Australian Government
Department of Agriculture
and Water Resources

An updated CPUE Index based on a GAMM

F. Helidoniotis

Research by the Australian Bureau of Agricultural and Resource Economics and Sciences

August 2016

Working Paper CCSBT-ESC/1609/12 prepared for the CCSBT Extended Scientific Committee for the 21st Meeting to the Scientific Committee

5–10 September 2016, Kaohsiung, Taiwan

© Commonwealth of Australia 2016

Ownership of intellectual property rights

Unless otherwise noted, copyright (and any other intellectual property rights, if any) in this publication is owned by the Commonwealth of Australia (referred to as the Commonwealth).

Creative Commons licence

All material in this publication is licensed under a Creative Commons Attribution 3.0 Australia Licence, save for content supplied by third parties, logos and the Commonwealth Coat of Arms.

Creative Commons Attribution 3.0 Australia Licence is a standard form licence agreement that allows you to copy, distribute, transmit and adapt this publication provided you attribute the work. A summary of the licence terms is available from creativecommons.org/licenses/by/3.0/au/deed.en. The full licence terms are available from creativecommons.org/licenses/by/3.0/au/legalcode.

Cataloguing data

Helidoniotis, F 2016, An updated CPUE index based on a GAMM, ABARES, Canberra, August. CC BY 3.0.

Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES)

Postal address GPO Box 858 Canberra ACT 2601
Switchboard +61 2 6272 2010|
Facsimile +61 2 6272 2001
Email info.abares@agriculture.gov.au
Web agriculture.gov.au/abares/

Inquiries regarding the licence and any use of this document should be sent to: copyright@agriculture.gov.au.

The Australian Government acting through the Department of Agriculture and Water Resources, represented by the Australian Bureau of Agricultural and Resource Economics and Sciences, has exercised due care and skill in preparing and compiling the information and data in this publication. Notwithstanding, the Department of Agriculture and Water Resources, ABARES, its employees and advisers disclaim all liability, including for negligence and for any loss, damage, injury, expense or cost incurred by any person as a result of accessing, using or relying upon information or data in this publication to the maximum extent permitted by law.

Acknowledgements

The author thanks the CPUE working group and Ilona Stobutzki (ABARES) for comments on the paper and Mark Chambers for advice on understanding his previous work. This research is supported by the Department of Agriculture and Water Resources, through the Fisheries Resources Research Fund and ABARES.

Contents

Summary

Catch per unit effort (CPUE) for southern blue fin tuna (SBT) was estimated using a generalised additive mixed model (GAMM) for SBT, ages 4 years and above. Results show that the CPUE index has been steadily increasing since 2008. The most recent index (2015) continued along this trend with the CPUE index being similar to early 1980 levels.

The age frequency results indicate differences in age frequencies between years. Some years have a higher proportion of older/larger animal and other years (i.e the more recent years) have fewer larger animals. The CPUE index however is numbers based, not weight based, and therefore relying solely on inter-annual trends in CPUE may mislead comparisons because the CPUE, does not take into account the size structure of the fished stock. In recent years there was an increase in numbers in the 4-10 age groups with the 0-3 age groups and the 11+ age groups (the spawning stock) being lower than in previous years (2006-2009).

The proportion of zero catches in the data was approximately 37%. There was a notable increase in the proportion of zero catches between the years 1989 to 1992. This change in zero catches suggest that the possibility that the time series be broken into two series: one prior to 1989 and another after 1991.

Introduction

A generalised additive mixed model (GAMM) was used to estimate a catch per unit effort (CPUE) index for southern blue fin tuna (SBT, ages 4+), based on a previously developed method (Chambers 2014). The CPUE working group recommended the GAMM index had value as an additional monitoring series and so it is a requirement of the Commission for the Conservation of Southern Bluefin Tuna (CCSBT) Data Exchange. Other methods for estimating CPUE include: the nominal CPUE index (Patterson and Stobutzki 2015) and weighted averages from the proxy geostatistical and the proxy B-ratio series (Itoh, 2009). This paper presents the method and results for the GAMMs method for estimating CPUE monitoring index updated with the most recent data, to the year 2015, for discussion and input for the CPUE working group and Extended Scientific Committee.

1 Methods

Data

The file containing the data was made available by the secretariat of the CCSBT, located in the file 'CPUEInputs_6515.txt' within the 'CPUEInputs_6515.mdb' database.

The data used consists of observations from the Japanese longline commercial fishery which has been adjusted and reported as raised catch at age in the form of age frequency.

Data were aggregated to gear, month and 5x5 degree cell resolution. Catch and effort is reported as number of fish caught per 1000 hooks. The data are collected on a spatial scale that is aggregated to 5˚ increments of latitude and longitude: one record represents a 5*5 degree lat/long cell for a given month and a given year. The Japanese longline fleet targets bluefin tuna aged between 0 - 20 years. The data of the 4+ years age group is used in the standardisation. More information on how the Japanese longline data is adjusted is provided in attachment 10 of the Report of the Eighteenth Meeting of the Scientific Committee (p6), but in summary, data collected south of 50 ˚S are deleted, area 5 and 6 are combined into one area (Area 56) and CPUE values > 120 are deleted.

The data are subsetted to include observations limited to the Japan's longline adjusted data (JAP_ADJ) records. In addition, observations with latitudes between 25 ˚S and 55 ˚S and months between March and October are selected to encompass the core months (4-9: April - September) and core areas. Observations before 1969 are excluded due to reduced reliability of these early data. In total, 20 278 records are used in the GAMMs fitting.

It should be noted that in this case the modelled data include catch and effort from vessels not considered part of the core fleet of Japanese longline vessels (Itoh et al. 2013).

A simple transformation of the longitude value, which ranged between -180° and 75°, is required for continuity about the international dateline. This is achieved by converting to a 0-360° longitude range.

Model

As distinct from other models fitted to catch and effort data, longitude, latitude and month are continuous variables within the smoothing term.

The response variable is the natural log transformed CPUE of 4+ SBT defined as:

$$\log .CPUE_{lo,la,m,y} = \log\left(\frac{SBT_{lo,la,m,y}}{Hooks_{lo,la,m,y}} \times 1000 + 0.2 \right)$$

Equation 1

A constant of 0.2 was selected because 0.2 is approximately 10% of the mean of the nominal CPUE (from Campbell et al 1996 cited in CCSBT-ESC/0809/09 p. 38)

Variables

Table 1. The variables used in the CPUE data.

Variable	Type	Description
$\log .CPUE_{lo,la,m,y}$	Numeric continuous response variable	The catch per unit effort. Defined in Equ 1.
$Cell_{lo,la,m,y}$	Categorical	Spatial unit (5° longitude (lo) and 5° latitude (la)) for each record of catch and effort.
$Cell_{lo,la}$	Categorical	Spatial unit (5° longitude (lo) and 5° latitude (la)) that is invariant of time.
$LONG_{lo}$	Continuous covariate	Longitude (in 5° increments)
LAT_{la}	Continuous covariate	Latitude (in 5° increments)
$MONTH_{m}$	Continuous covariate	Month
$Cell_{lo,la,m}$	Categorical	Spatial unit (5° longitude (lo) and 5° latitude (la)) for a given month = m (all years).
$SBT_{lo,la,m,y}$	Continuous covariate	No. of 4+ SBT captured in $Cell^{lo, la, m, y}$
$HOOKS_{lo,la,m,y}$	Continuous covariate	Total no. of hooks in $Cell^{lo, la, m, y}$
$YEAR_{Y}$	Categorical	Calendar Year

The following model was used:

$$\log .CPUE_{lo,la,m,y} = t(LONG_{lo}, LAT_{la}, MONTH_{m}) + YEAR_{y} + 1|(Cell_{lo,la} : YEAR_{y}^{*}) + e_{lo,la,m,y}$$

Equation 2

The model uses a tensor smoother over latitude, longitude and month to estimate the average spatio-temporal distribution of CPUE over the Japanese longline fishing season. The tensor 'te' smoother is part of the 'mgcv' package (Wood, 2011) in R (R Core Team 2016). The categorical YEAR fixed effect provides estimates of how the average annual CPUE varies between years. Differences in the spatial distribution of CPUE between years are handled by a random interaction effect between 5*5 degree ($Cell_{lo,la}$) and Year. The residual error structure, $e_{lo,la,m,y}$ between the observed values and those predicted by the model is assumed to be normally distributed.

Variation in the spatio-temporal distribution of CPUE between years is modelled with the random interaction effect between categorical variables $YEAR_{y}$ and $Cell_{lo,la}$

$$1|(Cell_{lo,la} : YEAR_{y}) \sim N(0, \sigma_{v}^{2})$$

The observations are weighted to address non-constant variance. Observation weighting of $w_{lo,la,m,y}$, are inversely related to their respective variance:

$$e_{lo,la,m,y} \sim N\left(0, \frac{\sigma^2}{w_{lo,la,m,y}}\right) \qquad \textbf{Equation 3}$$

The outcomes from the weighted GAMM model is used to predict the CPUE index over areas 4-9 and months April-September consistent with the Laslett Core area. This definition of month and area is also consistent with the nominal CPUE (Patterson and Stobutzki 2015). In this way the GAMM model with revised observation weights was used to predict log CPUE in the Laslett Core Area (Laslett, 2001) for the population of SBT aged four years and above, harvested by the Japanese longline fleet.

2 Results and Discussion

CPUE indices

In addition to the GAMM CPUE index, three additional indices, provided by other scientific organisations, are presented: a nominal CPUE index, and two weighted averages (w0.5, and w0.8) from the proxy geostatistical and the proxy B-ratio series (Itoh, 2009). The four indices are fairly consistent in their trends (Figure 1). The GAMM index and the nominal CPUE index do not differ markedly from the w0.5 and w0.8 weighted CPUE indices. The consistency in results may not be surprising. Firstly the main difference between the GAMMs model and the nominal CPUE is that the GAMM includes a smoothing term ($LONG_{lo}$, LAT_{la}, $MONTH_{m}$) and an interaction term $1|(\mathrm{Cell}_{lo,la}{:}\mathrm{YEAR}_{y})$. However the similarity between the nominal and GAMMs is that they both include only 'Year' as the main covariate. It appears that both the smoothing term and the interaction term have a slight effect on the CPUE index, given the data that was used in the analyses.

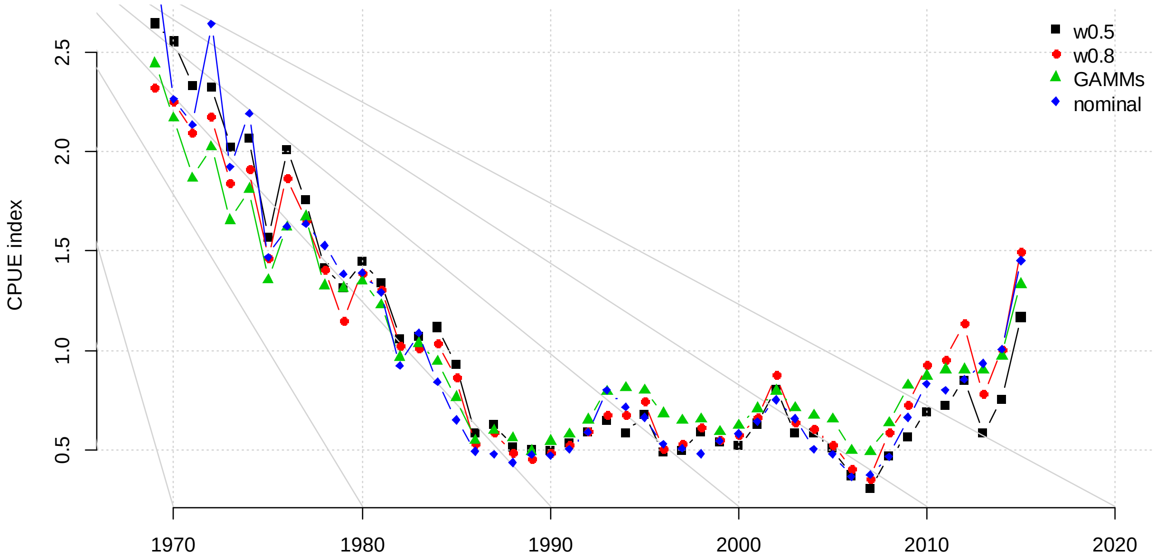

Figure 1. CPUE indices 1969-2013 for ages 4+ in areas 4-9 and months April – September. Ages 0 – 3 years were removed but the number of total hooks remain unchanged. Three other methods for estimating CPUE (References) are also presented for comparison with the GAMM.

Although the 0 – 3 year olds (y.o.) catches were removed from the data, the number of total hooks that were occupied by that age group are not removed. By removing the 0-3 years age group and not removing the associated number of hooks implies that these hooks had no fish, which is not the case, and the inclusion of these artificial zeros may artificially inflate the number of hooks and thereby falsely reduce the CPUE estimate if they are sufficiently abundant in the data set. However the effect of including the 0-3 yo age group hooks on the CPUE was examined and found to be negligible (Appendix)

Frequency of age classes

The age frequency results (Figure 2) indicate that the 0-3 age group has been gradually decreasing in the data between 2011 and 2015. However the 4+ age group in 2015 has been relatively higher than in previous years.

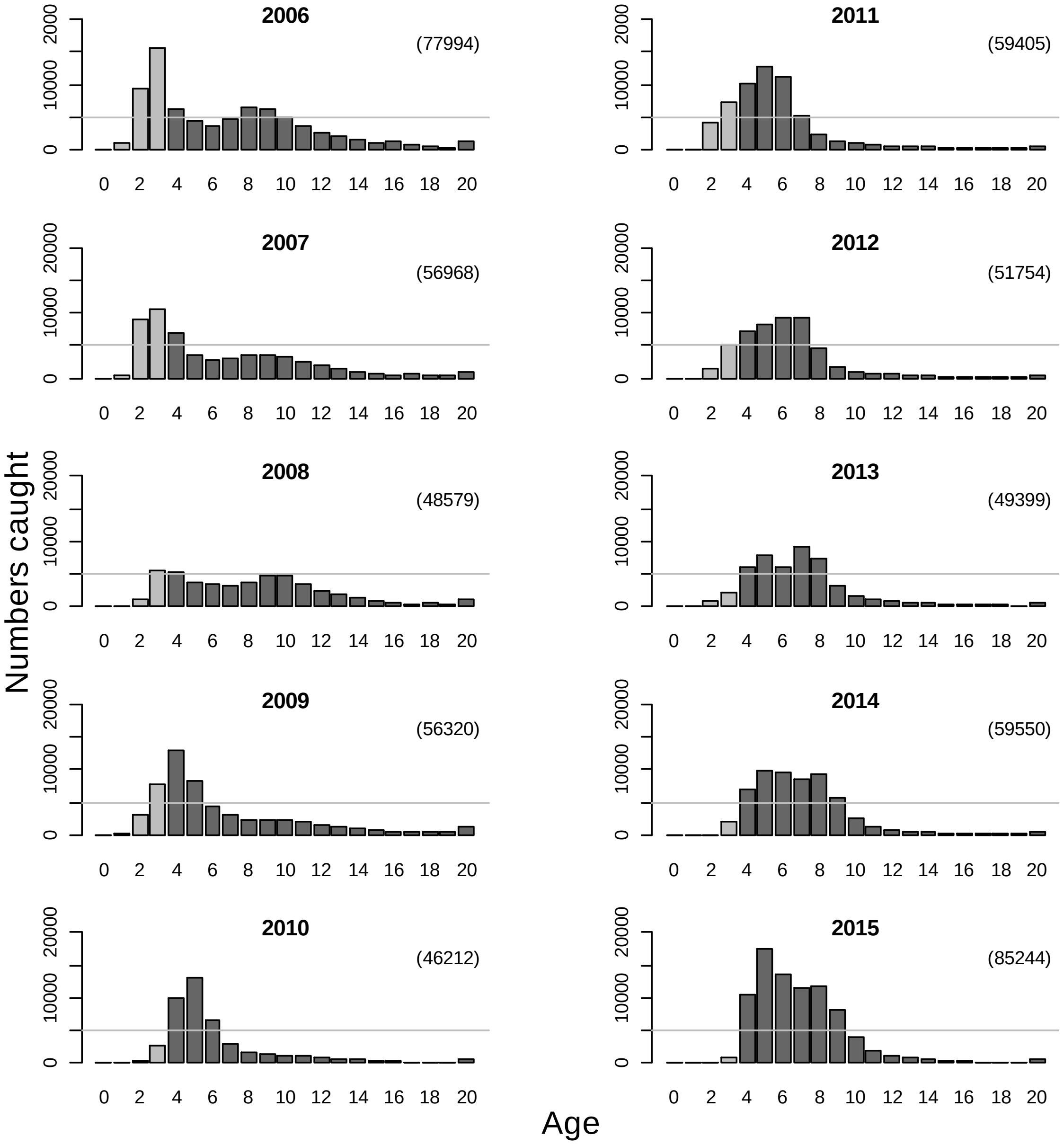

Figure 2. The total number of SBT caught in each age class in the Japanese longline fishery between the months of March and September and in areas 4–9 (see map Figure 4). The CPUE data only uses the age 4+ age groups and is coloured dark grey to distinguish between the data used to estimate CPUE and the 0–3 age groups which are not used in the estimate of CPUE. The grey horizontal line at numbers caught = 5000, serves as reference line to aid comparisons between plots.

Implication to biomass

Outside of a formal model based assessment, CPUE indices are routinely used as a proxy for biomass.The CPUE index for the Japanese longline fishery is numbers based not weight based and therefore the CPUE may not be useful indicator of relative biomass through time in terms of weight. Despite the high CPUE in 2015, a high CPUE that consist of smaller size animals is likely to represent a lower biomass (in term of weight) than the same CPUE consisting of larger or older size animals. Therefore the CPUE index does not take into account the implication to biomass as a result of the frequency of age or size classes. Since 2010 the results also indicate a lower proportion older size animals compared to 2006 – 2009. The low proportion of larger size animals (relating to the 12+ age group) may be indicative of a relatively lower level of spawning biomass. The mismatch in numbers versus weight is taken into account in the Management Procedure 3 (The Bali Procedure, in CCSBT-OMMP/1006/4 Hillary *et. al.* 2012*)* which recommends the global TAC.

Zero catches

Given the data consist of a notable proportion of zero catches (approximately37%), the interannual variation in zero catches was further explored. A distinct increase in the number of zero catches occurs between the years 1989 to 1992 (Figure 3). This change may reflect a change in monitoring to the Real Time Monitoring Program. Given the difference in proportion of zero catches the question arises whether or not the CPUE should be broken up into two time series, pre-1989 and post-1991, and possibly omit 1990 as this may have been the transition year. The properties of these data suggest that the possibility that the time series was broken into two in about 1990, perhaps by the change in monitoring, should be considered.

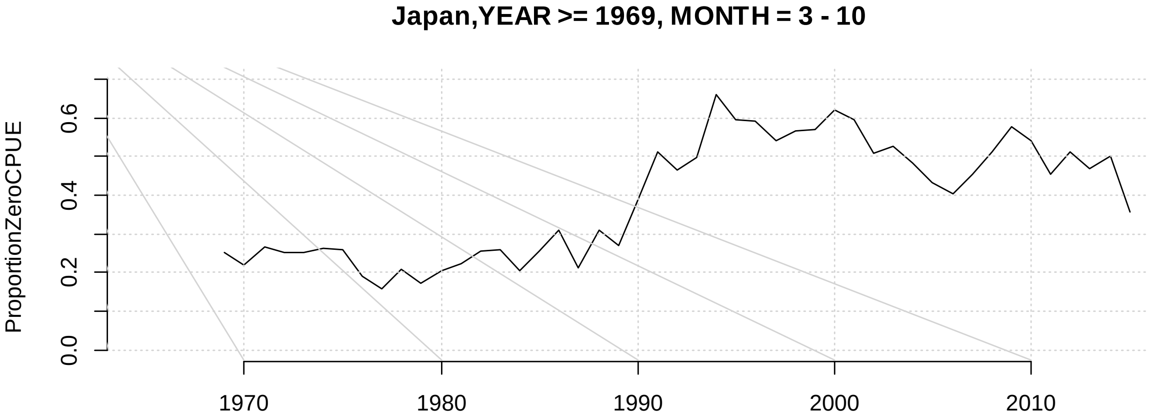

Figure 3. The proportion of zero CPUE due to zero catches (zero number of hooks were already removed). The data is for the Japanese longline fishery between the months of March and September and in areas 4–9 (see map Figure 4) for the 4+ age (i.e 4–20 y.o) group.

Conclusion

A GAMM CPUE index was produced for SBT, ages 4 year and above. The results show that the CPUE index has been steadily increasing since 2008. The most recent index (2015) continued along this trend with the CPUE index being similar to early 1980 levels.

The most recent analysis suggested a few points for consideration. The TAC is weight based whereas the CPUE is numbers based. Despite the high CPUE in 2015, a high CPUE that consist of smaller size animals is likely to represent a lower available biomass (in term of weight) than the same CPUE consisting of larger size individuals or older age groups. Therefore only considering changes in CPUE through time does not take into account the implication to biomass as result of having a size composition that consist of mainly smaller sizes and ages composition of the catch.

The increase in zero catches from 1989 onward might suggest the need for two CPUE time series: one prior to 1989 and another after 1991.

References

Chambers, M 2014, A CPUE Index based on a GAMM: a proposed monitoring series. ABARES Commonwealth Department of Agriculture and water Resources Canberra, Australia.

Hillary Rich, Preece Ann, Davies Campbell 2012. Developing a management procedure based recovery plan for southern bluefin tuna FRDC 2011/034 January 2012

Itoh T, Lawrence E & Pope JG 2009, 'The development of new agreed CPUE series for future MP work', CCSBT-ESC/0809/09

Itoh T, Sakai O & Takahashi N 2013, 'Description of CPUE calculation from the core vessel data for southern bluefin tuna 2013', CCSBT-ESC/1309/29.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Laslett, GM 2001, 'Exploratory analysis of the SBT CPUE data using smoothing splines', CCSBT-SC/0103/06.

Patterson H & Stobutzki I 2015, 'Fisheries Indicators for the southern bluefin tuna stock 014-15', CCSBT-ESC/1509/11

Wood, S.N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalised linear models. Journal of the Royal Statistical Society (B) 73(1):3-36

Zuur AF 2012, *A Beginner's Guide to Generalized Additive Models with R.* Highland Statistics Ltd Newburgh, United Kingdom.

Zuur AF, Savaliev A & Ieno EN 2014, *A Beginner's Guide to Generalized Additive Models with R.* Highland Statistics Ltd Newburgh, United Kingdom.

Appendix A - Additional information

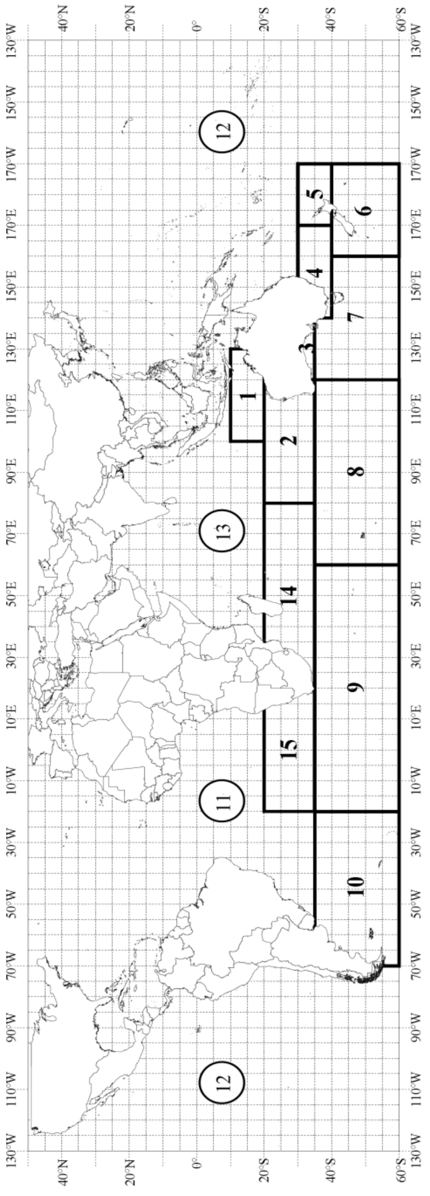

Figure 4 Map of CCSBT statistical areas from https://www.ccsbt.org/sites/ccsbt.org/files/userfiles/file/docs_english/operational_resolutions/Resolution_CDS.pdf

Data exploration

The presence of potential outliers are presented in Figure 5. There are a few data points with relatively large numbers of abundance of tuna greater than 4 years of age (SBT.4p panel) however given the very high total number of record these outlier are unlikely to have a bearing on the results and were therefore not removed. Other variables do not consist of conspicuous outliers.

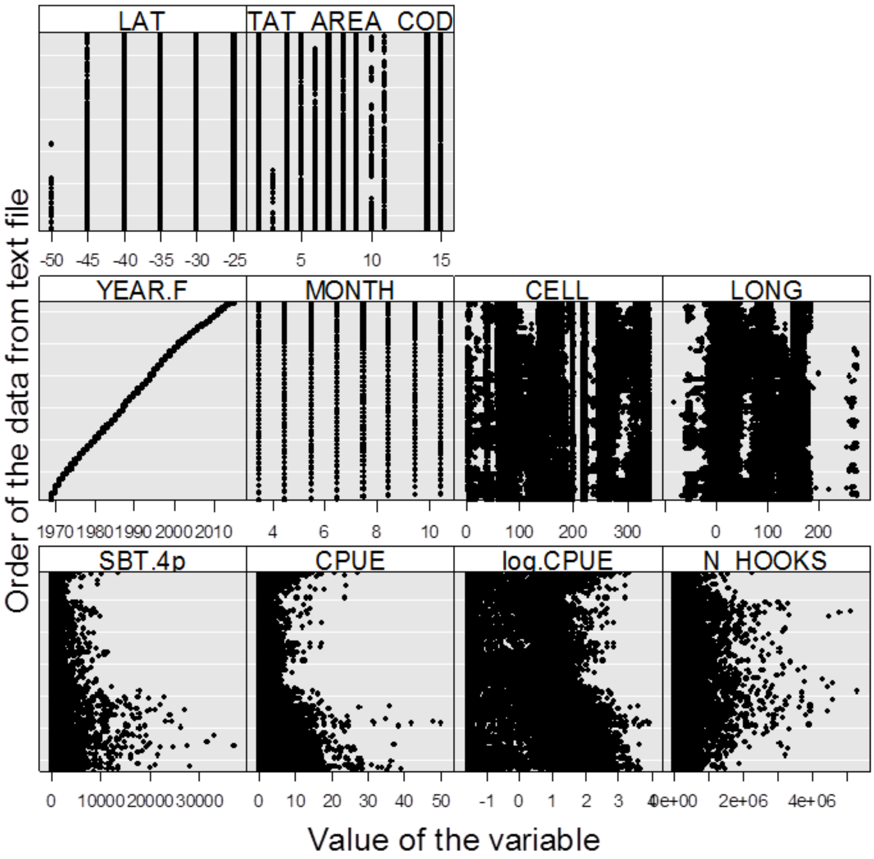

Figure 5. Dotplot of for all variables in the dataset for the Japanese logline fishery between the months of March and September and in areas 4–9 (see map Figure 4) The data includes the age 4+ age (i.e. 4–20 y.o) groups only.

The dataset consist of a notable proportion of zero catches; 37.6% of the data contains zero catches.

Collinearity

Collinearity (the correlation between covariates) should be avoided because any evidence of collinearity increases the standard errors of estimated regression parameters and therefore increases the p-values compared to models where no collinearity exists (Zuur, 2012).

Collinearity is examined using multi-panel scatterplots and Pearson correlation coefficients Figure 6. In Figure 6. The first row and the first column is for the response variable versus the covariates: the other panels are for collinearity. Results indicate that the highest correlation was between the log CPUE (response variable of interest) and latitude (0.7) indicating a strong pattern between these two variables.

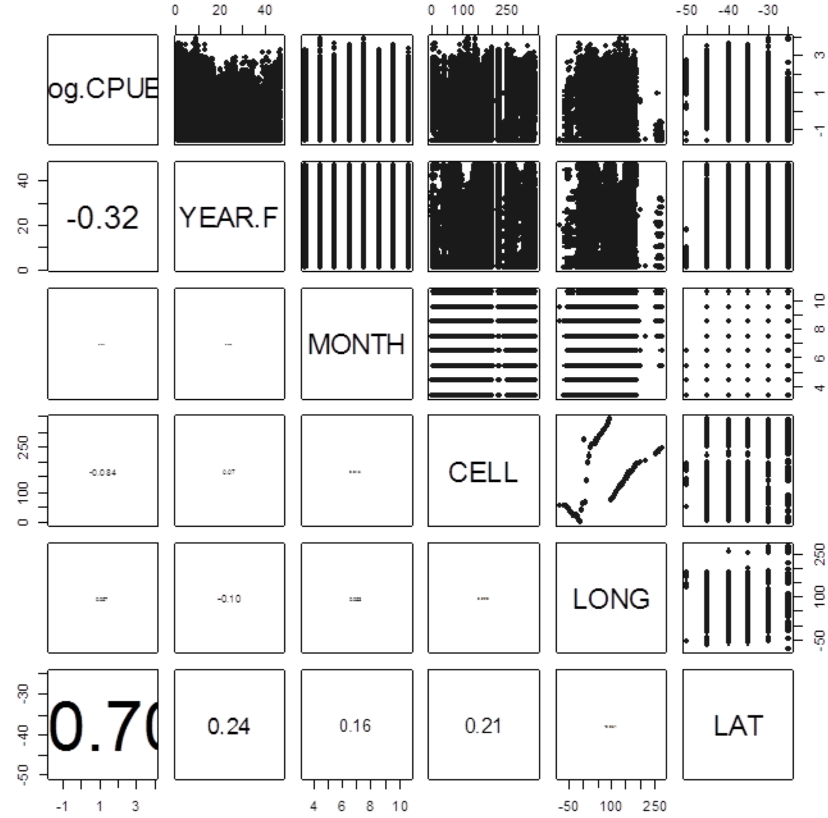

Figure 6. Scatterplot and Pearsons correlation between all variables in the dataset for the Japanese longline fishery between the months of March and September and in areas 4–9 (see map Figure 4) for the 4+ age (i.e. 4–20 y.o) group.

With the exception of the response variable (LogCPUE) and latitude, the pattern between the LogCPUE and other variables is weak. Therefore, a further analysis on the variance inflation factors (VIF) is considered worthwhile in order to identify even the smallest amount of collinearity (Table 2). It is considered that a covariate with a VIF value great than 3 should be removed from further analysis (Zuur et al. 2014). All the main covariates (year, month, cell, latitude and longitude) can remain since they all had a VIF smaller than 3 (Table 2).

Table 2. Variance inflation factors for the main explanatory variables. The variable logCPUE is not included because it is a response variable.

Variable	VIF
YEAR.F	1.070951
MONTH	1.030616
CELL	1.04697
LONG	1.015244
LAT	1.130863

The scatterplot between the response variable and logCPUE and fitted with a LOESS smoother provide additional information about the relationship between the response variable and the main covariates (Figure 7). With the exception of latitude results indicate that there is no strong patter between the response variable and the main covariates.

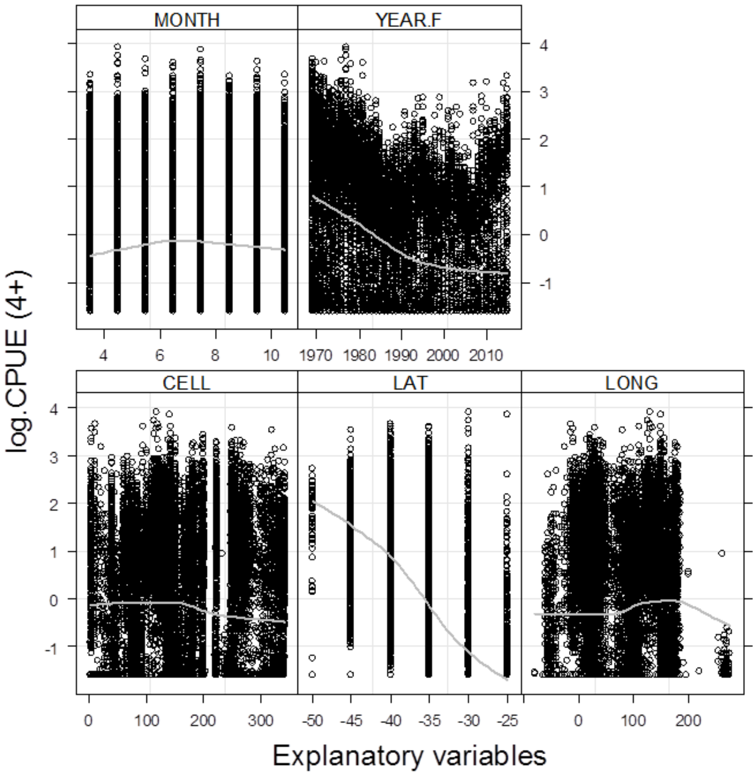

Figure 7. Scatterplots of the response variable (LogCPUE) against the 5 main covariates (month, year, 5*5˚ cell, latitude and longitude). Included is a LOESS smother to aid in visual interpretation. The data is for the Japanese longline fishery between the months of March and September and in areas 4–9 (see map Figure 4) for the 4+ age (i.e. 4–20 y.o) group.

Model validation

The CPUE indices presented in involved a two stage process of the GAMM model of Equation 2 the first stage consisted of the GAMM model being generates and the second stage involved using the outputs of the GAMM to generate a weighting and then rerunning the GAMM model with the weighting applied. The Pearson residuals versus the fitted values are presented in Figure 8 for the unweighted and weighted GAMM. The results show that with the exception of the straight line pattern there is no remaining pattern in the data. At this stage it is not known whether this straight line pattern present a problem. The straight line pattern is most likely due to the zero CPUE value where a constant (of 0.2) is added. Zero catches represent 37.6% of the data.

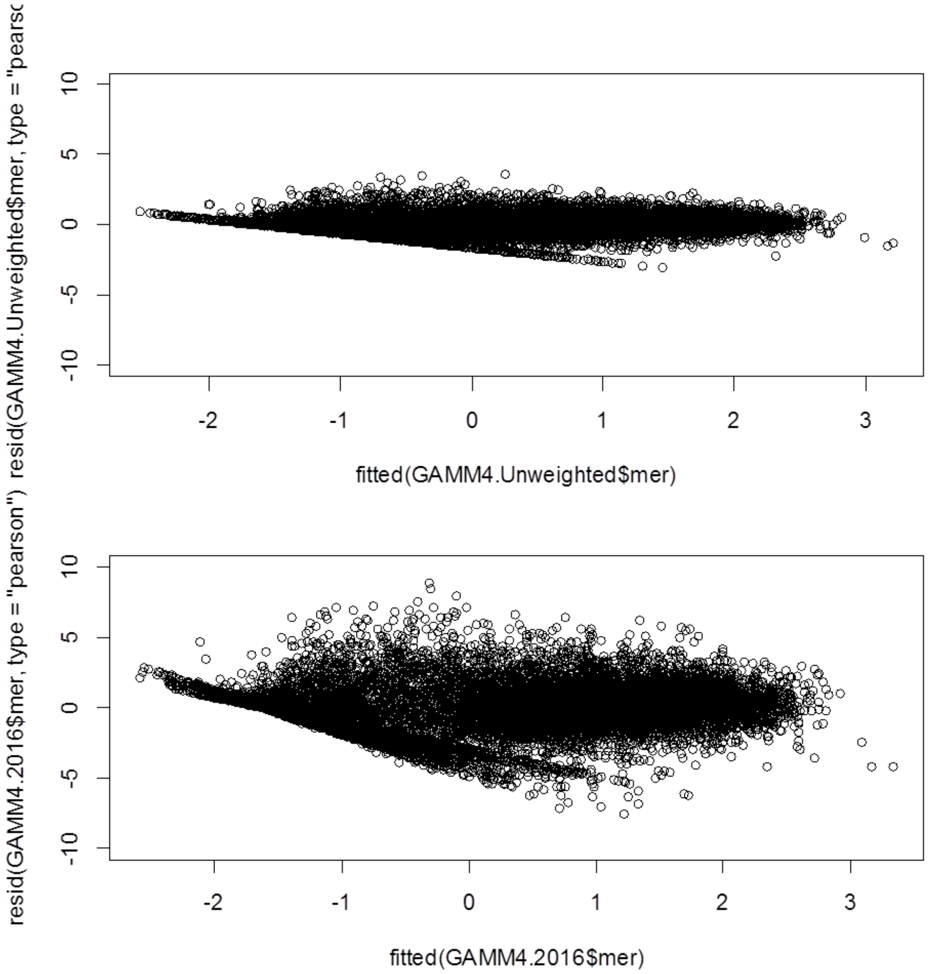

Figure 8. Pearson residuals versus fitted values for the GAMM model in Equation 2. The top panel represents the unweighted model and the lower panel represents the weighted model. The weighted model is used to generate the submitted CPUE indices. The data is for the Japanese longline fishery between the months of March and September and in areas 4–9 (see map Figure 4) for the 4+ age (i.e. 4–20 y.o) group.

Comparison with previous years submitted CPUE index

A comparison of the CPUE index submitted last year in 2015 with the CPUE index submitted this year (2016) is presented in Figure 9. The two indices are very similar when plotted on the same axes. This indicates that the data series did not vary between the two years of data exchange and also the method for calculating the CPUE index remains unchanged from last year.

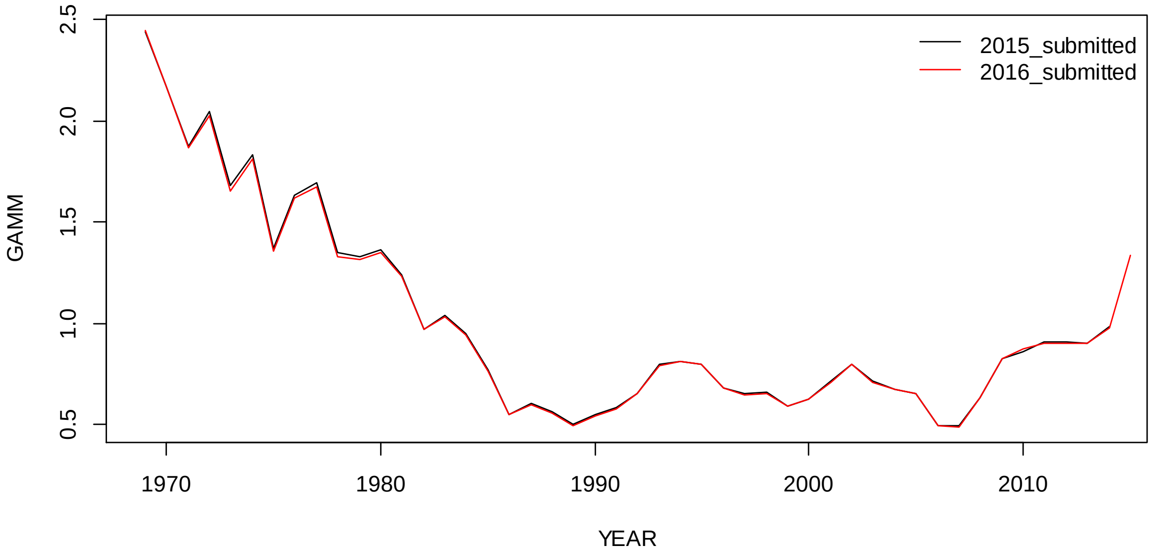

Figure 9. The GAMM submitted index. The index submitted in 2015 and the most recent index submitted in 2016

Implications of the removal of the 0–3 year olds on relative CPUE estimates

Although the 0 – 3 year olds (y.o.) catches were removed from the data, the number of total hooks that were occupied by that age group are not removed. The percentage of 0-3 y.o removed from the catches is not consistent from year to year (grey line Figure 10). By removing the 0-3 years age group and not removing the associated number of hooks implies that these hooks had no fish, which is not the case, and the inclusion of these artificial zeros may falsely reduce the CPUE estimate if they are sufficiently abundant in the data set. A visual comparison between the percentage of 0-3 y.o in the catch and CPUE suggest a relationship: the higher the percentage the lower CPUE estimate (Figure 10). The inclusion of hooks occupied by the 0-3 age group may therefore lead to complications in interpreting relative CPUE estimates for the 4+ age group especially since the proportion of 0-3 y.o in the catch is not consistent from year to year. One way to address the issue of the non removal of hooks on CPUE is to compare the CPUE for the 4+ age group with all age groups included. For simplicity the nominal CPUE was used in this comparison (Figure 11). The nominal CPUE series was comparable to the GAMM CPUE series () and so it is assumed that the effect on hooks on the nominal CPUE can apply to the GAMM CPUE index.

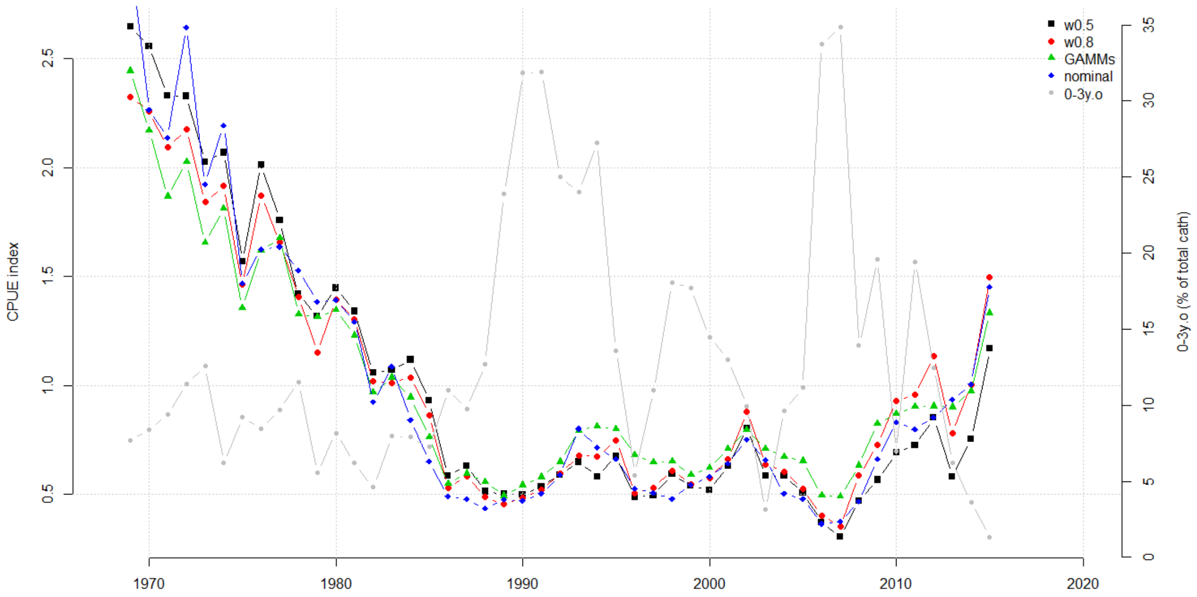

Figure 10. CPUE indices 1969-2013. Similar to Figure 1 but also shown is the 0 - 3 y.o which are omitted from the catches. CPUE indices are for 1969-2013 for ages 4–20 in areas 4–9 and months April – September. Three other methods for estimating CPUE are also presented for comparison with the GAMMs.

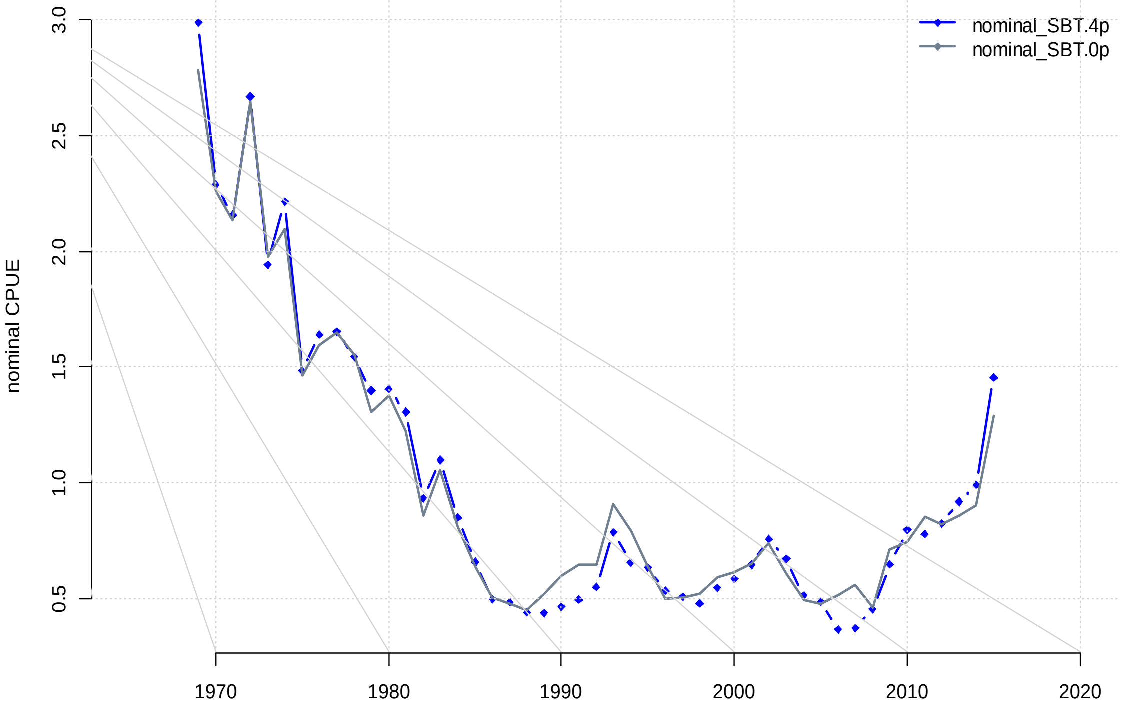

Figure 11.Comparison of Nominal CPUE indices 1969-2013 for two age groups: the 4+age group (nominal_SBT.4p, blue) and all ages (nominal_SBT.0p, grey). The nominal_SBT.4p is

identical the nominal results in Figure 1 and Figure 10. CPUE indices are for 1969-2013 for two different age groups (4-20 and 0-20) in areas 4–9 and months April – September.

Miscellaneous

The running of the GAMM model takes several hours. The analysis was run on a Hewlett Packard with Intel core i7, 3.4 Ghz CPU and 8G of RAM ona 64 bit operating system. After running the GAMM using R software (version 3.2.4) the following warning message was generated

```
Warning message:

In optwrap(optimizer, devfun, getStart(start, rho$lower, rho$pp),  :

 convergence code 1 from bobyqa: bobyqa -- maximum number of function evaluations exceeded.
```