Japanese Longline CPUE Standardization (1979-2022) for black marlin (*Makaira indica*) in the Indian Ocean using Bayesian hierarchical spatial model

Takayuki Matsumoto, Kenji Taki, Hirotaka Ijima and Mikihiko Kai

Fisheries Resources Institutes (FRI), Japan Fisheries Research and Education Agency

Summary

CPUE of black marlin caught by Japanese longliners during 1979-2022 was standardized. Area definition is the same as that in the previous studies. Time-period was divided into two, 1979-1993 and 1994-2022. Bayesian hierarchical spatial models were applied. Considering high zero catch ratio, zero-inflated Poisson generalized linear mixed model (ZIP-GLMM) was used with the R-INLA package. Best model was selected from multiple models mainly using Widely Applicable Bayesian Information Criterion (WAIC). Gradual annual declining trend with interannual variation were observed for the standardized CPUE during 1979-1993, while stable annual trends were observed for that during 1994-2022. The trend of the CPUE for 1994-2022 was similar to that for the previous study.

1. Introduction

The IOTC Working Party on Billfish (WPB) conducted the stock assessment of black marlin (*Makaira indica*) in the Indian Ocean. In this stock assessment, production models such as ASPIC, BSPM and JABBA (Yokoi and Nishida 2016, Andrade 2016, Parker 2021) were used. Ijima (2018) and Taki et al. (2021) standardized the CPUE of black marlin caught by Japanese longliners in the Indian Ocean using zero-inflated Negative Binomial generalized linear mixed model (ZINB-GLMM) without considering the spatial random effect. It is generally thought that the abundance indices of Japanese longliners are very critical for the stock assessment.

Integrated nested Laplace approximations (INLA) methodology and its powerful application to the modelling of complex datasets have recently been introduced to a wider nontechnical audience (Illian et al. 2013). As opposed to Markov Chain Monte Carlo (MCMC) simulations, INLA uses an approximation for inference and hence avoids the intense computational demands, convergence, and mixing problems that are sometimes encountered by MCMC algorithms (Rue and Martino 2007). Additionally, R-INLA includes the stochastic partial differential equations (SPDE) approach (Lindgren et al. 2011) which is another statistical development. This approach enables us to model spatial random effect (Gaussian random field, GRFs) and to construct flexible fields that are better adept to handle datasets with complex partial structure (Lindgren and Rue 2013). This is often the case with fisheries data, since fishermen tend to aggregate particular fishing grounds, resulting in clustered spatial patterns and a lack of data at large regions. Together, these new statistical methods and their implementation in R allow scientists to fit considerably faster and more reliably complex spatiotemporal models (Rue et al. 2009, Cosandey-Godin et al. 2015).

The aim of this paper is to estimate the annual trends in abundance indices of black marlin (*Makaira indica*) caught by Japanese longliners in the Indian Ocean from 1979 to 2022 for the stock assessment of this species using the same method as that in the previous study (Taki et al. 2021). A zero-inflated Bayesian hierarchical approach was applied in consideration with spatial changes in the fishery and the species.

2. Materials and methods

Data sets

Japanese longline logbook data was used for the CPUE standardization of black marlin in the Indian Ocean. The logbook data has information about the resolution of fishing location at 1 x 1 degree grid scale. We used the data from 1979 onwards because the number of hooks between floats and the vessel name, which largely affect the CPUE standardization, are completely available since then. We divided the time-period into two, 1979-1993 and 1994-2022, as the gear configuration of Japanese longline fishery such as number of hooks between floats and gear material had drastically changed in the early period of 1990s. At the same time, the quality and quantity of logbook data were improved by adding new items to the logsheet as well. We defined the same area of the analysis as Ijima (2018) and Taki et al. (2021), considering spatial CPUE and body weight information (Figure 1). In this area, Japanese longliners tended to catch similar body weight of black marlin in all time. Japanese longliners have operated throughout the Indian Ocean from the 1990s to the 2000s, but in the 2010s, the fishing ground was shrunk rapidly (Figure 2). There are two main reasons for this, that is, the influence of piracy activities in the northwest Indian Ocean, and the target shift of fishermen to southern bluefin tuna in the Southern Indian Ocean. The target shift makes it difficult to catch black marlins staying frequently in the shallower depths.

Statistical models

We applied Bayesian hierarchical spatial models, but we did not directly consider the spatiotemporal effects in the model because this approach is computationally intensive and the Widely Applicable Bayesian Information Criterion (WAIC; Watanabe, 2012) did not differ so much between spatial and spatiotemporal models in the preliminary analysis. Since the catch data is countable and characterized by many zeros (Figure 3), we used a zero-inflated Poisson GLMM (ZIP-GLMM). The zero-inflated model is useful because it can estimate "true" zero catch. As an alternative way, it is possible to use ZINB-GLMM but we did not use the model because the ZINB tended to cause underdispersion (Ijima and Kanaiwa, 2019).

The explanatory variables of fixed effect are year (yr) and quarter (Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec; qtr), and those of random effect are area (5 x 5 degree scale; latlon), month (month), vessel ID (vessel name; jp_name), and gear configuration (number of hooks between floats; hpb). The hpb increased remarkably in the early 1990s in the defined area (Figure 4). Most variables were treated as the categorical variable, but the autoregressive model (AR1) was applied to year effect for two spatial models to consider the autocorrelation. The latest SPDE models using AR1 tended to show smaller WAIC as compared to those using year as fixed effect (e.g., Ijima and Koike 2020). The use of these random effects in the model seems more appropriate to raise the accuracy of estimation (Ijima and Kanaiwa 2019). The random effects are also expected to remove the pseudo-replication by each effect (vessel, gear configuration, month, and area).

All analyses were performed using R, specifically the R-INLA package. The INLA procedure, in accordance with the Bayesian approach, calculates the marginal posterior distribution of all random effects and parameters involved in the model. We applied a half Cauchy distribution as a prior for the random effect. We plotted latent spatial field to indicate the expected CPUE distribution. Best candidate model was selected based on WAIC for the defined area in each period.

3. Result and discussion

We compared the WAIC among eight different structurer's models for the defined area and each period (Table 1). The best model (yellow marker) was selected based on the lowest WAIC and by confirming that the results are reasonable (i.e. credible interval for CPUE is not too broad).

The predicted CPUE was higher in the northwestern coastal part in the defined area during 1979-1993, while that was lower for the same part during 1994-2022 (Figure 5). The annual predicted CPUE showed a gradual decline trend for 1979-1993, while no apparent trend was observed for 1994-2022 with a recent decrease. (Figure 6, Table 2). The 95% credible intervals were wide due to the inclusion of spatial effect (Taki et al., 2021).

Figure 7 shows a comparison of annual trend of standardized CPUE (relative values) between present and previous (Taki et al., 2021) studies for the defined area. The trend of CPUE is usually very similar between two studies. There was some difference in CPUE for 2019 (terminal year in the previous analysis). This may be due to updating catch and effort data, in which 2019 data was preliminary at the previous analysis.

4. References

Andrade, H.A. 2016. Preliminary stock assessment of black marlin (*Makaira indica*) caught in the Indian Ocean using a Bayesian state-space production model. IOTC–2016–WPB14–28, 16 pp.

Cosandey-Godin, A., Krainski, E. T., Worm, B. and Flemming, J. M. 2015. Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. Can. J. Fish. Aquat. Sci. 72: 186–197. dx.doi.org/10.1139/cjfas-2014-0159.

Ijima, H. 2018. Standardized CPUE of the Indian Ocean black marlin (*Makaira indica*) caught by Japanese longline fisheries. IOTC-2018-WPB16-26.

Ijima, H. and Kanaiwa, M. 2019. Japanese longline CPUE of the striped marlin (*Kajikia audax*) in the WCNPO. ISC/19/BILLWG-01/07.

Ijima, 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.

Illian, J.B., Martino, S., Sørbye, S.H., Gallego-Fernández, J.B., Zunzunegui, M., Paz Esquivias, M., and Travis, J.M.J. 2013. Fitting complex ecological point process models with integrated nested Laplace approximation. Methods Ecol. Evol. 4(4): 305–315. doi:10.1111/2041-210x.12017.

Lindgren, F., and Rue, H. 2013. Bayesian spatial and spatio-temporal modelling with R-INLA. Journal of Statistical Software. [In press.] Rue, H., and Martino, S. 2007. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. J. Stat. Plann. Infer. 137(10): 3177–3192. doi:10.1016/j.jspi.2006.07.016.

Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73(4): 423–498. doi:10.1111/j.1467-9868.2011.00777.x.

Parker, D. 2021. Assessment of the Indian Ocean black marlin (*Makaira indica*) stock using JABBA. IOTC-2021-WPB19-18, 17 pp.

Rue, H., and Martino, S. 2007. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. J. Stat. Plann. Infer. 137(10): 3177–3192. doi:10.1016/j.jspi.2006.07.016.

Rue, H., Martino, S., and Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71(2): 319–392. doi:10.1111/j.1467-9868. 2008.00700.x.

Taki, K., Ijima, H. and Kai, M. 2021. Japanese Longline CPUE Standardization (1979-2019) for black marlin (*Makaira indica*) in the Indian Ocean using Bayesian hierarchical spatial model. IOTC-2021-WPB19-26.

Watanabe, S. 2013. A widely applicable Bayesian information criterion. Journal of Machine Learning Research 14: 867-897.

Yokoi, H., and Nishida, T. 2016. Stock assessments of black marlin (*Makaira indica*) in the Indian Ocean using A Stock-Production Model Incorporating Covariates (ASPIC). IOTC–2016–WPB14– 24_Rev1, 13 pp.

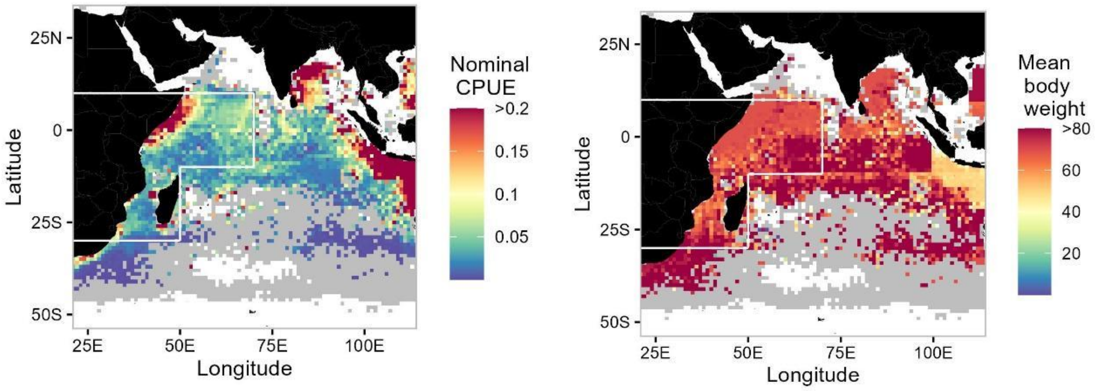

Figure 1. Spatial distributions of nominal CPUE and mean body weight for black marlin caught by Japanese longliners in the Indian Ocean. The area used for CPUE standardization (water inside of white line) in the present study.

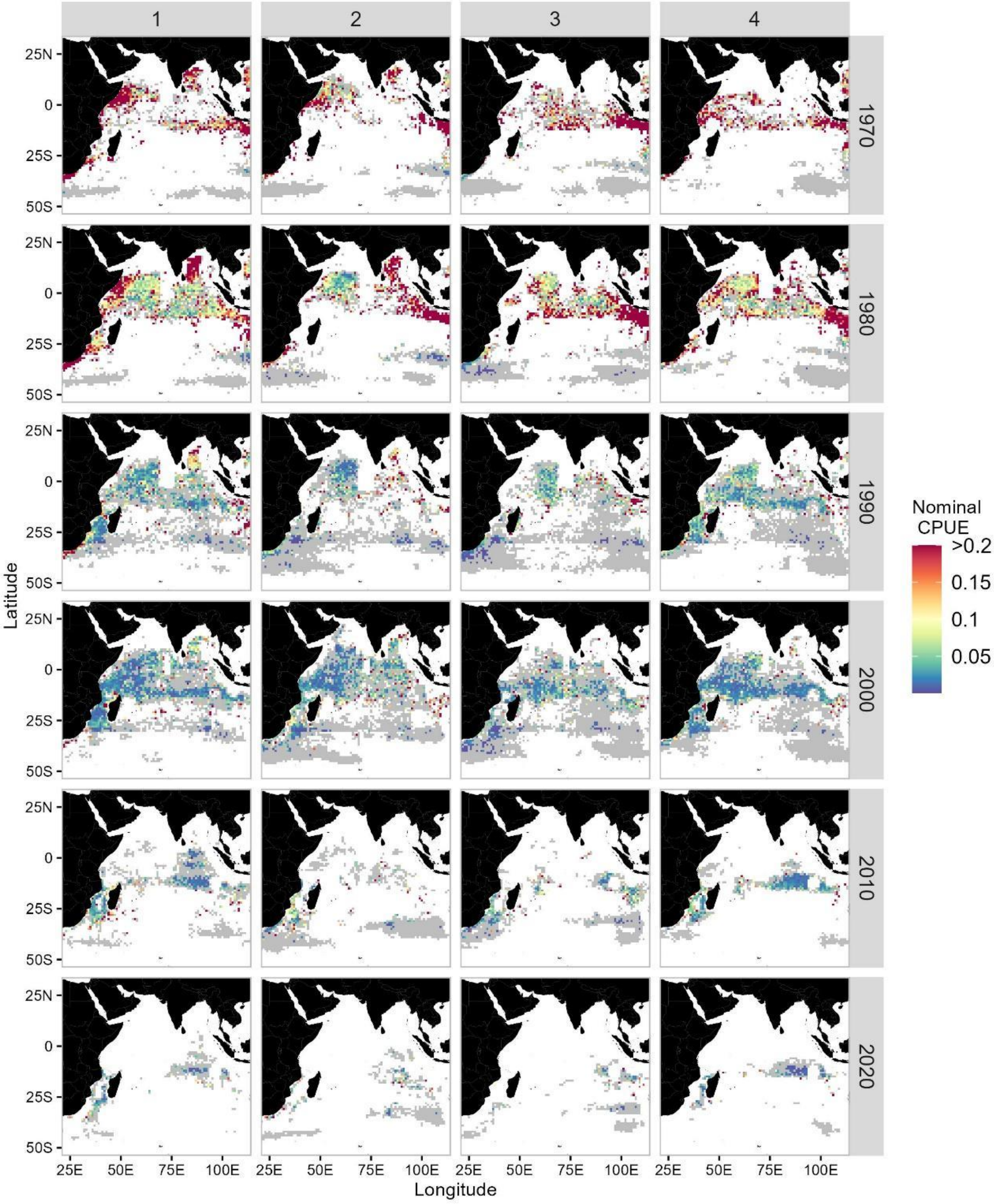

Figure 2. Spatial-temporal (seasonal and decadal) changes in the nominal CPUE for black marlin caught by Japanese longliners in the Indian Ocean. 1: Jan-Mar, 2: Apr-Jun, 3: Jul-Sep, 4: Oct-Dec.

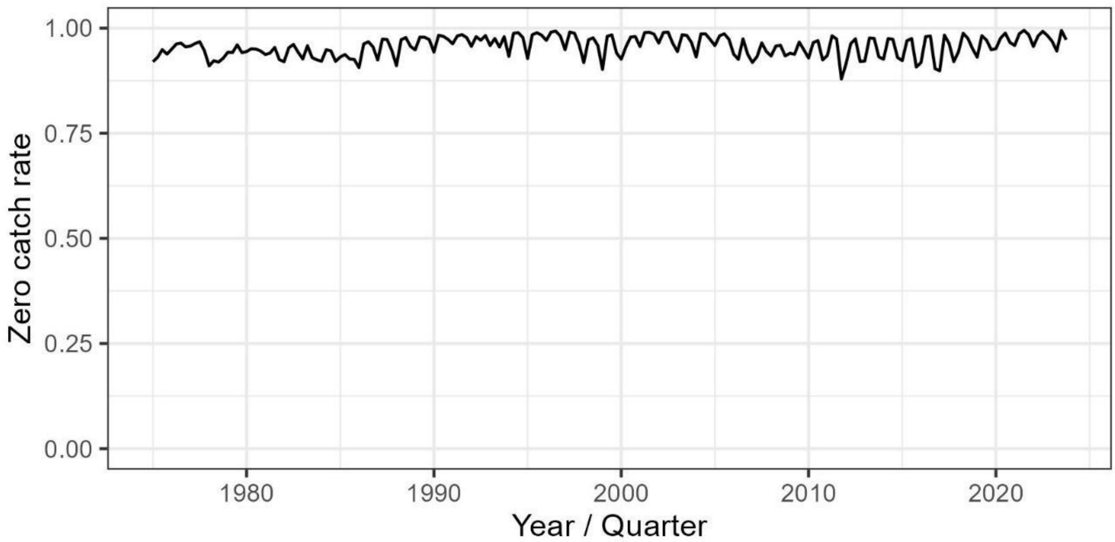

Figure 3. Annual changes in zero catch ratio of black marlin caught by Japanese longliners in the defined area of the Indian Ocean.

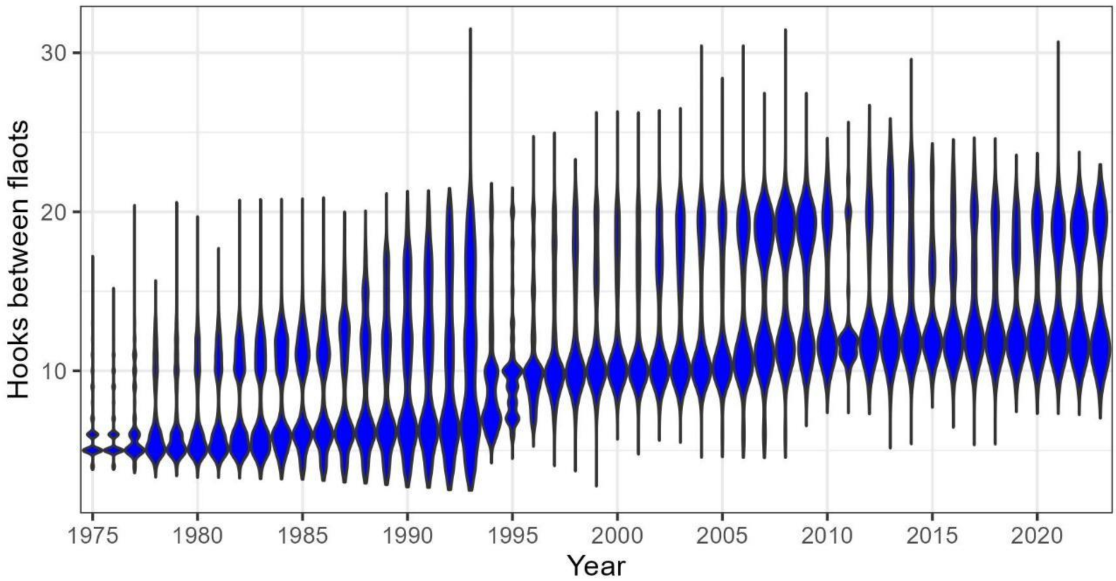

Figure 4. Historical changes in the gear configuration (number of hooks between floats) in the defined area of the Indian Ocean. Vertical range of the plots shows the range of the data, and width shows frequency of the data.

Table 1. The models and their WAIC values for two time periods of analyses. Selected models corresponded to those with the smallest values yellow-highlighted.

Model*	1979-1993	1994-2022
m_null = inla (blm~1, data=d,offset=log(hooks/1000), family="poisson")	57409	75288
m_glm = inla (blm~yr + latlon,data=d,offset=log(hooks/1000), family="poisson")	>10^{19}	75092
m_glmm = inla (blm~yr + qtr + f(latlon,model="iid", hyper=hcprior) + f(jp_name,model="iid")+f(hpb,model="iid"), data=d,offset=log(hooks/1000), family="poisson")	51434	75219
m_zip_glmm = inla (blm~yr + qtr + f(latlon, model="iid") + f(jp_name, model="iid"), data=d,offset=log(hooks/1000), family="zeroinflatedpoisson1")	51070	68970
m_spde = inla (blm~0 + intercept + yr + qtr + f(hpb, model="iid") + f(jp_name, model="iid") + f(w,model=spde), data=inla.stack.data(StackFit), offset=log(hooks/1000), family="poisson")	51063	73203
m_spde2 = inla (blm~0 + intercept + f(yr,model="ar1") + f(month, model="iid",hyper=hcprior) + f(hpb,model="iid",hyper=hcprior) + f(jp_name, model="iid",hyper=hcprior) + f(w,model=spde), data=inla.stack.data(StackFit2),offset=log(hooks/1000),family="poisson")	50948	73081
m_zip_spde = inla (blm~0 + intercept + yr + qtr + f(hpb,model="iid") +f(jp_name,model="iid") + f(w,model=spde), data=inla.stack.data(StackFit), offset=log(hooks/1000), family="zeroinflatedpoisson1")	50590	68726
m_zip_spde2 = inla (blm~0 + intercept + f(yr, model="ar1") + f(month, model="iid",hyper=hcprior) + f(hpb, model="iid") + f(jp_name, model="iid") + f(w, model=spde), data=inla.stack.data(StackFit2),offset=log(hooks/1000),family="zeroinflatedpoisson1")	50483	68654

* blm: catch of black marlin in number, hooks: number of hooks, yr: year, qtr: quarter, latlon: 5 x 5 degree latitude and longitude, hpb: number of hooks between floats, jp_name: vessel ID (vessel name), iid: Gaussian random effects, ar1: auto-regressive model of order 1, spde: stochastic partial differential equations, hyper: hyperparameters, hcprior: halfcauchy prior, family: likelihood family, d: catch and effort data set used in the program code. StackFit, StackFit2: stacked data for INLA.

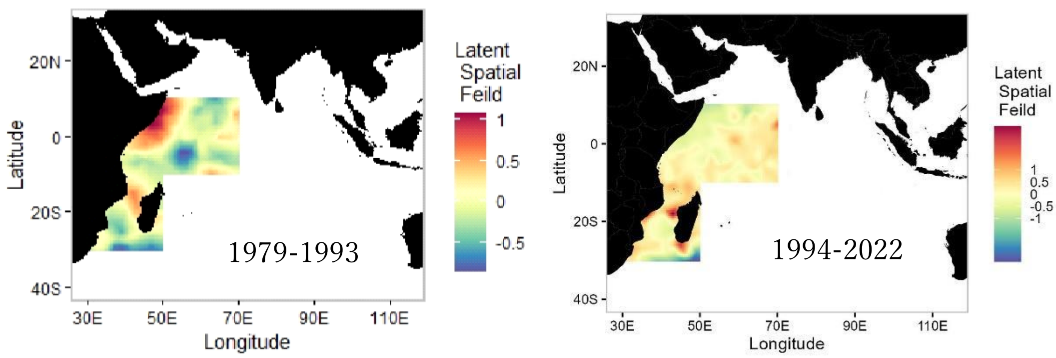

Figure 5. Spatial distribution in standardized CPUE (mean latent spatial field) of black marlin for two periods in the defined area of the Indian Ocean.

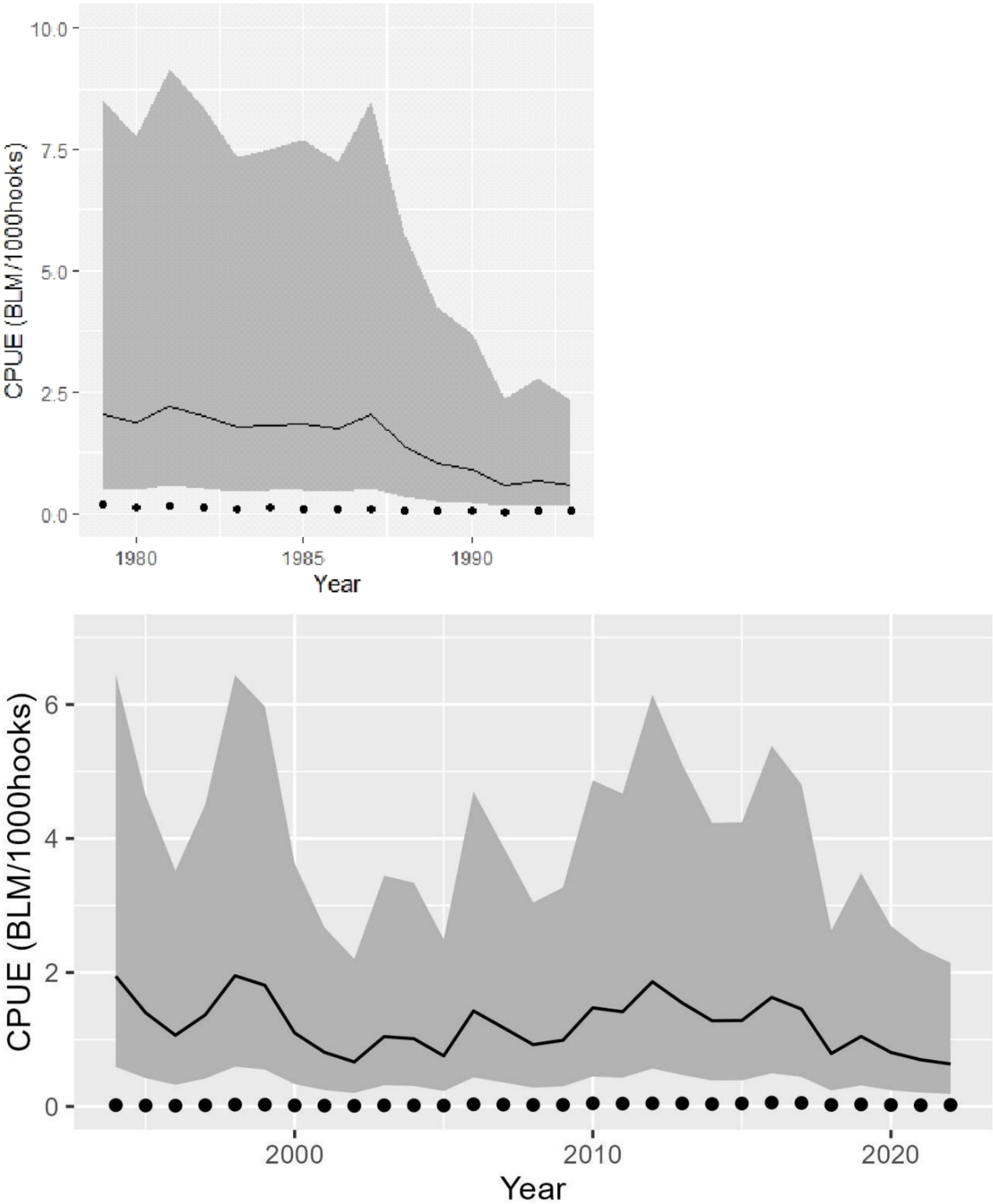

Figure 6. Historical changes in the CPUEs of black marlin for two periods in the defined area of the Indian Ocean. Thin line and filled points denote point estimates of standardized and nominal CPUE, respectively. Gray shadows denote 95% credible intervals. Note that the scale of y-axis is different between upper and lower figures.

Table 2. Nominal and standardized CPUEs of black marlin for two periods; 1979-93 and 1994-2022 in the defined area of the Indian Ocean.

year	nominal	Standardized	2.50%	97.50%	year	nominal	Standardized	2.50%	97.50%
1979	0.19	2.05	0.52	8.52	1994	0.02	1.94	6.45	0.59
1980	0.12	1.87	0.47	7.77	1995	0.01	1.40	4.64	0.42
1981	0.14	2.21	0.56	9.16	1996	0.01	1.06	3.52	0.32
1982	0.11	2.02	0.51	8.36	1997	0.02	1.37	4.51	0.42
1983	0.09	1.77	0.45	7.35	1998	0.03	1.95	6.44	0.59
1984	0.10	1.81	0.46	7.51	1999	0.03	1.81	5.97	0.55
1985	0.09	1.85	0.47	7.68	2000	0.02	1.10	3.63	0.33
1986	0.08	1.75	0.44	7.24	2001	0.01	0.81	2.67	0.25
1987	0.08	2.04	0.52	8.47	2002	0.01	0.66	2.20	0.20
1988	0.06	1.39	0.35	5.79	2003	0.02	1.04	3.44	0.32
1989	0.04	1.03	0.26	4.26	2004	0.02	1.01	3.34	0.31
1990	0.04	0.89	0.22	3.68	2005	0.01	0.76	2.50	0.23
1991	0.03	0.57	0.14	2.37	2006	0.03	1.43	4.70	0.43
1992	0.04	0.67	0.17	2.79	2007	0.03	1.18	3.88	0.36
1993	0.04	0.56	0.14	2.33	2008	0.02	0.92	3.04	0.28
					2009	0.02	0.99	3.27	0.30
					2010	0.05	1.47	4.87	0.45
					2011	0.04	1.41	4.67	0.43
					2012	0.05	1.86	6.15	0.57
					2013	0.05	1.55	5.11	0.47
					2014	0.04	1.28	4.23	0.39
					2015	0.04	1.28	4.24	0.39
					2016	0.06	1.63	5.38	0.49
					2017	0.05	1.45	4.81	0.44
					2018	0.02	0.79	2.63	0.24
					2019	0.03	1.05	3.48	0.31
					2020	0.02	0.81	2.69	0.24
					2021	0.02	0.70	2.35	0.21
					2022	0.02	0.63	2.14	0.19

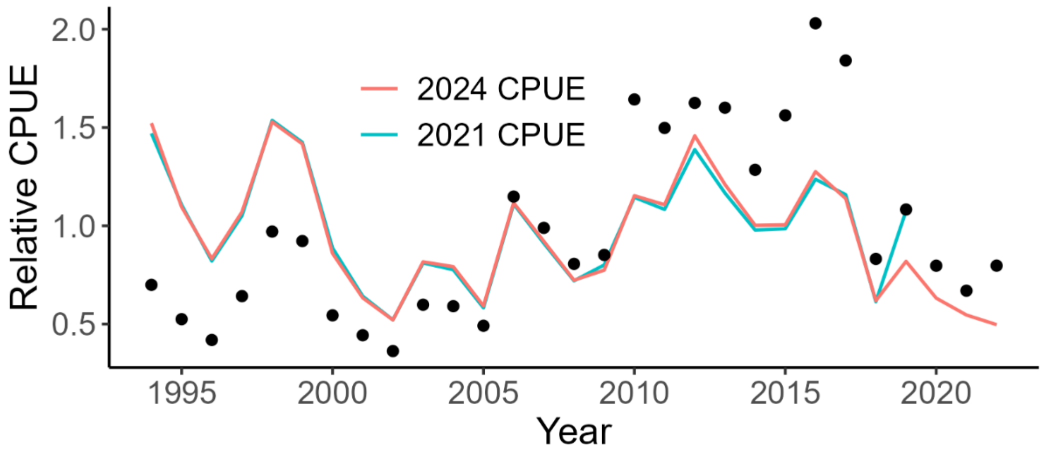

Figure 7. Comparison of annual standardized CPUE of black marlin (relative to its mean value for 1994-2019) between present (red line) and previous (blue line, Taki et al., 2021) studies for the defined area of the Indian Ocean. Black dots denote nominal CPUE.