

**CPUE STANDARDIZATION FOR PACIFIC SWORDFISH (*Xiphias gladius*) CAUGHT BY THE JAPANESE LONGLINE FISHERY:
A GLMM ANALYSIS USING THE R SOFTWARE PACKAGE R-INLA**

¹Marko Jusup, ²Haruko Koike, ¹Hirotaka Ijima

¹Highly Migratory Resources Division, Fisheries Stock Assessment Center, Fisheries Resources Institute, Japan Fisheries Research and Education Agency, 2-12-4 Fukuura, Kanazawa-ku, Yokohama 236-8648, Kanagawa, Japan.

²Fishery Solution

Email: jusup_marko00@fra.go.jp



Abstract

We analyzed the logbooks recorded by the Japanese longline vessels to obtain the abundance index required for the Western and Central North Pacific Ocean (WCNPO) swordfish stock assessment. Considering the transition of Japanese longline fishing gear and the change of the logbook format, we separated the logbook into two time series (1976 to 1993 and 1994 to 2021). Using the R-INLA package, we constructed multiple GLMMs, including the spatiotemporal model with and without gear effect. We selected the best model using the information criteria WAIC and LOOCV. Model selection preferred spatiotemporal models without gear effect for both time series. Upon obtaining the spatiotemporally resolved standardized CPUE, we calculated the averages for each management area.

Introduction

Japanese longline fleets have been operating in the Pacific Ocean since the 1950s, and these operations have been logged for more than 60 years. Fishing ground and fishing gears have changed over these years, causing spatiotemporal bias. Catchability of the longline fishery might have changed as well (Miyabe, N. and Nakano 2004). Thus, it is necessary to correct for these biases and remove any historical catchability change through the process of CPUE standardization.

The catch size of highly migratory fishes such as swordfish is likely dependent on fishing grounds (Ijima and Kanaiwa 2018), indicating an age-specific migration. It is therefore beneficial to consider age-specific CPUE for the stock synthesis model that assumes age-structured population dynamics for swordfish. In the previous assessment, subareas were defined based on the catch size, and area-specific CPUE was calculated for CPUE standardization (Kanaiwa and Ijima 2018). Although this analysis successfully accounted for the age-specific migration, it did not improve the issue of the data bias caused by the change of fishing ground and gear settings. It has recently been reported that such problems can be better ameliorated using geostatistical models (Ijima and Koike 2020, Ijima and Koike 2021, Ijima et al., 2022, Ijima 2022).

We first examined the historical change in Japanese longline operations in this study. We then constructed multiple GLMMs, including geostatistical models, and selected the best model based on information criterion values. We finally calculated the standardized CPUE using the estimated parameters.

Materials and Methods

Japanese longline operational data

Japanese longline operational data have been available since 1952. However, the available information has changed over the years. For example, the hooks between floats were not recorded until 1975, and vessel names were also not organized until the mid-1970s. In addition, the data format was changed in 1993, which may have altered the accuracy of the data compared to previous years. The fields regarding details on the gear configuration and catch in weight have been added from the year 1994. One notable change in the longline fishery is that the material of the lines was changed from hemp to nylon in the mid-1990s, and the method of operation may have changed to maintain bait depth. However, no such information was recorded in the logbook. In order to maintain consistency in the data format and quality, this study splits the data into two-time series groups, “early” (1976-1993) and “late” (1994-2021), to standardize the CPUE. The data consistency is considered to have been maintained during these two time periods.

Data preparation

We used the data from WCPNO Area 1 and Area 2, defined by Ijima and Kanaiwa 2018 (Figure 1). We kept this old delineation to allow comparison with our previous study. However, we removed a small section between 150°W to 135°W and 5°N to 10°N from our study area to reflect the new delineation of the WCNPO defined in ISC 22.

We extracted the fishing data for these two areas and filtered for vessels above 20 tons and operations with 3 to 21 hooks between floats. Although the area and periods were divided, the operational data was still vast, and the cost of time was high to perform the analysis with a spatiotemporal model. Therefore, operational data were aggregated by year, month, $1^\circ \times 1^\circ$ grid, vessel name, and hooks between floats (hereafter HBF) to reduce the computation time.

Fishing Effort and nominal CPUE

Japanese longline fishing ground seems to be concentrated around three areas, one of which is located east of Japan, another along the equatorial line, and the last one around Hawaii (Figure 2). The fishing area seems to have reached maximum coverage during the 1980s and 1990s, but it started shrinking from the 2000s. By 2010, the Hawaii fishing ground has disappeared, and the current fishery is operating in a minimal area, not reaching the eastern side of the Pacific due to

COVID-19 effects.

Although fishing effort has been decreasing, especially in the EPO area, the nominal CPUE has increased in this area (Figure 3). Additionally, for Area 1, nominal CPUE used to be high around the latitude line between 20°N and 30°N, east of 140°E, but since 2000 high CPUE area in the north around 40°N has appeared.

Gear setting

A high number of HBFs is generally thought to allow the bait to sink deeper and affect catchability between species distributed at different depths (Ward and Hindmarsh 2007). We see a drastic change in HBF around 1994 (Figure 4). Furthermore, we see that the number of branch lines has increased since 1994, which may also have the purpose of sinking the lighter line deeper. However, it is important to note other gear changes when using the HBF as a proxy for depth because the material change in line could significantly alter the number of HBF.

Japanese long-liners also change their gear settings depending on the locations. For instance, fishermen fishing in the northern part of area 1 use fewer HBFs even after the line material change (Figure 4). Fishermen also choose shorter branch line and buoy line in area 1 than area 2 (Figure 5).

Considering all of the above it is safe to say that fishing gears/settings have changed over the years and are heavily influenced by fishing locations and other species' catch. We therefore took caution when we used a simple HBFs proxy for the fishing gears effect.

Spatial pattern of fish size (depicted in mean-body weight)

Along with the spatial dependence on gear, we also noticed that the change in catch size in latitude revealed that very large fish were caught in Area 1 (Figure 6). In addition, a mix of large and small fish (including juvenile fish) were caught in the mid-Pacific area in Area 2 (Figure 6).

Statistical models

We used the R software package R-INLA (Lindgren and Rue 2015) for the CPUE standardization. We constructed several generalized linear mixed models (GLMMs) considering the spatiotemporal effect and examined the combination of multiple fixed effects to construct the optimal model (Table 1). We tested the theory of gear setting simply reflecting fishing location differences by comparing models with and without HBF. Two distributions (zero-inflated Poisson; ZIP and negative binomial;

NB) were considered for the response variable, which was the number of fish caught with offset terms of the number of hooks (1,000 hooks). The explanatory variables were year, quarter, fishing location, vessel name, and HBF. The effect of year and quarter were set as fixed effects of categorical variables. The vessel name and HBF were set as random effects. The spatial effects were either considered as correlations by location (stochastic partial differential equation approach; SPDE) or defined as a spatiotemporal model that considered correlations by location and time (stochastic partial differential equation approach with ar1; spde_time). The spatiotemporal GLMM in negative binomial distribution could be written as:

$$Catch_i \sim NB(r, p)$$

$$E(Catch_i) = \mu_i = r(1 - p)/p$$

$$\text{Hyperparameter (size)} = r = \mu(p/(1 - p))$$

$$\log(\mu_i) = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{A}_i \mathbf{u} + \log(1000 * hooks_i),$$

where $Catch_i$ is a number of swordfish caught by fishing set i , μ_i is the mean value of swordfish caught by set i , r is the size and p is the probability of success for negative binomial distribution and can be rewritten as:

$$r = \frac{\mu^2}{\sigma^2 - \mu} \text{ and } p = \frac{\mu}{\sigma^2}.$$

The same model with zero-inflated Poisson distribution could be written as:

$$Catch_i \sim ZIP(\mu_i, \pi p)$$

$$E(Catch_i) = \mu_i(1 - \pi)$$

$$\log(\mu_i) = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{A}_i \mathbf{u} + \log(1000 * hooks_i).$$

where $Catch_i$ is number of swordfish caught by fishing set i and π is the probability of zero catch.

In either model, \mathbf{X}_i is the covariate matrix row in the set i , $\boldsymbol{\beta}$ are fixed effect coefficients vector for each covariates, \mathbf{Z}_i is the model matrix row for the random effects in set i and $\boldsymbol{\delta}_i$ is the random effect coefficients vector that is $\boldsymbol{\delta}_i \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Psi})$, where

Ψ is the covariance matrix that depends on the number of random effects variables. \mathbf{A}_i is a projector matrix row in the set i , and \mathbf{u} is a spatial effect.

The elements of spatial effect at location s in year t ($u_{s,t}$) follows the AR1 process $u_{s,t} = \rho u_{s,t-1} + v_{s,t}$, where ρ is an auto-correlation parameter and $v_{s,t}$ is a spatial Gaussian random field $v_{s,t} \sim GMRF(0, \Sigma)$. The element of covariance is $\Sigma_{j,k} = \sigma_v^2 Cor_M(V(s_j), v(s_k))$ and correlation function is the Matérn correlation function

$$Cor_M(V(s_j), v(s_k)) = \frac{2^{1-v}}{\Gamma(v)} (\kappa \| s_j - s_k \|)^v K_v(\| s_j - s_k \|).$$

R-INLA estimates the posterior of parameters using Bayesian inference. The prior coefficient vector for each covariate (β) was set as a default value of the INLA package's Guassian prior. We used the half-Cauchy distribution truncated at zero to serve as a prior for the standard deviation sigma for random effect coefficients. Penalized Complexity (PC) priors were used for the spatial effects parameters, auto-correlation parameter, and size parameter of negative binomial. The practical range and marginal deviation were used to check for spatial effects parameters (Krainski et al 2018, Fuglstad et al. 2019).

Model selection

One of the benefits of using R-INLA is that it can calculate Widely Applicable Information Criterion (WAIC) and Leave One Out Cross Validation (LOOCV) since it uses the Bayesian inference technique (Watanabe 2010, Vehtari et al., 2017). These two information criterions are known to be more suitable for comparing complex models (e.g., random effect models and hierarchical models) since normal AIC cannot accurately reflect the model complexity (Watanabe 2010). In this study, model selection was performed using WAIC and LOOCV, and the value of over-dispersion was also confirmed.

Standardized CPUE

The standardized CPUE was calculated using the estimated fixed effect parameters and the potential spatial field using INLA's prediction function. CPUEs were estimated for each mesh node within the study area. Because each node also corresponds to a generating point in the Voronoi diagram, the indices for each area and period were area-weighted evenly and allowed us to avoid giving locations with

more sets a disproportionate weight in the final CPUE indices.

Results and Discussion

Results of model selection

As a result of model selection using information criterion and visual model diagnostics, spatiotemporal models without gear effect (HBF) were selected for both time series and areas (Table 2). The negative binomial models were chosen for the final model because 1) information criteria scores were better for all models with NB distribution, 2) randomized quantile residual plots showed a clustered pattern for the ZIP model, and 3) most ZIP models could not calculate LOOCV that indicate some convergence issues. Spatiotemporal models showed better information criterion scores for all time periods and areas compared to simple spatial models. Between the two spatiotemporal models (with and without HBF), both WAIC value and LOOCV value were slightly better for the spatiotemporal model with gear effect (HBF). However, the difference was very small (only 1.5% score change) which is likely due to spatial patterns explaining many HBF effects. Furthermore, the estimated random effect parameter of HBF was of enormous value, making us wonder about the possibility of this parameter absorbing the potential year effects. Therefore, in the end, we chose to use the spatiotemporal model without gear effects (HBF).

Model output

All models regardless of the time period and area, year effect was close to zero for most years and did not function well as a parameter (Figure 7 ~ 10). Other fixed effects and hyperparameters were unlikely to be zero based on the posterior distribution and were well estimated.

Standardized CPUE showed lower value estimated for both time periods of Area 1, and similar values for Area 2 (Figure 11). Vessel effect was the only effect removed from models for standardization, thus we examined the CPUE prediction when vessel effect was included (Figure 12). This showed that vessel effect was in fact the reason nominal CPUE is high compared to the standardized CPUE. This indicates a possibility that certain vessels with high CPUE are likely over-representing the trend when without standardization.

References

- Fuglstad, G.A., Simpson, D., Lindgren, F. and Rue, H., 2019. Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, 114(525), pp.445-452.
- Iijima, H., 2022. CPUE STANDARDIZATION FOR ATLANTIC SWORDFISH CAUGHT BY JAPANESE LONGLINE FISHERY: THE GLMM ANALISIS USING R SOFTWARE PACKAGE R-INLA. *Collect. Vol. Sci. Pap. ICCAT*, 79(2), pp.156-179.
- Iijima, H. and Kanaiwa, M. 2018. *Pattern recognition of population dynamics for North Pacific swordfish (*Xiphias gladius*): the operational data analysis of Japanese longline fishery using the finite mixture model*. ISC/18/BILLWG- 01/9, pp.12.
- Iijima, H. and Koike, H. 2020. *Preliminary analysis for the CPUE standardization of the Pacific blue marlin using Japanese longline logbook and the R software package R-INLA*. ISC/20/BILLWG-03/02, pp.10.
- Iijima, H. and Koike, H. 2021. *CPUE Standardization for Striped Marlin (*Kajikia audax*) using Spatio-Temporal Model using INLA*. ISC/21/BILLWG-03/01, pp.14.
- Iijima, H., Minte-Vera, C., Xu, H., and Maunder, M., 2022. The Japanese logbook analysis for the South Eastern swordfish abundance indices using R-INLA. INTER-AMERICAN TROPICAL TUNA COMMISSION SCIENTIFIC ADVISORY COMMITTEE. 13TH MEETING.
- Kanaiwa, M. and Iijima, H. 2018. *Abundance indices of Swordfish (*Xiphias gladius*) by the Japanese offshore and distant- water longline fishery in the North-Western Central Pacific*. ISC/18/BILWG-1/7, pp.44.
- Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F. and Rue, H., 2018. *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall/CRC.
- Lindgren, F. and Rue, H., 2015. Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19), pp.1-25.
- Miyabe, N. and Nakano, H., 2004. Historical trends of tuna catches in the world (No. 467). Food & Agriculture Org.

Vehtari, A., Gelman, A. and Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, 27(5), pp.1413-1432.

Ward, P. and Hindmarsh, S., 2007. An overview of historical changes in the fishing gear and practices of pelagic longliners, with particular reference to Japan's Pacific fleet. *Reviews in Fish Biology and Fisheries*, 17(4), pp.501-516.

Watanabe, S., 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12).

Table 1. Statistical model list for CPUE standardization considered for all time-period and area combinations (Area1×early, Area1×late, Area2×early, Area2×late)

No	Model	INLA function
001	simple spatial model + gear (hbf)	swo ~ -1 + intercept + yr + qtr + f(hbf, model="iid", hyper=hcprior) + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde)
002	spatio-temporal model + gear (hbf)	swo ~ -1 + intercept + yr + qtr + f(hbf, model="iid", hyper=hcprior) + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1"))
003	spatio-temporal model	swo ~ -1 + intercept + yr + qtr + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1"))

Table 2. WAIC and LOOCV values for each model

Early data (1976 – 1993).

Model	Distribution		Area1	Area2
001 Simple spatial model + gear (hbf)	NB	waic	725,465.3	533,806.1
		loocv	725,920.4	535,030.4
	ZIP	waic	1,071,908.9	775,959.3
		loocv	NA	NA
002 Spatio-temporal model + gear (hbf)	NB	waic	714,714.8	
		loocv	715,205.7	
	ZIP	waic	999,784.0	
		loocv	NA	
003 Spatio-temporal model	NB	waic	721,592.3	525,215.8
		loocv	722,175.2	526,389.8
	ZIP	waic	1,035,782.6	723,221.2
		loocv	NA	NA

Late data (1994 – 2021).

Model	Distribution		Area1	Area2
001 Simple spatial model + gear (hbf)	NB	waic	473,988.6	245,184.7
		loocv	474,339.7	245,196.4
	ZIP	waic	681,307.0	266,868.9
		loocv	NA	NA
002 Spatio-temporal model + gear (hbf)	NB	waic	464,577.0	
		loocv	464,990.1	
	ZIP	waic	614,289.9	
		loocv	NA	
003 Spatio-temporal model	NB	waic	471,878.5	241,613.7
		loocv	472,021.9	241,653.4
	ZIP	waic	633,743.9	258,887.5
		loocv	NA	NA

Table 3. Nominal CPUE (n CPUE), standardized CPUE (std CPUE), and standard deviation for each region (early)

year	Area 1			Area2		
	nCPUE	stdCPUE	SD	nCPUE	stdCPUE	SD
1976	0.9651	0.5295	0.1014	0.2784	0.3309	0.0306
1977	0.9752	0.4540	0.0795	0.2045	0.3146	0.0396
1978	0.8401	0.4663	0.0839	0.2189	0.2614	0.0233
1979	0.8937	0.4116	0.0779	0.2185	0.2512	0.0246
1980	0.6511	0.2787	0.0448	0.3057	0.4566	0.0506
1981	0.6709	0.2733	0.0376	0.3181	0.2577	0.0166
1982	0.7351	0.4668	0.0917	0.2341	0.2856	0.0263
1983	0.9974	0.4732	0.0788	0.2905	0.3411	0.0324
1984	0.8698	0.3728	0.0664	0.3122	0.3571	0.0308
1985	1.3197	0.4698	0.0908	0.3423	0.3487	0.0346
1986	1.3003	0.5425	0.1156	0.2730	0.3595	0.0362
1987	1.2963	0.6133	0.1337	0.3477	0.4342	0.0417
1988	1.0893	0.4867	0.1136	0.3801	0.4101	0.0439
1989	0.9030	0.4334	0.0919	0.3050	0.3404	0.0345
1990	0.9744	0.3705	0.0709	0.2694	0.3374	0.0308
1991	0.7888	0.4243	0.0952	0.2118	0.2756	0.0320
1992	1.0562	0.4546	0.1163	0.1613	0.2564	0.0316
1993	1.0709	0.4849	0.1117	0.1388	0.2266	0.0291

Table 3. Nominal CPUE (n CPUE), standardized CPUE (std CPUE), and standard deviation for each region (late)

year	Area 1			Area2		
	nCPUE	stdCPUE	SD	nCPUE	stdCPUE	SD
1994	1.0966	0.4296	0.0747	0.1325	0.2316	0.0274
1995	1.0013	0.3763	0.0745	0.1055	0.1657	0.0178
1996	1.1709	0.3974	0.0779	0.1387	0.2821	0.0373
1997	1.0700	0.4474	0.0944	0.0956	0.2229	0.0316
1998	0.9947	0.3652	0.0857	0.0918	0.2161	0.0293
1999	1.0335	0.3813	0.0716	0.1061	0.2184	0.0288
2000	1.1436	0.4551	0.0879	0.1340	0.2272	0.0317
2001	1.0345	0.4236	0.1215	0.1496	0.2310	0.0277
2002	1.0958	0.4020	0.0789	0.1620	0.2043	0.0214
2003	0.8821	0.2876	0.0502	0.1486	0.1701	0.0144
2004	1.0078	0.2874	0.0411	0.2055	0.1696	0.0166
2005	1.1874	0.3074	0.0381	0.1608	0.1536	0.0171
2006	1.1851	0.3473	0.0506	0.2202	0.2008	0.0253
2007	1.1147	0.3914	0.0567	0.2268	0.2418	0.0305
2008	0.7892	0.2787	0.0611	0.1755	0.2448	0.0336
2009	1.0327	0.3819	0.0890	0.1702	0.2117	0.0320
2010	0.9027	0.3467	0.0928	0.2627	0.1912	0.0297
2011	0.7880	0.3336	0.0934	0.1272	0.1555	0.0225
2012	0.7942	0.3249	0.0955	0.1300	0.1482	0.0228
2013	0.7794	0.3221	0.0813	0.1292	0.1228	0.0192
2014	0.8157	0.3375	0.0937	0.1102	0.1176	0.0187
2015	0.9795	0.3464	0.0996	0.1811	0.1776	0.0335
2016	1.0467	0.4600	0.1001	0.2360	0.1735	0.0308
2017	1.0232	0.5503	0.1233	0.1624	0.1207	0.0198
2018	1.1394	0.7167	0.1790	0.1635	0.1482	0.0253
2019	1.0356	0.5339	0.1435	0.1126	0.1549	0.0276
2020	1.1394	0.5273	0.1595	0.1604	0.2415	0.0497
2021	1.0180	0.4558	0.1093	0.4972	0.2748	0.0576

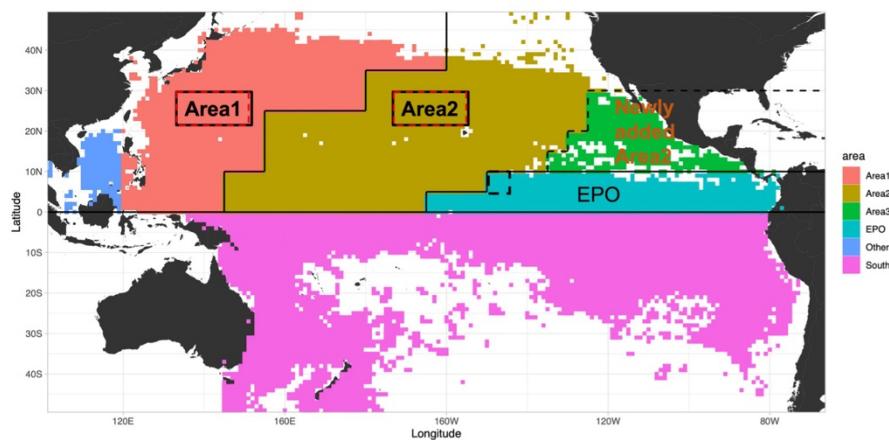


Figure 1. Japanese longline operational data points overlapped with WCNPO and EPO boundaries. The area 1 highlighted in pink and area 2 highlighted in yellow are the two areas assessed in this study.

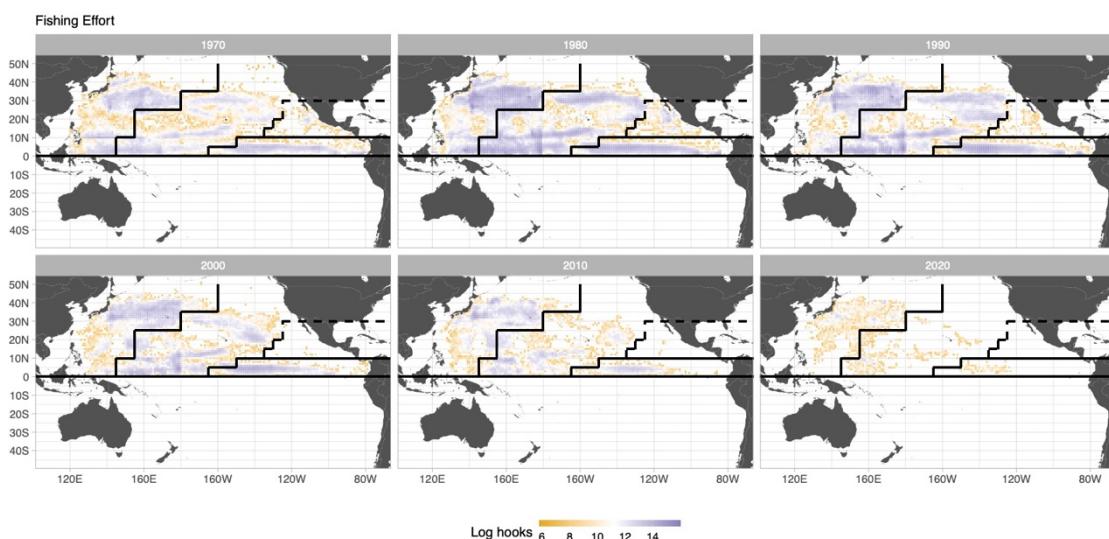


Figure 2. Japanese long line fishing effort measured by number of hooks ($\log(\text{hooks})$) for the area of study. Number of hooks were summarized by decade and $1^\circ \times 1^\circ$ grid resolution.

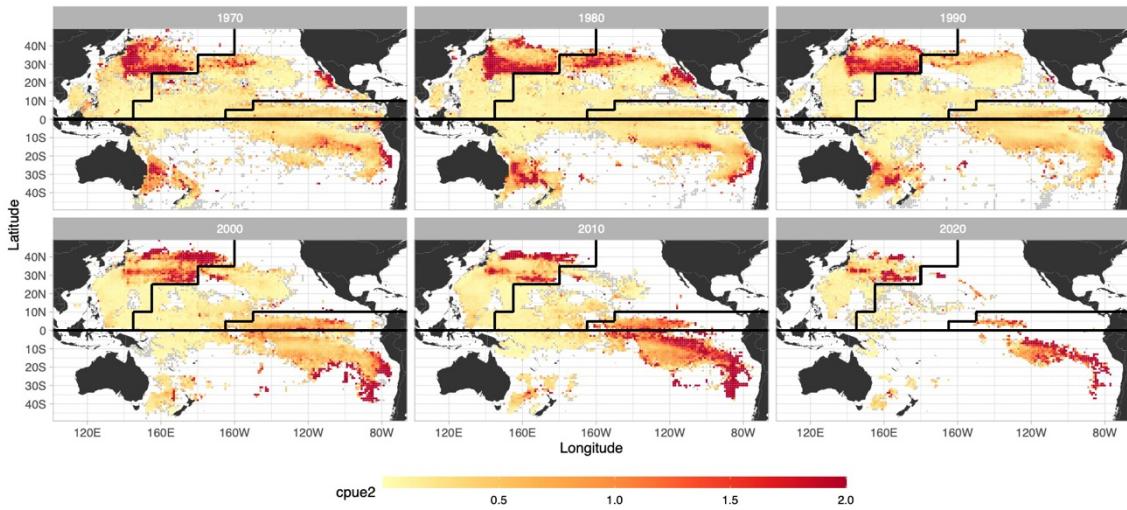


Figure 3. Historical changes in nominal swordfish CPUE (number of sword fish caught / 1,000 hooks) by Japanese longline fishery (1976-2020). The gray area denotes zero catch during the summarizing period. Nominal CPUE was summarized by decade and $1^\circ \times 1^\circ$ grid resolution.

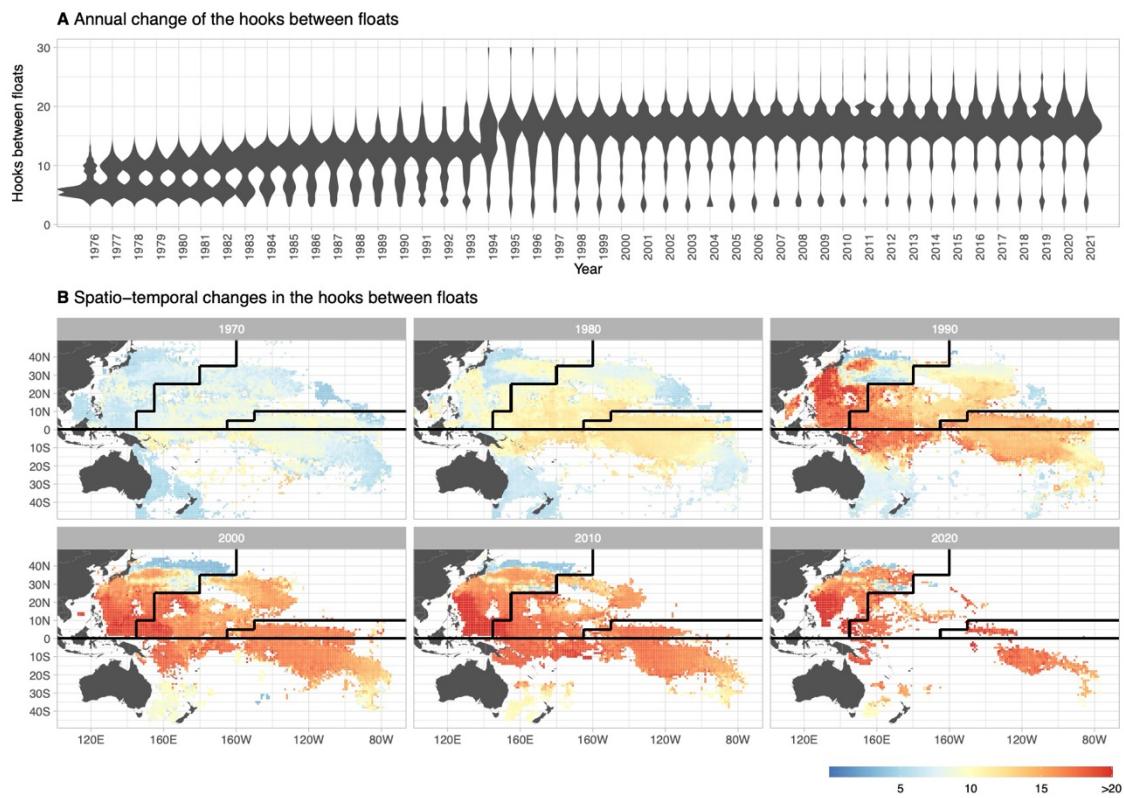


Figure 4. Changes in the number of hooks between floats (HBF) for Japanese longline fishery for each decade and $1^\circ \times 1^\circ$ grid resolution. HBF quickly changes after 1990s and the HBF seems to be associated with the fishing location.

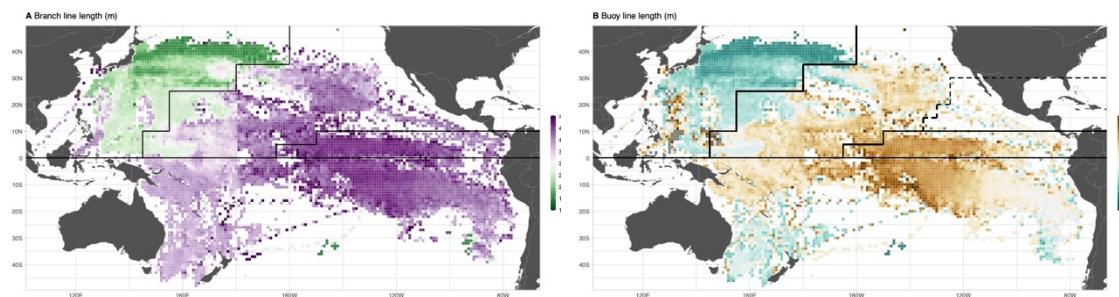


Figure 5. Spatial pattern of the branch line length (A) and buoy line length (B) for Japanese longline fishery.

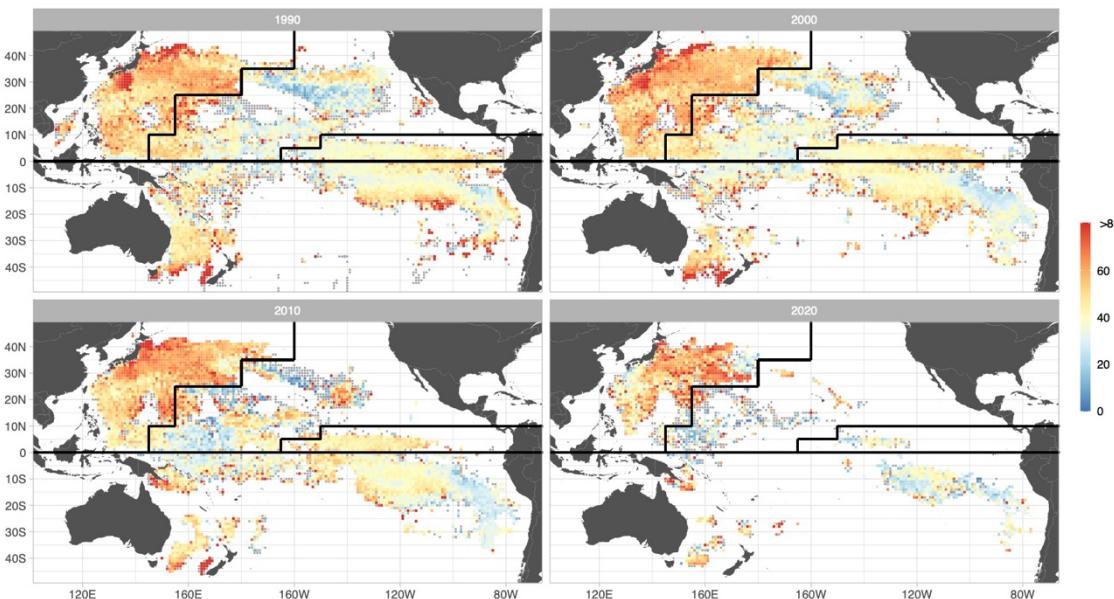


Figure 6. Decadal mean semi-dress weight (kg) of swordfish caught by Japanese longline fishery (1994-2021). The gray area denotes zero catch during the summarizing period. Mean semi-dress weight was summarized by decade and $1^\circ \times 1^\circ$ grid resolution.

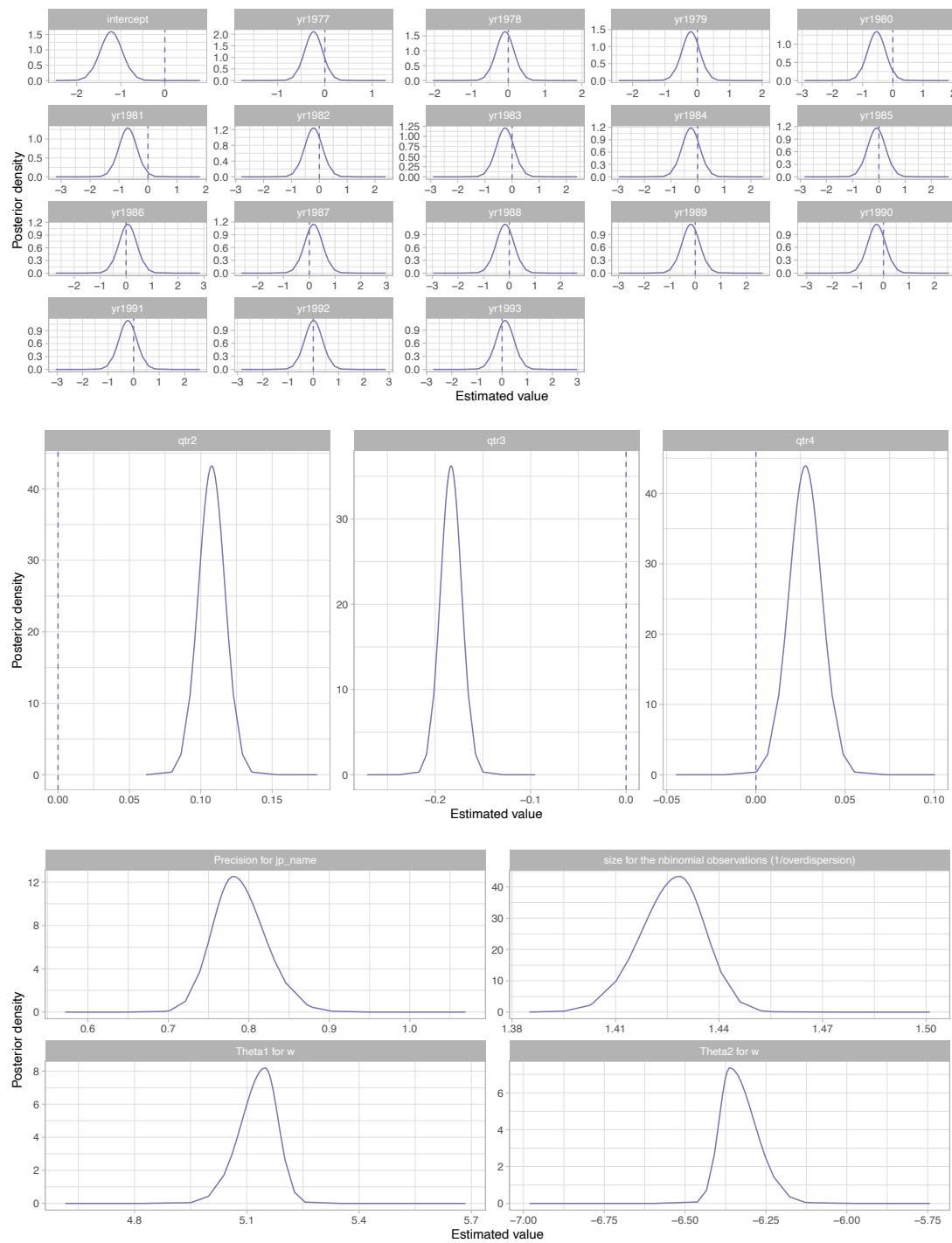


Figure 7. Posterior distribution of early period model for Area 1 (1976-1993); Top: fixed year effect; Middle: fixed quarter effect; Bottom: random-effect parameters.

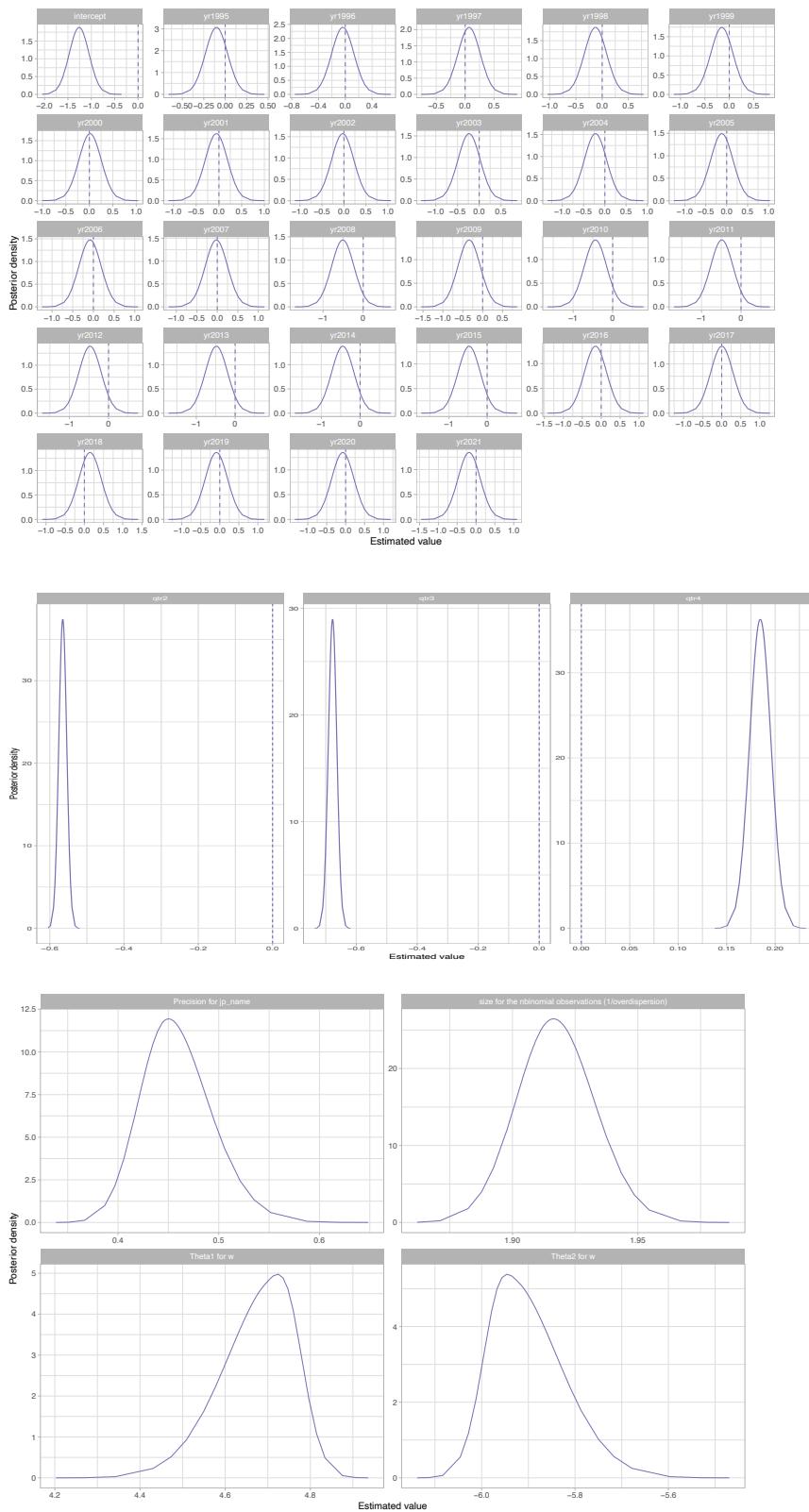


Figure 8. Posterior distribution of late period model for Area 1 (1994-2021); Top: fixed year effect; Middle: fixed quarter effect; Bottom: random-effect parameters.

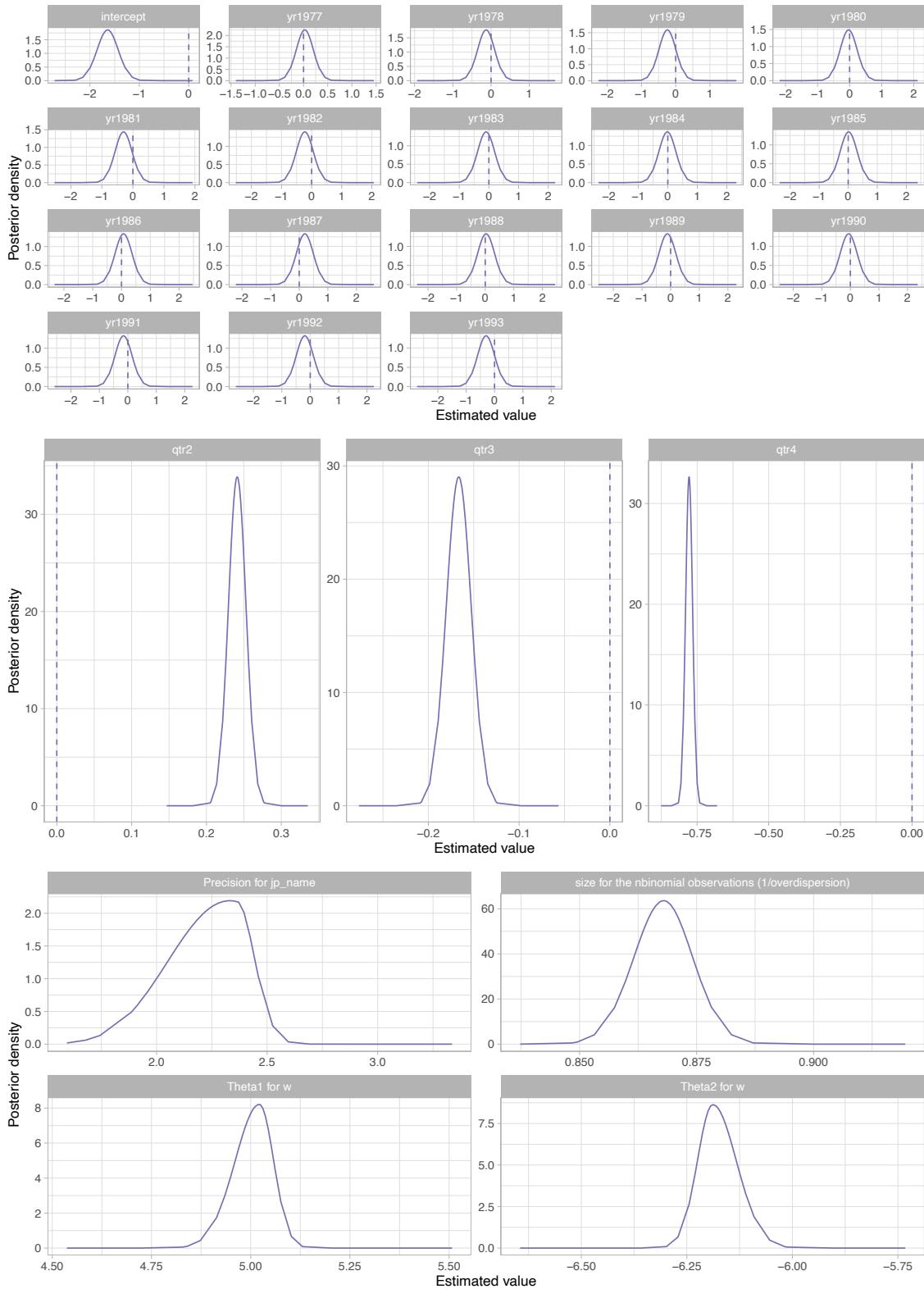


Figure 9. Posterior distribution of early period model for Area 2 (1976-1993); Top: fixed year effect; Middle: fixed quarter effect; Bottom: random-effect parameters.

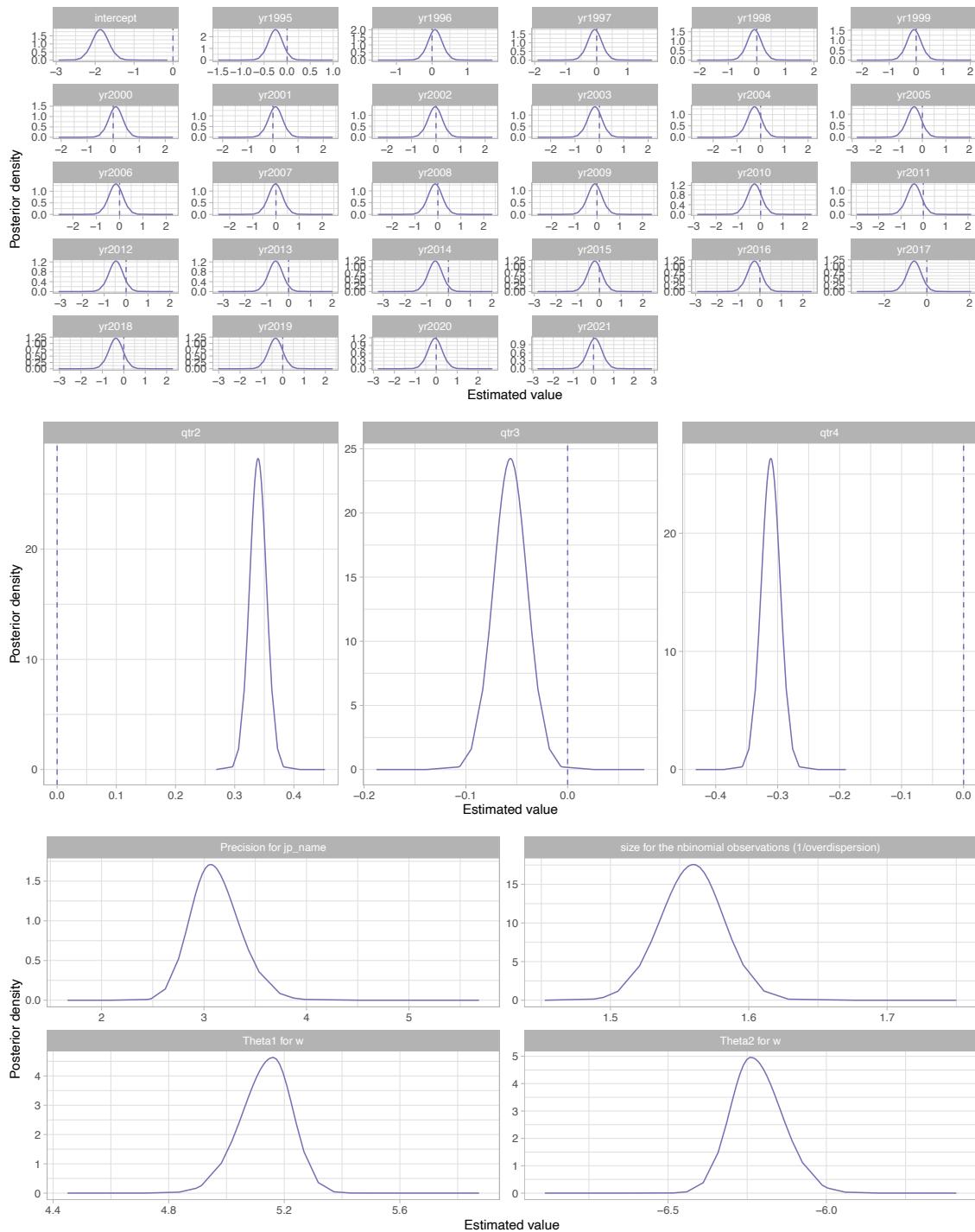


Figure 10. Posterior distribution of late period model for Area 2 (1994-2021); Top: fixed year effect; Middle: fixed quarter effect; Bottom: random-effect parameters.

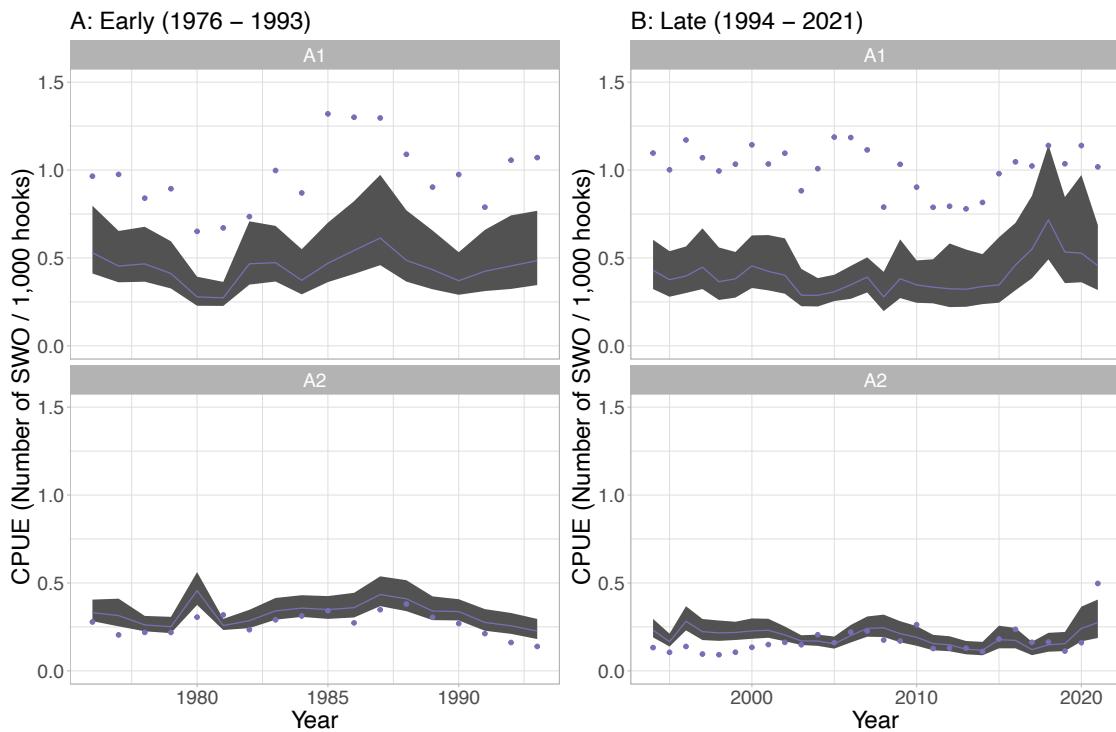


Figure 11. Standardized Japanese longline CPUE. Left: early time period (1976 – 1993). Right: late time period (1994 – 2021).

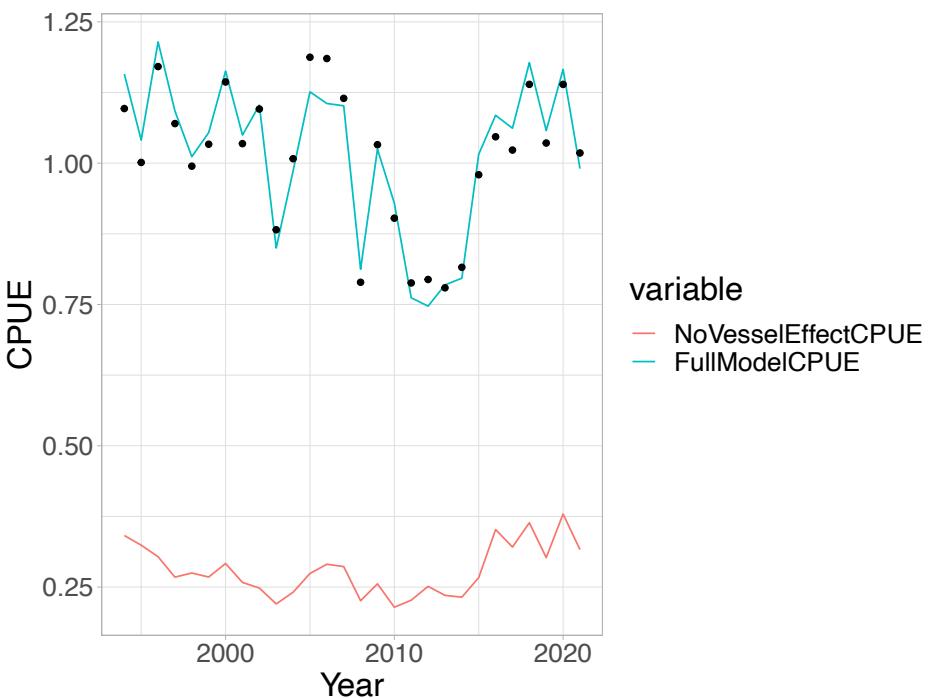


Figure 12. CPUE prediction with and without vessel effect.