Catch per Unit of Effort (CPUE) Standardization of Black Marlin (*Makaira Indica*) Caught by Indonesian Tuna Longline Fishery in the Eastern Indian Ocean

Bram Setyadji[1*], Teja Arief Wibawa[2] and Zulkarnanen Fahmi[1]

[1]Research Institute for Tuna Fisheries, Bali
[2]Institute for Marine Research and Observation, Bali, Indonesia

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 Generalized Linear 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 2017 through scientific observer program (2005-2017) and national observer program (2016-2017). Most of the vessels monitored were based in Benoa Port, Bali. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to select the best models among all those evaluated. Zero-inflated negative binomial (ZINB) model and simple negative binomial (NB) model had the lowest AIC and BIC value, respectively. Time trends of standardized CPUE as calculated using NB and ZINB models were similar from 2005 to 2016, however, time trends are conflictive at the very end (2017). At this stage, there is no strong motivation to choose one of the two models (NB or ZINB), hence sensitivity analysis concerning between time series are an alternative when running stock assessment models.

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 of industrial and artisanal fisheries in Indonesian tuna longline fishery. It ranked second after swordfish in term of catch composition (Setyadji *et al.*, 2012). It is also known to have high commercial value in the tropical and subtropical Indian and Ocean Pacific (Nakamura, 1985). In Indian Ocean it has been caught between 20°N and 45°S, but more often off the western coast of India and the Mozambique Channel (IOTC, 2015).

* Corresponding Author: Jl. Mertasari No. 140, Sidakarya, Denpasar, Bali, Indonesia-80224
e-mail: bram.setyadji@gmail.com, phone: +62 361 726201

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, 2015). Contribution of black marlin from Indonesian fleet between 2011-2014 was around 18% (~2,500 tons) of total catch in Indian Ocean, ranked fourth after Iran, Sri Lanka and India (IOTC, 2015). Results of latest stock assessment as calculated using Stock Reduction Analysis (SRA), which is a data poor method, suggest that black marlin stock of the Indian Ocean is not overfished but subject to overfishing (IOTC, 2015). 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 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 hampers the calculation of standardized CPUE for black marlin. The last estimation was calculated using Japanese longline fishery statistics for 1967-1997 (Uozumi, 1998). However, since 2005, Indonesia through scientific observer program has been providing information concerning black marlin caught by longline boats operating in the east of Indian Ocean (Setyadji *et al*., 2014). In this paper we have 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

Fishery and Environmental Data

A total of 2,887 set-by-set data span in detail 1x1 degree latitude and longitude grid from August 2005 to December 2017 were obtained from Indonesia scientific observer and national observer program, 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. The response variable in the models was the catch of black marlin in number of fish. Year and quarter were used as categorical (factor) explanatory variables. Additional information was used as explanatory variables as follows:

a. Fishing area (AreaTree)

   Area stratification method was applied using GLM-tree approach proposed by Ichinokawa and Brodziak (2010); The algorithm showed that the area was divided into four categories (Figure 1).

b. Number of hooks between floats (HBF)

   Number of hooks between floats was set as a categorical variable in the model. It was assigned as 1 if HBF <10 hooks (surface longline), and 2 if HBF $\geq$10 hooks (deep longline) following Sadiyah *et al.* (2012);

c. Soak time

   Soak time was calculated as the time elapsed between the start of the fishing setting and the start of hauling of the longline. Soak time in the model was treated as continuous variable, thus the values were rounded to the nearest integer;

d. Moon phase (29.5 days) were categorized into two periods, as light and dark, and assumed the demilunes (first/last quarters), waxing and waning gibbous and full moon as light period, while new moon, waxing and waning crescent considered as dark period (Akyol, 2013).

e. Daily Mean Sea Surface Temperature (SST) was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/. The spatial resolution was ¼ degree global grid. To address any possibility of non-linear (quadratic) relationship between CPUE and SST, it was assigned as a quadratic variable (expressed in R as poly (SST, 2)) (Sadiyah *et al*., 2012) and incorporated as a continuous variable.

f. Daily Mean Sea Surface Height (SSH) was extracted from Copernicus Marine Service Products, namely GLOBAL_REANALYSIS_PHY_001_025 for 2005-2015 datasets and GLOBAL_ANALYSIS_FORECAST_PHYS_001_015 for 2016-2017 datasets. The spatial resolution was ¼ degree global grid. To address any possibility of non-linear (quadratic) relationship between CPUE and SST, it was assigned as a quadratic variable (expressed in R as poly (SST, 2)) and incorporated as a continuous variable.

CPUE standardization

We considered six GLM models for modeling the number of black marlin for modelling the nominal catch (number of fish) as response variable while effort was included in the models as an offset caught. These models are Poisson (P) and negative binomial (NB), which we refer to as the standard models, zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), Poisson hurdle (PH), and negative binomial hurdle (NBH) models.

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) and Bayesian Information Criterion (BIC) (Schwarz, 1978) were used to select the final models for the six approaches. We also rely in AIC and BIC to compare these models.

The qualities of the fittings were assessed by comparing the observed frequency distributions of the number of fishes caught to the predicted frequency distribution, as calculated using the selected models. Kolmogorov-Smirnov test was used to assess if the difference of the two distributions (observed and predicted) were significant. Maps were produced using QGIS version 2.14 (QGIS Developer Team, 2009) and the statistical analyses were carried out using R software version 3.3.3 (R Core Team, 2016), particularly the package *pscl* (Zeileis *et al*., 2008), *lsmeans* (Lenth, 2016), *MASS* (Venables & Ripley, 2002), *Hmisc* (Harrell Jr, 2017), and *statmod* (Giner & Smyth, 2016).

Results

Descriptive Catch Statistic

Scientific observers and national observers recorded catch and operational data at sea following Indonesian tuna longline commercial vessels from 2005-2017 and 2016-2017, respectively. The combined dataset contained 115 trips, 2887 sets, 3499 days-at-sea, and more than 3.5 million hooks deployed, respectively (Table 1). The spatial data distributed mainly in eastern Indian Ocean with most of the observation were conducted in the area south of Indonesian waters, between 0°-35° S and 75°-125° E.

CPUE data characteristics

BLM nominal CPUE series is presented in Figure 2. In general, the catches of BLM during the last decade were highly variable, but showing an increasing trend. The lowest CPUE recorded was in 2005 (0.05$\pm$0.19), as the highest was in 2009 (0.22$\pm$0.59). On the other hand, the proportion of zero catch for BLM was also very high, varying annually between a minimum of 0.82$\pm$0.38 in 2011 and a maximum of 0.95$\pm$0.23 in 2017 with average value 0.89$\pm$0.30 (Figure 3).

CPUE standardization

The number of parameters (k), AIC, BIC, logarithm of the likelihood (logLik), number of predicted zero catches, and *p* values of Kolmogorov-Smirnov test as calculated using six model structures (P, NB, ZIP, ZINB, HP and HNB are shown in Table 2. Interactions among variables were excluded to avoid overfitting on the models. A difference of 2 units in the AIC values is not strong evidence that one model is better than the other (Burnham and Anderson, 2002). Hence, zero-inflated negative binomial (ZINB) model and simple negative binomial (NB) model were the model which had the lowest AIC (2314.08) and BIC value (2499.11), respectively (see Table 2).

The number of zero catches in the database is 2,575. Hurdle models always predict the correct number of zeros due to its structure, but these models are more complex in the sense there are more parameters to estimate. Poisson and NB simple models are the more biased as indicated by the differences between the observed and the predicted number of zero catches. However, bias of all the models including the simple ones, were not strong as indicated by the p values calculated using Kolmogorov-Smirnov test to compare the observed and predicted distributions of number of fish. NB and ZINB models were selected to calculate standardized catch rate indices for black marlin as they were the models with low values of AIC and BIC, and because there were no evidence whether they biased based on Kolmogorov-Smirnov test. Hereafter only the results of NB and ZINB are showed. Summary of parameter estimations of NB and ZINB models are in Table 2 and 3, respectively. If we rely in AIC and BIC, four categorical and two quantitative explanatory variables were included in the NB model, though estimations of parameters for the levels of factors were not significantly different from zero. ZINB model contained five categorical and two quantitative explanatory variables, but the number of parameters is larger because it includes two sets of estimations, one for the binomial and one for the negative binomial part of the model.

Estimations of standardized catch rates are shown in Figure 4. Time trends of standardized CPUE as calculated using NB and ZINB models were similar from 2005 to 2016, however, time trends are conflictive at the very end (2017). As there is no strong reason to select one of these two standardized time series for stock assessment purposes, a sensitivity analysis is an alternative.

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 input on developing the R script.

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control AC 19: 716-723.

Akyol, O. (2013). The influence of the moon phase on the CPUEs of swordfish gillnet fishery in the Aegean Sea, Turkey. Turkish Journal of Fisheries and Aquatic Sciences, 13, 355-358. https://dx.doi.org/10.4194/1303-2712-v13_2_18

Burnham, K. P. and Anderson, D. R. 2002. Model Selection and Multimodel Inference. Second Edition. New York, Springer.

Giner, G & Smyth, GK (2016). statmod: probability calculations for the inverse Gaussian distribution. R Journal. 8(1): 339-351

Harrell Jr, F.E. 2017. Hmisc: Harrell Miscellaneous. R package version 4.0-3. https://CRAN.R-project.org/package=Hmisc

Ichinokawa, M., & Brodziak, J. (2010). Using adaptive area stratification to standardize catch rates with application to North Pacific BLMrdfish (*Xiphias gladius*). *Fisheries Research, 106*, 249–260. https://dx.doi.org/10.1016/j.fishres.2010.08.001

IOTC (Indian Ocean Tuna Commission). 2015. Report of the 13th Session of the IOTC Working Party on Billfish. Olhão, Portugal, 1–5 September 2015. *IOTC–2015–WPB13–R[E]:* 98 pp.

Lenth, R.V. 2016. Least-Squares Means: The R Package lsmeans. Journal of Statistical Software. 69(1): 1-33. doi:10.18637/jss.v069.i01

Nakamura, I. 1985. FAO species catalogue. Vol. 5. Billfishes of the world. An annotated and illustrated catalogue of marlins, sailfishes, spearfishes and BLMrdfishes known to date. *FAO Fish. Synop.*, 5 (125): 65.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Sadiyah, L., Dowling, N. and B.I. Prisantoso. 2012. Develoving Recommendations for Undertaking CPUE Standardisation using Observer Program Data. *Ind. Fish. Res. J.* 18(1): 19-33.

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

Setyadji, B., Jumariadi and B. Nugraha. 2012. Catch estimation and size distribution of billfishes landed in Port of Benoa, Bali. *Ind.Fish.Res.J.* 18(1): 35-40

Setyadji, B., Jatmiko, I., Wujdi, A. and B. Nugraha. 2014. Preliminary analysis of length – weight relationship of BLMrdfish (*Xiphias gladius*), black marlin (*Makaira indica*), and blue marlin (*Makaira nigricans*) caught by Indonesian longliners in the Indian Ocean. *Paper presented in the 12th Session of the IOTC Working Party on Billfish. Yokohama, Japan, 21–25 October 2014*. *IOTC–2014–WPB12–13*: 12 pp.

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

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Zeileis, A., Kleiber, C. & Jackman, S. 2008. Regression Models for Count Data in R. Journal of Statistical Software 27(8). URL http://www.jstatsoft.org/v27/i08/.

Table 1. Summary of observed fishing effort from Indonesian tuna longline fishery during 2005–2016. Results are pooled and also presented by year of observation. Operational parameters are means (upper entries) and standard deviations (lower parenthetical entries).

Year	Trips	Sets	Days at Sea	Total Hooks	Hooks per Set		Hooks per Float		Mean Latitude		Mean Longitude	
2005	9	108	117	157,065	1,454.31	(151.8)	18.6	(1.5)	14.3ºS	(1.0º)	111.8ºE	(2.1º)
2006	13	401	401	577,243	1,439.51	(214.9)	11.2	(3.9)	16.9ºS	(6.0º)	113.4ºE	(5.4º)
2007	13	265	258	406,135	1,532.58	(326.5)	14.0	(4.4)	17.0ºS	(6.4º)	103.5ºE	(13.3º)
2008	15	370	404	483,662	1,307.19	(385.9)	13.0	(4.5)	14.2ºS	(2.6º)	107.3ºE	(14.1º)
2009	13	283	288	323,042	1,141.49	(234.7)	12.1	(4.9)	11.4ºS	(3.3º)	113.2ºE	(5.6º)
2010	6	165	152	220,394	1,335.72	(457.5)	13.6	(5.2)	12.0ºS	(3.3º)	113.3ºE	(6.0º)
2011	3	105	111	110,384	1,051.28	(173.9)	12.0	-	13.7ºS	(0.9º)	117.4ºE	(1.3º)
2012	8	198	192	290,265	1,465.98	(559.1)	14.1	(2.3)	18.9ºS	(7.8º)	104.5ºE	(10.8º)
2013	7	225	198	252,919	1,124.08	(210.4)	12.7	(2.1)	12.4ºS	(1.1º)	114.6ºE	(6.6º)
2014	5	167	265	193,740	1,160.12	(176.9)	15.0	(2.0)	11.0ºS	(1.7º)	105.7ºE	(7.5º)
2015	5	148	241	172,463	1,165.29	(145.2)	14.1	(3.2)	10.8ºS	(2.7º)	103.8ºE	(8.1º)
2016	8	244	383	324,068	1,314.89	(146.4)	15.2	(6.4)	10.6ºS	(3.8º)	107.5ºE	(9.4º)
2017	10	218	489	279,204	1.214.04	(395.3)	17.2	(4.8)	11.8ºS	(8.9º)	99.1ºE	(4.4º)

Table 2. Summary of indicators as calculated using six model structures: Poisson (P), Negative Binomial (NB), Zero-inflated with Poisson (ZIP), Zero-inflated with Negative Binomial (ZINB), Hurdle with Poisson (HP), and Hurdle with Negative Binomial (HNB). The terms in the column at left indicate: number of parameters (k), Akaike (AIC) and Bayesian (BIC) Information Criteria, logarithm of the likelihood (logLik), number of predicted zero catches (zero), and *p* values as calculated using a Kolmogorov-Smirnov test.

Parameters	Model Structure					
	P	NB	ZIP	ZINB	HP	HNB
k	26	25	46	52	52	44
AIC	2454.96	2343.94	2354.98	2314.08	2378.12	2371.62
BIC	2610.12	2499.11	2629.50	2624.42	2688.45	2634.21
logLIk	-1201.48	-1145.97	-1131.49	-1104.04	-1137.06	-1140.81
zero	2547	2610	2571	2575	2575	2575
p.value	1.00	0.98	1.00	1.00	1.00	1.00

Table 3. Summary of parameter estimations of Negative Binomial model. Terms: SE – standard error, p – p values as calculated using Z test to assess difference from zero.

	Estimate	SE	p	
(Intercept)	-9.788	0.439	2.00E-16	***
AreaTree2	0.707	0.183	0.000106	***
AreaTree3	0.086	0.309	0.781503	
AreaTree4	1.157	0.267	1.43E-05	***
Year2006	0.667	0.432	0.122708	
Year2007	0.082	0.473	0.86219	
Year2008	-0.462	0.461	0.316088	
Year2009	0.621	0.436	0.154681	

	Estimate	SE	p	
Year2010	-0.710	0.515	0.168226	
Year2011	0.323	0.500	0.51756	
Year2012	0.272	0.488	0.57649	
Year2013	0.595	0.456	0.19186	
Year2014	0.029	0.480	0.951563	
Year2015	0.542	0.465	0.243423	
Year2016	0.856	0.443	0.053298	.
Year2017	0.360	0.557	0.517981	
poly(SSH,2)1	15.143	6.092	0.012933	*
poly(SSH,2)2	3.821	5.144	0.457575	
Moon2Light	-0.313	0.122	0.010202	*
Cat_HBFShallow	0.627	0.185	0.000702	***
Quarter2	0.038	0.203	0.852511	
Quarter3	-0.399	0.219	0.068383	.
Quarter4	0.029	0.219	0.89545	
poly(SST,2)1	23.548	10.182	0.020743	*
poly(SST,2)2	-14.483	7.760	0.061986	.

Table 4. Summary of parameter estimations of Zero Inflated Negative Binomial model. Terms: SE – standard error, p – p values as calculated using Z test to assess difference from zero.

	Zero			
	Estimate	SE	p	
(Intercept)	-11.03993	0.75206	2.00E-16	***
AreaTree2	0.15063	0.23561	0.52262	
AreaTree3	0.81846	0.61879	0.18594	
AreaTree4	0.59601	0.33978	0.07941	.
Year2006	-0.21315	0.5983	0.72165	
Year2007	-0.39151	0.61796	0.52637	
Year2008	-1.29747	0.60624	0.03234	*
Year2009	0.16807	0.60519	0.78124	
Year2010	-1.03802	0.68943	0.13216	
Year2011	-0.44349	0.64027	0.48852	
Year2012	-0.63369	0.6572	0.33493	
Year2013	0.25934	0.61932	0.6754	
Year2014	0.1659	0.67599	0.80613	
Year2015	-0.84008	0.64915	0.19562	
Year2016	-0.11405	0.60792	0.85119	
Year2017	0.36434	0.86745	0.67448	
poly(SSH,2)1	5.95115	0.57	0.56863	
poly(SSH,2)2	16.36711	1.735	0.08267	.

Zero				
	Estimate	SE	p	
Moon2Light	-0.24751	0.14123	0.07968	.
Cat_HBFShallow	0.58042	0.19988	0.00369	**
Quarter2	0.09931	0.22401	0.65751	
Quarter3	-0.53308	0.27562	0.0531	.
Quarter4	0.37895	0.25094	0.13102	
Soak_Time	0.21572	0.05317	4.97E-05	***
poly(SST,2)1	11.94188	0.622	0.53417	
poly(SST,2)2	-7.38654	-0.575	0.56528	
Log(theta)	-0.40771	0.20618	0.04799	*
Positive				
	Estimate	SE	p	
(Intercept)	-14.7828	3.0152	9.45E-07	***
AreaTree2	-6.8259	2.5574	0.00761	**
AreaTree3	3.6722	2.1267	0.08422	.
AreaTree4	0.4577	1.3733	0.73889	
Year2006	-5.3699	2.3718	0.02357	*
Year2007	-3.7284	2.2705	0.10057	
Year2008	-7.7604	4.2945	0.07075	.
Year2009	-0.9789	2.8347	0.72985	
Year2010	-1.2914	2.2251	0.56165	
Year2011	-4.2212	4.4577	0.34366	
Year2012	-5.4257	2.8354	0.05567	.
Year2013	-0.2084	2.3455	0.92921	
Year2014	-0.757	2.1562	0.72553	
Year2015	-9.544	5.2447	0.0688	.
Year2016	-5.8151	3.6009	0.10633	
Year2017	-3.048	2.7832	0.27346	
poly(SSH,2)1	51.5717	34.3238	0.13297	
poly(SSH,2)2	38.4424	24.5566	0.11747	
Moon2Light	0.8719	0.7363	0.23639	
Cat_HBFShallow	-2.6216	1.7099	0.12524	
Quarter2	-1.2905	1.2653	0.30778	
Quarter3	-2.1562	1.3929	0.12163	
Quarter4	1.2206	1.2115	0.31372	
Soak_Time	0.8739	0.3319	0.00846	**
poly(SST,2)1	-75.8664	42.868	0.07677	.
poly(SST,2)2	2.5716	30.8813	0.93363	

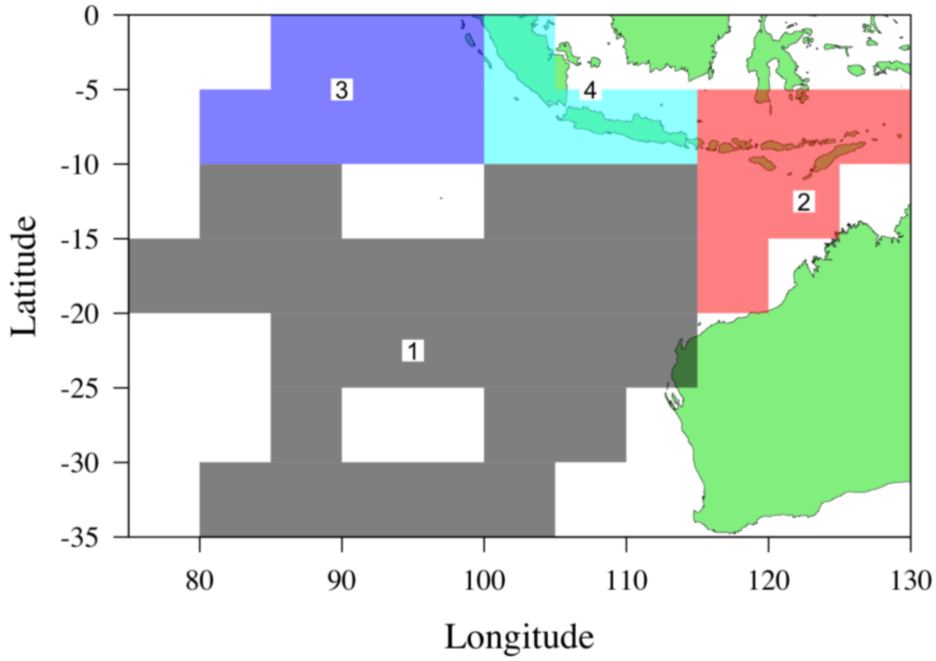

Figure 1. Area stratification used in the analysis based on GLM tree algorithm.

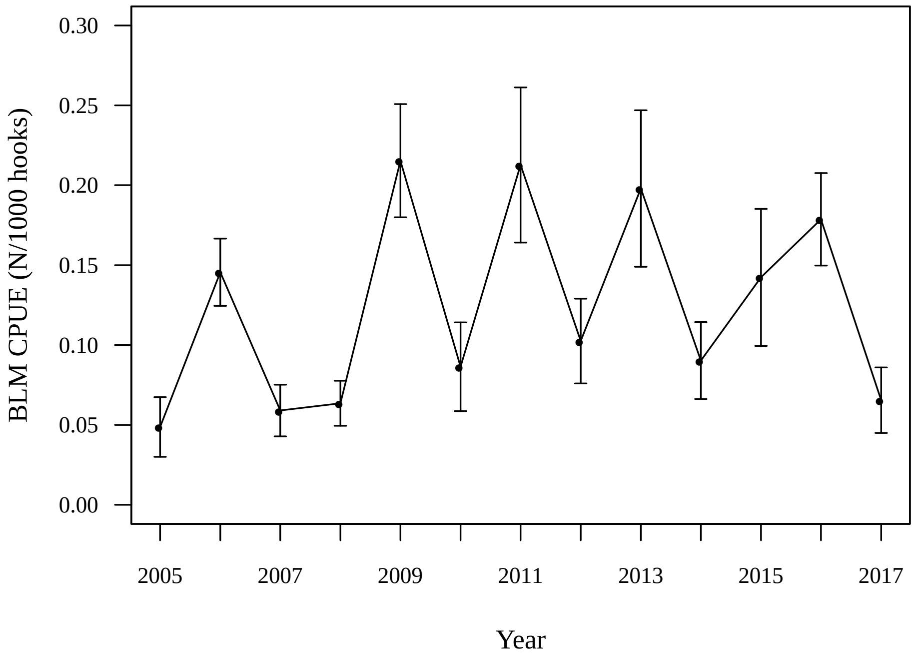

Figure 2. Nominal CPUE series (N/1000 hooks) for BLM from 2005 to 2017. The error bars refer to the standard errors.

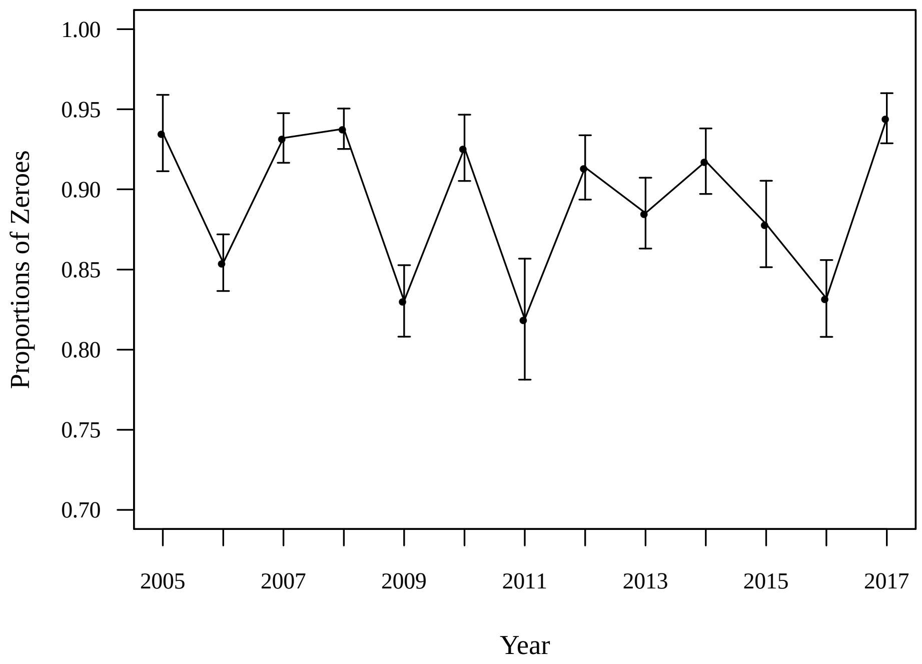

Figure 3. Proportion of zero BLM catches from 2005 to 2017. The error bars refer to the standard errors.

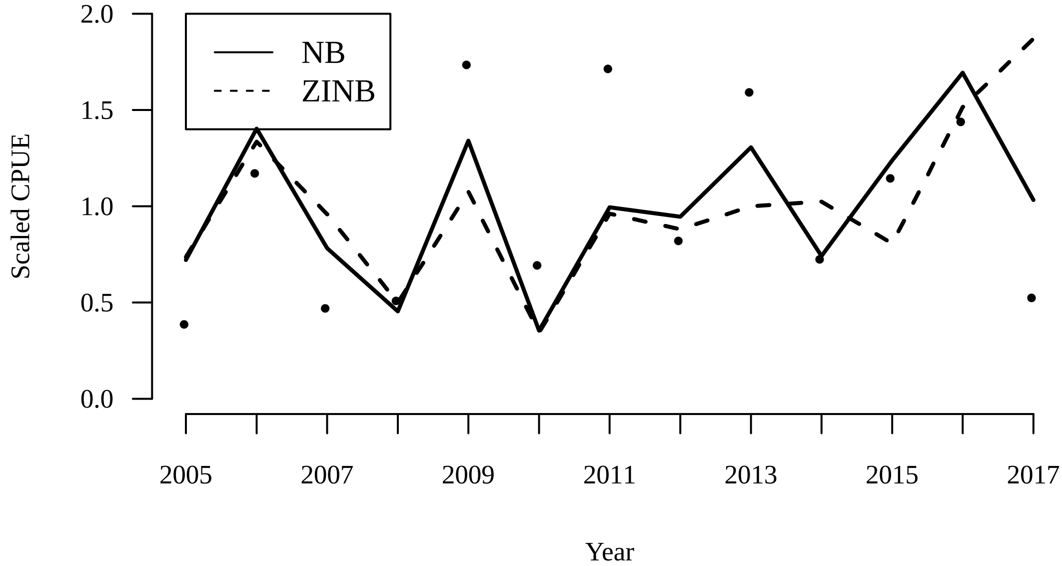

Figure 4. Standardize catch per unit effort (CPUE) calculated using Negative Binomial (NB) and Zero Inflated Negative Binomial (ZINB) models. Values were scaled by dividing them by their means.