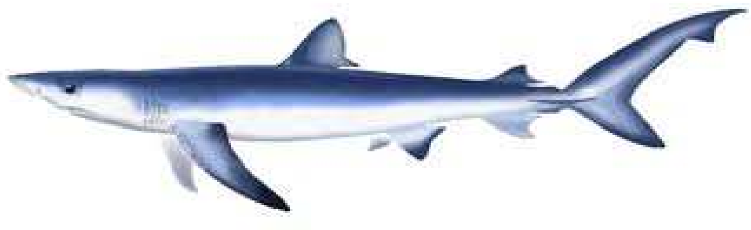
ISC/21/SHARKWG-2/03

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

Mikihiko Kai[2]

[2] Fisheries Resources Institute, Japan Fishery Research and Education Agency
5-7-1 Orido, Shimizu-ku, Shizuoka 424-8633, JAPAN
Email: kaim@affrc.go.jp



[1] Working document submitted to the ISC Shark Working Group Workshop, 9-19 November 2021, Web-meeting. **Document not to be cited without author's permission.**

Abstract

This working paper provides a standardized CPUE of blue shark caught by Japanese research and training vessels (JRTVs) longline fishery for 1994-2020 in the western and central North Pacific. A statistical filtering method was used to remove unreliable set-by-set data after 2000s collected by JRTVs. The nominal CPUE of the JRTVs was then standardized using the spatio-temporal generalized linear mixed model (GLMM) to provide the annual changes in the abundance indices in the North Pacific Ocean. The predicted abundance indices of blue shark revealed a downward trend until 2008 and an upward trend thereafter with a stable trend in recent years. The CPUE trends predicted from the fishery-independent data widely collected in the North Pacific Ocean is a very useful information about the abundance in this region.

Introduction

In the previous stock assessment in 2017, the ISC shark working group (WG) used standardized CPUE of blue shark (*Prionace glauca*) caught by Japanese offshore and distant water shallow-set longline fishery in the northwestern Pacific from 1994 to 2015 (Kai and Shiozaki, 2016) as an abundance indices of base case model (ISC, 2017). The CPUE was chosen as the best available indicators of stock abundance for late time. However, a few issues were raised with respect to the fishery-dependent data and/or an insufficient analysis of the CPUE standardization (e.g., area stratification and target effect in the generalized linear model). It is therefore essential to develop alternative abundance indices using a more sophisticated tool such as a spatial-temporal model (Thorson et al., 2015) with fishery-independent data.

The JRTVs data were collected from research vessels belonging to, or chartered to, national/prefectural fisheries research institutes, and vocational training vessels attached to fisheries high schools throughout Japan. The data are treated as one data set because both vessels are not commercial fisheries, their operation overlap, and their gear configuration is similar (i.e., deep-set fishery). Since the JRTVs data are fishery-independent data, it is expected that the data would be accurately reported and has no issue of target shifts. However, one issue of the JRTVs data was raised by Clarke et al. (2011) for the lower reporting ratio of sharks after 2000s. Kai (2019) mentioned that the main reason for the under-reporting between 2001 and 2013 is possibly due to a reduced recording of discarded sharks resulting from a revision of the input for in 2000 (an increase in input items). In 2013, the Japan Fishery Agency instructed to record accurately the number of all sharks caught, including the discard, and then the reporting rate of sharks increased after 2013.

The objective of this working paper is to provide standardized CPUE of blue shark caught by JRTVs longline fishery for 1994-2020 in the western and central North Pacific Ocean. First, temporal changes in the reporting rate are analyzed and unreliable set-by-set data are removed using a statistical filtering method. Then, the nominal CPUE is standardized using a spatio-temporal GLMM for the filtered data.

Materials and Methods

The author used similar statistical filtering method and spatio-temporal model as those used in the paper (Kai, 2019).

Data sources

Set-by-set longline logbook data collected from JRTVs in the western and central North Pacific Ocean from 1992 to 2020 were used and the data includes information on species of sharks, operation time, catch numbers, number of hooks, number of branch line between floats (HBF), location of sets by latitude-longitude resolution of 1° ×1°, and trip identity. As the JRTVs mostly use deep sets (i.e., 6-16 HBF), two types of deep sets (shallower and deeper deep sets; HBF < 11 and 10 < HBF) were used in this analysis. The four seasons (quarters (Q) 1 to 4) of the year were defined: Q1: JAN-MAR; Q2: APR-JUN; Q3:JUL-SEP; Q4:OCT-DEC in the analysis.

Data filtering

Preliminary filtering was conducted to remove incomplete and insufficient data that have little or no information about HBF and locations (latitude and longitude), number of hooks that were less than 800, HBF that were less than 6 (i.e., shallow sets), and operations that were operated in waters other than the North Pacific Ocean.

In addition, follow-up filtering was also conducted to remove unreliable set-by-set data caused by under-reporting of actual shark catches. The author used a statistical filtering method based on the information on shark presence in the catch (Hoyle et al., 2017; Kai, 2019) and applied to JRTV data from 2001 to 2013 to accommodate a clear decline in annual reporting rates during this period (**Fig. 1a**).

CPUE standardization with Spatio-temporal model

The spatio temporal model is consisted of two components of encounter probability and positive catch in a delta model. The first predictor was fixed at a constant value because of high positive catches (> 94%). Second predictor was modeled using a negative binomial (NB) model to account for the datasets with over-dispersion (variance/mean =14.56):

$$c \sim NegBin\ (c^*, c^*(1+\sigma_1) + {c^*}^2\sigma_2),$$
$$\log\ (d) = d_0(t) + \gamma(s) + \theta(s,t) + \epsilon(v) + \textstyle\sum_{j=1}^{n_j} \beta_j x_j, \tag{1}$$

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), $\epsilon(v)$ represents random variation in catchability for the vth vessel, and β_j represents the impact of covariate j with value x_j on catchability. The shallow and deep sets and three-month quarters (i.e. $n_j = 2$, $x_j = hbf\ and\ q$) are used as covariates (changing the catchability) corresponding to Eq. (1).

The VAST (Vector Autoregressive Spatio-Temporal; version VAST_v13_0_0) software package for R (Thorson, 2019) was used to standardize the nominal CPUE. To be consistent with the period of late time series for Japanese longline fishery, the author used the JRTVs data after 1993 in the CPUE standardization. Annual abundance index I was estimated as:

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

where n_s is total number of knots (i.e., sampling location in this study) and f is fishing effort (number of hooks) at location s.

Model selection and diagnostics

To select the best model, the explanatory variable was sequentially removed from the full model in Eq (1). The best model was selected using the AIC (Akaike 1973). 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

Summary of data filtering

The preliminary filtering reduced the number records for this analysis from 38,720 sets to 34,810 sets. The follow-up filtering reduced the number of records for this analysis from 34,810 sets representing 1,425 trips to 31,045 sets representing 1,261 trips. The follow-up filtering appeared to be reasonable because the reduction of catch rates between 2001 and 2013 disappeared (**Fig. 1**). The difference of annual changes in number of catches, number of hooks, and nominal CPUE between the data with and without follow-up filtering are shown in **Fig. 2**.

Selection of the best model and annual trends in CPUE

All models were reasonably converged with the positive definite of hessian matrix and a small value of maximum gradient (**Table 1**). The saturated model (M_6) including spatial, spatio-temporal variances, and variation over vessel as random effects were identified by AIC as the most parsimonious model (**Table 1**). The predicted CPUE changed substantially if random effect components were sequentially added to the null model which had no random effects (M_1) (**Fig. 3**). The fixed effect components (HBF and quarter) had a small effect on the annual trends in the CPUE but those decreased the values of AIC (decrease of AIC for HBF and quarter was 1099 and 464, respectively). Lists of all parameters and estimates of the best models are shown in **Table 2**. The predicted annual CPUE of the best model revealed a declining trend for 1994 to 2008, and then it increased gradually until 2016 with a stable trend in recent years (**Fig. 4**). Uncertainty (CV) in the CPUE estimates was substantially larger in 1999 and 2020 due to the fact that the fishing efforts (number of hooks) were smaller than those of any other years (**Table 3**).

Model diagnostics

Diagnostic plots of goodness-of-fit for the best model didn't show a serious deviation from normality and model misspecification (**Fig. 5**). These results suggested that the fitting of the best model to the data was good.

Discussions

This document paper predicted annual abundance indices of blue shark in the western and central North Pacific Ocean from 1994 to 2020 (**Fig. 4**). The author applied a spatio-temporal GLMM to the fishery independent JRTVs data after the author removed unreliable set-by-set data with low reporting rates of sharks in 2000s using a statistical filtering method.

The hotspots of BSH appeared mainly in the temperate waters as shown in past studies (Hiraoka et al., 2016; Kai et al., 2017) and in the subtropical areas off the southwest coast of Hawaii (**Fig. 6**) where adult blue shark occur widely in low-latitude waters (Nakano and Stevens, 2008) and Hawaii-based pelagic longline vessels frequently operate in those waters (ISC, 2017).

The predicted annual CPUE could be a candidate of abundance indices for the base-case model in the upcoming stock assessment of North Pacific blue shark. There are three advantages: 1) the wide area coverage in the North Pacific Ocean compared to the other CPUEs (e.g. Hawaiian longline fleet), 2) reliability of the reporting ratio compared to commercial fisheries, and 3) the statistical soundness of the spatio-temporal model compared to the conventional GLM approach (Shelton et al., 2014). It is essential, however, to carefully choose the abundance indices for the base-case model, as the time series is of deep-set data, which typically catch fewer sharks than shallow-set data, has poor coverage in areas with high catch rates in the temperate waters.

Although the author didn’t consider the seasonal changes in the catch rates as random effect due to the limitation of time, it might be useful to include such effect in the spatio-temporal model in the future work because adult and subadult blue sharks are well known to have seasonal migration (Fujinami et al., 2021; Kai et al. 2017).

References

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

Clarke, S.C., Yokawa, K., Matsunaga, H., Nakano, H. 2011. Analysis of North Pacific shark data from Japanese commercial longline and research/training vessel records. WCPFC-SC7-2011/EB-WP-02.

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

Fujinami, Y., Shiozaki, K., Hiraoka, Y., Semba, Y., Ohshimo, S., Kai, M. 2021. Seasonal migrations of pregnant blue sharks Prinace glauca in the northwestern Pacific. Mar. Ecol. Prog. Ser. 658, 163−179.

Hiraoka, Y., Kanaiwa, M., Ohshimo, S., Takahashi, N., Kai, M., Yokawa, K. 2016. Relative abundance trend of the blue shark *Prionace glauca* based on Japanese distant-water and offshore longliner activity in the North Pacific. Fish. Sci. 82, 687−699.

Hoyle, S.D., Semba, Y., Kai, M., Okamoto, H. 2017. Development of Southern Hemisphere porbeagle shark stock abundance indicators using Japanese commercial and survey data. New Zealand Fisheries Assessment Report 2017/07.

ISC 2017. Stock assessment and future projections of blue shark in the North Pacific Ocean through 2015. ISC 17 Plenary report and document.

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., Shiozaki, K. 2016. Update of Japanese abundance indices for blue shark caught by Japanese offshore and distant water shallow-set longliner in the North Pacific. ISC/16/SHARKWG-1/10.

Kai, M., Thorson, J.T., Piner, K.R., Maunder, M.N. 2017. Predicting the spatio-temporal distributions of pelagic sharks in the western and central North Pacific. Fish. Oceanogr. 2, 569−582

Lindén, A., Māntyniemi, A. 2011. Using the negative binomial distribution to model overdispersion in ecological count data. Ecology 92, 1414−1421.

Nakano, H., Stevens, J.D. 2008. The biology and ecology of the blue shark, *Prionace glauca*, in: Camhi, M.D., Pikitch, E.K., Babcock, E.A. (Eds.), Sharks of the open ocean: biology, fisheries and conservation. Blackwell Publishing Ltd., Oxford, pp. 140–151.

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.doi. 10.1139/cjfas-2013-0508.

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

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.

Tables

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.

Model	Catch rate predictors of random effect	Number of parameters	Deviance	ΔAIC	Maximum gradient
M-1	Null	32	172095	19371	< 0.0001
M-2	Vessel	33	168620	15898	< 0.0001
M-3	Station	36	161991	9276	< 0.0001
M-4	Vessel + Station	37	159614	6827	< 0.0001
M-5	Station + Year and station	38	157882	2081	< 0.0001
M-6	Vessel + Station + Year and station	39	152710	0	< 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	0.0023
2	Variation over vessel	σ_{ϵ}	Fixed	1.37
3	Northings anisotropy	h_1	Fixed	1.44
4	Anisotropic correlation	h_2	Fixed	0.91
5	Parameter governing pointwise variance (Spatial random effect)	η_{γ}	Fixed	1.56
6	Parameter governing pointwise variance (Spatio-temporal (year) random effect)	η_{θ}	Fixed	0.51
7	Parameter governing autocorrelation (Spatio-temporal: year random effect)	ρ_{θ}	Fixed	1.38
8	Residual variation 1 of negative binomial model	σ_1	Fixed	0.14
9	Residual variation 2 of negative binomial model	σ_2	Fixed	0.20
10	Coefficient of hooks between floats	β_1	Fixed	-0.551
11	Coefficient of three month quarters	β_2	Fixed	0.117
12-37	Intercept for year	d_0	Fixed	Not shown
38	Vessel effect	ϵ	Random	Not shown
39	Spatial residuals	γ	Random	Not shown
40	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. Values are predicted using the best fitting model and scaled by average CPUE.

Year	Predicted CPUE	Nominal CPUE	CV	Number of hooks (millions)
1994	1.48	1.27	0.10	4.83
1995	1.44	1.03	0.12	4.63
1996	1.39	1.23	0.10	4.52
1997	1.44	1.47	0.10	4.25
1998	1.39	1.24	0.12	2.76
1999	1.44	1.58	0.19	0.86
2000	1.24	0.95	0.12	2.73
2001	1.17	1.10	0.10	2.91
2002	1.09	1.05	0.10	3.03
2003	1.05	1.33	0.11	2.75
2004	0.96	0.99	0.10	3.09
2005	0.78	0.83	0.12	2.34
2006	0.72	0.71	0.12	2.31
2007	0.64	0.72	0.14	1.51
2008	0.41	0.51	0.13	1.44
2009	0.58	0.69	0.13	0.69
2010	0.79	0.98	0.15	0.75
2011	0.66	0.84	0.15	0.83
2012	0.59	0.84	0.15	0.85
2013	0.79	0.84	0.15	1.17
2014	1.04	0.90	0.16	1.47
2015	0.83	0.83	0.15	1.24
2016	1.09	1.10	0.13	1.19
2017	1.06	1.01	0.12	1.19
2018	0.98	0.92	0.13	1.13
2019	0.98	1.22	0.15	0.91
2020	0.97	0.81	0.17	0.52

Figures

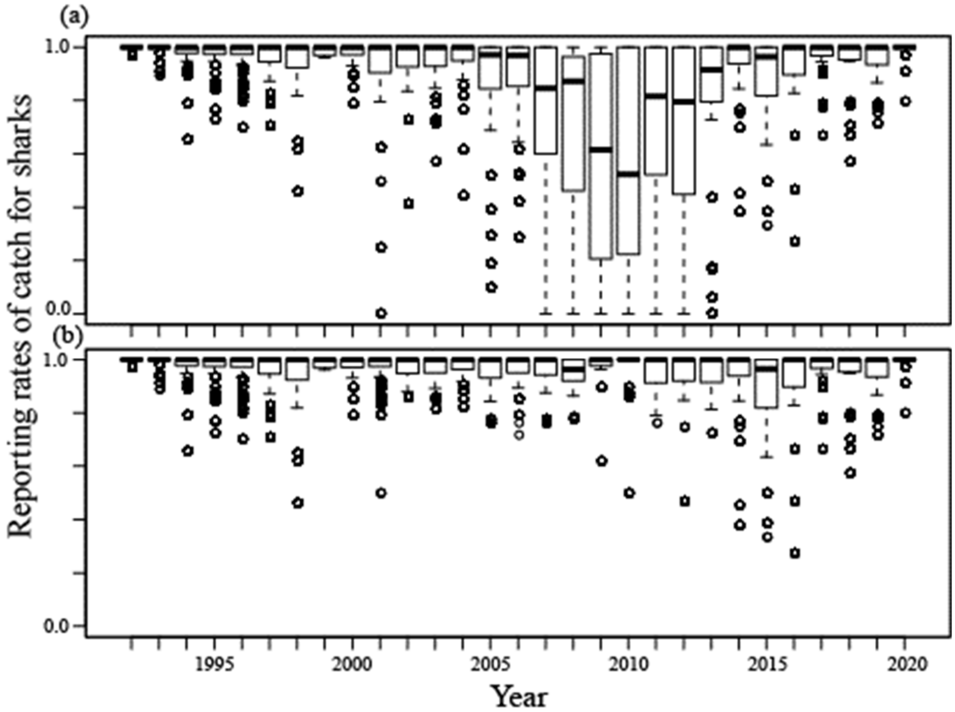


Fig. 1 Annual reporting rates of catch for sharks (a) before filtering and (b) after filtering.

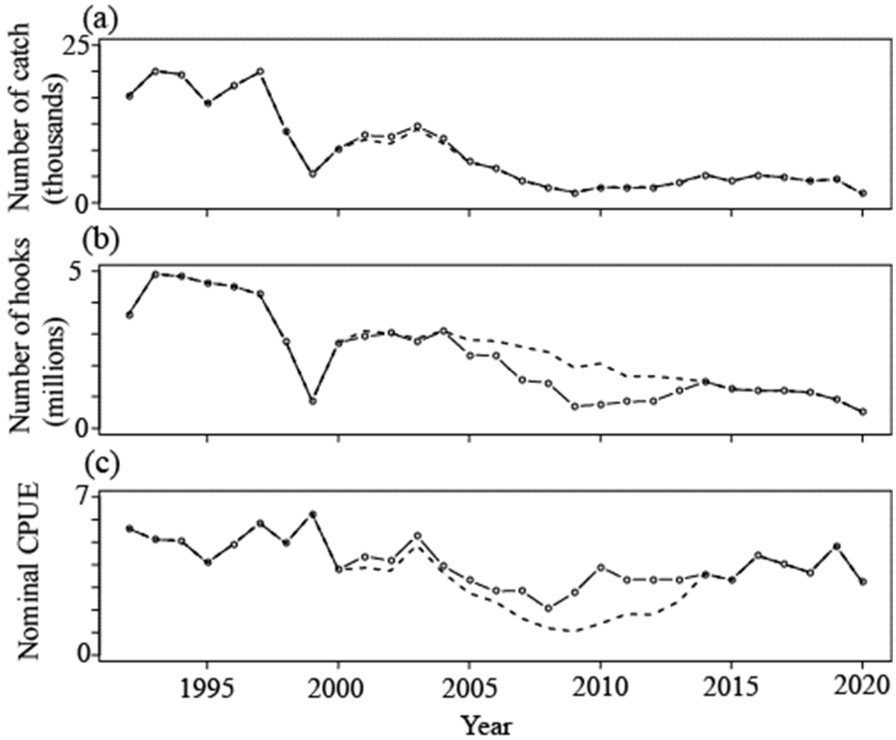


Fig. 2 Annual catch in numbers (thousands) (a), number of hooks (millions) (b), and nominal CPUE (per 1000 hooks) (c) for blue shark before filtering (broken line) and after filtering (solid line with open circle).

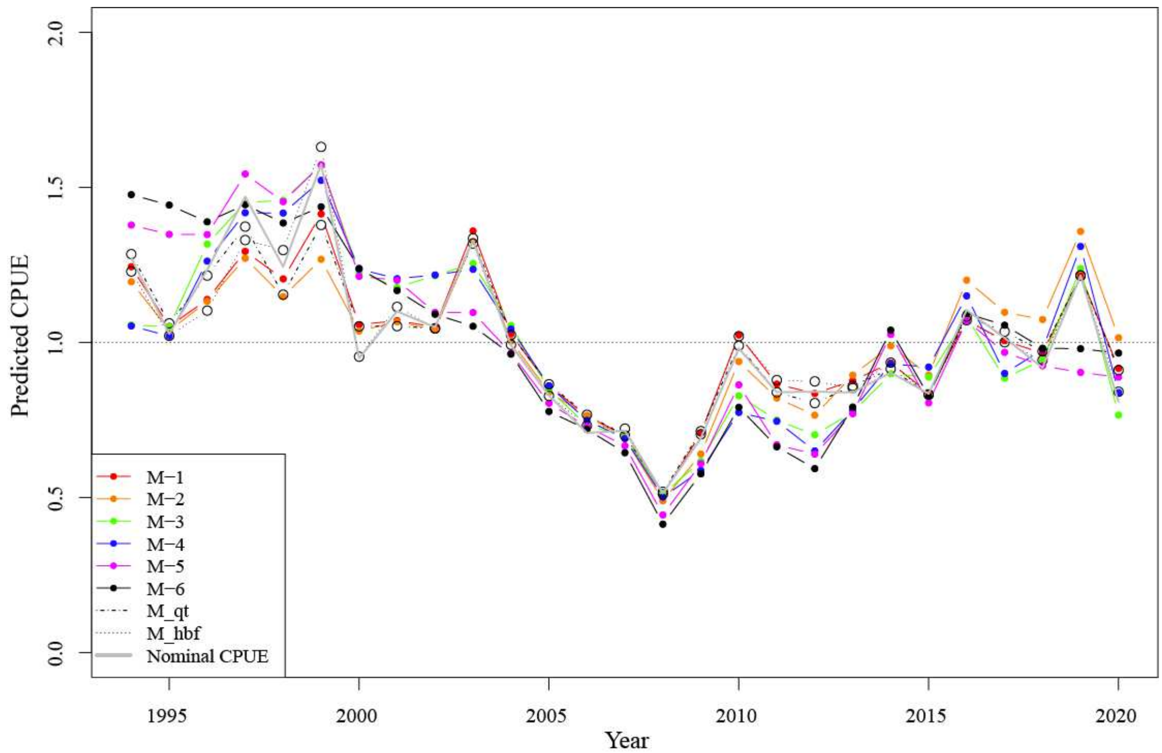


Fig. 3 Comparisons of annual predicted CPUE relative to its average among different structures of the spatio-temporal model. M_qf and M_hbf denotes the null model with effect of season and hbf, respectively. For the details of other models, see table 1.

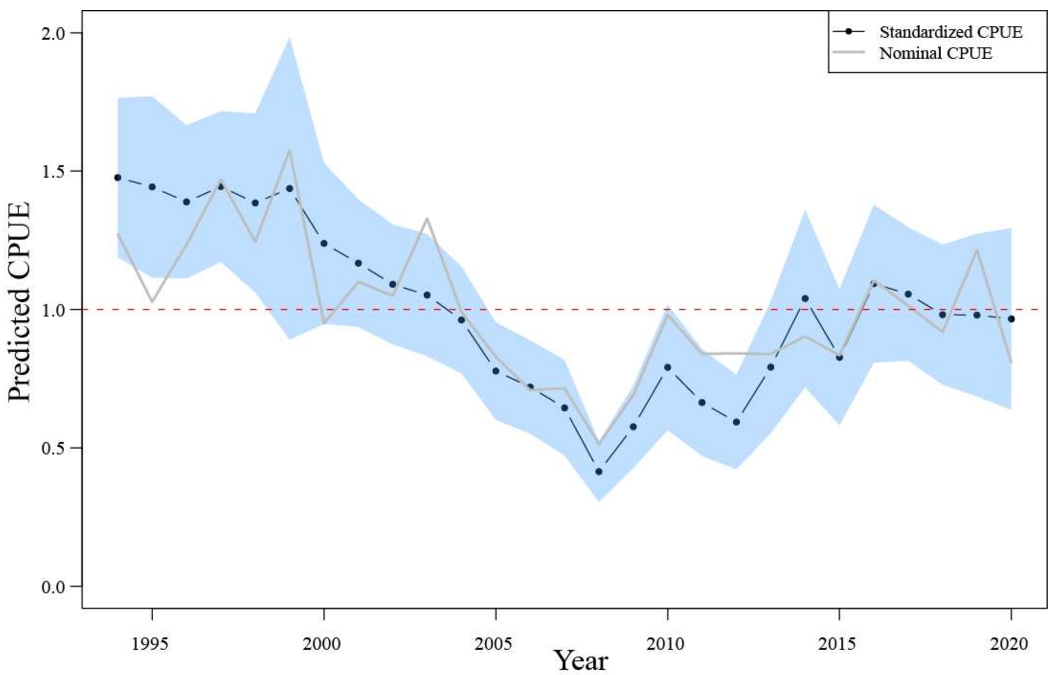


Fig. 4 Annual predicted CPUE relative to its average. Gray solid line denotes nominal CPUE relative to its average, shadow denotes 95% confidence intervals, and horizontal dotted line denotes mean of relative values (1.0).

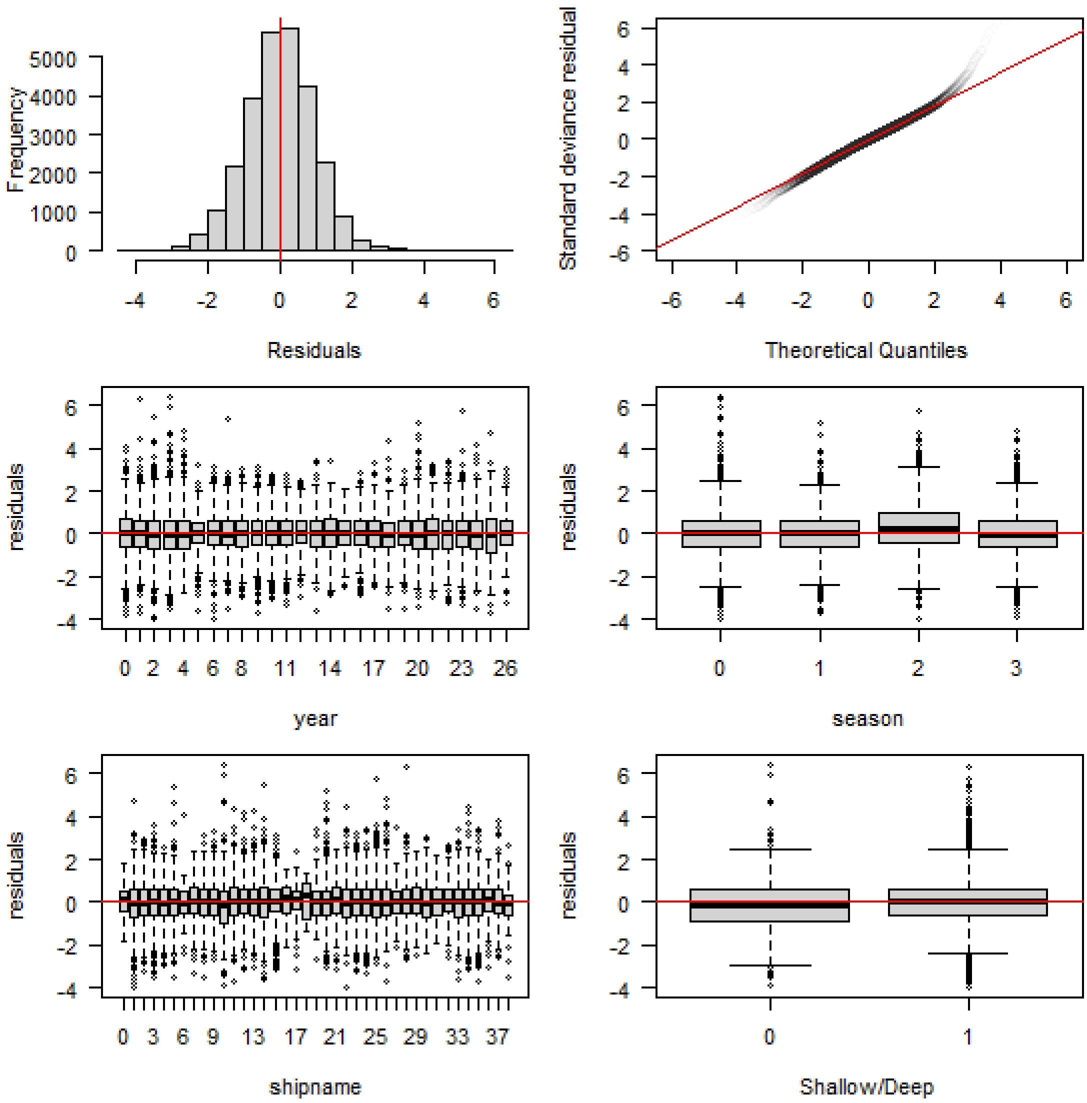


Fig. 5 Diagnostic plots of goodness-of-fit for the most parsimonious model.

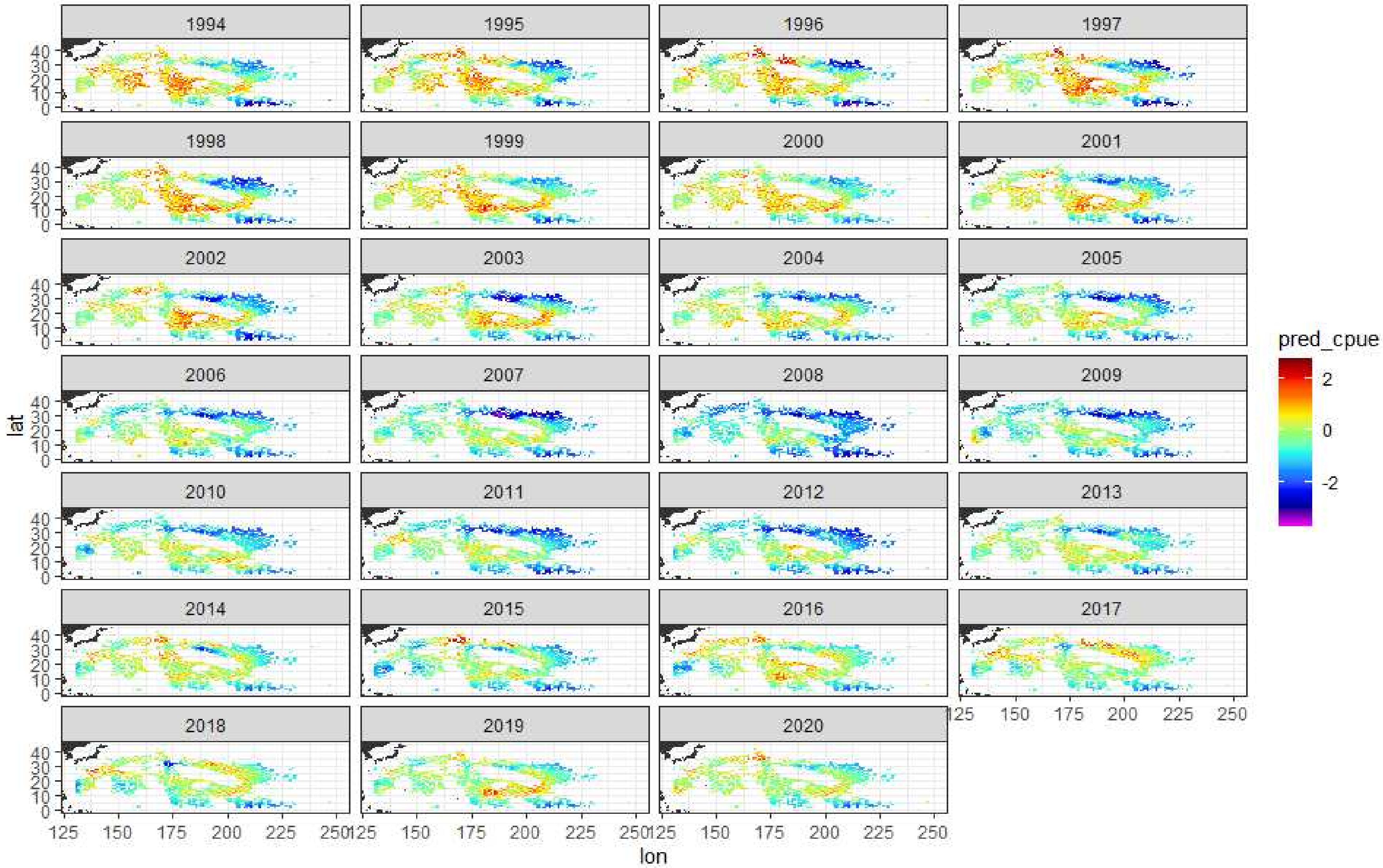


Fig. 6 Annual spatial distribution of log-scaled predicted CPUE for blue shark. Number of knot (1480) is the same as that of sampling location.