NPFC-2017-TWG CMSA01-WP02

# CPUE standardization for the Pacific chub mackerel historical catch in the Northwest Pacific Ocean

Old Russian Vessel Monitoring System (old VMS) recorded changes in position of all operations of Russian vessels. Each vessel reported its catch in metric tons by species for each operation (2-4 per day in average).

## References

Chub mackerel prefers SST at 25°C and is usual up to 27°C.

Unfortunately, we could not find data from *in situ* measurements of SST and other potentially useful hydrographic measurements in Russian VMS records. Therefore, we reconstructed SST values for each day using coordinates of catches from Daily High-Resolution-Blended Analyses for Sea Surface Temperature (Reynolds et al. 2007) converted to degrees Celsius from Kelvin without bias correction.

## Temporal and spatial scales for data grouping for CPUE standardization

The data is divided by such factors as Year and Month on the time scale. We selected years with catches from 1980 to 1989. We didn’t use preselected regions for spatial grouping, because we have spatial information about every single catch. Therefore, we could include coordinates into model directly.

## Statistical assumptions in the full models

We didn’t include 0 catches, because such “catches” are just reports about positions of vessels during searching or other non-fishing operations. So, catches can be positive only and we’ll need a logarithmic link. We checked overdispersion using optimization of power parameter in Tweedie family in mgcv package (Wood 2011) and found out that it is very close to 2 (*p* = 1.98). It means that Gamma family is the best candidate.

## Select and evaluate the models

Model selection was made using An Information Criterion – AIC (Akaike 1974) and Schwarz's Bayesian criterion – BIC (Sakamoto et al. 1986). All models were tuned in mgcv package using maximum likelihood. We used Generalized linear models, or GLMs. Common part of GLMs, which were used, can be expressed as follows: 

where — is the link function (natural logarithm), which establishes the connection between the linear predictor, *η*, and the mean of the distribution, *µ*, in such way, that the inverse of link function equals to the expectation *E* of catches *Y* given the group of observations (*t*) from catches (*y*) in tons per day distributed according to the variance function with scale parameter . The variance function was from Gamma (Г) exponential family. Therefore, GLMs distinguished only by linear predictor and scale (Table 1). We couldn’t find daily SST in 1980, thus we had to repeat model optimization for a subset with SST values (Table 2).

Table 1 – Models and their statistics for a full set of data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| №  |  |  | Adj. R2 | Dev. expl.% | AIC | BIC | df |
| 1 |  | 0.58 | 0.19 | 21.0 | 124643 | 125943 | 166.1 |
| 2 |  | 0.58 | 0.20 | 21.2 | 124592 | 125916 | 169.3 |
| 3 |  | 0.65 | 0.12 | 12.8 | 126487 | 127158 | 85.8 |
| 4 |  | 0.59 | 0.18 | 19.7 | 124929 | 126061 | 144.6 |
| 5 |  | 0.58 | 0.20 | 21.2 | 124592 | **125916** | 169.3 |
| 6 |  | 0.58 | 0.20 | 21.4 | **124570** | 125978 | 179.9 |

*β*0 – intercept, *β*1 – coefficient of the model’s offset (swept length (*L*) in km), *f1…3* – thin plate regression splines (TPRS) estimated by using the Generalized Cross Validation (Wood 2003) for years, months and coordinates in UTM55 kilometer grid,  coefficient of *i*-th type of vessel (*vestypei*),  coefficient of *i*-th type of gear (*geari*), – coefficient of *i*-th unique name of a vessel (*vesi*),  – coefficient of *i*-th year (*yeari*), – coefficient of *i*-th month (*monthi*), – TPRS for months independent for each year.

Finally, the best by BIC is GLM No 5 (see Table 1). Final estimates of the coefficients for GLM No 5 were found using restricted maximum likelihood as it is recommended (Wood 2011).

The dataset with SSTs contains 1.8 times less observations (10375 vs 18519). Nevertheless it was sufficient to find out that the best by BIC and AIC is GLM No 16 (see Table 2).

Table 2 – Models and their statistics for a set of data with SSTs

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| №  |  |  | Adj. R2 | Dev. expl.% | AIC | BIC | df |
| 7 |  | 0.53 | 0.24 | 22.9 | 70024 | 70897 | 120.5 |
| 8 |  | 0.53 | 0.24 | 23.5 | 69940 | 70868 | 128.0 |
| 9 |  | 0.61 | 0.14 | 14.1 | 71138 | 71658 | 71.8 |
| 10 |  | 0.55 | 0.20 | 21.4 | 70221 | 71025 | 111.0 |
| 11 |  | 0.53 | 0.24 | 23.5 | 69940 | 70868 | 128.0 |
| 12 |  | 0.52 | 0.25 | 24.5 | 69809 | 70787 | 134.9 |
| 13 |  | 0.53 | 0.24 | 22.9 | 70018 | 70909 | 123.0 |
| 14 |  | 0.55 | 0.23 | 21.3 | 70202 | 70912 | 98.0 |
| 15 |  | 0.53 | 0.24 | 23.6 | 69937 | 70875 | 129.5 |
| 16 |  | 0.52 | 0.25 | 24.8 | **69779** | **70779** | 138.0 |

 – TPRS for SST at each place of operations.

GLM No 16 predicts higher catches when SST is closer to the optimum of 25°C as we expected, though we didn’t find temperatures higher then 23°C during the main months of catch October to February (Fig. 1).



Figure 1 – Thin plate regression spline with SST estimated with REML in GLM No 16

## Evaluate if distributional assumptions are satisfied

Gamma distribution suited very well to capture over dispersion, so we don’t see patterns in residuals of both final models (Fig. 2 and 3).



Figure 2 – Diagnostic of GLM No 5



Figure 3 – Diagnostic of GLM No 16

## Nominal and standardized CPUE over time

Year’s coefficients in GLM No 5 plus intercept on a link scale after inverse operation to the original scale can be considered as indices of abundance, because they are not included as interaction term anywhere in the model No 5. The result of inverse operation is below (Fig. 4).



Figure 4 – CPUEi from GLM No 5, where red line connects inversed estimates of Year’s coefficients, green line connects means of raw catch values (tons per operation) used for standardization, while blue line connects published estimates of Russian catch per vessel per year divided by 50 for comparative purpose

Interaction term of Month given factor Year complicates the use of Year’s coefficients itself as indices of abundance in GLM No 16. To overcome this difficulty, we expanded a grid which included all used levels of months with the most numerous quantities of catches (October to February), all years (1981-1989), all vessels (68 levels), all gears (22 levels) and median values for SST and coordinates for each given month. Then we predicted catches using GLM No 8 and summarized them (Fig. 5).



Figure 5 – Violin and box plots for catches per operation (CPUEi) predicted from GLM No 16, where black line connects medians of predictions and red line connects means of predictions, green line connects means of raw catch values (tons per operation) used for standardization, while blue line connects published estimates of Russian catch per vessel per year divided by 50 for comparative purpose

Summary statistics of predicted values is given below (Table 2) as well as statistics of raw data used for standardization (Table 3) in GLM No 16 as well as in GLM No 5 (Table 4).

Table 2 – Summary statistics of predicted 7480 CPUEi by GLM No 16 for each year

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | Mean | SD | SE | Median | Trimmedmean | MAD | Min | Max | skew | kurtosis |
| 1981 | 20.27 | 28.20 | 0.33 | 11.57 | 14.63 | 11.93 | 0.01 | 446.07 | 4.5 | 33.4 |
| 1982 | 20.39 | 23.39 | 0.27 | 13.35 | 16.00 | 10.69 | 0.03 | 341.84 | 4.0 | 25.9 |
| 1983 | 25.93 | 29.90 | 0.35 | 16.90 | 20.32 | 13.39 | 0.04 | 460.99 | 4.0 | 27.5 |
| 1984 | 30.57 | 37.89 | 0.44 | 19.31 | 23.29 | 16.82 | 0.03 | 609.42 | 4.4 | 32.3 |
| 1985 | 27.23 | 33.33 | 0.39 | 17.16 | 20.88 | 14.30 | 0.04 | 568.18 | 4.5 | 35.7 |
| 1986 | 19.51 | 24.09 | 0.28 | 12.31 | 14.91 | 10.43 | 0.03 | 405.20 | 4.5 | 35.2 |
| 1987 | 15.14 | 18.08 | 0.21 | 9.72 | 11.72 | 7.91 | 0.02 | 299.12 | 4.3 | 32.6 |
| 1988 | 12.70 | 15.57 | 0.18 | 8.05 | 9.74 | 6.62 | 0.02 | 269.20 | 4.6 | 36.9 |
| 1989 | 19.44 | 22.31 | 0.26 | 12.76 | 15.25 | 10.37 | 0.03 | 302.43 | 3.9 | 24.9 |

Table 3 – Summary statistics of raw catch values used for CPUE standardization by GLM No 16

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | n | Mean | SD | SE | Median | Trimmedmean | MAD | Min | Max | skew | kurtosis |
| 1981 | 1368 | 14.35 | 10.99 | 0.30 | 10.05 | 12.65 | 7.49 | 0.30 | 92.00 | 2.16 | 7.24 |
| 1982 | 193 | 10.73 | 7.20 | 0.52 | 10.00 | 9.79 | 7.41 | 1.00 | 45.00 | 1.41 | 2.71 |
| 1983 | 803 | 12.84 | 10.41 | 0.37 | 10.00 | 11.25 | 7.41 | 0.30 | 115.00 | 2.93 | 16.15 |
| 1984 | 669 | 17.89 | 15.22 | 0.59 | 13.20 | 15.42 | 10.08 | 0.60 | 103.20 | 1.75 | 3.78 |
| 1985 | 2143 | 14.89 | 11.38 | 0.25 | 10.20 | 13.20 | 7.71 | 0.10 | 100.00 | 1.91 | 5.51 |
| 1986 | 1553 | 12.12 | 10.28 | 0.26 | 10.00 | 10.54 | 7.41 | 0.10 | 72.45 | 1.86 | 4.80 |
| 1987 | 2704 | 10.66 | 9.29 | 0.18 | 9.00 | 9.08 | 5.93 | 0.20 | 99.00 | 2.45 | 9.53 |
| 1988 | 819 | 12.52 | 12.32 | 0.43 | 10.00 | 10.37 | 9.64 | 0.30 | 82.00 | 1.92 | 4.58 |
| 1989 | 123 | 11.97 | 10.65 | 0.96 | 10.00 | 10.26 | 7.41 | 0.10 | 60.20 | 1.62 | 3.03 |

Table 4 – Summary statistics of raw catch values used for CPUE standardization by GLM No 5

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | n | Mean | SD | SE | Median | Trimmedmean | MAD | Min | Max | skew | kurtosis |
| 1980 | 7654 | 12.24 | 10.96 | 0.13 | 10.00 | 10.38 | 7.41 | 0.10 | 90.00 | 2.14 | 6.46 |
| 1981 | 1830 | 13.57 | 11.10 | 0.26 | 10.00 | 11.86 | 7.41 | 0.10 | 92.00 | 2.07 | 6.64 |
| 1982 | 193 | 10.73 | 7.20 | 0.52 | 10.00 | 9.79 | 7.41 | 1.00 | 45.00 | 1.41 | 2.71 |
| 1983 | 816 | 12.76 | 10.36 | 0.36 | 10.00 | 11.19 | 7.41 | 0.30 | 115.00 | 2.94 | 16.34 |
| 1984 | 669 | 17.89 | 15.22 | 0.59 | 13.20 | 15.42 | 10.08 | 0.60 | 103.20 | 1.75 | 3.78 |
| 1985 | 2143 | 14.89 | 11.38 | 0.25 | 10.20 | 13.20 | 7.71 | 0.10 | 100.00 | 1.91 | 5.51 |
| 1986 | 1553 | 12.12 | 10.28 | 0.26 | 10.00 | 10.54 | 7.41 | 0.10 | 72.45 | 1.86 | 4.80 |
| 1987 | 2719 | 10.63 | 9.28 | 0.18 | 8.64 | 9.05 | 5.40 | 0.20 | 99.00 | 2.45 | 9.55 |
| 1988 | 819 | 12.52 | 12.32 | 0.43 | 10.00 | 10.37 | 9.64 | 0.30 | 82.00 | 1.92 | 4.58 |
| 1989 | 123 | 11.97 | 10.65 | 0.96 | 10.00 | 10.26 | 7.41 | 0.10 | 60.20 | 1.62 | 3.03 |

Finally, GLM No5 and GLM No16 were highly correlated with biomass estimates made by Japanese and Russian scientists (Fig. 6 and Fig. 7)



Figure 6 – Pearson’s correlations with GLM No5 (GLM5), where “\*\*\*” indicates *p* < 0.001, “\*\*” – *p* < 0.01, “\*” – *p* < 0.05 and “∙” – *p* < 0.1 for the period from 1980 to 1989 years



Figure 7 – Pearson’s correlations with GLM No 16 (GLM16), where “\*\*\*” indicates *p* < 0.001, “\*\*” – *p* < 0.01, “\*” – *p* < 0.05 and “∙” – *p* < 0.1 for the period from 1981 to 1989 years

REFERENCES

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19(6): 716–723. doi:10.1109/TAC.1974.1100705.

Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., and Schlax, M.G. 2007. Daily High-Resolution-Blended Analyses for Sea Surface Temperature. J. Clim. 20(22): 5473–5496. doi:10.1175/2007JCLI1824.1.

Wood, S.N. 2003. Thin plate regression splines. J. R. Stat. Soc. Ser. B 65(1): 95–114. Available from http://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00374/full [accessed 28 April 2013].

Wood, S.N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. Ser. B (Statistical Methodol. 73(1): 3–36. doi:10.1111/j.1467-9868.2010.00749.x.

Sakamoto, Y., Ishiguro, M., and Kitagawa G. 1986. Akaike Information Criterion Statistics. D. Reidel Publishing Company.