CPUE Standardization for Striped Marlin (*Kajikia audax*) using Spatio-Temporal Model using INLA

Hirotaka Ijima* and Haruko Koike

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

Email: ijima@affrc.go.jp

This working paper was submitted to the ISC Billfish Working Group Workshop, 13, 15-18 December 2021 held by webinar.

Abstract

Using the Japanese logbook data, we addressed to standardize the CPUE of the Western Central North Pacific striped marlin. Before the CPUE standardization, we analyzed the relationship between detailed fishing gear settings and fishing grounds using a geostatistical model. Adding the effect of fishing gear to the geostatistical model did not significantly improve the WAIC. This result indicates that the gear setting also depends mainly on the location, and we did not use gear information for CPUE standardization. The fleet definition was based on the previous stock assessment and assumed that the catch size would change depending on the area and season. We used R software package INLA for these analyses, and model selection was performed using the WAIC and the LOOCV obtained from Bayesian estimation. As a result of model selection, a spatiotemporal model was selected. To standardize CPUE, we estimated spatial CPUE annually, averaged according to the fleet definition. Striped marlin might be migrating seasonally in the North Pacific. In this study, we tried to build a spatiotemporal model considering seasonality. However, most models had technical problems. For example, some models did not converge, and the calculation crashed in the middle. Some converged model indicated that the spatial distribution of latent spatial field fluctuates greatly depending on the season. Thus, it needs to develop a spatiotemporal model considering seasonality for the future.

Introduction

Striped marlin is a bycatch species in the Japanese longline fishery, and its catch is smaller than other species, with most records being zero. On the other hand, the operation area of Japanese longline vessels is shrinking year by year. As a result, information on striped marlin catches is becoming sparse in location. In addition, the fish size of striped marlin has been reported to vary spatially and seasonally, and the ISC BILLWG has been using definition of the spatially and seasonally variable fleet. (Ijima and Kanaiwa 2019a, Figure. 1). However, these data characteristic poses a significant problem in standardizing the catch per unit effort (CPUE) data. For example, it is not well known how the effect of gear setting on bycatch species is affected. The reported gear setting seems to be also spatially dependent (Figure 2). For spatial deficits, geostatistical models are increasingly being used to correct spatiotemporal bias (Ijima and Koike 2020). However, geostatistical models have not yet been used for Western Central North Pacific (WCNPO) striped marlin (Ijima and Kanaiwa 2019b). In addition, the model selection is essential because stock assessment requires scientific consensus. However, several studies indicate that AIC does not work for the hierarchical model (Watanabe and Opper 2010). Also, there are few examined in detail the extent to which explanation for data has been improved in complex hierarchical CPUE standardized models.

In this study, we first build the geostatistical model using Japanese longline logbook data, where gear settings are recorded in detail, and discuss the relationship between CPUE and gear settings for striped marlin. Next, multiple geostatistical models were constructed, and the best model was selected by comparing WAIC and LOOCV. Finally, using the selected model, we calculated standardized CPUE based on the previous fleet definition.

Material and methods

- Longline logbook data

We used the offshore and distant water longline logbook data from 1976 onwards. The Japanese longline logbook has been recorded since 1952, but vessel names became available after 1976. The format of Japanese longline logbook data was changed around 1994. Thus, we set two-time series, 1976-1993 and 1994-2020, as in other billfish species analyses (Kanaiwa and Ijima 2018, Ijima and Koike 2020). Regarding the gear configuration, the number of hooks between floats (HBF) was described after 1975, and since 1994 some vessels have reported buoy length, length of a branch line, and length between floats. The reporting rate of this detailed information was very low until 2010, but a relatively large number of vessels have reported it in recent years. Thus, we analyzed the fishing gear effect for striped marlin CPUE using only data from vessels reporting detailed gear setting.

Before the analysis, we organized the logbook and checked the trend of the CPUE. We chose the logbook with over 300 hooks operation and used the range of HBF between 3 and 30. The spatial distribution of nominal CPUE varies with the season, and the average fish size differs greatly between the north and south Pacific Oceans (Figure. 3). Nominal CPUE shows some spatiotemporal variation, with areas where no striped marlin was caught appearing (Figure. 4 and 5). The fishing area in the Japanese longline is shrinking year by year (Figure. 4 and 5).

- Statistical model

In analyzing the gear setting effect for the CPUE, we first constructed a geostatistical model in which the response variable was a zero-inflated Poisson distribution, and the covariates were the spatial effect and intercept. Next, we added gear effects (hooks between floats, branch line length, and buoy line length) step-by-step with categorical variables, and a total of seven models were constructed (Table 1). We used number of hooks for the effort in these models as an offset term.

In terms of the statistical model for CPUE standardization, we used the zero-inflated Poisson distribution model, similar to the fishing gear analysis, and we used the covariates year, quarter, vessel name, and location. We treated these covariates as fixed or random effects, and multiple models were constructed (Table 2). In the geostastical model, location data was treated by the Stochastic Partial Differential Equations (SPDE) approach.

- Model selection and validation

In a hierarchical model such as a geostatistical model, WAIC or LOOCV have been recommended for model selection instead of AIC (Watanabe and Opper 2010, Vehtari et al. 2017). In order to calculate WAIC and LOOCV, we estimated the parameters using INLA that works Bayesian estimation. We plotted the posterior distribution of the parameters for the fixed effects and checked how much the posterior distribution contains zero.

- Standardized CPUE

We could not calculate standardized CPUE and Bayesian credible intervals due to the technical problems within INLA. Usually, standardized CPUE is calculated by the least-squares means. The least-squares means needs estimate data made by all combinations of variables and give estimate data to the INLA package. However, estimate data was too big to run the model. Therefore, we calculated standardized CPUE outside of the INLA package. At first, we estimated the spatiotemporal CPUE and calculated standardized CPUE that arithmetic averaged by year according to the fleet definition.

Result and discussion

- Effect of fishing gear on striped marlin CPUE

The simple geostatistical model improved the value of WAIC by 6.8% over the GLMM with the information of coordinates as random effects (Table. 4). When the effect of the HBF was added to this simple geostatistical model, the value of WAIC increased by 0.5% (Table. 4). On the other hand, when we added branch line length and buoy line length, both WAIC values decreased by 0.4% (Table. 4). These fixed effects were thought to indicate the gear depth of the longline, but they did not contribute significantly to the improvement of the model. The setting of the gear depth of the longline is considered to be dependent on the ocean environment, such as the mixed layer depth. Thus, the gear effect may be included by the latent spatial field of the geostatistical model. It is also possible that the effects of the fishing gear may not have occurred in the first place because striped marlin is not a target species. Considering these results, we did not add gear effects to the statistical model for the CPUE standardization.

- Standardized CPUE of WCNPO striped marlin

In the late period (1993-2020), the WAIC of model 004 was the smallest (Table 5). This model incorporates the Metern function in the time step and sets the knot to reduce the calculation cost. In other words, we set to a year or year-quarter time step, and the one latent spatial field in the year was estimated. However, we could not obtain all the knot for the time step we assumed. Thus, we selected the spatiotemporal model (008) with the second-lowest WAIC in this study (Table 6). The early period (1976-1993) model was constructed similarly, and the lowest WAIC model was obtained, but the spatiotemporal model (013) was selected because there was a difference between the input knot and the output knot. Zeros values were not included regarding the posterior distribution of the estimated parameters, and they are generally well estimated (Figures 6-7). The latent spatial field showed large inter-annual variability, with the high spatial effect area shrinking after 2000 (Figure 8-9).

The spatiotemporal model analysis results were used to calculate the standardized CPUE corresponding to the fleet definition of stock synthesis 3 (Figure 10). Although these fleets were assumed to catch different cohorts, two CPUE show similar trends (Figure 10). If the definition of fleets was correct and other cohorts were selected, the trends of the two indices should be different. Therefore, the fleet definition needs to be examined in the future.

We could not output an annual knot in the seasonal spatiotemporal model, but we could estimate the latent seasonal spatial fields in model 012 that accounted for seasonal variation (Figure 11). There was a tendency for the positive and negative spatial effects to reverse between the second and third quarters (Figure 11). This might reflect the seasonal migration of striped marlin, which may significantly impact the CPUE standardization. Thus, we should be continued to develop a seasonal variation model.

Although there are still various problems to be solved in this analysis, we propose using the standardization results for the next stock assessment because the current model has considerably improved WAIC over the GLMM used in the previous stock assessment.

References

Ijima, H., Kanaiwa M 2019a Size-dependent distribution of Pacific Striped Marlin (*Kajikia audax*): The analysis of Japanese longline fishery logbook data using the finite mixture model. ISC/19/BILLWG-1/09.

Ijima H., 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-3/02.

Ijima, H., Kanaiwa M 2019b Japanese longline CPUE of the striped marlin (*Kajikia audax*) in the WCNPO. ISC/19/BILLWG-1/07.

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

Kanaiwa, M., Ijima, 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/BILLWG-1/07.

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.

Table 1. Statistical model list for the gear effect analysis. Operational data that report detailed gear settings were used (1994-2020).

No	Model	INLA function
001	non-spatial model + no gear	stm ~ 0 + intercept + yr + qtr + f(jp_name, model="iid", hyper=hcprior)
002	non-spatial model + latlon as random effect	stm ~ yr + qtr + f(jp_name2, model="iid", hyper=hcprior) + f(latlon, model="iid")
003	non-spatial model + gear(buoy length)	stm ~ 0 + intercept + yr + qtr + f(jp_name, model="iid", hyper=hcprior) + buoy
004	simple spatial model + no gear	stm ~ 0 + intercept + yr + qtr + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde)
005	simple spatial model + gear(hpb)	stm ~ 0 + intercept + yr + qtr + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde) + hpb
006	simple spatial model + gear(buoy length)	stm ~ 0 + intercept + yr + qtr + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde) + buoy
007	simple spatial model + gear(length_branch_line)	stm ~ 0 + intercept + yr + qtr + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde) + branch_line

Table 2. Statistical model list for the CPUE standardization in the late period (1994-2020).

No	Model	INLA function
001	simple spatial model	stm ~ 0 + intercept + yr + qtr + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde)
002	simple spatial with no year effect	stm ~ 0 + intercept + qtr + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde)
003	simple spatial with no qtr effect	stm ~ 0 + intercept + yr + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde)
004	spatiotemporal (AR1 yr (with t mesh knot at Qtr1)	stm ~ 0 + intercept + yr + qtr + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1"))
005	spatiotemporal (AR1 yr (with t mesh knot at Qtr1) w/out fixed yr)	stm ~ 0 + intercept + qtr + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1"))
006	spatiotemporal and seasonal (AR1 yr (with t mesh knot at Qtr1), iid Qtr)	stm ~ 0 + intercept + yr + qtr +f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1")) + f(q, model=spde, group=q.group, control.group=list(model="iid"))
007	spatiotemporal and seasonal (AR1 yr (with t mesh knot at Qtr3), iid Qtr)	stm ~ 0 + intercept + yr + qtr +f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1")) + f(q, model=spde, group=q.group, control.group=list(model="iid"))
008	spatiotemporal (AR1 yr No t-mesh) w/out fixed yr	stm ~ 0 + intercept + qtr +f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1"))
009	spatiotemporal and seasonal (AR1 yr, iid Qtr)	stm ~ 0 + intercept +f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1")) + f(q, model=spde, group=q.group, control.group=list(model="iid"))

Table 3. Statistical model list for the CPUE standardization in the early period (1976-1993).

No	**Model**	**INLA function**
010	spatiotemporal (AR1 yr (with t mesh knot at Qtr1)	stm ~ 0 + intercept + yr + qtr + f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1"))
011	spatiotemporal and seasonal (AR1 yr (with t mesh knot at Qtr1), iid Qtr)	stm ~ 0 + intercept + yr + qtr +f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1")) + f(q, model=spde, group=q.group, control.group=list(model="iid"))
012	spatiotemporal and seasonal (AR1 yr (with t mesh knot at Qtr3), iid Qtr)	stm ~ 0 + intercept + yr + qtr +f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1")) + f(q, model=spde, group=q.group, control.group=list(model="iid"))
013	spatiotemporal (AR1) No t-mesh No fixed yr	stm ~ 0 + intercept + qtr +f(jp_name, model="iid", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="ar1"))

Table 4. Model comparison several gear effect models. The rate of change of WAIC was calculated based on a simple spatial model.

No	Model	WAIC	LOOCV	%Change of WAIC
001	non-spatial model + no gear	414,312	412,385	13.4
002	non-spatial model + latlon as random effect	390,220	386,820	6.8
003	non-spatial model + gear(buoy length)	415,560	413,382	13.8
004	**simple spatial model + no gear**	**365,207**	**362,296**	**0**
005	simple spatial model + gear(hpb)	366,869	363,904	0.5
006	simple spatial model + gear(buoy length)	363,766	360,999	-0.4
007	simple spatial model + gear(length_branch_line)	363,731	360,682	-0.4

Table 5. Model selection result 1994-2020. No. 008 was the selected model. "t mesh kont" model could not estimate annual latent spatial field.

No	Model	WAIC	LOOCV
001	Simple spatial model	792,459	790,622
002	simple spatial with no year effect	819,056	817,113
003	simple spatial with no qtr effect	797,882	796,088
004	spatiotemporal (AR1 yr (with t mesh knot at Qtr1))	716,028	714,228
005	spatiotemporal (AR1 yr (with t mesh knot at Qtr1) w/out fixed yr)	760,357	757,671
006	spatiotemporal and seasonal (AR1 yr (with t mesh knot at Qtr1), iid Qtr)	Killed	
007	spatiotemporal and seasonal (AR1 yr (with t mesh knot at Qtr3), iid Qtr)	726,127	723,309
008	**spatiotemporal (AR1 yr No t-mesh) w/out fixed yr**	**724,080**	**721,616**
009	spatiotemporal and seasonal (AR1 yr, iid Qtr)	Crashed	

Table 6. Model selection result 1976-1993. No. 013 was the selected model.

No	**Model**	**WAIC**	**LOOCV**
010	spatiotemporal (AR1 yr (with t mesh knot at Qtr1)	1,546,768	1,537,948
011	spatiotemporal and seasonal (AR1 yr (with t mesh knot at Qtr1), iid Qtr)	1,459,512	1,451,576
012	spatiotemporal and seasonal (AR1 yr (with t mesh knot at Qtr3), iid Qtr)	1,464,964	1,457,088
013	**spatiotemporal (AR1) No t-mesh No fixed yr**	**1,477,973**	**1,468,642**

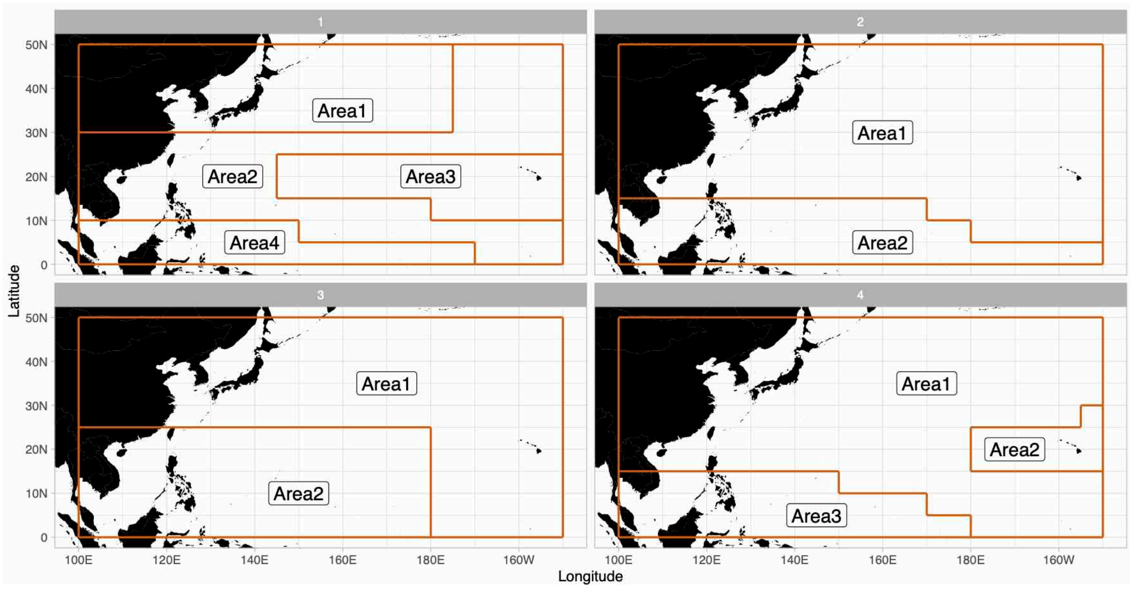
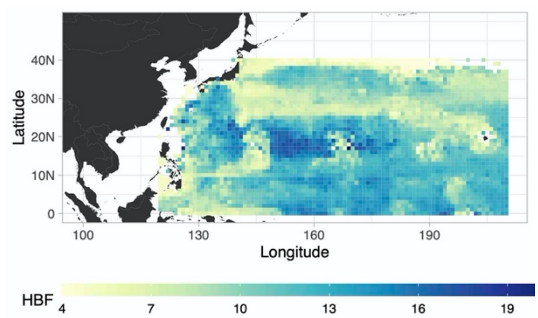
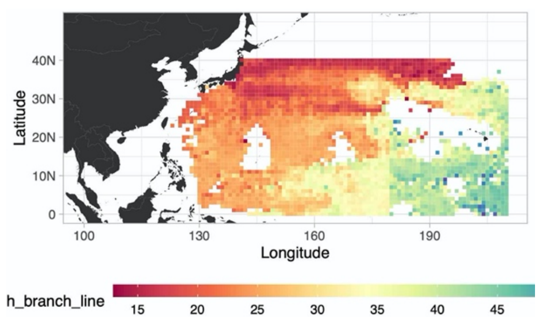


Figure 1. Area-seasonal fleet definition of Japanese longline fishery.

A Hooks between floats



B Branch Line Length



C Buoy Line Length

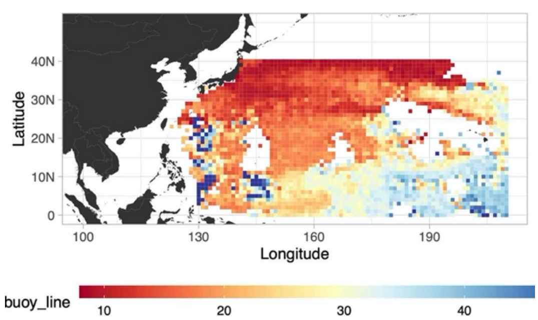
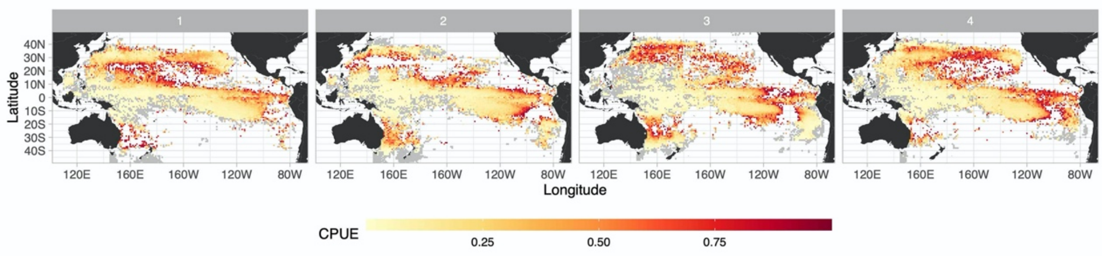


Figure 2. Spatial differences in gear configuration of Japanese longline fishery.

A Nominal CPUE



B Mean semi-dress weight (Kg)

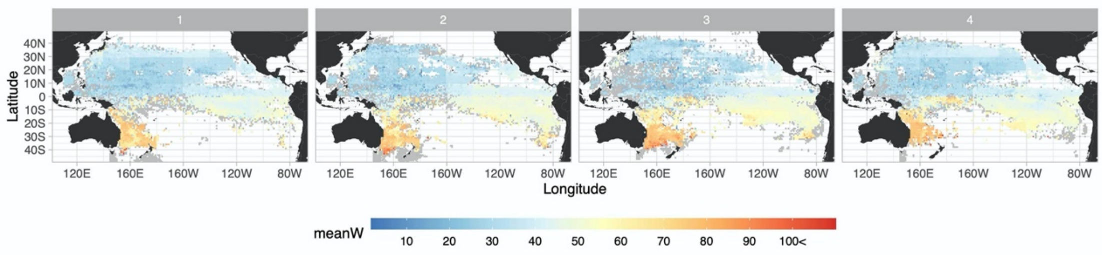


Figure 3. Spatial pattern of CPUE and fish size (1994-2020).

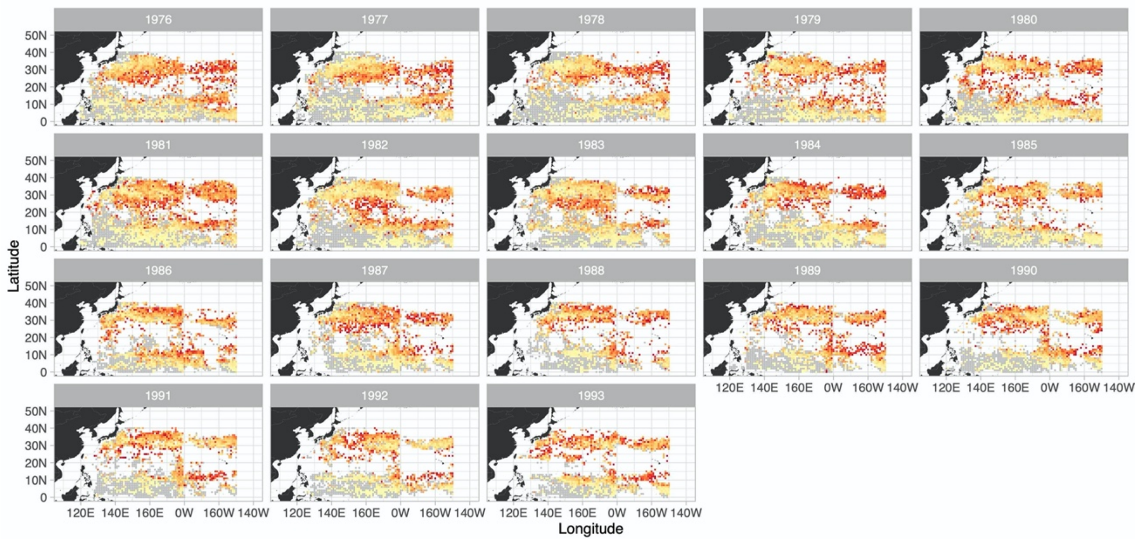


Figure 4. Spatiotemporal trends of Japanese longline CPUE (1976-1993).

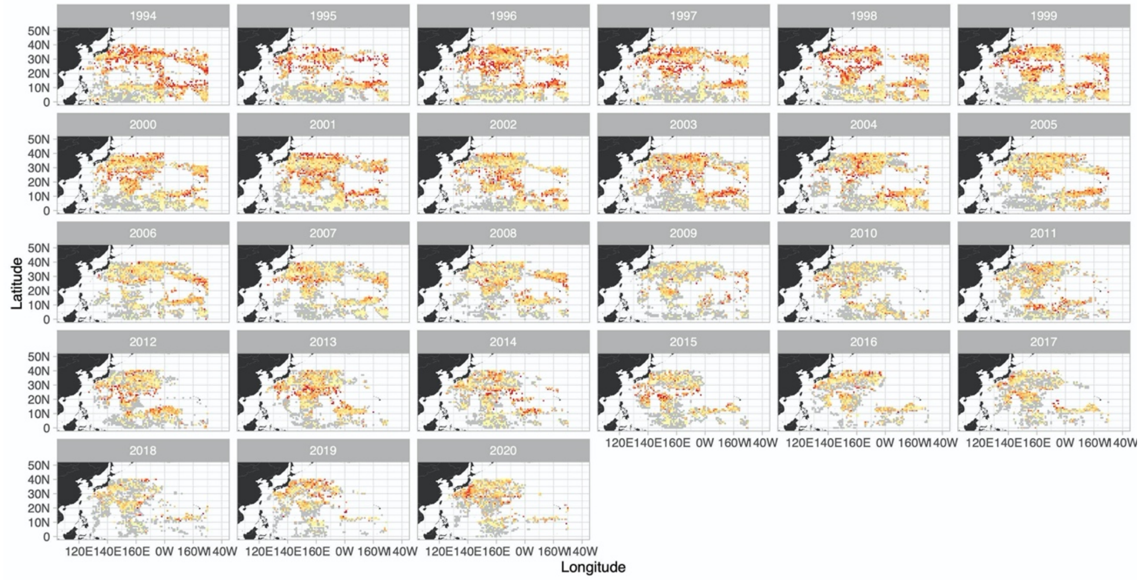


Figure 5. Spatiotemporal trends of Japanese longline CPUE (1994-2020).

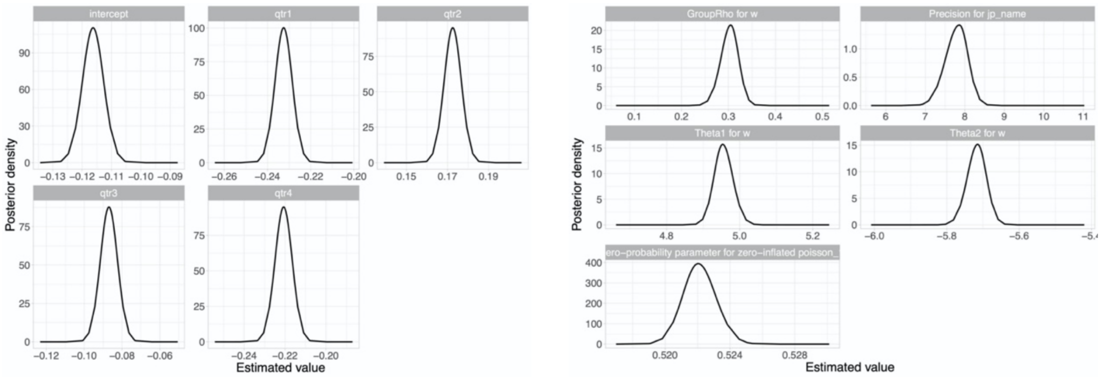


Figure 6. Posterior distribution of early period model (1976-1993); Left fixed effect, Right random effect parameter.

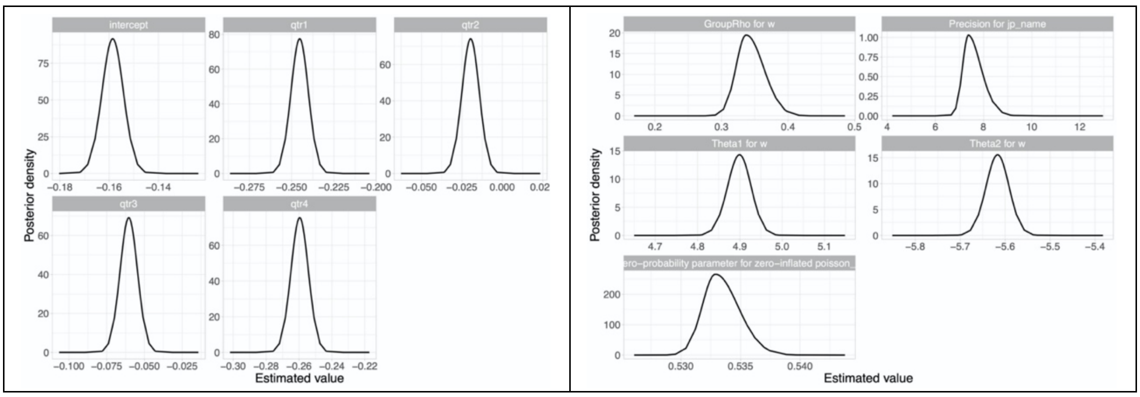


Figure 7. Posterior distribution of late period model (1994-2020); Left fixed effect, Right random effect parameter.

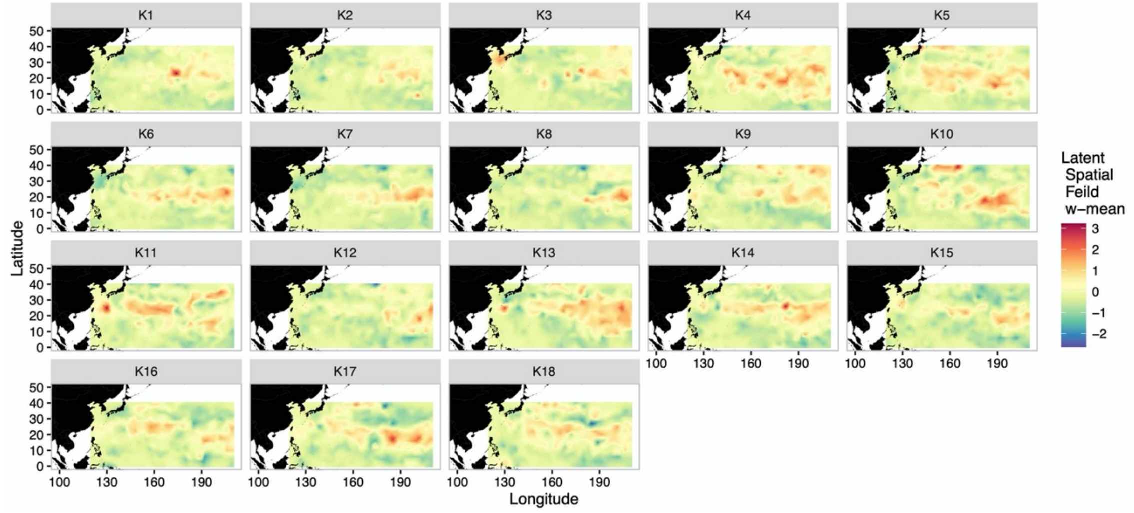


Figure 8. Annual trends of latent spatial field for early period model (1976-1993).

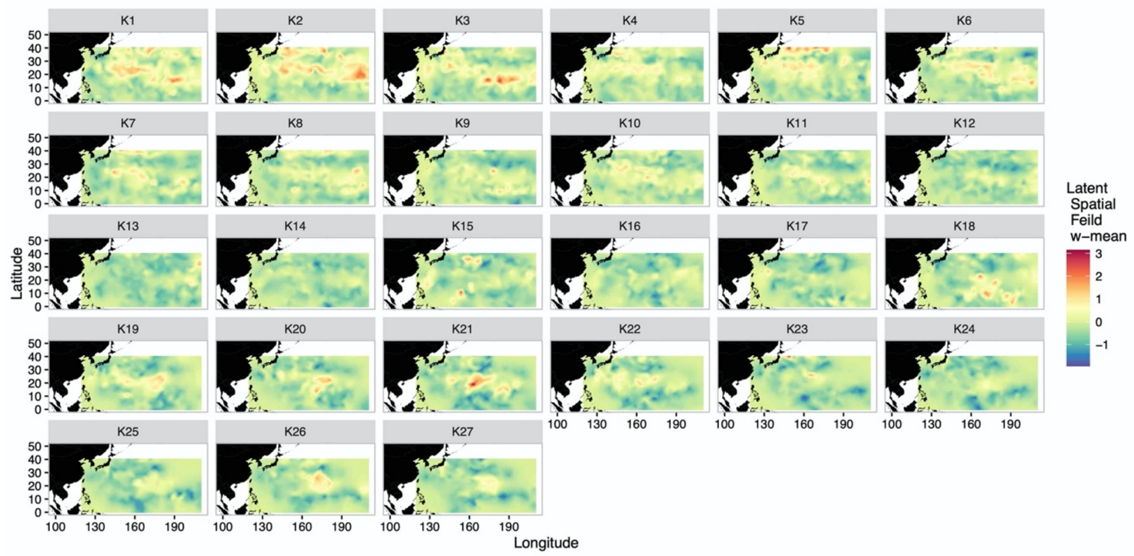


Figure 9. Annual trends of latent spatial field for late period model (1994-2020).

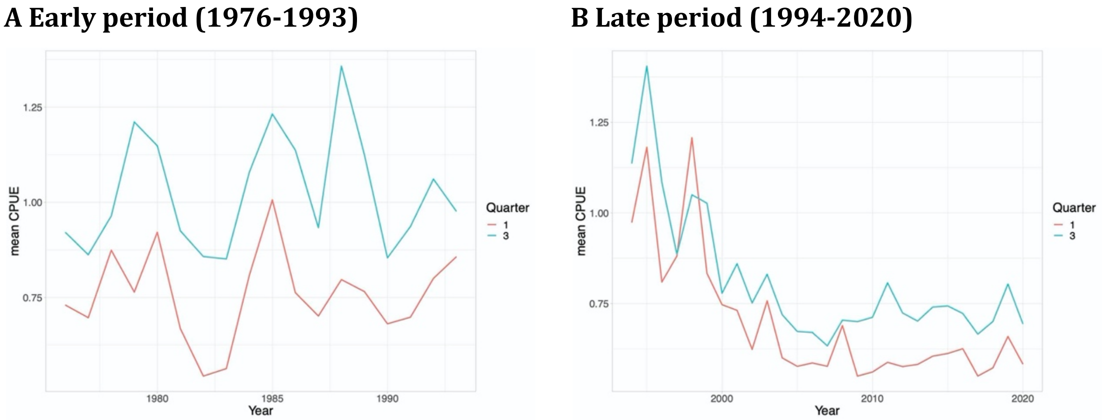


Figure 10. Standardized Japanese longline CPUE. Left: early time period (1976-1993). Right: late time period (1994-2020).

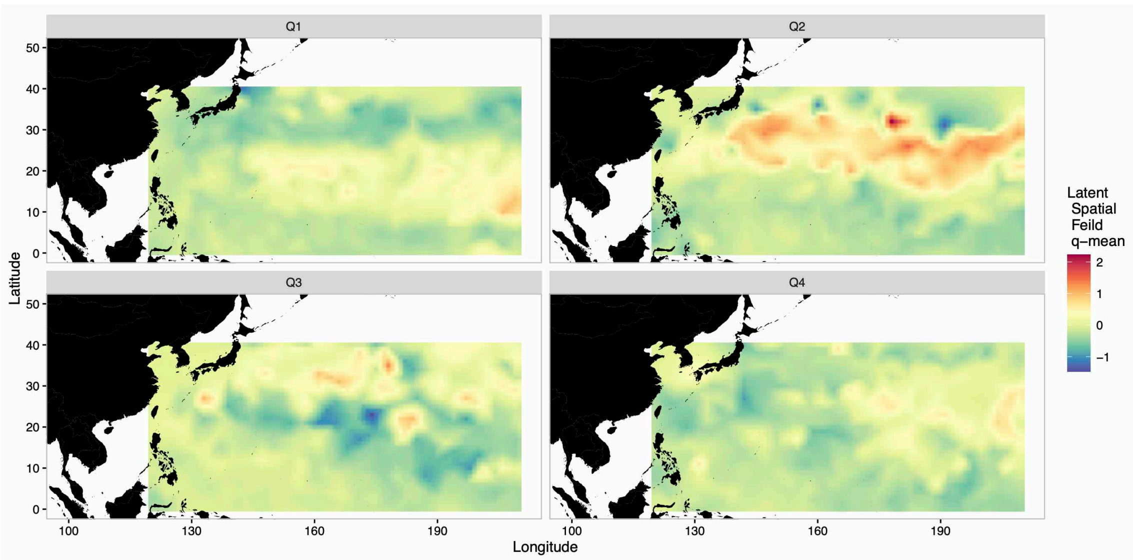


Figure 11. Estimated seasonal latent spatial field using model 012.