NPFC-2022-TWG CMSA05-WP03

# Standardized CPUE for Chub mackerel (*Scomber japonicus*) caught by Russian pelagic trawl fishery in 2015-2021

Igor Chernienko, Emilya Chernienko
Pacific branch of VNIRO (TINRO), Russia

**Summary**

Standardization of chub mackerel (*Scomber japonicus*) trawl catches in Russian Federation waters has been conducted based on 2015-2021 fisheries statistics. Production and natural factors were used as predictors. To analyze the influence, generalized additive models (GAM) were used. The choice of the best model was made using the AIC and BIG information criteria. The selected model includes coordinates, day of the year, vessel length, engine power, number of fishing vessels and SST. Interpretation for the influence of considered factors on catch per effort is given.

## 1. BACKGROUND FOR CHUB MACKEREL FISHERY

Chub mackerel is an important fish species for Russian fishery in the Northwest Pacific. Russian fisheries for chub mackerel was resumed in 2015. In 2021, the chub mackerel catch by the Russian fleet amounted to 87,387.988 tons. In this work, we used data from trawl catches of Russian vessels from 2015 to 2021 during autumn, when mackerel feeding migrations occur in Russian national waters (Fig. 1).

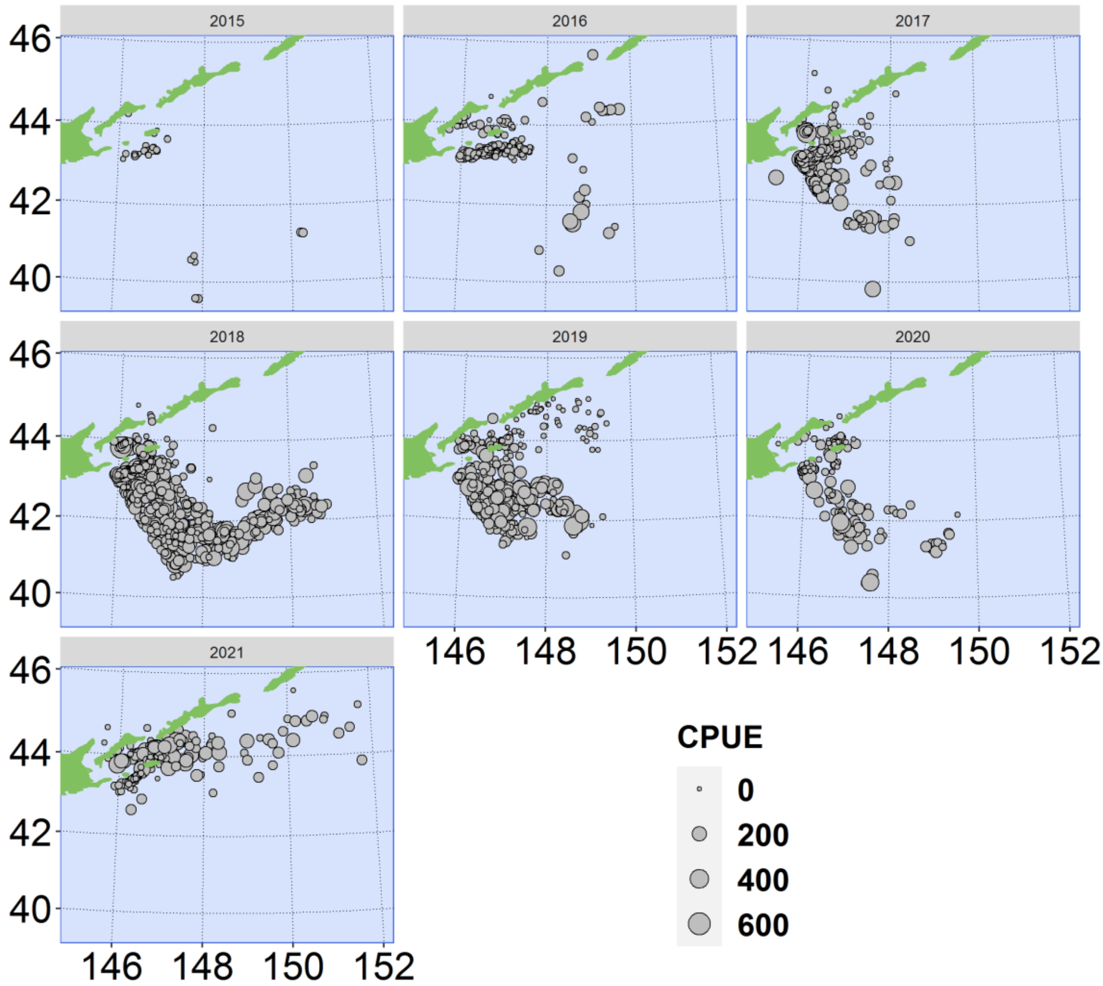


Figure 1. CPUE for Chub Mackerel in Russian national waters in 2015-2021

## 2. METHODS

### 2.1 The data

We used the 2015-2021 fisheries statistics for oceanic area off the South Kuril Islands, based on daily vessel reports and the positions of the vessels of the Industry Monitoring System of the Federal Agency for Fisheries (Pyrkov, 2015). Vessel characteristics were taken from the same source: vessel type, vessel length, engine power. CPUE was a catch per day per vessel; daily effort was also used, which is defined as the number of fishing vessels. Only target fishing operations were chosen (over 50% of mackerel in the catch), and mid-water trawls were taken as most frequently used gear. The fishing period is September-December (Fig. 2). Depth data was obtained from the General Bathymetric Chart of the Oceans GEBCO Web Map Service (WMS) (Becker et al., 2009). SST data were obtained from the GHRSST Multi-Product Ensemble (GMPE) SST (Chin et al., 2017). The spatial-temporal resolution of the SST data is daily at 0.01°×0.01°.

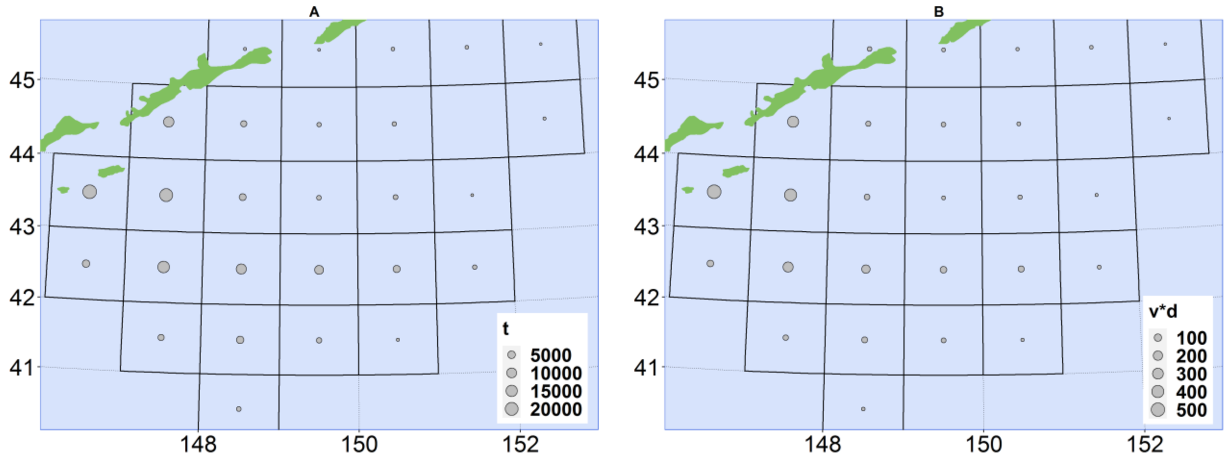


Figure 2. Catch distribution (A) and fishing effort (B) for Chub mackerel fishing fleets in Russian national waters in 2015-2021

## 2.2 Full model description and model selection

Generalized additive models (GAM) (Wood, 2003, 2011) were used to standardize CPUE. We considered the influence of a number of natural and production factors: spatial factors - latitude, longitude, depth, production factors - vessel type as a factor, daily effort (number of vessels in the fishery), vessel length and engine power, natural factor - SST. We used models with different numbers and combinations of factors. The description of independent variables, which were used to standardize CPUE, is given in Table 1.

Table 1 Summary of explanatory variables used for candidate models

| Variable | Notation | Units | Details |
|---|---|---|---|
| Year | $\beta_i^Y$ | categorical | 7 years from 2015 to 2021 |
| Vessel type | $\beta_i^V$ | categorical | 20 types of fishing vessel |
| Longitude | $x$ | decimal degrees | |
| Latitude | $y$ | decimal degrees | |
| Depth | $h$ | meters | |
| Day of year | $d$ | | Serial day of year, ranged from 227 to 366 corresponds to fishing season (in September-December). |
| Daily fishing effort | $E$ | | Number of vessels per day |
| Vessels length | $L_V$ | meters | |
| Engine power | $P_V$ | kWt | Engine power of vessel |
| Sea surface temperature | $SST$ | Celsium degres | Sea surface temperature at vessel position |

Model selection was performed using the AIC as information criterion and the Bayesian information criterion (BIC) criterion (Burnham and Anderson, 2002). All models were tuned in the mgcv package for the R programming language using the maximum likelihood method (Wood, 2017). The common part of the GAMs used can be expressed as follows:

$$g_\mu(\mu_i) = \eta(t_i), \mu_i = E(Y_i), Var(Y_i) = \phi v(\mu), Y_i \mid t_i \sim Tw(y_i), i = 1,...,n,$$

where $g_\mu$ - the link function (natural logarithm) that relates the linear predictor, η, to the mean of the distribution, μ, such that the inverse link function is equal to the mean E of catches Y for a given group of observations (t) by catches (y) in tons per day, distributed according to the dispersion function $v(\mu) = \mu^2$ with the scale parameter $\phi$. The variance function was from the exponential

Twidy (Tw) family (Wood, 2017). The candidate models are listed below:

$$\eta(t_i) = \beta_0 + \beta_i^Y + f_1(x) + f_2(y) + f_3(d) + \beta_{vestype_i}^V \quad (1)$$

$$\eta(t_i) = \beta_0 + \beta_i^Y + f_1(x) + f_2(y) + f_3(d) + f_4(h) + \beta_{vestype_i}^V \quad (2)$$

$$\eta(t_i) = \beta_0 + \beta_i^Y + f_1(x) + f_2(y) + f_3(d) + f_4(h) + f_5(E) + \beta_{vestype_i}^V \quad (3)$$

$$\eta(t_i) = \beta_0 + \beta_i^Y + f_1(x) + f_2(y) + f_3(d) + f_4(h) + f_5(E) + f_6(L_V) + f_7(P_V) \quad (4)$$

$$\eta(t_i) = \beta_0 + \beta_i^Y + f_1(x,y) + f_2(d) + f_3(h) + f_4(E) + f_5(L_V) + f_6(P_V) \quad (5)$$

$$\eta(t_i) = \beta_0 + \beta_i^Y + f_1(x) + f_2(y) + f_3(d) + f_4(h) + f_5(E) + f_6(L_V) + f_7(P_V) + f_8(SST) \quad (6)$$

$$\eta(t_i) = \beta_0 + \beta_i^Y + f_1(x,y) + f_2(d) + f_3(h) + f_4(E) + f_5(L_V) + f_6(P_V) + f_7(SST) \quad (7)$$

$$\eta(t_i) = \beta_0 + \beta_i^Y + f_1(x,y) + f_2(d) + f_3(E) + f_4(L_V) + f_5(P_V) + f_6(SST) \quad (8)$$

where $\beta_0$ – free parameter, $f_1$ – tensor product of coordinates, $f_j$ – thin-walled spline functions (TPRS) estimated using generalized cross-validation (Wood, 2003, 2011) $\beta_{vestype_i}^V$ - coefficient for the ship type factor, $\beta_{year_i}^Y$ – coefficient for the year factor.

Model No. 8 was obtained by excluding depth from the number of predictors, which did not show a significant effect on CPUE. The selected model includes coordinates, day of the year, vessel length, engine power, daily effort (number of vessels in the fishery) and SST. Standardized values were obtained by substituting the median values of the independent variables into the resulting model. For all calculations, the R programming language was used (R Core Team, 2021). Graphs and maps were plotted using the ggplot2 library for the R programming language (Wickham, 2016).

## 3. Results and discussion

Correlation analysis showed that some of the selected factors are quite significantly correlated (Fig. 3). Depth, closely related to coordinates, was excluded from the final model. Based on this, another model No. 8 was built. At the same time, it was decided to keep the engine power correlated with the length of the vessel, since the relationship between these factors is non-linear. In general, the selected variables do not have a high level of dependence.

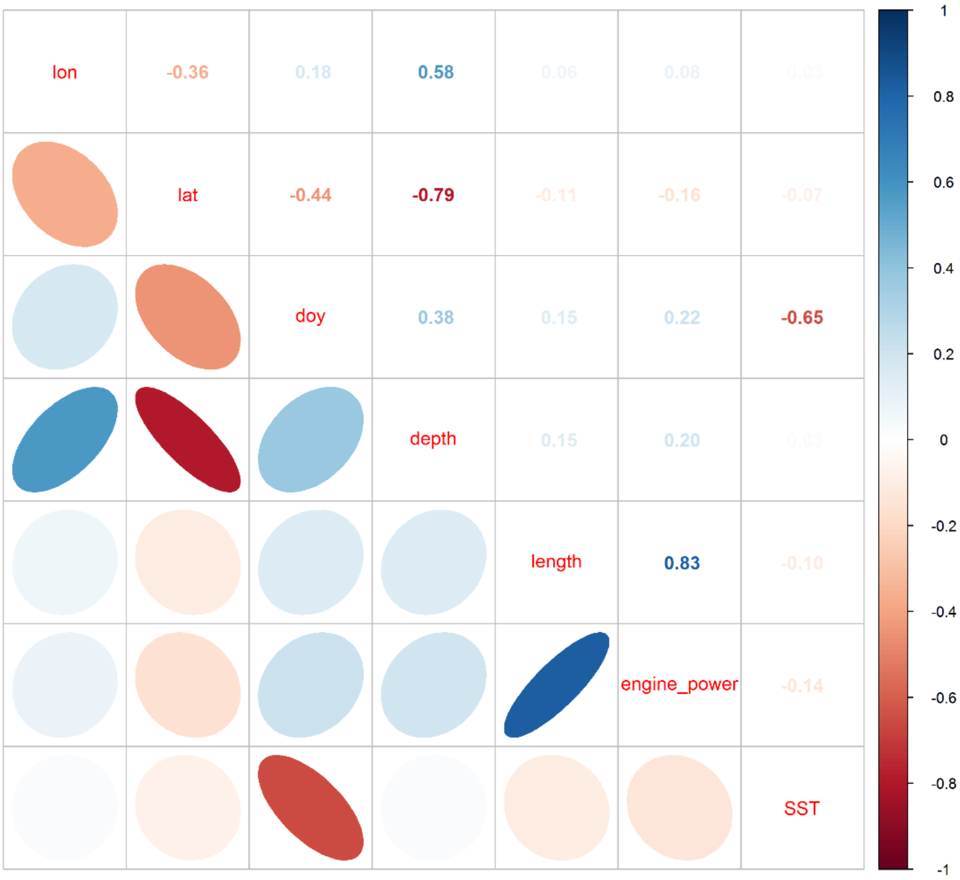


Figure 3. Correlation matrix of explanatory variables used in the analysis.

Table 2 shows values of the AIC and BIC information criteria and explained the variations for the candidate models. The minimum value of AIC (19645) and BIC (19895) and the maximum explained variation (0.63) were marked for model No. 8.

Table 2 – Information criterion values for candidate models

| Model | AIC | BIC | Explained deviance |
|---|---|---|---|
| 1 | 19860 | 20098 | 0,54 |
| 2 | 19860 | 20099 | 0,54 |
| 3 | 19817 | 20091 | 0,55 |
| 4 | 19645 | 19898 | 0,63 |
| 5 | 19665 | 19917 | 0,62 |
| 6 | 19646 | 19899 | 0,63 |
| 7 | 19646 | 19922 | 0,63 |
| **8** | **19645** | **19895** | **0,63** |

For model No. 8, the distribution of residuals corresponds to the predicted (Fig. 4).

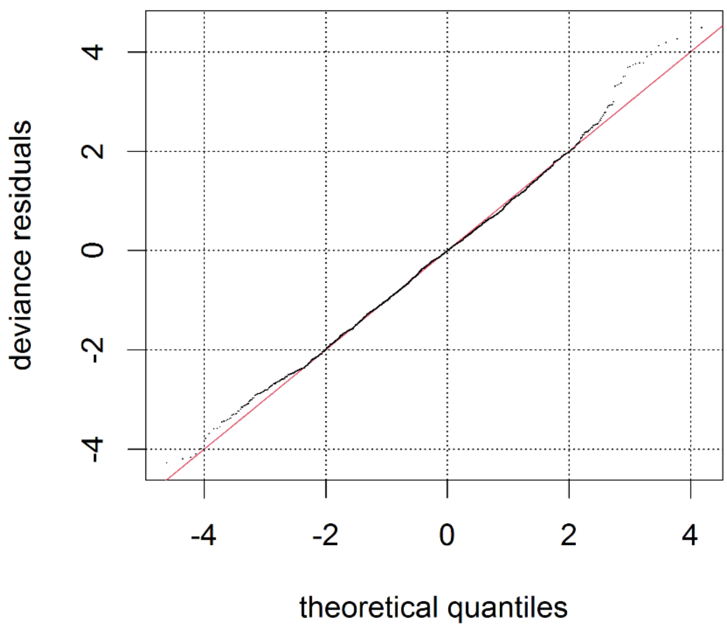


Figure 4. Tweedy distribution check, Q-Q plot for selected model

Fig. 5 shows estimates of the influence of the considered factors on the linear predictor, which can be considered as their contribution to the target variable CPUE. Figure 5A clearly shows an increase in catches with increasing distance from the coast, which is explained by the peculiarities of mackerel migration in this area. Figure 5B is consistent with the dynamics of catch per effort within the framework of the fishing season in the territorial waters of the Russian Federation. Figure 5C shows an increase in catch per effort with the number of vessels in the mackerel fishery. The most interesting effect was noted for the relationship between catch per effort and vessel length (Figure 5D), where, as the vessel length increases beyond 110 m, CPUE decreases sharply. We believe that this is due to the fact that mackerel and sardine concentrations in Russian waters are mostly mixed, catches by large vessels are less selective, and part of the catches of such vessels was not assigned to the target catches (more than 50% of the catch) and was not included in the dataset. The dependence of CPUE on engine power is described by a curve with saturation (Fig. 5E). SST showed a significant effect; however, the nature of the curve is apparently due to the fact that, in Russian national waters, during considered period, temperature values, that could provide negative effect on the concentration of mackerel, were not reached (Fig. 5F).

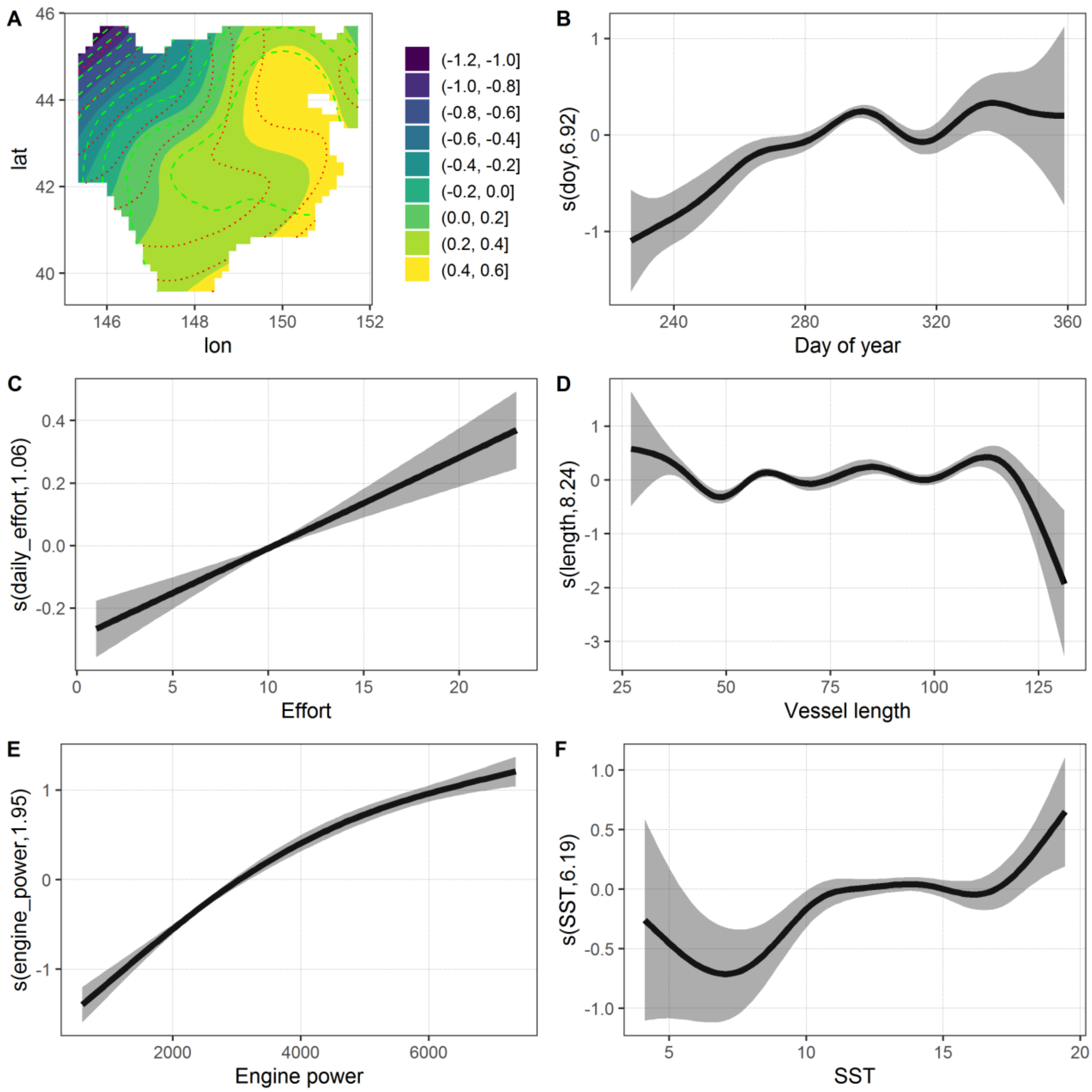


Fig. 5 Component smooth functions for the GAM No 8. A – Tensor product smooths with longitude and latitude; B – Thin plate regression spline with day of year; C – Thin plate regression spline with daily effort; D – Thin plate regression spline with vessel length; E – Thin plate regression spline with engine power; F – Thin plate regression spline with SST.

The results (Fig. 6) demonstrate that, under the influence of production and natural factors, the dynamics of nominal and standardized CPUE differed significantly. During the reviewed period, both the number of fishing vessels and water area covered by the fishery changed (Fig. 1). In addition, the introduction of modern vessels types in chub mackerel fishery and experience of the crews of fishing vessels has a significant impact.

Table 3 shows the nominal CPUE, its standardized values, standard deviation and 95% confidence interval.

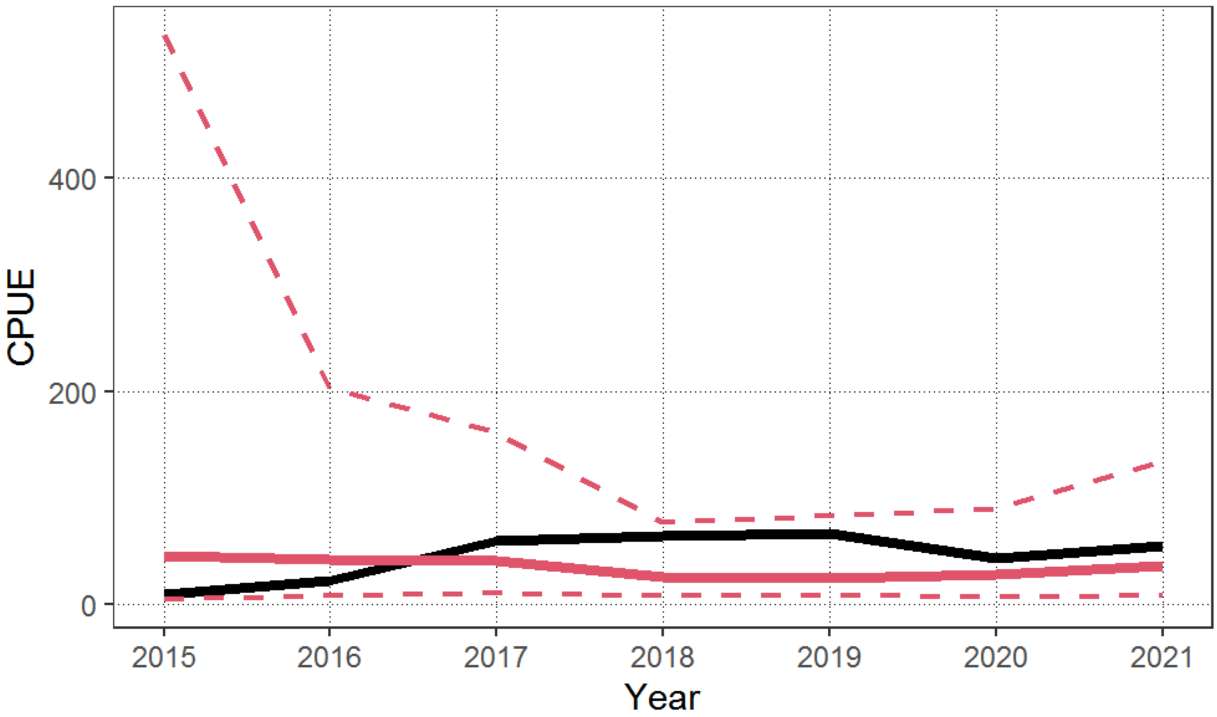


Fig. 6 Annual changes in nominal and standardized CPUEs. Black solid line -- nominal CPUE, red solid line -- sdandardized CPUE, red dashed lines -- 95% confidense interval for standardized CPUE

Table 3 Nominal and standardized CPUE from 2015 to 2021

| Year | Nominal CPUE | Stindirdized CPUE | SD | 95% CI |
|---|---|---|---|---|
| 2015 | 9.41 | 46.08 | 9.25 | [5.08 534.07] |
| 2016 | 22.78 | 42.80 | 3.73 | [9.05 202.52] |
| 2017 | 60.03 | 41.34 | 2.91 | [10.58 161.6] |
| 2018 | 64.91 | 26.10 | 1.95 | [8.84 77.01] |
| 2019 | 67.10 | 25.00 | 2.19 | [8.12 82.65] |
| 2020 | 43.23 | 27.82 | 2.49 | [7.39 90.24] |
| 2021 | 55.57 | 36.03 | 3.07 | [8.12 134.43] |

## 4. REFERENCES

Becker, J.J., Sandwell, D.T., Smith, W.H.F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S.-H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Trimmer, R., Von Rosenberg, J., Wallace, G., and Weatherall, P. 2009. Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS. Marine Geodesy 32:355–371. https://doi.org/10.1080/01490410903297766.

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York, NY., 2. ed., [4. printing}.

Chin, T.M., Vazquez-Cuervo, J., and Armstrong, E.M. 2017. A multi-scale high-resolution analysis of global sea surface temperature. Remote Sensing of Environment 200:154–169. https://doi.org/10.1016/j.rse.2017.07.029.

GEBCO Web Map Service (WMS) GEBCO. https://www.gebco.net/data_and_products/gebco_web_services/web_map_service/

Пырков В.Н., Солодилов А.В., Дегай А.Ю. Создание и внедрение новых спутниковых технологий в системе мониторинга рыболовства // Современные проблемы дистанционного зондирования Земли из космоса. — 2015. — Т. 12, No 5. — С. 251–262.

R Core Team 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer International Publishing : Imprint: Springer, Cham., 2nd ed. 2016. https://doi.org/10.1007/978-3-319-24277-4.

Wood, S.N. 2003. Thin plate regression splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65:95–114. https://doi.org/10.1111/1467-9868.00374.

Wood, S.N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models: Estimation of Semiparametric Generalized Linear Models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73:3–36.

Wood, S.N. 2017. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC., 2nd ed. https://doi.org/10.1201/9781315370279.