Preliminary analysis for the CPUE standardization of the Pacific blue marlin using Japanese longline logbook and the R software package R-INLA.

Hirotaka Ijima[1] and Haruko Koike[2]

[1] Highly Migratory Resources Division, Fisheries Stock Assessment Center, Fisheries Resources Institute (FRI), Japan Fisheries Research and Education Agency.
[2] Seafood Legacy LTD.

Email: ijima@affrc.go.jp

This working paper was submitted to the ISC Billfish Working Group Data Preparatory Meeting, 6-7 and 10 November 2020 held by webinar.

Abstract
We analyzed Japanese longline logbook data to obtain indicators of the historical trends of the Pacific blue marlin. We applied the spatiotemporal model for the CPUE standardization because the Japanese longline area coverage shrinks year by year. We used an R-INLA package and WAIC to make an appropriate model selection for the random effect model. At first, we tried the pan-Pacific analysis similar to the Habitat model used in the previous stock assessment. However, this model did not converge. Secondly, considering the average catchweight spatial pattern, we extracted the area that fish of the size corresponds to the SS3 model's selectivity. The smallest WAIC among the converged models was the seasonal geostatistical model. However, various problems have been identified with this model. The randomized quantile residuals indicated overestimation in the 1990s population. In detail, the spatial trends of randomized quantile residuals differed between 1994 and 2018. In other words, the model validation suggests the need to build a spatiotemporal model. However, the spatiotemporal model could not be estimated the fixed effect of season and intercept. Also, we need to perform a statistical analysis to determine the analysis area because the trend of standardized CPUE strongly depends on the area definition. From these results, we judged that the results of this study are preliminary.

Introduction
The ISC BILLWG conducted a stock assessment of the Pacific blue marlin in 2016 (ISC 2016). This stock assessment used a Japanese abundance index estimated by a habitat model (Kai et al., 2016). The habitat model considers environmental variability and the corresponding distribution of the Pacific blue marlin. On the other hand, in recent years, CPUE standardization methods using the spatiotemporal models are being established (Kai et al., 2017). It is possible to estimate the habitat model with the same accuracy because the spatiotemporal model handles the location information as a latent spatial field. Besides, R software package R-INLA can calculate Widely Applicable Information Criterion (WAIC) in the Bayesian estimation process, and the randomized quantile residuals (RQR) can be used to diagnose the model's estimation accuracy (Lindgren & Rue 2015). In this study, we attempted CPUE standardization for the Pacific blue marlin using R-INLA.

Material and methods

1. Data source

Logbook data

The fleet of Japanese offshore and distant water longline was divided into the early and the late series. The Japanese longline logbook format has changed since 1994, and the quality of the data changes before and after. The catchability has also changed significantly in the 1990s because the mainline material was changed to nylon. As a result, the branch line did not sink deeply unless the hooks between floats (HBF) were increased. For example, the five HPB in the 1990s and five HPB in the 1970s have a different effect on the blue marlin catch. The operating area in the pacific ocean of Japanese longline fishery is shrinking year by year. Thus, a spatiotemporal model is required for the CPUE standardization. We used the R software package R-INLA to apply the spatiotemporal model (Lindgren & Rue 2015). The body size (cohort) of blue marlin varies depending on the fishing area. "Area as fleet approach" may be needed to reflect such an area dependent size selectivity (Waterhouse et al., 2014). This study considered that the length-frequency data mode is 150 cm and extracted the area where small size fish are caught. For reference, we also analyzed in a wide area, the same as the last stock assessment.

Data screening

In carrying out the analysis, we screened the logbook data as follows:

- Area in WCPO: -10=<lat<25, 125=<lon<210 (see. Figure 1)
- Time period: 1994 ~ 2018
- Number of hooks: >200
- Hooks between floats: 3-36

To reduce the data set size, we aggregated the data set by location, vessel name, year, month, and HBF.

2. Model developing

Model assumption

We assumed a linear or non-linear regression to explain the number of blue marlin catch (*bum*) as, *bum = intercept + trend + noises + offset(1,000 hooks).*

The number of blue marlins caught (observed value: bum) follows zero-inflated Poisson distribution (ZIP) or Poisson distribution (PO). ZIP can eliminate the overdispersion of Poisson distribution. Negative binomial distribution (NB) can also consider the overdispersion of count data. However, we did not use the NB because Minami et al., 2007 reported that the estimated CPUE trend tends to be steeper in the NB model. We also did not use the delta model to compare with the simple one-step model.

We assumed the Japanese longline logbook had the annual population "trend" that depends on the selectivity of the longline fishery. To examine our assumption, we tested the three models:

- The fixed effect model of the year
- The autoregressive model of order 1 (ar1)
- The random walk model of order 1 (rw1).

Various "noises" are affecting the number of the catch of blue marlin. We used four covariates that are the HBF (or gear deep and shallow), season, vessel name, and locations to remove the "noises". Each covariate was applied to a fixed or a random effect. We compared four types of spatial covariates that are the fixed effect of a 5°x5 ° grid, the random effect of a 5°x5 ° grid, the stochastic partial differential equation (SPDE) model, spatiotemporal SPDE model, and seasonal SPDE model. The vessel name is considered a random effect. HBF effect was applied fixed effect (deep: HBF>8, shallow: HBF<=8) or random effect.

The list of models examined in this study is shown in Table 1.

Model selection

There are various debates about the accuracy of AIC in the random effect model. This time, we used the WAIC that supports the random effect model. The WAIC can only be calculated by Bayesian estimation. Thus, we used the R-INLA package.

Model validation

We checked how much the overdispersion had been eliminated. We also clarified with the Randomized Quantile Residuals (RQR) for any unnatural tendencies in the estimates.

3. Standardized CPUE

When estimating the response variable with INLA, it is necessary to create a data set of covariates used for estimation at the same time as parameter estimation. However, the Japanese longline data is so large, and this function did not work. Therefore, we extracted the posteriors' mean for both the required fixed and random effects and multiplied them by all year, quarter, and location combinations to calculate the annual average CPUE.

The same calculation is performed for the standard error (SE). However, instead of this method, it is necessary to calculate the parameters' variance-covariance matrix and calculate the approximate value of SE by the delta method (probably INLA outputs variance-covariance matrix).

Result and discussion

We built multiple models and attempted to standardize CPUE for Pacific blue marlin caught by Japanese longline fishery. We first tried to standardize in the same area as the habitat model, but the SPDE model could not estimate the fixed effect parameters. Then, we cut out the area where the juvenile fish was caught corresponding to the 2016 stock assessment's selectivity and conducted the analysis.

For the WAIC, the spatiotemporal ZIP model was the smallest, followed by the seasonal SPDE model (Table. 2). However, the spatiotemporal ZIP model could not estimate the fixed effect (e.g., year effect)(Table.2). Thus, we selected the seasonal SPDE model, where all parameters could be estimated. Using the selected model, we confirmed the residual tendency of the seasonal SPDE model. The RQR for the predicted value did not vary uniformly and showed an unnatural tendency (Figure. 2). Looking at the RQR by year, the median was biased negative in the 1990s (Figure. 3). It indicates that the predicted value was overestimated. Focus on the RQR in 1994 and 2018, the tendency of the RQR showed a difference (Figure 4). Also, the estimated seasonal spatial field showed that the location changes seasonally (Figure 5). From these results, this model can explain the spatial effect that changes with the seasons. However, it was considered that we need to consider the annual fluctuation further in the spatial effect. For example, R-INLA can specify AR(k) as an option of the spatiotemporal model, so we can create a model that takes into account the spatial field that fluctuates from year to quarter. However, such complex models are computationally expensive and crash in some cases. Thus, we need a high-performance PC essentially.

The standardized CPUE tended to decline from 1994 to 2018 (Figure 6). However, the nominal CPUE shows the opposite trend (Figure 6). Considering the annual trend of the RQR, the estimated results were overestimated in the 1990s. Therefore, further examination is required for this downward CPUE trend. On the other hand, it is necessary to pay attention to the area range specification. Because standardized CPUE is very much affected by data screening, it is advisable to perform a separate analysis in advance when dividing areas (Ijima and Kanaiwa 2018). From these results, we propose that the analysis results of this standardized CPUE are preliminary, and it is desirable to use the habitat model for the next stock assessment.

For the future study, we need to i) examine the "Area as Fleet Approach" and ii) build a model that incorporates seasonal fluctuations into the spatiotemporal model.

References

ISC 2016. Stock Assessment Update for Blue Marlin (*Makaira nigricans*) in the Pacific Ocean through 2014. ISC16 Plenary Report

Kai, M., Okamoto, H., Shiozaki, K., and Hinton, M. 2016. Update of Japanese longline CPUE for Blue Marlin Makaira nigricans in the Pacific Ocean standardized applying habitat model. ISC/16/BILLFISHWG-1/06

Kai, M., Thorson, J.T., Piner, K.R. and Maunder, M.N., 2017. Spatiotemporal variation in size-structured populations using fishery data: an application to shortfin mako (*Isurus oxyrinchus*) in the Pacific Ocean. Canadian Journal of Fisheries and Aquatic Sciences, 74(11), pp.1765-1780.Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. Journal of Statistical Software, 63(19), 1-25.

Waterhouse, L., Sampson, D.B., Maunder, M. and Semmens, B.X., 2014. Using areas-as-fleets selectivity to model spatial fishing: asymptotic curves are unlikely under equilibrium conditions. Fisheries research, 158, pp.15-25.

Minami, M., Lennert-Cody, C. E., Gao, W., & Román-Verdesoto, M. (2007). Modeling shark bycatch: the zero-inflated negative binomial regression model with smoothing. Fisheries Research, 84(2), 210-221.

Ijima, 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

Table 1. Candidate models.

Data Area	Model	Distribution	Model description
WCPO	po_yrFIX_glm	PO	bum ~ yr2 + qtr + gear + latlon5
WCPO	po_yrFIX_GLMM	PO	bum ~ yr2 + qtr + gear + f(latlon5, model="iid", hyper=hcprior) + f(jp_name2, model="iid", hyper=hcprior)
WCPO	zip_yrFIX_GLMM	ZIP	bum ~ yr2 + qtr + gear + f(latlon5, model="iid", hyper=hcprior) + f(jp_name2, model="iid", hyper=hcprior)
WCPO	zip_rw1_GLMM	ZIP	bum ~ f(yr, model="**rw1**", hyper=hcprior) + qtr + gear + f(latlon5, model="iid", hyper=hcprior) + f(jp_name2, model="iid", hyper=hcprior)
WCPO	zip_ar1_GLMM	ZIP	bum ~ f(yr, model="**ar1**", hyper=hcprior) + qtr + gear + f(latlon5, model="iid", hyper=hcprior) + f(jp_name2, model="iid", hyper=hcprior)
WCPO	zip_ar1_SPDE	ZIP	bum ~ 0 + intercept + f(yr, model="**ar1**", hyper=hcprior) + qtr + gear + f(w, model=spde) + f(ves_eff, model="iid", hyper=hcprior)
All Area	zip_rw1_SPDE	ZIP	bum ~ 0 + intercept + f(yr, model="**rw1**") + gear + f(w, model=spde) + f(ves_eff, model="iid")
WCPO	zip_yrFIX_SPDE	ZIP	bum ~ 0 + yr + qtr + f(w, model=spde) + f(ves_eff, model="iid", hyper=hcprior)
All Area	p_GLMM_SPDE	PO	bum ~ 0 + intercept + yr + qtr + gear + f(w, model=spde) + f(ves_eff, model="iid")
WCPO	zip_ar1_SpTmp_25yrs	ZIP	bum ~ 0 + intercept + f(yr, model="**ar1**", hyper=hcprior) + qtr + gear + f(w, model=spde, group=w.group, control.group=list(model="ar1", hyper=h.spec))+ f(ves_eff, model="iid", hyper=hcprior) *w.group = 25 yrs
WCPO	zip_ar1_SpTmp_4season	ZIP	bum ~ 0 + intercept + f(yr, model="ar1", hyper=hcprior) + f(w, model=spde, group=w.group, control.group=list(model="iid")) + f(hpb, model="iid", hyper=hcprior) + f(ves_eff, model="iid", hyper=hcprior) *w.group = 4 seasons

Table 2. Result of parameter estimation by INLA.

Model	WAIC	Dispersion	Conversion (checking the posterior distribution)
po_yrFIX_glm	1.00E+18	6.842779	yes
po_yrFIX_GLMM	786,187	3.600052	yes
zip_yrFIX_GLMM	739,485	2.096098	yes
zip_rw1_GLMM	739,485	2.095213	no(yr) yes(other parameters)
zip_ar1_GLMM	739,484	2.095476	no(yr) yes(other parameters)
zip_ar1_SPDE	732,146	2.035494	no (gear & intercept & yr_precision) yes(qtr, vessel effect, w, zero prob)
zip_rw1_SPDE	1,630,475	1.627742	no (yr_precision) yes(gear & intercept, vessel effect, w, zero prob)
zip_yrFIX_SPDE	732,158	2.036302	yes
p_GLMM_SPDE	1,170,076	86.09476	no (yr, gear, intercept), yes(qtr, ves_eff, w)
zip_ar1_SpTmp_25yrs	**688,082**	1.728719	yes(yr, ves_eff, w, ZeroProb) no(gear, qtr, intercept)
zip_ar1_SpTmp_4season	**702,962**	1.867761	yes

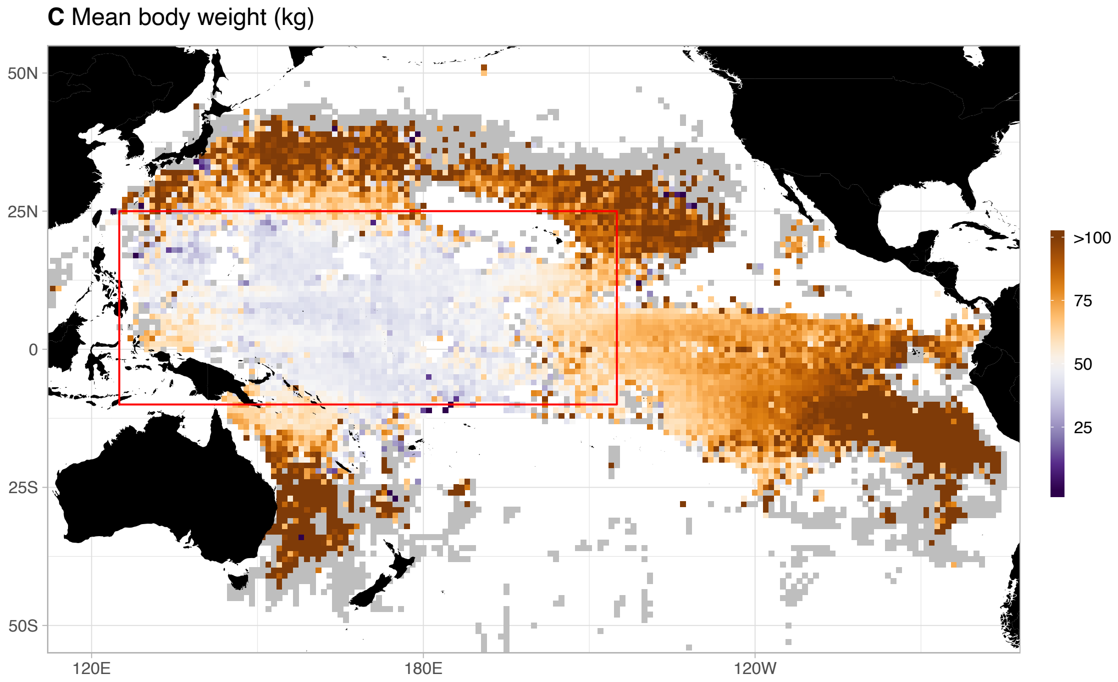

Figure 1. Analysis area in this study.

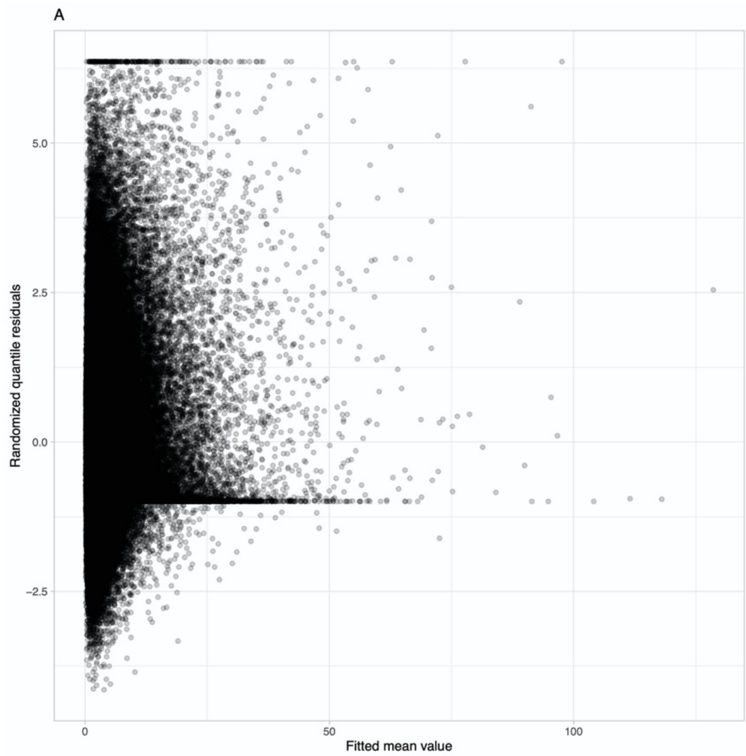

Figure 2. Randomized quantile residuals of zip_ar1_SpTmp_4season model.

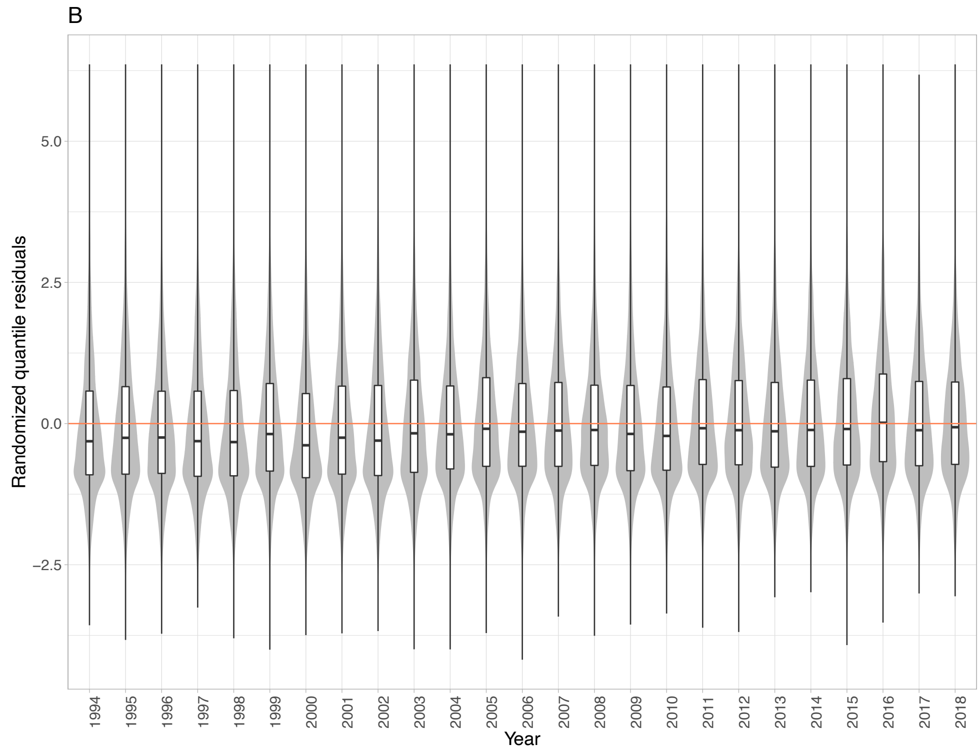

Figure 3. Randomized quantile residuals of zip_ar1_SpTmp_4season model summarized by year.

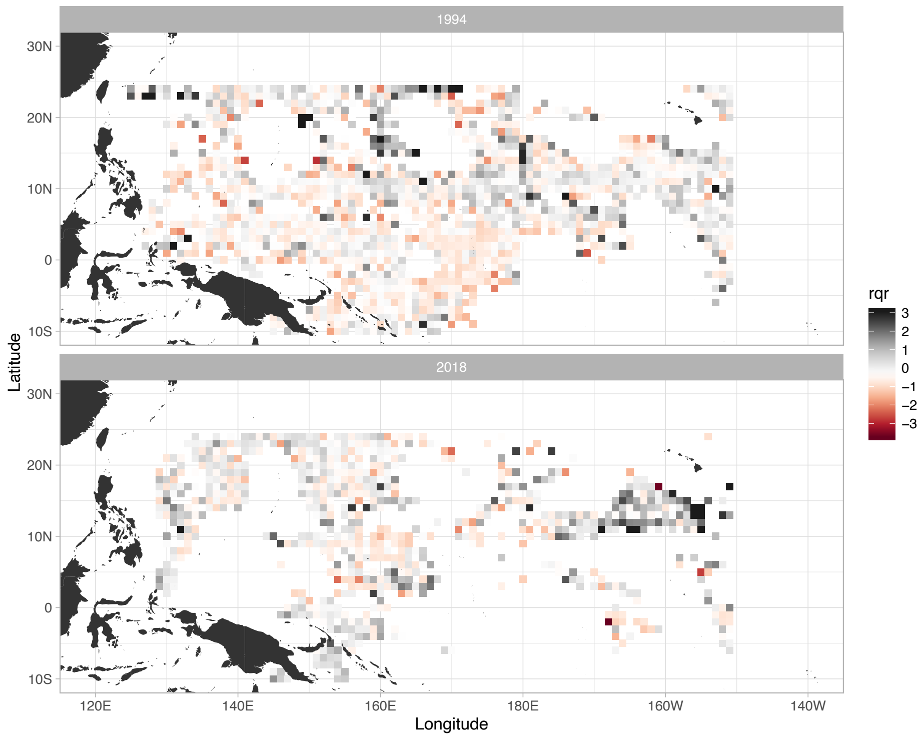

Figure 4. Randomized Quantile Residuals of zip_ar1_SpTmp_4season model in 1994 and 2018.

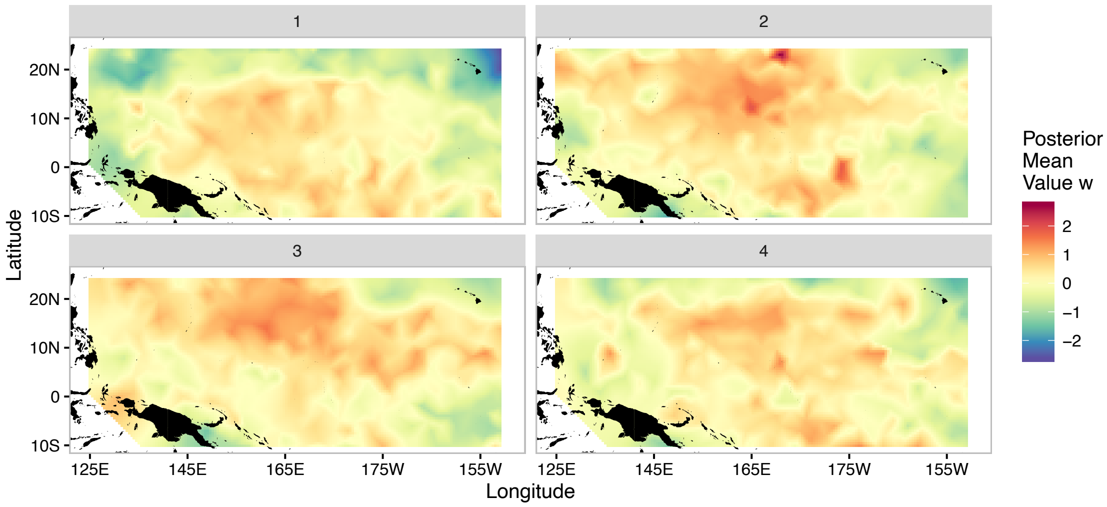

Figure 5. Estimated spatial field.

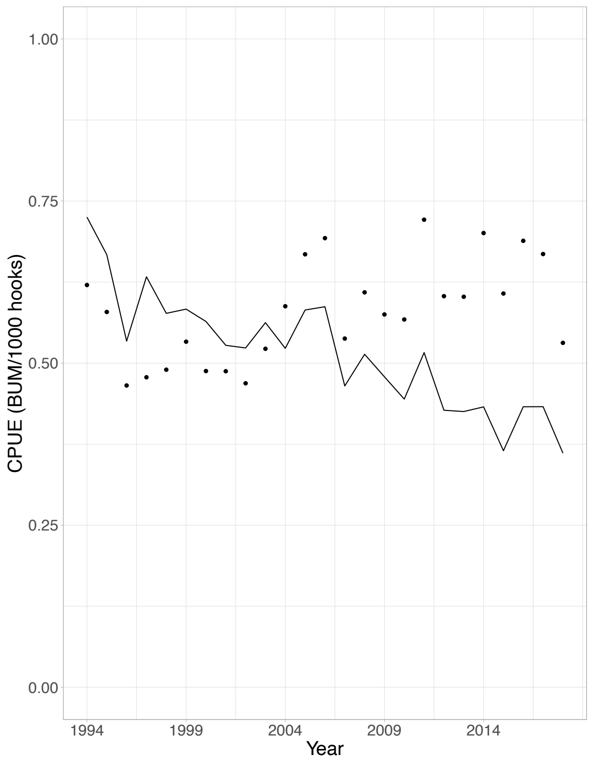

Figure 6. Standardized CPUE. Solid line denote standardized CPUE using xx model. Dotted value is the nominal CPUE.