North Pacific Fisheries Commission

NPFC-2019-SSC PS05-WP10

Joint CPUE standardization of the Pacific saury in the Northwest Pacific Ocean during 2001-2017 by using the conventional and geostatistical approaches

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 an index standardization by using the joint CPUE (catch-per-unit-effort) data of the Pacific Saury in the Northwest Pacific Ocean during 2001 and 2017. Furthermore, we provided a comparison of the CPUE standardization between VAST and the conventional generalized linear model. The objective is to make a suggestion for the appropriate specification of the joint CPUE index to be used in the future Pacific saury stock assessment. The results indicated that VAST performs better than the GLM with less residuals depart from zero and smaller residual variance. We recommend using VAST for deriving the standardized joint index as improved input data in the stock assessment. The analysis we presented is generally applicable and should be considered as a standard tool in the CPUE standardization.

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).

In addition, observations that occur 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 modeling 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).

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 by using the joint CPUE data of the Pacific saury in the Northwest Pacific Ocean during 2001 and 2017. Furthermore, we provide a comparison of the CPUE standardization between VAST and the conventional model (i.e., GLM). The objective is to make a suggestion for the appropriate specification of the joint CPUE index to be used in the future Pacific saury stock assessment. Progress in joint standardized CPUE should result in better assessment and management of the stock.

2. Materials and methods

2.1 Joint CPUE dataset

The joint CPUE data of stick-held dip net fisheries was collected from each member including Chinese Taipei, China, Japan, Korea, Russia and Vanuatu in the North Pacific Fisheries Commission (NPFC). This 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 – 171°E from 2001 to 2017. Data grooming was applied prior to the standardization to remove the monthly observation with less than 10 operation days. In total, 0.2% records have been removed. CPUE was defined as a catch of Pacific saury in metric ton per operating day fished.

2.2 Conventional CPUE standardization

We use a log offset GLM to standardize CPUE as the following description:

$$\log(C(s,t)) \ \sim \ YearMonth(t) + cell(s) + \sum_{j=1}^{n_j} \gamma_j x_j(s,t) + \sum_{k=1}^{n_k} \lambda(k) Q(k) + \log(Op\_day(s,t))$$

where $C(s,t)$ is the prediction of Pacific saury catch (in metric ton) in the 1°×1° cell s and year-month t, $YearMonth(t)$ is the fixed effect for each year-month t (131 time steps), $cell(s)$ is the fixed effect for the 1°×1° spatial cell s (274 cells), γ_j represents the impact of covariate j (i.e., linear impact of SST, $n_j = 1$;) with value $x_j(s,t)$ on catch for cell s and year-month t. λ_k is the coefficient for the catchability covariate $Q(k)$ (i.e., fleet, $n_k = 1$), and $Op\_day(s,t)$ is the fishing effort (operating day) as a log offset in cell s and year-month t.

2.3 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 used 100 spatial knots (see **Figure 1** for the configuration) to approximate the spatial and spatio-temporal autocorrelated variations.

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 et al., 2019). The logarithm prediction of Pacific saury biomass density, $p(s,t)$, in knot s and year-month t is described below:

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

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 spatial-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. λ_k is the coefficient for the catchability covariate $Q(k)$ (i.e., fleet, $n_k = 1$).

2.4 Model diagnostics

Histograms of the residuals were used to assess normality for the GLM and VAST, in addition, the quantile-quantile normal probability plots (Normal Q-Q plot) for both of them. 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 GLM and VAST with respect to the estimated CPUE indices.

2.5 Standardized CPUE trends

Predictions of standardized Pacific saury biomass 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. The standardized index for GLM and VAST is respectively described as below:

1) Conventional CPUE model (GLM)

$$\log(C(s,t)) = YearMonth(t) + cell(s) + \sum_{j=1}^{n_j} \gamma_j x_j(s,t)$$

$$B(t) = \sum C(s,t) \times a(s)$$

where $B(t)$ is the area re-weighted biomass density in year-month t throughout the population domain, $a(s)$ is the area of the 1°×1° spatial cell s (110 km^2).

2) Geostatistical CPUE model (VAST)

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

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

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.

3. Results and discussion

3.1 Model diagnostic

The histogram and Q-Q plots of both models based on the lognormal distributions appear normal in GLM and VAST for all fleets (**Figs 2 and 3**), which confirms the assumption of error distribution is appropriate for the CPUE standardization. **Figure 4** shows that there is no significant residual pattern for each fixed effect in the GLM, except there are the larger residuals are found in the catchability effect for Chinese Taipei and Russia (**Fig 4b**). For the VAST, a similar result of the residual pattern was found (**Fig. 5**). However, Russia has larger negative residuals compared to other fleets. The results revealed that the VAST yielded higher R^2 (0.59) than did the GLM (0.31). Generally, the VAST performed better than the GLM with less residuals depart from zero and smaller residual variance.

3.2 Comparision of the standardized indices

Step plots provided a clear indication that there are incremental changes in the indices when effects were introduced into the GLM successively (**Fig. 6**). However, for the VAST, the knot variable has a major influence on standardized CPUE compared to the other effects (**Fig. 7**).

The estimated total biomass density value from GLM is lower than the VAST (**Fig. 8a**). Generally, the results of total biomass density from GLM and VAST showed similar trends across time (**Fig. 8b**).

Although there is no clear difference in the annual trends of standardized CPUE indices between the GLM and VAST, we recommend using VAST for deriving the joint index to be used in the future Pacific saury stock assessment according to Grüss et al. (2019). The study has suggested that the spatio-temporal modeling platform VAST achieved the best performance by using the simulation testing, namely generally had one of the lowest biases, one of the lowest mean absolute errors, and 50% confidence interval coverage closest to 50%. Furthermore, we 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).

References

Guisan, A., Edwards Jr., T.C., 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.

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. (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., & 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.

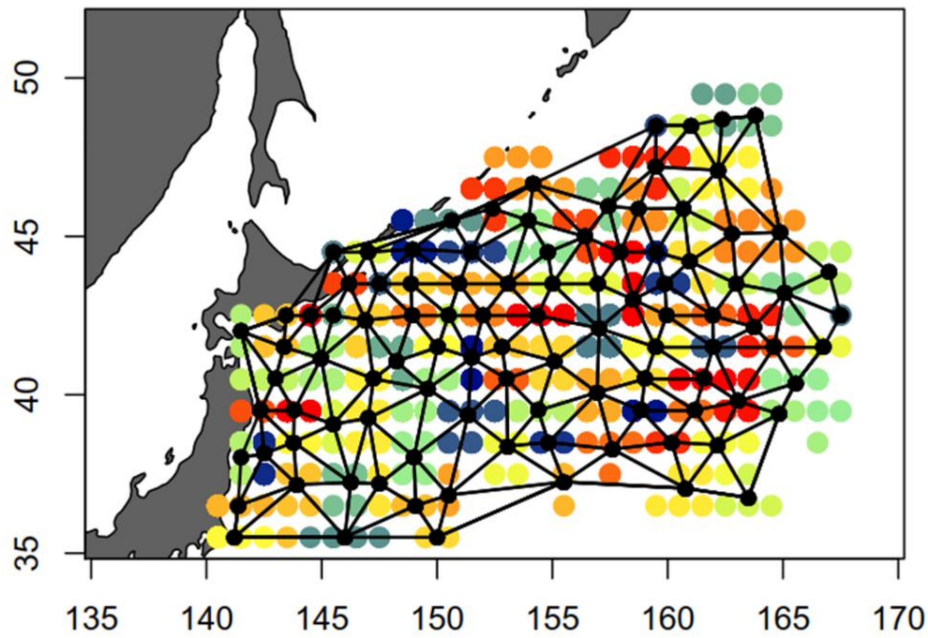

Figure 1. Mesh used to fit the geostatistical model (VAST). An effect is estimated for each of the 100 core knots (black).The colored circles grouped by knots indicate the locations of spatial observations of the Pacific saury from 2001 to 2017 within the 1°×1° grid.

`

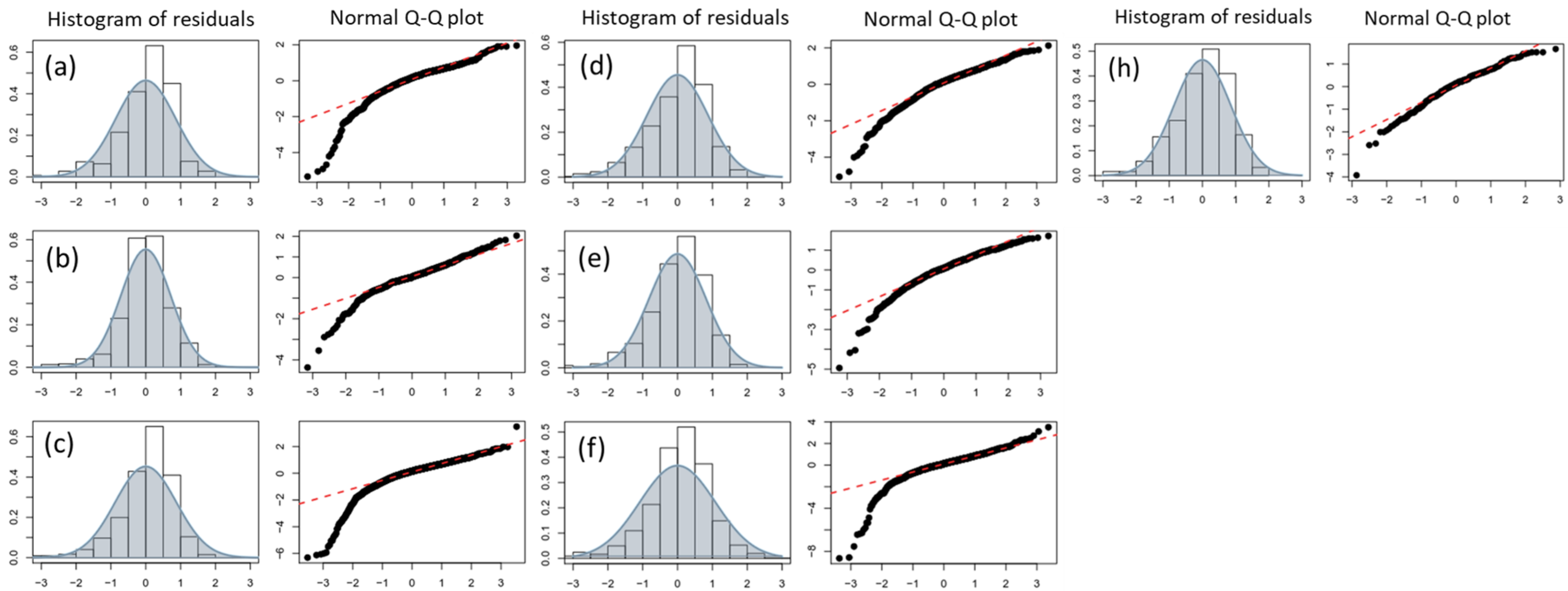

Figure 2. Diagnostic plots of the fitted GLM. The histogram of residuals (left) and Q-Q plot (right) from (a) Japanese fisheries by vessels of <100; (b) Japanese fisheries by vessels of >= 100; (c) Chinese Taipei; (d) Korea; (e) China; (f) Russia, and (g) Vanuatu fisheries.

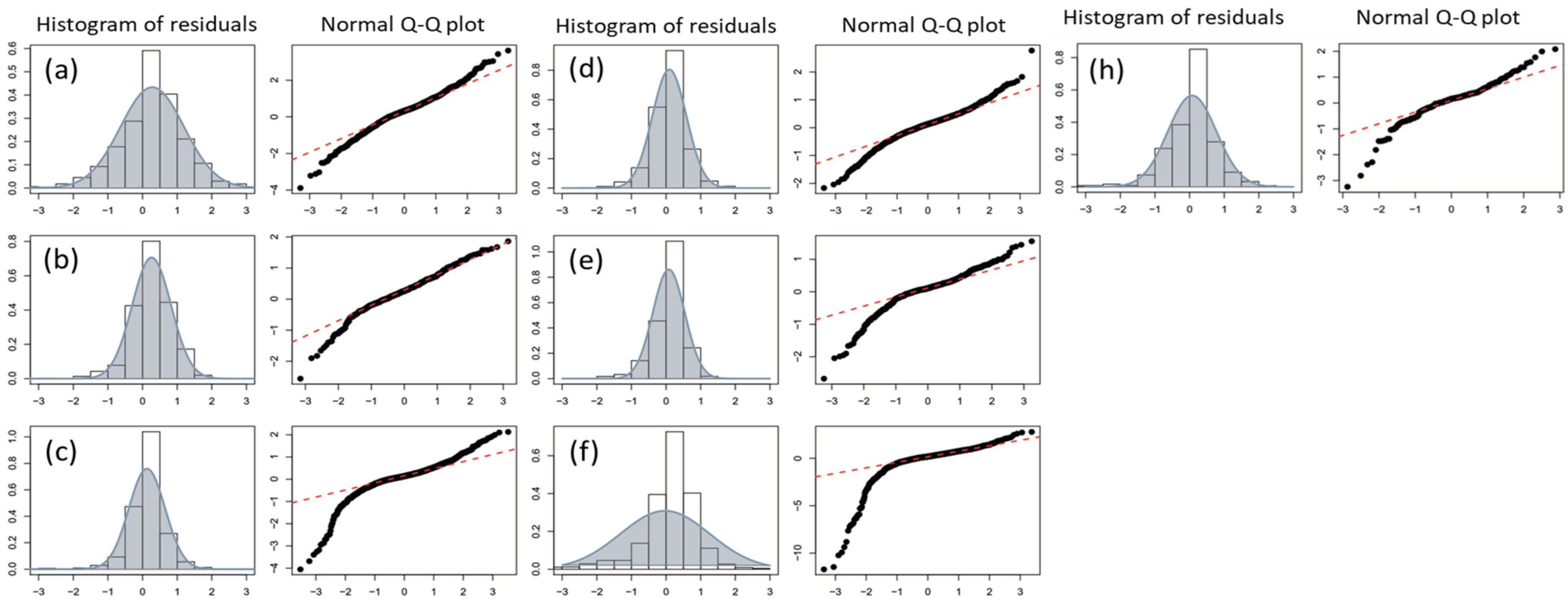

Figure 3. Diagnostic plots of the fitted VAST. The histogram of residuals (left) and Q-Q plot (right) from (a) Japanese fisheries by vessels of <100; (b) Japanese fisheries by vessels of >= 100; (c) Chinese Taipei; (d) Korea; (e) China; (f) Russia, and (g) Vanuatu fisheries.

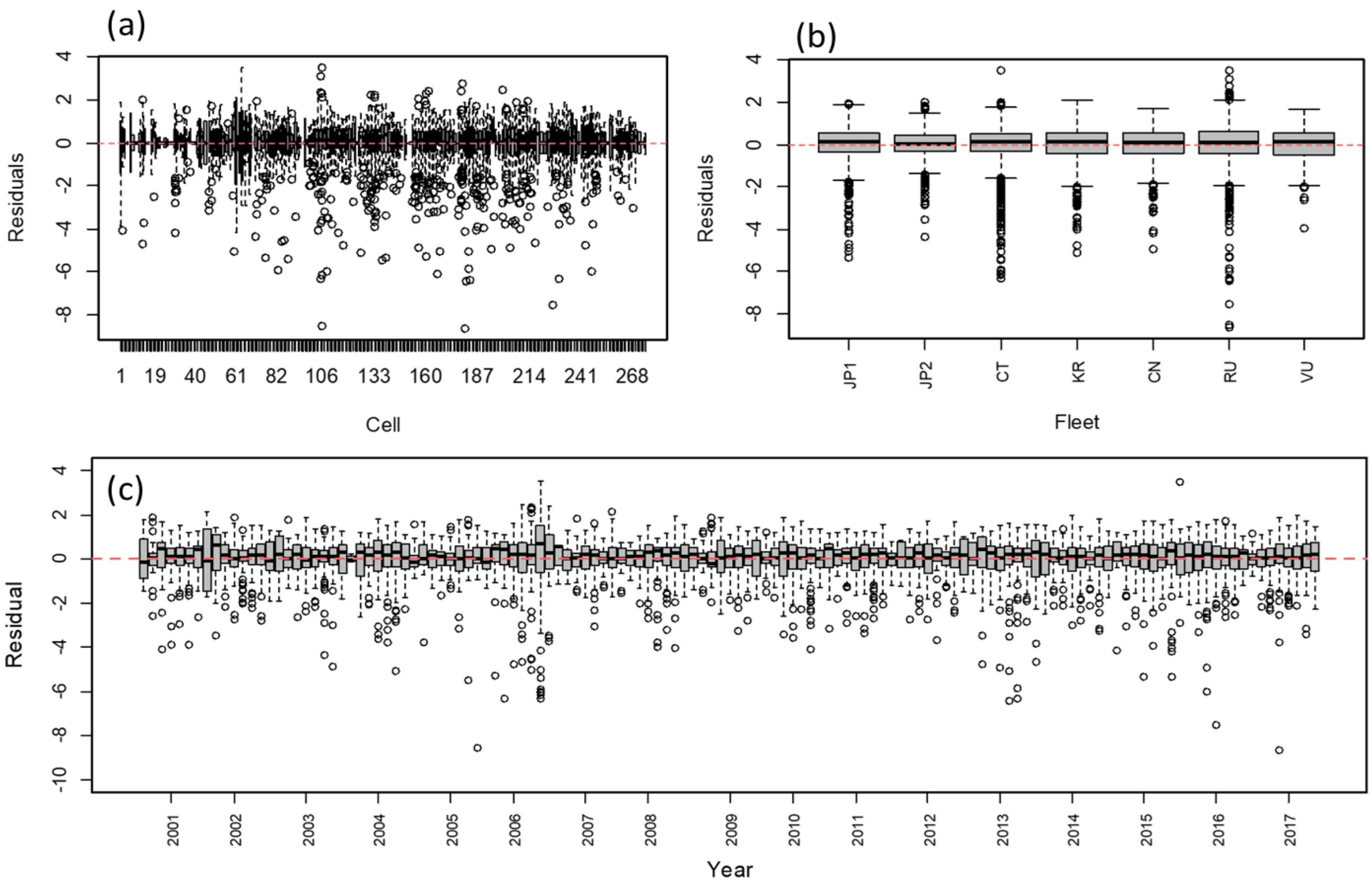

Figure 4. Boxplots of residuals by (a) cells, (b) fleets, and (c) year-month of the fitted GLM. JP1 is Japanese fisheries by vessels of <100; JP2 is Japanese fisheries by vessels of >= 100; CT is Chinese Taipei; KR is Korea; CN is China; RU is Russia, and VU is Vanuatu

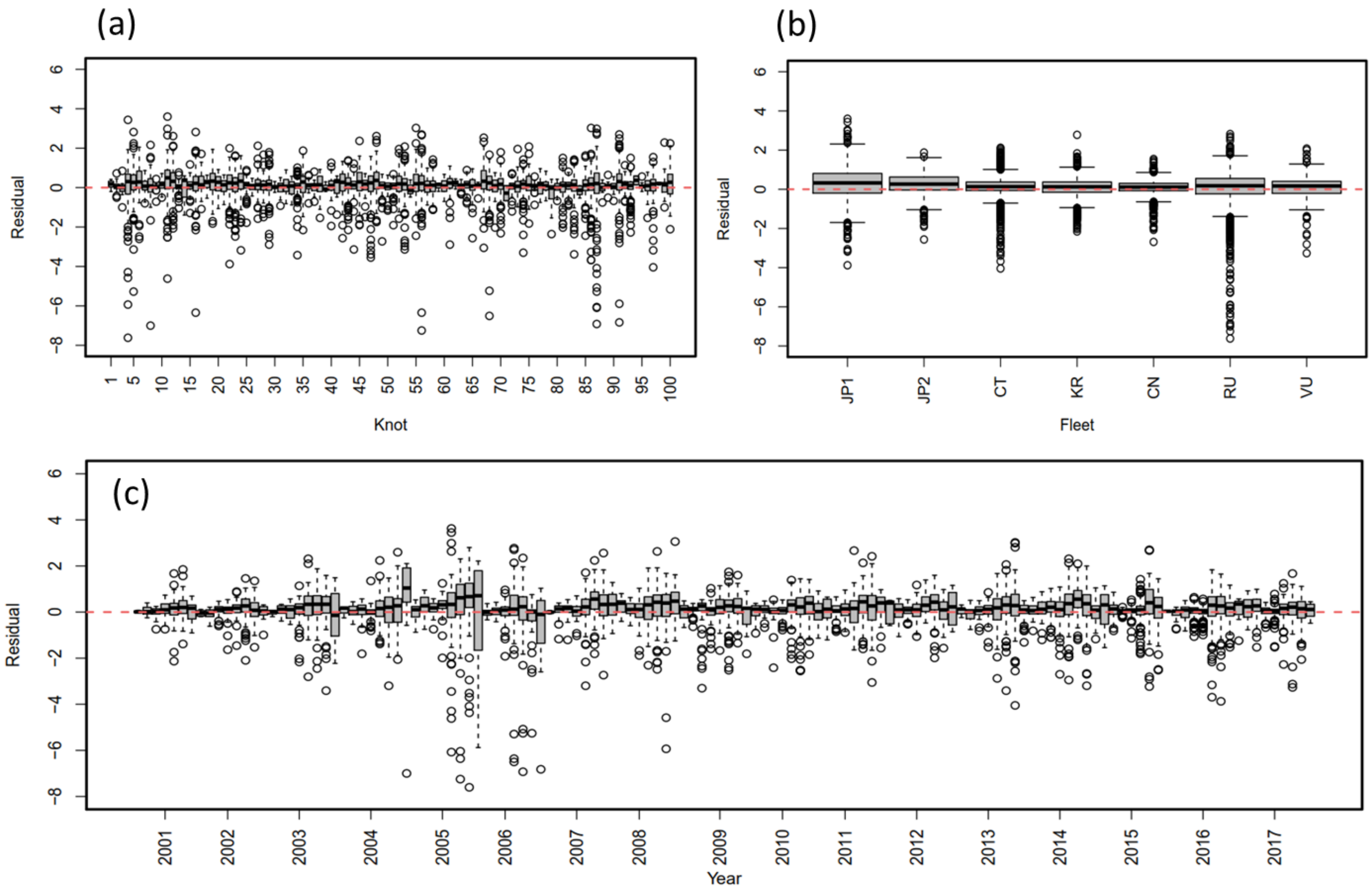

Figure 5. Boxplots of residuals by (a) cells, (b) fleets, and (c) year-month of the fitted VAST. JP1 is Japanese fisheries by vessels of <100; JP2 is Japanese fisheries by vessels of >= 100; CT is Chinese Taipei; KR is Korea; CN is China; RU is Russia, and VU is Vanuatu.

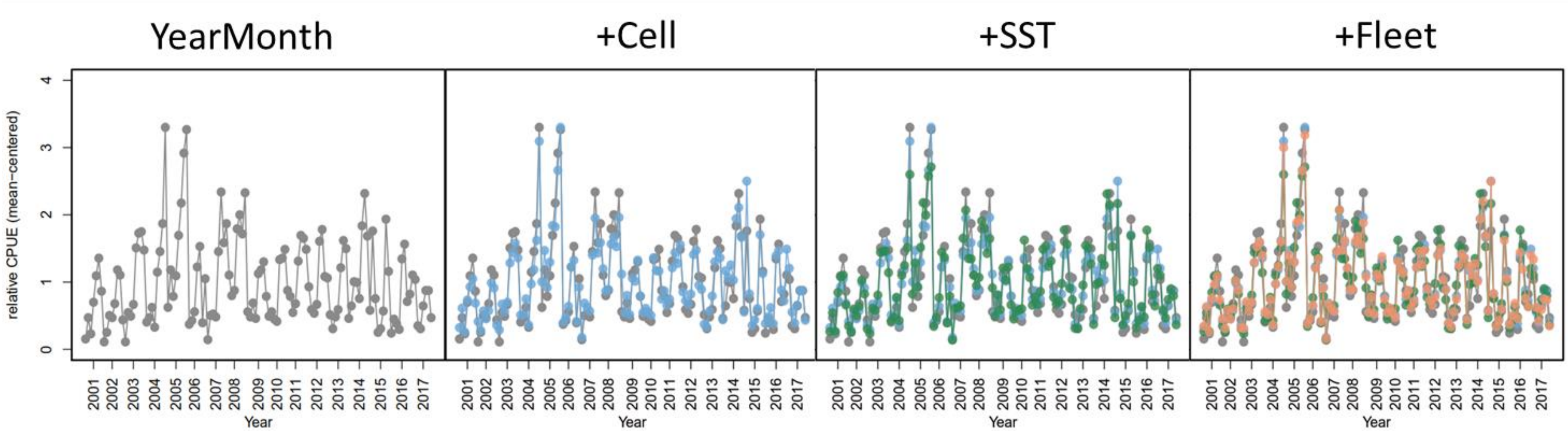

Figure 6. Step plots showing the effects of removing individual factors from the GLM with respect to the estimated CPUE indices.

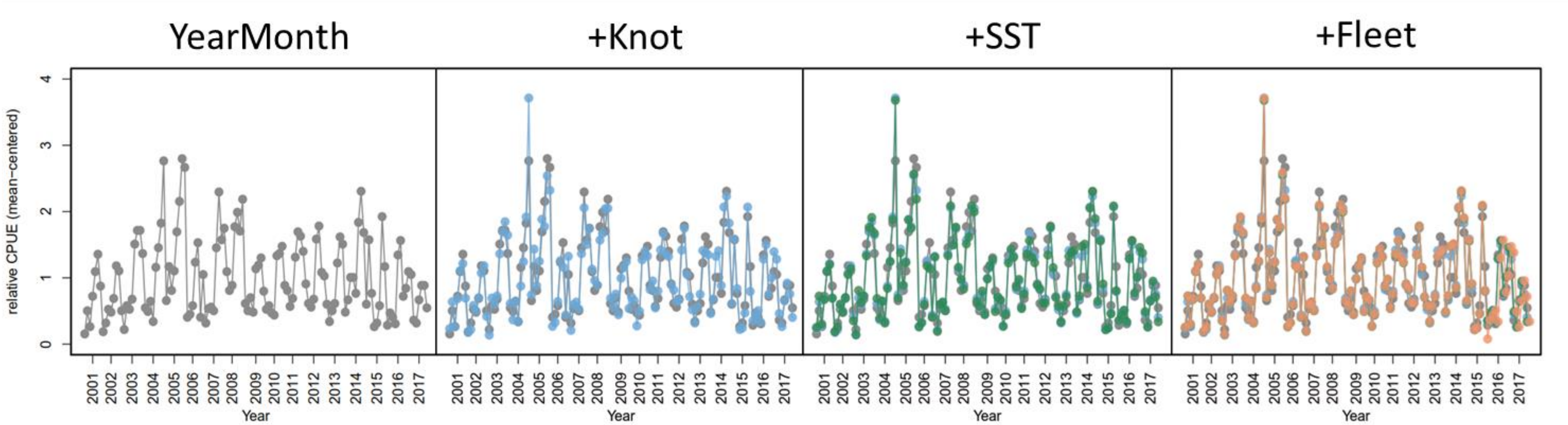

Figure 7. Step plots showing the effects of removing individual factors from the VAST with respect to the estimated CPUE indices.

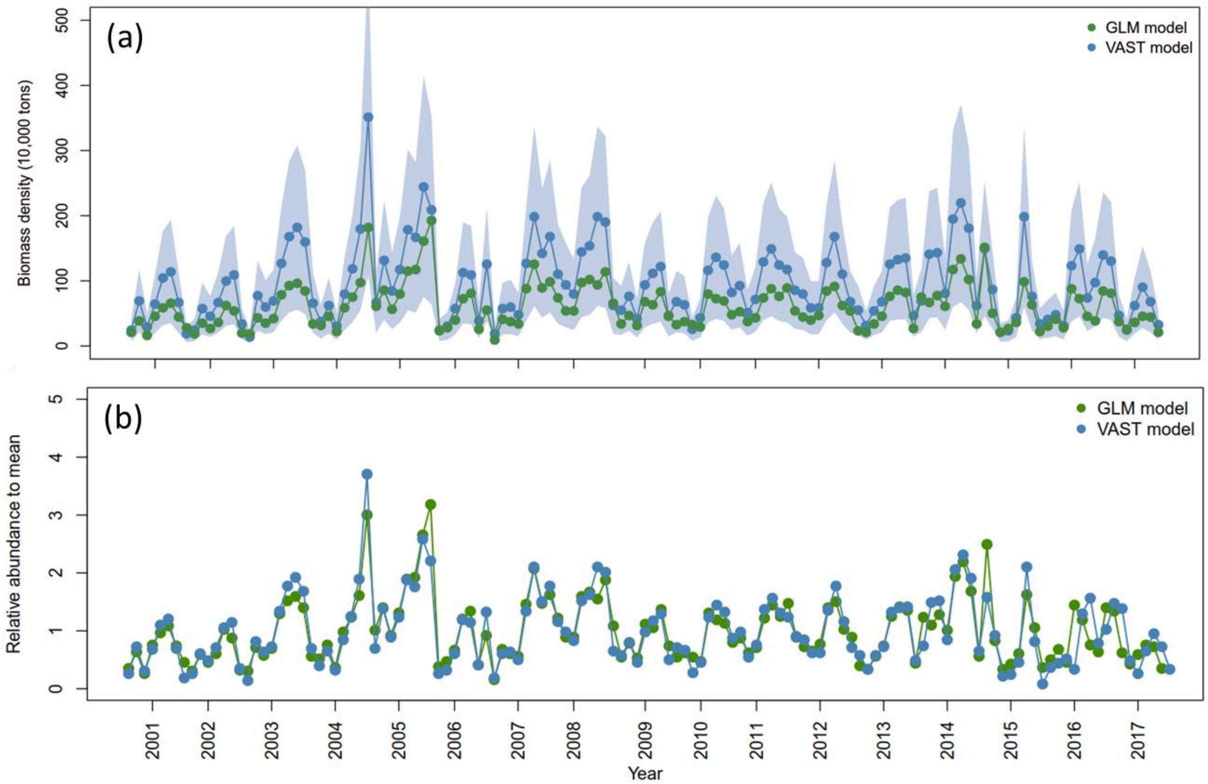

Figure 8. Time-series of year-month (a) absolute, and (b) relative (relative to mean) standardized indices from the GLM (green points) and VAST (blue points) for the Pacific saury in Northwest Pacific Ocean from 2001 to 2017. The blue polygon denote the 95% confidence intervals by VAST.