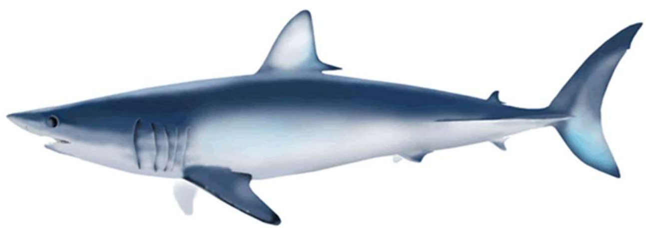
ISC/23/SHARKWG-1/3

Spatio-temporal model for CPUE standardization: Application to shortfin mako caught by longline of Japanese research and training vessels in the western and central North Pacific[1]

Mikihiko Kai[2]

.

[2]Fisheries Resources Institute, Highly Migratory Resources Division,
Japan Fishery Research and Education Agency
2-12-4 Fukuura, Kanazawa, Yokohama, Kanagawa 236-8648, JAPAN
Email: kai_mikihiko61@fra.go.jp



[1] Working document submitted to the ISC Shark Working Group Workshop, 29-30 November- 1-2, 4-7 December 2023, Yokohama, Kanagawa, Japan. **Document not to be cited without author's permission.**

Abstract

This document paper provides annual changes in standardized catch per unit of effort (CPUE: catch number per 1,000 number of hooks) for shortfin mako caught by longline fishery of Japanese training and research vessels during 1994 and 2022 in the western and central North Pacific. Since the reporting rates of sharks during 2001 and 2013 are clearly lower than those before 2000, the author removed the data with lower reporting rates using a statistical filtering method based on the prediction of the binomial generalized linear model (GLM). The nominal CPUE was then standardized using spatio-temporal generalized linear mixed models (GLMMs) to provide the annual changes in the abundance indices in the North Pacific Ocean. The estimated abundance indices of shortfin mako revealed a flat trend from 1994 to 2005, and then showed two times up- and down- trends for 2009-2013, 2013-2017 and was stable thereafter. The CPUE trends estimated from the fishery-independent data widely collected in the North Pacific Ocean is a very useful information about the abundance in this region.

Introduction

The National Research Institute of Far Seas Fisheries in Japan has been collecting the research and training vessel (JRTV) data since 1992. The JRTV data was collected from the research vessels belonging to, or chartered to, national or prefectural fisheries research institutes, and vocational training vessels attached to fisheries high schools throughout Japan. The JRTV commonly operates the water around Hawaii due to reputedly calm sea conditions and the attractiveness to students of Honolulu port call (WCPFC, 2011). Although this survey is not well designed for spatiotemporal changes in the operational patterns, this survey is fishery independent and there is no issue about the targeting shift and the significant differences of the catchability by ships. In addition, it is expected to report the data with accuracy. However, past examination of the data revealed that reporting rate for sharks defined by the number of sets recorded with sharks divided by the total number of sets for a trip (Nakano and Clarke, 2006) appeared to decrease after 2000, and it suggested that JRTV had released or discard sharks without recording them (WCPFC, 2011).

In the previous stock assessment in 2017, a statistical filtering method was used to remove unreliable set-by-set data collected by JRTVs during 2001 and 2013 (Kai, 2017a). The nominal CPUE of the JRTVs was then standardized using two-part model (Zuur et al., 2009) to account for the occurrence of excess zeros (Annual mean: 85% of zero catch) and small dispersion ratios (variance/mean) of the catch for shortfin mako (Annual mean: 1.54). A binomial and a Poisson (PO) generalized linear model (GLM) was applied to the first and second stage of the two-part model, respectively. The two response variables (positive catch ratio and mean CPUE for positive catch) were then combined to calculate the annual trends in the abundance indices of shortfin mako in the western and central North Pacific. The CPUE was used in the stock assessment in 2018 based on the statistical soundness, long time-span, extensive spatial coverage, and reliability of record. However, one of the issues was insufficient CPUE modeling regarding the main interaction term such as a year and area due to a lack of data for some subarea in some years. The subareas used in the GLM were too large to explain the influence of the spatio-temporal changes in the catch rate. To address this issue, spatio-temporal generalized linear mixed model (GLMM) was applied to the JRTV data (Kai, 2019). The spatio-temporal model enables us to predict the spatial changes in species distribution and temporal variations in a population range and density in a fine scale such as a resolution of 1 x 1 degrees, based on spatial and temporal autocorrelation among catch rates and correlations with various biotic and abiotic environmental factors (Thorson, 2019; Thorson and Barnett, 2017; Kai, 2017b). The spatiotemporal model therefore may yield more precise, biologically reasonable, and interpretable estimates of abundance than the commonly used design-based models or spatially stratified models (Shelton et al., 2014; Thorson et al., 2015; Cao et al., 2017).

The main objective of this study is to provide the annual changes in catch rates of shortfin mako in the western and central North Pacific using the spatio-temporal GLMMs with fishery-independent data (i.e., longline logbook data of JRTVs). This information may contribute to improvements in the stock assessments of North Pacific shortfin mako through an understanding of the spatial and temporal changes in the hotspots and temporal changes in catch rates. Firstly, temporal changes in the reporting rate are analyzed and unreliable set-by-set data are removed using the same statistical filtering method as applied in the previous analysis. Secondly, the nominal CPUE is standardized using the spatio-temporal GLMMs for filtered data during 1994 and 2022.

Materials and Methods

Data sources

This study used the longline logbook data mainly collected from JRTVs in the western and central north Pacific (mainly 0–40 °N and 130 °E–140 °W) from 1994 to 2022 (**Fig. 1**). To be consistent with the time-period of late time series of Japanese longline fishery, the JRTVs data after 1993 was used. Set-by-set operational data used in this study includes information on species of pelagic sharks, operation time (year, month), catch numbers, amount of effort (number of hooks), number of branch lines between floats (hooks between floats: HBF) as a proxy for gear configuration, location of sets by latitude-longitude resolution of 1° × 1°, and trip identity. Deep-set data was used in this analysis because the JRTVs mostly use deep-sets. A deep-set is identified by the number of HBF, which determines fishing depth (Nakano et al., 1997). A deep-set fishery was defined as one that uses a large number of HBF (6–16 hooks). The number of HBF with the most catches for SFM was between 12 and 13, and a small change in gear configuration was observed (**Fig. A1**). The four seasons (quarters (Q) 1 to 4) of the year were defined as follows: Q1 was spring from January to March; Q2 was summer from April to June; Q3 was fall from July to September; and Q4 was winter from October to December.

Data filtering

Incomplete and insufficient data were filtered, as were sets that have little or no information about HBF and locations (latitude and longitude), numbers of hooks that were less than 800, HBF that were less than 6 (i.e., shallow-sets), and operations that were conducted in waters other than the North Pacific. In this document paper, this filtering step is referred to as "preliminary filtering". In addition, to remove errors and biases of the set-by-set data caused by under-reporting of actual shark catches, unreliable set-by-set data were further removed based on the information on shark presence in the catch (Kai, 2019). The author applied the statistical filtering method based on a GLM with binomial error distribution to JRTV data from 2001 to 2013 to accommodate a clear decline in annual reporting rates during this period (upper figure of **Fig. 2**). In this document paper, this filtering step is referred to as "follow-up filtering". The details of the filtering method can be seen in the previous papers (Kai, 2017a; Kai, 2019)

CPUE standardization with spatio-temporal model

In the previous analysis (Kai, 2019), the zero-inflated Poisson (ZIP) and zero-inflated negative-binomial (ZINB) model were used, and the ZINB model was selected as the most parsimonious model. ZINB model is therefore solely used in this study. The spatio-temporal ZINB model is consisted of two components of encounter probability and positive catch in a delta model. The first predictor was models using a binomial model to account for the encounter probability of low positive catch (mean positive catch rate = 15 %). However, the random effects were not used because of the convergence issue. Second predictor was modeled using a negative binomial (NB) model to account for the count data with over-dispersion (variance/mean =1.52):

$$c \sim NegBin\ (c^{*}, c^{*}(1+\sigma_1) + {c^{*}}^{2}\sigma_2),$$

$$\log\ (d) = d_0(t) + \gamma(s) + \theta(s,t) + \varepsilon(v) + \sum_{j=1}^{n_j} \beta(j) \times x(j), \quad (1)$$

where c is observed catch, *NegBin* (a, b) is a negative binomial distribution with mean a and variance b (Lindén and Mäntyniemi, 2011), c^{*} is an expected catch and a function of density d and fishing effort f (number of hooks = 1), σ_1 and σ_2 are residual variations, $d_0(t)$ represents temporal variation (the intercept for each year t), $\gamma(s)$ represents spatial variation (s), $\theta(s,t)$ represents spatio-temporal variation (station s and year t), $\varepsilon(v)$ represents random variation in catchability for the v th vessel, and β_j represents the impact of covariate j with value x_j on catchability. The three-month quarters and HBF (i.e. $n_j = 2$, $x_j = q\ and\ l$) are used as covariates (changing the catchability) corresponding to Eq. (1).

The VAST (v3.10.1) was used to standardize the nominal CPUE. Temporal abundance index I was estimated as:

$$I(t) = \sum_{s=1}^{n_s} f(s) \times c^{*}(s,t) / \{\sum_{t=1}^{n_t} \sum_{s=1}^{n_s} f(s) \times c^{*}(s,t)\}, \quad (2)$$

where n_s is total number of knots at location s. The number of knots (n_s = 400) was specified in a balance between computational speed and spatial resolution.

Model selection and diagnostics

To select the best model, the explanatory variable was sequentially added to the random effect model. The best model was selected using the AIC (Akaike, 1973) and BIC (Schwarz, 1978). Given the different model is selected by AIC and BIC, the model selected by BIC is chosen to avoid the overfitting that the AIC tends to choose the complex model with a large number of data (Shono, 2005). For the best model, the goodness of fits was examined using the Pearson residuals and QQ-plot. The residuals were computed using a randomized quantile (Dunn and Smyth, 1996) to produce continuous normal residuals.

Results

For the analyses of the follow-up filtering, the model including the factor of month, latitude by 5 degrees, and longitude by 5 degrees was selected by BIC as the most parsimonious model (**Table A1**). The lower and upper 95% confidence intervals of shark reporting reliability (SR) for 1994-2000, 2014–2022 were estimated as 0.925 and 1.263, respectively, and the lower bound was used as a cut-off point (**Fig. A2**). The threshold (i.e., 0.925) appeared to be reasonable because the reduction of catch rates between 2001 and 2013 disappeared (lower panel of **Fig. 2**). The preliminary filtering reduced the number of records for this analysis from 39,571 sets to 35,421 sets. The follow-up filtering reduced the number of records for this analysis from 35,421 sets representing 1,469 trips to 30,803 sets representing 1,256 trips. The differences of annual changes in number of catch, number of hooks, and nominal CPUE between the data with and without follow-up filtering are shown in **Fig. A3**.

Selection of the best model

All models except for M1 and M2 were reasonably converged with the positive definite of hessian matrix and a small value of maximum gradient (**Table 1**). The model (M-7) including spatial (station), spatio-temporal variances (year and station) and overdispersion (vessel effects) as random effects and year and quarter as fixed effects were identified by AIC and BIC as the most parsimonious model (**Table 1**). The estimated CPUE changed substantially if random effect components were sequentially added to the simplest model (M_1) which has no random effect (**Fig. 3**). Diagnostic plots of goodness-of-fit for the best model didn't show serious deviations from normality and model misspecification (**Fig. 4**). These results suggested that the fitting of the best model to the data was good. Lists of all parameters and estimates of the best models are shown in **Table 2**.

Temporal trends in CPUE

The estimated annual changes in the CPUE of shortfin mako revealed a flat trend from 1994 to 2005, and then showed two times up- and down- trends for 2009-2013, 2013-2017 and was stable in recent years (**Fig. 5**). The 95% confidence intervals were wider in 1999, 2007, 2015, and 2021 compared to those in the other years (**Fig. 5**).

Spatio-temporal trends in CPUE

The annual spatial maps of predicted CPUE clearly showed that the higher CPUEs of shortfin mako at the higher latitudes (30-40° N, 130° E -160° W) in the temperate water (**Fig. 6**). Meanwhile, the lower CPUEs of shortfin mako were observed in the sub-tropical and tropical areas, however, the CPUEs in the north-east water near Hawaii islands were higher. These results suggested that the shortfin mako prefer to staying in the temperate water in the western and central North Pacific Ocean.

Discussions

This study presented the annual changes in the standardized CPUE of shortfin mako caught by longline gear of JRTV in the North Pacific from 1994 to 2022. The data with lower reporting rates was removed and the nominal CPUE was standardized using spatio-temporal GLMM models. The filtering appeared to be reasonable because the lower reporting rates by vessel-trip between 2001 and 2013 were disappeared (**Fig. A2**). The results of the standardization of CPUE suggested that the abundance trends of shortfin mako in the North Pacific appeared to be stable in recent years but the CPUE level was lower than those in 1990s and 2000s (**Fig. 5**). Although JRTV mainly operate in the sub-tropical and temperate areas near Hawaii using deep-set longline gear, the result doesn't support the slightly increasing abundance trends in Hawaii deep-set time series (Calvalho, 2021). This inconsistent trends between two CPUEs in addition to the large fluctuations of predicted CPUE from 2009 to 2017 in this study might be attributed to shrinkage of operational areas around Hawaii islands in recent decade due to the continuous decline of fishing effort of JRTV since 2000 (**Fig. A3**).

References

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. *In* Proceedings of the Second International Symposium on Information Theory, pp. 267–281. Eds. by B. N. Petrov and F. Csaki. Akademiai Kiado, Budapest.

Carvalho, F. 2021. Standardized catch rates of shortfin mako shark (*Isurus oxyrinchus*) caught by the Hawaii-based pelagic longline fleet (1995-2019). ISC/21/SHARKWG-1/10

Cao, J., Thorson, J.T., Richards, R.A., Chen, Y., 2017. Spatiotemporal index standardization improves the stock assessment of northern shrimp in the Gulf of Maine. Can. J. Fish. Aquat. Sci. 74, 1781–1793. https://doi.org/10.1139/cjfas-2016-0137.

Dunn, P.K., Smyth, G.K., 1996. Randomized quantile residuals. J. Comput. Graph. Stat. 5, 236–244.

Kai, M. 2017a. CPUE of shortfin mako, *Isurus oxyrinchus*, caught by Japanese research and training vessels in the North Pacific. ISC/17/SHARKWG-3/8

Kai, M., Thorson, J. T., Piner, K. R. and Maunder, M. N. 2017b. Predicting the spatio-temporal distributions of pelagic sharks in the western and central North Pacific. *Fish. Oceanogra.* doi:10.1111/fog-12217

Kai, M., Thorson, J. T., Piner, K. R. and Maunder, M. N. 2017c. Spatio-temporal variation in size-structured populations using fishery data: an application to shortfin mako (*Isurus oxyrinchus*) in the Pacific Ocean. *Can. J. Fish Aquat. Sci.* doi:10.1139/cjfas-2016-0327

Kai, M. 2019. Spatio-temporal changes in catch rates of pelagic sharks caught by Japanese research and training vessels in the western and central North Pacific. Fish. Res. 216: 177−195.

Kai, M. 2023. Spatio-temporal model for CPUE standardization: Application to shortfin mako caught by Japanese offshore and distant water shallow-set longliner in the western and central North Pacific. ISC/24/SHARKWG-1/2.

Linden, A., Māntyniemi, A., 2011. Using the negative binomial distribution to model overdispersion in ecological count data. Ecology 92, 1414–1421. https://doi.org/10.1890/10-1831.1.

Nakano, H., and Clarke, S. 2006. Filtering method for obtaining stock indices by shark species from species-combined logbook data in tuna longline fisheries. Fish. Sci. 72: 322−332.

Nakano, H., Okazaki, M. and Okamoto, H. 1997. Analysis of catch depth by species for tuna longline fishery based on catch by branch lines. *Bull. Nat. Res. Inst. Far Seas Fish*. 34:43–62.

Sippel, T., Ohshimo, S., Yokawa, K., Semba, Y., Kai, M., Carvalho, F., Kinney, M., and Suzanne, K. 2014. Spatial and temporal patterns in the size and sex of shortfin mako sharks from US and Japanese commercial fisheries: a synthesis to guide future research. ISC/14/SHARKWG-3/INFO-02.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics*, 6: 461–464.

Semba, Y., Liu, K.M., Su, S.H. 2021. Revised integrated analysis of maturity size of shortfin mako (*Isurus oxyrunchus*) in the North Pacific. ISC/17/SHARKWG-1/22

Shelton, A.O., Thorson, J.T., Ward, E.J., Feist, B.E., 2014. Spatial semiparametric models improve estimates of species abundance and distribution. Can. J. Fish. Aquat. Sci. 71, 1655–1666. https://doi.org/10.1139/cjfas-2013-0508.

Shono, H. 2005. Is model selection using Akaike's information criterion appropriate for catch per unit effort standardization in large samples? Eish. Sci. 71: 978−986.

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.

Thorson, J.T., Barnett, L.A.K., 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES J. Mar. Sci. 74, 1311–1321. https://doi.org/10.1093/icesjms/fsw193.

Thorson, J.T., Shelton, A.O., Ward, E.J., Skaug, H., 2015a. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J. Mar. Sci. 72, 1297–1310. https://doi.org/10.1093/icesjms/fsu243.

WCPFC. 2011. Analysis of North Pacific Shark Data from Japanese Commercial Longline and Research/Training Vessel Records. WCPFC-SC7-2011/EB-WP-02. Available at: https://www.wcpfc.int/system/files/EB-WP-02%20%5BAnalysis%20of%20NP%20shark%20data%20form%20Japan%5D.pdf

Zuur, A.E., Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G.M. 2009. Zero truncated and zero inflated models for count data. In Mixed effects models and extensions in ecology with R. New York: Springer Science + Business Media LLC; pp. 261–293.

Table

Table 1. Summary of model structure and outputs among different models. All models include fixed effects. "Δ" denotes a difference between the value of criteria and the minimum value for AIC and BIC.

Model	Catch rate predictors of random effect	Fixed effect	Number of parameters	Deviance	ΔAIC	ΔBIC	Maximum gradient
M-1	Null	Year	60	28739	3910	3844	< 0.002
M-2	Vessel	Year	61	27791	2886	2829	< 0.04
M-3	Station	Year	64	25701	880	847	< 0.0003
M-4	Vessel + Station	Year	65	25107	287	263	< 0.0001
M-5	Station + Year and station	Year	65	25460	641	616	< 0.0001
M-6	Vessel + Station + Year and station	Year	66	24896	79	62	< 0.0001
M-7	Vessel + Station + Year and station	Year + Quarter	68	24813	0	0	< 0.0001
M-8	Vessel + Station + Year and station	Year + Quarter + HBF	70	24815	6	22	< 0.0001

Table 2. List of all parameters and estimates of the selected model.

No	Parameter name	Symbol	Type	Estimates
1	Distance of correlation (Spatial random effect)	κ	Fixed	0.0015
2	Variation over vessel	σ_ϵ	Fixed	1.83
3	Northings anisotropy	h_1	Fixed	2.10
4	Anisotropic correlation	h_2	Fixed	0.93
5	Parameter governing pointwise variance (Spatial random effect)	η_γ	Fixed	2.56
6	Parameter governing pointwise variance (Spatio-temporal (year) random effect)	η_θ	Fixed	1.74
7	Residual variation 1 of negative binomial model	σ_1	Fixed	0.02
8	Residual variation 2 of negative binomial model	σ_2	Fixed	0.19
9	Coefficient of three month quarters for 1st predictor	β_1	Fixed	1.13
10	Coefficient of three month quarters for 2nd predictor	β_2	Fixed	0.75
11-72	Intercept for year	d_0	Fixed	Not shown
73	Vessel effect	ϵ	Random	Not shown
74	Spatial residuals	γ	Random	Not shown
75	Spatio-temporal (year) residuals	θ	Random	Not shown

Table 3. Summary of annual CPUE predicted by spatio-temporal model along with corresponding estimates of the coefficient of variation (CV), annual nominal CPUE, and number of hooks in millions. CPUEs are predicted using the best fitting model and scaled by the average CPUE.

Year	Predicted CPUE	Nominal CPUE	CV	Number of hooks (millions)
1994	1.09	0.93	0.20	4.83
1995	0.99	0.89	0.19	4.63
1996	1.03	0.79	0.21	4.52
1997	1.03	1.03	0.18	4.25
1998	1.09	1.24	0.22	2.76
1999	1.33	1.17	0.50	0.86
2000	1.37	1.20	0.27	2.73
2001	1.01	0.90	0.20	2.69
2002	1.10	1.09	0.20	2.89
2003	1.17	1.12	0.21	2.66
2004	1.10	1.10	0.20	2.89
2005	1.09	1.03	0.21	2.08
2006	1.37	1.24	0.26	2.08
2007	1.74	1.08	0.38	1.45
2008	1.07	1.08	0.24	1.30
2009	0.86	0.64	0.25	0.67
2010	0.93	0.77	0.30	0.66
2011	0.67	0.69	0.19	0.80
2012	0.71	0.97	0.17	0.76
2013	0.34	0.52	0.11	1.07
2014	0.76	0.93	0.19	1.47
2015	1.32	1.68	0.36	1.24
2016	1.09	1.58	0.23	1.19
2017	0.75	0.83	0.17	1.19
2018	0.85	1.05	0.22	1.13
2019	0.78	1.10	0.24	0.91
2020	0.67	0.70	0.23	0.52
2021	0.92	0.90	0.33	0.35
2022	0.79	0.74	0.28	0.57

Figures

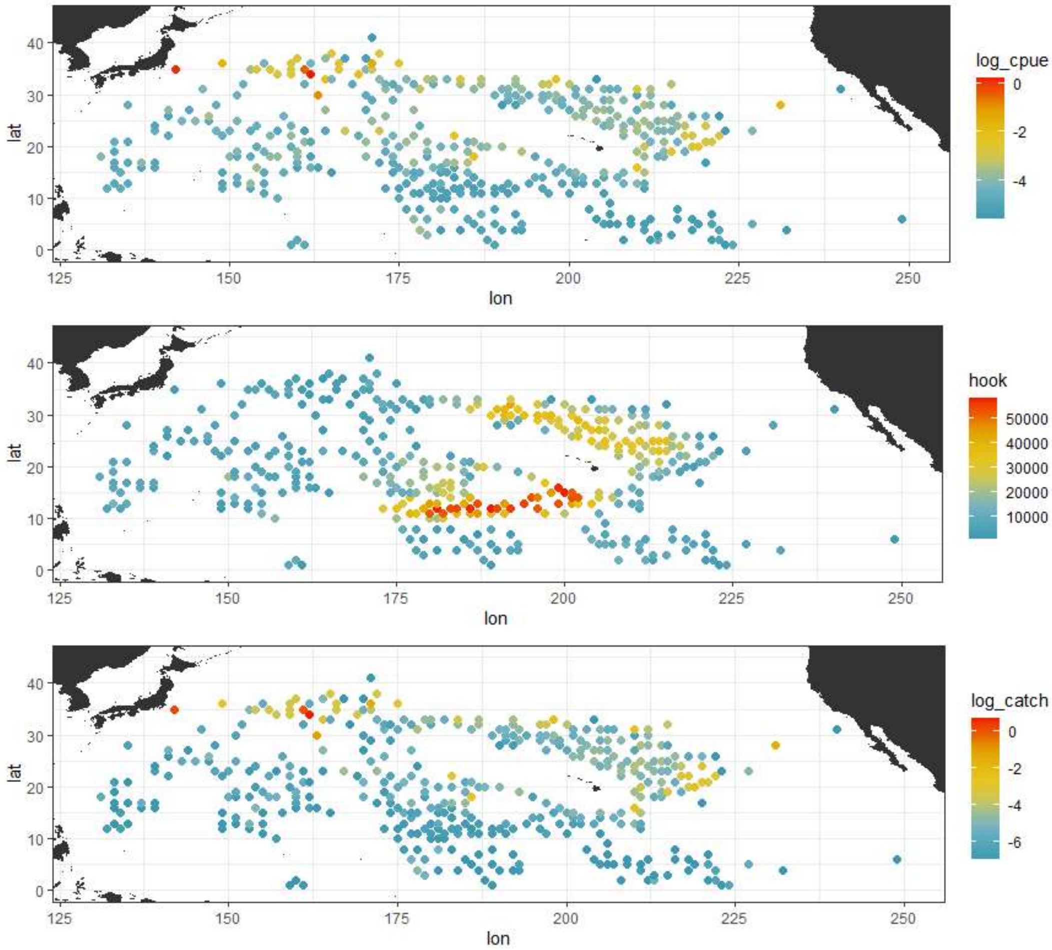
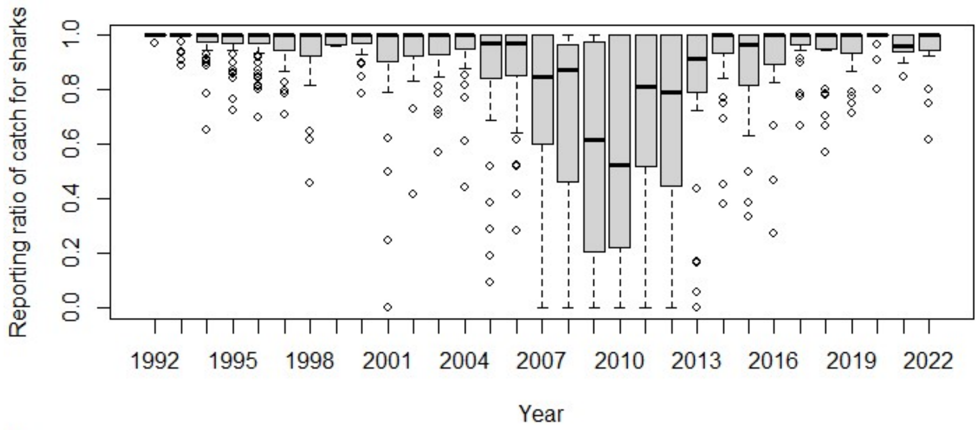


Figure 1. Spatial distributions of log-scaled nominal CPUE (upper), fishing effort (number of hooks in millions) (middle), and log-scaled catch (lower) combined from 1994 to 2022 for shortfin mako in the North Pacific. Each point denotes the location of knot.



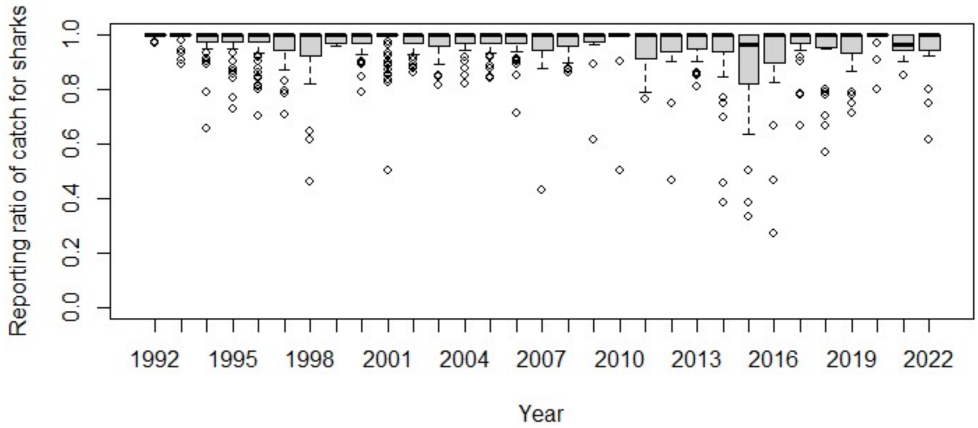


Figure 2. Boxplots of annual changes in reporting rates of catch for sharks before (upper) and after (lower) filtering.

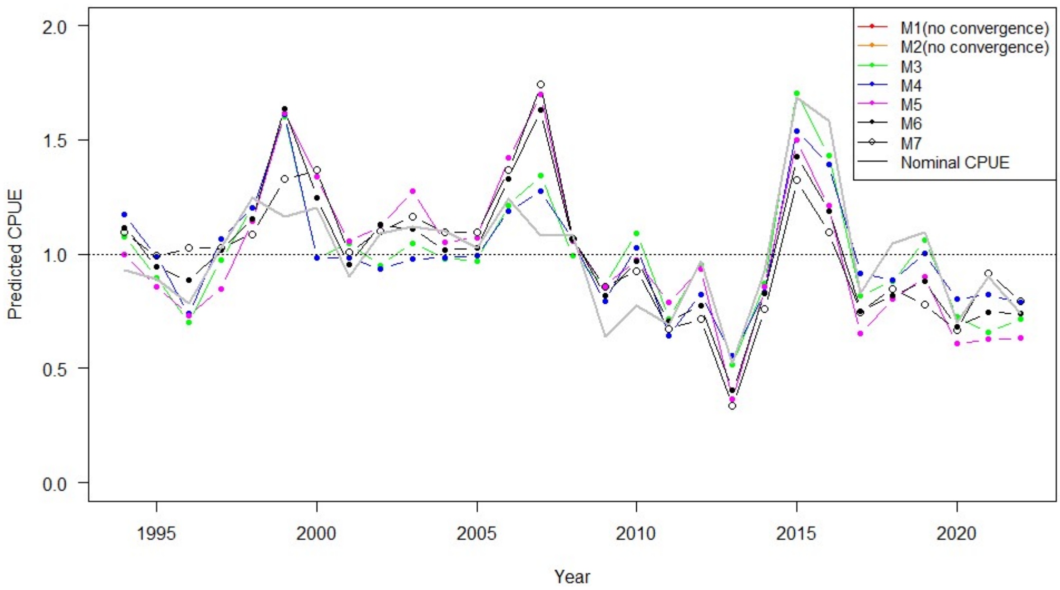


Fig. 3 Comparisons of annual predicted CPUE relative to its average among different model structures. For the details of the models, see table 1. The horizontal dotted line denotes mean of relative values (1.0). M8 was removed from plot due to large annual fluctuations.

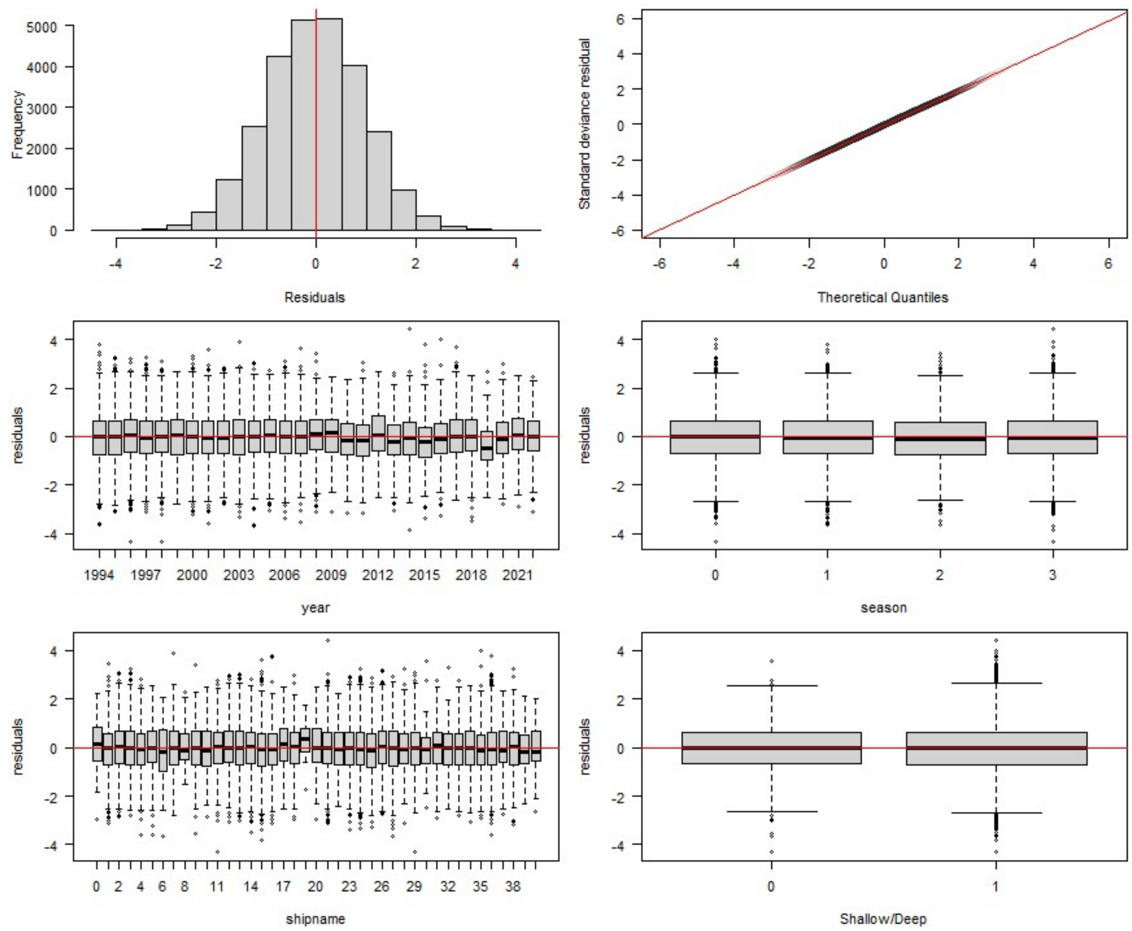


Fig. 4 Diagnostic plots of goodness-of-fit for the most parsimonious model (M7).

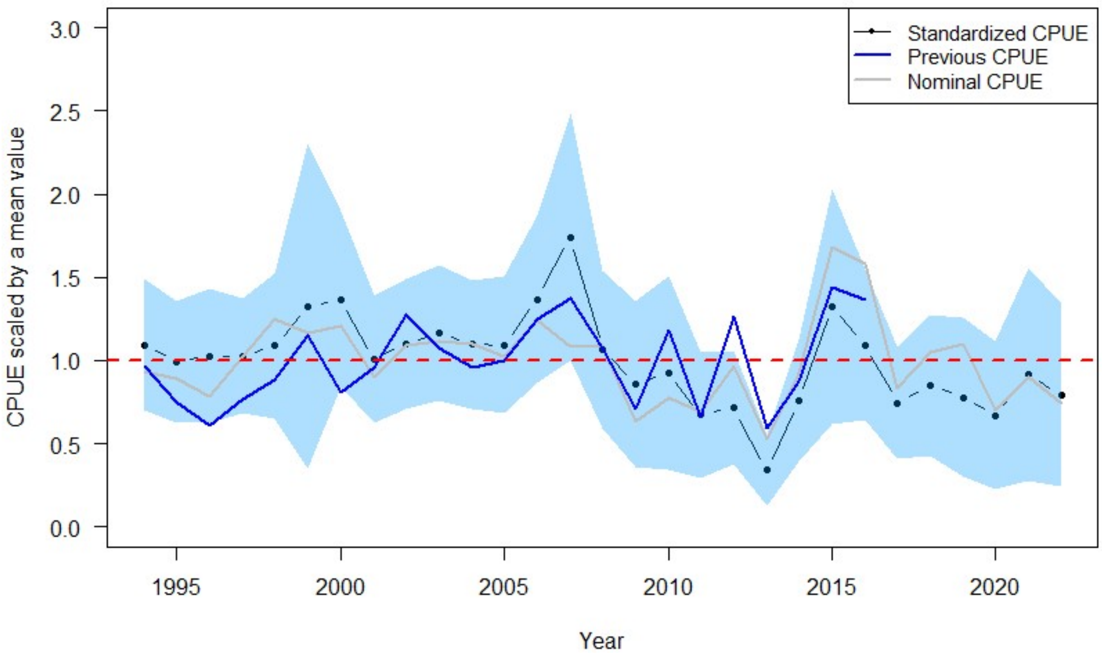


Fig. 5 Annual predicted CPUE relative to its average of the best model (M-6). Gray solid line denotes nominal CPUE relative to its average, shadow denotes 95% confidence intervals, blue solid line denotes standardized CPUE used in the previous assessment in 2018 and horizontal dotted line denotes mean of relative values (1.0).

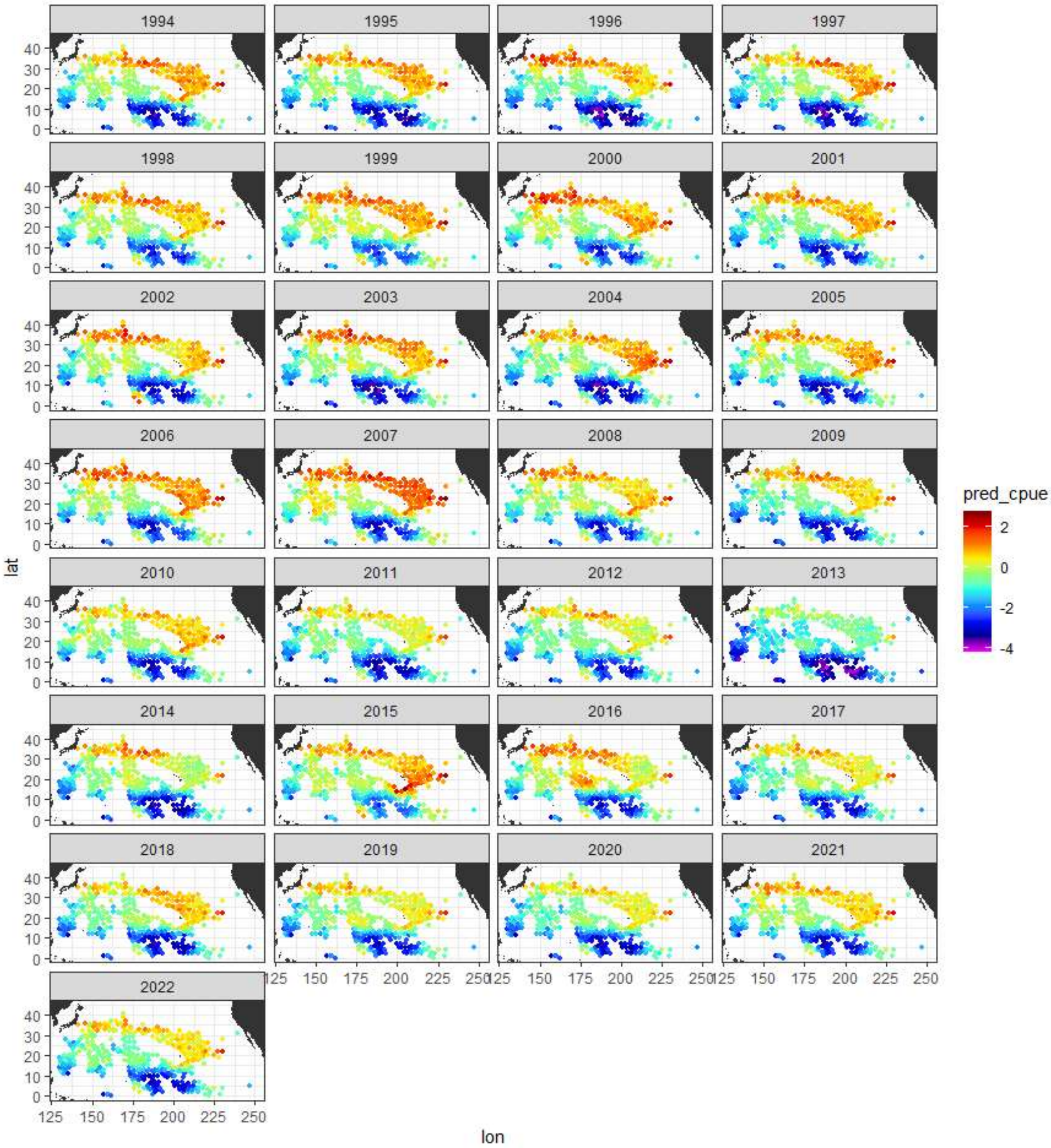


Fig. 6 Year specific spatial distribution of log-scaled predicted CPUE for shortfin mako from 1994 to 2022. Each point denotes the location of knot.

Appendix table

Table A1. Summary of model selection for binomial model with different combination of explanatory variables. Δ denotes the reduction in AIC from the best fitting model.

Model	Binomial model	Number of parameters	Deviance	ΔAIC	ΔBIC
M-1	Null	1	5744.9	914	529
M-2	Month	12	5374.6	566	267
M-3	Month +Lon5	35	5080.5	318	201
M-4	Month + Lat5	21	5099.5	309	81
M-5	Month + Lat5 + Lon5	44	4790.0	45	0
M-6	Month + Lat5 + Lon5 + HBF	54	4724.7	0	34

Appendix figures

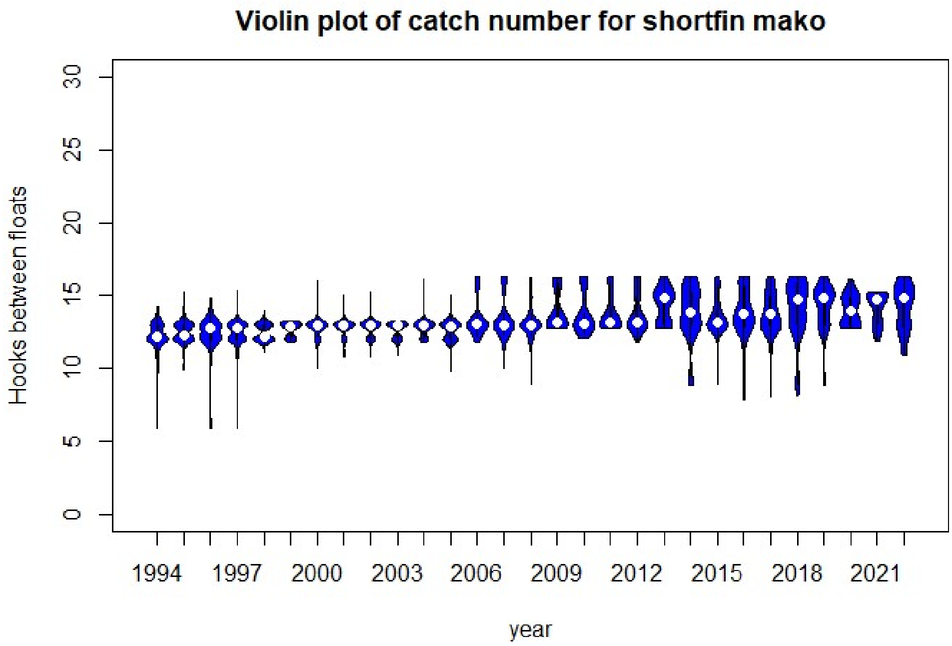


Figure A1. Annual change in number of hooks between floats and catch number of shortfin mako.

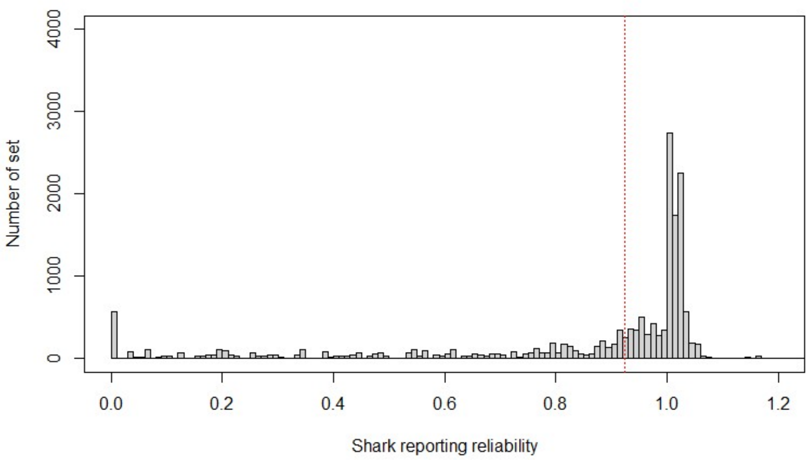


Figure A2. Number of set against shark reporting reliability for the data from 2001 to 2013.

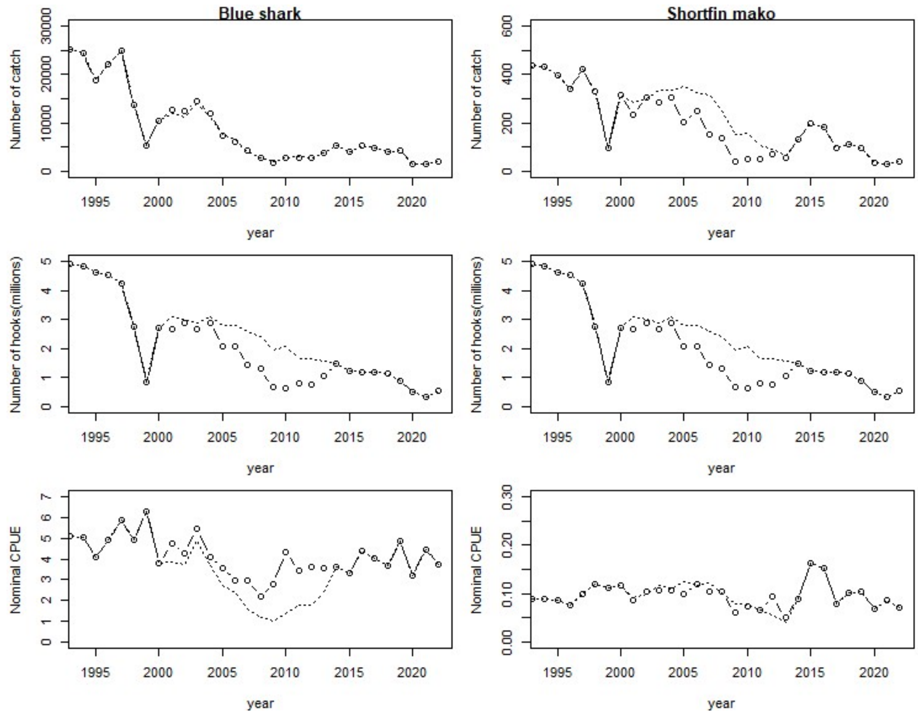


Figure A3. Yearly changes in number of catches, number of hooks (millions) and nominal CPUE (/1000hooks) for shortfin mako before (solid line with open circle) and after (broken line) filtering.