Japanese Longline CPUE Standardization (1979-2018) for Swordfish (*Xiphias gladius*) in the Indian Ocean using zero-inflated Bayesian hierarchical spatial model

Kenji Taki*, Hirotaka Ijima, and Yasuko Semba *takisan@affrc.go.jp

National Research Institute of Far Seas Fisheries, Fishery Research Agency, Shimizu, Shizuoka, Japan

Summary

To grasp the historical trajectory of Swordfish stock abundance, we addressed standardizing the CPUE of Swordfish in the Indian Ocean by Japanese longliners using their logbook data for the period 1979-2018. We divided the time-period into two periods, 1979-1993 and 1994-2018 for the analysis for four areas (NW, NE, SW, SE) of Indian Ocean because of apparent change of data-format of logbook around in 1994 and the change of fishing methods (e.g. materials of stem and branch lines and gear configuration such as number of hooks between floats) related to catchability: q not detailed in the logbook during the mid-1990s. In this analysis, we applied Bayesian hierarchical spatial models. Since the catch data are counts characterize by many zeros, we evaluated zero-inflated Poisson GLMM (ZIP-GLMM). 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 half Cauchy distribution as a prior for the random effect. Best candidate models were selected based on Widely Applicable Bayesian Information Criterion (WAIC). From the lowest value of WAIC, spatial Poisson GLMM with autoregressive (AR1) modelled for the year trend (i.e. m_zip_spde2 model) was selected as the best candidate for each area except for SE area. No apparent trend in interannual variation of standardized CPUE was generally observed for each area. The uncertainties are much larger for the current spatial models due to consideration of spatial effect as compared to the past non-spatial models (Ijima 2017) although the trend of point estimates is similar. We will improve the models dealing with the appropriate catchability (q), applying the state space model and/or latent variable model in the future.

1. Introduction

Hierarchical Bayesian models have traditionally relied on Markov chain Monte Carlo (MCMC) simulation techniques, which are computationally expensive and technically

challenging, consequently limiting their use. However, a new statistical approach is currently readily available, namely integrated nested Laplace approximations (INLA) via the R-INLA package (http://www.r-inla.org). INLA methodology and its powerful application to modelling complex datasets has recently been introduced to wider nontechnical audience (Illian et al. 2013). As opposed to MCMC simulations, INLA uses an approximation for inference and hence avoids the intense computational demands, convergence, and mixing problems sometimes encountered by MCMC algorithms (Rue and Martino 2007). Moreover, included in R-INLA, the stochastic partial differential equations (SPDE) approach (Lindgren et al. 2011) is another statistical development that models spatial random effect (Gaussian random field, GRFs) much faster as well as constructs flexible fields that are better adept to handle datasets with complex partial structure (Lindgren 2013). This is often the case with fisheries data, since fishermen tend to target fishing grounds, resulting in clustered spatial patterns and large regions without any values. Together, these new statistical methods and their implementation in R allows scientists to fit considerably faster and more reliably complex spatiotemporal model (Rue et al. 2009, Cosandey-Godin et al. 2015).

The aim of this paper is to grasp the historical trajectory of Swordfish (*Xiphias gladius*) stock abundance in the Indian Ocean by Japanese longliners during 1979-2018 for the four areas (Northwest, Northeast, Southwest, and Southeast), applying zero-inflated Bayesian hierarchical spatial models fitted using these two novel techniques.

2. Materials and methods

Data sets

Japanese longline logbook data was used for the CPUE standardization of the Swordfish in the Indian Ocean. We used the data from 1979 onwards because the number of hooks between floats and the vessel name, which affect largely the CPUE standardization, are completely available since then. The format of the logbook was changed around 1994 and the fishing methods (e.g. materials of stem and branch lines and gear configuration such as number of hooks between floats) related to catchability: q, which is not detailed in the logbook, was changed during the mid-1990s. Therefore, we divided the time-period into two periods, 1979-1993 and 1994-2018 for the analysis. The resolution of the logbook is 1x1 grid scale. We used the four analysis areas (NW, NE, SW, SE) of Indian Ocean set in the 9th session of the IOTC working party on billfish for the standardization analysis of Swordfish (IOTC 2014; Figure 1).

Statistical models

In this analysis, we applied Bayesian hierarchical spatial models. We did not apply the spatiotemporal models because this method is computationally expensive and the Widely

Applicable Bayesian Information Criterion (WAIC; Watanabe, 2012) did not differed so much between spatial and spatiotemporal models. Since the catch data are counts characterize by many zeros (Figure 2), we evaluated zero-inflated Poisson GLMM (ZIP-GLMM). The zeroinflated model is useful because this model can estimate "true" zero catch. To apply zeroinflated negative binomial GLMM (ZINB-GLMM) is another way to consider the many zero issue, but the ZINB tends to cause underdispersion (e.g. Ijima 2017), thus we think zeroinflated Poisson GLMM (ZIP-GLMM) is more appropriate to use for the CPUE standardization.

The explanatory variables of fixed effect part are majorly the year (yr) and quarter (Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec; qtr), and random effect part are area (latlon), gear configuration; number of hooks between floats (hpb), month (month), fleet (jp_name). The number of hooks between floats generally increased to the mid-1990s (Figure 3). Most variables were treated as the categorical variables but the autoregressive model (AR1) was applied to years for some spatial models to avoid the large uncertainties. Considering the random effect is appropriate because there are a lot of variables for the vessel name (jp_name), gear configuration (hpb), 5x5 area (latlon) effect. The random effect model can also remove the pseudo-replication by vessel, gear configuration and operating 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 half Cauchy distribution as a prior for the random effect. Best candidate models were selected based on Widely Applicable Bayesian Information Criterion (WAIC).

3. Result and discussion

We examined the total of eight models. From the lowest value of WAIC, spatial Poisson GLMM with autoregressive (AR1) modelled for the year trend (i.e. m_zip_spde2 model) was selected as the best candidate for each area except for SE area (Table 1).

Northwest

The posterior probability distribution is shown in Figure 4. The southeastern part has been negatively correlated with the Swordfish CPUE during 1994-2018, while the western coastal part has been positively correlated with the CPUE throughout the period (Figure 5). No apparent trend in interannual variation of standardized CPUE was observed (Figure 6, Table 2).

Northeast

The posterior probability distribution is shown in Figure 7. The western offshore part (south of India) has been positively correlated with the Swordfish CPUE throughout the period (Figure 8). No apparent trend in interannual variation of standardized CPUE was observed (Figure 9, Table 3).

Southwest

The posterior probability distribution is shown in Figure 10. The western coastal part has been positively correlated with the Swordfish CPUE throughout the period (Figure 11). No apparent trend in interannual variation of standardized CPUE was observed Figure 12, Table 4).

Southeast

From the WAIC, spatial and non-zero-inflated model (m_spde2) was selected during 1979-1993, while non- spatial and zero-inflated model (m_zip_glmm) was selected during 1994-2018. The posterior probability distribution is shown in Figure 13. Southern part has been negatively correlated with the Swordfish CPUE during 1979-1993 (Figure 14). No apparent trend in interannual variation of standardized CPUE was generally observed for each area (Figure 15, Table 5).

Figure 16 shows the comparison of interannual variations of relative standardized CPUE between this study and the past study by Ijima (2017) who did not use the spatial model. The trend of point estimates is similar for each area between the two studies. However, the uncertainties are much larger for the spatial models due to consideration of spatial effect.

We will improve the models dealing with the appropriate catchability (q), applying the state space model (e.g. Yin et al. 2019) and/or latent variable model (e.g. Warton et al. 2015) in the future.

4. References

- 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. 2017. CPUE standardization of the Indian Ocean swordfish (*Xiphias gladius*) by Japanese longline fisheries: Using negative binomial GLMM and zero inflated negative binomial GLMM to consider vessel effect. IOTC-2017-WPB15-19.
- 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.

- IOTC. 2014. Report of the 12th session of the IOTC working party on billfish. IOTC-2014-WPB-R [E].
- 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.
- 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.
- Warton, D, I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., and Hui, F. K.C. 2015. So many variables: joint modeling in community ecology. Trends in Ecology and Evolution. 30 (12): 766-779.
- Watanabe, S. 2013. A widely applicable Bayesian information criterion. Journal of Machine Learning Research 14: 867-897.
- Yin, Y., Aeberhard, W. H., Smith, S. J., and Flemming, J. M. 2019. Identifiable state-space models: A case study of the Bay of Fundy sea scallop fishery. Canadian Journal of Statistics 47 (1): 27-45.



Figure 1. Four analysis areas for the Swordfish CPUE standardization in the Indian Ocean set in the 9th session of the IOTC working party on billfish (IOTC 2014).



Figure 2. Zero catch rate of Swordfish caught by Japanese long line fishery.



Figure 3. Historical change of the gear setting (hooks between floats) in Indian Ocean. Gear configuration is different between North and South Indian Ocean because Japanese longliners are targeting Southern Bluefin tuna in the South Indian Ocean.

Table 1. Eight models and their WAIC values for two time periods of four areas. Selected r	models cor	respondec	l to those	with the s	smallest va	alues yello	w-highlig	nted.
	NW(79-93)	NW(94-18)	NE(79-93)	NE(94-18)	SW(79-93)	SW(94-18)	SE(79-93)	SE(94-18)
m_null = inla (swo~1,data=d,offset=log(d\$hooks/1000),family="poisson")	165589	280178	68131	163162	320465	349816	134496	327635
m_glm = inla (swo~yr + latlon,data=d,offset=log(d\$hooks/1000),family="poisson")	154089	254423	>10 ¹⁸	>10 ¹⁸	>10 ¹⁸	286990	>10 ¹⁶	>10 ¹⁸
m_glmm = inla (swo~yr + qtr + f(latlon,model="iid",hyper=hcprior) + f(jp_name,model="iid")+f(hpb,model="iid"),data=d,offset=log(d\$hooks/1000),family="poisson")	140089	233079	59084	135636	191961	256884	46025	182335
m_zip_g mm = inla (swo~yr + qtr + f(latlon,model="iid") + f(jp_name,model="iid"), data=d,offset=log(d\$hooks/1000),family="zeroinflatedpoisson1")	137367	223494	57837	131982	183257	242387	43373	154266
m_spde = inla (swo~0 + intercept + yr + qtr + f(hpb,model="iid") + f(jp_name,model="iid") + f(w,model=spde), data=inla.stack.data(StackFit),offset=log(d\$hooks/1000),family="poisson") ↔	138833	230180	58380	134313	178481	243911	44402	170572
m_spde2 = inla (swo-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(d\$hooks/1000),family="poisson")	138155	229650	58302	133796	176021	243289	43013	163426
m_zip_spde = inla (swo~0 + intercept + yr + qtr + f(hpb,model="iid") +f(jp_name,model="iid") + f(w,model=spde), data=inla.stack.data(StackFit),offset=log(d\$hooks/1000),family="zeroinflatedpoisson1")	135626	220290	57164	130300	168193	230009	52478	163012
m_zip_spde2 = inla (swo~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(d\$hooks/1000),family="zeroinflatedpoisson1")	135008	219831	<mark>57109</mark>	<mark>129899</mark>	166239	229598	I	I

IOTC-2020-WPEB18-14



Figure 4. Northwest. Posterior marginal distribution of the fixed (intercept), precision for random effects, temporal correlation term (Rho), and spatial field parameters (Thetas).



Figure 5. Northwest. Mean of latent spatial field.



Figure 6. Northwest. Historical changes of CPUEs. Line is standardized CPUE and filled area is 95% credible interval. Points denote nominal CPUE. Note the different scale of y axis for CPUE between the periods.

year	nominal	Standardized	2.50%	97.50%	year	nominal	Standardized	2.50%	97.50%
1979	0.71	1.00	0.62	1.59	1994	0.93	1.68	0.76	3.65
1980	0.50	0.75	0.47	1.19	1995	0.77	1.27	0.58	2.77
1981	0.49	0.77	0.48	1.23	1996	0.69	1.08	0.49	2.35
1982	0.64	0.85	0.53	1.35	1997	0.73	1.06	0.48	2.30
1983	0.51	0.82	0.52	1.31	1998	0.70	1.12	0.51	2.43
1984	0.59	0.99	0.62	1.57	1999	0.49	0.84	0.38	1.82
1985	0.89	1.35	0.84	2.14	2000	0.50	0.83	0.38	1.80
1986	0.74	1.20	0.75	1.91	2001	0.60	1.01	0.46	2.20
1987	0.70	1.30	0.81	2.06	2002	0.51	0.91	0.41	1.99
1988	0.86	1.56	0.98	2.48	2003	0.42	0.70	0.32	1.51
1989	0.58	1.10	0.69	1.75	2004	0.39	0.61	0.28	1.33
1990	0.66	1.18	0.74	1.88	2005	0.40	0.58	0.26	1.26
1991	0.50	0.95	0.59	1.51	2006	0.41	0.54	0.25	1.18
1992	0.92	1.16	0.72	1.85	2007	0.51	0.72	0.32	1.56
1993	1.06	1.16	0.72	1.85	2008	0.48	0.66	0.30	1.44
					2009	0.46	0.73	0.33	1.59
					2010	0.44	0.85	0.38	1.85
					2011	NA	0.90	0.38	2.10
					2012	0.74	0.96	0.43	2.11
					2013	0.60	0.68	0.31	1.49
					2014	0.35	0.59	0.27	1.29
					2015	0.56	0.98	0.44	2.15
					2016	0.62	0.87	0.39	1.91
					2017	0.86	1.10	0.50	2.41
					2018	0.51	0.71	0.32	1.55

Table 2. Northwest. Nominal and standardized CPUEs for periods 1979-93 and 1994-2018.



Figure 7. Northeast. Posterior marginal distribution of the fixed (intercept), precision for random effects, temporal correlation term (Rho), and spatial field parameters (Thetas).



Figure 8. Northeast. Mean of latent spatial field.



Figure 9. Northeast. Historical changes of CPUEs. Line is standardized CPUE and filled area is 95% credible interval. Points denote nominal CPUE. Note the different scale of y axis for CPUE between the periods.

year	nominal	Standardized	2.50%	97.50%	year	nominal	Standardized	2.50%	97.50%
1979	0.35	0.89	2.41	0.33	1994	0.38	3.32	8.95	1.23
1980	0.46	0.93	2.52	0.34	1995	0.38	3.11	8.38	1.15
1981	0.37	0.75	2.01	0.27	1996	0.41	3.58	9.66	1.33
1982	0.36	0.78	2.10	0.29	1997	0.46	3.59	9.68	1.33
1983	0.34	0.76	2.05	0.28	1998	0.41	3.07	8.29	1.14
1984	0.34	0.82	2.20	0.30	1999	0.38	2.93	7.90	1.09
1985	0.41	0.94	2.53	0.35	2000	0.32	2.35	6.33	0.87
1986	0.39	0.85	2.29	0.31	2001	0.25	1.89	5.10	0.70
1987	0.55	0.90	2.42	0.33	2002	0.21	1.73	4.67	0.64
1988	0.50	0.92	2.47	0.34	2003	0.27	2.15	5.80	0.79
1989	0.42	0.90	2.42	0.33	2004	0.20	1.72	4.66	0.64
1990	0.35	0.85	2.28	0.31	2005	0.26	2.08	5.63	0.77
1991	0.43	0.92	2.47	0.34	2006	0.23	1.68	4.52	0.62
1992	0.30	0.65	1.75	0.24	2007	0.33	2.10	5.67	0.78
1993	0.43	0.98	2.65	0.36	2008	0.36	2.08	5.60	0.77
					2009	0.31	2.19	5.91	0.81
					2010	0.27	2.09	5.64	0.77
					2011	0.27	1.96	5.29	0.73
					2012	0.25	1.96	5.29	0.73
					2013	0.36	2.30	6.22	0.85
					2014	0.50	3.02	8.15	1.12
					2015	0.53	3.02	8.16	1.12
					2016	0.69	4.23	11.42	1.57
					2017	0.49	3.04	8.20	1.13
					2018	0.56	3.26	8.81	1.21

Table 3. Northeast. Nominal and standardized CPUEs for periods 1979-93 and 1994-2018.



Figure 10. **Southwest.** Posterior marginal distribution of the fixed (intercept), precision for random effects, temporal correlation term (Rho), and spatial field parameters (Thetas).



Figure 11. Southwest. Mean of latent spatial field.



Figure 12. **Southwest.** Historical changes of CPUEs. Line is standardized CPUE and filled area is 95% credible interval. Points denote nominal CPUE. Note the different scale of y axis for CPUE between the periods.

year	nominal	Standardized	2.50%	97.50%	year	nominal	Standardized	2.50%	97.50%
1979	0.51	0.78	0.26	2.28	1994	0.62	2.42	12.10	0.46
1980	0.33	0.81	0.27	2.38	1995	0.39	1.54	7.70	0.29
1981	0.37	0.73	0.25	2.13	1996	0.34	1.38	6.91	0.26
1982	0.24	0.65	0.22	1.91	1997	0.34	1.40	6.99	0.27
1983	0.25	0.80	0.27	2.34	1998	0.26	1.00	4.98	0.19
1984	0.42	1.25	0.42	3.69	1999	0.22	0.80	4.02	0.15
1985	0.66	1.58	0.53	4.63	2000	0.25	0.72	3.60	0.14
1986	0.39	1.22	0.41	3.60	2001	0.18	0.63	3.17	0.12
1987	0.47	1.28	0.43	3.76	2002	0.15	0.62	3.10	0.12
1988	0.71	1.33	0.45	3.92	2003	0.11	0.53	2.64	0.10
1989	0.54	1.10	0.37	3.22	2004	0.16	0.69	3.46	0.13
1990	0.79	1.50	0.51	4.41	2005	0.19	0.78	3.91	0.15
1991	0.66	1.10	0.37	3.23	2006	0.25	0.86	4.31	0.16
1992	0.64	1.43	0.48	4.20	2007	0.23	0.69	3.45	0.13
1993	0.60	1.37	0.46	4.03	2008	0.30	0.84	4.20	0.16
					2009	0.38	1.04	5.21	0.20
					2010	0.35	1.13	5.67	0.22
					2011	0.37	0.99	4.97	0.19
					2012	0.31	0.97	4.84	0.18
					2013	0.28	0.83	4.13	0.16
					2014	0.26	0.77	3.86	0.15
					2015	0.28	0.88	4.41	0.17
					2016	0.43	1.28	6.38	0.24
					2017	0.46	1.30	6.52	0.25
					2018	0.37	1.24	6.19	0.24

Table 4. Southwest. Nominal and standardized CPUEs for periods 1979-93 and 1994-2018.



Figure 13. **Southeast.** Posterior marginal distribution of the fixed (intercept), precision for random effects, temporal correlation term (Rho), and spatial field parameters (Thetas). Note no spatial model with no autoregressive (m_zip_glmm) during 1994-2018.



Figure 14. **Southeast.** Mean of latent spatial field during 1979-1993. The lack of the figure during 1994-2018 is due to the non-spatial model (m_zip_glmm) during the period.



Figure 15. **Southeast.** Historical changes of CPUEs. Line is standardized CPUE and filled area is 95% credible interval. Points denote nominal CPUE. Note the different scale of y axis for CPUE between the periods.

year	nominal	Standardized	2.50%	97.50%	year	nominal	Standardized	2.50%	97.50%
1979	0.12	0.66	4.52	0.10	1994	0.16	0.12	0.10	0.14
1980	0.13	0.90	6.14	0.13	1995	0.18	0.12	0.10	0.14
1981	0.15	0.77	5.21	0.11	1996	0.25	0.13	0.11	0.16
1982	0.08	0.72	4.91	0.10	1997	0.34	0.14	0.12	0.17
1983	0.14	0.76	5.16	0.11	1998	0.14	0.12	0.10	0.14
1984	0.22	0.78	5.28	0.11	1999	0.20	0.13	0.11	0.15
1985	0.26	0.90	6.15	0.13	2000	0.19	0.12	0.10	0.14
1986	0.05	0.81	5.51	0.12	2001	0.13	0.09	0.07	0.11
1987	0.12	0.90	6.11	0.13	2002	0.16	0.10	0.08	0.12
1988	0.14	1.35	9.16	0.20	2003	0.13	0.10	0.08	0.12
1989	0.10	1.00	6.78	0.14	2004	0.20	0.10	0.08	0.12
1990	0.11	0.74	5.01	0.11	2005	0.10	0.08	0.06	0.09
1991	0.18	0.48	3.27	0.07	2006	0.21	0.09	0.07	0.11
1992	0.08	0.35	2.40	0.05	2007	0.36	0.11	0.09	0.14
1993	0.22	0.54	3.72	0.08	2008	0.28	0.09	0.07	0.10
					2009	0.24	0.08	0.06	0.09
					2010	0.37	0.08	0.07	0.10
					2011	0.46	0.10	0.08	0.12
					2012	0.43	0.08	0.07	0.10
					2013	0.50	0.10	0.08	0.12
					2014	0.52	0.10	0.08	0.12
					2015	0.42	0.09	0.07	0.10
					2016	0.31	0.06	0.05	0.08
					2017	0.28	0.06	0.05	0.07
					2018	0.27	0.06	0.05	0.08

Table 5. Southeast. Nominal and standardized CPUEs for periods 1979-93 and 1994-2018.



Figure 16. Interannual variations of relative standardized CPUE of Swordfish for the four areas in the Indian Ocean by Japanese longline fisheries during the two periods: 1979-1993 and 1994-2018 from this study (blue solid lines: point estimates, blue broken lines: 95% credible interval) and Ijima (2017) who made the analysis using the same logbook during 1976-2015 (red solid lines: point estimates, red broken lines: 95% confidence interval). The horizontal black broken lines for the mean relative standardized CPUEs equaled to 1 during the periods are inserted.