Evaluation of the influence of spatial treatments on catch-per-unit-effort standardization: A fishery application and simulation study of Pacific saury in the Northwestern Pacific Ocean

Jhen Hsu[1], Yi-Jay Chang[1,2*], Nicholas D. Ducharme-Barth[3]

[1] Institute of Oceanography, National Taiwan University, Taipei 10617, Taiwan

[2] Institute of Fisheries Science, National Taiwan University, Taipei 10617, Taiwan

[3] NOAA National Marine Fisheries Service, Pacific Islands Fisheries Science Center, 1845 Wasp Boulevard, Building 176, Honolulu, Hawaii, USA 96818

* Corresponding author: tel: +886-2-3366-1392

e-mail: yjchang@ntu.edu.tw

Highlights

- Various spatial treatments to standardize CPUE data were evaluated within a Generalized Linear Mixed Model (GLMM) framework.
- Some spatial stratification approaches may yield biased estimates of abundance indices under non-random spatial sampling.
- VAST yielded the best performance for index estimation as indicated by the lowest model mean square error and bias.
- Influence analysis was expanded to VAST and used to explain differences between unstandardized and standardized indices.

Abstract

Fishery-dependent catch-per-unit-effort (CPUE) data often exhibit spatial heterogeneity over space and time, which means that the spatial treatment in statistical models used to standardize CPUE is critically important. We evaluated several spatial treatments to standardize CPUE data using Generalized Linear Mixed Models (GLMMs). Results include a real-world application and a simulation based on the Taiwanese stick-held dip net fishery for Pacific saury in the Northwestern Pacific Ocean. We compared the performance of three spatially stratified approaches in GLMMs, (i) Ad hoc; (ii) Binary (binary recursive area partitioning based on model selection criteria); and (iii) Spatial clustering (partitioning of grids into discrete strata based on the spatial proximity and average CPUE in each grid), to a spatio-temporal GLMM (VAST). An influence analysis was constructed to quantify discrepancies between unstandardized and standardized indices that assisted in identifying the annual influence of explanatory variables in GLMMs. We developed a simulation to corroborate the results from the case study and evaluated the four spatial treatments using data generated from two contrasted, random and preferential, sampling scenarios. Results from the real-world application indicated that VAST was statistically superior to the other approaches, based on conditional deviance explained, conditional Akaike Information Criterion, and five-fold cross-validations. The influence analysis indicated that the interaction of year and spatial effect or spatio-temporal variable had a major influence on the standardized CPUE. Both simulation scenarios showed that VAST performed the best, with the lowest model error (measured by root mean square error) and bias, for estimating relative abundance indices. Although the spatial clustering approach created a flexible shape for the area strata, the simulation results under preferential samplings showed that clustering with a stronger emphasis placed on average CPUE could lead to bias in estimated abundance indices. However, spatial clustering that balanced average CPUE with spatial proximity could be a reasonable alternative if it is not possible to apply a spatio-temporal approach. The importance of conducting influence analysis and the greater performance of a spatio-temporal approach are highlighted.

Keywords: CPUE standardization, simulation-testing, influence analysis, spatio-temporal modelling approach, area stratification, Pacific saury

1. Introduction

Commercial catch-per-unit-effort (CPUE) data are often standardized for use as an index of relative abundance, particularly in fisheries where a regular survey has not been feasible (Hilborn and Walters, 1992; Maunder and Punt, 2004). To be useful as an index of relative abundance, raw CPUE data requires standardization to adjust for the effects of spatial and temporal dynamics in gear configuration, fishing power, and fishing behaviour (Campbell, 2004; Maunder and Punt, 2004; Maunder et al., 2006). The non-random nature of spatial distributions of fish density can also cause CPUE data from preferred fishing grounds to exert a disproportionate effect on the estimated fish abundance (i.e., preferential sampling; Rose et al., 1991; Conn et al., 2017; Ducharme-Barth et al., 2022). Therefore, adjusting for the spatial heterogeneity of CPUE data is essential for CPUE standardization. The confounding effect of spatial heterogeneity has commonly been addressed by the inclusion of categorical grids (e.g., 5 × 5 grids) or area strata within Generalized Linear Models (GLMs), Generalized Additive Models (GAMs), and Generalized Linear Mixed Models (GLMMs) used for CPUE standardizations (Maunder and Punt, 2004). In theory, the area strata are assumed to represent spatial heterogeneity in fish density, which may be treated as categorical variables in the standardization model to adjust for differences in CPUE associated with each stratum. Ideally, an appropriate area stratum is a region where fish density is homogeneous (Bishop, 2006), but in reality this is seldom the case. Many studies take ad hoc sets of grids as the area strata in standardizing CPUE analysis, or use area strata defined using either the spatial distribution of nominal CPUE and fishing effort or oceanographic conditions (e.g., Quinn et al., 1982; Nakano, 1998). However, it is difficult to construct appropriate area strata objectively using an ad hoc approach, and spatial misspecification in the CPUE standardization process could potentially result in a biased index of relative abundance (Bishop, 2006).

Several spatially stratified approaches have been developed to define objective criteria for the area stratifications in standardizing fishery CPUE. For example, a binary recursive partitioning approach was developed to automatically stratify the study area based on information criteria (e.g., Akaike or Bayesian Information Criteria, AIC or BIC)

(Ichinokawa and Brodziak, 2010). This approach created area stratifications more effectively than the area strata determined in an ad hoc manner and achieved better GLM fits to the CPUE data of North Pacific swordfish (*Xiphias gladius*). However, the binary recursive partitioning approach is limited to generate rectangular area strata due to its binary partitioning nature. In the real world, the distribution of fish density is unlikely to be structured as rectangular shapes. Ono et al. (2015) proposed an alternative area stratification approach known as spatial clustering to improve upon the rectangular area strata shape seen in Ichinokawa and Brodziak (2010). The spatial clustering approach applies a *k*-medoids algorithm to cluster grids of CPUE data according to the similarity of the spatial proximity and average CPUE of grids. This approach creates a flexible shape rather than a rectangular shape for the area strata which may better match the population structure inferred from the CPUE data. Ono et al. (2015) also suggested, based on simulation experiments, that the spatial clustering approach could reduce bias in abundance indices compared to the ad hoc approach, but this would not be expected to occur if the spatial distribution of fishing grounds had shifted over time.

The spatial effect in conventional GLMs for CPUE standardization assumes that the adjacent strata/grids are independent of each other. However, the variation of fish abundance and availability is often continuous and correlated with biotic and abiotic environmental factors over space (i.e., spatially structured). Therefore, it seems appropriate to incorporate spatial autocorrelation into a standardization model as a continuous covariate to more reasonably reflect the spatial heterogeneity of fish distributions (Thorson et al., 2015; Thorson and Barnett, 2017). Recent years have seen the emergence of spatio-temporal models (e.g., Vector-Autoregressive Spatio-Temporal, VAST; Thorson, 2019) for standardizing CPUE data (e.g., Xu et al., 2019; Maunder et al., 2020; Ducharme-Barth et al., 2022), because this approach could provide a more sophisticated treatment of spatial variation by accounting for not only the long-term spatial autocorrelation (i.e., spatial autocorrelation that is constant over time) but also the spatio-temporal autocorrelation (i.e., spatial autocorrelation that is specific to each year of the study period) in the CPUE standardization. Specifically, a spatio-temporal approach allows for the spatial and spatio-temporal effects to be treated as continuous Gaussian

112 Markov random fields (GMRFs), which may yield more precise, biologically reasonable,
113 and interpretable estimates of abundance than common area strata factors in GLMs
114 (Shelton et al., 2014; Thorson et al., 2015).

115 Multiple comparative studies (Grüss et al. 2019; Zhou et al. 2019) have shown the
116 benefits of using spatio-temporal models for standardizing CPUE data relative to other
117 regression models. Grüss et al. (2019) used a simulation experiment to show that VAST
118 minimized residual variance over space without defining the area strata in advance and
119 usually had the lowest error and bias relative to the eight other linear models considered.
120 Zhou et al. (2019) showed similar results when comparing a spatio-temporal approach to
121 GLMs and GAMs. However, the Grüss et al. (2019) and Zhou et al. (2019) studies only
122 considered a fixed ad hoc area stratification (10 areas and 7 areas, respectively) *a priori*
123 to CPUE standardization. Consequently, it was not possible to evaluate whether their area
124 stratification approaches were appropriate and sufficient to standardize CPUE. To our
125 knowledge, there has not been a study which explicitly compares the performance of
126 different spatial treatments (Ad hoc, Binary, Spatial clustering, and spatio-temporal
127 approaches) where a consistent model structure with the same sets of covariates has been
128 applied.

129 Most CPUE standardization studies have concentrated on removing the effects of
130 predictors to obtain an unbiased index of abundance. Few studies have focused on
131 understanding the differences between standardized and unstandardized CPUE indices
132 (Holdsworth et al., 2017; Hoyle et al., 2019; Feenstra et al., 2019). Bentley et al. (2012)
133 suggested that it is necessary when conducting an explanatory analysis (e.g., R package
134 *influ*; https://github.com/trophia/influ) to explore how the CPUE standardization model
135 removes confounding effects by including each explanatory variable in models rather
136 than simply accepting the relative abundance index arising from a model. Furthermore,
137 although considering the interaction effect of year and area in GLMs is a common way to
138 standardize spatiotemporal patterns in CPUE data, the influence analysis of the
139 interaction effect has rarely been examined.

140 Given the identified gaps in the literature, our analytical objective was to evaluate
141 the effects of spatial treatments on CPUE standardization, using a consistent model

structure and with the same sets of covariates within a GLMM framework. This objective was achieved by using the CPUE data of a commercially important migratory species, Pacific saury (*Cololabis saira*), as an example. The Pacific saury fishery in the Northwestern Pacific Ocean exploited by the Taiwanese stick-held dip net vessels provides an example of spatial heterogeneity of CPUE data, with more than 98,000 fishery operations recorded during 1997 – 2019 over a large geographical range (Chang et al., 2019). We applied a GLMM framework, with a year effect, spatial, and spatio-temporal variation terms, a vessel effect, and the effect of oceanographic conditions (i.e., sea surface temperature). We then compared the statistical performances (e.g., conditional deviance explained, conditional AIC, and five-fold cross-validations) of the four alternative sets of spatial treatments (three area stratification approaches and the spatio-temporal approach). We also used influence plots (Bentley et al., 2012) to clarify how a GLMM can remove confounding effects of explanatory variables. Finally, we conducted a simulation study using two contrasted spatial sampling scenarios to evaluate whether various spatial treatments could result in model misspecification in CPUE standardization. Our study was developed and illustrated in the context of Pacific saury. However, it should be broadly applicable to other fisheries for which similar data are available and could provide useful guidance for the appropriate treatment of the spatial effect in CPUE standardization analyses.

2. Materials and methods

2.1 Pacific saury fishery dataset and data filtering

Pacific saury fishery logbook data from the Taiwanese stick-held dip net vessels operating in the Northwestern Pacific Ocean (mainly 35 °N – 49 °N and 140 °E – 173 °E) during the main fishing season (August to November) from 1997 to 2019 were used in this study. Daily logbook data included the catch (in metric tons), the vessel identification (vessel ID), the amount of effort (number of hauls), and the location of sets by latitude-longitude at a resolution of 0.25°. These logbook records were obtained from the Overseas Fisheries Development Council of Taiwan. This dataset does not include any

zero catches because Pacific saury are targeted by vessels fishing at night using lamps to attract schools of fish. Data filtering was applied before the standardization to remove incomplete and insufficient data, such as sets with no information about the date, vessel ID, and locations. The final dataset included 98,738 fishing operations, after removing 2% of the sets from the original dataset.

2.2 Spatially stratified GLMMs

The application of GLMMs has been one of the most frequently employed modelling approaches for CPUE standardization (Maunder and Punt, 2004). GLMMs (Pinheiro and Bates, 2000) extend the GLM method by allowing some of the explanatory variables in the linear predictor to be treated as random variables. GLMMs have proven particularly useful for dealing with the year × spatial interaction in the CPUE standardization process (Miyabe and Takeuchi, 2003; Forrestal et al., 2019; Grüss et al., 2019), and were therefore considered as the most suitable framework for this study. Various studies have indicated that the spatial and temporal distribution of Pacific saury could be affected by changes in sea surface temperatures (SST; Chang et al., 2019; Hsu et al., 2021). Specifically, a fishery-independent survey for Pacific saury showed that it prefers a habitat with SST between 7 and 15°C (Hashimoto et al., 2020). Therefore, we used SST at fishing and its squared value (SST^2) as continuous variables in the GLMM standardization models, which allowed for the assumption that observed CPUE had a peak at an intermediate level of SST (Thorson and Barnett, 2017; Hashimoto et al., 2019). Given that fishermen could attract fish schools with fishing lamps at nighttime, the use of light (e.g., number or power of lights) could vary among fishing vessels and affect the catchability, hence it should be included in the CPUE standardization. However, such information is not available in our logbook data for the stick-held dip net fishery. Vessel ID was considered as a proxy for fishing power and was used as a covariate in the CPUE standardization procedure (Punt et al., 2000; Glazer and Butterworth, 2002; Battaile and Quinn, 2004).

199 Two metrics of fishing effort data (i.e., haul numbers vs. fishing days) could be
200 extracted to define the CPUE from the Pacific saury logbook dataset, although they were
201 not consistently reported throughout the study duration. Haul data were first recorded in
202 2003 and were missing from 48% of the logbook records from 2003 – 2006. Using effort
203 data defined as the number of fishing days allowed for a longer period of analysis (1997 –
204 2019), but this definition could potentially be hyperstable (Hilborn and Walters, 1992) if
205 an increase in daily hauls was used to maintain a daily catch-rate when abundance was
206 low. As a sensitivity, we developed standardized indices using both effort definitions
207 (**Fig. S1**). Because the trends were similar, CPUE with the fishing day effort definition
208 (metric tons/day) was used throughout the analysis.

209 Positive CPUE data were assumed to follow either a lognormal or gamma
210 distribution (Ortiz and Arocha, 2004); the lognormal distribution was assumed in the
211 current study for positive CPUE data following Ichinokawaa and Brodziak (2010) and
212 Winker et al. (2013). An exploratory analysis did not indicate that the current analysis
213 was sensitive to the choice of error distribution as the standardized CPUE indices and the
214 deviance explained (e.g., conditional R^2) were similar for the two error distributions
215 within the GLMM considered. The residual frequency distributions of the two error
216 assumptions derived from the four GLMMs indicated no violation of the assumed
217 statistical distributions. Accordingly, we normalized the CPUE response by the natural
218 logarithm transformation, $log(CPUE)$, which is a common procedure in CPUE
219 standardization (Winker et al., 2013; Grüss et al., 2019). To focus on comparing the
220 influences of various spatial treatments on the CPUE standardization, we only included
221 five explanatory variables in the GLMMs, and we did not conduct any model selection
222 procedure. However, the statistical significance of each variable was examined.
223 Explanatory variables considered in the GLMMs included *Year*, *Area* (**see section 2.3**),
224 *Vessel*, quadratic water temperature effects (*SST* and SST^2), and the interaction of year
225 and spatial effects ($Year \times Area$; to take into account vessel targeting behavior or fish
226 distribution shift). *SST* and SST^2 were treated as continuous variables while the remaining
227 covariates were treated as factors. We fit GLMMs to standardize CPUE for Pacific saury
228 as described in the following equation (Ono et al., 2015; Thorson et al., 2015):

$$\log(p_i) = \beta_{Year(i)} + \beta_{Area(i)} + \beta_{Year(i)\times Area(i)} + \beta_{Vessel(i)} + \beta_{SST(i)} \times SST_{(i)} + \beta_{SST^2(i)} \times SST^2_{(i)} \quad (1)$$

where p_i is the predictor for observations i (metric tons/day), and β is the estimated coefficient of its subscript (e.g., *Year*, *Area*, etc.). We treated the year effect, area effect, and quadratic water temperature effects as fixed effects. The vessel and the year × spatial interaction were treated as random effects. Model fitting was carried out with the "lmer" functions provided in the R statistical platform (Bates et al., 2015; R Core Team, 2021).

2.3 Spatial treatments in the spatially stratified GLMMs

Three area stratification approaches were used in the GLMMs (i.e., spatially stratified GLMMs) for standardizing Pacific saury CPUE data. These alternative area stratifications are described below:

(i) Ad hoc approach

The definition of the four area strata was modified based on Huang et al. (2007; 2020) which grouped the $1^o \times 1^o$ grids of Pacific saury CPUE data based on the bathymetric (depth) contours derived from the Centenary Edition of the GEBCO (General Bathymetric Chart of the Oceans) Digital Atlas (IOC-IHO-BODC, 2003). This approach is currently used for the CPUE standardization of Pacific saury by the North Pacific Fisheries Commission (Huang et al., 2021; Hashimoto et al., 2021).

(ii) Binary recursive partitioning approach

The binary recursive partitioning approach (i.e., Binary approach) developed by Ichinokawa and Brodziak (2010) is an algorithm that sequentially and recursively divides the whole studied area into several strata. This binary approach was applied to partition Pacific saury CPUE data using the following three steps. First, the algorithm divided the whole study domain into all possible pairs of strata, assuming a fixed spatial resolution (0.25^o) defined by a set of regularly spaced dividing lines. Second, a GLMM (see **section 2.2**) was applied to fit the CPUE data under each of the possible stratifications. Third, the information criterion (i.e., AIC; Burnham and Anderson, 2002) was used to determine the

stratification that produced the best fit over the set of possible stratifications. This three-step procedure was repeated recursively until the AIC value could not be improved by an increase in the number of areas. The maximum area strata number was fixed to six to have a comparable set of strata among various area stratification approaches.

(iii) Spatial clustering approach

The spatial clustering approach applied a *k*-medoids algorithm to partition the $0.25^{o} \times 0.25^{o}$ grids of Pacific saury CPUE into discrete area strata based on the spatial proximity and average CPUE in each grid (Ono et al., 2015). In this approach, the first step entailed calculating the Euclidean distance and the difference in mean values of CPUE observations between any two grids. Next, a *k*-medoids cluster analysis (Kaufman and Rousseeuw, 1990) was run where medoids are objects within a cluster for which the average dissimilarity to the remaining objects in the cluster is minimal. Furthermore, the dissimilarity matrix was multiplied by a weighting factor w to emphasize (or de-emphasize) the magnitude of spatial proximity with respect to the difference in underlying fish population density (Ono et al., 2015). This process makes the shape of the resulting stratified area irregular to match the population structure as inferred from the CPUE data. Following the approach of Ono et al. (2015), we set two weighting factors, w =1 and w = 0.1, to illustrate two shapes for the stratified area, namely "Spatial 1" and "Spatial 0.1". The "Spatial 1" approach assigns equal weight to spatial proximity and average fish density. The "Spatial 0.1" approach sets less weight to spatial proximity, representing the difference of one standardized unit in either spatial coordinate with 100 times less influence than one unit change in population density. We applied the *pam*() function in the *cluster* package (Maechler et al., 2014) provided in the R programming environment for the *k*-mediod cluster analysis. A visual inspection was used to identify the inflection point in the change in AIC values as the number of clusters increased. The number of clusters (choice between 2 and 6 strata) corresponding to the inflection point was selected as the optimal number of area strata.

2.4 Spatio-temporal GLMM

285 The spatio-temporal modelling approach used herein was adapted from the R
286 package *VAST* (version 3.2.2) (https://github.com/James-Thorson-NOAA/VAST)
287 developed by Thorson et al. (2015). By default, VAST implements a delta-generalized
288 linear mixed modelling framework, where the probability distribution for the catch is
289 decomposed into two components representing the probability of encounter and the
290 expected catch rate, given that catch occurs (Thorson, 2019). We only included the
291 positive catch rate (i.e., observed CPUE) component because the Pacific saury dataset did
292 not contain zero CPUE data. This model structure is possible in *VAST* by specifying the
293 user-controlled vector for observation models as ObsModel = 1 (the distribution for
294 positive catch rates is lognormal) and 3 (the encounter probability equals 1 for any year)
295 in VAST and by turning off all model structure associated with the probability of
296 encounter component (e.g., FieldConfig = c("Omega1"= 0, "Epsilon1"= 0, "Omega2"=
297 1, "Epsilon2"= 1) and OverdispersionConfig = c("Eta1"= 0, "Eta2"= 1)). The positive
298 catch rate was approximated using a lognormal GLMM with a log-link function and
299 linear predictors, which included GMRFs to model the spatial and spatio-temporal
300 effects. The assumption of using a log-transformed positive catch rate in VAST has been
301 commonly used (e.g., Xu et al., 2019; Grüss et al., 2019; Sculley and Brodziak, 2020).
302 Other positive continuous distributions such as the Gamma distribution could also be
303 considered.

304 VAST requires the previous definition of spatial knots *s,* which are the points
305 where the correlations for spatial and spatio-temporal effects are estimated. We specified
306 100 spatial knots to approximate the spatial and spatio-temporal auto-correlated
307 variations. This is the default configuration in VAST, which entails spatial allocation of
308 the knots with a density proportional to the sampling intensity by applying a *k*-means
309 algorithm (e.g., Xu et al., 2019; Grüss et al., 2019; Sculley and Brodziak, 2020).
310 However, it has previously been suggested that a uniform allocation of knots across a
311 pre-defined spatial domain was preferable to a proportional allocation based on fishing
312 intensity using a *k*-means analysis (Ducharme-Barth et al., 2022). As a sensitivity, we
313 developed two standardized indices using both knot configurations (**Fig. S2**). A similar
314 trend could be observed between both knot configurations, but the coefficients of

variation (CVs) of the standardized indices calculated from the uniform knot distribution were greater than those from the proportional knot distribution model (**Fig. S3**). Given the similarity of standardized indices, we used the proportional knot distribution for the comparison to other spatial treatments in this study. We also confirmed using an exploratory analysis that our results are qualitatively similar when using different numbers of spatial knots (100, 150, and 200 knots). The logarithm prediction of the Pacific saury CPUE is described below:

$$\log(p_i) = \beta(t_i) + \omega(s_i) + \varepsilon(s_i, t_i) + \delta(v_i) + \sum_{j=1}^{n_j} \gamma(j) X(s_i, t_i, j) \quad (2)$$

where $p(i)$ is the predictor for observation i (metric tons/day), $\beta(t_i)$ is the intercept for year t_i as a fixed effect and independent among years, $\omega(s_i)$ is the time-invariant spatial variation at location s_i (i.e., each of the 100 knots), and $\varepsilon(s_i,t_i)$ is the time-varying spatio-temporal variation for location s_i in year t_i, and $\delta(v_i)$ is the vessel effect as a mean-zero random effect with a standard deviation of one (Thorson, 2019). In addition, $\gamma(j)$ is the j^{th} catchability covariate $X(s_i, t_i, j)$ on location s_i in year t_i (i.e., the impact of SST on daily observed Pacific saury CPUE; Hashimoto et al., 2021), n is the number of catchability covariates. The marginal likelihood of fixed effect parameters is calculated with Template Model Builder using the Laplace approximation to integrate across random effect parameters (Kristensen et al., 2016), and fixed effect parameters are then estimated by maximizing the marginal likelihood within the R computing environment (R Core Team, 2021). Convergence was checked by ensuring that the absolute value of the final gradient of the log-likelihood function at the maximum likelihood estimate was less than 0.001 for all parameters and that the Hessian matrix of the likelihood function was positive definite.

2.5 Statistical performance

The conditional R^2 (Nakagawa and Schielzeth, 2013) and the conditional AIC (Greven and Kneib, 2010) were calculated to represent the performance of the five GLMMs using various spatial treatments (subsequently called the Ad hoc, Binary, Spatial

343 1, Spatial 0.1 GLMMs, and VAST) for Pacific saury. We also conducted repeated five-
344 fold cross-validations to compare the performance of each GLMM (Winker et al., 2013;
345 Shono, 2014). A stratified random sampling approach using years as strata was used in
346 the five-fold cross-validation procedure to maintain the basic structure of the dataset
347 because the year effect was the variable of interest. Each dataset was randomly resampled
348 without replacement under each of the year strata. The first 80% of the randomly drawn
349 data were used as "training data" to fit the model, and the estimated model coefficients
350 were then applied to predict the CPUE from the covariate information in the remaining
351 20% of the "testing data". We utilized the averaged values of the explained deviances
352 (conditional R^2) and the Pearson's correlation coefficients (derived from the observed and
353 the predicted CPUE) estimated from the testing data to compare statistical performances
354 of the five GLMMs. The five-fold cross-validations were repeated 10 times and were
355 used with the averaged values of the conditional R^2 and the Pearson's correlation
356 coefficients derived from the 10 replicates to evaluate the performances of the five
357 GLMMs.

358

359 *2.6 Standardized abundance indices*

360 Standardized abundance indices derived from spatially stratified GLMMs (Ad
361 hoc, Binary, Spatial 1, and Spatial 0.1) were calculated in two steps. First, CPUE was
362 predicted with fitted GLMMs for all combinations of years and areas. We then used the
363 mean of the assumed normal distribution of the year × spatial random effect (i.e., zero) to
364 impute the missing year × spatial values to derive the standardized CPUE for each year
365 and area (Campbell, 2015):

366 $$\hat{CPUE}_{Area,Year} = \exp\left(\beta_{Year} + \beta_{Area} + \beta_{Year\times Area}\right) \quad (3)$$

367 A bias-corrected estimate for the standardized CPUE in each year and area was
368 calculated as $\exp\left(\hat{CPUE}_{Area,Year} + \hat{\sigma}^2/2\right)$, where $\hat{\sigma}$ is the estimated model standard
369 deviation (residual standard error) (Maunder and Punt, 2004). In the second step,

we calculated the standardized abundance indices, with and without area weighting, and compared their difference(s) for each spatially stratified GLMM. The index without the area weighting was computed by using an equal weight for each area (i.e., arithmetic mean). The area-weighted index ($\hat{CPUE}_{Year}$) was obtained by summing over all stratified areas within a year ($\hat{CPUE}_{Area,Year}$) following Campbell (2015):

$$\hat{CPUE}_{Year} = \sum_{Area=1}^{n} SA_{Area} \times \hat{CPUE}_{Area,Year} \quad (4)$$

where n is the number of area strata (which differ among various area stratification approaches); SA_{Area} is the proportion of the surface area for a given $Area$ to the whole studied domain.

Unlike the spatially stratified approaches, the explicit spatial correlation modelled in VAST was used to predict the Pacific saury density across all spatial cells in the study area. The standardized abundance index in year t across the studied area was described as follows:

$$\hat{CPUE}(t) = \sum_{s=1}^{n} SA(s) \times \exp\left(\beta(t) + \omega(s) + \varepsilon(s,t)\right) \quad (5)$$

where s is the number of knots s; $\beta(t)$ is the year effect in t year; $\omega(s)$ is the spatial effect at s knot; $\varepsilon(s,t)$ is the spatio-temporal effect at s knot in t year, and $SA(s)$ is the surface area of the triangulated mesh associated with knot s. Annual relative abundance indices without area weighting were also calculated for VAST. The bias-correction estimator within VAST (Thorson and Kristensen, 2016) was employed to account for retransformation bias when predicting the abundance.

Uncertainties about the annual relative abundance indices deriving from the four spatially stratified GLMMs were estimated based on the method used by Campbell (2015; see section 6 for details). For VAST, the uncertainty of the predicted index was computed using a generalization of the delta method (Thorson et al., 2015; Thorson and Barnett, 2017). Each of the standardized abundance indices was normalized to its mean

for the studied period (1997 – 2019), respectively, in order to facilitate comparison (Winker et al., 2013; Kai et al., 2017).

2.7 Quantifying the influence of explanatory variables

In order to quantify the influence that each explanatory variable had for explaining the difference between standardized and unstandardized (arithmetic mean of observed CPUE observations) CPUE in each year (i.e., year effect), we used the influence analysis described by Bentley et al. (2012). The annual influence of an explanatory variable can be quantified as the combination of the model coefficients and CPUE data distributional changes over the years based on the Coefficient Distribution-Influence plot (CDI) (Bentley et al., 2012).

To calculate such a measure of influence, the normalized coefficient associated with an explanatory variable (ρ) was calculated as:

$$\rho = \sum_{i=1}^{n} \alpha_i \Big/ n \tag{6}$$

where α_i is the estimated coefficient of a variable (factor) corresponding to observation i, and n is the number of CPUE observations. Then, the annual influence value for a variable in year y (AI_y) was the exponentiation of the mean difference between the coefficients corresponding to all observations in that year and the normalized coefficient for a multiplicative GLMM (since the log-link function was used, i.e., $AI_y = \exp(\delta_y)$):

$$\delta_y = \frac{\sum_{i=1}^{n} \alpha_i - \rho}{n} \tag{7}$$

where n is the number of observations in year y. If the AI_y value for a variable is larger than one, it implies that adding this variable to the model will make the standardized CPUE less than the unstandardized CPUE in year y. Conversely, an AI_y value less than one indicates that adding the variable will make the standardized CPUE higher than the

unstandardized CPUE in year y. If AI_y equals one, the variable has no impact on the difference between unstandardized and standardized CPUE in year y.

The overall influence metric of a variable across all years ($\overline{AI}$) was calculated as:

$$\overline{AI} = \exp\left(\frac{\sum_{y=1}^{m}\delta_y}{m}\right) - 1 \quad (8)$$

where m is the number of years.

The CDI plot first presented by Bentley et al. (2012) combines the information of the normalized coefficient values, the distributional changes of the data record, and the resulting AI values into a single plot (see **Fig. 4** as an example). Specifically, the top panel of the CDI plot provides normalized coefficients for a variable and their standard errors, bubbles in the bottom left panel indicate the annual distribution of data records from each level of the variable and the AI values for a variable over years were shown in the bottom right panel. When extending the CDI plot to VAST, we first grouped the 100 spatial knots into 20 "grouped" knots according to their coefficients from low to high in order to simplify the visualization of the data distribution and coefficients for the spatial random effect. A CDI plot was then created for the grouped knots. Since knots with similar coefficients correspond to spatial areas with similar levels of predicted abundance, this allowed us to identify changes in the index corresponding to shifts in sampling between areas of high or low estimated abundance. Additionally, through an exploratory analysis we noted that spatial knots in the same group were generally grouped together spatially.

Although the year × spatial interaction was treated as a random effect (i.e., normally distributed with a mean value of zero), the influence of the interaction could still be examined because the coefficients may have a pattern varied from their overall average. For a better understanding of the information contained in the coefficients of year × spatial interaction across years, we modified the CDI plot of Bentley et al. (2012) by examining the normalized coefficients of the year × spatial interaction (see Fig. 6 as

445 an example, solid circles with color in the top panel) by area. Then, the AI_y value for the
446 year × spatial interaction in year y (shown in the bottom right panel) could be calculated
447 similarly as previously described for a factor variable. The CDI plot for the spatio-
448 temporal effect in VAST also utilized 20 grouped knots.

449 In this study, influence analysis was only used to explain the difference between
450 the standardized index without area weighting and the unstandardized index. The reason
451 being that the area-weighting process was independent of the estimated coefficients and
452 distributional changes of CPUE observations over the years.

453

454 *2.8 Evaluation using simulated data*

455 Simulation testing is a powerful tool to evaluate the performance of CPUE
456 standardization methods (Ono et al., 2015; Grüss et al., 2019; Ducharme-Barth et al.,
457 2022). The advantage of this approach is that the simulated abundance trends are known
458 so that the standardization method can be tested in terms of how well it predicts "true"
459 abundance trends. Abundance indices derived from fishery-dependent CPUE data are
460 known to have the potential for bias, due to the usual non-random nature of fisheries
461 spatial distributions ("preferential sampling"; Clark and Mangel, 1979; Rose and Leggett,
462 1991; Rose and Kulka, 1999; Swain and Sinclair, 1994). Therefore, we developed a
463 simulation framework with two spatial sampling scenarios (i.e., random and preferential
464 spatial sampling patterns) to evaluate which spatial treatments could effectively address
465 the spatial heterogeneity in the underlying data (i.e., leading to unbiased results compared
466 to the true value). We then simulated Pacific saury data similar to that available in the
467 Northwestern Pacific Ocean and then fit both the spatially stratified GLMMs (i.e., Ad
468 hoc, Binary, and Spatial clustering approaches) and VAST to the simulated CPUE data
469 under two sampling scenarios: (i) a random; and (ii) a preferential spatial sampling
470 pattern. We then extracted the estimate of total abundance and compared this with the
471 true value.

472

2.8.1 Sampling patterns

The spatial domain of the simulation covered the spatial extent of the fishing ground for Pacific saury from 140° E – 170° E longitude and from 35° N – 50° N latitude (Hsu et al., 2021). Following the approach by Thorson et al. (2015), true biomass was simulated by fitting a VAST model without effects of vessel and sea surface temperature to generate a base biomass distribution of Pacific saury $B(s,t)$ (metric tons/day) from 1997 to 2019, where s denotes a cell of 0.25° spatial resolution (7,023 cells in the spatial domain) and t denotes the yearly time step (illustrated in **Fig. S4**a). The biomass summed across all 0.25° spatial cells over the whole spatial domain for the year t was referred to as the "true" index (T_t). For each of the different spatial sampling patterns (i.e., random and preferential sampling; **Fig. S4**b - c), observation error η was incorporated to produce the "observed" data $\hat{d}(s,t)$ at each spatial cell s and year t:

$$\hat{d}(s,t) = B(s,t) \times e^{\eta} \tag{9}$$

$$\eta \sim Normal(0, 0.2) \tag{10}$$

In the random sampling pattern, each spatial cell s had an equal probability of being selected, regardless of the underlying Pacific saury density distribution. The preferential sampling pattern was based on the principle that fishers are more likely to fish in areas of higher fish density (Allen and McGlade, 1986; Hilborn and Walters, 1987). For both the random and preferential sampling scenarios, sampling of 300 spatial cells within a time step was done with replacement. Using an exploratory analysis, we confirmed that a total of 300 observations could ensure recovering the "true" index under the random sampling pattern. In contrast to the random sampling pattern, the probability of a spatial cell s being selected ($P_{pref,s}$) in any given year for the preferential sampling pattern was proportional to simulated fish density (Ducharme-Barth et al., 2022):

$$P_{pref,s} = \frac{(B_s)^{\varphi}}{\sum_{s=1}^{n} (B_s)^{\varphi}} \tag{11}$$

where the probability exponent φ controls the magnitude of preferential sampling. When $\varphi = 0$ all spatial cells have an equal probability of being sampled (e.g., random sampling). We set $\varphi = 8$ to impose a very strong degree of preferential sampling (extreme preferential sampling, Ducharme-Barth et al., 2022). Both fishing effort sampling patterns were simulated 40 times, resulting in a total of 80 datasets that were used to estimate abundance indices. It should be noted that the defined area strata in the Binary and Spatial clustering approaches and knot configurations in VAST would differ slightly among simulation runs since the simulated data varied with each iteration.

2.8.2 CPUE standardization and performance evaluation

Model performance in all simulations was evaluated relative to the "true" index. Prior to assessing model performance, each of the derived indices with area weighting (both estimated and true indices) was rescaled to a mean of 1 by dividing by its overall mean, respectively. Model performance was assessed based on three metrics: (1) the relative error metric in year y (RE_y); (2) the root mean squared error (RMSE) and (3) the bias metric described below.

The relative error in each year was calculated as (Ono et al., 2015):

$$RE_y = \frac{(\hat{I}_y - T_y)}{T_y} \tag{12}$$

where $\hat{I}_y$ and T_y are the estimated and true indices in year y.

The measurement of model error was calculated as below (Stow et al., 2009; Ducharme-Barth et al., 2022):

$$RMSE = \sqrt{\frac{\sum_{y=1}^{n} (\hat{I}_y - T_y)^2}{n}} \tag{13}$$

where n is the number of years. A higher RMSE represents a greater error in the estimated index of abundance (Stow et al., 2009; Ducharme-Barth et al., 2022).

The bias metric we considered was the coefficient β of the following the linear model (Thorson et al., 2015; Grüss et al., 2019; Ducharme-Barth et al., 2022):

$$\hat{I}_y = \alpha + \beta \times T_y + \varepsilon_y \quad (14)$$

$$\varepsilon_y \sim Normal(0, \sigma_\varepsilon^2) \quad (15)$$

where α is the intercept, and β is the slope parameter from a linear model between the true index T_y and the estimated index of abundance $\hat{I}_y$. A β of 1 is indicative that changes in the true index are reflected accurately by the estimated index (i.e., unbiased), while a β greater than 1 (lower than 1) indicates that $\hat{I}_y$ underestimates (overestimates) changes in the true index (Wilberg et al., 2010; Thorson et al., 2015). We also examined if the best model selected by conditional AIC and conditional R^2 in the simulation test could provide the least biased representation of the true index pattern.

3. Results

3.1 Area stratifications and knots configuration

The geographic boundary and variability of observed CPUE of the area strata obtained using the area stratification approaches are shown in **Fig. 1**. The studied area was separated into four strata by both the ad hoc method and the binary recursive algorithm (**Fig. 1**a and c). The inflection point in the change in AIC values was used to determine the optimal number of clusters (i.e., five area strata) in the Spatial 1 and Spatial 0.1 approaches, respectively (results not shown). The shape of the resulting area strata in Spatial 1 was less patchy than Spatial 0.1, due to the higher importance of spatial proximity with respect to the difference in underlying observed CPUE (**Fig. 1**e and g). For each area stratification approach, the mean observed CPUE among the area strata were significantly different (ANOVA test, $p < 0.001$) (**Fig. 1**b, d, f and h). For the VAST model, the 100 spatial knots were distributed in proportion to the density of observations which resulted in relatively fewer knots towards the periphery of the spatial domain.

548 Model convergence was confirmed by the fact that the Hessian matrix was positive
549 definite and the maximum gradient component was smaller than 0.001. Most of the
550 observed CPUE with higher values were located in the area where more spatial knots
551 were distributed (**Fig. 1**i). This was not unexpected given the choice to allocate spatial
552 knots in proportion to the underlying density of the fishery dependent samples, and
553 implies some level of preferential sampling is present in the Pacific saury data.

554

555 *3.2 Comparison of the statistical performance*

556 All explanatory variables considered in the GLMMs were statistically significant
557 (likelihood-ratio test, $p < 0.005$) (**Table S1**). The year × spatial random-effect
558 coefficients for each GLMM have been assessed to follow a normal distribution (
559 $p > 0.05$). In general, the VAST model had a higher conditional R^2 and a lower
560 conditional AIC value than the spatially stratified GLMMs (**Table 1**). In addition, by
561 using the testing data from the five-fold cross-validations, the results suggested that
562 VAST had the highest mean values of the deviance explained (mean conditional R^2 =
563 65%) and Pearson's correlation coefficient (mean $\rho = 0.54$) compared to the other
564 GLMMs (**Table 1**). Overall, VAST outperformed the other GLMMs, with the best fit to
565 the data, whereas the Ad hoc GLMM performed the worst. Besides the VAST model, the
566 two spatial clustering GLMMs appeared to perform better than the other spatially
567 stratified models based on the results of the goodness-of-fit and five-fold cross-
568 validations (**Table 1**).

569

570 *3.3 Trends in nominal and standardized indices*

571 Trends in nominal and standardized indices for Pacific saury without area-
572 weighting illustrated large fluctuations (**Fig. 2a**). Annual trends of nominal and
573 standardized CPUE were consistent with each other during 1997 – 2008, except for the
574 Binary GLMM in 2000. An increasing trend of nominal and standardized CPUE was
575 observed during 2009 – 2014, but the standardized CPUE exhibited a pattern of large

interannual variability. Both nominal and standardized CPUE showed a decreasing trend during 2015 – 2019. The CVs of annual relative abundance indices derived from the Binary GLMM (0.23 – 0.33) and VAST (0.10 – 0.21) showed a decreasing temporal trend while the CVs from the other GLMMs were relatively stable in the late time period (**Fig. 2b**).

The area-weighted abundance indices calculated from all GLMMs had a smaller variation than the abundance indices without the area-weighting approach (**Fig. S5**). We compared the derived abundance indices with and without the area-weighting approach from the spatially stratified GLMMs (**Fig. S6**). For the Ad hoc and Spatial 1 GLMMs, each area strata had a similar surface area (i.e., similar spatial weighting) which caused the derived weighted standardized abundance indices to be very similar to the un-weighted indices (**Fig. S6**a and c). However, the derived abundance indices from the Binary and Spatial 0.1 GLMMs had more weighting from Area III and Area IV, respectively. Therefore, the weighted abundance index diverged slightly from the un-weighted standardized abundance index (**Fig. S6**b and d).

3.4 Influence of explanatory variables on annual relative abundance indices

The AI plots of the GLMMs, which present the influence of each variable in relation to the differences between standardized and unstandardized CPUE are shown in **Fig. 3**. AI values for the spatial and year × spatial effects varied among the five GLMMs. In contrast, the AI values for the vessel and quadratic SST effects, respectively, had consistent trends among GLMMs. This result demonstrated the importance of the spatial treatment in the GLMMs (for example, see **Fig. 2a**). The substantial discrepancy between annual relative abundance indices compared to the unstandardized CPUE during 2009 – 2014 could be explained by the larger variation of the AI values for the year × spatial effect (0.81 – 1.45) and the increasing AI values of the vessel effect (0.94 – 1.10) over the time period. The influence of vessel effect showed an obvious increasing trend in recent years (**Fig. 3**). The higher AI values for the vessel effect (1.15 – 1.27) among GLMMs compared to the effects of spatial and year × spatial effect (spatio-temporal

605 effect) (0.80 – 1.24) during 2015 – 2019 caused the annual standardized CPUE to be less
606 than the unstandardized CPUE.

607 Overall influence for each variable indicated that the spatial, interaction of year ×
608 spatial, and vessel effects had greater overall influences on annual relative abundance
609 indices than the quadratic SST effect (overall influence of SST is the smallest) (**Table 2**).
610 Although the spatial and vessel effects generally had greater explanatory power
611 (conditional R^2) than the year × spatial effect in the spatially stratified GLMMs, its
612 overall influence was less than that of the year × spatial effect. This result suggested that
613 the variable with a higher explanatory power (i.e., spatial and vessel effects) may not
614 necessarily indicate that it is influential on the difference between standardized and
615 unstandardized CPUE.

616

617 *3.5 Exploration of the spatial treatments in influencing annual relative abundance*
618 *indices*

619 The composite CDI plots of the spatial effects for spatially stratified GLMMs are
620 shown in **Fig. 4** and the spatial random effect for VAST is shown in **Fig. 5**. The AI
621 values of spatial effects for the spatially stratified GLMMs were greater than one (vertical
622 line in the right bottom panel of **Fig. 4**) before 2000, except for the Spatial 0.1 GLMM
623 because a greater proportion of data was distributed in the area stratum (Area I for Ad
624 hoc; Area III for Binary; and Area I for Spatial 1, respectively) with coefficients (spatial
625 effect) larger than one (horizontal line in the top panel of **Fig. 4**). Conversely, a higher
626 proportion of data concentrated in the area stratum (Area I) with a coefficient lower than
627 one for the Spatial 0.1 GLMM resulted in the AI values of the spatial effect to be below
628 one during 1998 – 2000. The data were distributed uniformly across all area strata for all
629 spatially stratified GLMMs during 2006 – 2014, thus the AI values of spatial effect were
630 stable around one. The AI values of the spatial effect were consistently below one since
631 2015 for all spatially stratified GLMMs because a larger proportion of data had then
632 shifted to the area stratum with a smaller coefficient (Area IV for Ad hoc; Area II for
633 Binary; Area V for Spatial 1; and Area II and V for Spatial 0.1, respectively). For the
634 VAST model, the AI values of the spatial random effect were above one before 2004

because a greater proportion of data tended to distribute in the grouped knots (grouped knots 17 – 20) with coefficients larger than one (**Fig. 5**).

The composite CDI plot of the year × spatial effect for spatially stratified GLMMs is presented in **Fig. 6**. A substantially higher AI value for the year × spatial effect was observed in 2000 for the Binary GLMM and in 2006 for all spatially stratified GLMMs, respectively. This is due to a higher proportion of data that was distributed in Area III with a high coefficient (3.30) in 2000 for the Binary GLMM and in Areas III, II, IV, and V in 2006 for the Ad hoc (2.33), Binary (3.95), Spatial 1 (2.30), and Spatial 0.1 (1.63) GLMM, respectively. The CDI plot of the spatio-temporal effect for the VAST model is shown in **Fig. 7**. The higher AI values were observed in 2006 and 2007 for VAST because the coefficients were generally higher for those years. A smaller AI value was also observed in 2015 which resulted from a higher proportion of the data which had shifted to the grouped knots 1 – 3 with lower coefficients.

3.6 Results of simulation test

Relative errors (REs) from the simulation results showed that the VAST model could yield relative abundance indices closer to the "true" index (REs fluctuated around zero overtime without an apparent trend) for both sampling scenarios compared to other GLMMs (**Fig. 8**). VAST also had the smallest mean RMSE and the lowest bias (i.e., close to one) under either random or preferential sampling scenarios (**Fig. 9**). Ad hoc and Binary GLMMs generally had larger REs overtime (beyond or close to ± 20%) (**Fig. 8**a) and higher mean RMSE values (0.14 and 0.13) and provided the least accurate estimates of relative abundance index (mean bias = 1.02 and 0.97) than other approaches under the random sampling scenario (**Fig. 9**a). We noted that the Spatial 1 and 0.1 GLMMs generally performed equally as well as VAST under the random sampling scenario. However, the Spatial 0.1 GLMM had a time trend in REs with substantial variation (**Fig. 8**i), the largest model error (mean RMSE = 0.18), and the highest bias (overestimate; mean bias = 0.89) compared to other GLMMs under the preferential sampling scenario (**Fig. 9**b). In general, all GLMMs showed higher model errors (i.e., larger mean RMSE)

under the preferential sampling compared to the random sampling, however a pattern of larger mean bias was not observed (but SD of bias had increased) (**Fig. 9**).

The model selection criteria (i.e., conditional R^2 and conditional AIC) evaluated by the simulation results, the RMSE, and bias, consistently showed that VAST was the best model under both sampling scenarios (**Fig. S7**). However, under the preferential sampling scenario, Spatial 0.1 GLMM had a greater value for the conditional R^2 and a lower conditional AIC value than the Ad hoc and Binary GLMMs (**Fig. S7**b). It was also the most inaccurate and biased representation of the true index based on the simulation results of RE, RMSE, and bias metrics (**Fig. 8i** and **9b**). This suggested that model selection criteria did not produce a consistent ranking of models when compared to their performance in fitting the true index.

4. Discussion

We applied an influence analysis to understand the standardization effects of explanatory variables included in the five GLMMs with Pacific saury fishery data. We noted that an increasing influence trend of vessel effect was observed in recent years. This implied a temporal change in catchability in the Taiwanese stick-held dip net fishery. Catchability changes occur in most fisheries (Wilberg et al., 2010), which if unaccounted for in the CPUE standardization could lead to an error of overestimating the Pacific saury abundance index in recent years. This study also extended the original *influ* package to accommodate the interaction between year × spatial effects in spatially stratified GLMMs and to the spatio-temporal effect of spatio-temporal model (i.e., VAST). The year × spatial or spatio-temporal effect was identified as the most influential effect, with large coefficients and substantial interannual variation, resulting in large variations of AI values in the influence plot. It is worth noting that the annual influence values of year × spatial interaction in the spatially stratified GLMMs had substantial AI values in 2000 and 2006 (**Fig. 3**b). This could imply that the assumed spatial structure was mis-specified and resulted in estimated coefficients that were extremely large in certain years.

693 Our results also showed that the variable with the most explanatory power
694 according to deviance explained (i.e., vessel or area variable of Pacific saury example)
695 may not have had the greatest influence on the difference between unstandardized and
696 standardized CPUE. Overall, we conclude that the difference between standardized and
697 unstandardized CPUE in the Pacific saury fishery could primarily be explained by the
698 year × spatial or spatio-temporal effect which could be related to changes in fishing
699 location or shifts in fish distribution over time (Hashimoto et al., 2020; Hsu et al., 2021).
700 There was also evidence for shifts in the fleet toward more efficient

701 In the real-world application using Pacific saury fishery data, VAST had a
702 superior performance in terms of the conditional AIC values, deviance explained (in
703 conditional R^2), and five-fold cross-validations (mean conditional R^2 and Pearson's
704 correlation coefficient) than the other GLMMs. Furthermore, under both sampling
705 scenarios, the VAST model achieved a better performance in terms of lower model error
706 (the smallest RMSE) and bias (bias value is close to one) than the other spatially
707 stratified GLMMs. This is because VAST can efficiently define how fish density varies
708 continuously across space (Kristensen et al., 2014) and describe the patterns in density
709 distribution over time (unexplained variability; Thorson and Barnett, 2017; Hsu et al.,
710 2021; Han et al., 2022) using spatial and spatio-temporal random effects implemented
711 with a stochastic partial differential equation approach (Lindgren et al., 2011). We
712 concluded that a spatio-temporal model such as the one implemented using VAST can
713 more appropriately capture the spatial heterogeneity of CPUE data based on the
714 simulation. Additionally, VAST is more robust to deviation from random sampling than
715 the other GLMMs used in this study. A study for pelagic sharks by Kai (2019) also
716 indicated that the inclusion of spatio-temporal random effects in a spatio-temporal model
717 could minimize residual variability and the fitting of a GLMM with spatial random
718 effects was better than that of GLMs with fixed spatial effects. Furthermore, the VAST
719 model represented a more direct approach to a conventional CPUE analysis because it is
720 not necessary to conduct the area stratification in advance. The present study contributed
721 to a growing literature suggesting that using a spatio-temporal approach such as VAST is
722 more statistically efficient (i.e., have greater precision for a given amount of data; **Fig.**

723 **2**b) than analyzing data using spatially stratified models (Thorson et al., 2015; Kai, 2019;
724 Grüss et al., 2019; Zhou et al., 2019).

725 The ad hoc GLMM showed the poorest performance in terms of its conditional
726 R^2, conditional AIC, and five-fold cross-validations among the spatially stratified
727 GLMMs for the Pacific saury fishery data. For this reason, we argue that the currently
728 used four area strata, separated in an ad hoc manner according to bathymetric contours
729 may not accurately reflect the spatial heterogeneity of Pacific saury density. This
730 inference was supported by our simulation experiment which showed that, regardless of
731 the sampling patterns, the Ad hoc GLMM had greater model errors of index estimation
732 than the Binary and Spatial 1 approaches. Noting that most CPUE standardization
733 analyses often use an ad hoc approach (e.g., Nakano, 1998; Haltuch et al., 2013), the
734 most immediate implication of this study is the need for caution when an ad hoc method
735 is used. However, we noted that other ad hoc spatial structures exist (e.g., 5 × 5 grid)
736 beyond the one explored in this analysis and their use for CPUE standardization should
737 be evaluated in the context of this study in the future.

738 The binary recursive algorithm based on AIC can be considered an objective area
739 stratification approach. Our simulation results showed that a binary approach generally
740 had a poor performance in index estimation (large RMSE and bias). This may reflect the
741 fact that the binary partitioning approach is constrained to generate rectangular area
742 strata. However, in the real world, the distribution of fish density is unlikely to be
743 structured as rectangular shapes. Of the remaining approaches evaluated, we found a
744 higher conditional R^2 and lower conditional AIC by the Spatial 0.1 and Spatial 1 GLMMs
745 than the Binary GLMM fitted to the Pacific saury fishery data (**Table 1**). The simulation
746 results also indicated a good performance in index estimation by the Spatial 0.1 and
747 Spatial 1 GLMMs under the random sampling scenario. In general, the spatial clustering
748 approach creates a flexible shape for the area strata (based on similarity of average CPUE
749 and spatial proximity of areas) that may reflect the patchy nature of Pacific saury habitats
750 (associated with eddies and SST fronts, Kuroda and Yokouchi, 2017; Ichii et al., 2018)
751 better than rectangular area strata. However, we noted that the simulation result of the
752 Spatial 0.1 approach (highly density-weighted) under the preferential sampling scenario

753 had the largest model error and bias of index estimation relative to the other GLMMs. In
754 addition, a simulation study by Ono et al. (2015) suggested that the Spatial 0.1 approach
755 could lead to poorer index estimation when there was a directional change in the fishing
756 ground over time. It should be noted that such patterns of directional change in the
757 fishing ground were also observed in the Pacific saury fishery. Hashimoto et al. (2020)
758 and Hsu et al. (2021) reported that the centroid of gravity of the Pacific saury fishing
759 ground has slightly moved eastward following a shift in the fish distribution.

760 We identified inconsistencies in the results based on the model selection criteria
761 (conditional R^2 and conditional AIC) compared to the simulation results for the Spatial
762 0.1 approach, which implied that model selection criteria are not necessarily the best way
763 to finding the model that produces the most unbiased index of abundance. This also
764 highlights the fact that simulation testing is an important aspect in evaluating CPUE
765 standardization methods (Ono et al., 2015; Grüss et al., 2019; Ducharme-Barth et al.,
766 2022), and we recommend its use. Overall, we recommended that a highly density-
767 weighted spatial cluster-based approach (i.e., Spatial 0.1) may not be useful when there is
768 a non-random spatiotemporal distribution of fishers relative to that of fish populations
769 and fish distribution shift. However, based on the simulation results, the Spatial 1
770 approach could be a reasonable alternative for defining spatial strata if a spatio-temporal
771 analysis is not possible.

772 In the current VAST analysis, we assumed an unbalanced spatial knot structure
773 where knots were allocated spatially using a *k*-means algorithm in proportion to sampling
774 intensity. A sensitivity to this choice was explored and though it was determined that
775 specifying the spatial knots in proportion to sampling intensity versus uniformly across
776 the spatial domain did not meaningfully impact the mean index, it did impact the
777 associated uncertainty estimates with the uniform knot distribution resulting in larger
778 estimated CVs. This is likely due to the sparsity (or absence) of samples corresponding to
779 knots at the periphery of the spatial distribution resulting in larger estimates of
780 uncertainty. This may be a desirable property if the spatial domain of the model is
781 correctly specified to match the spatial distribution of the underlying population, with the
782 increased uncertainty representing a lack of sampling of the population in certain areas.

However, if the spatial domain is mis-specified (e.g., overly broad relative to the spatial distribution of the underlying population) then this increased variance may be inappropriate. We noted that previous studies (Grüss et al., 2019; Ducharme-Barth et al., 2022) have indicated that using a uniform allocation of knots in VAST may improve the index estimation when applied to spatially-imbalanced data (e.g., fishery-dependent data). To our knowledge, however, there has not been a study that explicitly compares the performance of different knot structures in VAST. We recommend that future studies investigate the impact of different configurations of knot allocation in VAST on the index estimation and associated uncertainty to assist in further improvement.

The Taiwanese stick-held dip net logbook data does not contain zero catch values due to fishers attracting Pacific saury schools with fishing lamps. In addition to the previously mentioned issue with the effort definition (hauls v.s days), not including information on fishing lamps may also lead to non-proportionality involving standardized CPUE remaining high while abundance declines. This is known as hyperstability (Hilborn and Walters, 1992) and can lead to overestimation of biomass and underestimation of fishing mortality (Crecco and Overholtz, 1990). Such information on the number and/or power of lamps may be implicitly addressed by the random vessel effect in the current study. However, the inclusion of a vessel random effect can likely only account for differences in fishing power across vessels and can not account for temporal changes in the fishing power of a single vessel. We encourage the collection of these key catchability information in the logbook so that they can be included in the CPUE standardization explicitly. We also recommend that future studies examine the impact on CPUE standardization model performance of the various treatments for the vessel random effect (e.g., random covariate in spatially stratified GLMMs and mean-zero random effect with a standard deviation of one in VAST; Xu et al., 2019; Ducharme-Barth et al., 2022) as it was beyond the scope of this study, although results from the influence analysis (**Fig. 3**c) indicated that the impact on the mean index is likely similar.

5. Conclusions

CPUE standardization is essential to stock assessment, and a standardized CPUE index is often the major source of data to assess the stock status in commercial or recreational fisheries (Hilborn and Walters, 1992). We have provided a real-world example (fishery-dependent CPUE of Pacific saury) and a simulation experiment to compare the performance of various spatial treatments within a consistent GLMM framework. Our results build upon a growing body of work (Grüss et al., 2019; Zhou et al., 2019) which indicates that explicitly modelling the spatial autocorrelation structure of the data using GMRFs in a spatio-temporal modelling framework (such as VAST) can account for and provide a more precise treatment to address the spatial heterogeneity of CPUE data. In contrast, the use of area stratifications determined in an ad hoc manner or constrained to rectangular grids may cause the underlying fish density distribution to be misinterpreted. Although the most extreme spatial clustering approach (Spatial 0.1) created a flexible shape for the area strata, this approach may lead to a substantial model error and bias in index estimation if the spatiotemporal distribution of fishers is non-random and non-stationary relative to that of the underlying fish distribution. A less extreme spatial clustering approach (Spatial 1) could be a reasonable alternative for defining spatial areas, based on the simulation results, if it is not possible to apply a spatio-temporal modelling framework.

Most CPUE standardization analyses have focused on the resulting abundance indices arising from the CPUE standardization model(s), but did not explore how the inclusion of each explanatory variable in the model affected the standardized index. The influence analysis of each explanatory variable can be a useful tool in helping analysts understand the difference(s) between unstandardized and standardized indices. In this study, we extended the original *influ* package to a spatio-temporal modelling framework such as VAST and also considered the influence analysis for the year × spatial interaction term. Although this study was focused on Pacific saury, the methodology should be broadly applicable to other fisheries for which similar data are available. Hence, the methods and analysis we have presented can now be considered as standard tools for conducting CPUE standardization analyses.

Acknowledgments

We thank the Overseas Fisheries Development Council (OFDC) of Taiwan for providing data from the Taiwanese stick-held dip net fishery and the anonymous reviewers for comments on the manuscript. We also thank William Walsh for comments on an earlier version of this manuscript. We have no conflict of interest to declare. This study was partially financially supported by the Ministry of Science and Technology and the Fishery Agency of Council of Agriculture of Taiwan through the research grants 111-2611-M-002-009- and 111AS-6.4.1-FA-F6 to Yi-Jay Chang. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect those of National Oceanic and Atmospheric Administration (NOAA) or the U.S. Department of Commerce.

Appendix A. Supplementary materials

The R codes utilized to conduct the binary and spatial clustering area stratification and the influence analyses for the spatial and spatio-temporal random effects for VAST are accessible at the following GitHub repository: https://github.com/jhenhsuNTU/spatial.treatment.influ.analysis.manuscript.

References

Allen, P.M., McGlade, J.M., 1986. Dynamics of discovery and exploitation: the case of the Scotian shelf groundfish fisheries. Can. J. Fish. Aquat. Sci. 43 (6), 1187–1200.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Software 67, 1–48.

Battaile, B. C., Quinn II, T. J. 2004. Catch per unit effort standardization of the eastern Bering Sea walleye pollock (*Theragra chalcogramma*) fleet. Fish. Res., 70, 161–177.

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, 84–88.

871 Bishop, J. 2006. Standardizing fishery-dependent catch and effort data in complex
872 fisheries with technology change. Rev. Fish Biol. Fish., 16, 21–38.

873 Burnham, K. P., and Anderson, D. R. 2002. A practical information-theoretic approach.
874 Model selection and multimodel inference. Second edition. Springer, New York. 488 pp.

875 Campbell, R. A. 2004. CPUE standardisation and the construction of indices of stock
876 abundance in a spatially varying fishery using general linear models. Fish. Res., 70, 209–
877 227.

878 Campbell, R. A. 2015. Constructing stock abundance indices from catch and effort data:
879 Some nuts and bolts. Fish. Res., 161, 109–130.

880 Chang, Y. J., Lan, K. W., Walsh, W. A., Hsu, J., Hsieh, C. H., 2019. Modelling the
881 impacts of environmental variation on habitat suitability for Pacific saury in the
882 Northwestern Pacific Ocean. Fish. Oceanogr. 28, 291–304.

883 Clark, C.W., Mangel, M., 1979. Aggregation and fishery dynamics - theoretical study of
884 schooling and the purse seine tuna fisheries. Fish. Bull. 77, 317–337.

885 Conn, P.B., Thorson, J.T., Johnson, D.S., 2017. Confronting preferential sampling when
886 analysing population distributions: diagnosis and model-based triage. Methods Ecol.
887 Evol. 8, 1535–1546.

888 Crecco, V., and Overholtz, W.J. 1990. Causes of density-dependent catchability for
889 Georges Bank haddock *Melanogrammus aeglefinus*. Can. J. Fish. Aquat. Sci. 47(2):385–
890 394.

891 Ducharme-Barth, N. D., Grüss, A., Vincent, M. T., Kiyofuji, H., Aoki, Y., Pilling, G.,
892 Hampton, J., Thorson, J. T. 2022. Impacts of fisheries-dependent spatial sampling
893 patterns on catch-per-unit-effort standardization: A simulation study and fishery
894 application. Fish. Res., 246, 106169.

895 Forrestal, F.C., Schirripa, M., Goodyear, C.P., Arrizabalaga, H., Babcock, E.A., Coelho,
896 R., Ingram, W., Lauretta, M., Ortiz, M., Sharma, R., 2019. Testing robustness of CPUE
897 standardization and inclusion of environmental variables with simulated longline catch
898 datasets. Fish. Res. 210, 1–13.

Feenstra, J., McGarvey, R., Linnane, A., Haddon, M., Matthews, J., Punt, A. E., 2019. Impacts on CPUE from vessel fleet composition changes in an Australian lobster (*Jasus edwardsii*) fishery. N. Z. J. Mar. Freshwater Res., 53, 292–302.

Glazer, J. P., Butterworth, D. S. 2002. GLM-based standardization of the catch per unit effort series for South African west coast hake, focusing on adjustments for targeting other species. Afr. J. Mar. Sci., 24, 323–339.

Greven, S., Kneib, T. 2010. On the behaviour of marginal and conditional AIC in linear mixed models. Biometrika, 97, 773–789.

Grüss, A., Walter III, J. F., Babcock, E. A., Forrestal, F. C., Thorson, J. T., Lauretta, M. V., 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.

Haltuch, M.A., Ono, K., Valero, J.L., 2013. Status of the U.S. Petrale Sole Resource in 2012. Pacific Fishery Management Council, Portland, Oregon.

Han, Q., Grüss, A., Shan, X., Jin, X., Thorson, J. T. 2021. Understanding patterns of distribution shifts and range expansion/contraction for small yellow croaker (*Larimichthys polyactis*) in the Yellow Sea. Fish. Oceanogr., 30, 69–84.

Hashimoto, M., Nishijima, S., Yukami, R., Watanabe, C., Kamimura, Y., Furuichi, S., Ichinokawa, M., Okamura, H., 2019. Spatiotemporal dynamics of the Pacific chub mackerel revealed by standardized abundance indices. Fish. Res. 219, 105315.

Hashimoto, M., Kidokoro, H., Suyama, S., Fuji, T., Miyamoto, H., Naya, M., Kitakado, T., 2020. Comparison of biomass estimates from multiple stratification approaches in a swept area method for Pacific saury *Cololabis saira* in the western North Pacific. Fish Sci. 1–12.

Hashimoto, M., Naya, M., Suyama, S., Nakayama, S.I., Fuji, T., Miyamoto, H., Kawabata, A., Nakatsuka, S. 2021. Standardized CPUE of Pacific saury (*Cololabis saira*) caught by the Japanese stick-held dip net fishery up to 2020. NPFC-2021-SSC PS07-WP07.

927 Hilborn, R., Walters, C.J., 1987. A general model for simulation of stock and fleet
928 dynamics in spatially heterogeneous fisheries. Can. J. Fish. Aquat. Sci. 44, 1366–1369.

929 Hilborn, R., and Walters, C.J., 1992. Quantitative Fisheries Stock Assessment and
930 Management: Choice, Dynamics and Uncertainty. Chapman & Hall, New York, 570 pp.

931 Holdsworth, J. C., Saul, P. J., 2017. Striped marlin catch and CPUE in the New Zealand
932 sport fishery 2013-14 to 2015-16. New Zealand Fisheries Assessment Report, 18, 27.

933 Hoyle, S. D., Huang, H., Kim, D. N., Lee, M. K., Matsumoto, T., Walter, J., 2019.
934 Collaborative study of bigeye tuna CPUE from multiple Atlantic Ocean longline fleets in
935 2018. Collect. Vol. Sci. Pap. ICCAT, 75, 2033–2080.

936 Hsu, J., Chang, Y. J., Kitakado, T., Kai, M., Li, B., Hashimoto, M., and Park, K. J. 2021.
937 Evaluating the spatiotemporal dynamics of Pacific saury in the Northwestern Pacific
938 Ocean by using a geostatistical modelling approach. Fish. Res., 235, 105821.

939 Huang, W. B., Lo, N. C., Chiu, T. S., and Chen, C. S. 2007. Geographical distribution
940 and abundance of Pacific saury, *Cololabis saira* (Brevoort) (Scomberesocidae), fishing
941 stocks in the northwestern Pacific in relation to sea temperatures. Zool. Res., 46, 705.

942 Huang, W. B., Chang, Y. J., Hsieh, C. H., 2020. Standardized CPUE of Pacific saury
943 (*Cololabis saira*) caught by the Chinese Taipei stick-held dip net fishery up to 2019.
944 NPFC-2020-SSC PS06-WP05.

945 Huang, W. B., Chang, Y. J., Hsieh, C. H., 2021. Standardized CPUE of Pacific saury
946 (*Cololabis saira*) caught by the Chinese Taipei stick-held dip net fishery up to 2020.
947 NPFC-2021-SSC PS07-WP14.

948 Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between Gaussian fields
949 and Gaussian Markov random fields: the stochastic partial differential equation approach.
950 J. R. Stat. Soc., B: Stat. Methodol., 73, 423–498.

951 Ichinokawa, M., and Brodziak, J. 2010. Using adaptive area stratification to standardize
952 catch rates with application to North Pacific swordfish (*Xiphias gladius*). Fish. Res., 106,
953 249–260.

954 Ichii, T., Nishikawa, H., Mahapatra, K., Okamura, H., Igarashi, H., Sakai, M., Suyama,
955 S., Nakagami, M., Naya, M., Usui, N., Okada, Y., 2018. Oceanographic factors affecting
956 interannual recruitment variability of Pacific saury (*Cololabis saira*) in the central and
957 western North Pacific. Fish. Oceanogr. 00, 1–13.

958 IOC-IHO-BODC, A., 2003. Centenary Edition of the GEBCO Digital Atlas. The
959 Intergovernmental Oceanographic Commission, The International Hydrographic
960 Organization and the British Oceanographic Data Centre, Liverpool, UK.

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

964 Kai, M. 2019. Spatio-temporal changes in catch rates of pelagic sharks caught by
965 Japanese research and training vessels in the western and central North Pacific. Fish.
966 Res., 216, 177–195.

967 Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster
968 Analysis. Wiley, New Jersey.

969 Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. TMB:
970 Automatic differentiation and Laplace approximation. Journal of Statistical Software, 70:
971 1–21.

972 Kristensen, K., Thygesen, U.H., Andersen, K.H., Beyer, J.E., 2014. Estimating
973 spatiotemporal dynamics of size-structured populations. Can. J. Fish. Aquat. Sci. 71,
974 326–336.

975 Kuroda, H., Yokouchi, K., 2017. Interdecadal decrease in potential fishing areas for
976 Pacific saury off the southeastern coast of Hokkaido, Japan. Fish. Oceanogr., 26, 439–
977 454.

978 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2014. Cluster: Cluster
979 Analysis Basics and Extensions R Package Version 1.15.2.

980 Maunder, M. N., Punt, A. E. 2004. Standardizing catch and effort data: a review of recent
981 approaches. Fish. Res., 70, 141–159.

Maunder, M. N., Sibert, J. R., Fonteneau, A., Hampton, J., Kleiber, P., Harley, S. J. 2006. Interpreting catch per unit effort data to assess the status of individual stocks and communities. ICES J. Mar. Sci., 63, 1373–1385.

Maunder, M.N., Thorson, J.T., Xu, H., Oliveros-Ramos, R., Hoyle, S.D., Tremblay-Boyer, L., Lee, H.H., Kai, M., Chang, S.-K., Kitakado, T., Albertsen, C.M., Minte-Vera, C.V., Lennert-Cody, C.E., Aires-da-Silva, A.M., Piner, K.R., 2020. The need for spatiotemporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models. Fish. Res., 229, 105594.

Miyabe, N., Takeuchi, Y. 2003. Standardized bluefin CPUE from the Japanese longline fishery in the Atlantic and Mediterranean Sea up to 1999. Collect. Vol. Sci. Pap. ICCAT, 52, 1130–1144.

Nakano, H. 1998. Stock status of Pacific swordfish, *Xiphias gladius*, inferred from CPUE of the Japanese longline fleet standardized using general linear models. US Nat. Mar. Fish. Serv., NOAA Tech. Rep. NMFS, 142, 195–209.

Nakagawa, S., Schielzeth, H. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods Ecol. Evol., 4, 133–142.

Ono, K., Punt, A. E., Hilborn, R. 2015. Think outside the grids: An objective approach to define spatial strata for catch and effort analysis. Fish. Res., 170, 89–101.

Pinheiro, J. C., Bates, D. M. 2000. Linear mixed-effects models: basic concepts and examples. Mixed-effects models in S and S-Plus, 3–56.

Punt, A. E., Walker, T. I., Taylor, B. L., Pribac, F. 2000. Standardization of catch and effort data in a spatially-structured shark fishery. Fish. Res., 45, 129–145.

Quinn II, T. J., Hoag, S. H., Southward, G. M. 1982. Comparison of two methods of combining catch-per-unit-effort data from geographic regions. Can. J. Fish. Aquat. Sci., 39, 837–846.

R Core Team, 2021. R: a language and environment for statistical computing [online]. R Foundation For Statistical Computing, Vienna, Austria. Available from http://www.R-project.org/.

Rose, G.A., Leggett, W.C., 1991. Effects of biomass range interactions on catchability of migratory demersal fish by mobile fisheries - An example of Atlantic cod (*Gadus morhua*). Can. J. Fish. Aquat. Sci. 48, 843–848.

Rose, G.A., Kulka, D.W., 1999. Hyperaggregation of fish and fisheries: how catch-per-unit-effort increased as the northern cod (*Gadus morhua*) declined. Can. J. Fish. Aquat. Sci. 56, 118–127.

Sculley, M. L., Brodziak, J. 2020. Quantifying the distribution of swordfish (*Xiphias gladius*) density in the Hawaii-based longline fishery. Fish. Res., 230, 105638.

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

Shono, H., 2014. Application of support vector regression to CPUE analysis for southern bluefin tuna *Thunnus maccoyii*, and its comparison with conventional methods. Fish Sci 80, 879–886.

Stow, C.A., Jolliff, J., McGillicuddy, D.J., Doney, S.C., Allen, J.I., Friedrichs, M.A., Rose, K.A., Wallhead, P., 2009. Skill assessment for coupled biological/physical models of marine systems. J. Mar. Syst. 76, 4–15.

Swain, D.P., Sinclair, A.F., 1994. Fish Distribution and catchability - What is the appropriate measure of distribution. Can. J. Fish. Aquat. Sci. 51, 1046–1054.

Thorson, J. T., Shelton, A. O., Ward, E. J., Skaug, H. J., 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., 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.

1036 Thorson, J. T., Barnett, L. A., 2017. Comparing estimates of abundance trends and
1037 distribution shifts using single-and multispecies models of fishes and biogenic
1038 habitat. ICES J. Mar. Sci. 74, 1311–1321.

1039 Thorson, J. T., 2019. Guidance for decisions using the Vector Autoregressive Spatio-
1040 Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fish.
1041 Res. 210, 143–161.

1042 Wilberg, M.J., Thorson, J.T., Linton, B.C., Berkson, J., 2010. Incorporating time-varying
1043 catchability into population dynamic stock assessment models (Incorporating Time-
1044 Varying Catchability into Population Dynamic Stock Assessment Models). Rev. Fish.
1045 Sci. 18, 7–24

1046 Winker, H., Kerwath, S. E., Attwood, C. G. 2013. Comparison of two approaches to
1047 standardize catch-per-unit-effort for targeting behaviour in a multispecies hand-line
1048 fishery. Fish. Res., 139, 118–131.

1049 Xu, H., Lennert-Cody, C. E., Maunder, M. N., Minte-Vera, C. V. 2019. Spatiotemporal
1050 dynamics of the dolphin-associated purse-seine fishery for yellowfin tuna (*Thunnus*
1051 *albacares)* in the eastern Pacific Ocean. Fish. Res., 213, 121–131.

1052 Zhou, S., Campbell, R.A., Hoyle, S.D., 2019. Catch per unit effort standardization using
1053 spatio-temporal models for Australia's Eastern Tuna and Billfish Fishery. ICES J. Mar.
1054 Sci. 76, 1489–1504.

Table 1. Summary statistics of Ad hoc, Binary, Spatial 1, Spatial 0.1 GLMMs, and VAST fitted to Pacific saury CPUE data from Northwestern Pacific Ocean during 1997 – 2019. The conditional R^2 denotes the conditional explained residuals, the ΔAIC is the difference between the conditional AIC and that of the best model. The mean conditional R^2 is the explained value of the training data and the mean rho value (ρ) is the Pearson's correlation coefficient derived from the observed and predicted CPUE of the testing data based on the five-fold cross-validations by repeated 10 times. The CV represents the coefficients of variation for the conditional R^2 and ρ from the five-fold cross-validations, respectively.

Model	Goodness-of-fit			Cross-validations			
	conditional R^2	conditional AIC	Δ AIC	mean conditional R^2	CV	mean ρ	CV
Ad hoc	0.27	534365	14063	0.25	0.023	0.43	0.014
Binary	0.34	529145	8843	0.32	0.018	0.48	0.016
Spatial 1	0.37	525713	5411	0.36	0.011	0.52	0.014
Spatial 0.1	0.39	524864	4562	0.39	0.015	0.53	0.017
VAST	0.68	520302	0	0.65	0.005	0.54	0.008

Table 2. Summary of the explanatory power (in conditional R^2) and overall influence for each variable considered in the Ad hoc, Binary, Spatial 1, Spatial 0.1 GLMMs, and VAST by using the Pacific saury data in Northwestern Pacific Ocean during 1997 - 2019.

Model	Variable	Conditional R^2	Overall influence
Ad hoc	*Year*	0.052	-
	SST+*SST*^{2}	0.028	0.017
	Vessel	0.087	0.169
	Area	0.042	0.152
	Year×*Area*	0.065	0.259
Binary	*Year*	0.047	-
	SST+*SST*^{2}	0.045	0.017
	Vessel	0.104	0.170
	Area	0.082	0.135
	Year×*Area*	0.064	0.301
Spatial1	*Year*	0.045	-
	SST+*SST*^{2}	0.021	0.017
	Vessel	0.098	0.169
	Area	0.121	0.122
	Year×*Area*	0.087	0.221
Spatial0.1	*Year*	0.040	-
	SST+*SST*^{2}	0.017	0.017
	Vessel	0.094	0.170
	Area	0.154	0.140
	Year×*Area*	0.087	0.225
VAST	*Year*	0.040	-
	SST+*SST*^{2}	0.032	0.017
	Vessel	0.091	0.169
	Spatial	0.230	0.128
	Spatio-temporal	0.288	0.252

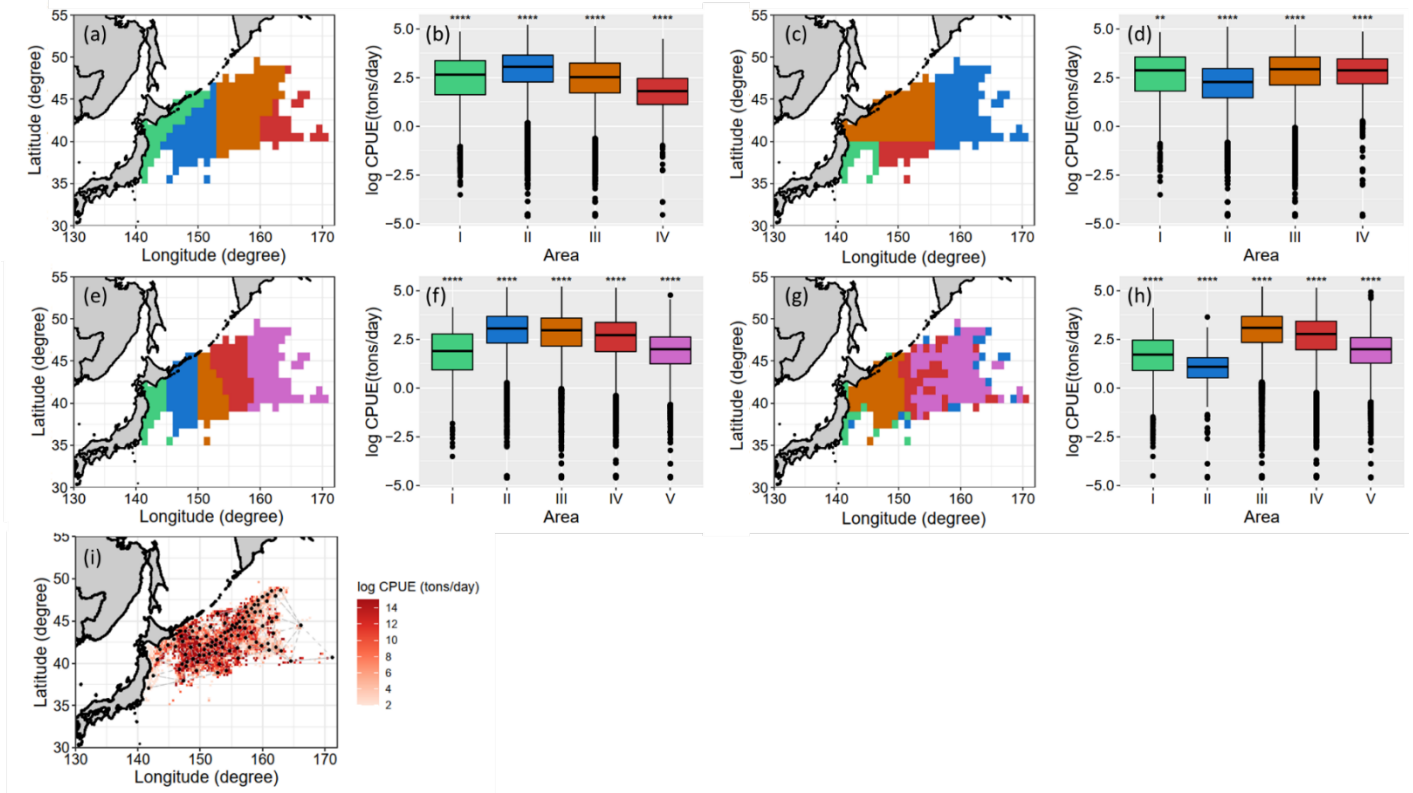
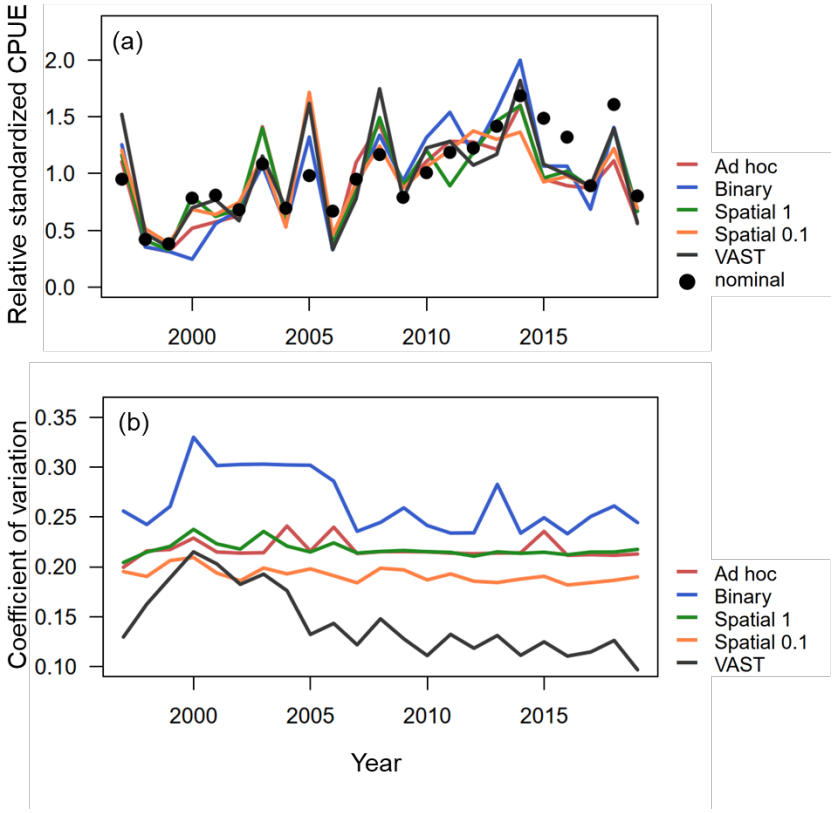
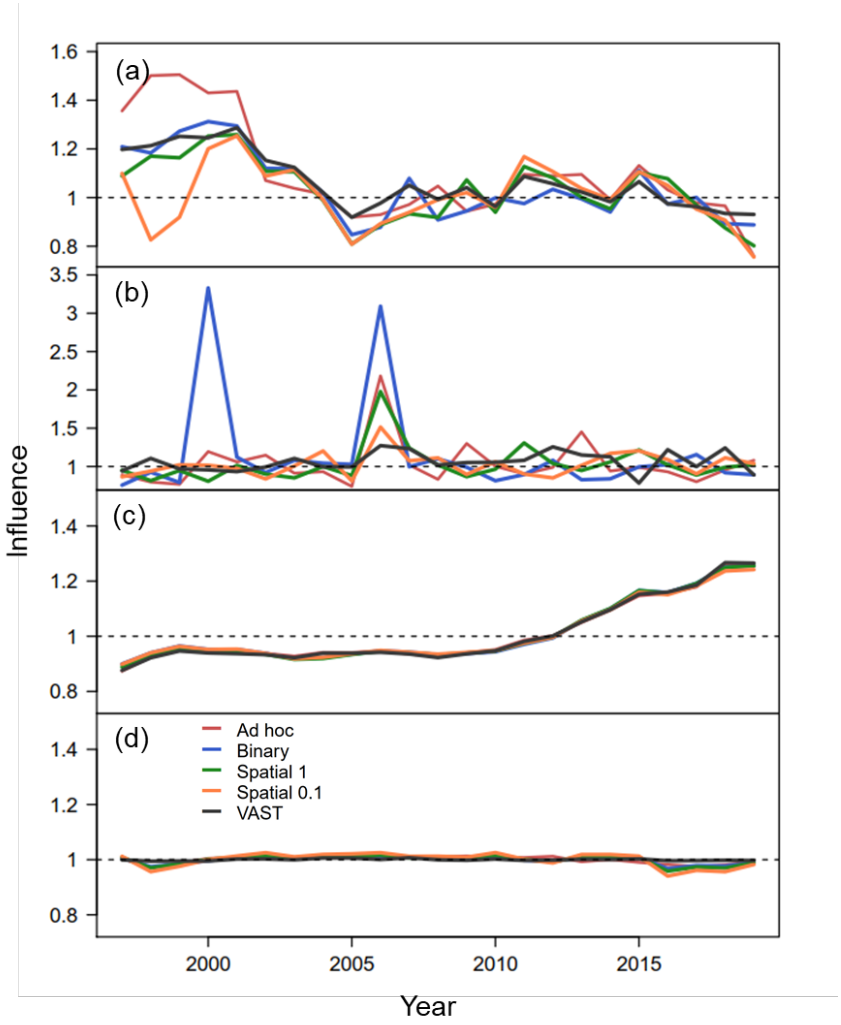


Figure 1. The resulting area stratifications with boxplots of observed CPUEs for each area strata are determined by (a-b) the Ad hoc approach (four area strata), (c-d) the Binary approach (four area strata), (e-f) the Spatial 1 (five area strata), (g-h) the Spatial 0.1 (five area strata); and the (i) knot (in black points) configuration of VAST with the observed CPUE (in red colors) by using the Pacific saury data in Northwestern Pacific Ocean during 1997 - 2019. Star symbols on boxplot represent the significance levels with the overall mean of observed CPUEs (****: $p <= 0.0001$, ***: $p <= 0.001$, **: $p <= 0.01$) by using the ANOVA test.



1078

1079 **Figure 2**. Annual trends of (a) the relative standardized abundance indices without area-
1080 weighting (relative to mean), and (b) the coefficient of variation of standardized CPUE
1081 indices without area-weighting for the Ad hoc, Binary, Spatial 1, Spatial 0.1 GLMMs, and
1082 VAST for Pacific saury in the Northwestern Pacific Ocean during 1997 – 2019. Solid black
1083 points represent the annual nominal CPUEs for Pacific saury.



1084

1085 **Figure 3**. The annual influence of (a) spatial, (b) interaction of year and spatial/spatio-
1086 temporal random effect, (c) vessel, and (d) quadratic water temperature effects for Ad hoc,
1087 Binary, Spatial 1, Spatial 0.1 GLMMs, and VAST for Pacific saury in the Northwestern
1088 Pacific Ocean during 1997 – 2019.

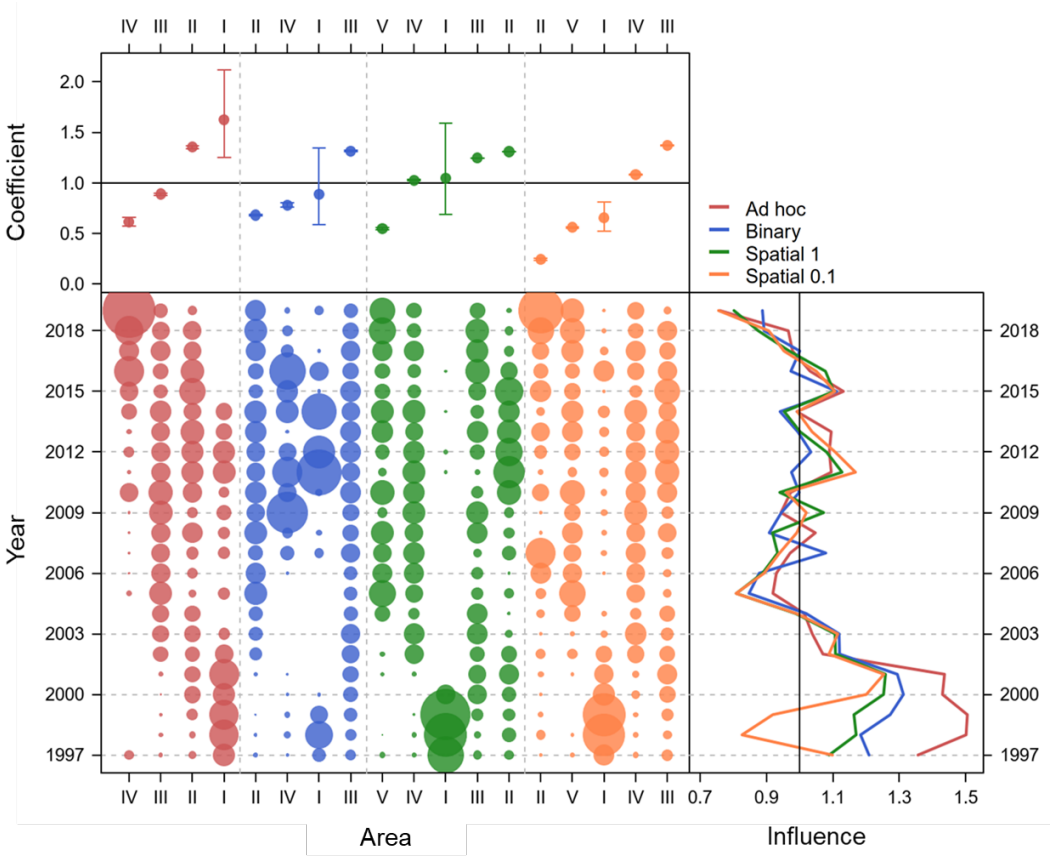


Figure 4. The coefficient-distribution-influence (CDI) plot of the spatial effect for Ad hoc, Binary, Spatial 1 and Spatial 0.1 GLMMs for Pacific saury in the Northwestern Pacific Ocean during 1997– 2019. The top panel of the CDI plot provides the normalized coefficients and their standard errors for each area stratum (four area strata for Ad hoc, Binary GLMMs, and five area strata for Spatial 1 and Spatial 0.1 GLMMs). In the bottom left panel, the bubbles indicate the annual distribution of observed CPUEs from each area stratum. A larger bubble represents a larger number of data records. The bottom right panel shows the annual value of influence for the spatial effect.

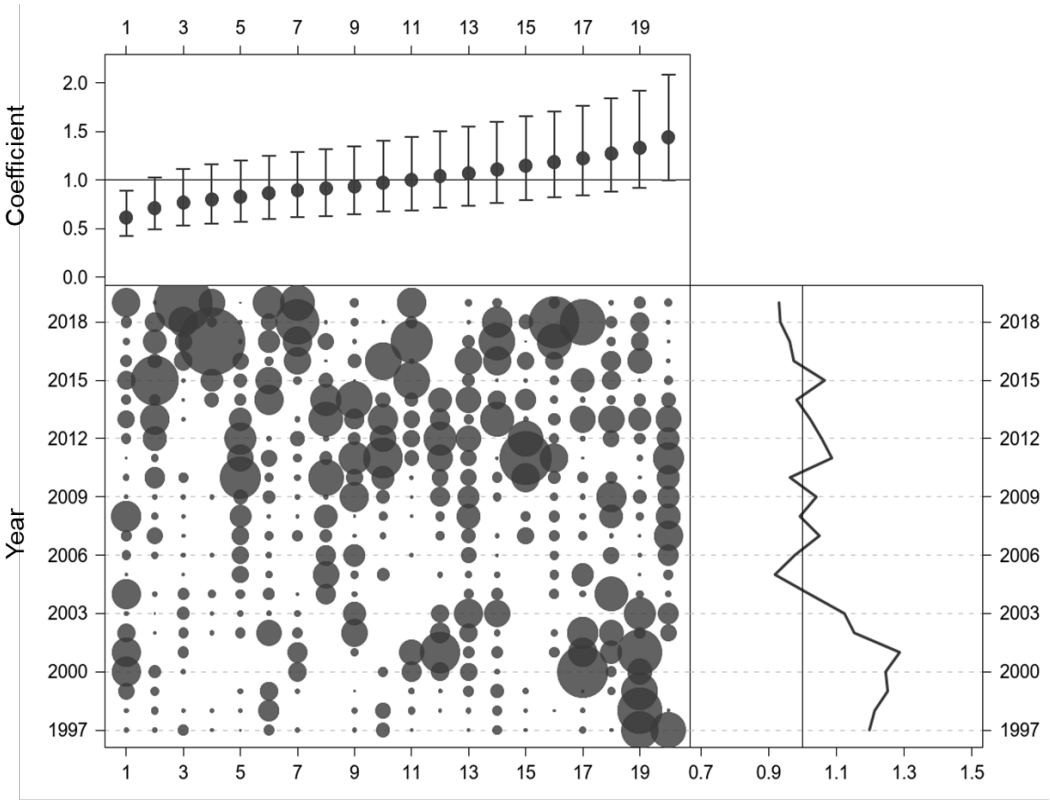
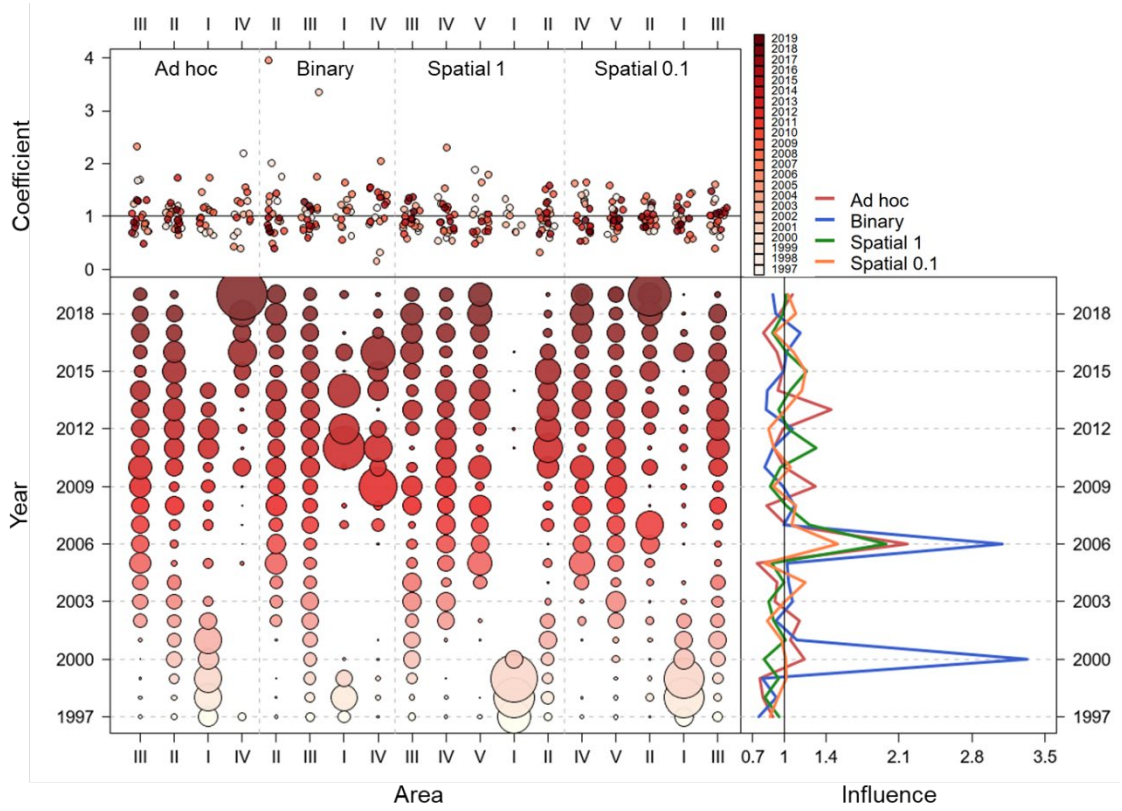


Figure 5. The coefficient-distribution-influence (CDI) plot of the spatial random effect for the VAST model for Pacific saury in the Northwestern Pacific Ocean during 1997 – 2019. The top panel of the CDI plot provides the normalized coefficients and their standard errors. In the bottom left panel, bubbles indicate the annual distribution of observed CPUEs from each grouped knot (every five knots). A larger bubble represents a larger number of data records. The bottom right panel shows the annual value of influence for the spatial random effect.



1105

1106 **Figure 6**. The coefficient-distribution-influence (CDI) of the year × spatial interaction
1107 random effect for the Ad hoc, Binary, Spatial 1, and Spatial 0.1 GLMMs for Pacific saury in
1108 the Northwestern Pacific Ocean during 1997 – 2019. The top panel of the plot provides the
1109 normalized coefficients for each area stratum (four area strata for Ad hoc, Binary GLMMs,
1110 and five area strata for Spatial 1 and Spatial 0.1 GLMMs). The solid color points represent the
1111 coefficient for each year. The area coefficient in each year was jittered with small random
1112 noise for graphical visualization. In the bottom left panel, bubbles indicate the annual
1113 distribution of observed CPUEs from each area stratum. A larger bubble represents a larger
1114 number of data records. The bottom right panel shows the annual value of influence for the
1115 year × spatial interaction random effect.

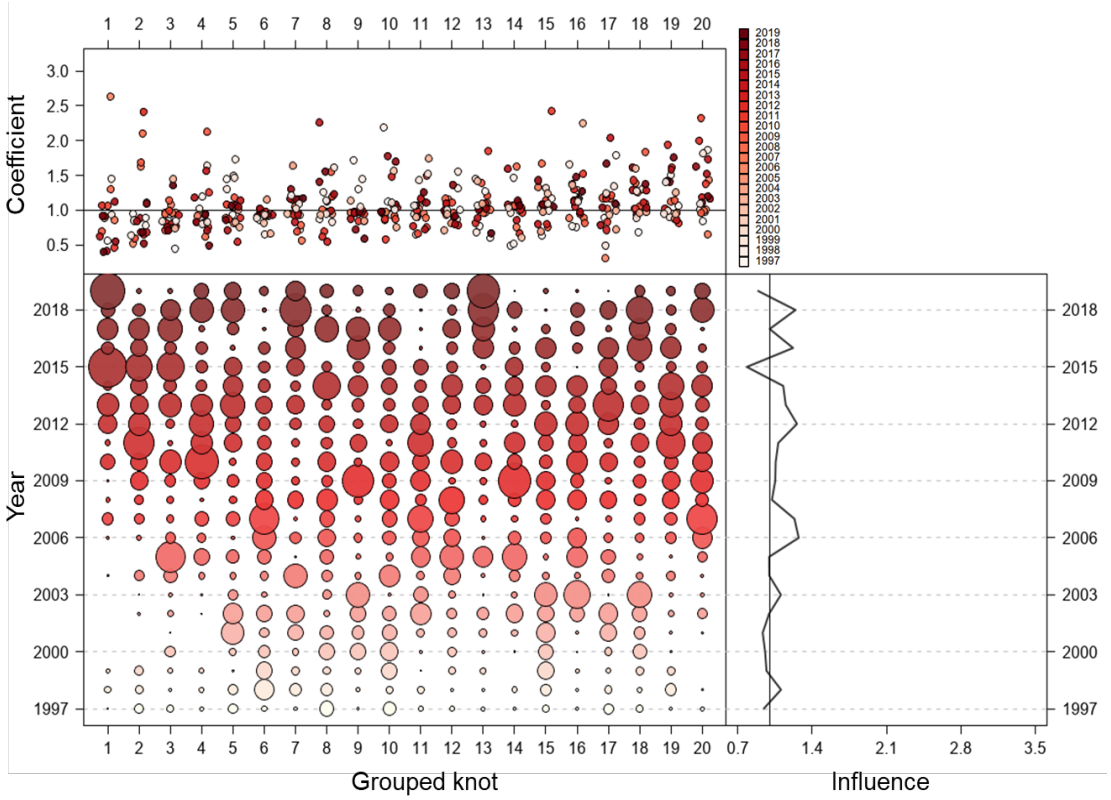
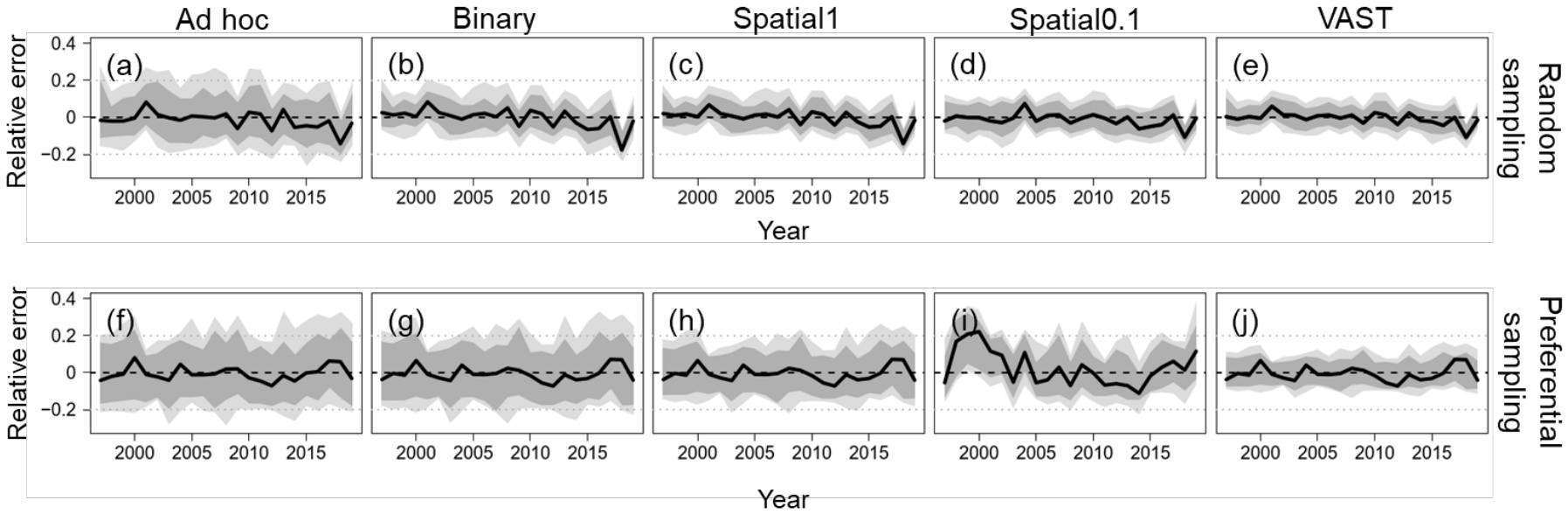


Figure 7. The coefficient-distribution-influence (CDI) plot of spatio-temporal random effect for the VAST model for Pacific saury in the Northwestern Pacific Ocean during 1997 – 2019. The top panel of the plot provides the normalized coefficient for each grouped knot (every five knots). The solid color points represent the coefficient for each year. The grouped knots coefficient in each year was jittered with small random noise for graphical visualization. In the bottom left panel, bubbles indicate the annual distribution of observed CPUEs from each grouped knot. A larger bubble represents a larger number of data records. The bottom right panel shows the annual value of influence for the spatio-temporal random effect.



1125

1126 **Figure 8**. Time series of relative error (RE) in the index of abundance. The dark and light grey polygon represents the 80% and 95% confidence interval,
1127 respectively. RE is calculated for Ad hoc, Binary, Spatial 1 and Spatial 0.1 GLMMs, and VAST with the two sampling scenarios (random, and preferential
1128 samplings). The grey horizontal dashed line is the reference line (±0.2) for the relative errors.

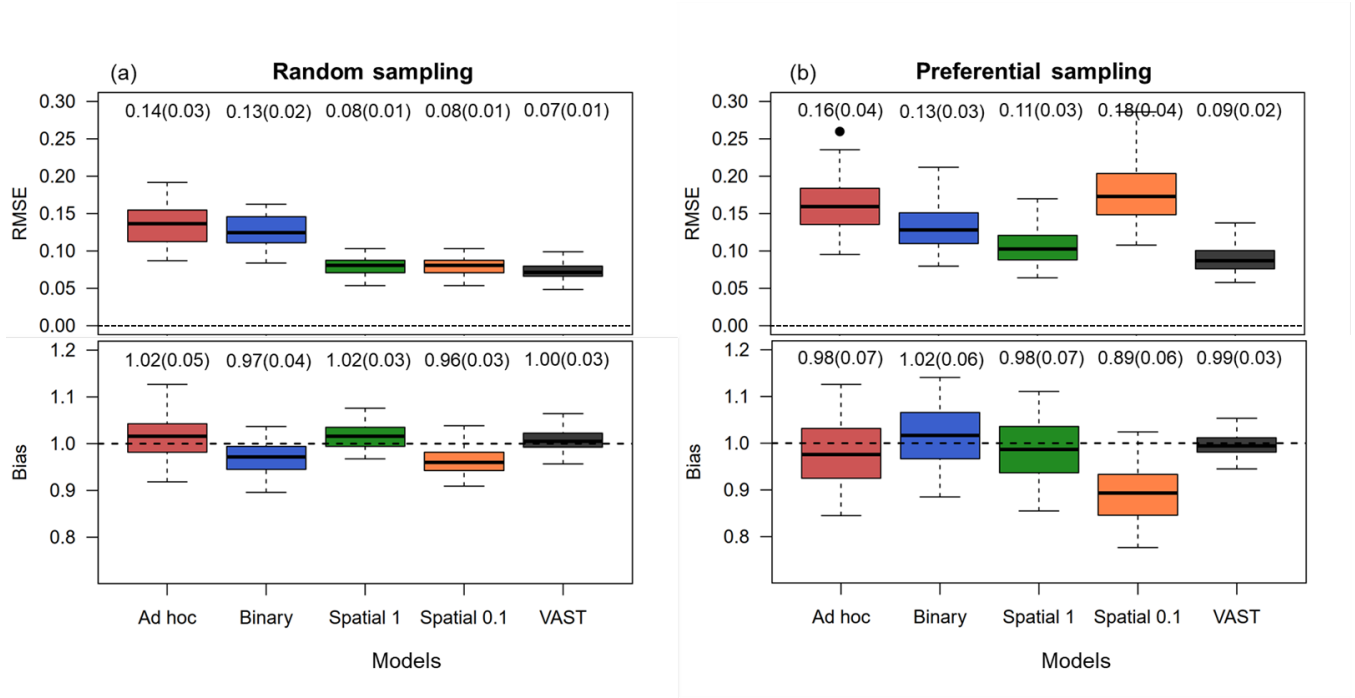
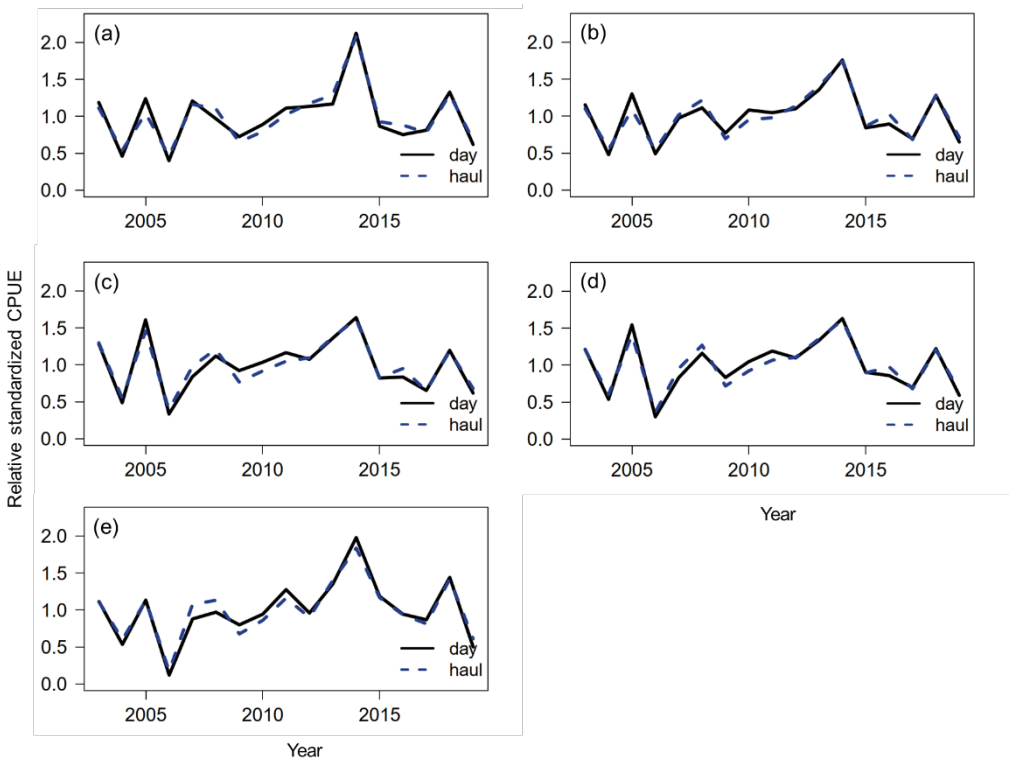


Figure 9. The boxplots of performance metrics for the (a,b) root mean squared error (RMSE) and (c,d) bias metrics calculated across all replicates under the two spatial sampling scenarios (random and preferential samplings) for the Ad hoc, Binary, Spatial 1, Spatial 0.1 GLMMs and VAST. The numbers above the boxplots indicate the mean value of RMSE and bias metrics. Values in the parentheses are the standard deviations. The horizontal dashed line is the reference line for the bias metrics (one represents no bias).

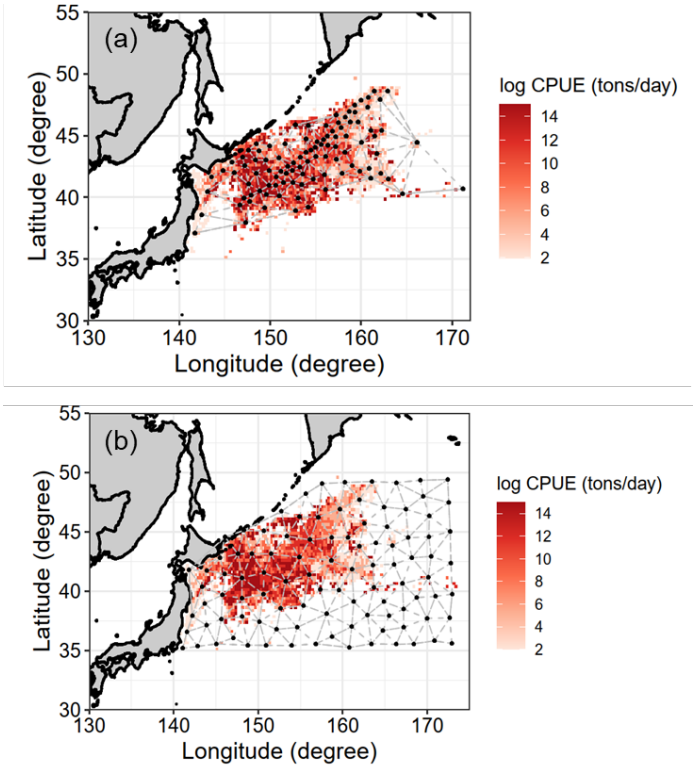
Table S1. Summary of model selection information on Pacific saury data by using the likelihood ratio test for each explanatory variable included in the Ad hoc, Binary, Spatial 1, Spatial 0.1 GLMMs and VAST for Pacific saury in the Northwestern Pacific Ocean during 1997 - 2019.

Model	Variables	Deviance	Number of parameters	*p*-value
Ad hoc	-*Area*	286992	28	<0.001
	-*SST*+*SST*^{2}	287179	29	<0.001
	-*Vessel*	290535	30	<0.001
	-*Year*×*Area*	289624	30	<0.001
	Full	284748	31	<0.001
Binary	-*Area*	281147	28	<0.005
	-*SST*+*SST*^{2}	281159	29	<0.001
	-*Vessel*	284923	30	<0.001
	-*Year*×*Area*	285834	30	<0.001
	Full	280846	31	<0.001
Spatial 1	-*Area*	278613	28	<0.001
	-*SST*+*SST*^{2}	278628	30	<0.001
	-*Vessel*	282355	31	<0.001
	-*Year*×*Area*	283293	31	<0.001
	Full	278420	32	<0.001
Spatial 0.1	-*Area*	277783	28	<0.001
	-*SST*+*SST*^{2}	277622	30	<0.001
	-*Vessel*	281343	31	<0.001
	-*Year*×*Area*	281157	31	<0.001
	Full	275648	32	<0.001
VAST	-*Spatial*	1064595	45	<0.001
	-*SST*+*SST*^{2}	1064501	45	<0.001
	-*Vessel*	1066833	44	<0.001
	-*Spatio-temporal*	1071044	44	<0.001
	Full	1050998	46	<0.001



1139

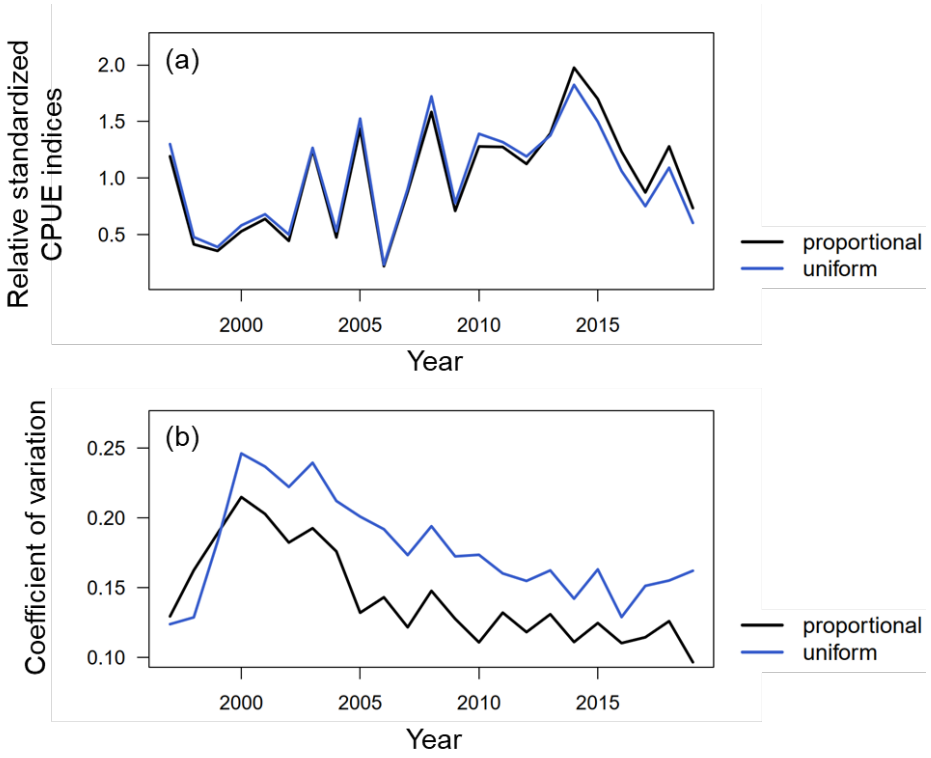
1140 **Figure S1**. Annual trends of the relative standardized abundance indices (normalized to their
1141 mean) calculated by using the effort metrics of fishing day (in black line), and haul (dashed
1142 line) for (a) Ad hoc, (b) Binary, (c) Spatial 1, (d) Spatial 0.1 GLMMs and (e) VAST for
1143 Pacific saury in the Northwestern Pacific Ocean during 1997 - 2019.



1144

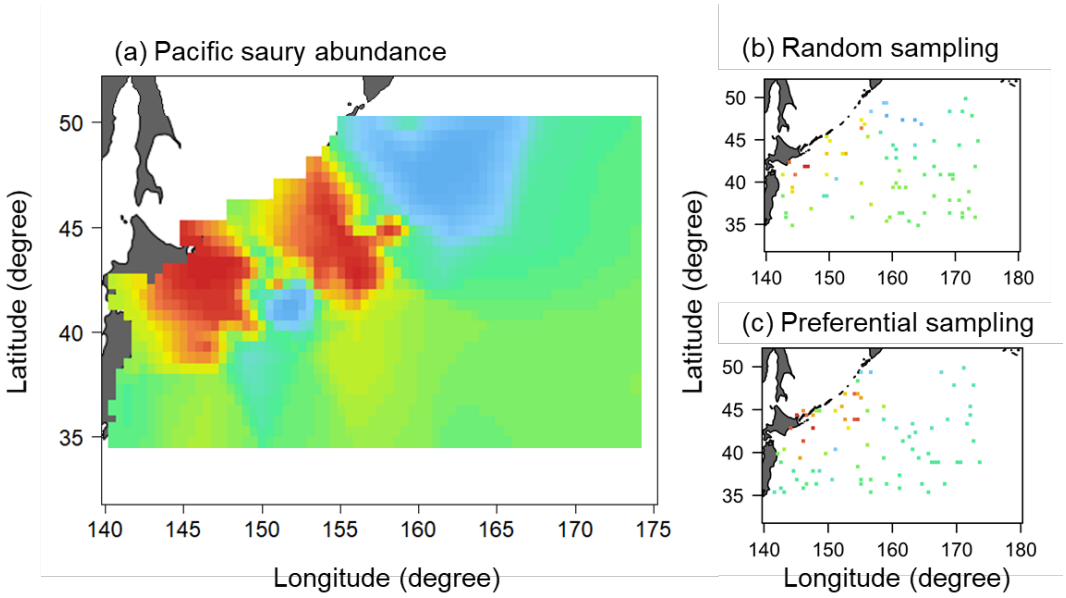
1145

1146 **Figure S2**. The defined spatial distribution of knots (in black points) based on (a) the *k*-means
1147 algorithm and (b) uniform allocation (in red colors) for the Pacific saury data in the
1148 Northwestern Pacific Ocean during 1997 – 2019.



1149

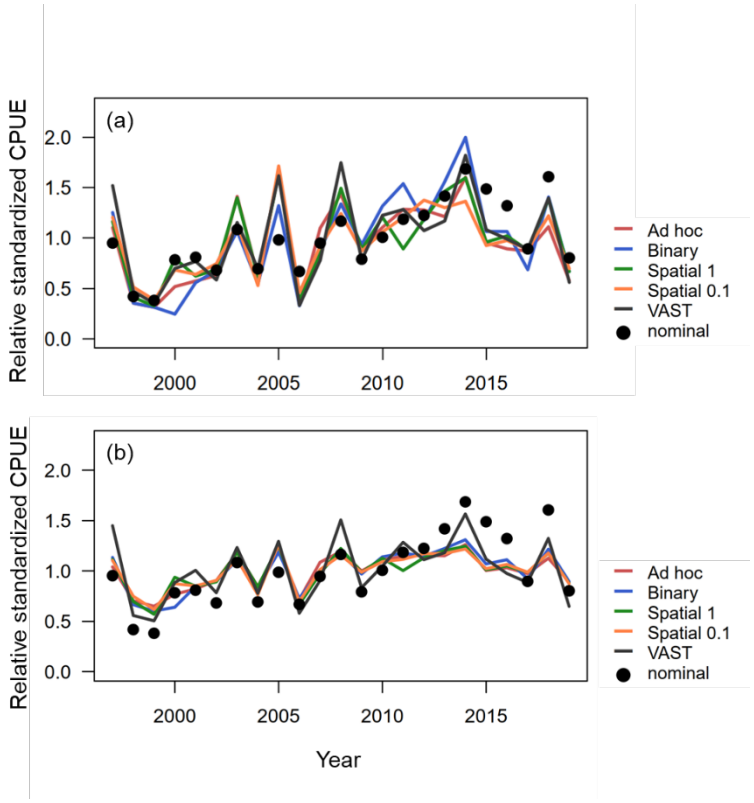
1150 **Figure S3.** Results of (a) standardized CPUEs (normalized to their mean) and (b) coefficient
1151 of variations derived from VAST with proportional knot distribution (in black line) and the
1152 uniform knot distribution (in blue line).



1153

1154 **Figure S4.** (a) Simulated spatial distribution of Pacific saury abundance in the Northwestern
1155 Pacific Ocean. Red (blue) colors indicate greater (smaller) levels of abundance. The simulated
1156 distributions of fishery data for the Pacific saury with (b) random sampling and (c)
1157 preferential sampling scenarios.

1158



1159

1160 **Figure S5**. Annual trends of (a) the relative standardized abundance indices (normalized to
1161 their mean) without area-weighting, and (b) with area-weighting for the Ad hoc, Binary,
1162 Spatial 1, Spatial 0.1 GLMMs, and VAST for Pacific saury in the Northwestern Pacific Ocean
1163 during 1997 – 2019. Solid black points represent the annual nominal CPUEs for Pacific saury.

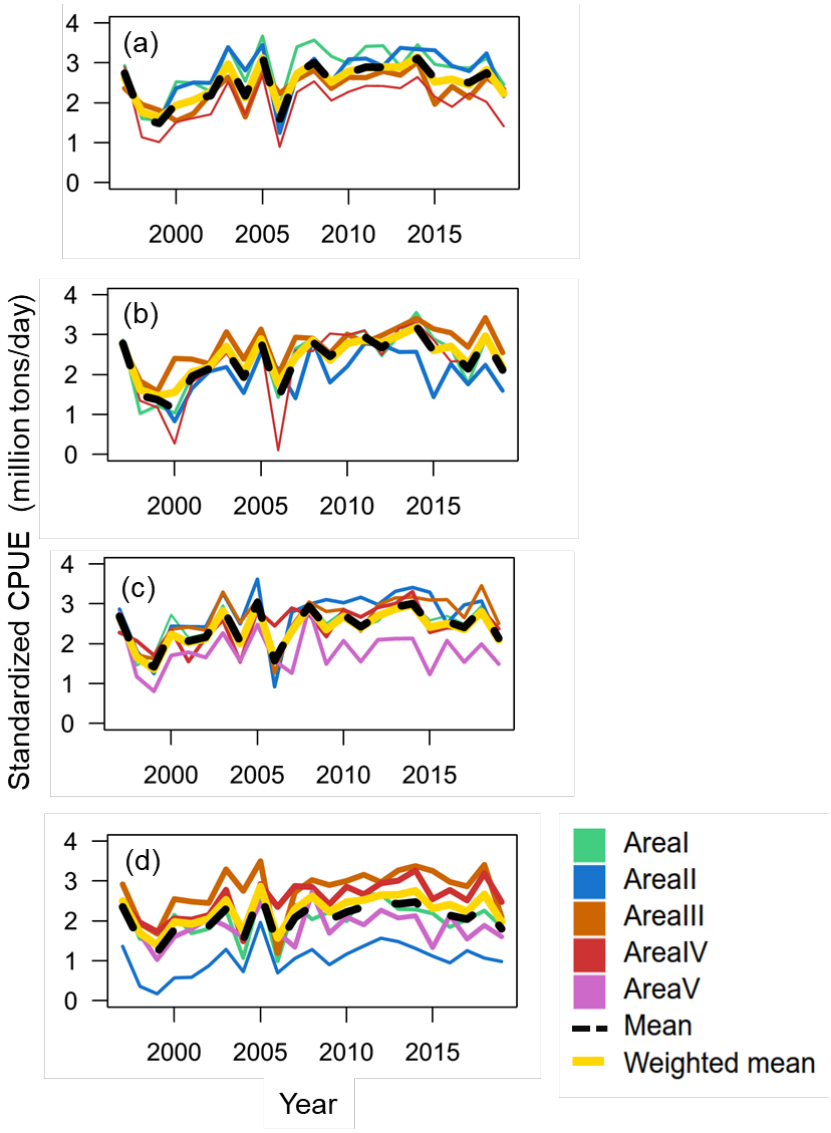
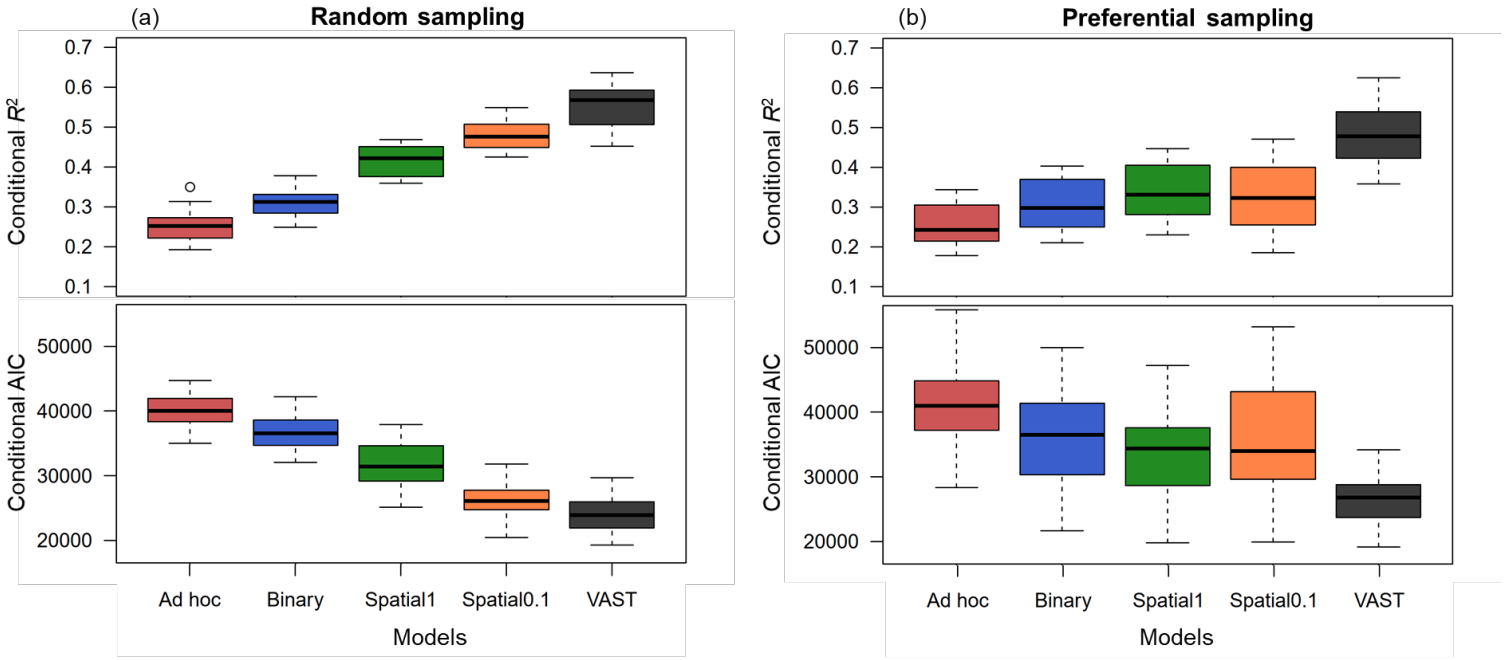


Figure S6. Time series of the standardized abundance indices for each area derived from the (a) Ad hoc, (b) Binary, (c) Spatial1 and (d) Spatial 0.1 GLMMs for Pacific saury in the Northwestern Pacific Ocean during 1997 – 2019. Black dash line represents the arithmetic mean of standardized abundance over areas. Yellow solid line represents the weighted mean of standardized abundance over areas. Thickness of line represents the size of the surface area for each area strata (a larger surface area has a thicker line).



1171

1172 **Figure S7**. The boxplots of model selection criteria for the conditional R^2 and conditional AIC derived from all replicates under the (a) random and (b)

1173 preferential sampling scenarios for the Ad hoc, Binary, Spatial 1, Spatial 0.1 GLMMs and VAST.