North Pacific Fisheries Commission

NPFC-2024-TWG CMSA08-WP07 (Rev.1)

Standardized CPUE of Chub mackerel (*Scomber japonicas*) caught by the China's lighting purse seine fishery up to 2022

Yongchuang Shi, Heng Zhang, Haibin Han

East China Sea Fisheries Research Institute, Chinese Academy of Fishery Science. Shanghai 200090, China

Summary

Catch per unit fishing effort (CPUE) standardization is an important approach to obtaining accurate indices of resource abundance by removing the influence of external factors. Chub mackerel (*Scomber japonicas*) is an economically important small pelagic fish inhabiting 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 CPUE of Chub mackerel 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), fishery variables (vessel length and proportion of chub mackerel) and environmental variables (SST and Chla). 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 (MSE) were observed by GAM. Therefore, we prefer to choose the best GAM model to estimate 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 (Yatsu et al., 2005). The annual catches of Chub mackerel recorded in 2022 were about 81,181 tons in China, which accounted for about 30% of the global production. Now, about 100 Chub mackerel vessels from China operate in the Northwest Pacific Ocean. The distribution of Chub mackerel fishing grounds shows large variation during the fishing period (April–November) each year (Yatsu et al., 2002), therefore, temporal variables (year and month), spatial variables (longitude and latitude) were included in the analysis. The fishing ground of the Chub mackerel is tightly associated with the marine environment (Zhang et al., 2009). Thus, Sea surface temperature (SST) and Chlorophyll-a concentration (Chla) were included in the analysis. In addition, the vessel length and the proportion of chub mackerel may affect the quantity of the catch, which were also included in this study.

2. Method

2.1. The Data

Full-commercial fishery data (logbook) were from 2014 to 2022, which were derived from Technical Group for Chub mackerel Fishery, Distant-water Fishery Society of China. The catch and effort of CPUE Fleet were aggregated by monthly at 1°×1° grid, with good representativeness of the whole fishery (Table 1). The Table 2 represents the filter "rules" used on data for CPUE standardization and the effect on the overall sample size. Annual spatial distribution patterns of catch, effort and nominal CPUE were presented in the Figure 1.

Summary of explanatory variables used for CPUE standardization were listed in the Table 3. *Year* is a categorical variable of 9 years (2014—2022). *Month* is a categorical variable including the 10 calendar months from March 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 variable of 44—61 m, which will affect the catchability. The proportion of chub mackerel in the daily catch is the continuous variable (Table 3).

SST and Chla data were derived from the Copernicus Marine Service products (http://marine.copernicus.eu). The spatial-temporal resolution of the SST and Chla data are monthly at 0.25°×0.25° grid. The environmental data was matched with the fishery data for the further analysis. The environmental factors such as SST, Chla have been recognized as important drivers of chub mackerel distribution (Torrejon-Magallanes et al., 2021). SST influences fish physiology, metabolism, production rates, and migration patterns, and Chla reflects primary productivity (Lee et al., 2018; Okunishi et al., 2020). These factors play crucial roles in shaping the distribution and abundance of fishery resources. Therefore, they should be considered in CPUE standardization.

The scatter plots/ box plots of explanatory variables were presented in Figure 2, and the correlation matrix of explanatory variables used in the analysis was shown in Figure 3.

2.2 Full model description and model selection

Both generalized linear model (GLM) and generalized additive model (GAM) were used to estimate standardized CPUE.

The full GLM model was:

log(CPUE+1*) =Year + Month + Longitude + Latitude + Sst + Chla +Vessellength +Proportion+ interaction+ε*

The full GAM model was:

log(CPUE+1*)=Year+ Month+ Longitude+ Latitude + s(Sst) + s(Chla) + s(Vessellength) + s(Proportion) +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 and Latitude.

The optimal model was selected using the Bayesian information criterion (BIC) based on forward selection. Spearman's correlation and mean squared errors (MSE) between the predicted and observed CPUEs were calculated by 5 fold cross-validation with repeated 5 times to select well-performance model between two optimal models. All the model construction and data analysis were used the R(4.0.3) software (packages mgcv and nlme).

2.3 Yearly trend extraction

Time series of standardized CPUE was estimated using the well-performance model. Expanded grid function in R was used to generate a series of spatial homogeneous explanatory variables and the area of each 1°×1° grid cell was considered the same. Then, annual values of ln(*CPUE*) for each area (1°×1°) were predicted. Finally annual standardized CPUE were calculated as the mean of CPUE_y:

$$\overline{CPUE}_y = \frac{1}{n_y} \times \sum_{k=1}^{n_y} CPUE_k^{fitted}$$

where, $\overline{CPUE_y}$ is CPUE indices in *y*th year, n_y is the spatial homogeneous explanatory variables number in *y*th year, $CPUE_k^{fitted}$ is the *k*th fitted CPUE data in *y*th year.

The fitted CPUE and 95% confidence intervals of optimal model were calculated by bootstrap resampled residuals with 1000 replications.

3 Result and Discussion

In this study, we used two models to standardize the CPUEs. The result of the best GLM and GAM model selections were shown in Table 4 and Table 5, respectively. Comparing the results of cross validation tests in GLM and GAM analyses (Table 6), higher Spearman's correlation and lower 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 of Chub mackerel. The summary of fitting a GAM for the optimal model is shown in Table 7. All explanatory variables are highly significant (p<0.01) except for Longitude and Chla. Residuals from the best GAM model showed an approximately normal distribution around 0, which indicated that the model assumptions were satisfied (Figure 4). The estimated relationship between response and explanatory variables were shown in the Figure 5, and the estimated values of main parameters and uncertainty in the parameters were presented in Table 8.

Table 9 and Figure 6 shows the annual changes of nominal CPUE and standardized CPUE by the optimal GAM model. There is similar trend between nominal CPUE and standardized CPUE by GAM. In conclusion, we prefer to choose the best GAM model to estimate standardized CPUE of Chub mackerel fishery.

We standardized CPUE in accordance with the standardization protocol. The checklist is shown in Appendix 1.

References

Lee, D., Son, S., Kim, W., Park, J.M., Joo, H., Lee, S.H., 2018. Spatio-temporal variability of the habitat suitability index for chub mackerel (Scomber japonicus) in the east/Japan sea and the south sea of South Korea. Remote Sens. 10 (6), 938.

Okunishi, T., Yokouchi, K., Hasegawa, D., Tanaka, T., Setou, T., Yukami, R., Takasuka, A., 2020. Relationship between sea temperature variation and fishing ground formations of chub mackerel in the Pacific Ocean off Tohoku. Jpn. Soc. Fish. Oceanogr. 84 (4), 271–284.

Torrejon-Magallanes, J., Angeles-Gonzalez, L.E., Csirke, J., Bouchon, M., Morales- Bojorquez, E., Arreguin-Sanchez, F., 2021. Modeling the Pacific chub mackerel (Scomber japonicus) ecological niche and future scenarios in the northern Peruvian current system. Prog. Oceanogr. 197, 10.

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.
Yatsu, A. Watanabe, T. Ishida, M. Sugisaki, H. Jacobson, L.D., 2005. Environmental effects on recruitment and productivity of Japanese sardine Sardinops melanostictus and Chub mackerel Scomber japonicus with recommendations for management. Fish. Oceanogr. 14, 263–278.
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.

Tables:

Table 1. Catch and effort information by CPUE FLEET

Year	Number of observations	% Coverage of CPUE FLEET(catch)	% Coverage of CPUE FLEET(effort)	Total Catch of CPUE FLEET (MT)	Total Effort for CPUE FLEET and unit	Percentage of overall catch by member (across all fleets/gears)
2014	1477	80%	75%	30030	1477 vessel days	71%
2015	5605	74%	85%	93884	5605 vessel days	67%
2016	6644	82%	89%	98132	6644 vessel days	69%
2017	9578	92%	95%	133632	9578 vessel days	86%
2018	6617	81%	90%	98142	6617 vessel days	75%
2019	2504	81%	90%	43364	2504 vessel days	67%
2020	5158	82%	94%	69543	5158 vessel days	75%
2021	14239	93%	96%	88550	14239 vessel days	82%
2022	13723	70%	90%	75341	13723 vessel days	68%

Table 2. Filter "Rules" used on data for CPUE standardization and the effect on the overall sample size.

Filter Applied	Number of Records Remaining	Number Removed	Number of Records with Chub Mackerel Catch >0
Initial Data set	67518	-	63481
Remove records <2°C & >26°C	65545	1973	63228
Final Data Set	65545	1973	63228

Table 3. Summary of explanatory variables used for GLM and GAM analysis.

Variables		Categorical or continuous	Details	Note
Year	*Year*	9 categories	9 years from 2014 to2022	
Month	*Month*	10 categories	10 months from March to December	
Longitude	*Longitude*	20 categories	145°≤Longitude＜146° ; 146°≤Longitude＜147°; 147°≤Longitude＜148°;…, 164°≤Longitude＜165°	at intervals of 1°
Latitude	*Latitude*	11 categories	35°≤Latitude＜36°; 36°≤Latitude＜37°; …, 45°≤Latitude＜46°	at intervals of 1°
Sea surface temperature	*SST*	spline		
	SST_c	20 categories	3℃≤SST＜4℃;4℃≤SST＜5℃; 5℃≤SST＜6℃; …, 25℃≤SST＜26℃	at intervals of 1℃
Chlorophyll-a concentration	*Chla*	continues		
Vessel length	*Vessellength_c*	10 categories	45m≤Vessellength＜47m; 47m≤Vessellength＜49m …, 61m≤Vessellength＜63m	at intervals of 2m
Proportion of chub mackerel	*Proportion*	continues		

Table 4. Result of GLM model selection

No	GLM model	R^2	BIC	Explained deviance
1	*Ln*(CPUE+1)~*Intercept+Year+Month+Longitude+Latitude+Sst_c +Chla+Vl_c+Proportion*	0.6085	18855.6	60.52%
2	*Ln*(CPUE+1)~*Intercept+Year+Month+Longitude+Latitude+Sst_c +Chla+Vl_c+ Proportion +Year:Month*	**0.6340**	**18797.8**	**62.83%**
3	*Ln*(CPUE+1)~*Intercept+Year+Month+Longitude+Latitude+Sst_c +Chla+Vl_c+ Proportion +Year:Month +Year:Longitude*	0.6413	19485.6	63.20%
4	*Ln*(CPUE+1)~*Intercept+Year+Month+Longitude+Latitude+Sst_c +Chla+Vl_c+ Proportion+ Year:Month +Year:Latitude*	0.6410	19237.31	63.28%
5	*Ln*(CPUE+1)~*Intercept+Year+Month+Longitude +Latitude+Sst_c +Chla+Vl_c+ Proportion +Year:Month+Year: Longitude + Year: Latitude + Month: Longitude + Month: Latitude + Longitude: Latitude*	0.6619	19527.66	64.20%

Table 5. Result of GAM model selection

No	GAM model	R^2	BIC	Explained deviance
1	*Ln*(CPUE+1)~*Intercept+Year+Month+Longitude+Latitude+Sst+Chla+Vl_c+Proportion*	0.6535	17591.2	65.62%
2	*Ln*(CPUE+1)~*Intercept+Year+Month+Longitude+Latitude+Sst+Chla+Vl_c+Proportion +Year:Month*	**0.6712**	**17578.5**	**67.59%**
3	*Ln*(CPUE+1)~*Intercept+Year+Month+Longitude+Latitude+Sst+Chla+Vl_c+Proportion +Year:Month+Year:Longitude*	0.6732	18271.7	68.10%
4	*Ln*(CPUE+1)~*Intercept+Year+Month+Longitude+Latitude+Sst+Chla+Vl_c+Proportion +Year:Month +Year:Latitude*	0.6744	18004.7	68.13%
5	*Ln*(CPUE+1)~*Intercept+Year+Month+Longitude +Latitude+Sst+Chla +Vl_c+Proportion +Year:Month+ Year: Longitude + Year: Latitude + Month: Longitude + Month: Latitude + Longitude: Latitude*	0.6814	184264.5	69.86%

Table 6. The Five-fold cross validation for the best GLM

case	cor_GLM_test	MSE_GLM_test	cor_GAM_test	MSE_GAM_test
1	0.5832	1.1547	0.6125	1.0212
2	0.5889	1.1971	0.6275	1.0347
3	0.5961	1.1048	0.6291	1.0382
4	0.5882	1.1017	0.6159	1.0579
5	0.5957	1.1643	0.6264	1.0561

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

Table 7. Anova test for best GAM model

Parametric Terms:

	df	*F*	P-value	
factor(Year)	8	30.20	< 2.2E-16	***
factor(Month)	9	8.19	3.25E-12	***
factor(Longitude)	19	2.91	0.0432	*
factor(Latitude)	10	4.25	0.00046	***
factor(Vl_c)	9	26.39	< 2.2E-16	***
factor(Year):factor(Month)	69	10.56	< 2.2E-16	***

Approximate significance of smooth terms:

	Edf	Ref.df	*F*	P-value	
s(SST)	2.63	3.43	3.72	0.008	**
s(Chla)	4.82	6.02	1.99	0.04	*
s(Proportion)	8.66	8.97	809.98	< 2.2E-16	***

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

Table 8. The estimated coefficients in the best GAM models for CPUE standardization

Explanatory variable	Coefficient	SE	Explanatory variable	Coefficient	SE
Year2015	-0.225	0.041	Month5	0.655	0.156
Year2016	-0.292	0.037	Month6	0.292	0.149
Year2017	-0.714	0.360	Month7	0.403	0.161
Year2018	-0.204	0.390	Month8	0.277	0.178
Year2019	0.901	0.148	Month9	0.952	0.178
Year2020	-0.180	0.167	Month10	0.782	0.156
Year2021	-0.048	0.210	Month11	0.563	0.170
Year2022	-0.052	0.045	Month12	0.428	0.108
Month4	0.388	0.132			

Table 9. Nominal and standardized CPUEs of CPUE FLEET from 2014 to 2022

Year	Nominal CPUE	Standardized CPUE by GAM	CV (%)	95% CI by GAM
2014	22.33	17.28	1.72	[17.02 17.45]
2015	16.75	12.83	0.56	[12.71 12.94]
2016	14.77	11.44	0.98	[11.30 11.62]
2017	13.92	10.20	1.51	[9.95 10.41]
2018	14.83	12.72	1.24	[12.58 12.89]
2019	17.32	15.19	1.38	[15.01 15.42]
2020	13.48	10.27	0.85	[10.08 10.48]
2021	6.22	4.64	0.94	[4.41 4.87]
2022	5.49	4.10	1.02	[3.88 4.42]

Figures:

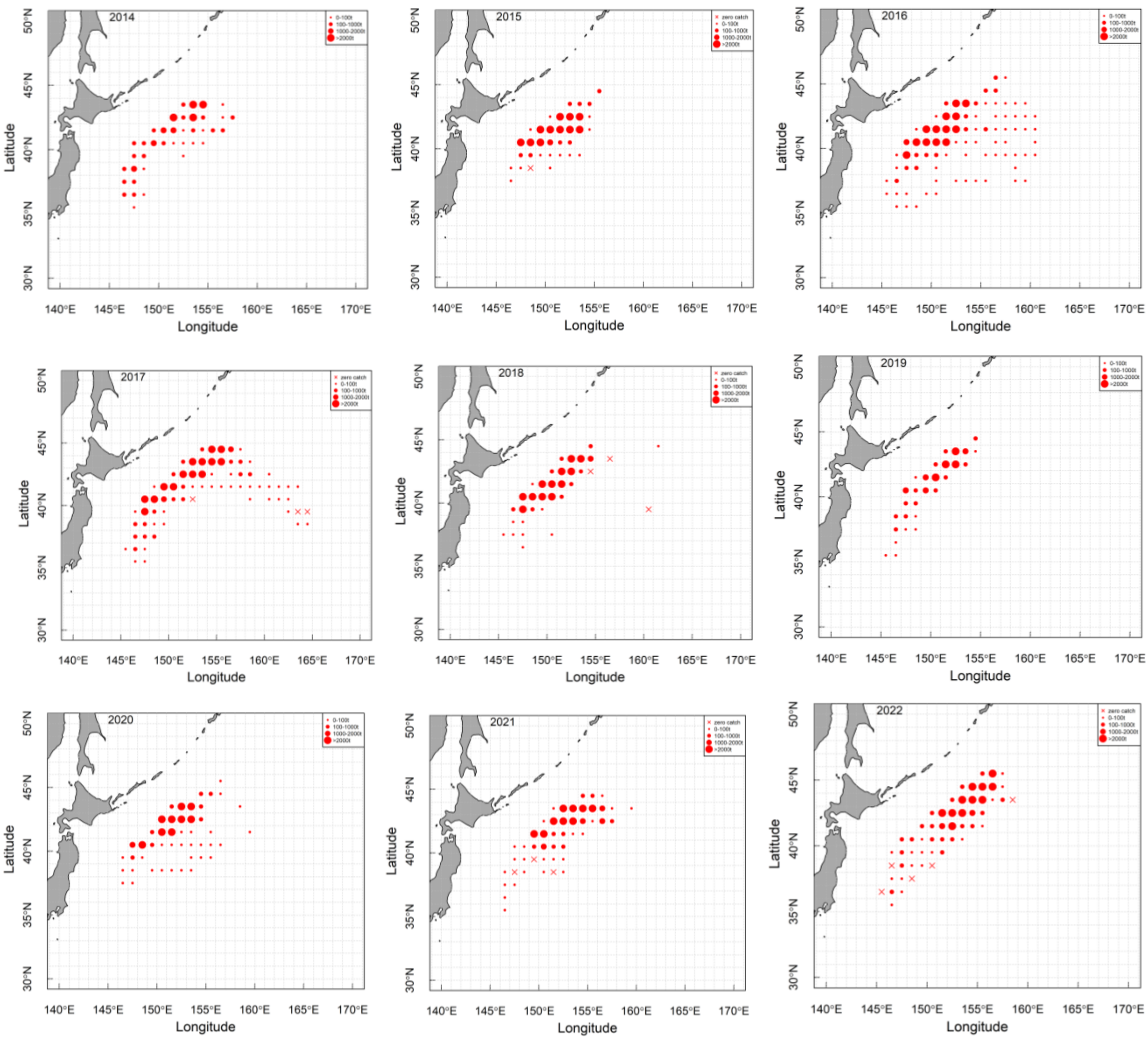


Fig. 1a. Spatio-temporal distribution of the total catch of CPUE fleet (metric tons).

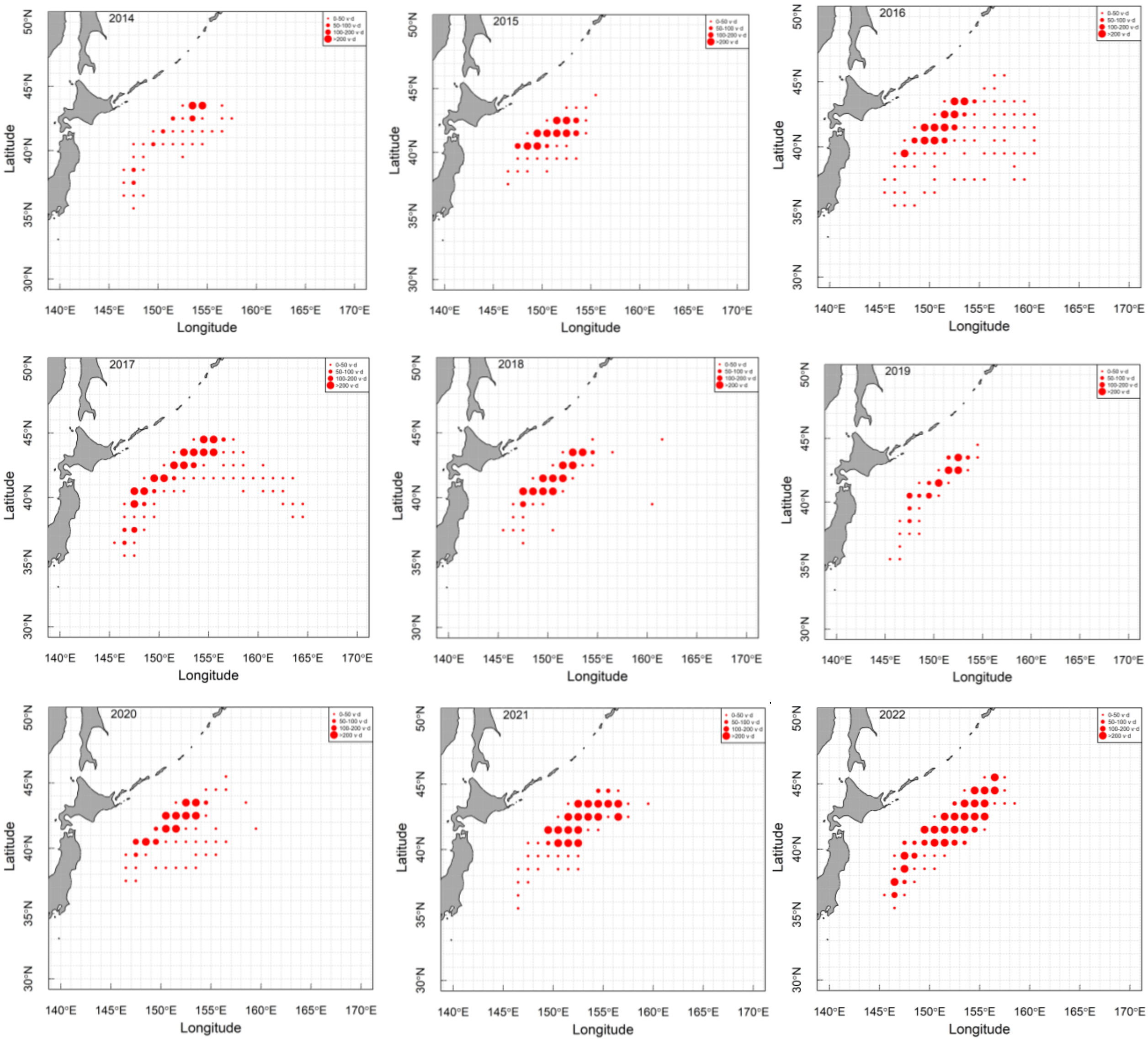


Fig. 1b. Spatio-temporal distribution of efforts by CPUE FLEET (vessel·day).

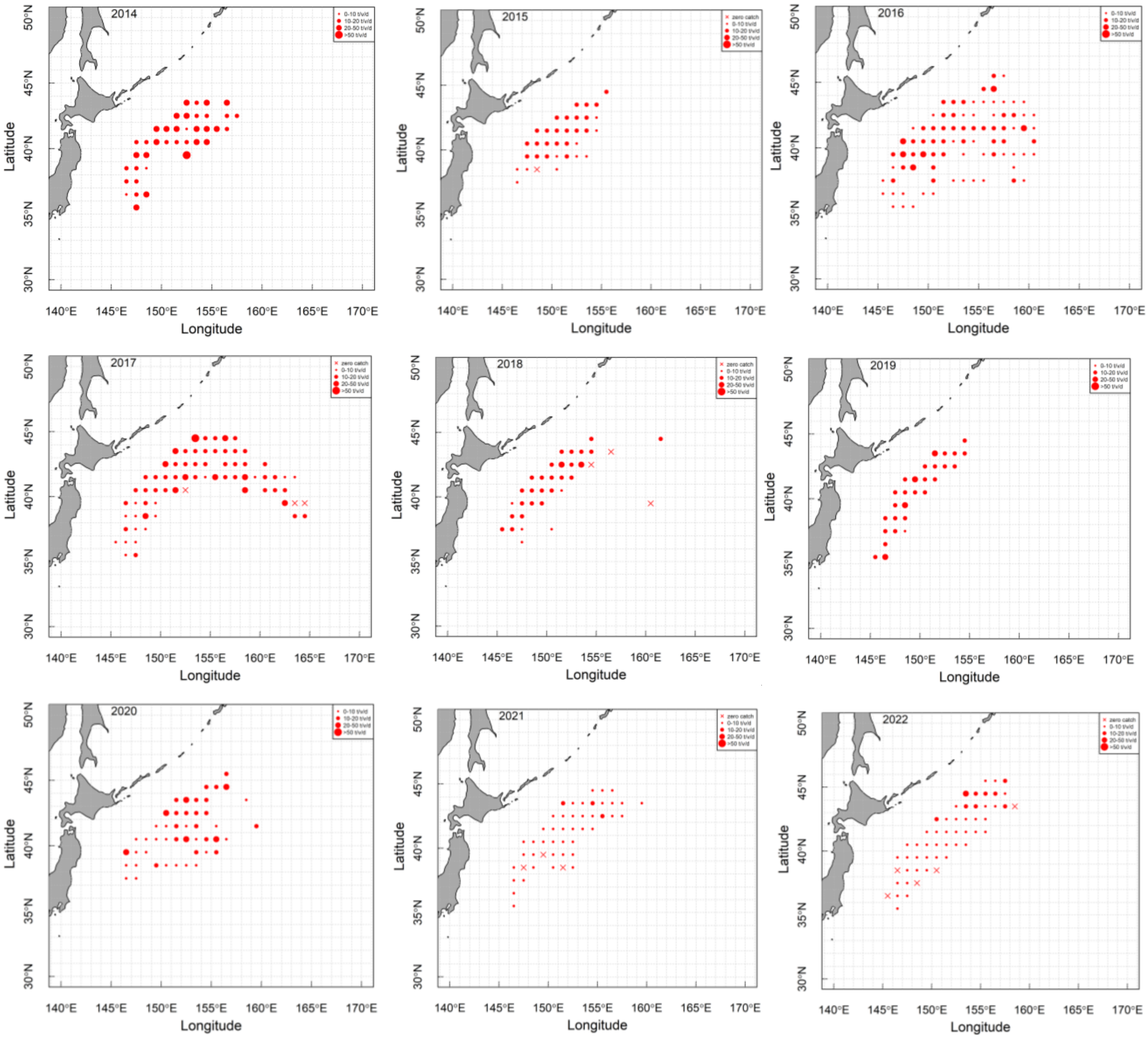
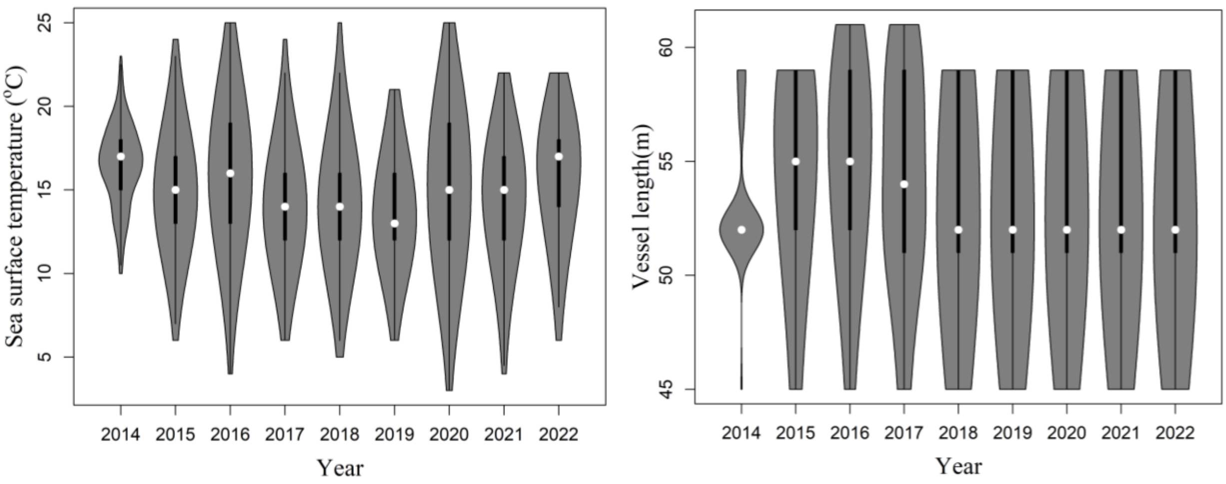


Fig. 1c. Spatio-temporal distribution of nominal CPUE of CPUE Fleet (t/v/d).

(a)



(b)

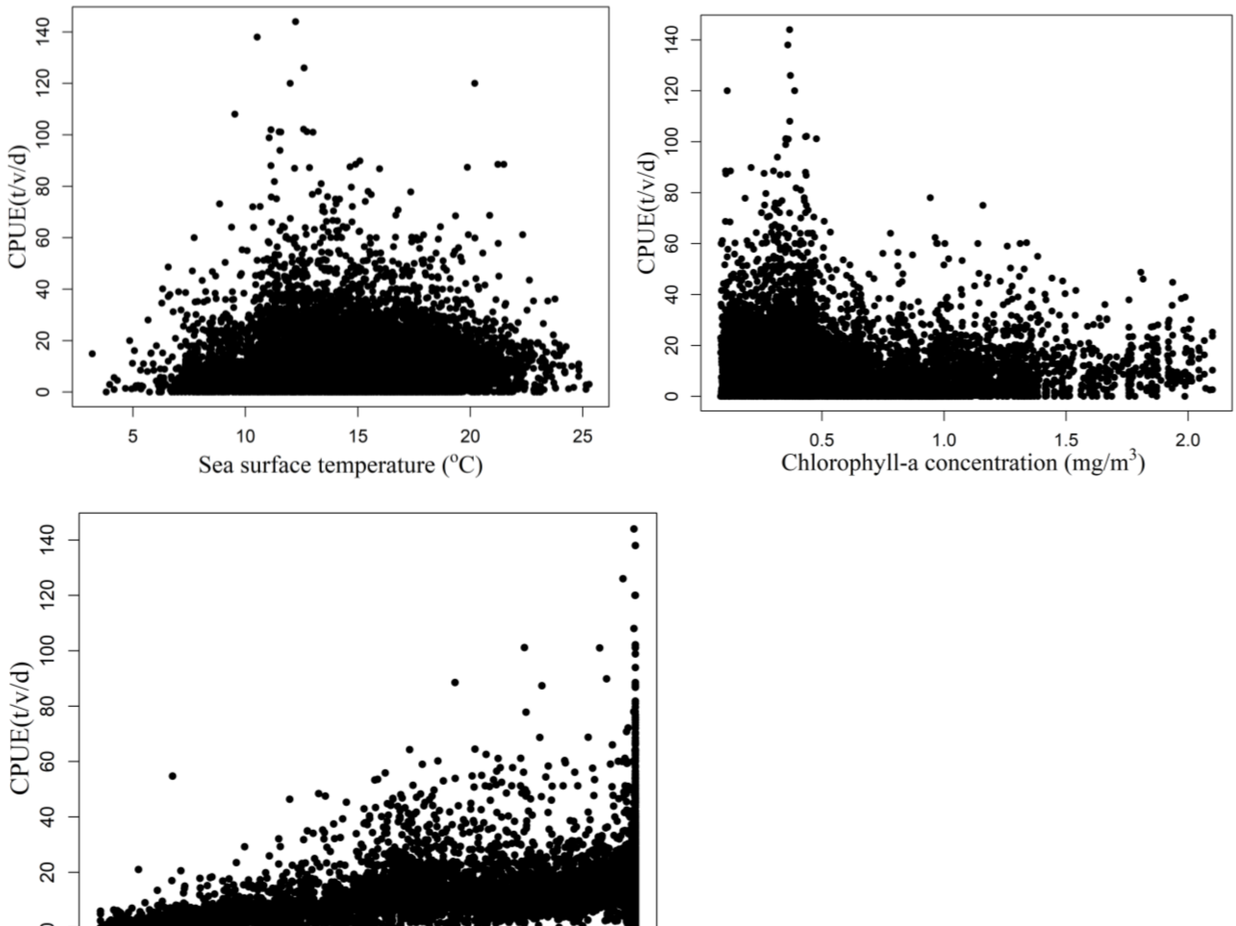


Fig. 2. Plots of explanatory variables of sea surface temperature (SST) and Vessel length by year (a) and scatter plots between CPUE and SST, Chla and proportion of Chub mackerel (b).

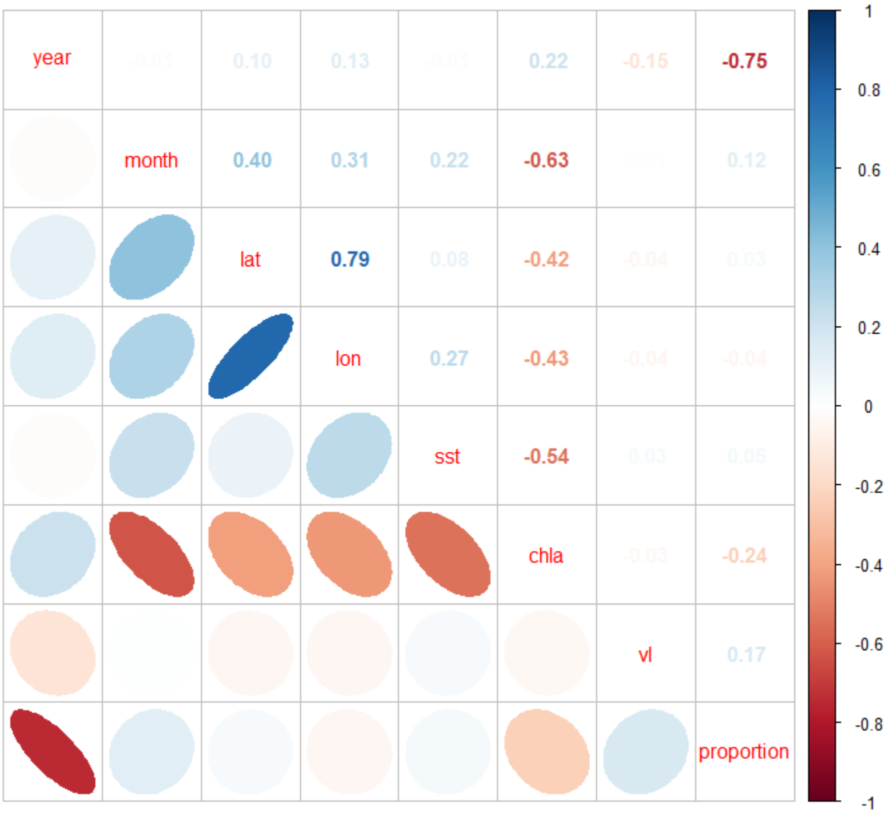


Fig. 3. Correlation matrix of explanatory variables used in the analysis

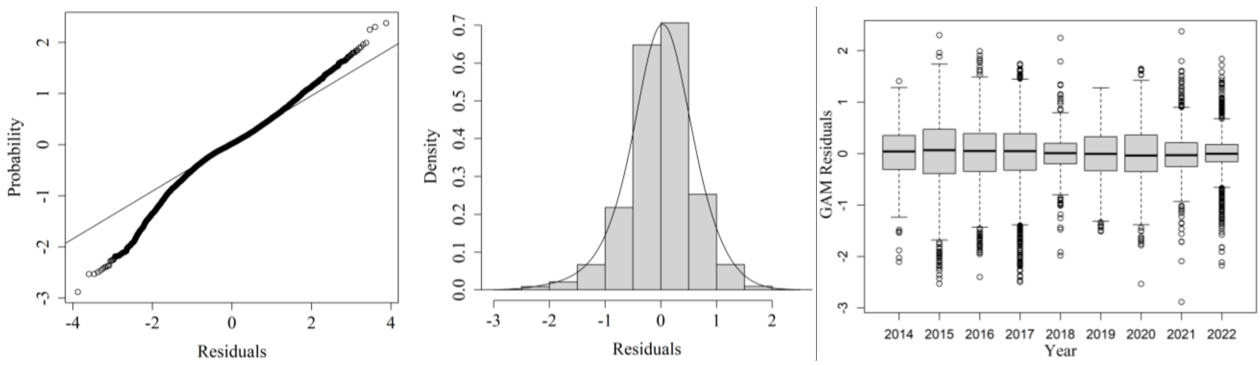


Fig. 4. Q-Q plot, histogram of residuals and residual plots across years for the best GAM.

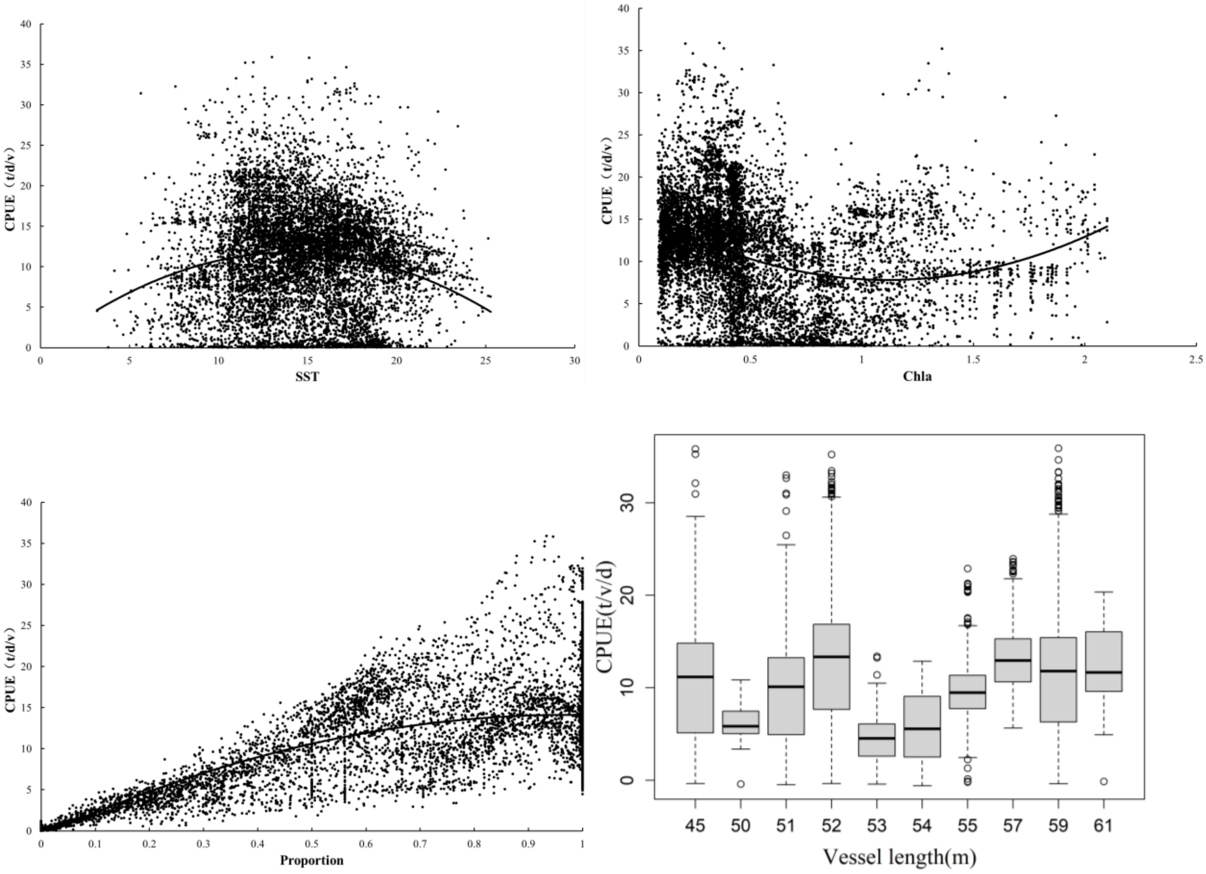


Fig. 5. Estimated relationships between response and explanatory variables.

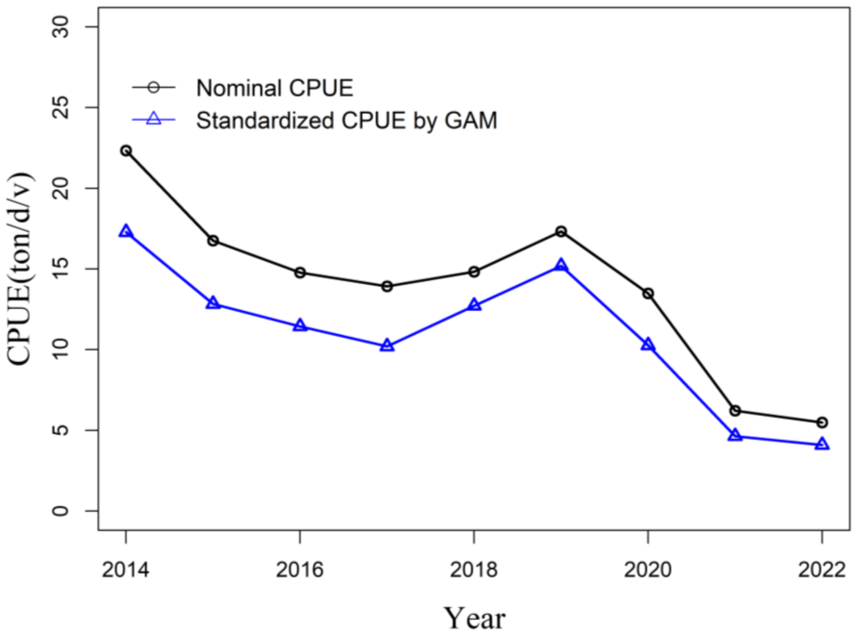


Fig.6. The nominal CPUE and standardized CPUE of Chub mackerel by best GAM up to 2022.

APPENDICES

Appendix1. Checklist for the CPUE standardization protocol

No.	Step-by-step protocols	yes/no	Note
1	Provide a description of the type of data (logbook, observer, survey, etc.), and the "resolution" of the data (aggregated, set-by-set etc..). This description should also include the representativeness of the data in two tables: (1st table) Number of observations, % Coverage of CPUE fleet (catch), % Coverage of CPUE fleet (effort), Total Catch CPUE fleet (mt), Total Effort CPUE fleet, Percentage of overall catch by member (across all fleets/gears); and (2nd table) Number of records remaining, Number removed, Number of records with chub mackerel catch >0;	Yes	See section *2.1* ([page 2-3]) and Tables 1, [page 6] and 2, [page 6]
2	Conduct a thorough literature review to identify potential explanatory variables (i.e., spatial, temporal, environmental, and fisheries variables) that may influence CPUE values;	Yes	See sections 1 and 2.1 ([page 2])
3	Plot annual/monthly spatial catch, effort and nominal CPUE distributions and determine temporal and spatial resolution for CPUE standardization	Yes	See Fig. 1, [page 11-13]
4	Make scatter plots (for continuous variables) and/or box plots (for categorical variables) and present correlation matrix if possible to evaluate correlations between each pair of those variables;	Yes	See Figs 2, [page 14] and 3 [15]
5	Describe selected explanatory variables based on (2)-(4) to develop full model for the CPUE standardization;	Yes	See section *2.2.* ([page 3]) and Table 3, [page 7]
6	Specify model type and software (packages) and fit the data to the assumed statistical models (i.e., GLM, GAM, Delta-lognormal GLM, Neural Networks, Regression Trees, Habitat based models, and Statistical habitat based models);	Yes	See section *2.2.* ([page 3])
7	Evaluate and select the best model(s) using methods such as likelihood ratio test, information criterions, cross validation etc.;	Yes	See Table 4, [page 8] and Table 5, [page 8] and section 3
8	Provide diagnostic plots to support the chosen model is appropriate and assumption are met (QQ plot and residual plots along with predicted values and important explanatory variables, etc.);	Yes	See Table 7, [page 9] and Fig. 4, [page 15]
9	Present estimated values of parameters and	yes	See Table 8, [page 10]

	uncertainty in the parameters in table;		
10	Present the relationship between dependent variable and independent variables. Check whether it is interpretable.	Yes	See Fig. 5, [page 16]
11	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. Provide details on how the CPUE index was extracted.	Yes	See section 2.3. ([page 3])
12	Calculate uncertainty (SD, CV, CI) for standardized CPUE for each year. Provide detailed explanation on how the uncertainty was calculated;	Yes	See section 2.3 (page 4), Table 9, [page 10] and Fig. 6, [page 16]
13	Provide a table and a plot of nominal and standardized CPUEs over time. When the trends between nominal and standardized CPUE are largely different, explain the reasons (e.g. spatial shift of fishing efforts), whenever possible.	Yes	