Joint CPUE standardization of the Pacific saury in the Northwest Pacific Ocean during 2001 - 2020 using the Vector-Autoregressive Spatiotemporal Model

Jhen Hsu and Yi-Jay Chang

Institute of Oceanography, National Taiwan University

Abstract

Reliable indices of population abundance are an important type of data for stock assessment. We applied a Vector-Autoregressive Spatio-Temporal Model (VAST) to conduct index standardization update by using the joint CPUE (catch-per-unit-effort) data of the Pacific saury in the Northwest Pacific Ocean during 2001 and 2020. The results indicated that the annual standardized CPUE trend had a fluctuated pattern over studied periods, and the annual standardized CPUE value was at the lowest level below average (2001 - 2020) in 2020.

1. Introduction

Standardization of commercial catch and effort data is important in fisheries where standardized abundance indices based on the fishery-dependent data are a fundamental input to stock assessments. The nominal CPUE (catch-per-unit-effort) index, derived from yearly means of the raw CPUE data, can be severely biased due to the fishing fleets in specific locales using gear that increases catchability, low fishing effort in areas which give inaccurate average CPUE, oceanography conditions that increase catchability by, for instance, making fish more vulnerable to fishing gear, or simply chance. The most commonly used standardization procedures entail the application of Generalized Linear Models (GLMs) or Generalized Additive Models (GAMs), which aim to isolate temporal abundance trends from the total variation in the CPUE data by adjusting for confounding effects on the estimated abundance trends (Guisan et al., 2002; Maunder and Punt, 2004). It should be noted that observations occurred closer in space are more likely to be similar (spatial autocorrelation), which makes it harder to distinguish the real signal of a spatial effect by an explanatory variable. Recent years

have seen the emergence of spatiotemporal modelling methods for standardizing CPUE data (e.g., Walter et al., 2014; Thorson et al., 2015; Kai et al., 2017; Grüss et al., 2019), because they allow the spatial autocorrelation to be removed, which may yield more precise, biologically reasonable, and interpretable estimates of abundance than common methods such as GLM (Shelton et al., 2014; Thorson et al., 2015).

Pacific saury (*Cololabis saira*), a migratory small pelagic fish, is widely distributed and migrate over extensive areas of the Northwestern Pacific Ocean. (Fukushima, 1979). This species is commercially important in the Northwestern Pacific Ocean, targeted by stick-held dip net fisheries from several members of the North Pacific Fisheries Commission (NPFC) that the offshore fishing vessels by Japan and Russia operate mainly within the exclusive economic zones while the distant-water vessels of China, Korea, and Chinese Taipei operate mainly east of Hokkaido and the Kuril Islands in the Northwestern Pacific Ocean. In view of the fact that there is a conflict among the standardized CPUE indices derived by members, the 3rd Technical Working Group on the Pacific Saury Stock Assessment (TWG PSSA) aim to develop a single joint CPUE index for the Pacific saury from the catch and effort data by all members (i.e., joint CPUE data).

In this study, we apply a Vector-Autoregressive Spatio-Temporal Model (i.e., VAST; Thorson, 2019) to conduct an index standardization update by using the joint CPUE data of the Pacific saury in the Northwest Pacific Ocean during 2001 and 2020. The objective is to develop a single joint index for the use in the Pacific saury stock assessment. Progress in joint standardized CPUE should result in better assessment and management of the stock.

2. Methods

2.1 Joint CPUE dataset

The joint CPUE data of stick-held dip net fisheries was collected from each member including Japan, Chinese Taipei, China, Korea, Russia and Vanuatu in the North

Pacific Fisheries Commission (NPFC) during 1994 - 2020. The joint CPUE dataset also serves as a repository for the fisheries summary plots (including catch, operating day, and nominal CPUE; see **Appendix figures 1 - 6**). The original dataset was aggregated by year and month with a spatial resolution of 1° × 1° and covered the northwestern Pacific Ocean between 32 - 50 °N and 140 - 176 °E from 1994 to 2020. Data grooming was applied prior to the standardization to remove the monthly observations before 2001 and that less than 10 operation days. CPUE was defined as a catch of Pacific saury in metric ton per operating day fished. **Figure 1** illustrates the distribution of fishing effort (i.e., 1 by 1 grid) by different members during 1994 - 2020. The spatial and temporal pattern of the nominal joint CPUE data during 2001 and 2020 was shown in **Figures 2 - 4.**

2.2 Geostatistical CPUE standardization

The approach we used here is adapted from the R package VAST (https://github.com/James-Thorson-NOAA/VAST) developed by Thorson et al. (2015). VAST uses the Gaussian random fields to model the spatial autocorrelation with anisotropy (which means the relationship of spatial autocorrelation does not have to change at the same rate in all directions), and an interactive relationship between space and time (i.e., spatio-temporal autocorrelation). These Gaussian random fields are defined with a Matérn covariance function (see Thorson, 2019). VAST requires the previous definition of knots s which are points where the correlation of spatial and spatio-temporal effects are estimated. Each observation in the dataset then gets assigned to the knot which is the closest to them using the k-means. In this study, we specify 100 spatial knots (see **Figure 5** for the configuration) to approximate the spatial and spatio-temporal autocorrelated variations. We confirmed that our results are qualitatively similar when using various numbers of spatial knots (100, 150, and 200 knots) in the exploratory runs.

We give a brief description of how the VAST is applied to the Pacific saury joint CPUE dataset below and refer the readers to the original reference for more technical details (see also Thorson, 2019). The logarithm prediction of Pacific saury density, $p(s,t)$, in knot s and year-month t is described below:

$$p(s,t) = \beta(t) + \omega(s) + \varepsilon(s,t) + \sum_{j=1}^{n_j} \gamma_j x_j(s,t) + \sum_{k=1}^{n_k} Q(k) \tag{1}$$

where $\beta(t)$ is the intercept for each year-month t as a fixed effect, $\omega(s)$ is a time-invariant spatial autocorrelated variation for knot s (100 knots), and $\varepsilon(s,t)$ is a time-varying spatio-temporal autocorrelated variation for knot s and in year-month t (i.e., the interaction of spatial variation and time). γ_j represents the impact of covariate j (i.e., the linear impact of SST, $n_j = 1$) with value $x_j(s,t)$ on density for knot s and year-month t. $Q(k)$ are the fixed effects for catchability (e.g., fleet, $n_k = 1$). The detail information of explanatory variables used in VAST was shown in **Table 1**. The correlation matrix for these explanatory variables of VAST is shown in **Figure 6**.

2.4 Model selection and diagnostics

We used the Akaike Information Criterion (AIC; Akaike, 1973) to identify the best model which has the greatest support given available data within the VAST. Histograms of the residuals were used to assess normality by fleets for the best model, in addition, the quantile-quantile normal probability plots (Normal Q-Q plot). For a better understanding of CPUE standardization of Pacific saury, the "step plots" (Bishop et al., 2008) were conducted to understand the effects of removing individual factors from the VAST with respect to the estimated CPUE indices. We also calculated the influence index to quantify how much a variable can contribute to differences in CPUE patterns of standardized and unstandardized CPUE values (Bentley et al., 2012).

2.5 Standardized CPUE trend

Predictions of standardized Pacific saury density for observation i then excludes the value for the covariates linked to catchability, here is the fleet but otherwise retains the other predictors of density in space and time. Estimated values of fixed and random effects are used to predict the relative density $p(s,t)$ except the catchability variables (Thorson, 2019). Year-month density, $B(t)$, is calculated as the sum of the density of each station ($p(s,t)$):

$$p(s,t) = \beta(t) + \omega(s) + \varepsilon(s,t) + \sum_{j=1}^{n_j} \gamma_j x_j(s,t)$$

(2)

$$B(t) = \sum \exp\left(p(s,t)\right) \times a(s)$$

(3)

where $B(t)$ is the area re-weighted biomass density in year-month t throughout the population domain, $a(s)$ is the area of knot s. Furthermore, year-month density was bias-corrected by using the "epsilon bias-correction estimator" (Thorson and Kristensen, 2016) to correct for retransformation bias. Annual standardized CPUE is calculated as the averaged density across the month.

3. Results and discussion

3.1 Model selection and diagnostic

The convergence in optimization was confirmed for each model if the Hessian matrix was positive and the maximum gradient of each component was smaller than 0.0001. According to the AIC value, we used the most parameterized model (V-5) of VAST to predict the year-month changes in CPUEs of Pacific saury (**Table 2**). VAST model was shown to robustly fit the CPUE data (deviance explained = 67%). The spatial residuals, aggregated across months, at the level of the knot, exhibited a normal distribution around zero, indicating an overall reasonable fit to the data (**Fig. 7**). However, a notable spatial pattern in residuals, such that the area of 145 °E - 155°E and 40 °N - 45°N exhibited residuals apparently below zero during 2004 - 2006. Moreover, the peripheral areas generally exhibited spatial residual below zero, this pattern suggested that the density estimates may be slightly overestimated in peripheral regions. The histogram and Q-Q plots of model based on the lognormal distributions appear normal in VAST for all fleets (**Fig. 8**), which confirms the assumption of the error distribution is generally appropriate for the CPUE standardization.

3.2 Exploration of the standardization effects in variables influencing CPUEs

Step plots indicated that there are incremental changes in the indices when effects were introduced into the VAST successively (**Fig. 9**). The influence plot was shown in **Figure 10**, the spatial and SST had an apparently seasonal influence pattern on the temporal CPUE trend, but the SST effect had the smallest impact (influence value is close to one) among four variables. For the fleet effect, the influence value was close to 1 and variable during 2001 - 2014, however the influence index was increased after 2015. The spatiotemporal effect had the largest influence on the time series of estimated CPUE among all variables. Therefore, the temporal CPUE trend is slightly similar to the influence pattern of the spatiotemporal effect.

3.3 Standardized CPUE index

The estimated year-month density values from VAST were shown in **Figure 11**. The annual relative biomass density trend indicated there was a fluctuated pattern over studied periods (**Fig. 11**). The annual relative density was at the lowest level below average (2001 - 2020) in 2020. The summary of year-month and annual standardized CPUEs by the VAST compared with the nominal CPUEs were shown in **Figure 12** and **Table 3**. Compared with the results of the previous joint CPUE analysis, the estimated yearly trends of standardized CPUE are not much different (**Fig. 13**). It is noting that the uncertainties of estimated yearly CPUE for 2019 and 2020 are smaller than in previous years. It may be because the observed CPUE data of these two years had a wider coverage compared to data of 2001 - 2018 (**Figs. 2 - 4**), therefore more observed data could provide the model for estimation, especially for the periphery of the study area.

Previous study has suggested that the spatio-temporal modeling platform VAST achieved the best performance among nine CPUE standardization methods by using the simulation testing, namely generally had one of the lowest biases, one of the lowest mean absolute errors, and the probability of the true index been included by the estimated 50% confidence interval is closest to 50% (Grüss et al., 2019). We also recommend using VAST from a practical standpoint that the regional weights, the year-quarter standardized indices, and the corresponding standard errors can be estimated directly as part of the

modelling procedure, so no additional step is required to produce them (often not been reported).

For most of the standardizing CPUE analysis, they focused on the resulting abundance indices arising from the standardization CPUE models, while ignoring to understand the standardization effects achieved by including each of the explanatory variables in the model. This study used the influence analysis to tease apart the effects of a CPUE standardization model to gain more insight and hence generate greater confidence in the standardization process. We recommend that the results should be considered in the Pacific saury stock assessment.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.

Bishop, J., Venables, W. N., Dichmont, C. M., and Sterling, D. J. (2008). Standardizing catch rates: is logbook information by itself enough?. *ICES J. Mar. Sci.*, 65(2), 255-266.

Bentley, N., Kendrick, T. H., Starr, P. J., and Breen, P. A. (2012). Influence plots and metrics: tools for better understanding fisheries catch-per-unit-effort standardizations. *ICES J. Mar. Sci.*, 69(1):84-88.

Fukushima, S. (1979) Synoptic analysis of migration and fishing conditions of saury in the northwest Pacific Ocean. Bull. Tohoku Reg. *Fish. Res. Lab.* 41: l-70. (In Japanese, English abstract.)

Guisan, A., Edwards Jr., T.C., and Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Mod.* 157, 89-100

Grüss, A., Walter III, J. F., Babcock, E. A., Forrestal, F. C., Thorson, J. T., Lauretta, M. V., and Schirripa, M. J. (2019). Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. *Fish Res.*, 213, 75-93.

Kai, M., Thorson, J. T., Piner, K. R., and Maunder, M. N. (2017). Spatiotemporal variation in size-structured populations using fishery data: an application to shortfin mako (*Isurus oxyrinchus*) in the Pacific Ocean. *Can. J. Fish Aquat. Sci.*, 74(11), 1765-1780.

Maunder, M. N., and Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fish Res*., 70(2-3):141-159.

NPFC TWG-PSSA. 2018. Report of 3rd Meeting of the Technical Working Group on Pacific Saury Stock Assessment. NPFC-2018-TWG PSSA03-Final Report. 29 pp.

R Development Core Team (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. Retrieved from http://www.R-project.org

Shelton, A. O., Thorson, J. T., Ward, E. J., and Feist, B.E. (2014).Spatial semiparametric models improve estimates of species abundance and distribution. *Can. J. Fish Aquat. Sci.* 71(11): 1655-1666.

Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES J. Mar. Sci*., 72, 1297-1310.

Thorson, J. T., and Kristensen, K. (2016). Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. *Fish Res*., 175, 66-74.

Thorson, J. T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fish Res.*, 210:143-161.

Walter, J. F., Hoenig, J. M., and Christman, M. C. (2014). Reducing bias and filling in spatial gaps in fishery-dependent catch-per-unit-effort data by geostatistical prediction, I. Methodology and simulation. *N. Am. J. Fish Manage.*, 34(6), 1095-1107.

Table 1. Summary of explanatory variables used in VAST.

Variables		Number of categories	Detail	Note
Year-month	t	156	$t = 1$ (2001/May) - $t = 156$ (2020/Dec)	
Spatial knot	s	100	32 - 50°N and 140 - 176°E	See **Figure 5**
Sea surface temperature	*SST*	1	Continues variable (3 - 25°C)	
Fleet	Fleet	7	JP1: Japanese vessel less than 100 GRT; JP2: Japanese vessel larger than 100 GRT; CT: Chinese Taipei; CN: China; RS: Russia; KR: Korea; VU: Vanuatu	

Table 2. Summary of the model selection information from VAST.

Model No.	Model structure	Number of parameters (fixed effect)	Deviance	AIC	Maximum gradient
V-1	*Year-month*	156	42771	85551	< 0.0001
V-2	*Year-month + Spatial*	160	42383	85092	< 0.0001
V-3	*Year-month + Spatial + Spatio-temporal*	161	41921	84170	< 0.0001
V-4	*Year-month + Spatial + Spatio-temporal + Fleet*	168	41680	83705	< 0.0001
V-5	*Year-month + Spatial + Spatio-temporal + Fleet + SST*	169	41669	83685	< 0.0001

Table 3. Annual relative (relative to mean) nominal and standardized indices from VAST for Pacific saury during 2001 and 2020 in the Northwestern Pacific Ocean. Std. CPUE = standardized CPUE, SD = standard error, lower and upper = lower and upper limits of the 95% confidence intervals.

Year	Nominal CPUE	Std. CPUE	SD	Lower	Upper	CV
2001	0.81	0.72	0.22	0.50	0.94	0.31
2002	0.66	0.63	0.19	0.44	0.82	0.30
2003	1.11	1.21	0.36	0.85	1.58	0.30
2004	1.09	1.04	0.31	0.73	1.35	0.30
2005	1.65	1.72	0.51	1.21	2.23	0.30
2006	1.33	0.78	0.21	0.57	0.99	0.27
2007	1.33	1.24	0.34	0.90	1.59	0.28
2008	1.59	1.68	0.47	1.20	2.15	0.28
2009	0.99	0.99	0.29	0.70	1.28	0.29
2010	0.91	0.92	0.27	0.65	1.19	0.29
2011	1.07	1.24	0.40	0.84	1.63	0.32
2012	0.96	1.06	0.35	0.71	1.41	0.33
2013	0.96	0.85	0.23	0.61	1.08	0.28
2014	1.27	1.36	0.35	1.01	1.72	0.26
2015	0.87	0.84	0.25	0.59	1.09	0.30
2016	0.75	0.75	0.21	0.54	0.95	0.28
2017	0.60	0.85	0.26	0.59	1.11	0.31
2018	1.17	1.37	0.41	0.96	1.79	0.30
2019	0.48	0.45	0.10	0.35	0.56	0.23
2020	0.40	0.29	0.08	0.21	0.37	0.28
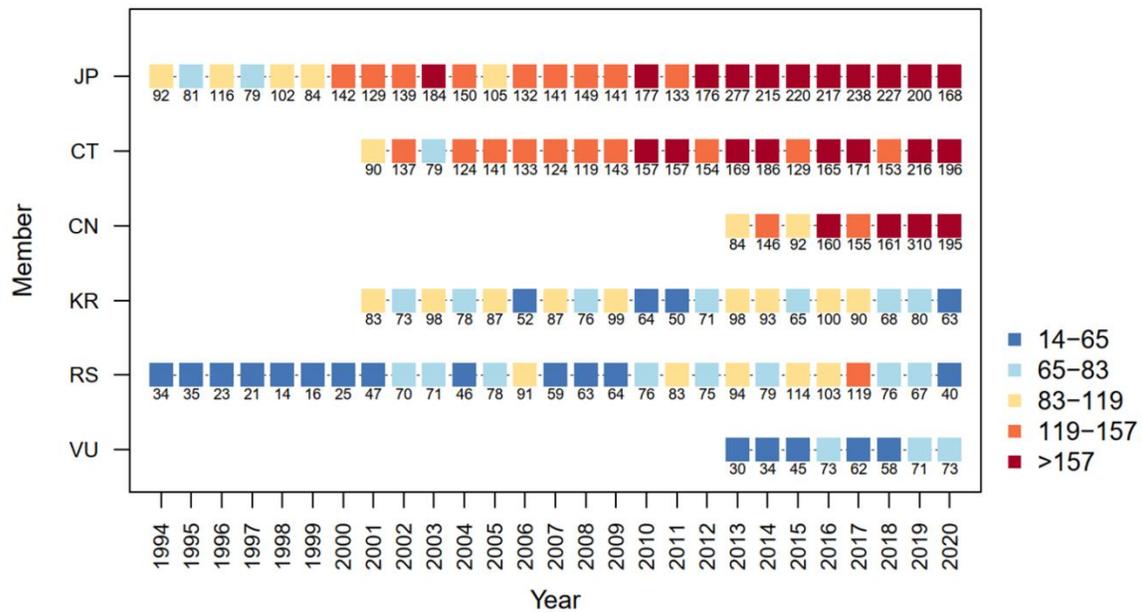
Figure 1. Distribution of fishing effort (i.e., 1 by 1 grid) by members (JP = Japan; CT = Chinese Taipei; CN = China; KR = Korea; RS = Russia; VU = Vanuatu) derived from the joint CPUE dataset of Pacific saury during 1994 - 2020.
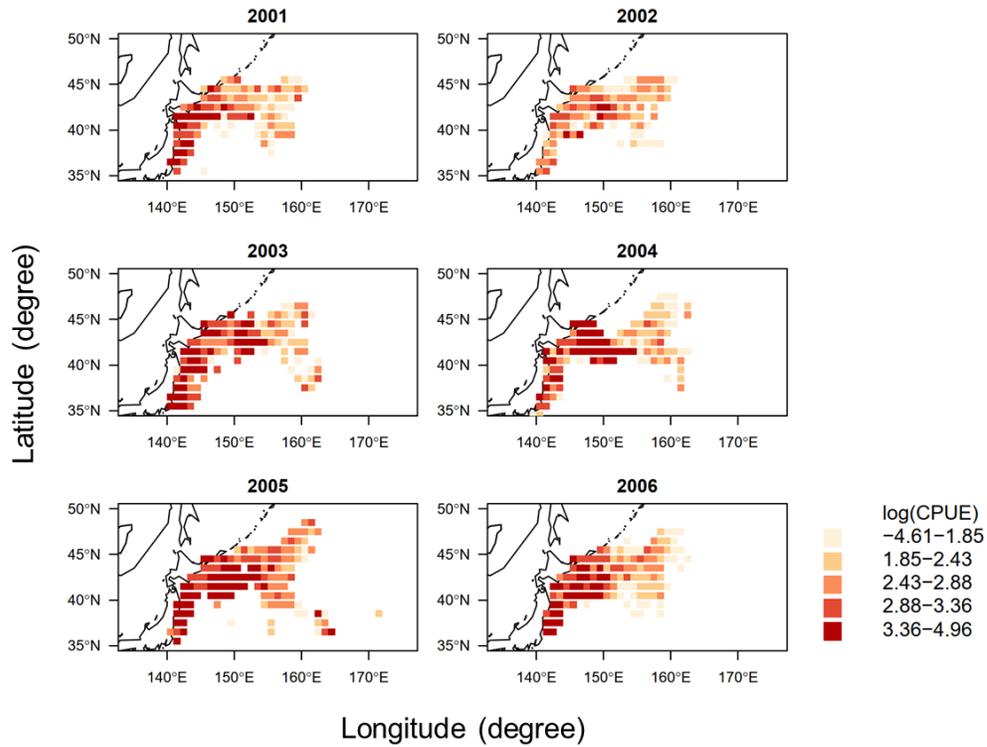
Figure 2. Spatial and temporal distribution of the logarithmic nominal CPUE (in metric ton per operating day fished) of Pacific saury during 2001 - 2006 in the Northwestern Pacific Ocean.
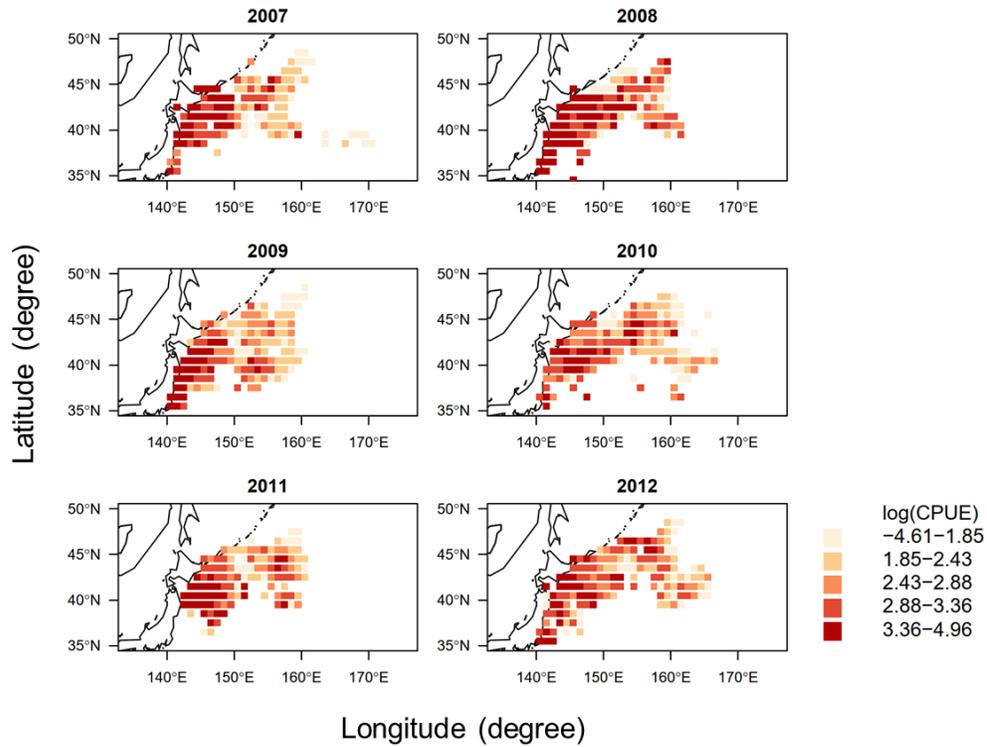
Figure 3. Spatial and temporal distribution of the logarithmic nominal CPUE (in metric ton per operating day fished) of Pacific saury during 2007 - 2012 in the Northwestern Pacific Ocean.
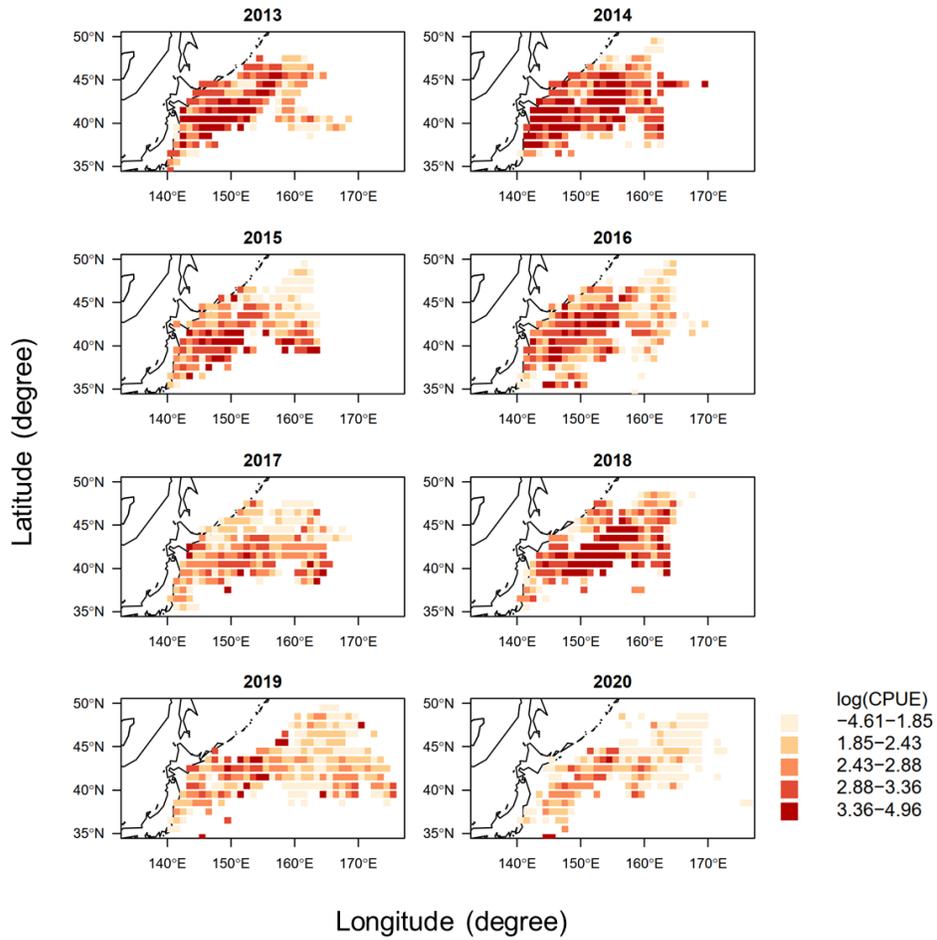
Figure 4. Spatial and temporal distribution of the logarithmic nominal CPUE (in metric ton per operating day fished) of Pacific saury during 2013 - 2020 in the Northwestern Pacific Ocean.
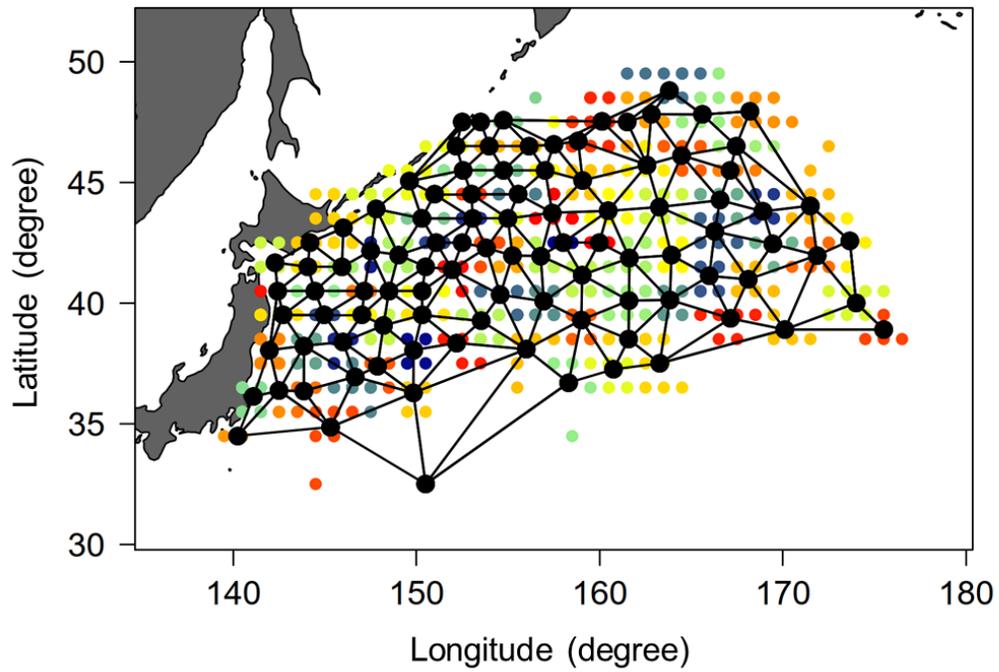
Figure 5. Mesh used to fit the geostatistical model (VAST). An effect is estimated for each of the 100 spatial knots (black).The colored circles grouped by knots indicate the locations of spatial observations of the Pacific saury from 2001 to 2020 within the 1°×1° grid.
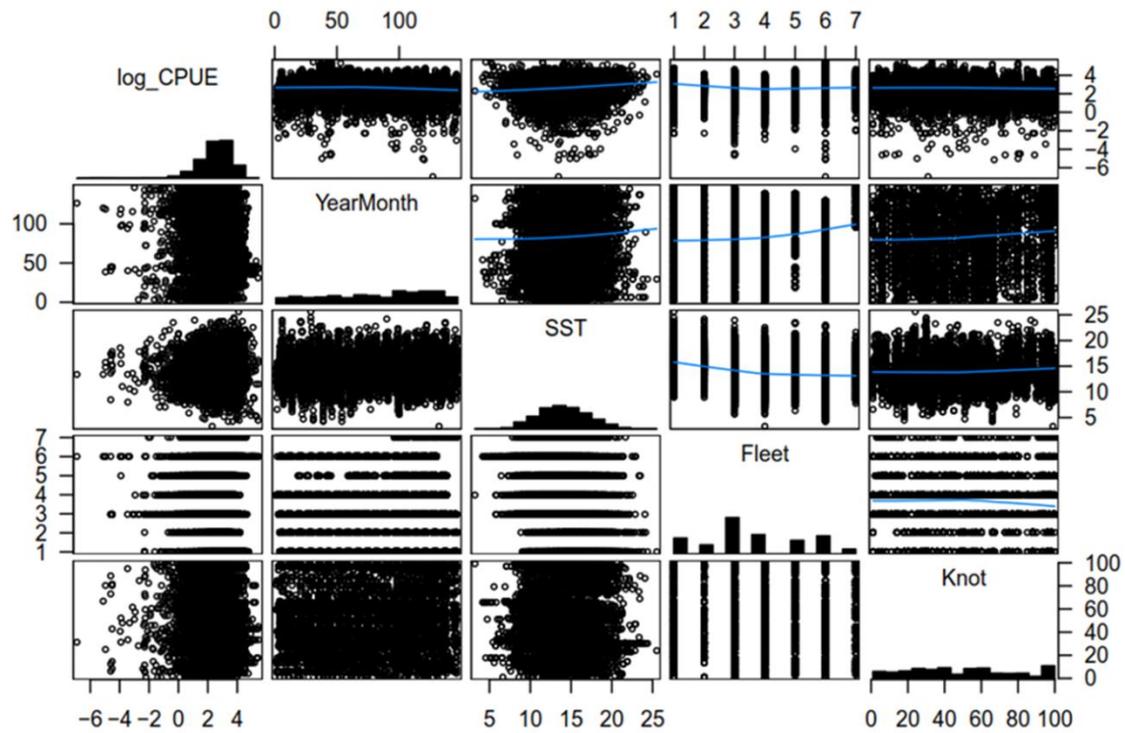
Figure 6. Correlation matrix of explanatory variables used in VAST analysis. The blue
curves in the upper triangular matrix denote the loess smooth curves.
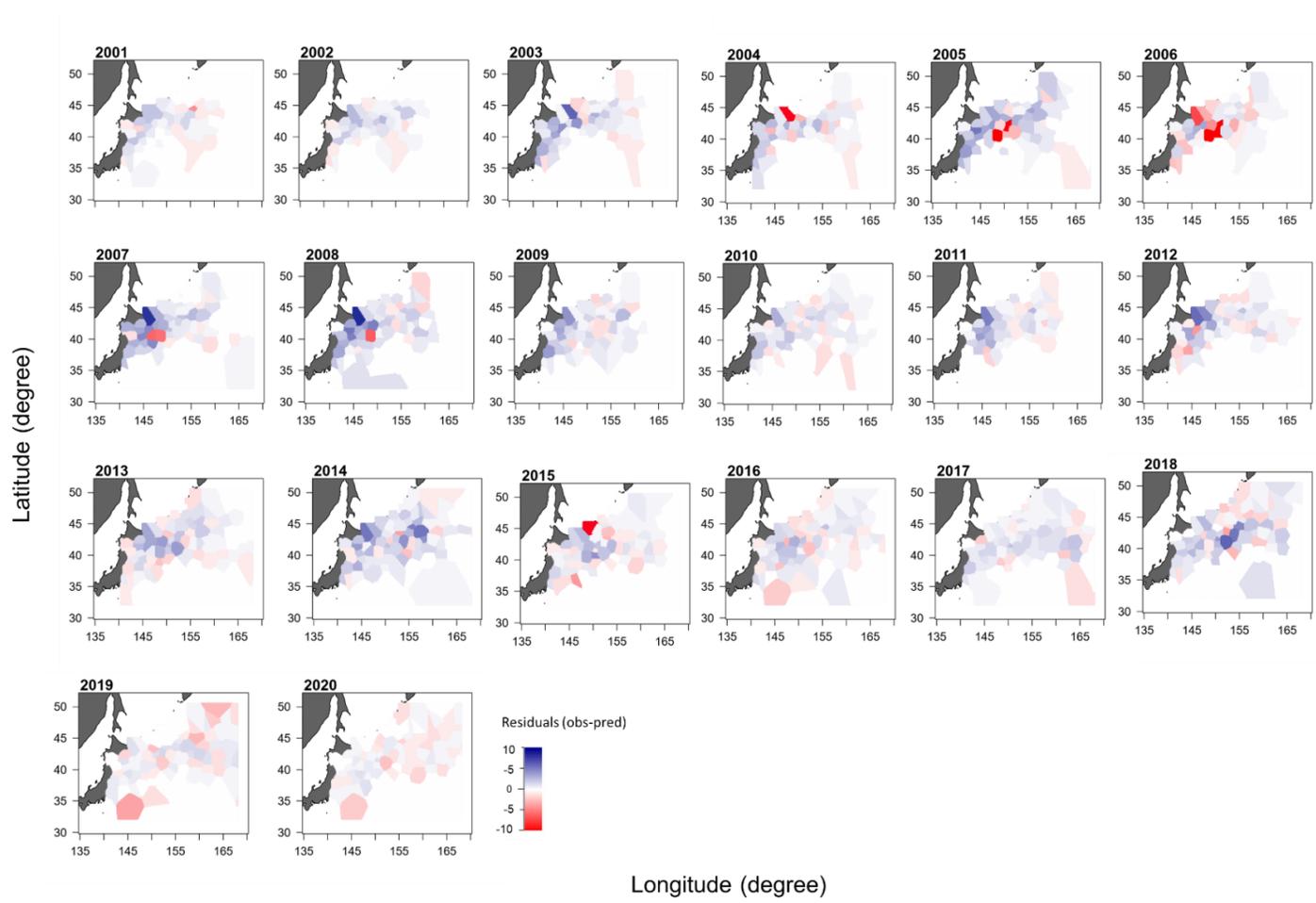
Figure 7. Spatial distribution of yearly aggregated residuals of Pacific saury derived from the VAST from 2001 to 2020.
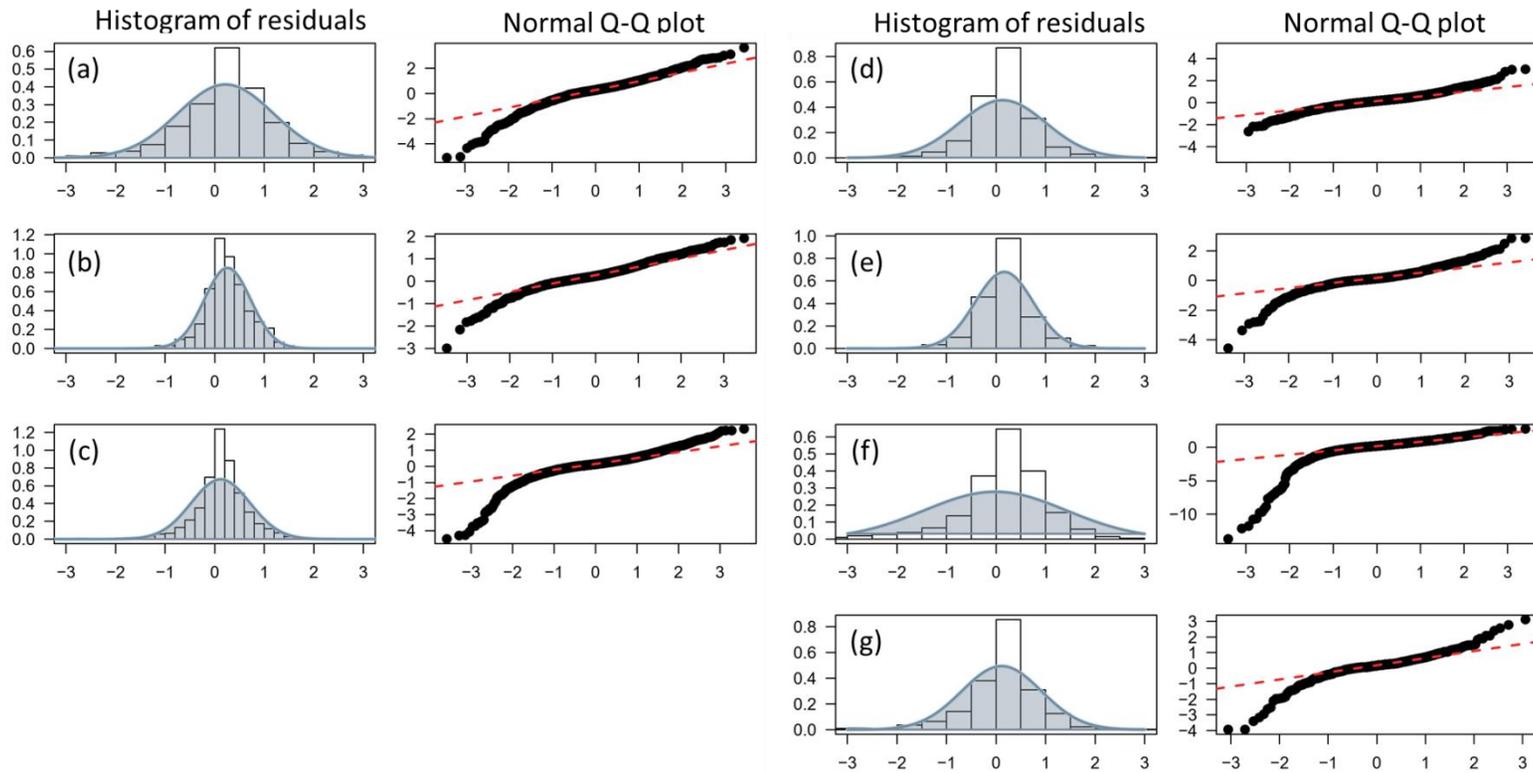
Figure 8. Diagnostic plots of the fitted VAST. The histogram of residuals (left) and Q-Q plot (right) from (a) Japanese fisheries by vessels of <100 GRT; (b) Japanese fisheries by vessels of >= 100 GRT; (c) Chinese Taipei; (d) Korea; (e) China; (f) Russia, and (g) Vanuatu fisheries.
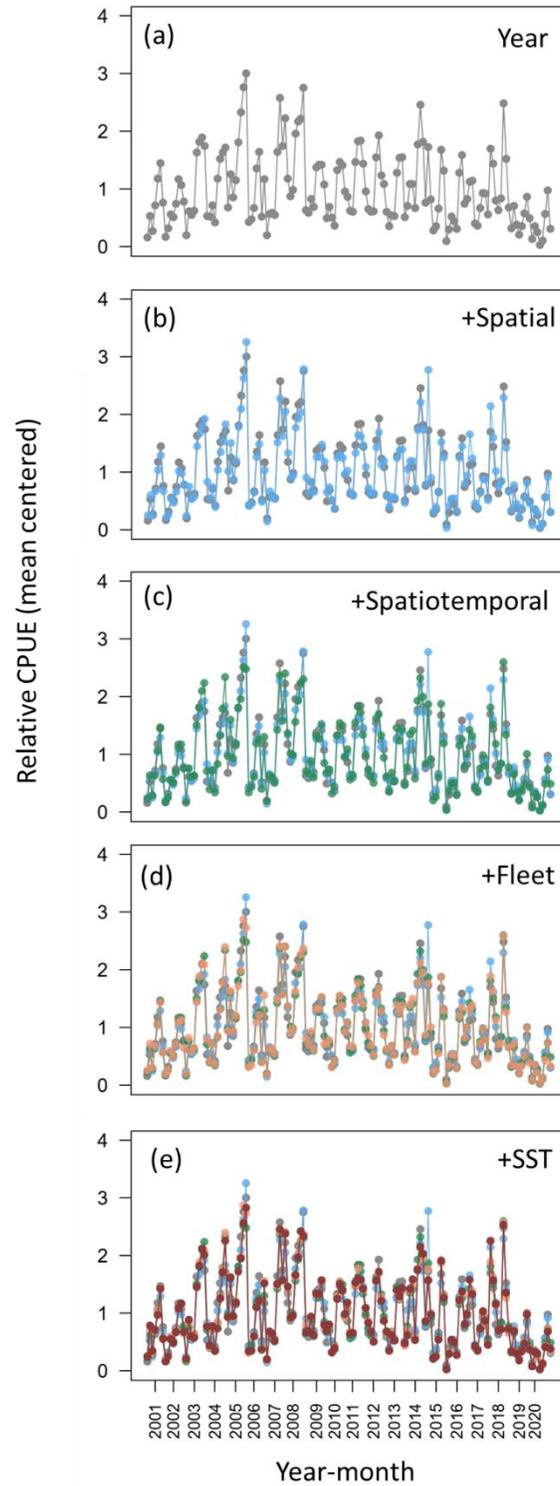
Figure 9. Step plots showing the effects of individual factors with respect to the estimated CPUE indices.
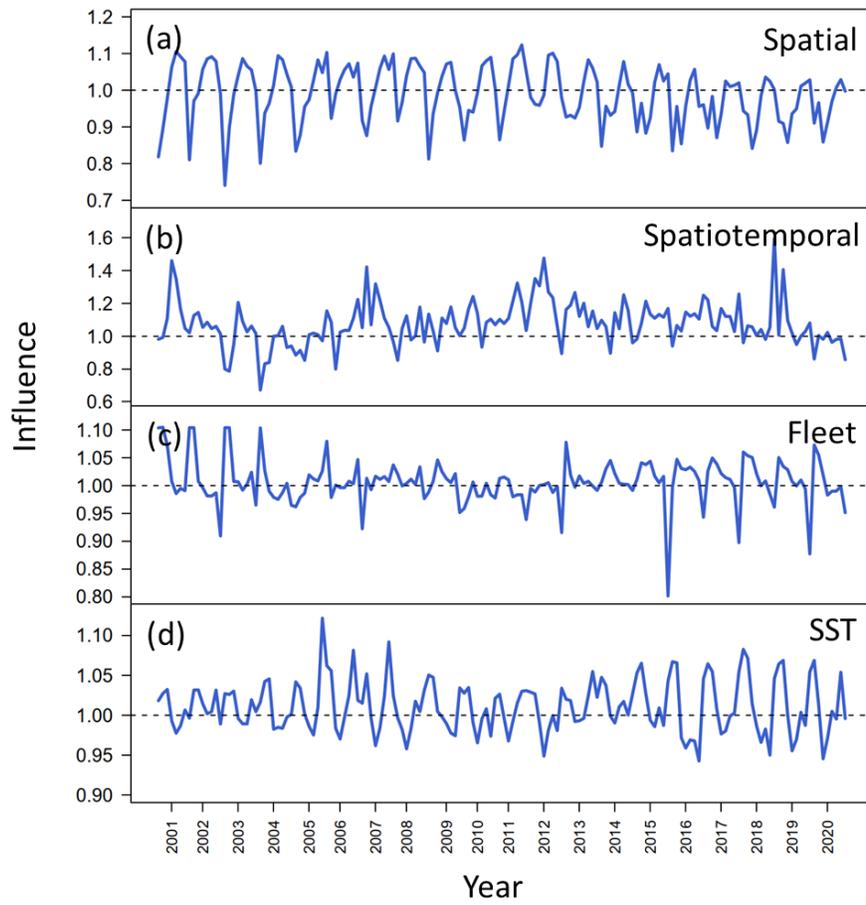
Figure 10. Time-series of influence indices of (a) spatial, (b) spatiotemporal, (c) fleet and (d) SST effect for the VAST during 2001 - 2020.
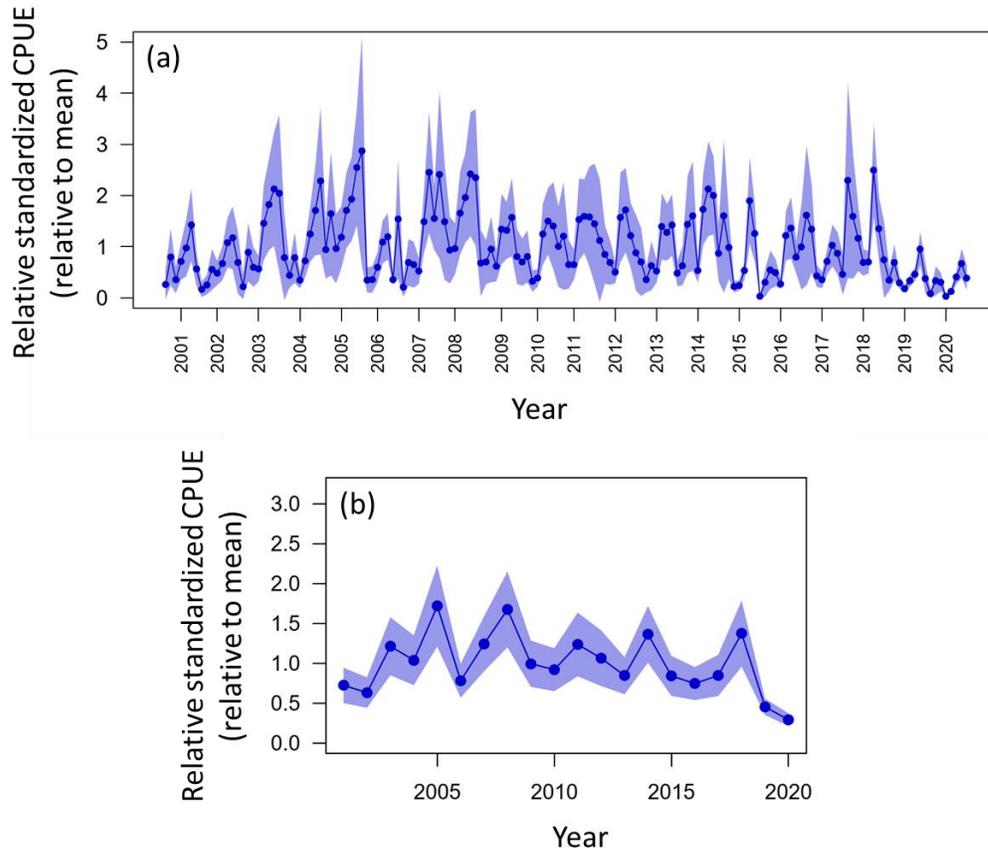
Figure 11. Time-series of (a) year-month, and (b) yearly relative standardized indices (relative to mean) from the VAST for the Pacific saury in Northwest Pacific Ocean from 2001 to 2020. The shaded areas denote the 95% confidence intervals.
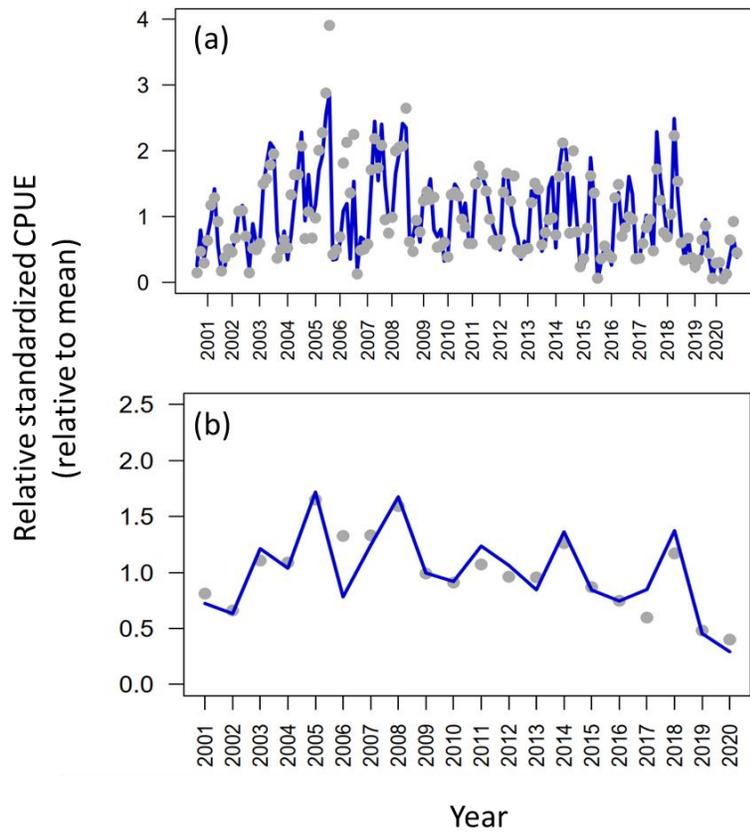
Figure 12. Time-series of (a) year-month and (b) yearly relative standardized indices in comparison to the nominal indices for the Pacific saury in Northwest Pacific Ocean from 2001 to 2020.
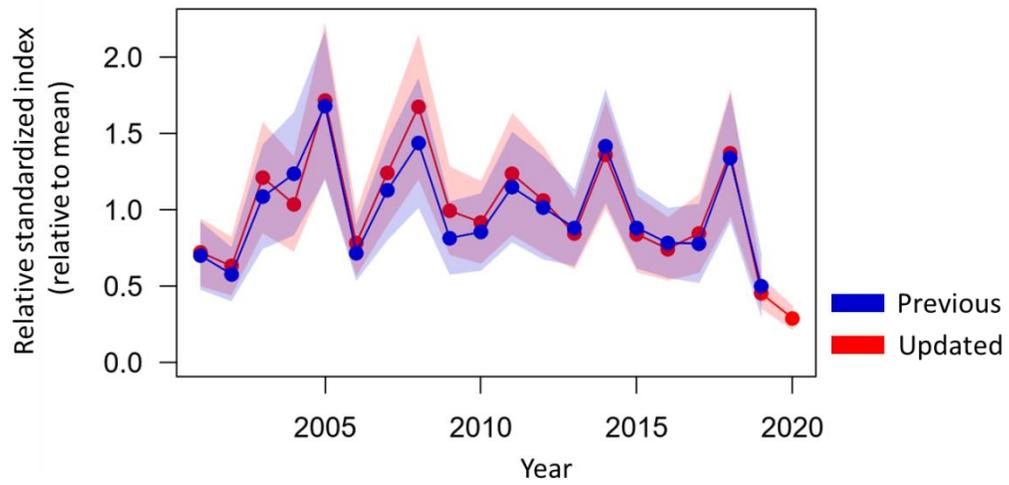
Figure 13. Comparison of time series of annual relative standardized indices (relative to mean) from the previous (2001 - 2019) and updated (2001 - 2020) joint CPUE dataset for the Pacific saury in the Northwest Pacific Ocean. The shaded areas denote the 95%.

Appendix figures
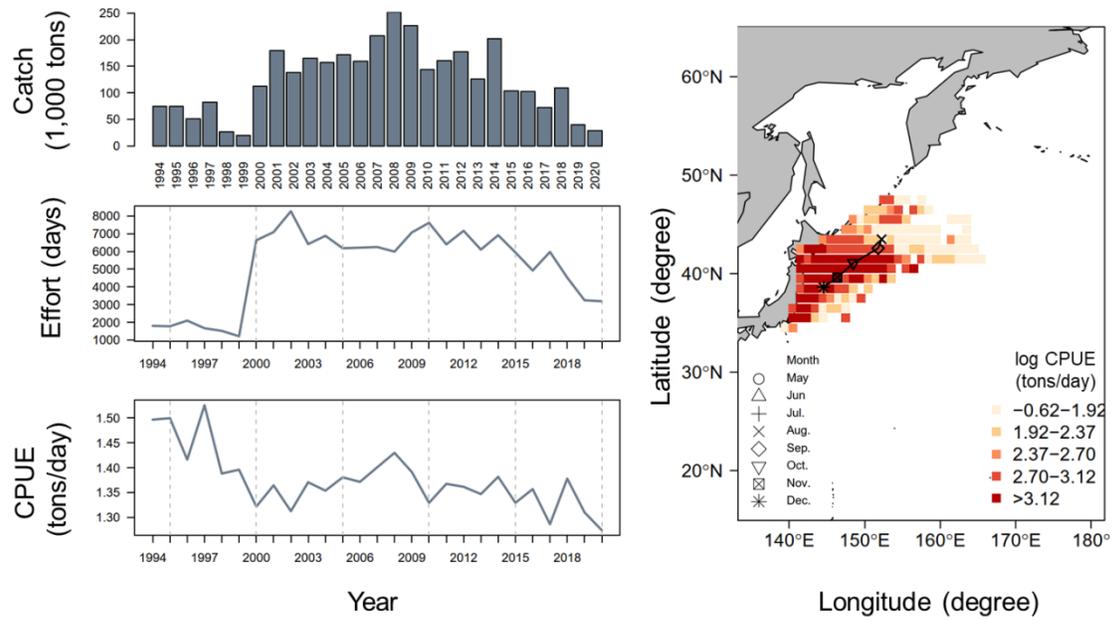
Figure A1. The summary plot of catch (in 1,000 tons), effort (in operating day), nominal CPUE (in tons/day), and the spatial distribution of CPUE from 1994 - 2020 for Pacific saury collected from Japan. The symbol of map represents the monthly centroid of gravity of nominal CPUE over years.
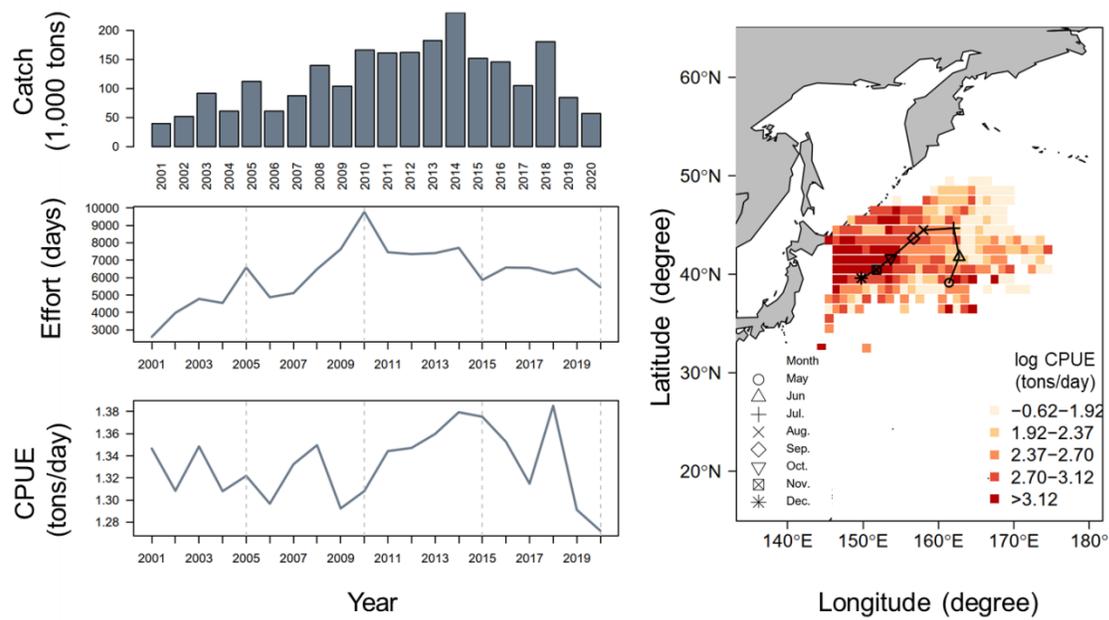
Figure A2. The summary plot of catch (in 1,000 tons), effort (in operating day), nominal CPUE (in tons/day), and the spatial distribution of CPUE from 2001 - 2020 for Pacific saury collected from Chinese Taipei. The symbol of map represents the monthly centroid of gravity of nominal CPUE over years.
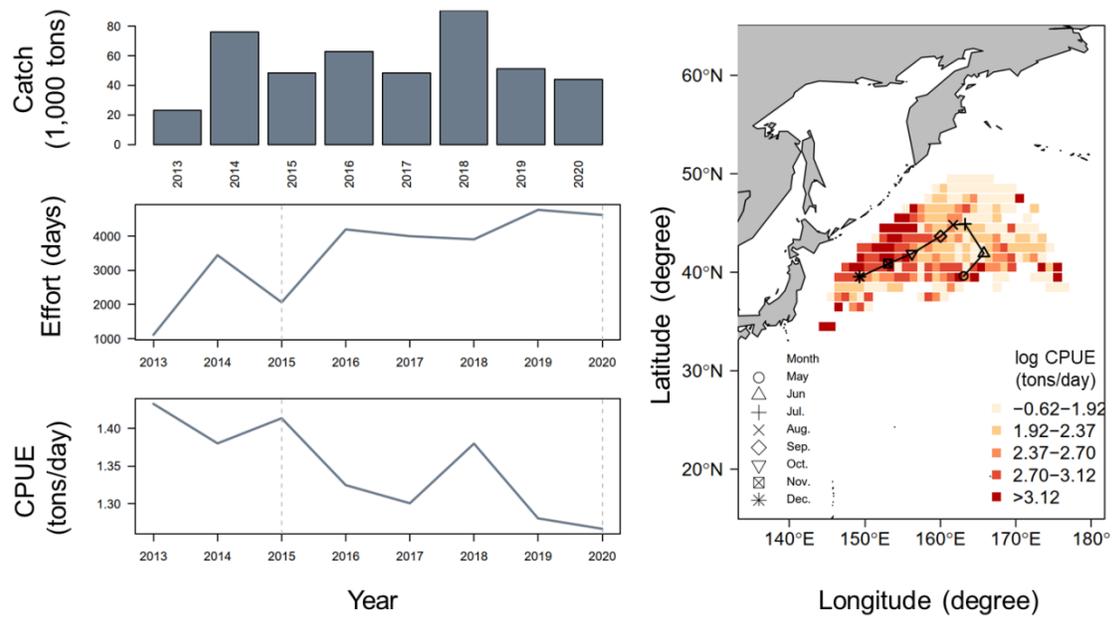
Figure A3. The summary plot of catch (in 1,000 tons), effort (in operating day), nominal CPUE (in tons/day), and the spatial distribution of CPUE from 2013 - 2020 for Pacific saury collected from China. The symbol of map represents the monthly centroid of gravity of nominal CPUE over years.
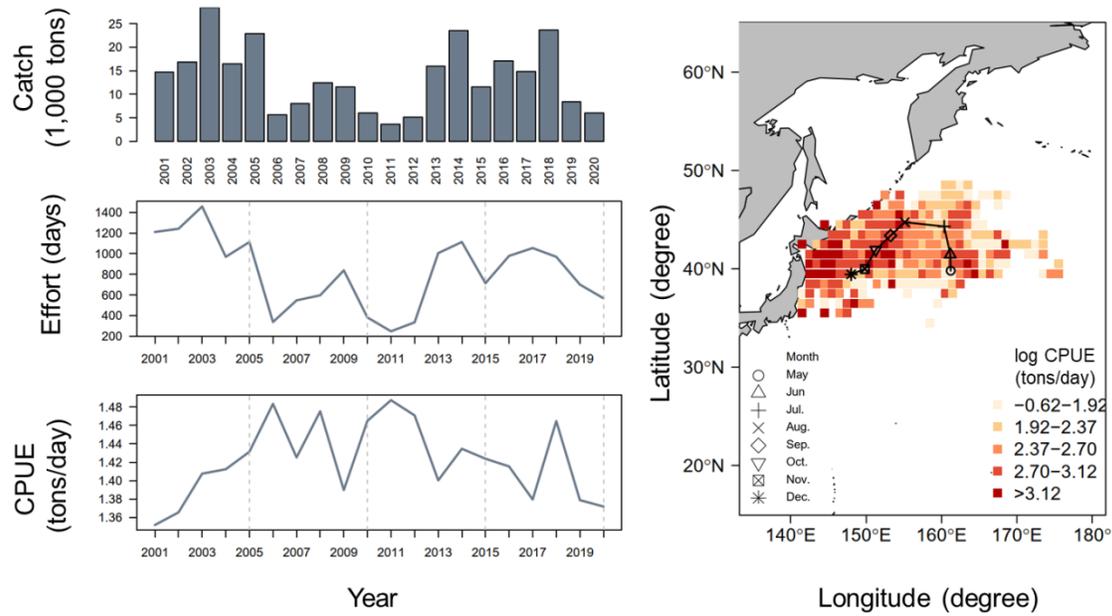
Figure A4. The summary plot of catch (in 1,000 tons), effort (in operating day), nominal CPUE (in tons/day), and the spatial distribution of CPUE from 2001 - 2020 for Pacific saury collected from Korea. The symbol of map represents the monthly centroid of gravity of nominal CPUE over years.
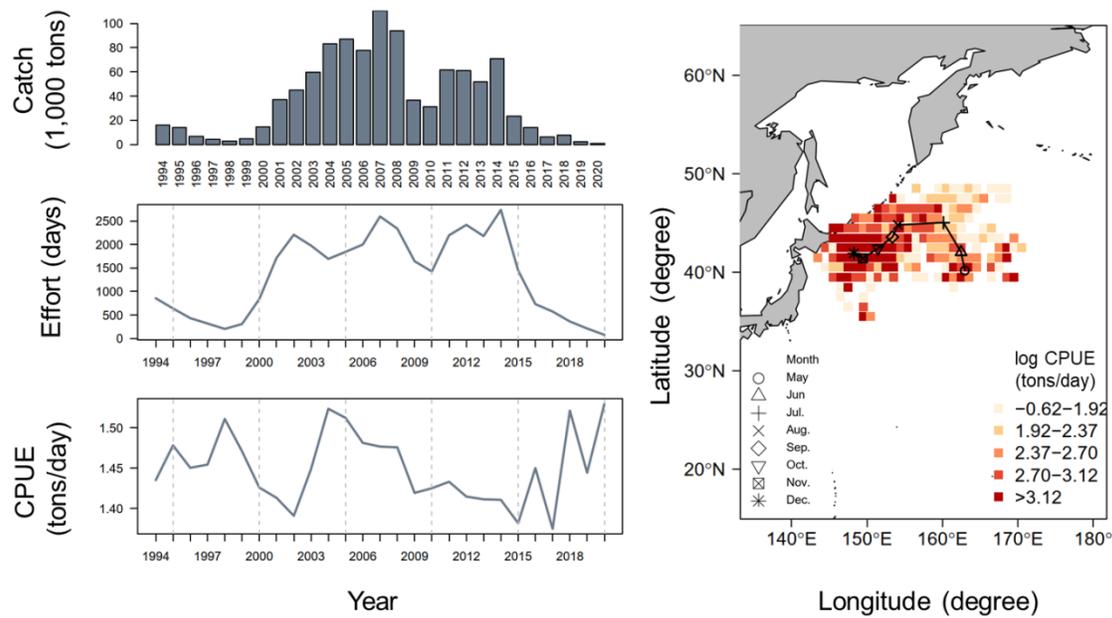
Figure A5. The summary plot of catch (in 1,000 tons), effort (in operating day), nominal CPUE (in tons/day), and the spatial distribution of CPUE from 1994 - 2020 for Pacific saury collected from Russia. The symbol of map represents the monthly centroid of gravity of nominal CPUE over years.
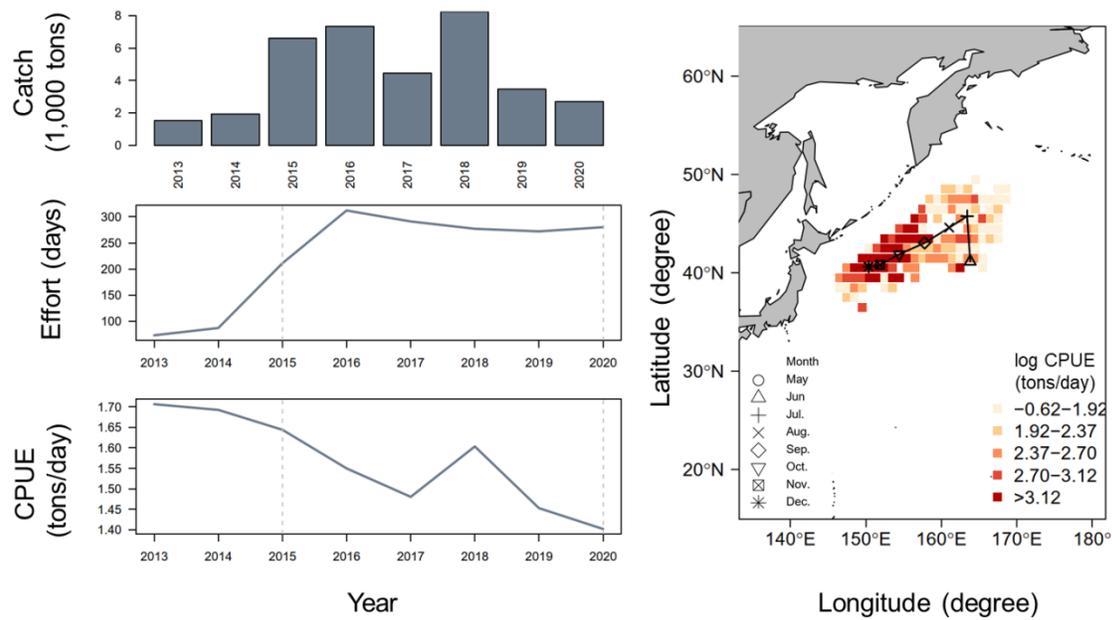
Figure A6. The summary plot of catch (in 1,000 tons), effort (in operating day), nominal CPUE (in tons/day), and the spatial distribution of CPUE from 2013 - 2020 for Pacific saury collected from Vanuatu. The symbol of map represents the monthly centroid of gravity of nominal CPUE over years.