CCSBT-ESC/1509/10

Australian Government
Department of Agriculture
ABARES

Estimates of non-member catch of SBT in the Indian and Pacific Oceans

Mark Chambers and Simon Hoyle[1]

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

August 2015

[1] National Institute of Water and Atmospheric Research, 217 Akersten St, Port Nelson 7010, New Zealand.

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/

Acknowledgements

The analyses described in this report are based on publically available data from the Indian Ocean Tuna Commission (IOTC), provided by its contracting parties (members) and cooperating non-contracting parties, and from the Western and Central Pacific Fisheries Commission (WCPFC), provided by its members, participating territories and cooperating non-members. The analyses also use the CCSBT data provided by members and cooperating non-members. The CCSBT Secretariat provided assistance in accessing data. This work was funded by the Fisheries Resources Research Fund and Australian Bureau of Agricultural and Resource Economics and Sciences (Australia) and the Ministry for Primary Industries (New Zealand).

Contents

Summary

The 2014 CCSBT Extended Scientific Committee has expressed the need to obtain estimates of unaccounted mortality of southern bluefin tuna resulting from potentially unreported catch by non-member fleets. This recommendation was reiterated at the later Extended Commission meeting. This paper describes analysis that builds on previous work that described overlap of longline effort in the Indian Ocean recorded in IOTC databases with regions where SBT is harvested, to provide estimates of catch that could be used in sensitivity analyses. We use the random forests machine learning algorithm to fit prediction models for catch of SBT in the Indian and Pacific Oceans. The models use characteristics of the effort and the catch per unit effort (CPUE) of other tuna and billfish species assumed to be reliably recorded for all fleets. The model is fitted to aggregated CPUE data from CCSBT member nations, and used to predict catch of SBT by non-member nations for the years 2007-2013, based on their recorded effort and the catch rates of IOTC or WCPFC target species. Random forest methods are also used in a classification guise to assign selectivity of the non-member SBT catch, to enable inference of the age classes of SBT captured by the non-member fleets as necessary for incorporation in stock assessment scenarios. The estimates should be considered indicative rather than authoritative, but may be appropriate for analysis of the sensitivity of the operating model to plausible non-member catch.

1 Introduction

The Extended Commission for the Conservation of Southern Bluefin Tuna (CCSBT) requested that the Extended Scientific Committee (ESC) conduct sensitivity analyses around all sources of unaccounted mortality of southern bluefin tuna (SBT) as part of the 2014 stock assessment. The fifth meeting of the Operating Model and Management Procedure Working Group (OMMP5) discussed the request from the Extended Commission and noted that the working group was not necessarily in possession of the information required to construct the full range of plausible scenarios for unaccounted mortalities. The OMMP5 recommended some scenarios to be considered in the 2014 stock assessment.

The 2014 ESC considered the impacts on the stock assessment and projections from unaccounted mortality scenarios (Attachment 8, ESC19 2014). The ESC19 noted that if the total mortalities are as large as those considered in the 'added-catch' scenario, then the impacts on the rebuilding plan may be substantial. The ESC19 also noted that this scenario was potentially plausible given the available data, information and anecdotal market reports. The ESC19 requested that the Extended Commission and Compliance Committee urgently provide detailed information and data to properly assess impacts of unaccounted mortalities.

The 2015 meeting of the Extended Commission directed the ESC, Compliance Committee and members and cooperating non-members undertake analyses to provide estimates of non-member catch. The Extended Commission also agreed that the 2015 ESC work schedule included the collation of information on unreported mortalities and categorising this information in accordance with OM "fleets".

As discussed by the ESC (ESC19 2014), one potential source of unaccounted mortality not currently considered in the operating model is unreported catch of SBT by countries that are not members of CCSBT. Larcombe (2014) provided an initial attempt at examining the overlap of non-Member fleets in areas identified as peak SBT areas within the area of competence of the Indian Ocean Tuna Commission (IOTC) during peak SBT seasons. This analysis identified increasing fishing effort in these peak areas and times that may indicate catch of SBT as a bycatch/byproduct for fishing targeting other tuna and billfish species.

This paper and Hoyle & Chambers (2015) attempt to estimate levels of potential unreported catch of SBT by using information on catch rates of SBT assumed to be reliable in order to predict possible catch rates of non-members in the Pacific and Indian Ocean given details of effort. Co-operating non-members of CCSBT are expected to report any SBT catch by their vessels to CCSBT.

Catch rates of SBT by longline might be expected to depend on factors such as the location of fishing effort, seasonal effects, but also on the size of the exploitable population of SBT (relative abundance) and aspects of the effort related to species targeting which are not available to the analysis. At the same time, the relative catch rates of various species can be expected to reflect targeting behaviour taking into account recorded details of effort. It follows that catch rates of an appropriate set of species, assumed to be recorded accurately, might provide information on potential catch rates of SBT. However, interactions between the effects of catch rates of various tuna and billfish species and effort variables on the catch of SBT are likely to be highly complex. Statistical techniques commonly fitted to catch and effort data to standardize catch rates of target species, such as GLMs and GAMs have a limited capacity to handle complex interactions and correlations between the effects of predictor variables. By contrast, certain machine learning procedures such as those based on classification and regression trees can approximately model complex interactions among variables.

We use the machine learning procedure 'random forests' to model catch rates of SBT in the Indian and Pacific Oceans. The fitted models are used to predict catch rates of SBT by non-member fleets based on effort data submitted to the IOTC and WCPFC. The reliability of estimates provided depend crucially on the quality and completeness of the data upon which the estimates are based.

2 Method

Basic approach

The CCSBT database includes only information reported by CCSBT members and cooperating non-members. Furthermore, not all longline catch of tuna and billfish by CCSBT members and co-operating non-members is included on the CCSBT database. In this case, the approach used requires auxiliary data to be sourced in order to estimate possible non-member catch. Throughout the remainder of this paper we use the term non-member to mean fleets that are not members of the CCSBT. Data describing the effort of non-member fleets were sourced from the IOTC and WCPFC datasets.

We use the random forests machine learning algorithm in R to train a prediction model for SBT catch rates on explanatory variables that are available for both CCSBT member fleets and non-member longline fleets that fish in either the Indian or Pacific Oceans. Random forests models are fitted to monthly five-degree square observations for which SBT catch rates are assumed known (longline fleets from CCSBT member countries) and used to predict SBT catch rates associated with the effort of the non-member fleets. Predicted catch of SBT associated with each non-member observation is then calculated as the product of the predicted catch rate and number of hooks set.

In addition, random forests classification models were used to assign a selectivity to each non-member observation from the Indian Ocean. This was done by assigning the appropriate longline 'fishery' (LL1, LL2, LL3 or LL4) to each CCSBT monthly 5-degree square observation as designated in the CCSBT data document. The fishery variable was then used as the response variable for random forests classification models fitted to the CCSBT member observations and used to predict the fishery class of each non-member observation.

Random forests

Regression trees are a non-parametric modelling approach that has considerable flexibility for handling interactions. The fact that regression trees do not make parametric assumptions about the distribution of observed bycatch around their expected levels is also convenient when estimating bycatch. A disadvantage of individual regression trees is that they have been found to generally perform less well in prediction than many alternative modelling approaches. Methods based on `ensembles of regression trees', such as random forests (Breiman 2001), retain the flexibility of regression trees, but tend to give predictions with lower mean squared error (Hastie et al. 2009).

Random forests and other approaches based on ensembles of regression trees are commonly included in the larger class of machine learning (or data mining) procedures. Machine learning approaches have been found to be useful for estimating bycatch because of their flexibility and typically good predictive performance (see e.g. Lennert-Cody and Berk 2007; Pons et al. 2009). Recently Shono (2014) has described advantages of using support vector machines for modelling SBT CPUE. We use random forests as implemented in the R (R Development Core Team 2012) package randomForest (Liaw and Wiener 2002) to predict catch of SBT by non-member fleets between 2007 and 2013. The uncertainty of predictions is assumed to be given by the distribution of estimates provided by the individual regression trees that make up the random forest model. We also use random forests in a classification guise to assign effort to SBT selectivities used in the SBT stock assessment.

Estimating non-member catch

It is well known that the spatial distribution of areas fished by CCSBT fleets targeting SBT has declined in recent years (Anon. 2014; Itoh and Takahashi 2014). It is plausible that there are areas in the Indian and Pacific Oceans where SBT can be caught that are not currently being regularly fished by CCSBT member fleets. Without information on catch rate, predictions of non-member SBT catch in areas not fished by member fleets will be less reliable than areas that are fished. In order to provide some information on possible catch rates in these areas we define a variable, *HISTORY,* equal to the log transform of total historic reported longline catch of SBT in each 5-degree square. So the value of the *HISTORY* variable for any fishing in 5 degree square *i* is given by:

$$HISTORY_i = \log\left(\left\{\sum_{1965}^{2014} \text{Annual reported number of SBT retained in 5-degree square } i\right\} + 1\right)$$

The *HISTORY* variable allows catch rates observed in fished areas to inform predictions of catch rates in unfished areas. Alternative predictor variables could have been derived from historic catch and effort data, but intuitively, the *HISTORY* variable mapped in Figure 1 seems likely to be related to catch rates of SBT.

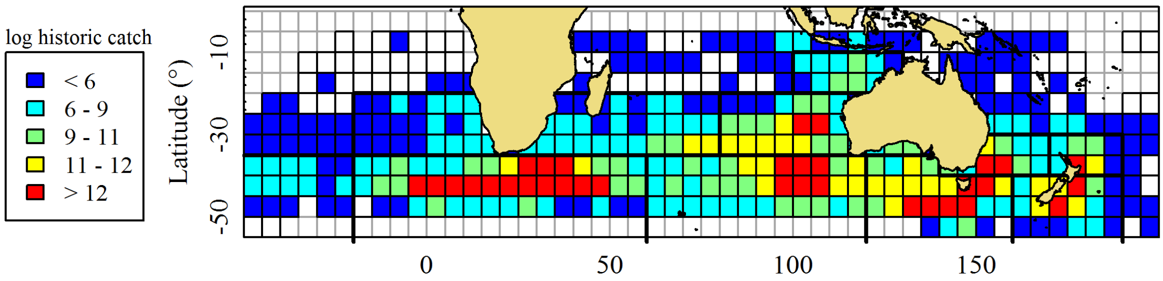

Figure 1 Logarithm of total reported longline catch of SBT (in numbers) 1965-2014.

Source: CCSBT database, AGGREGATED_CATCH_EFFORT table.

Estimating non-member selectivity

In order to assign estimated non-member catch to fisheries it was necessary to assign selectivities to the CCSBT member observations. This was done using the designation of 'fishery' described on page 16 of the 2015 CCSBT data documentation (Millar 2015). The designation of 'fishery' is dependent upon Fleet and the CCSBT Statistical Area fished (see Table 1). It can be seen that selectivity only needs to be modelled in the Indian Ocean. All catch of SBT in the Pacific Ocean can be assumed to have selectivity as defined by the 'LL1' fishery.

Table 1 Designation of Fishery (selectivity schedule) of catch of SBT by CCSBT member longline fleets.

	Designated Fishery	
Fleet	**Indian Ocean**	**Pacific Ocean**
JP	LL1, except in Area 1 (LL4) and Area 2 (LL3)	LL1
TW	LL2	LL1
KR	LL1	LL1

3 The Indian Ocean

Data

The primary data used for the analysis of Indian Ocean catch were sourced from the IOTC website (http://www.iotc.org/data/datasets). These data are freely available. Only longline data were considered. There is no evidence in the IOTC data that SBT is captured by purse seine vessels operating in the Indian Ocean other than that reported by the Australian surface fishery.

Data subsetting

We consider observations starting in 2007 due to known issues with reporting of SBT catch before this time. Although the data are available from the IOTC longline dataset for our purposes we also sourced catch and effort data from the CCSBT database for validation and editing purposes.

The IOTC longline data includes observations in Quadrants 1 (longitude 0° - 180° E, latitude 0° - 90° N) and 2 (longitude 0° - 180° E, latitude 0° - 90° S). We limit consideration to observations in Quadrant 2. SBT are also likely to be captured in the southern Atlantic Ocean, but this area falls under the International Commission for the Conservation of Atlantic Tunas (ICCAT). ICCAT data has not been included. In an effort to reduce the extent of spurious predictions, the region used for prediction was limited further to the region shown in Figure 2.

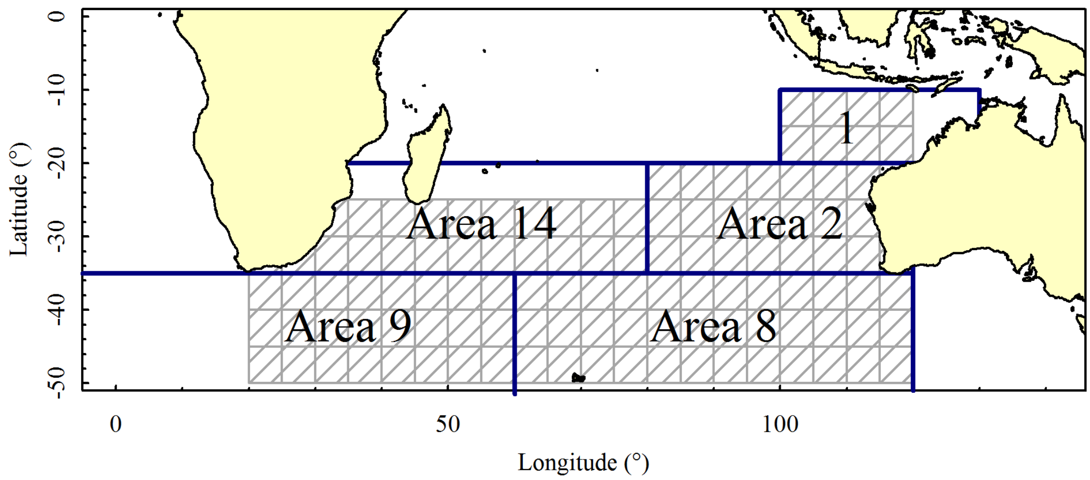

Figure 2 Area used to fit the prediction model and predict non-member catch in the Indian Ocean. Area considered indicated by hatched 5-degree squares.

The formatting of the IOTC longline records differed between fleets. In most cases longline effort was specified by number of hooks set, however, in some cases effort was specified in terms of number of longline sets or number of days fished. For some fleets catches were specified by number of individuals, whereas other fleets specified the total weight of individuals caught. In some cases both catch numbers and catch weights were reported (see Tables 2 and 3).

Generally, catch weights are recorded for the non-member fleets in the IOTC dataset rather than catch numbers (Table 3). On the other hand, the catch of the Japanese fleet is recorded only by number (Table 2). This means catch rates derived from Japanese data in the IOTC database are not directly comparable with the catch rates of most of the non-member fleets. For this reason, catch and effort data from the Japanese fleet are not directly used in the present study to model catch rates of SBT in the Indian Ocean. However, the catch of SBT by the Japanese longline fleet contributes indirectly to the modelling of catch rates via the dominant contribution of the Japanese fleet to the values of the *HISTORY* variable specific to 5-degree squares where catch of SBT has been recorded historically (Figure 1).

Table 2 Format of recorded data of CCSBT member fleets in the IOTC dataset.

Flag[†]	Catch in Numbers	Catch in tonnes	Effort Units
Australia	No	Yes	Hooks
Taiwan	Yes	Yes	Hooks
Japan	Yes	No	Hooks
Korea	Yes	Yes	Hooks

† Spatially explicit catch and effort data are not available for Indonesia.

Table 3 Format of recorded data of considered fleets that are not members of the CCSBT in the IOTC data. CCSBT co-operating non-members are identified by an asterisk.

Flag	Catch in Numbers	Catch in tonnes	Effort Units
China	Yes	Yes	Hooks
Seychelles	No	Yes	Hooks
Spain*	Yes (only SWO)	Yes (only SWO)	Hooks
Mauritius	No	Yes	Hooks
Portugal*	No	Yes	67% Hooks, 33% Days
France-Reunion*	No	Yes	Hooks
Thailand	No	Yes	36% Hooks, 64% Days

Data sourced from the CCSBT database were compared with the IOTC data reported against Taiwanese and Korean fleets. Observations were discarded when the number of hooks recorded in the IOTC database was less than 75 percent or more than 125 percent of the number recorded in the CCSBT database (see Fig. 3). Where the effort was consistent, but CPUE of SBT differed between the IOTC and CCSBT records, the CCSBT value was assumed. We suspect that unreported longline sets, evidenced by the poor correspondence between IOTC and CCSBT reported effort (Figure 3), are unlikely to be 'missing at random'. In particular, observed catch rates might of SBT might influence the probability of observations being reported to the CCSBT. Catch rates derived from these data may give misleading signals about abundance trends.

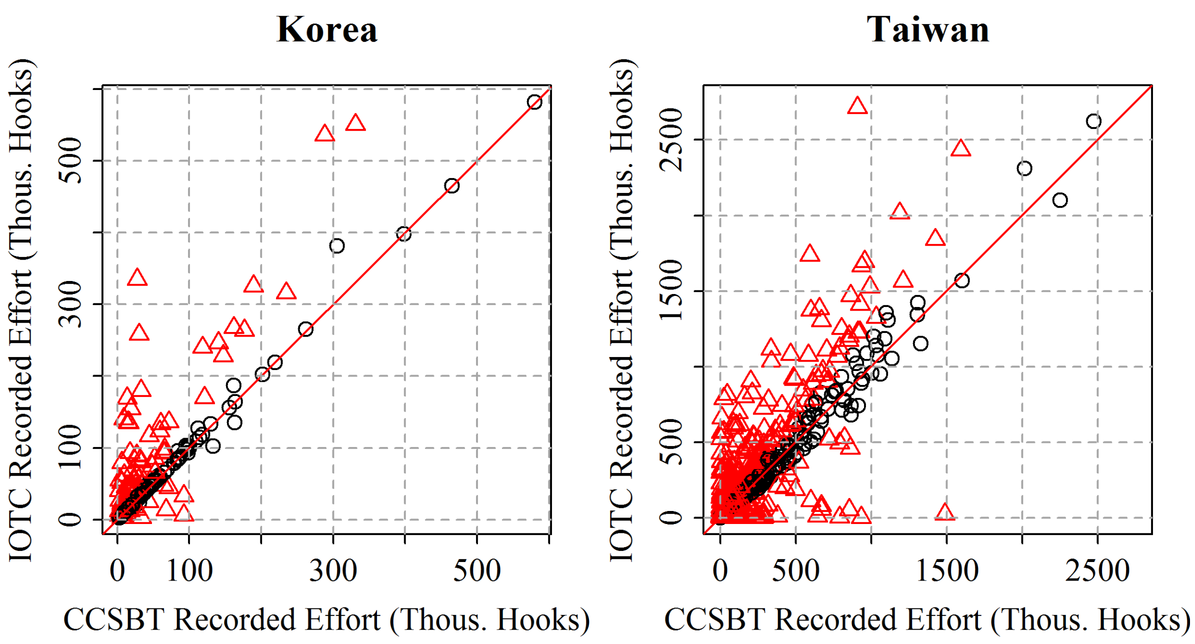

Figure 3 Correspondence of monthly sums of hooks set aggregated at the 5-degree square level according to IOTC and CCSBT data. Retained observations plotted as open circles, excluded observations plotted as open triangles.

Fitting models to CCSBT member data

CPUE models

The main consideration for model selection when fitting a random forests model is compiling a set of data with informative explanatory variables that are available for the training set and the prediction set.

For the Indian Ocean analysis, the base model fitted to the CCSBT member data is defined as:

$$SBT_{CPUE} \sim LONG + LAT + MONTH + YEAR + SWO_{CPUE} + BET_{CPUE} + ALB_{CPUE}$$

$$+ YFT_{CPUE} + YEAR + HISTORY$$

All of the explanatory variables are treated as continuous variables. The CPUEs of SBT, swordfish (SWO), bigeye (BET), albacore (ALB) and yellowfin (YFT) are specified in terms of tonnes per thousand hooks.

Catch for the Spanish fleet is recorded for swordfish only in the IOTC data obtained. We assume that the Spanish fleet catches the other target species included in the base CPUE model incidentally from time-to-time and these are not reported. It is not appropriate to use the base model to predict catch of SBT by the Spanish fleet in this case so we fit a simpler model to CCSBT member data to enable the prediction of SBT CPUE by the Spanish fleet. The simpler model is defined as:

$$SBT_{CPUE} \sim LONG + LAT + MONTH + YEAR + SWO_{CPUE} + YEAR + HISTORY$$

The variable importance measure reflects the predictive performance of models with each variable compared with models where the values of the variable in question are randomly permuted (Hastie et al. 2009). The estimated importance measures of the variables in the base model are shown in Figure 4.

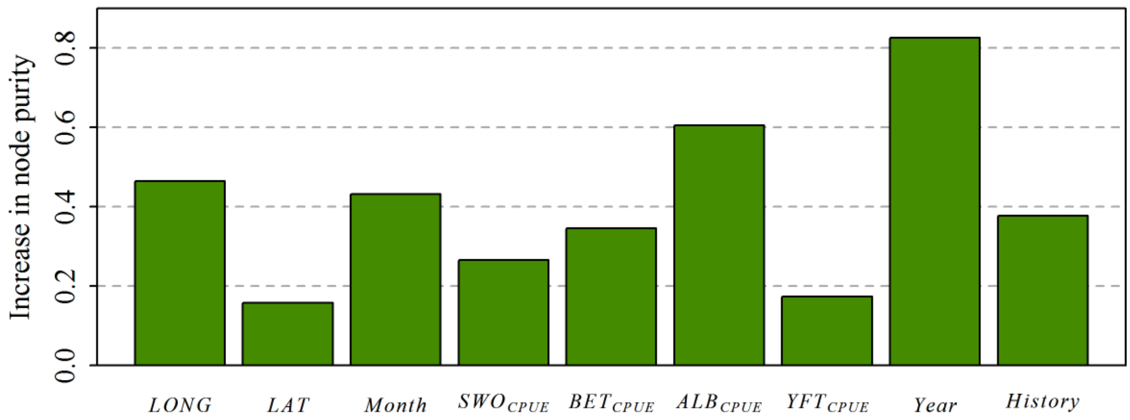

Figure 4 Variable importance plot of the base Indian Ocean random forests CPUE model.

Fitted random forests models are difficult to interpret comprehensively (Prasad et al. 2006). The partial effects plots (Figure 5) provide some indication of the effects of the individual predictor variables of the base model on SBT CPUE in the Indian Ocean.

The partial effect of *Year* increases markedly since 2011, which is consistent with the various monitoring series for SBT CPUE (see e.g. Chambers 2014; Itoh and Takahashi 2014). The effects of the target species CPUE variables generally appear to be fairly modest. The variable importance plot (Figure 4) suggests that albacore catch rates are informative, but their effect is not particularly clear from the partial effects plot (Figure 5). Increased SBT CPUE is expected as fishing moves from the western Indian Ocean to the eastern Indian Ocean. The effect of some variables might be quite different if SBT catch rates were modelled in terms of numbers of individuals instead of weight.

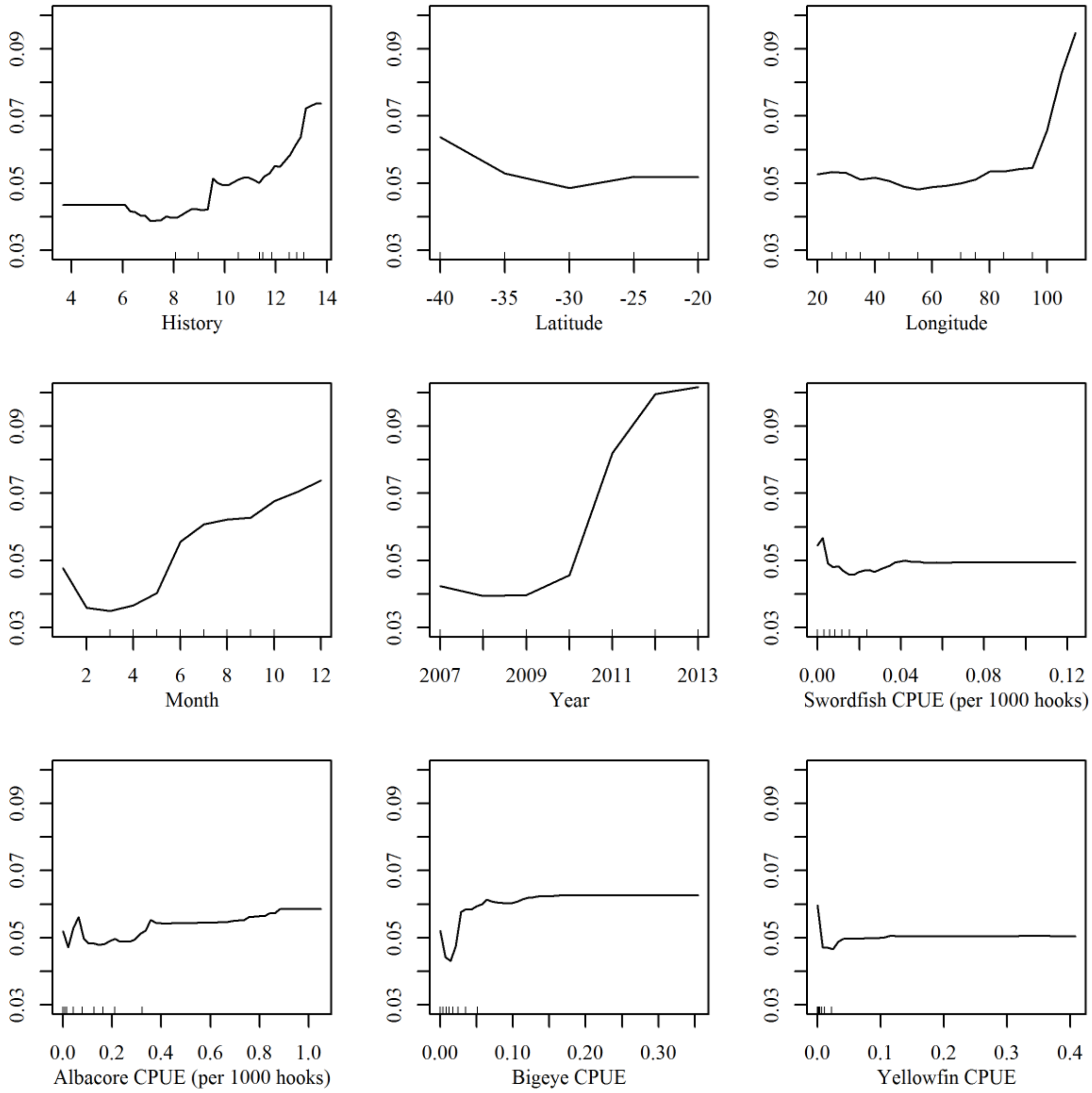

Figure 5 Partial effects of variables in the base Indian Ocean CPUE random forests model.

Selectivity models

In order to model selectivity it was necessary to assign selectivities to the CCSBT member observations. This was done using the designation of 'fishery' described on page 16 of the 2015 CCSBT data documentation (Millar 2015) and was summarised previously in Table 1.

The IOTC data on catch of the Japanese fleet is expressed only in terms of numbers of individuals (see Table 2) whereas catch of most non-members is available only by weight. In order to use Japanese data on catch composition to predict non-member SBT selectivity we define catch of IOTC target species as a fraction of total catch in weight, where possible, but otherwise as a fraction of total catch in number. For example the fraction of swordfish catch is defined as:

$$SWO_{frac} = \begin{cases} \dfrac{SWO_{wt}}{SWO_{wt} + BET_{wt} + YFT_{wt} + ALB_{wt}}, \text{where catch in weight is recorded} \\ \dfrac{SWO_{num}}{SWO_{num} + BET_{num} + YFT_{num} + ALB_{num}}, \text{where catch is recorded in number only.} \end{cases}$$

The species catch fraction terms when calculated in terms of weight will not be entirely consistent with values calculated in term of numbers. However, a compromise of some kind needs to be made in order for the Japanese data to be used to predict the selectivity of the non-member fleets.

The base selectivity model is defined as:

$$Fishery \sim LONG + LAT + MONTH + SWO_{frac} + BET_{frac} + YFT_{frac} + ALB_{frac} + YEAR$$

It might be recognised that the base selectivity model is over determined (the IOTC target species terms have unit sum for all observations). This is not a problem when using random forests because correlation between predictors is not problematic. Random forests classification models are not prone to overfitting (Hastie et al. 2009). Moreover, each component regression tree is fitted with a subset of the predictors, so usually the component regression trees will not be overdetermined in any case.

As was the case when modelling CPUE, prediction of selectivity of SBT captured by the Spanish fleet requires a model that does not include terms involving catch of bigeye, albacore and yellowfin. However, since catch of swordfish by the Spanish fleet is specified both in terms of numbers and weight (see Table 3), swordfish catch rate can be specified in terms of numbers per thousand hooks so that observations from Korea, Taiwan and Japan are consistent in the fitted model.

$$Fishery \sim LONG + LAT + MONTH + SWO_{NPUE} + YEAR$$

Where SWO_{NPUE} is defined as the number of swordfish reported per thousand hooks set.

Estimates of total non-member catch in the Indian Ocean

The fitted CPUE models described above were used to predict SBT catch corresponding to the monthly 5-degree observations of non-members recorded in the IOTC database using the randomForest 'predict' function. Predictions for each observation were obtained from each of the 500 bootstrapped regression trees making up the random forest model by setting the 'predict.all' argument to TRUE. Once SBT catch rates were predicted for the non-member observations, estimated catch of SBT in tonnes was calculated simply as the product of the predicted catch rate and recorded effort.

Table 4 Predicted annual catch of SBT by non-member countries in the Indian Ocean.

Year	Median Estimate (tonnes)	80% Prediction Interval (tonnes)
2007	346	(154, 669)
2008	279	(130, 566)
2009	330	(174, 613)
2010	675	(379, 1139)
2011	583	(242, 1300)
2012	779	(321, 1600)
2013	383	(140, 779)

Observations from the Portuguese and Thai fleets were not used in years where their effort was not expressed in terms of number of hooks set. In years where catch could be estimated for these fleets, the predicted catches were not high. Estimates of annual non-member catch of SBT calculated as described are given in Table 4 and plotted in Figure 6. The uncertainty bars shown in Figure 6 are based on predictions of catch rates for each non-member observation from each of the 500 component regression trees

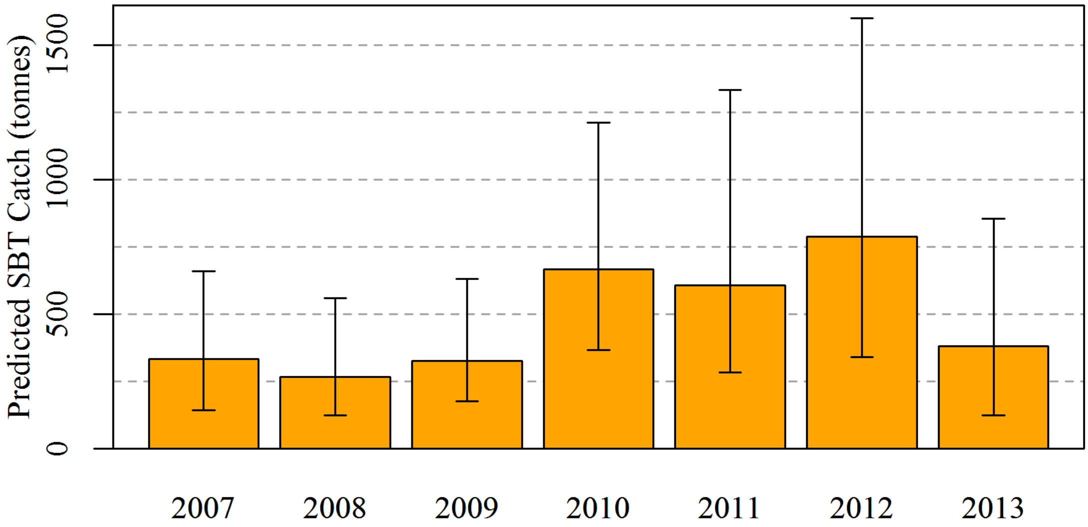

Figure 6 Estimated catch of SBT by non-member countries in the Indian Ocean. Uncertainty bars are estimated 80% prediction intervals based on the distribution of estimates from bootstrapped regression trees that form the random forests model.

Estimated selectivity

The fitted selectivity models were used to predict the 'fishery' class corresponding to each monthly 5-degree observation reported against non-member fleets in the IOTC database. The predicted fisheries were weighted by the mean predicted SBT catch corresponding to that observation to predict the proportion of total non-member catch that should be allocated to each fishery. Approximately 86 percent of the non-member catch across all years in the Indian Ocean was predicted to be consistent with the LL2 fishery and 14 percent with the LL1 fishery. A small number of non-member observations were predicted to be LL3, but these did not contribute a significant proportion of predicted non-member catch.

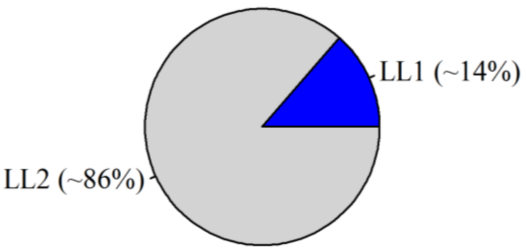

Figure 7 Estimated proportion of non-member catch assigned to the CCSBT selectivity schedules averaged over 2007-2013.

4 The Pacific Ocean

Data

The data describing the catch and effort of fleets targeting tuna and billfish under the jurisdiction of the WCPFC are not publicly available disaggregated by fleet and their provision is subject to confidentiality restrictions. For our purposes it was necessary to distinguish between member and non-member effort and so WCPFC catch and effort data by flag were requested by the CCSBT. Flag specific data corresponding to WCPFC effort south of 20°S that satisfied the confidentiality restrictions of the WCPFC were obtained. These data are explained in greater detail in Hoyle and Chambers (2015).

According to the data obtained from the WCPFC, there has been very little effort by non-member fleets within the area of interest. The WCPFC does not report catch of SBT, so CCSBT and WCPFC datasets needed to be combined if covariates from the WCPFC data are to be used for prediction of SBT catch rates. This being the case, the effort from the WCPFC and CCSBT observations need to be approximately the same so that the WCPFC and CCSBT catch rates can reasonably assumed to have resulted from mostly the same longline sets. In contrast to the IOTC data, all catch is reported in both numbers and weight for all fleets and effort in terms of hooks set is also available for all fleets notwithstanding the confidentiality restrictions.

Neither the WCPFC nor the CCSBT datasets indicate any longline fishing by Korea in the Pacific Ocean south of 20° S. Therefore the prediction model is fitted solely to catch and effort data from Japanese and Taiwanese fleets. We limit consideration to fishing effort in CCSBT Statistical Areas 4, 5 and 6 and Area 7 east of 140° E. Exploratory modelling suggests that predictions of non-member catch rates outside of this area is potentially unreliable.

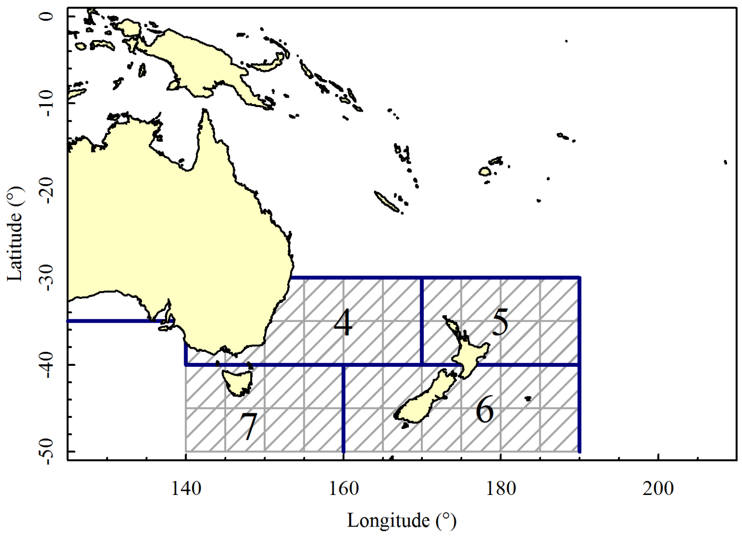

Figure 8 Area used to fit the prediction model and predict non-member catch in the Pacific Ocean. The numbers indicate CCSBT Statistical Areas.

All records of catch and effort data obtained from the WCPFC included catches by weight as well as by number of individuals. This meant the longline catch rates of bigeye, yellowfin, albacore, swordfish, striped marlin, blue marlin and black marlin could be used as covariates consistently for all fleets. However in order to model SBT catch by weight whilst including Japanese observations, it was necessary to convert reported Japanese catch of SBT in numbers to catch in weight. This was done using the method described in Hoyle and Chambers (2015). The correspondence in effort between the WCPFC and CCSBT datasets was checked and, as with the Indian Ocean data, observations that differed by more than 25 percent were excluded (see Fig. 9). After subsetting, 171 Japanese observations and 13 Taiwanese observations were available for analysis.

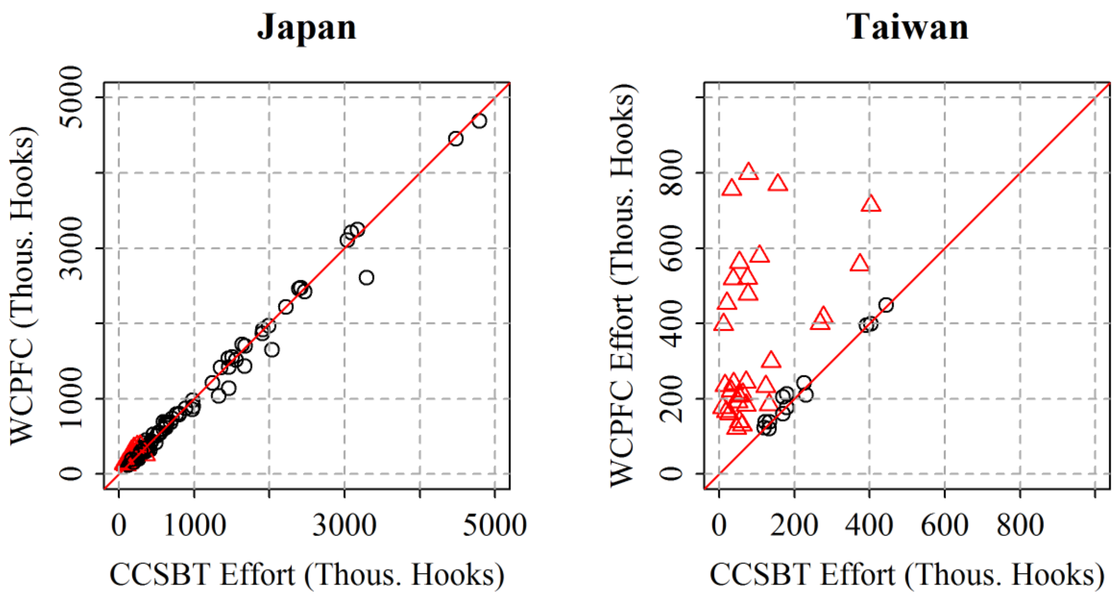

Figure 9 Correspondence of monthly sums of hooks set aggregated at the 5-degree square level according to WCPFC and CCSBT data. Retained observations plotted as open circles, excluded observations plotted as open triangles.

CPUE model

The model fitted is similar to that used for the Indian Ocean analysis except that a few additional catch rate terms are available. The fitted model can be defined as:

$$SBT_{CPUE} \sim LONG + LAT + MONTH + YEAR + SWO_{CPUE} + BET_{CPUE} + ALB_{CPUE} + YFT_{CPUE}$$

$$+ STM_{CPUE} + BLM_{CPUE} + BUM_{CPUE} + OTH_{CPUE} + HISTORY$$

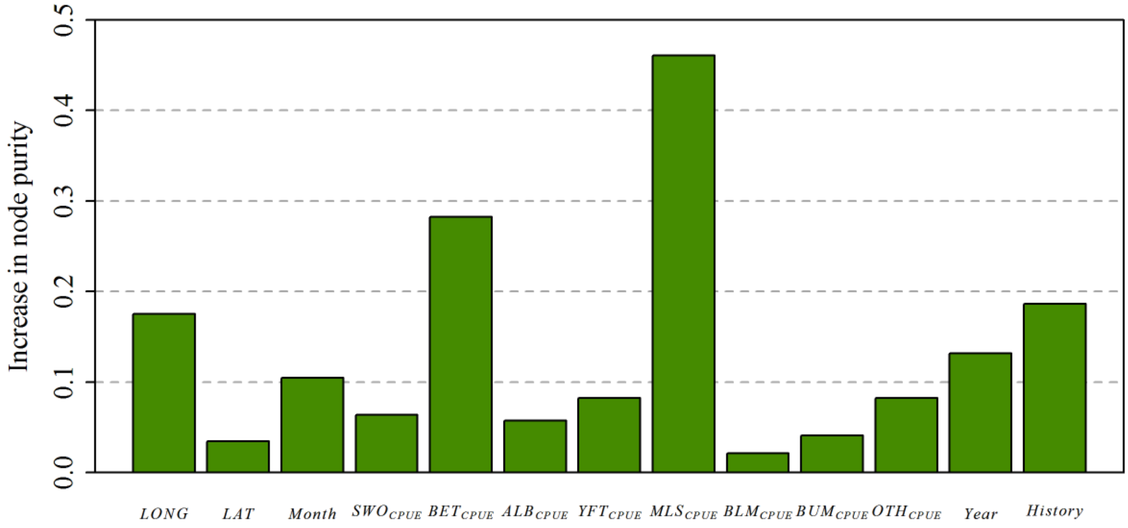

Figure 10 Variable importance plot for the Pacific Ocean CPUE model.

Partial effects plots from the fitted random forests model are shown in Figure 11. Catch of striped marlin produces sharply lower predictions of SBT everything else being equal. The *Year* effect appears to be less pronounced in the Pacific Ocean than in the Indian Ocean. It should be kept in mind, however, that Japanese data were not used in fitting the Indian Ocean CPUE model described in Section 3.

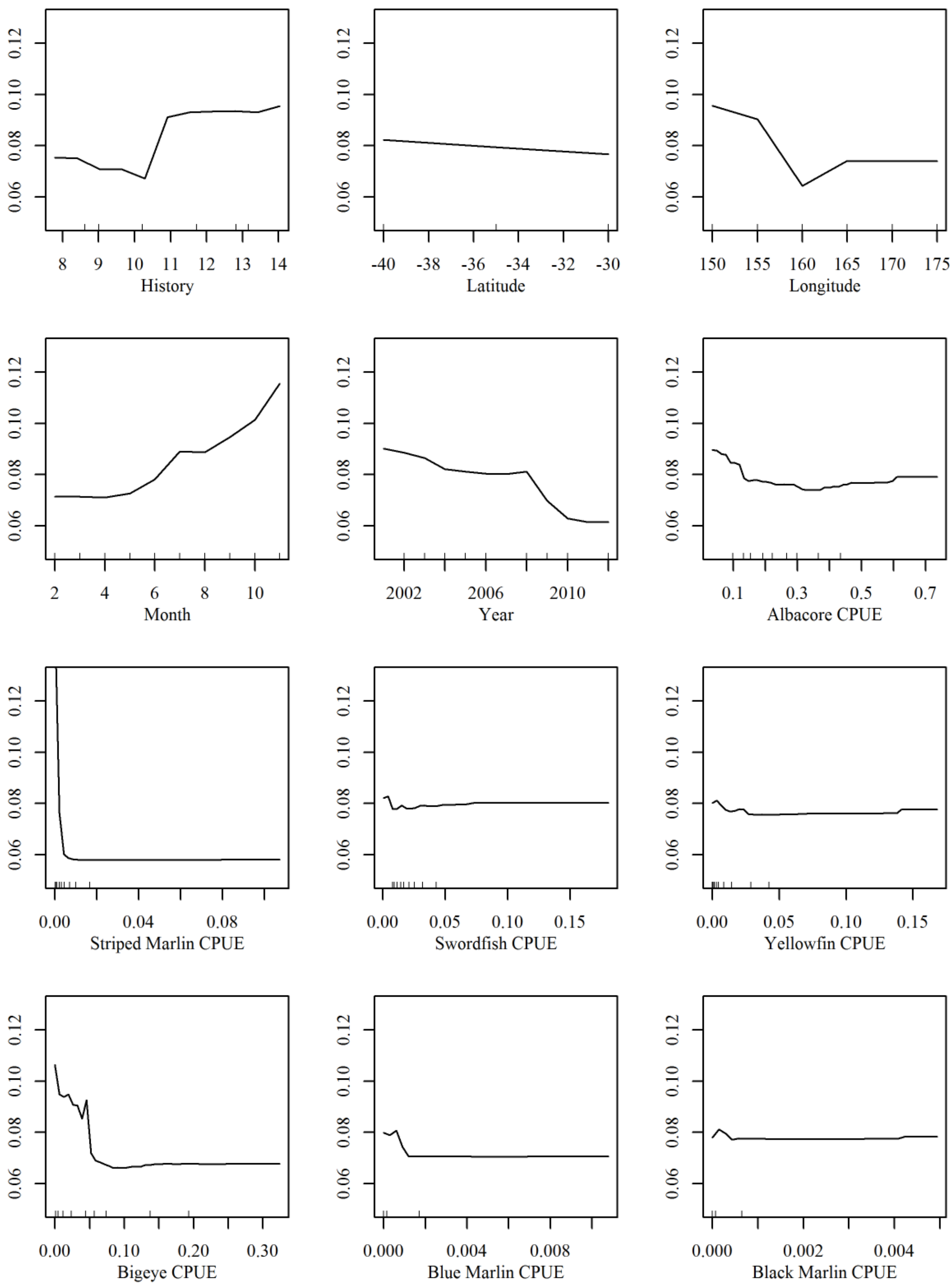

Figure 11 Partial effects plots for the Pacific Ocean CPUE model.

Validation

To examine the ability of the approach to estimate catch the predicted and estimated catch were compared for the Japanese and Taiwanese fleets in the Pacific Ocean. Figure 12(a) shows the predicted annual catches versus the estimated weight of the Japanese catch of SBT reported in numbers. Insufficient Taiwanese observations were retained for the analysis to justify a similar analysis of predicted Taiwanese catch. Figure 12(b) shows residuals versus predicted values for monthly 5-degree observations recorded against the Japanese and Taiwanese fleets. The predicted dataset includes observations where CCSBT effort differs substantially from WCPFC effort (i.e. the observations highlighted as excluded in Figure 9) so some large residuals are to be expected. However, Figure 9 highlights that WCPFC records of Taiwanese effort are generally greater than the CCSBT recorded values. This may be expected given the reporting requirements to CCSBT where, as a minimum, members provide catch and effort data for all operations where SBT are targeted or caught (CCSBT-ERS/1503/04). Nevertheless, since low catch rates of SBT seem to be associated with targeting other species (Figure 11), the pattern of negative residuals shown in Figure 12(b) is unexpected. Although it appears that the observed Taiwanese CPUEs might all be zero, in fact they are just very small compared with the predicted values. This anomaly needs to be better understood.

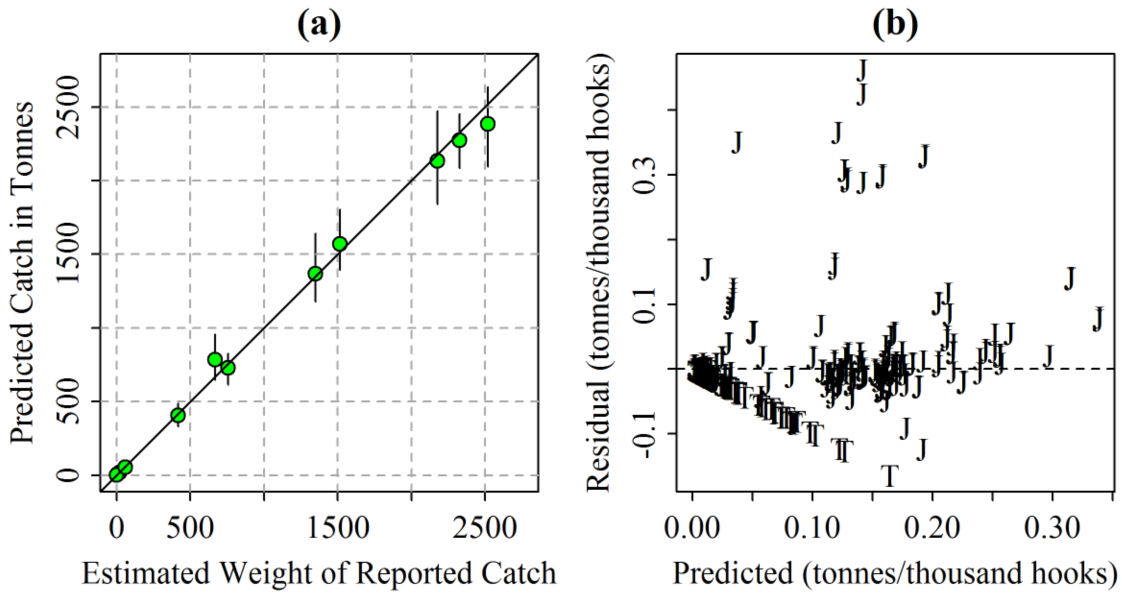

Figure 12 (a) Comparison of predicted annual catch of SBT of Japanese longline fleet in the Pacific Ocean with corresponding estimated from reported catch numbers; and (b) observation level residuals versus predicted values (J = Japan, T = Taiwan).

Non-member effort

The dataset obtained from the WCPFC contains very few non-member observations within the Pacific Ocean area used in this analysis (Figure 8). Overall, there are only six monthly 5-degree square observations from non-members between 2007 and 2012. However, in 2007 for instance, the reported number of hooks set by non-member fleets in the area of interest is high enough to warrant further investigation. It is possible that there was additional non-member effort in the Pacific Ocean area of interest between 2007 and 2012. Given the relatively low quantity of effort that did satisfy the WCPFC confidentiality restrictions it seems unlikely that any effort unable to be provided in this area was substantial.

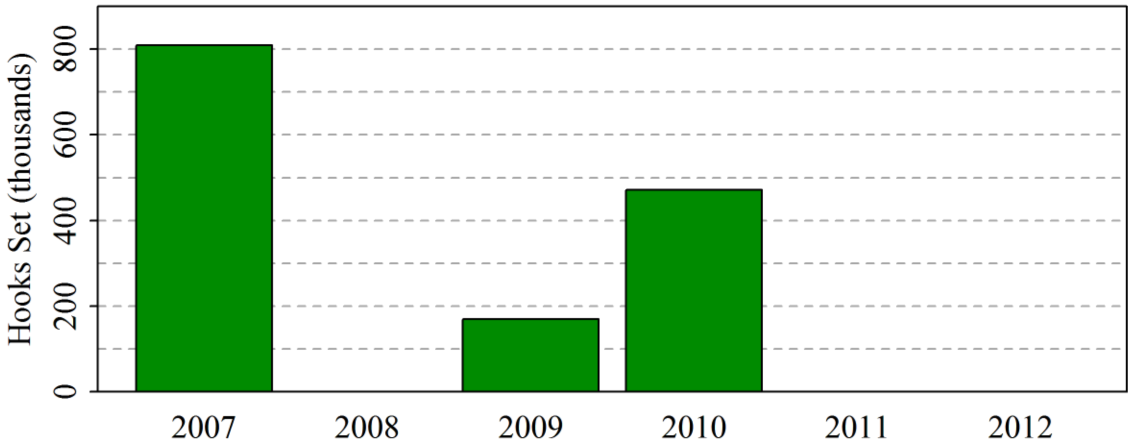

Figure 13 Hooks set by non-member fleets within the Pacific Ocean study area (2007-2012). No non-member effort was recorded in this area during 2008, 2011 or 2012.

Estimates of total non-member catch in the Pacific Ocean

As was the case with the Indian Ocean analysis, the fitted CPUE model is defined in terms of variables that are known for non-member fleets. The fitted model is then used to predict CPUE for the monthly non-member observations based on the values of explanatory variables sourced from the WCPFC. Also as before, catch is estimated for each aggregated observation simply by multiplying the catch rate by the reported number of hooks set. Each of the 500 bootstrapped regression trees that comprise the random forest model predict a slightly different set of annual non-member catches. Uncertainty in the predicted non-member catch is inferred from the distribution of predictions given by the individual regression trees. Estimated non-member catch of SBT in the Pacific Ocean between 2007 and 2012 are plotted in Figure 14 and summarised in Table 5.

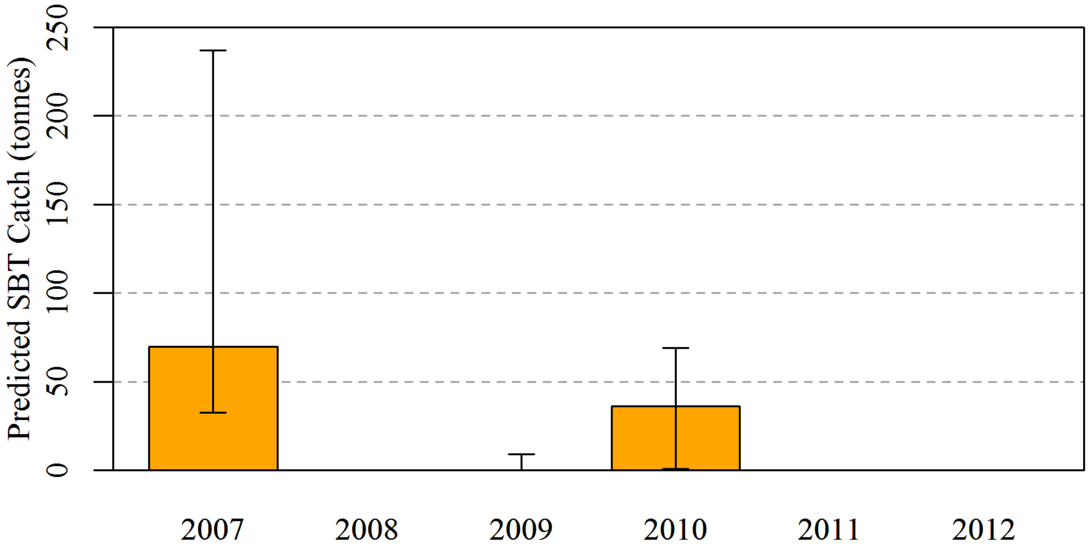

Figure 14 Estimates of annual non-member catch of SBT in the Pacific Ocean with 80 percent prediction intervals. Orange bars are estimates based on the medians from predictions derived from 500 bootstrapped regression trees of the random forests model.

Presumably all non-member catch in the Pacific Ocean would be assumed to have selectivity as defined by the LL1 fishery.

Table 5 Predicted non-member catch in the Pacific Ocean 2007-2012.

Year	Median Estimate (tonnes)	80% Prediction Interval (tonnes)
2007	70	(33, 237)
2008	0	-
2009	0	(0, 9)
2010	36	(1, 69)
2011	0	-
2012	0	-

5 Concluding remarks

We have described an approach for estimating catch of SBT by non-member longline fleets using data reported to the IOTC and WCPFC in combination with CCSBT data. We have provided some estimates of potential non-member catch of SBT between 2007 and 2012 given effort reported to the IOTC and WCPFC. The median estimated non-member catch is compared with the total global catch reported to the CCSBT in Figure 15.

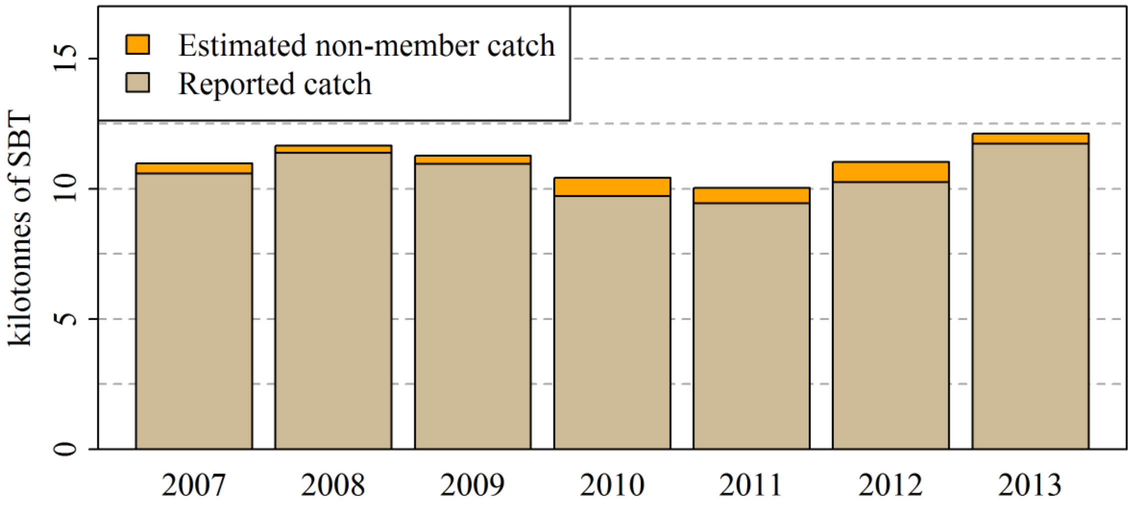

Figure 15 Barplot showing suggested sum of median estimated non-member catch and reported catch of SBT. The reported catch of the EU has been subtracted from the estimated non-member catch in the plot.

The reliability of the estimated non-member catch is uncertain because of the possibility that SBT catch rates of non-member fleets differ from CCSBT member fleets in ways that are not addressed by the models that we have applied. Whilst model diagnostics can be used to evaluate how well the model predict the catch rates of SBT by CCSBT members, it is likely that predictions of non-member catch are less accurate.

There are other, perhaps greater, sources of uncertainty as well. We have modelled the potential catch by non-member fleets, but cannot know what proportion of SBT captured is released alive. The inability to use the Japanese CPUE in the Indian Ocean is a potential weakness for that analysis. We have assumed that the non-member effort sourced from the IOTC and WCPFC datasets is comprehensive, but this is unlikely to be the case. For example, the IOTC data are prefaced with the statement 'Catches and effort are not available for all Nominal catches strata. When recorded, the catches in these datasets might represent the total catches of the species in the year for the fleet and gear concerned or represent simply a sample of those'. In addition, we have not considered the possibility of non-member catch of SBT in the Atlantic Ocean (area managed by ICCAT). It can be seen from Figure 1 that SBT have been frequently caught historically in the south east Atlantic Ocean.

This work has highlighted inconsistencies between the data held by the IOTC, the WCPFC and the CCSBT. The rationale for reporting and not reporting particular longline sets to particular RFMOs is not entirely clear and the current situation reduces the confidence that can be placed in analyses such as described here. For example, Hoyle and Chambers (2015) point out that CCSBT catch and effort data suggest decent average SBT catch rates well to the east of New Zealand despite the total reported catch of SBT from this area being very low. A better solution might be a single database that includes all catch and effort available to all tuna RFMOs where users can subset the data according to their individual needs.

Noting the uncertainties, the estimated non-member catch is small compared with the reported catch of SBT (Figure 15). Relatively low bycatch of SBT might be expected because its distribution tends to be temperate latitudes whereas alternative target species for longline fleets are caught farther north. There is little or no evidence of non-member effort on the SBT spawning grounds in the IOTC data since 2007.

References

Anon. (2014) Report of the nineteenth meeting of the Extended Scientific Committee, 1-6 September 2014, Auckland, New Zealand.

Breiman, L. (2001) Random forests. *Machine Learning* **45**, 5-32.

CCSBT (2015) Summary of the level and distribution of SBT effort. CCSBT-ERS/1503/04.

Chambers, M. (2014) A CPUE index based on a GAMM: a proposed monitoring series. CCSBT-ESC/1409/09.

Hastie, T., Tibshirani, R. And Friedman, J. (2009) The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd edn. Springer, New York.

Hoyle, S. and Chambers M. (2015) Estimating southern bluefin tuna catches by CCSBT non-members by combining reported effort with inferred sizes and catch rates. CCSBT-ESC/1509/21.

Itoh, T. and Takahashi N. (2014) Update of the core vessel data and CPUE for southern bluefin tuna in 2014. CCSBT-OMMP/1406/13.

Larcombe, J. (2014) Fleet overlap in the IOTC area. CCSBT-ESC/1409/13.

Lennert-Cody, C.E. and Berk, R.A. (2007) Statistical learning procedures for monitoring regulatory compliance: an application to fisheries data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **170**, 671-689.

Liaw, A. and Wiener, M. (2002) Classification and regression by randomForest. *R News* **2**, 18-22.

Millar, C. (2015) CCSBT Data CD Documentation.

Pons, M., Marroni, S., Machado, I., Ghattas, B. and Domingo A. (2009) Machine learning procedures: an application to by-catch data of the marine turtles *Caretta carreta* in the southwestern Atlantic Ocean. Collect. Vol. Sci. Pap. ICCAT, 64(7) 2443-2454.

Prasad, A.M., Iverson, L.R. and Liaw, A. (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* **9**, 181-199.

R Development Core Team. (2012) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Shono, H. (2014) Application of support vector regression to cpue analysis for southern bluefin tuna, *Thunnus maccoyii*, and its comparison with conventional methods. *Fisheries Science* **80**, 879-886.