Update on CPUE Standardization of Black Marlin (*Makaira indica*) from Indonesian Tuna Longline Fleets 2006-2020

Bram Setyadji[*,a], Denham Parker[b], Sheng-Ping Wang[c], Ririk Kartika Sulistyaningsih[a]

[a] *Research Institute for Tuna Fisheries, Bali, Indonesia*
[b] *Department of Agriculture, Forestry and Fisheries, South Africa*
[c] *Department of Environmental Biology and Fisheries Science, National Taiwan Ocean University, Keelung, Taiwan*

Abstract

Black marlin (*Makaira indica*) is commonly caught as frozen by-catch from Indonesian tuna longline fleets. Its contribution estimated 18% (~2,500 tons) from total catch in Indian Ocean. Relative abundance indices as calculated based on commercial catches are the input data for several to run stock assessment analyses that provide models to gather information useful information for decision making and fishery management. In this paper a Delta-Lognormal Model (GLM) was used to standardize the catch per unit effort (CPUE) and to calculate estimate relative abundance indices based on the Indonesian longline dataset. Data was collected from August 2005 to December 2020 through scientific observer program. Most of the vessels monitored were based in Benoa Port, Bali. The result showed that Year, latitude and longitude were the only variables remained and statistically significant in the lognormal model, whereas moon and latitude were excluded from the delta model. The positive catch of black marlin (BLM) was allegedly influenced by spatial (latitude and longitude) and temporal factor (year), an indication of sporadic catch instead of targeting. In addition, according to delta model, targeting effect (cluster) and horizontal movement (longitude) allegedly played significant part on possibility of catching BLM. Whereas, BLM was more likely caught when incorporated in cluster 1 (mixed targeting ALB, BET and YFT) and cluster 3 (targeting ALB). However, high uncertainties seemed as lingering issue, which is inevitable due to low coverage of scientific observer data.

Keywords: abundance indices, stock assessment, Generalized Linear Model (GLM), by-catch

Introduction

Black marlin (*Makaira indica*) is an apex predator, highly migratory species and considered as a non-target species from indonesian industrial and small-scale tuna fishery (Nugraha and Setyadji, 2013; Sulistyaningsih et al., 2011; Widodo et al., 2016). It is the second highest landed billfish species after swordfish (Setyadji and Nugraha, 2012), and known to have high commercial value, both in the tropical and subtropical Indian and Pacific Ocean (Nakamura, 1985). In general, it mostly caught between 20ºN and 45ºS, but more often off the western coast of India and the Mozambique Channel (IOTC-WPB18, 2020).

In Indian Ocean, black marlin was largely caught by gillnets (~59%), followed by longlines (~19%), with remaining catches recorded under troll and hand lines (IOTC-WPB18, 2020). Contribution of black marlin from Indonesian fleet between 2014-2018 was around 10% (under 2000 tons) of total catch in Indian Ocean, ranked fourth after Iran, India, and Sri Lanka (IOTC-WPB18, 2020). Results of latest stock assessment as calculated using JABBA (Parker et al., 2019), suggested that the stock is not subject to overfishing and is currently not overfished. However, these status estimates are subject to a high degree of uncertainty and should be interpreted with caution.

Estimations of relative abundance indices can support the use of more detailed models, which can provide important information concerning black marlin status stock. Statistical models such as Generalized Linear Models (GLM) can be used to "standardize" commercial catch per unit effort (CPUE) in order to calculate relative abundance indices, which are the input data for several stock assessment models. Estimations of

*Corresponding Author
Email addresses: bramsetyadji@kkp.go.id (Bram Setyadji), DParker@environment.gov.za (Denham Parker), wsp@mail.ntou.edu.tw (Sheng-Ping Wang), rk.sulistyaningsih.11@gmail.com (Ririk Kartika Sulistyaningsih)

standardized CPUE of Indian Ocean black marlin are limited, especially if compared to other billfish species as swordfish (*Xiphias gladius*), blue marlin (*Makaira mazara*), and striped marlin (*Tetrapturus audax*). Lack of detailed data was touted to be the key factor.

The initial standardized CPUE was calculated using Japanese longline fishery statistics for 1967-1997 (Uozumi, 1998). In response to data scarcity, Indonesia proposed a scientific observer program since mid-2005, which has been providing information concerning black marlin caught by longline boats operating in the northeastern Indian Ocean (Setyadji et al., 2018). In this paper we used a GLM to calculate standardized CPUE of black marlin caught by Indonesian longline fleet in the Eastern Indian Ocean. Results are useful to assess the status of the stock of black marlin, which is an important fishery resource 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 five billfish species (i.e., black marlin, blue marlin, striped marlin, sailfish, swordfish). Clustering a large dataset could be a major stumbling block. Sadiyah and Prisantoso (2011) 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 and 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 and 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 very high, close to 89%. In attempt to reduce it, several ways were conducted as follows:

1. Exclude 2005 data 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 0-33° S and 75-129° E. However, it lacks temporal coverage, especially in the area below 20S, therefore, the data trimmed only from 0-20° S to 75-129° E (northeastern Indian Ocean);

3. Exclude 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 was 3,050 and zero catch ratio were slightly reversed to ~85%. 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**, set as categorical variable (2006-2020);
b. **Quarter**, set as categorical variable (1-4);
c. **Cluster**, set as categorical variable (1-3);
d. **Moon**, referred 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**, defined as georeferenced information in 5x5 degree and presented in absolute value to avoid negative mark. 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:
\newline Gamma model for CPUE of positive catch:

$$log(CPUE) = +Year + Quarter + Cluster + Moon + Lat + Lon + ^{gamma}$$

Delta model for presence and absence of catch:

$$PA = +Year + Quarter + Cluster + Moon + Lat + Lon + ^{del}$$

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).

The area-specific standardized CPUE trends were estimated based on the exponentiation of the adjust means (least square means) of the year effects (Butterworth, 1996; Maunder and 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:

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

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.

Maps were produced using QGIS version 3.14 (QGIS Developer Team, 2020) and the statistical analyses were carried out using R software version 4.1.0 (R Core Team, 2021), particularly the package *emmeans* (Lenth, 2018), and *MASS* (Venables and Ripley, 2002).

Results

Cluster Result

Based on majority rules (Figure 1), the optimal number of clusters was three. Cluster 1 was consisted of mixed ALB, YFT and BET, whereas cluster 2 was dominantly filled with BET, and the biggest proportion in cluster 3 was ALB (Figure 2).

Descriptive Statistic

Observers recorded catch and operational data at sea (after cleaning) following Indonesian tuna longline commercial vessels from 2006-2019. The filtered dataset contained 80 trips, 2248 sets, and around 2.7 million hooks observed, respectively (Table 1). The distribution of sets mainly gathered in area of eastern Indian Ocean with most of the positive catches occurred in the area south of Indonesian waters, between 0°-20° S and 75°-125° E (Figure 3).

CPUE Data Characteristics

BLM nominal CPUE series is presented in Figure 4. In general, the catches of BLM during the last decade were fluctuating, with tendency of rising since 2014. The lowest CPUE recorded was in 2018 (0.06$\pm$0.02), as the highest was in 2009 (0.28$\pm$0.04). No data in 2017 and very low CPUE generated a year after was merely caused by data limitation (low coverage), rather than loss in abundance. Whereas, most of the observation on that particular year were conducted higher than 20° S, which included in the north-eastern Indian Ocean area. In addition, the proportion of zero catch for BLM was quite high. As opposed to nominal CPUE, the trend was varying annually between a maximum of 0.92% in 2008 and a minimum of 0.75% in 2020 with average proportion 0.84% year^{-1} (Figure 5).

CPUE Standardization

Year, latitude and longitude were the only variables remained and statistically significant in the lognormal model (Table 2), whereas moon and latitude were excluded from the delta model (Table 3). The positive catch of black marlin (BLM) was allegedly influenced by spatial (latitude and longitude) and temporal factor (year), an indication of sporadic catch instead of targeting. In addition, according to delta model, targeting effect (cluster) and horizontal movement (longitude) allegedly played significant part on possibility of catching BLM. Whereas, BLM was more likely caught when incorporated in cluster 1 (mixed targeting ALB, BET and YFT) and cluster 3 (targeting ALB).

Overall, the standardized CPUE trend was relatively stable over time with some variations in the last 5 years. However, high uncertainties seemed as lingering issue, which is inevitable due to low coverage of scientific observer data.

Acknowledgement

The Authors would like to thank to all scientific observers of Research Institute for Tuna Fisheries (RITF) 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 Center for International Agricultural Research (ACIAR) and the Research Institute for Capture Fisheries (RCCF) through research collaboration in the project FIS/2002/074: Capacity Development to Monitor, Analyze and Report on Indonesian Tuna Fisheries. Also, Fisheries Improvement Project (FIP) for their funding support during Covid-19 pandemic. 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, 716723. doi:https://doi.org/10.1109/TAC.1974.1100705

Butterworth, D.S., 1996. A possible alternative approach for generalised linear model analysis of tuna CPUE data. Collective Volume of Scientific Papers 45, 123124.

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, 136.

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, 147158.

IOTC-WPB16, 2018. Report of the 16th session of the IOTC working party on billfish. Cape Town, South Africa.

IOTC-WPB18, 2020. Report of the 18th Session of the IOTC Working Party on Billfish (Working {{Party Report}} No. IOTC–2020–WPB18–R[E]). Indian Ocean tuna Commission (IOTC), Microsoft Teams Online.

Kaufman, L., Rousseeuw, P.J., 1990. Finding groups in data: An introduction to cluster analysis. Wiley Series.

Lazaridis, E., 2014. Lunar: Lunar phase & distance, seasons and other environmental factors.

Lenth, R., 2018. Emmeans: Estimated marginal means, aka least-squares means.

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research 70, 141–159. doi: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, 274–295. doi:10.1007/s00357-014-9161-z

Nakamura, I., 1985. Billfishes of the world. An annotated and illustrated catalogue of marlins, sailfishes, spearfishes and swordfishes known to date. FAO species catalogue; FAO Fisheries Synopsis 5, 65.

Nugraha, B., Setyadji, B., 2013. Kebijakan pengelolaan hasil tangkapan sampingan tuna longline di Samudera Hindia. Jurnal Kebijakan Perikanan Indonesia 5, 67–71. doi:http://dx.doi.org/10.15578/jkpi.5.2.2013.67-71

Parker, D., Winker, H., Kerwath, S.E., 2019. Continuity runs of the Andrade (2016) Bayesian statespace surplus production model assessment of Indian ocean blue marlin (Makaira nigricans) stock using JABBA. Paper presented at 17th Working Party on Billfish, La Reunion, France, 9-12 September 2019. IOTC-2019-WPB17-20a 15 pp.

QGIS Developer Team, 2020. QGIS geographic information system.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sadiyah, L., Prisantoso, B.I., 2011. Fishing strategy of the indonesian tuna longliners in indian ocean. Indonesian Fisheries Research Journal 17, 2935. doi:http://dx.doi.org/10.15578/ifrj.17.1.2011.29-35

Setyadji, B., Andrade, H.A., Proctor, C.H., 2018. Standardization of Catch Per Unit Effort with High Proportion of Zero Catches: An Application to Black Marlin Istiompax indica (Cuvier, 1832) Caught by The Indonesian Tuna Longline Fleet in The Eastern Indian Ocean. Turkish Journal of Fisheries and Aquatic Sciences 19, 119–129. doi:10.4194/1303-2712-v19_2_04

Setyadji, B., Nugraha, B., 2012. By-catch of tuna longliner based at Benoa in the Indian Ocean. Jurnal Penelitian Perikanan Indonesia 18, 43–51. doi:http://dx.doi.org/10.15578/jppi.18.1.2012.43-51

Sulistyaningsih, R.K., Barata, A., Siregar, K., 2011. Perikanan pancing ulur tuna di kedonganan, bali. Jurnal Penelitian Perikanan Indonesia 17, 185–191. doi:http://dx.doi.org/10.15578/jppi.17.3.2011.185-191

Uozumi, Y., 1998. Standardization of catch per unit of effort for swordfish and billfishes caught by the Japanese longline fishery in the Indian Ocean. Paper presented at 7th Expert Consultation on Indian Ocean Tunas, Victoria, Seychelles, 9-14 November, 1998 13.

Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with s, Fourth. ed. Springer, New York.

Widodo, A.A., Prisantoso, B.I., Suprapto, S., 2016. Perikanan pancing ulur di samudera hindia: Hasil tangkapan ikan berparuh yang di daratkan di sendangbiru, malang, jawa timur. Jurnal Penelitian Perikanan Indonesia 18, 167173. doi:10.15578/jppi.18.3.2012.167-173

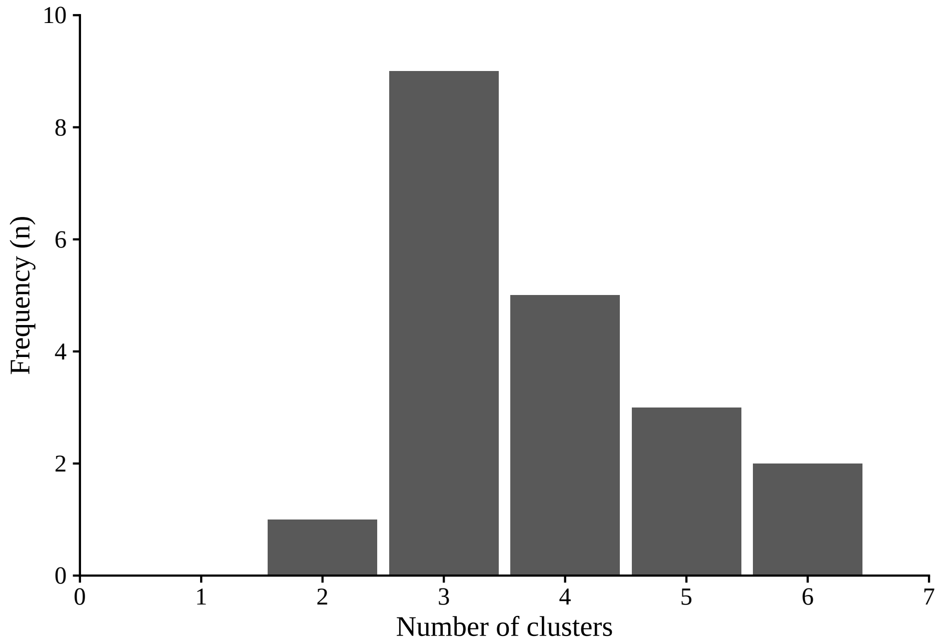

Figure 1. Selection of optimum number of clusters, based on the majority rules.

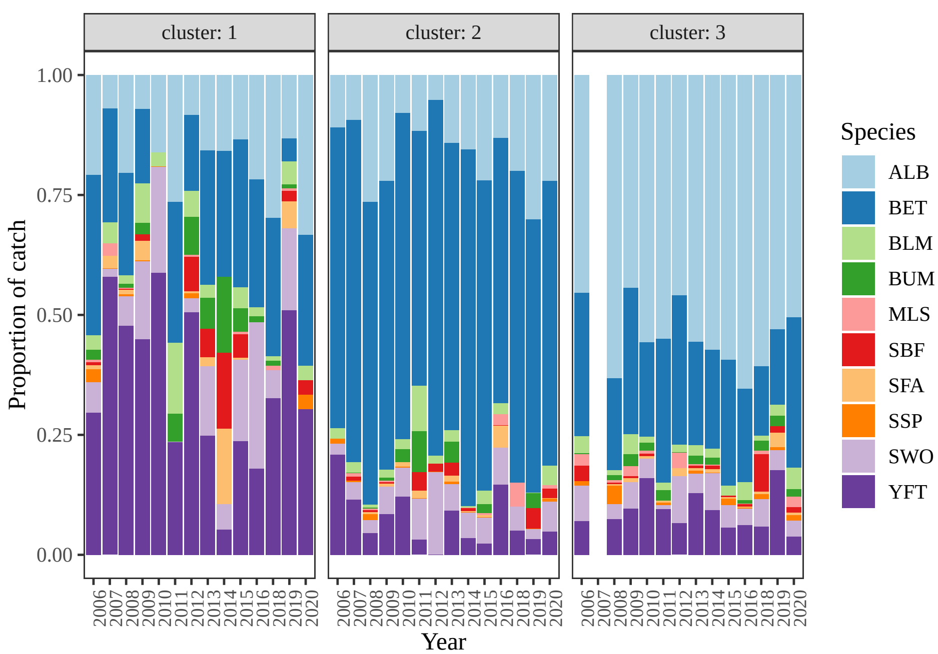

Figure 2. Catch proportions of SWO caught by Indonesian longline fleets operated in the north-eastern Indian Ocean.

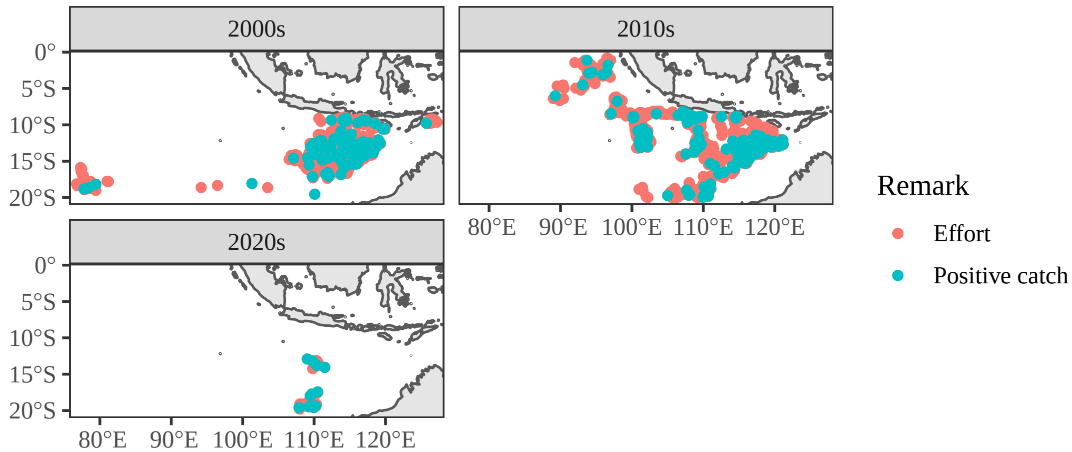

Figure 3. Effort and positive catch distribution of BLM by Indonesian longline fleets operated in the north-eastern Indian Ocean.

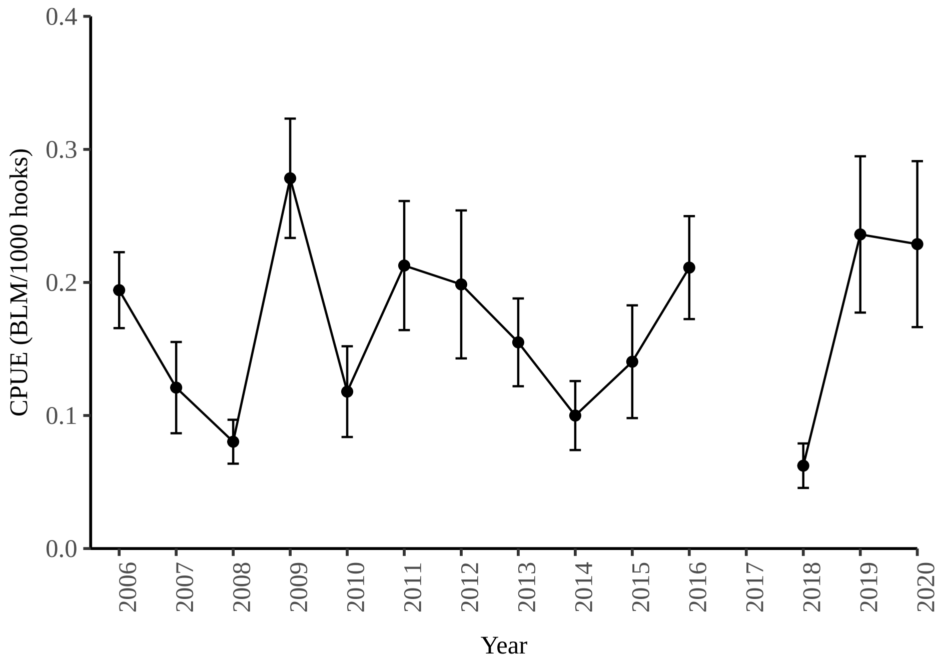

Figure 4. Nominal CPUE series (N/1000 hooks) for BLM from 2006 to 2020. The error bars refer to the standard errors.

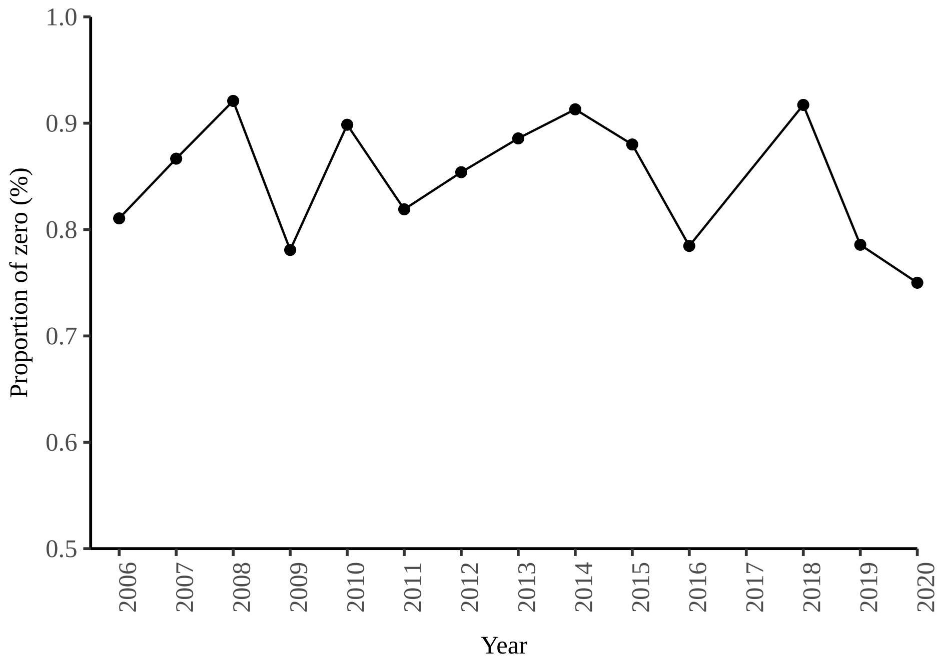

Figure 5. Proportion of zero-catch-per-set from 2006 to 2019. The error bars refer to the standard errors.

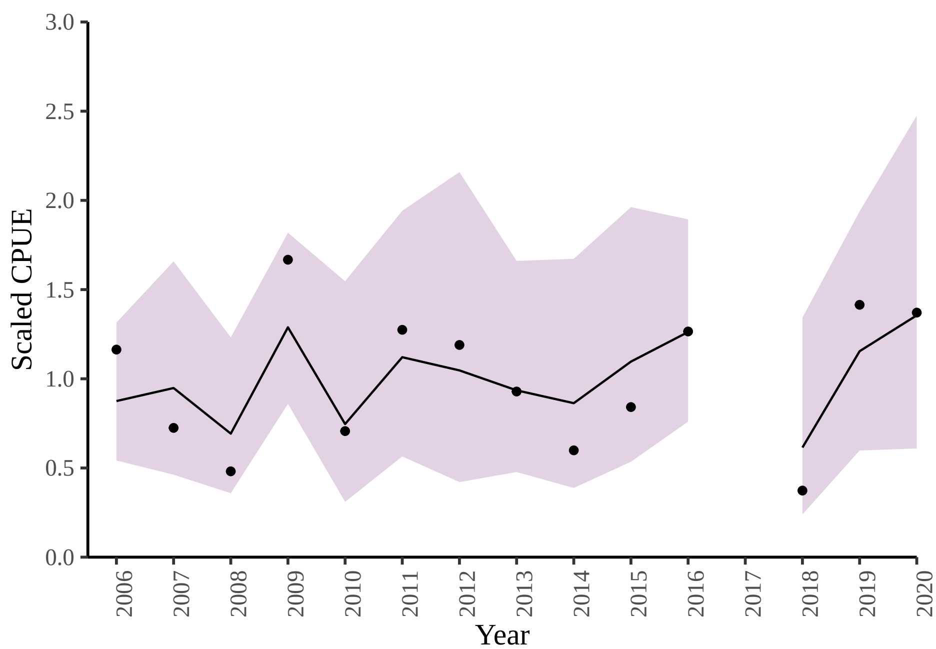

Figure 6. Standardized catch-per-unit-effort (CPUE) calculated using delta-lognormal model. Values were scaled by dividing them by their means.

Table 1. Summary of observed effort from Indonesian tuna longline fishery during 2006–2020. Results are pooled and also presented by year of observation

Year	Trips	Sets	Total Hooks	Mean Hooks	se	Mean HBF	se
2006	11	285	397081	1393.27	12.35	11.04	0.27
2007	7	120	159584	1329.87	19.23	14.38	0.54
2008	10	329	409116	1243.51	21.03	11.76	0.22
2009	7	219	258593	1180.79	15.42	11.30	0.35
2010	5	138	187674	1359.96	42.37	13.30	0.48
2011	3	105	110384	1051.28	16.97	12.00	0.00
2012	4	89	89538	1006.04	29.03	13.88	0.34
2013	7	210	231990	1104.71	14.11	12.40	0.15
2014	6	184	216705	1177.74	13.35	15.01	0.14
2015	5	150	174655	1164.37	11.81	14.15	0.26
2016	3	130	175868	1352.83	18.33	11.31	0.29
2018	4	157	211361	1346.25	16.00	14.90	0.22
2019	6	84	109150	1299.40	19.67	7.82	0.40
2020	2	48	65915	1373.23	17.80	13.62	0.14

Table 2. The deviance table for selected lognormal model.

	Df	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)
NULL	NA	NA	318	58.1860	NA	NA
Year	13	5.3300	305	52.8560	2.5002	0.0029
Lat2	1	1.4298	304	51.4262	8.7185	0.0034
Lon2	1	1.7372	303	49.6890	10.5931	0.0013

Table 3. The deviance table for selected delta model.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL	NA	NA	2247	1836.187	NA
Year	13	54.6699	2234	1781.517	0.0000
Quarter	3	22.3846	2231	1759.133	0.0001
cluster	3	10.2413	2228	1748.891	0.0166
Lon2	1	9.2891	2227	1739.602	0.0023