Standardized CPUE of swordfish (*Xiphias gladius*) from Indonesian tuna longline fleets in the north-eastern Indian Ocean

Bram Setyadji[1], Denham Parker[2], Sheng-Ping Wang[3] & Zulkarnaen Fahmi[1]

[1]*Research Institute for Tuna Fisheries, Bali, Indonesia*

[2]*Department of Agriculture, Forestry and Fisheries, South Africa*

[3]*Department of Environmental Biology and Fisheries Science, National Taiwan Ocean University, Keelung, Taiwan*

Abstract

Introduction

Swordfish (*Xiphias gladius*) is a large oceanic apex predator inhabits all the world's oceans. It is predominantly known as a subject of exploitation worldwide, mainly in the Pacific Ocean, Atlantic Ocean, and Mediterranean Sea (Tserpes & Tsimenides, 1995). Throughout the Indian Ocean, swordfish are primarily caught by longline fisheries, and the commercial harvest was first recorded by the Japanese in the early 1950s as bycatch of their tuna longline fisheries (IOTC-WPB16, 2018). Since 1990s the catches of swordfish increased sharply to a peak of 35,000 tons in 1998 (IOTC-WPB16, 2018) due to the growing shift of catching tunas to swordfish by Taiwanese longline fleets, the increasing number of longline fleets operations from various nations (e.g. Indonesia, Australia, La Reunion, Seychelles and Mauritius), and arrival of longline fleets from the Atlantic Ocean (e.g. Portugal, Spain and United Kingdom).

In recent years (2013-2017), Indonesian fleets are responsible for approximately 20% of the total catch of swordfish in the Indian Ocean (~8,000 MT), followed by Taiwan (17%), Sri Lanka (12%) and Spain (12%) (IOTC-WPB16, 2018). However, the total catch was revised to just under ~3,000 MT (9%) due to the refined methodology on catch estimation provided by the IOTC secretariat (IOTC-WPDCS14, 2018). In addition, the revision also aligned with the impact of Ministerial Regulation No. 56/2014 and No. 57/2014 about the moratorium on foreign fishing vessels and prohibition of transshipment at sea within Indonesia national jurisdiction, which resulted in a significant reduction of longline vessel operations from 584 in 2015 to 271 in 2016.

Our analytical objective was to investigate how the data-limited of swordfish fishery can construct a fairly robust relative abundance indices amid the "spatial gap" of the existing dataset for standardized CPUE in the eastern Indian Ocean (e.g., Japanese and Taiwanese longline dataset). We believe the results are valuable as an important information to assess the status of swordfish in the Indian Ocean.

Materials and Methods

Data Collection

This research analyzed the data gathered by the Indonesian scientific observers on commercial tuna longline vessels, which are mainly situated in Benoa Fishing Port, Bali. The observation program started in 2005 through an Australia-Indonesia collaboration (Project FIS/2002/074 of Australian Centre for International Agricultural Research), and since 2012 it has been conducted by the Research Institute for Tuna Fisheries (RITF Indonesia).

A total of 3,014 set-by-set data span in detail 1x1 degree latitude and longitude grid from January 2006 to December 2019 were obtained from Indonesia scientific observer, which covers commercial tuna longline vessels mostly based in Port of Benoa, Bali. Fishing trips usually last from three weeks to three months. Main fishing grounds cover from west to southern part of Indonesian waters, stretched from 75 ˚E to 35 ˚S (Figure 1). It also informed concerning the number of fish caught by species, total number of hooks, number of hooks between floats (HBF), start time of the set, start time of haul, soak time, and geographic position where the longlines were deployed into the water.

Cluster Analysis

Cluster analysis was performed based on species composition as proposed by (He et al., 1997). Further, for each set, the catch composition was calculated and expressed as proportions relative to the total of the four tuna species (e.g. albacore, bigeye tuna, southern bluefin tuna, and yellowfin tuna) and six billfish species (black marlin, blue marlin, striped marlin, sailfish, swordfish, short bill spearfish. Clustering a large dataset could be a major stumbling block. Sadiyah & Prisantoso (2011) and He et al. (1997) suggested to perform two step clustering methods, by using non-hierarchical k-means and followed by agglomerative hierarchical clustering. However, for this purpose the analyses were performed using NbClust package (Charrad et al., 2014), which was intended to perform k-means and hierarchical clustering with different distance measures and aggregation methods at one go.

The hierarchical cluster analysis with Ward minimum variance method (“ward.D2”) followed the criterion by Murtagh & Legendre (2014) was applied, which requires the dissimilarities to be squared before cluster updating. It then processed to the squared Euclidean distances across 15 indices in order to select the optimal number of clusters based on majority rule. The result then passed to CLARA (clustering large applications) under cluster package (Kaufman & Rousseeuw, 1990).

Data filtering

The major issue for modelling the abundance for billfishes from Indonesian tuna longline fishery was the high proportion number of zero-catch-per-set (Setyadji et al., 2018). It was acknowledged that predominance of zero catches could be driving the model outputs as the CPUE trends do not appear to be biologically plausible (IOTC-WPB16, 2018). Originally the mean annual proportion of zero catches from the data was quite high, around 70%. In attempt to reduce it, several ways were conducted as follows:

1. Data from 2005 was excluded from analysis, since it was the beginning of the scientific observer program, therefore it might contain species misidentification;
2. In general, spatial coverage of the scientific observer data covers from north eastern to south eastern Indian Ocean, ranged from 0-33S and 75-129E (Figure 1). However, it lacks temporal coverage, especially in the area below 20S, therefore, the data trimmed only from 0-17.5S to 75-129E (north-eastern Indian Ocean);
3. Excluding sets which doesn’t contain swordfish for the whole trip. As a result of the application of the procedures and criteria above, total number of sets used in the analysis is 2,935 and zero catch ratio were reversely up to 71%, however the filtering process was intended to find spatial consistency across years of observation.

CPUE Standardization

A delta-gamma GLM was applied to standardize the CPUE. As the approach of Wang (2018) with some modifications, the models were simply conducted with the main effects considered in this analysis were as follows:

a. Year : List of observation year (2006-2019), set as categorical variable;

b. Quarter : List of quarter of given year of observation (1-4), set as categorical variable;

c. Cluster : A result of cluster analysis (1-2), set as categorical variable;

d. Moon : Moon phase information is referring to the eight shapes of the directly sunlit portion of the moon that we can see from Earth. The moon phase was calculated using lunar package (Lazaridis, 2014);

e. Lat/Lon : Geographical information (latitude and longitude) is declared in actual condition (two decimal places) and squared to prevent any negative values. It incorporated as a continuous variable in the GLM analysis.

The interactions between main effects were not incorporated into the models to avoid overfitting. The gamma and delta models were conducted as follows:

Gamma model for CPUE of positive catch:

$$\log(CPUE) = \mu + Year + Quarter + Cluster + Moon + Lat^2 + Lon^2 + \varepsilon^{gamma} \quad \text{1)}$$

Delta model for presence and absence of catch:

$$\mathrm{PA} = \mu + Year + Quarter + Cluster + Moon + Lat^2 + Lon^2 + \varepsilon^{del} \quad \text{2)}$$

We used a forward approach to select the explanatory variables and the order they were included in the full model. The first step was to fit simple models with one variable at a time. The variable included in the model with lowest residual deviance was selected first. As second step the model with the selected variable then received other variables one at a time, and the model with lowest residual deviance was again selected. This procedure continued until residual deviance did not decrease as new variables were added to the previous selected model. Finally, all main effects and first order interactions were considered and a backward procedure based on Akaike Information Criterion (AIC) (Akaike, 1974), Bayesian Information Criterion (BIC) (Schwarz, 1978) and the values of the coefficient of determination (R^2) were used to select the final models.

The area-specific standardized CPUE trends were estimated based on the exponentiations of the adjust means (least square means) of the year effects (Butterworth, 1996; Maunder & Punt, 2004). The standardized relative abundance index was calculated by the product of the standardized CPUE of positive catches and the standardized probability of positive catches:

$$index = e^{\log(CPUE)}\left(\frac{e^{\tilde{P}}}{1+e^{\tilde{P}}}\right) \quad \text{3)}$$

where:

$CPUE$: is the adjust means (least square means) of the year effect of the gamma model;

$\tilde{P}$: is the adjust means (least square means) of the year effect of the delta model.

Result from other models, such as: negative binomial (NB), zero-inflated negative binomial (ZINB) also included for comparisons. Maps were produced using QGIS version 3.14 (QGIS Development Team, 2020) and the statistical analyses were carried out using R software version 4.0.2 (R Core Team, 2020) , particularly the package *pscl* (Jackman, 2017; Zeileis et al., 2008), *emmeans* (Lenth, 2018), *MASS* (Venables & Ripley, 2002), *Hmisc* (Harrell Jr. et al., 2018), *nlme* (Pinheiro et al., 2019) and *statmod* (Giner & Smyth, 2016).

Result and Discussion

Results

Cluster result

Based on majority rules (Figure 1), the optimal number of clusters was 2, cluster 1 consisted of ALB, followed by a relatively a balanced proportion of BET and YFT. Cluster 2 was dominated by BET, mostly operated on deep setting longline (Figure 2).

Descriptive Catch Statistic

Observers recorded catch and operational data at sea (after cleaning) following Indonesian tuna longline commercial vessels from 2006-2019. The combined dataset contained 94 trips, 2421 sets, 4,031 days-at-sea, and around 9 million hooks deployed, respectively (Table 1). The spatial data distributed mainly in eastern Indian Ocean with most of the positive catches occurred in the area south of Indonesian waters, between 0^{o}-17.5^{o} S and 75^{o}-125^{o} E (Figure 3).

CPUE data characteristics

SWO nominal CPUE series is presented in Figure 4. In general, the catches of SWO during the last decade were keep increasing, noticeably since 2011. The lowest CPUE recorded was in 2011 (0.12±0.04), as the highest was in 2019 (0.93±0.26). High spike of CPUE in 2019 was merely caused by filtering process, resulted in high positive catch proportion on less effort. Whereas, most of the observation on that particular year were conducted higher than 17.5 ^{o}S, which was the limit of north-eastern Indian Ocean area. On the other hand, the proportion of zero catch for SWO was relatively low for by-catch species. As opposed to nominal CPUE, the trend was lower in the last decade, varying annually between a maximum of 0.89±0.03 in 2011 and a minimum of 0.58±0.05 in 2019 with average value 0.71±0.02 (Figure 5).

CPUE standardization

Based on model selection, all effects were remained and statistically significant, except quarter for gamma model and longitude (Lon) for delta model. The deviance tables for selected gamma models are shown in the Table 2. The result indicated that the positive catch of swordfish strongly influenced by spatial (latitude and longitude) and temporal factors (year). Moreover, targeting effect and environmental variable (moon phase) also played significant part on the catch of SWO. The deviance tables for selected delta models are shown in the Table 3. Similar with gamma model, the catch probability of swordfish was probably be affected more by either spatial, temporal, environmental and targeting effect. However, latitude influenced the probability of catch better than longitude. The lower the latitude the higher the probability of catching SWO.

Estimations of standardized catch rates from several models are shown in Figure 6. Time trends of standardized CPUE as calculated using Poisson, negative binomial (NB), zero-inflated Poisson (ZIP) and delta-gamma (DELTA) models were similar from 2009 to 2019, however, time trend for ZIP was conflictive since 2017 and towards the end of the series. Zero-inflated negative binomial model experienced a non-convergence issue; therefore, it was not included in the final result. Moreover, DELTA model gave smoother trend, balanced trade-off and responded better to high proportion of zero catches and high CPUE spike in 2019 (Figure 7). However, further analysis is required for handling the high value of confidence interval on the final model.

Acknowledgement

The Authors would like to thank to all scientific observers of Research Institute for Tuna Fisheries (RITF) and national observers of Directorate General of Capture Fisheries (DGCF) for their contribution in collecting data throughout the years. We also would like to extend our gratitude to various organization, namely, Commonwealth Scientific and Industrial Research Organization (CSIRO), the Australian Centre for International Agricultural Research (ACIAR) and the Research Institute for Capture Fisheries (RCCF) for their funding support through research collaboration in the project FIS/2002/074: Capacity Development to Monitor, Analyze and Report on Indonesian Tuna Fisheries. The authors would also like to thank Dr. Humber Andrade, Dr. Rui Coelho and Dr. Philippe Sabarros for their valuable inputs on developing the script.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Butterworth, D. S. (1996). A possible alternative approach for generalised linear model analysis of tuna CPUE data. *Collective Volume of Scientific Papers*, *45*, 123–124.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*(6), 1–36.

Giner, G., & Smyth, G. K. (2016). statmod: Probability calculations for the inverse Gaussian distribution. *ArXiv Preprint ArXiv:1603.06687*.

Harrell Jr., F. E., Dupont, C., & Others. (2018). *Hmisc: Harrell Miscellaneous*. https://CRAN.R-project.org/package=Hmisc

He, X., Bigelow, K. A., & Boggs, C. H. (1997). Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery. *Fisheries Research*, *31*(1–2), 147–158.

IOTC-WPB16. (2018). *Report of the 16th Session of the IOTC Working Party on Billfish* (Working Party Report IOTC–2018–WPB16–R[E]; p. 97). Indian Ocean tuna Commission (IOTC).

IOTC-WPDCS14. (2018). *Report of the 14th Session of the IOTC Working Party on Data Collection and Statistics.* (Working Party Report IOTC–2018–WPDCS14–R[E]; p. 71). Indian Ocean tuna Commission (IOTC).

Jackman, S. (2017). *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory*. United States Studies Centre, University of Sydney. https://github.com/atahk/pscl/

Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data: An introduction to cluster analysis. In *Probability and Mathematical Statistics. Applied Probability and Statistics*. Wiley Series.

Lazaridis, E. (2014). *lunar: Lunar Phase & Distance, Seasons and Other Environmental Factors*. http://statistics.lazaridis.eu

Lenth, R. (2018). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans

Maunder, M. N., & Punt, A. E. (2004). Standardizing catch and effort data: A review of recent approaches. *Fisheries Research*, *70*, 141–159. https://doi.org/10.1016/j.fishres.2004.08.002

Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, *31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2019). *nlme: Linear and Nonlinear Mixed Effects Models*. https://CRAN.R-project.org/package=nlme

QGIS Development Team. (2020). *QGIS Geographic Information System* (3.14 (Phi)) [Computer software]. Open Source Geospatial Foundation Project.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing* (4.0.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Sadiyah, L., & Prisantoso, B. I. (2011). Fishing strategy of the Indonesian tuna longliners in Indian Ocean. *Indonesian Fisheries Research Journal*, *17*(1), 29–35.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Setyadji, B., Wibawa, T. A., & Fahmi, Z. (2018). Catch per Unit of Effort (CPUE) Standardization of Black Marlin (Makaira Indica) Caught by Indonesian Tuna Longline Fishery in the Eastern Indian Ocean. *Paper Presented on 16th Working Party on Billfish, Cape Town, South Africa, 4-8 September 2018, IOTC-2018-WPB16-12*, 14.

Tserpes, G., & Tsimenides, N. (1995). Determination of Age and Growth of Swordfish, Xiphias Gladius L., 1758, in the Eastern Mediterranean Using Anal-Fin Spines. *Fishery Bulletin*, *93*(3), 594–602.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). Springer. http://www.stats.ox.ac.uk/pub/MASS4

Wang, S.-P. (2018). CPUE standardization of striped marlin (Tetrapturus audax) caught by Taiwanese large scale longline fishery in the Indian Ocean. *Paper Presented on 16th Working Party on Billfish, Cape Town, South Africa, 4-8 September 2018, IOTC–2018–WPB16–18*, 31.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8), 1–25.

Table 1. Summary of observed effort from Indonesian tuna longline fishery during 2006–2019. Results are pooled and also presented by year of observation. Operational parameters are means (upper entries) and standard error (lower parenthetical entries).

Year	Trips	Sets	Days at Sea	Total Hooks	Hooks per Set		Hooks per Float	
2006	10	248	401	339,811	1,370	(14.3)	11.1	(0.3)
2007	11	181	258	263,473	1,456	(24.5)	14.5	(0.4)
2008	12	331	404	387,345	1,170	(16.3)	12.5	(0.3)
2009	11	268	288	312,702	1,167	(13.6)	12.1	(0.3)
2010	6	166	152	221,274	1,333	(35.5)	13.6	(0.4)
2011	3	105	111	110,384	1,051	(17.0)	12.0	(0.0)
2012	4	100	192	112,878	1,129	(43.4)	14.0	(0.3)
2013	8	237	198	255,562	1,078	(13.5)	12.3	(0.1)
2014	6	184	265	216,705	1,178	(13.4)	15.0	(0.1)
2015	5	150	241	174,655	1,164	(11.8)	14.1	(0.3)
2016	3	130	383	175,868	1,353	(18.3)	11.3	(0.3)
2017	3	107	489	128,228	1,198	(18.1)	16.0	(0.1)
2018	5	130	321	166,110	1,278	(14.6)	15.7	(0.2)
2019	7	84	328	104,266	1,241	(19.4)	7.8	(0.5)

Table 2. The deviance table for selected gamma model.

Variable	Df	Deviance Resid.	Df. Resid.	Deviance	F	Pr(>F)	
NULL			700	221.18			
Year	13	9.052	687	212.13	2.6226	0.0014	**
Moon	7	9.1369	680	202.99	4.9163	<0.0000	***
cluster	1	3.7667	679	199.22	14.1873	0.0001	***
Lat	1	2.8006	678	196.43	10.5485	0.0012	**
Lon	1	16.6815	677	179.74	62.8308	<0.0000	***

Significant codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 3. The deviance table for selected delta model.

Variable	Df	Deviance Resid.	Df. Resid.	Deviance	Pr(>F)	
NULL			2349	2864.2		
Year	13	68.335	2336	2795.9	<0.0000	***
Moon	7	43.312	2329	2752.6	<0.0000	***
Quarter	3	17.087	2326	2735.5	0.0006	***
cluster	1	10.391	2325	2725.1	0.0012	**
Lat	1	34.14	2324	2691.0	<0.0000	***

Significant codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

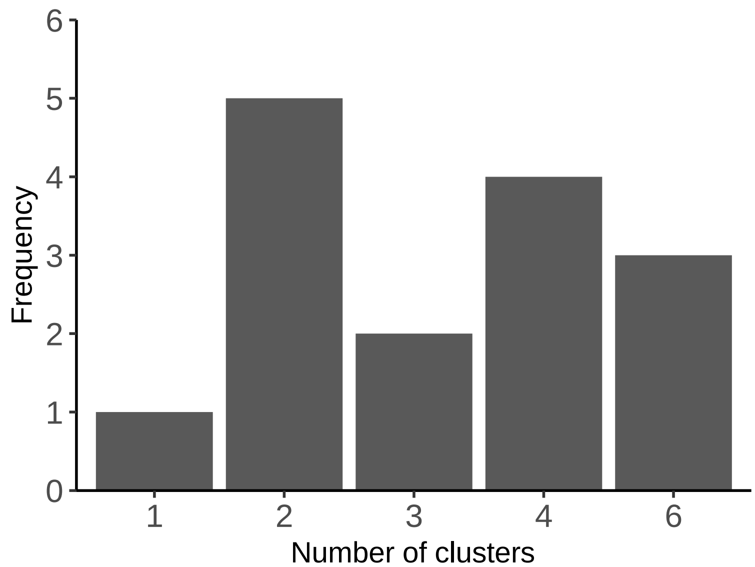

Figure 1. Selection on optimum number of clusters, based on the majority rules.

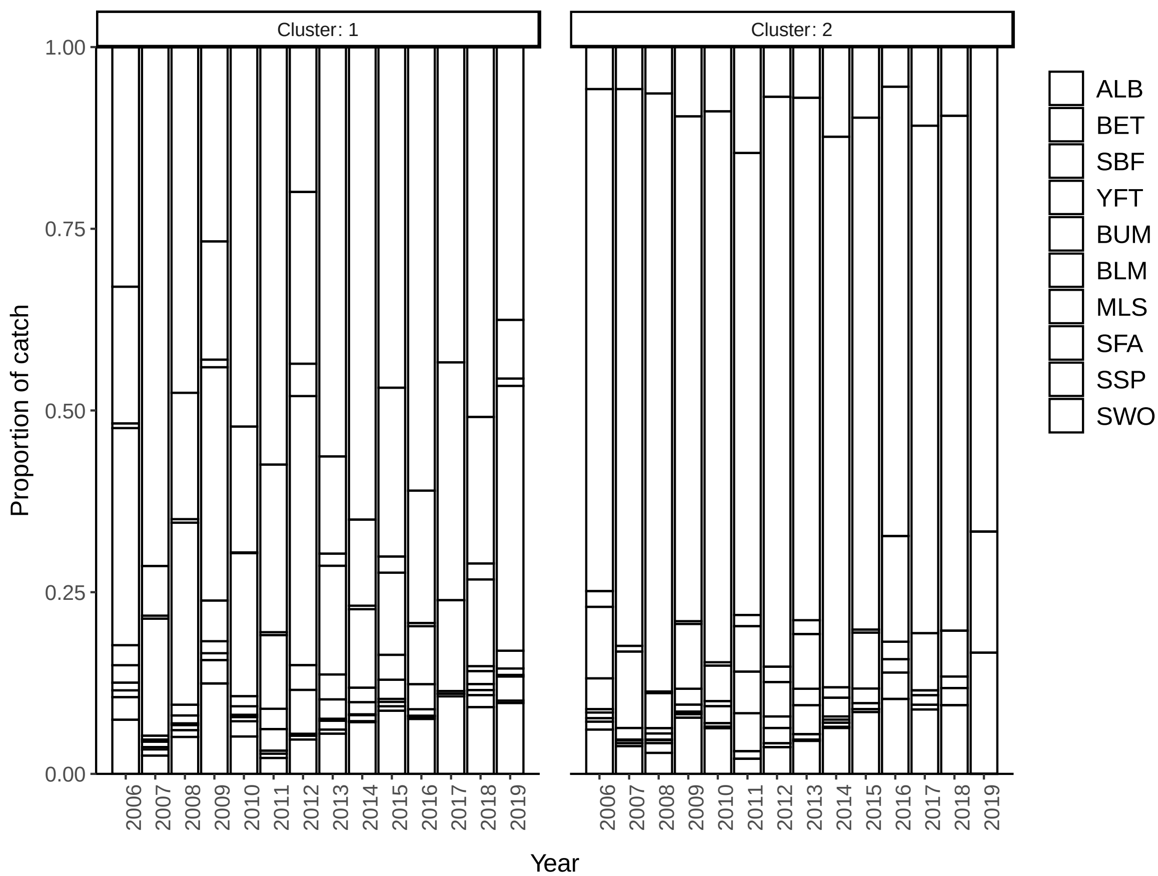

Figure 2. Catch proportions of SWO caught by Indonesian longline fleets operated in the north-eastern Indian Ocean.

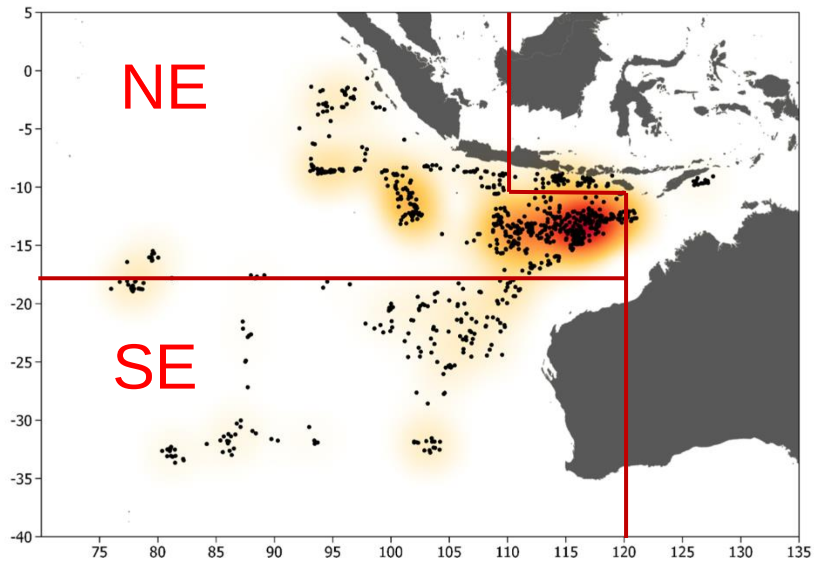

Figure 3. Area stratification used in the analysis (Wang, 2018) based on the aggregation of the relative sizes from nine IOTC statistics areas for swordfish in the Indian Ocean (Nishida & Wang, 2006)

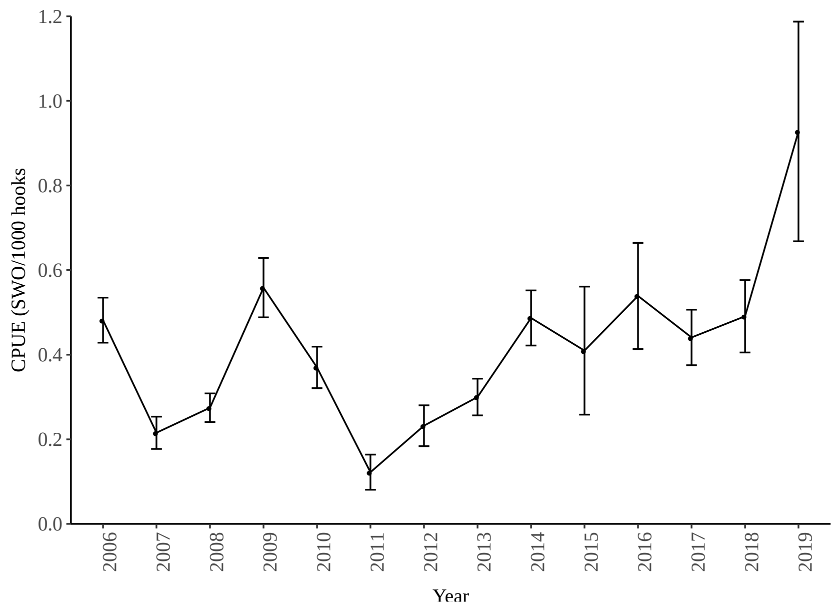

Figure 4. Nominal CPUE series (N/1000 hooks) for SWO from 2006 to 2019. The error bars refer to the standard errors.

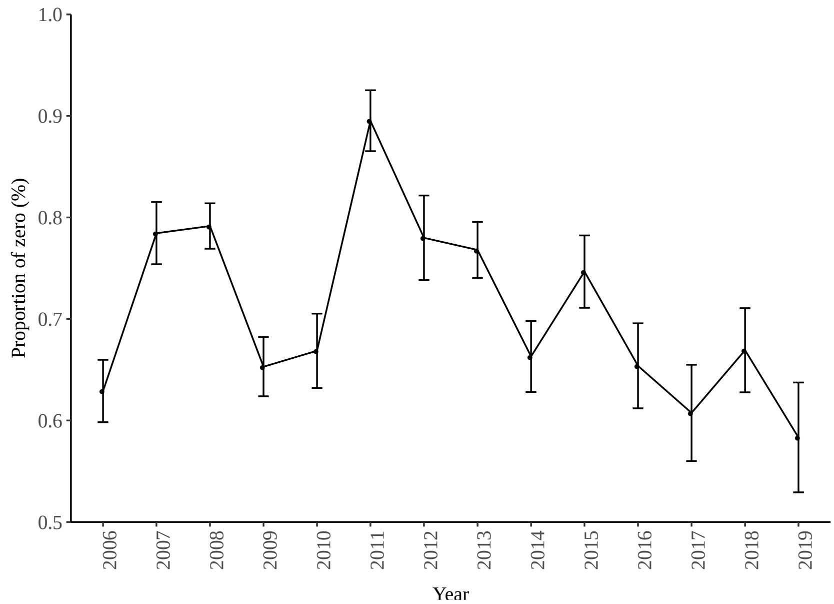

Figure 5. Proportion of zero-catch-per-set from 2006 to 2019. The error bars refer to the standard errors.

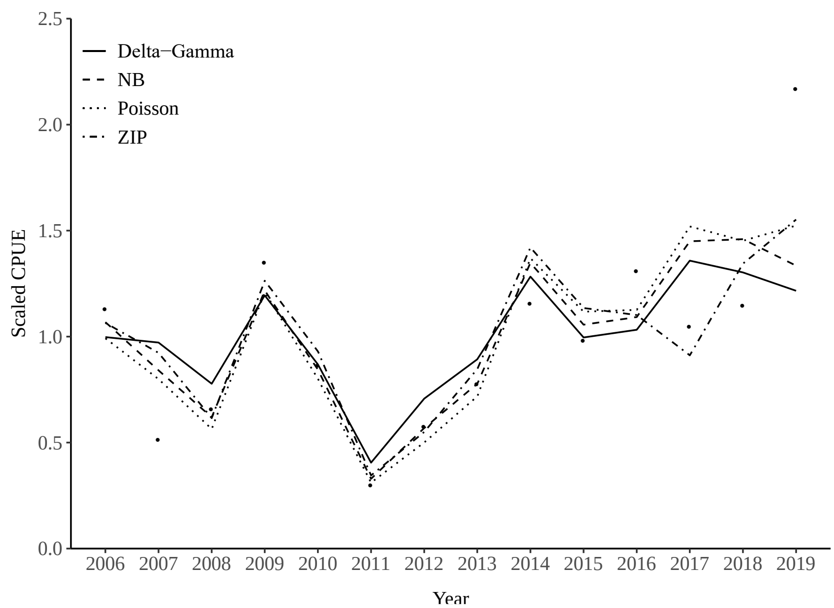

Figure 5. Standardized catch-per-unit-effort (CPUE) calculated using various models. Values were scaled by dividing them by their means.

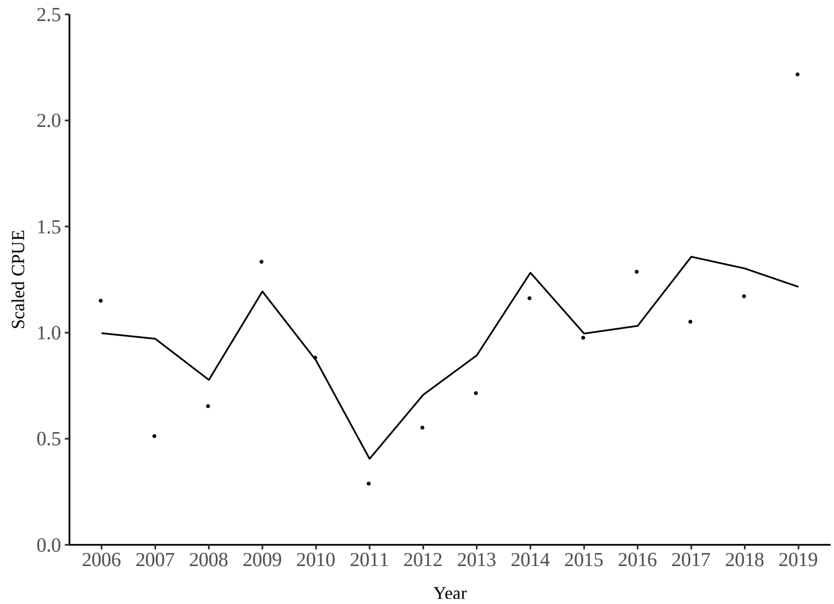

Figure 5. Final graph for standardized catch per unit effort (CPUE) of BUM calculated using delta-gamma model with 95% confidence interval (greyed area). Values were scaled by dividing them by their means.