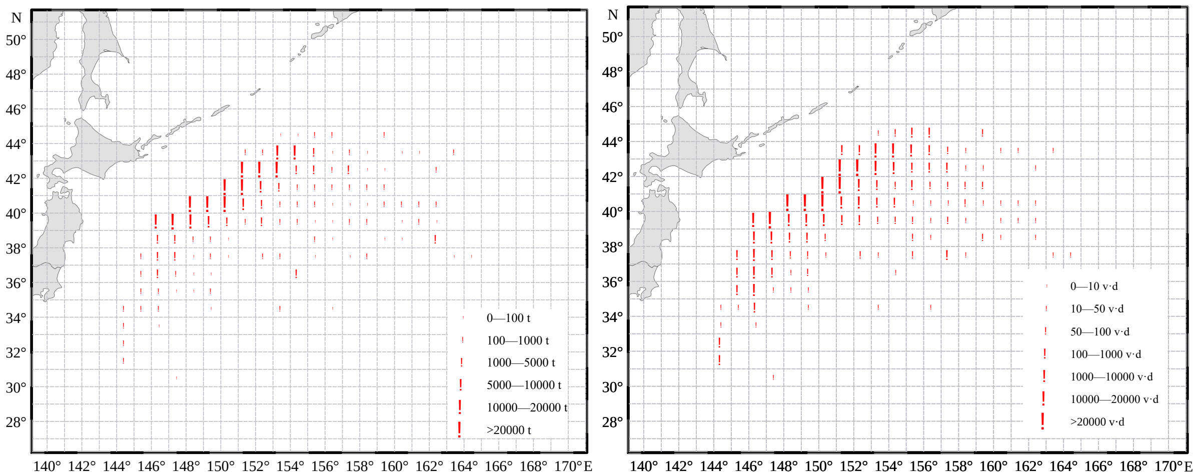


Fig. 1 The main fishing area of China



(a) (b)

Fig. 2 Distribution of catch (a) and fishing effort(b) for China Chub mackerel fishing fleets in the Northwestern Pacific Ocean from 2014 to 2020

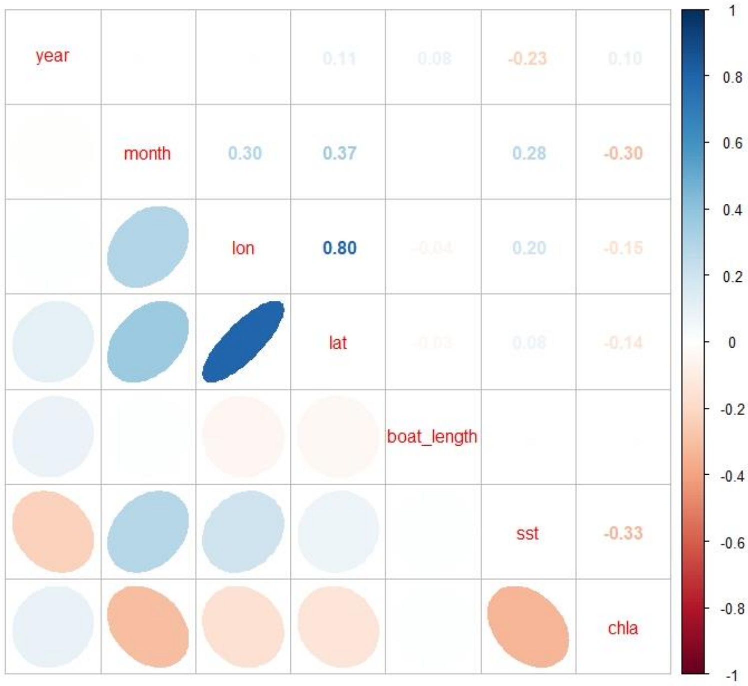


Fig. 3 Correlation matrix of explanatory variables used in the analysis

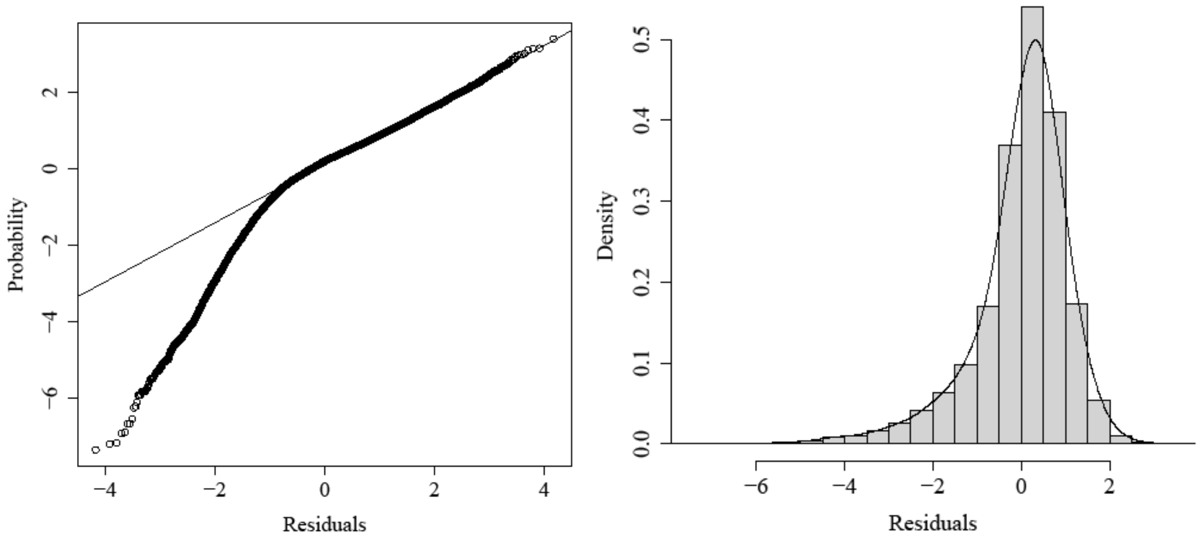


Fig. 4 Normal distribution checks, Q-Q plot and histogram of residuals for the GLM optimal model.

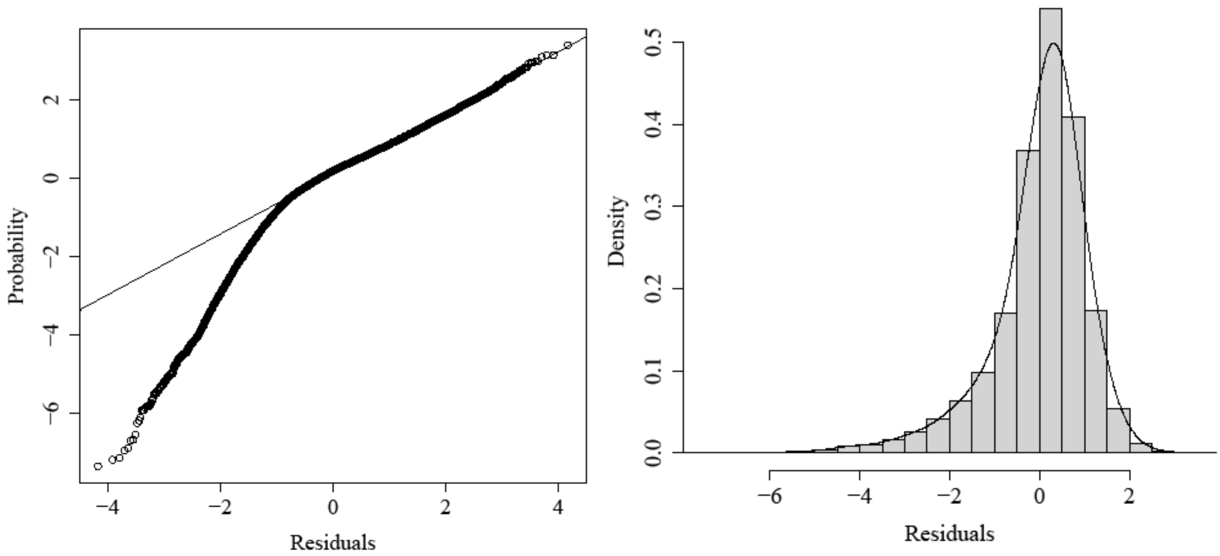


Fig. 5 Normal distribution checks, Q-Q plot and histogram of residuals for the GAM optimal model.

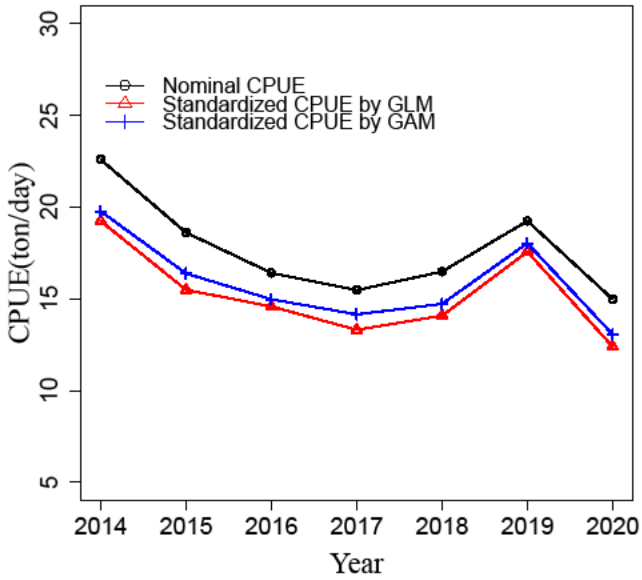


Fig.6 Annual changes in nominal, GAM and GLM estimated standardized CPUEs.