CPUE standardization for Taiwanese PBF fisheries using delta-GLMM and VAST, incorporating SST and size data

Tzu-Lun Yuan[1], Shui-Kai Chang[2], Yi Chang[2]

[1] Department of Applied Mathematics, Tunghai University, Taichung, Taiwan
[2] National Sun Yat-sen University, Kaohsiung, Taiwan

March 2023

Working document submitted to the ISC Pacific Bluefin Tuna Working Group, International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean (ISC), from 21 to 24 March 2023, Shinagawa, Japan.

Summary

Total catch of PBF of Taiwanese coastal and offshore fisheries (mainly from longline fishery) had been as high as 3,089 mt in 1999 but continuously declined to the lowest record of 214 mt (210 mt from longline fishery) in 2012. Thereafter, the catch slowly bounced back and stayed at the level of 400–550 mt during 2014–2019. Thereafter, the catch jumped up to 1,154 mt in 2020 and reached 1,475 mt in 2021 and 1497 mt in 2022. Number of registered PBF vessels fluctuated between 480 to 560 in recent ten years. Average size of PBF was around 212–220 cm before 2008; thereafter, the average in the North region stably maintained at 218–224 cm during 2008–2022; while in the South region, the average gradually increased to 235 cm in 2012 and declined to about 210 cm during 2020–2022. The substantial increase of average size in the South was resulting from the decline of recruitment to the fishing ground, and the decrease since 2013 was a response to more smaller fish recruited and more large fish removed from the region.

Three model designs were applied for PBF CPUE standardization: traditional delta-GLMM with SST effect, VAST with SST effect, and VAST incorporating size data. Standardized CPUE series did not show obvious difference for traditional delta-GLMM with or without SST effect. The VAST incorporating size data (converted to seven age groups) suggested that the South region has much higher density than the North, but the general trends are similar for both regions except for the latest one or two years for age groups older than 15 years old, and that age group 9-11 was the most dominant fish, followed by the 6-8 age group, for recent years. Relative CPUE series from GLMM and VAST with SST were compared and suggested that the two series in the South region is in similar trend with VAST series has higher CPUE in the beginning of the series (2007) and in later part of the series (after 2019). Relative CPUE from VAST incorporating size data showed relatively smoother trend comparing to VAST with SST and higher CPUE in the latest two years (2021 and 2022). All the standardized CPUE series suggested a decreasing trend from the beginning of the data series to the lowest level in 2011–2012 and a recovering thereafter to the recent year.

Introduction

PBF is an important seasonal target species for the Taiwanese coastal and offshore fisheries. Total catch (including catches from longline fishery and some small coastal fisheries) peaked in 1999 (3,089 mt) and then continuously declined to the lowest level of 214 mt in 2012, less than 10% of the peak catch, and recovered slowly thereafter. After stayed at a low level for a period, the catch recovered to 1,154 mt in 2020 (including 2 mt from south of the equator), and to 1,475 mt and 1,497 mt in 2021 and 2022, similar to the level of 2007 when the stock was relative healthy (Fig. 1).

For monitoring the status of PBF catch from Taiwanese offshore longline fishery, this report provides historical catch and size information as well as relative CPUE series standardized by delta-generalized linear mix model (delta-GLMM) from which the result was adopted for stock assessment purpose in previous PBFWG meetings, and from the vector-auto-regressive spatiotemporal model (VAST; Thorson and Barnett, 2017).

Large scale port sampling program to collect length data of PBF was implemented since 2010, over 95% of PBF landed in Taiwan were measured. Because of the high coverage of size measurements (length in cm) from the catch, size data from the port sampling program was directly used to represent the changes of catch at size.

For CPUE standardizations, two additional data were incorporated in the analyses. The first data of spatiotemporal sea surface temperature (SST) was used as a factor in the standardization models to see the effect of SST variation to the CPUE series. Vessel effect has been tested in the previous reports and SST effect is tested for the first time for this fishery. The second data was size measurement of individual PBF that was reported by fishmen and verified by port inspector, with spatiotemporal information. The size data was converted to age group data before being incorporated into the standardization models, using the von Bertalanffy growth equation (VBGE) of PBF that was developed by Shiao et al. (2016).

Materials and Methods

Catch and effort data (number of fish and fishing days per trip) was the same as that used in Chang et al. (2021, ISC/21/PBFWG-2/02), with additional one year data of 2022. The data was reconstructed and compiled using the approach documented in Chang et al. (2017). This study used three sets of data for performing three standardization models: (1) data of 2003–2022 for traditional delta-GLMM analyses, (2) 2007–2022 for spatiotemporal VAST analyses because the spatiotemporal information was available since 2007, and (3) 2010–2022 for spatiotemporal VAST models incorporating size data because the spatiotemporal size data of individual fish was available since 2010. Three model designs were applied for PBF CPUE standardization: traditional delta-GLMM with SST effect, VAST with SST effect, and VAST incorporating size data. Akaike Information Criterion (AIC) was applied to select the best model.

Traditional delta-GLMM with SST effect

The design of traditional non-spatial model is identical to the one used in ISC/21/PBFWG-2/02: standardizing the catch and effort data using delta-GLMM which separately estimates the proportion of positive PBF catches assuming a binomial error distribution (zero-proportion model) and the mean catch rate of positive catches by assuming a lognormal error distribution (positive-catch model). Akaike and Bayesian information criteria was used to determine the most favorable variable composition of standardization models. Covariates considered in the model included: year (2003–2022), month (May–July), fishing area (the North region and the South region separated by 24.3°N), vessel size (CT1–CT4), and SST. The SST was obtained were obtained from Copernicus Marine Environment Monitoring Service (CMEMS), which provides various forecasting products for marine conditions. Three standardization runs were performed: on the area-combined data, on the data from the South region and the North region. R package LME4 (version 1.1-31) of R (version 4.2.2) was used for GLMM calculations.

Spatiotemporal VAST with SST effect

The R package VAST (Thorson, 2019)(VAST, version 3.10.0) was applied to the data for spatiotemporal analyses. VAST is a delta-generalized linear mixed model that separately estimates the proportion of positive PBF catches and the mean catch rate of positive catches, and in this study, time-invariant spatial variations, time-varying spatiotemporal variations, effects of vessel and SST on catchability.

We modeled the encounter probability (p) for observation *i* using a logit-linked linear predictor:

$$\text{logit}(p_i) = \beta_1(t_i) + L_{\omega_1}\omega_1(s_i) + L_{\varepsilon_1}\varepsilon_1(s_i, t_i) + L_{\delta_1}\delta_1(v_i) + \gamma X(s_i, t_i)$$

and modeled the positive catch rate (λ) for observation i using a log-linked linear predictor:

$$\log(\lambda_i) = \beta_2(t_i) + L_{\omega_2}\omega_2(s_i) + L_2\varepsilon_2(s_i, t_i) + L_{\delta_2}\delta_2(v_i) + \gamma X(s_i, t_i)$$

where $\beta(t_i)$ is the intercept for in year t_i, $\omega(s_i)$ denotes time-invariant spatial variations at location s_i, $\varepsilon(s_i, t_i)$ denotes time-varying spatiotemporal variations at location s_i in year t_i, and $\delta(v_i)$ denotes the effect of vessel v_i on catchability and $\delta_i(v_i) \sim \text{Normal}(0,1), i = 1,2$. L_ω, L_ε and L_δ are the scaling coefficients of the random effect distributions and γ represents the impact of covariate (SST) with value $X(s_i, t_i)$ on density at location s_i in year t_i.

Spatiotemporal VAST incorporating size data

The spatio-temporal model incorporating size data of individual fish can predict density at unsampled locations, times and length classes and provide the relative trends of total abundance various age compositions (Thorson et al., 2017a; Kai et al. ,2017). Due to the limitation of computation capability, the size data was firstly converted to ages of fish using the VBGE of Shiao et al. (2016) [$L_t = 251.2 \times (1 - e^{-0.16(t+2.57)})$] and then grouped into seven age groups (age-bins) per 3 ages: 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, and >23 yrs. Each temporal, spatial, and age group has the encounter probability (p) for observation i:

$$\text{logit}(p_i) = \beta_1(t_i) + L_{\omega_1}\omega_1(s_i) + L_{\tau_1}\tau_1(l_i) + L_{\varepsilon_1}\varepsilon_1(s_i, t_i, l_i) + L_{\delta_1}\delta_1(v_i) + \gamma X(s_i, t_i, l_i)$$

and positive catch rate (λ) for observation i :

$$\log(\lambda_i) = \beta_2(t_i) + L_{\omega_2}\omega_2(s_i) + L_{\tau_2}\tau_2(l_i) + L_{\varepsilon_2}\varepsilon_2(s_i, t_i, l_i) + L_{\delta_2}\delta_2(v_i) + \gamma X(s_i, t_i, l_i)$$

where $\tau(l_i)$ denotes the impact of age group on the expected catch rate l_i, $\varepsilon(s_i, t_i, l_i)$ denotes interaction item of the spatio-temporal and age variation the age l_i at location s_i in year t_i, and $X(s_i, t_i, l_i)$ is a habitat covariates that explain variation in density at location s_i in year t_i for the l_i age. The spatio-temporal-at-age variation, $\varepsilon(s, t, l)$, is modeled by combining the Gaussian random field for spatial variation with a first-order autoregressive processes (AR1) for temporal and for age variation.

Standardized and nominal indices

The observed catch rate (c_i) for each observation is C_i/E_i. The probability function for c_i is

$$\Pr(c_i = c) = \begin{cases} 1 - p_i & \text{if } c = 0 \\ p_i \times \text{Lognormal}(c_i | \log(\lambda_i), \sigma^2) & \text{if } c > 0 \end{cases}$$

where σ^2 is a dispersion parameter.

Calculation the index of abundance of PBF (in year t) for delta-GLMM models:

$$I_{non}(t) = \text{logit}^{-1}(p_i) \times \exp(\lambda_i).$$

For spatio-temporal model with SST effect, the index is predicted using an area-weighted approach:

$$I_{st}(t) = \sum_{s=1}^{n_k} d(s,t)$$

where n_k denotes the number of locations and $d(s,t)$ is the predicted density for location s and year t:

$$d(s,t) = \text{logit}^{-1}(\beta_1(t_i) + L_{\omega_1}\omega_1(s_i) + L_{\varepsilon_1}\varepsilon_1(s_i,t_i)) \times \exp(\beta_2(t_i) + L_{\omega_2}\omega_2(s_i) + L_2\varepsilon_2(s_i,t_i))$$

For spatio-temporal model with additional size data, the index is the sum of abundance of each temporal, spatial, and age groups:

$$I_{sta}(t) = \sum_{k=1}^{n_k}\sum_{l=1}^{n_l} d(k,t,l)$$

where n_k denotes the number of locations, n_l denotes the number of ages, and $d(s,t,l)$ is the predicted density for the age l at location s in year t:

$$d(s,t,l) = \text{logit}^{-1}(\beta_1(t_i) + L_{\omega_1}\omega_1(s_i) + L_{\varepsilon_1}\varepsilon_1(s_i,t_i)) \times \exp(\beta_2(t_i) + L_{\omega_2}\omega_2(s_i) + L_2\varepsilon_2(s_i,t_i)).$$

Then, the average spatial distribution of predicted catch rate for each year is calculated as:

$$I_{sta}(t,l) = \sum_{s=1}^{n_k} d(s,t,l).$$

Essentially, the area-weighted approach computes total abundance as the weighted sum of estimated density across the pre-defined spatial domain of knots, with weights equal to the area associated with each knot. For computational purposes, k-means algorithm was used to cluster all the grid cells into 50 spatial knots and assumed that both the spatial and spatiotemporal random effects for a grid cell are from the closest knot in space.

Nominal CPUE is calculated as:

$$I_{nom}(t) = \frac{1}{n}\sum_{i \in t}\frac{c_i}{f_i}$$

Results and Discussions

Catch trend and size pattern

PBF catch had been as high as 3,089 mt in 1999 but continuously declined to the lowest record of 214 mt (210 mt from longline fishery) in 2012 (Fig. 1). Thereafter, the catch slowly bounced back and stayed at the level of 400–550 mt during 2014–2019. Thereafter, the catch jumped up to 1,154 mt in 2020 and reached 1,475 mt in 2021 and 1497 mt in 2022. Number of registered PBF vessels fluctuated between 480 to 560 in recent ten years.

PBF catches increased substantially in 2020 in both the North and the South regions. The increase of catch was mainly occurred in the South region with an annual proportion of 82% in 2020 but declined to 77% and 67% in 2021 and 2022, respectively. Comparing nominal CPUE of the vessels that have fished for PBF, 37% of them shows higher CPUE in 2022 than in 2021 (Fig. 2), less than the percentage of 57% for 2021 vs. 2020. This was reflected in the following nominal CPUE calculation but was different from the trend of standardized CPUE.

Average size of PBF caught by the Taiwanese longline fishery was around 212–220 cm before 2008 (Fig. 3). Thereafter, the average in the North region stably maintained at 218–224 cm during 2008–2022; while in the South region, the average size gradually increased to 235 cm in 2012 and declined to 210 cm and 209 cm in 2020 and 2021, respectively, showing a different trend from the North region. Average size in 2022 was 210 cm in the South region. The substantial increase of average size in the South region was considered resulting from the decline of recruitment to the fishing ground; and the decrease since 2013 was a response to more smaller fish recruited to the fishing ground and more large fish removed from the fishing ground (Fig. 4). Fig. 4 suggested that Taiwan fishery utilized mainly one or two cohorts of the stock.

Relative CPUE from delta-GLMM

All models with vessel effect (CT) have smaller AIC and so these models were selected to compare the effect of SST. Table 1 shows the best variable combinations of the delta-GLMM and their AIC/BIC for zero-proportion model (ZPM) and positive-catch model (PCM), by region. Models with SST factor for North and Whole regions have smaller AIC comparing to without SST factor, indicating models with SST factor have better statistical performance. However, for the South Area, including SST factor in the models resulted in slightly larger AIC, and the estimated CPUE series are almost identical (Fig. 5). This may be because the areal discrepancies of SST in the short

fishing season (April to June) was not substantial off Taiwan (annual standard deviation 0.8–1.5˚C), especially in the South region.

The GLMM runs fitted the data well (based on the qq-plots and the residual histograms, Fig. 6) for all the regions. From the AIC and BIC analyses, separate standardization by fishing grounds has better performance than the one combined both fishing grounds (Table 1); area-separate models were thus considered preferable because the size composition of fish of the two fishing grounds apparently different. The index of the South region was considered relatively better representing the PBF abundance index than the northern one considering its features of better data stability and with much higher proportion of historical catches; and was recommended to be used for the PBF stock assessment by the PBFWG since 2016.

Relative CPUE from VAST with SST effect

All the spatiotemporal models using VAST have successfully converged, which were confirmed by the fact that the Hessian matrix was positive definite, and the maximum gradient component was smaller than 0.001. Moreover, quantile diagnostics suggested the spatiotemporal model fitted the catch and effort data well (Fig. 7).

Distribution of fishing effort in the core area (fishing days, Fig. 8) showed two major fishing grounds. Pronounced spatiotemporal variations in density were predicted for the period of 2007–2022 (Fig. 9): the densities decreased from the starting year of the study (2007) to the lowest level in 2011 and 2012, and then started to increase gradually toward the end year of the study, while the increase in 2020–2022 apparently was substantial. The density pattern in the South region was similar to (for 2020), or higher than (for 2021 and 2022), that of 2007, with eastward expansion of high-density area, which suggested a recovery of the fish stock from 2011–2012.

Additional runs for data from 2007 to 2021, 2020, 2019, 2018 and 2017 were performed to see the sensitivity of the standardization model to the data series. The estimates of the standardized CPUE from each of the retrospective runs show consistent patterns over time for all regions (Fig. 10).

Relative CPUE from VAST incorporating size data

Density (upper box) and relative CPUE (bottom box) of PBF of the South region (S) and the North region (N) for 2010 - 2022 by the seven age groups were shown in Fig. 11. South region has much higher density than the North, but the general trends are similar for both regions except for the latest one or two years (increasing in the North region comparing to decreasing in the South region) for age groups older than 15 years old.

Age group 9-11 was the most dominant fish, and then the 6-8 age group, for recent years in the fishing ground. During the low stock status in 2010-2012, density of age groups of 6-8 and 9-11 were very low for both regions, started to increase slowly after around 2014, and then jumped up in 2020 when age groups >18-year-old were all at a

decreasing low level. The year that density of age group 12-14 started to increase was later than younger fish groups about 3 years and was also jumped up in 2020.

For age groups older than 15-year-old, all showed obvious declining trends since 2014 or 2015. Or, for age groups 18-20, 21-23 and >23 (large-matured fish), the density showed a concave up shape with the peak at around 2014 to 2016. Reasons for this observation is unknown.

Comparison of relative CPUE from all models

Relative CPUE series from GLMM and VAST with SST were compared in Fig. 12. Trend of the two series for the most important South region is similar with VAST series has higher CPUE in the beginning of the series (2007) and in later part of the series (after 2019). Relative CPUE from VAST incorporating size data (Fig. 13) showed relatively smoother trend comparing to VAST with SST and higher CPUE in the latest two years (2021 and 2022).

All the CPUE series suggested a decreasing trend from the beginning of the data series to the lowest level in 2011–2012 and a recovering thereafter to the recent year.

References

Shiao, J.-C., Lu, H.-B., Hsu, J., Wang, H.-Y., Chang, S.-K., Huang, M.-Y., Ishihara, T., 2016. Changes in size, age, and sex ratio composition of Pacific bluefin tuna (*Thunnus orientalis*) on the northwestern Pacific Ocean spawning grounds. ICES J. Mar. Sci. 74, 204–214. https://doi.org/10.1093/icesjms/fsw142

Thorson, J.T., 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fish. Res. 210, 143–161. https://doi.org/10.1016/j.fishres.2018.10.013

Table 1. Best variable combinations of the delta-lognormal mixed models for GLMM, and the Akaike information criterion (AIC) and Bayesian information criterion (BIC).(ZPM: zero-proportion model; PCM: positive-catch model).

Model type	Final Model formulation	AIC	BIC
South region			
ZPM	Year+Month+CT+ **sst** +Year*Month	19121.43	19301.25
PCM	Year+Month+CT+ **sst** +Year*Month	19917.03	20083.64
South region			
ZPM	Year+Month+CT+Year*Month	19120.19	19292.2
PCM	Year+Month+CT+Year*Month	19910.28	20069.95
North region			
ZPM	Year+Month+CT+ **sst** +Year*Month	8147.403	8299.862
PCM	Year+Month+CT+ **sst** +Year*Month	7740.698	7880.572
North region			
ZPM	Year+Month+CT+Year*Month	8188.803	8334.332
PCM	Year+Month+CT+Year*Month	7781.703	7915.496
Combined South and North regions			
ZPM	Year+Month+CT+Area+ **sst** +Year*Month	27592.54	27788.44
PCM	Year+Month+CT+Area+ **sst** +Year*Month	27830.72	28013.1
Combined South and North regions			
ZPM	Year +Month+CT+Area+Year*Month	27610.57	27798.32
PCM	Year+Month+CT+Area+Year*Month	27837.75	28012.83

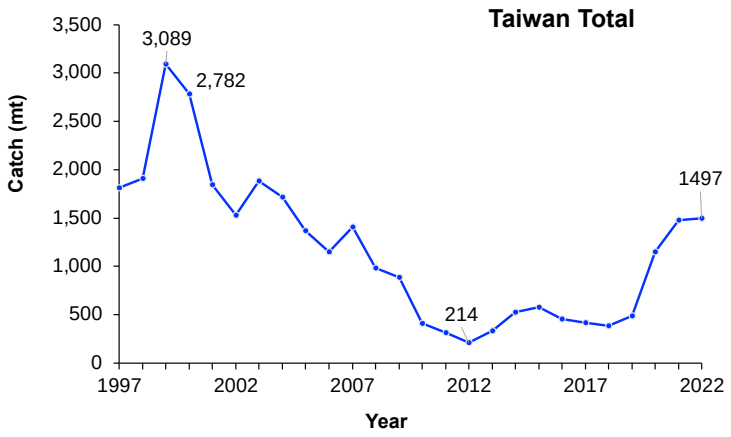

Fig. 1. Annual PBF catches by Taiwanese offshore fishery (mainly longline fishery with minor catches from other small coastal fisheries).

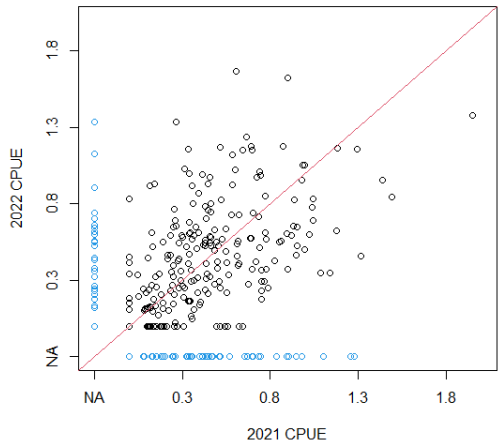

Fig. 2. Nominal CPUE of the vessels that have fished for PBF in 2021 and 2022 (black circles). The blue circles indicate the CPUE of vessels have either fishing in 2021 or 2022 only.

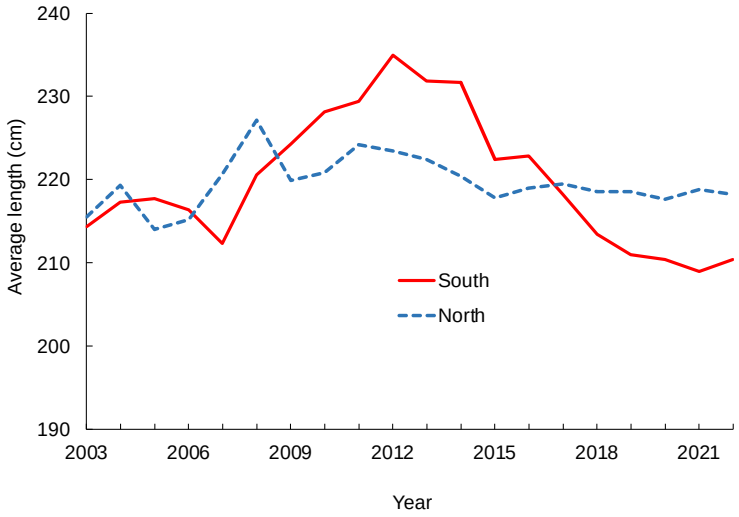

Fig. 3. Annual trend of average length of PBF of Taiwanese longline fishery, by region.

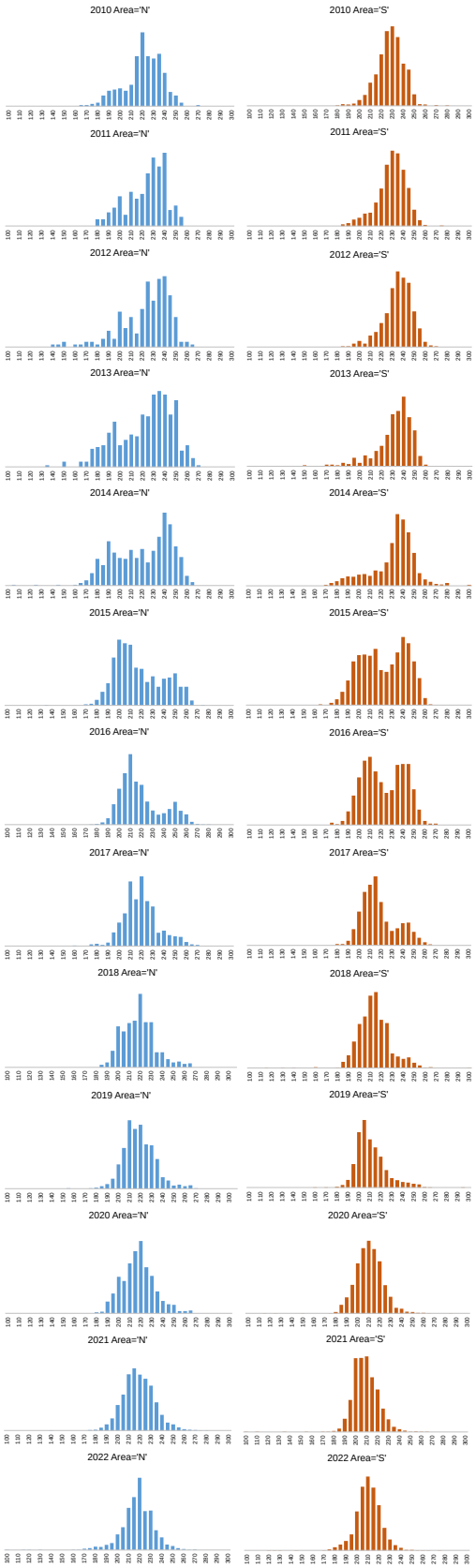

Fig. 4. Length frequencies of Taiwanese PBF during 2010 – 2022 for North region (left in blue) and South region (right in red).

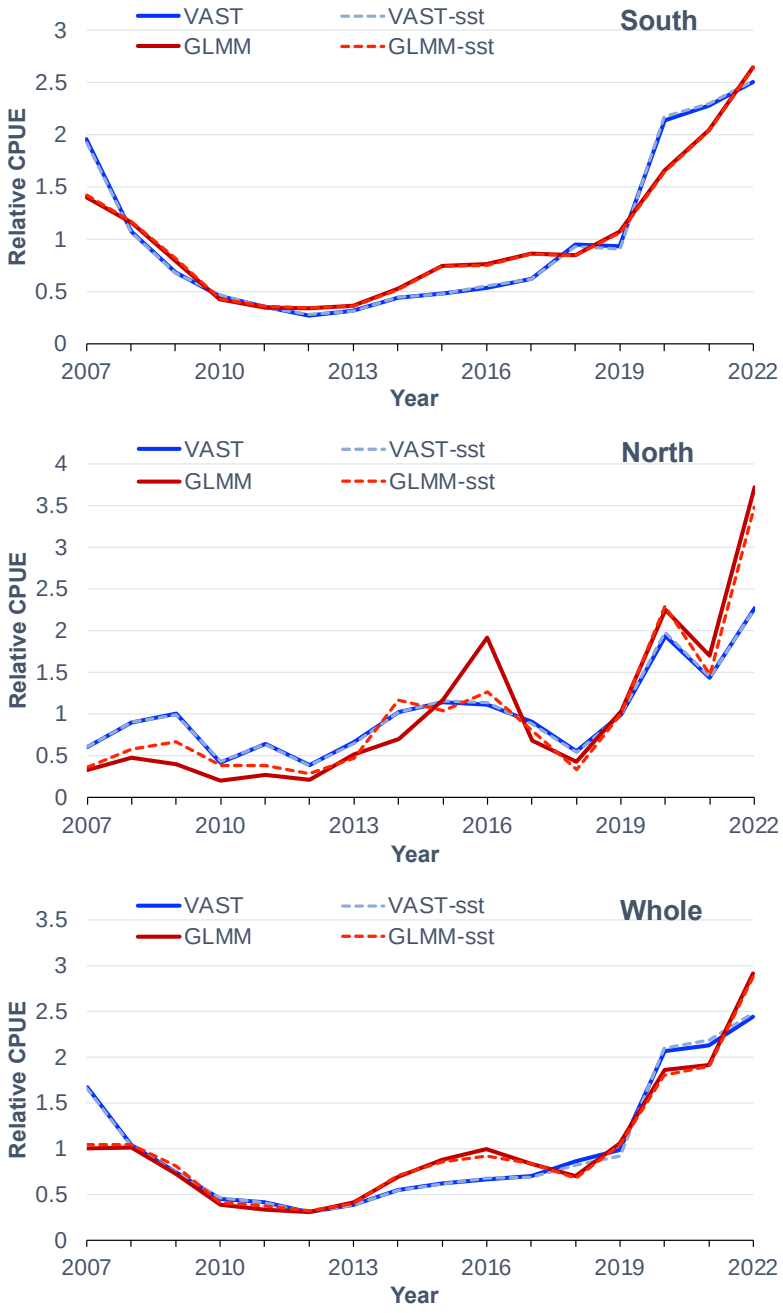

Fig. 5. Relative CPUE series obtained from GLMM and VAST with or without consideration of SST in the standardization models, by region.

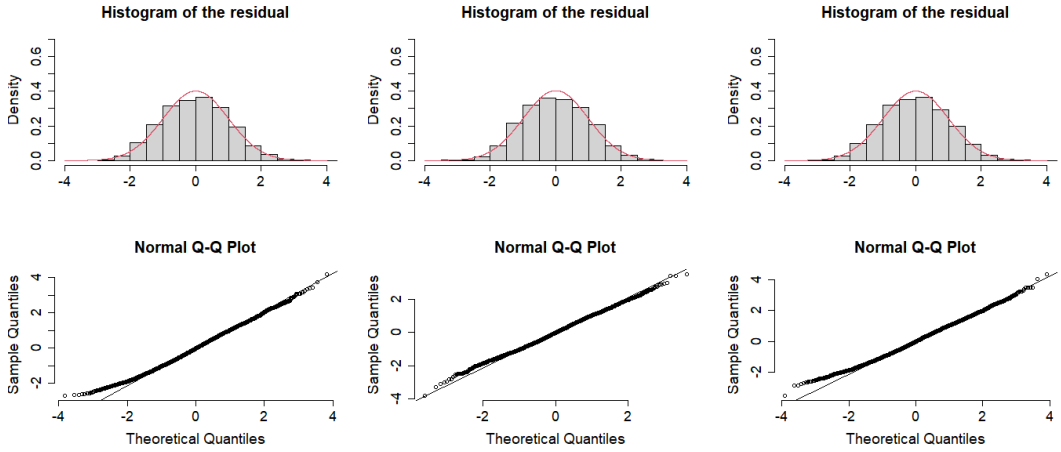

Fig. 6. Diagnostic residual plots (the posterior-predictive residual histogram and qq plot comparing the observed and predicted quantiles of CPUE given encounter) for the traditional delta-GLMM analyses. Panels from left to right are for the South and North regions and the whole fishing ground.

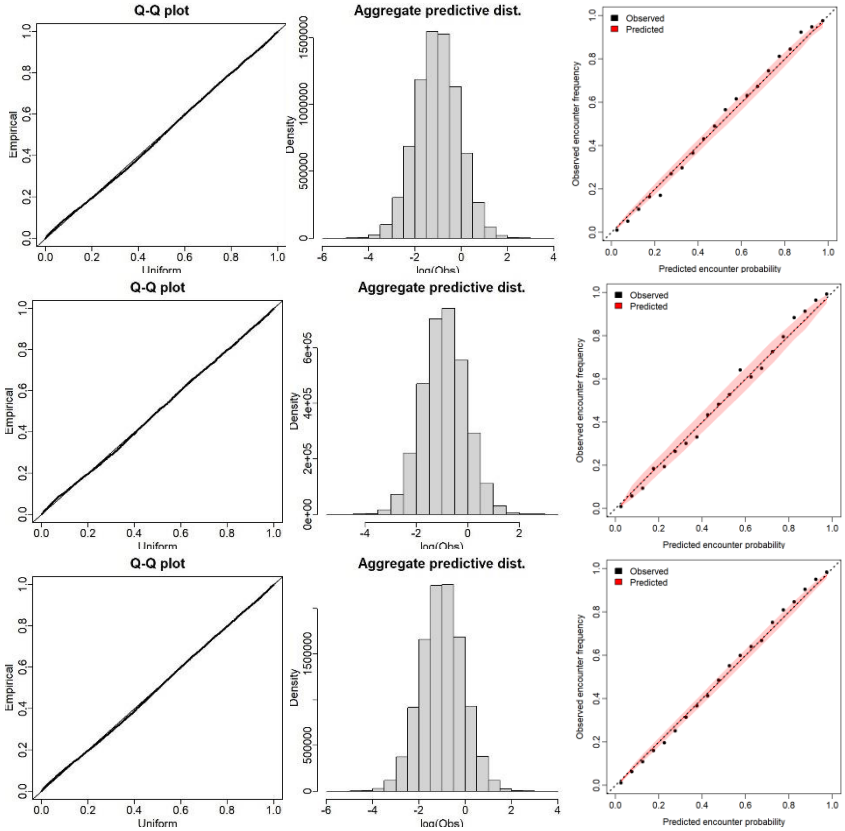

Fig. 7. Diagnostic residual plots for VAST analyses on, from top to bottom, the South, the North and the Whole regions, respectively). The graphs from left to right: the qq plot, the posterior-predictive residual histogram, and encounter probability.

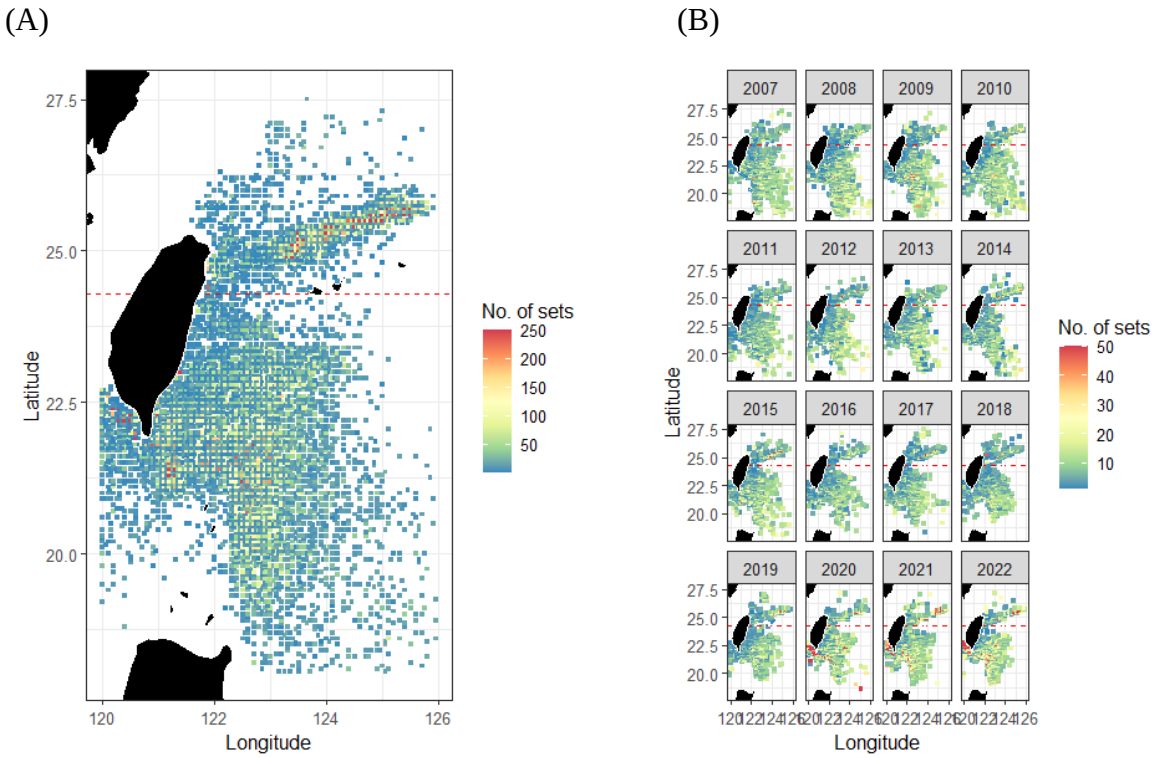

Fig. 8. Geographic distribution of fishing days per 0.1º×0.1º grid cell during 2007–2022: (A) whole period; (B) by year.

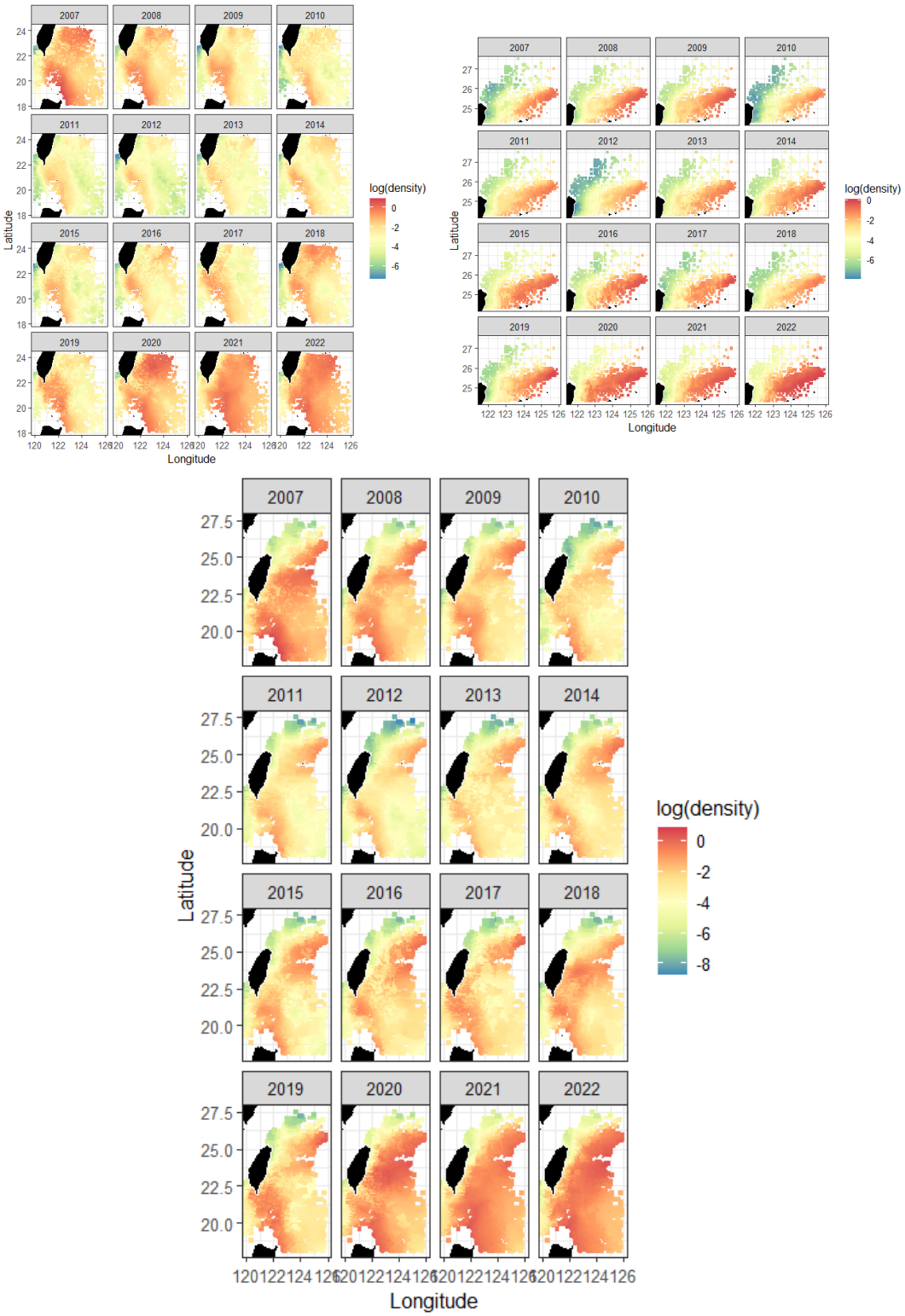

Fig. 9. Spatiotemporal distribution of predicted log density of PBF during 2007–2022 from VAST analyses (upper left – South, upper right – North, bottom: South and North combined).

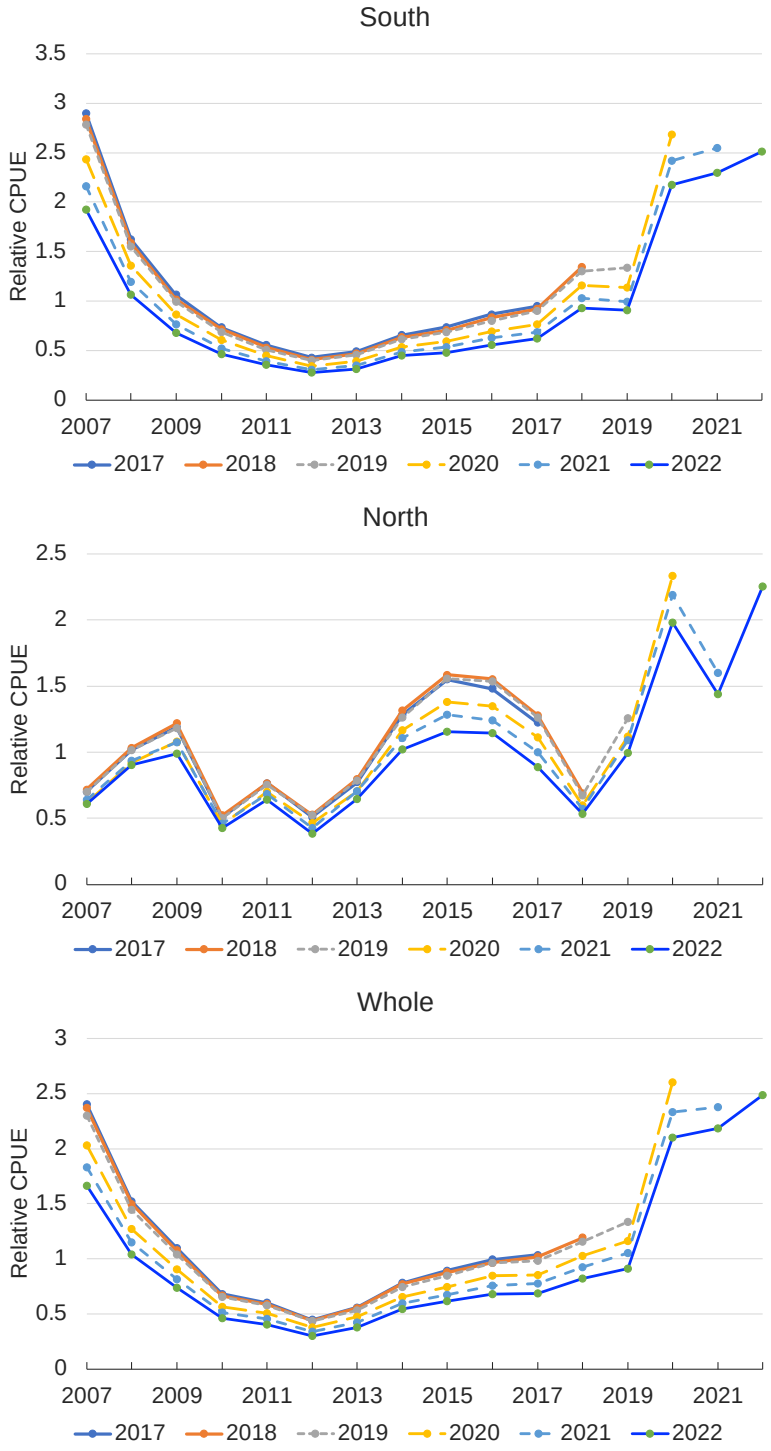

Fig 10. Retrospective CPUE standardizations from 2017 to 2022 using VAST with consideration of SST.

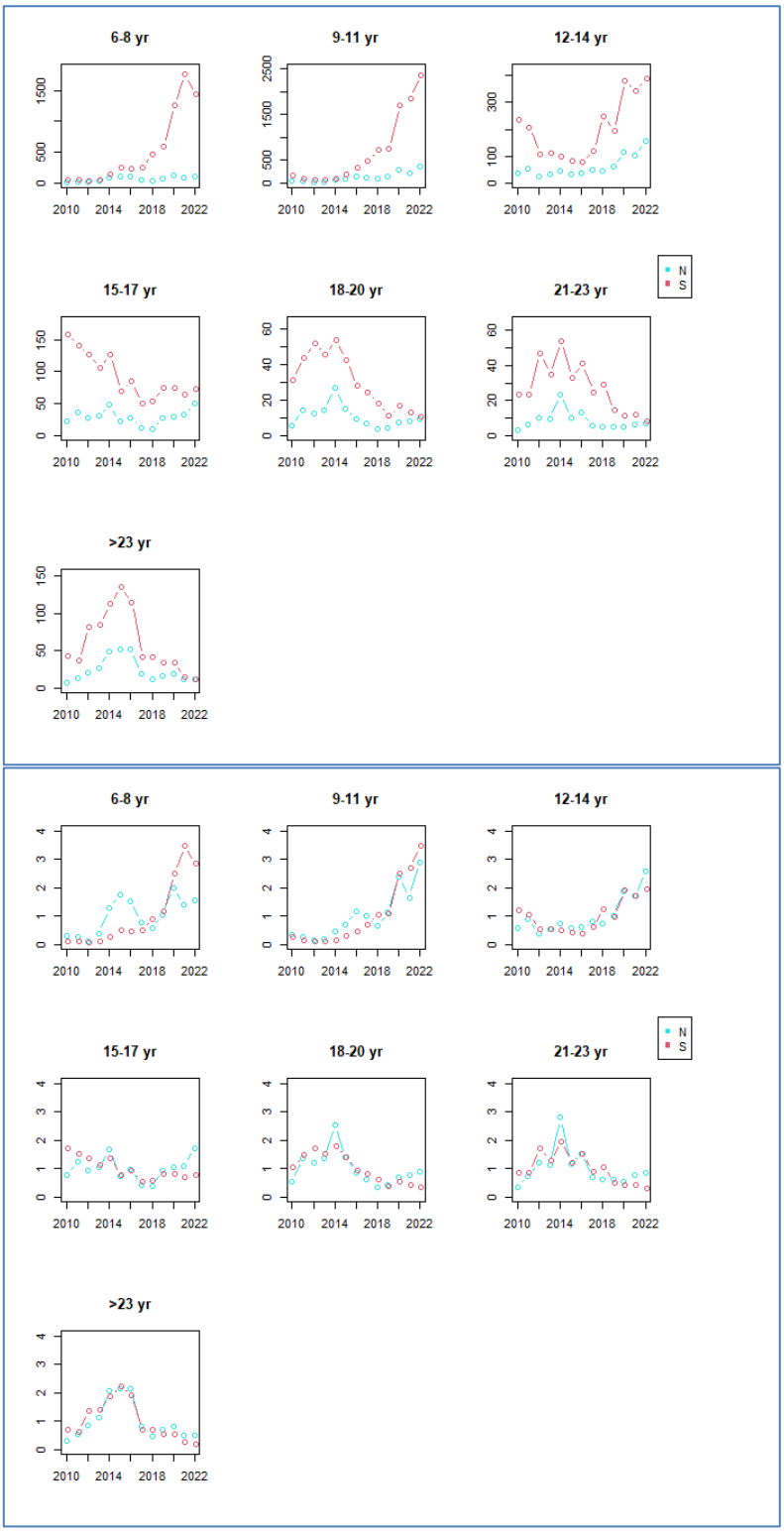

Fig. 11. Density (upper box, scale is different by age group) and relative CPUE (bottom box) of PBF of the South region (S) and the North region (N) for 2010 - 2022, by age group.

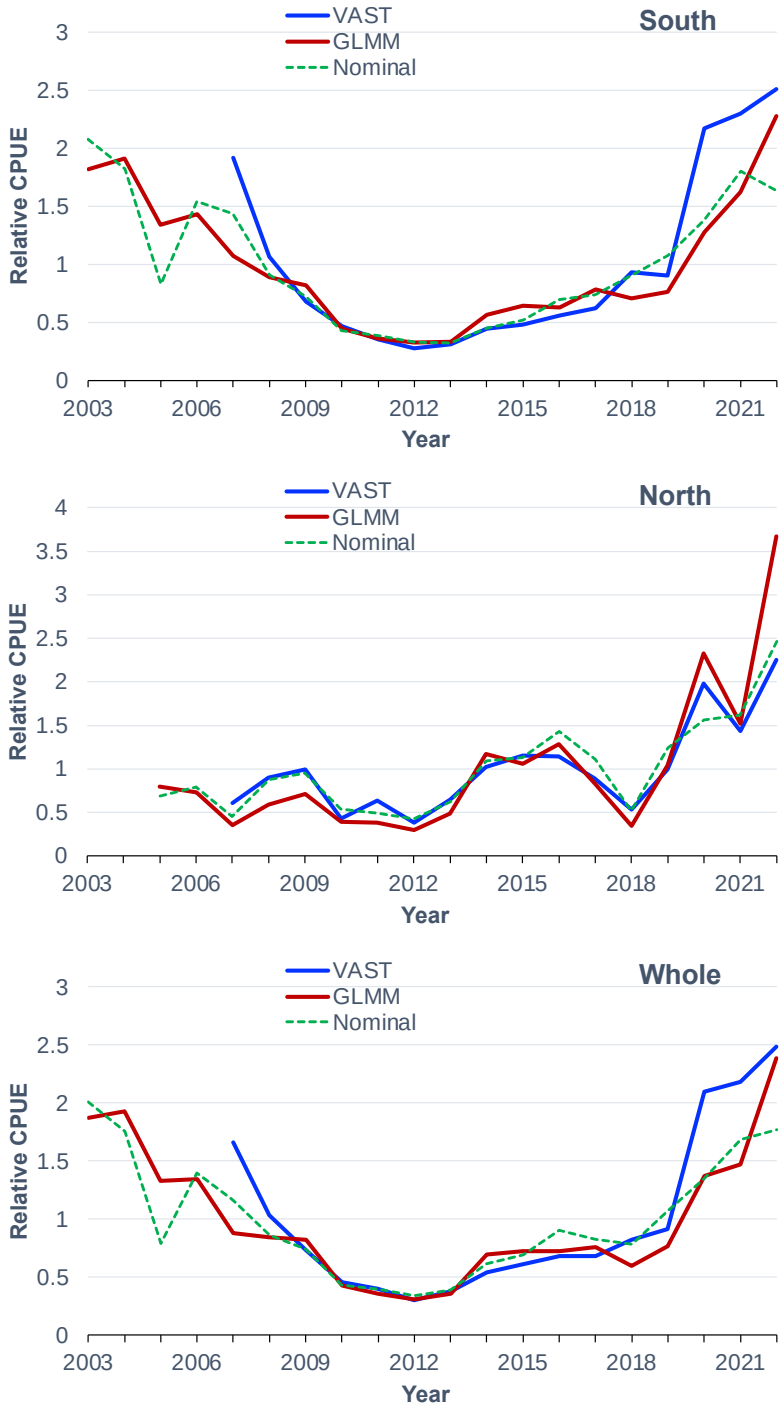

Fig. 12. Nominal CPUE and standardized CPUE series based on delta-GLMM and VAST analyses (with SST factor), by region.

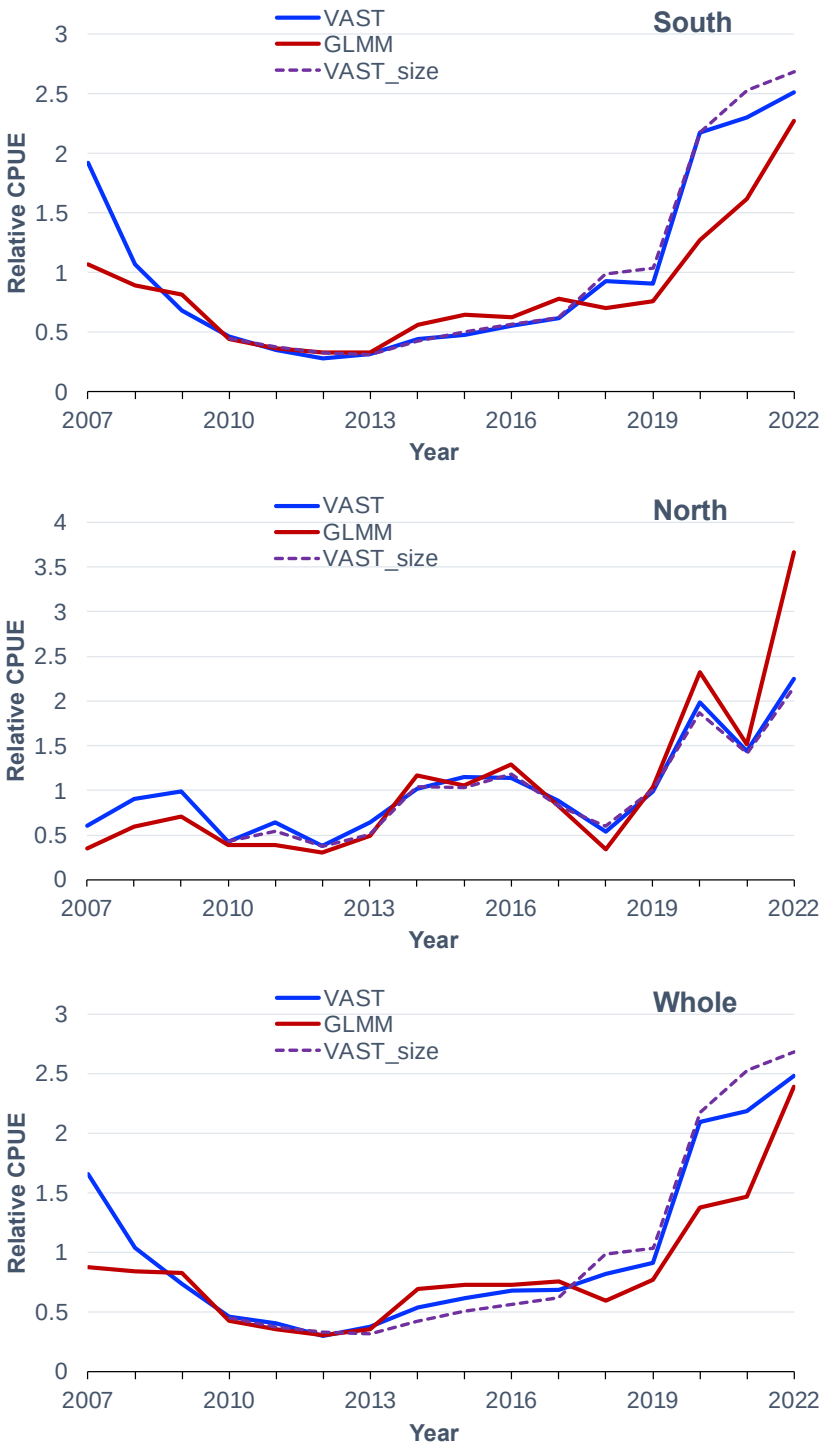

Fig 13. Relative CPUE for PBF estimated by GLMM, VAST (with SST factor), and VAST_size (VAST incorporated size information), by region.