Verifying FAD-association in purse seine catches on the basis of catch sampling

Steven R. Hare*, Shelton J. Harley, W. John Hampton

Secretariat of the Pacific Community, Oceanic Fisheries Programme, BP D5, 98848 Noumea, New Caledonia

A B S T R A C T

We investigate the potential of verifying whether individual purse seine sets were made in association with a fish aggregation device (FAD) or on an unassociated (FAD-free) tuna school, on the basis of low intensity catch sampling by onboard observers. The target tuna catch and length compositions and bycatch amounts were analyzed from more than 50,000 purse seine sets sampled by onboard observers who had, in addition to collecting sampling data on species and size composition of target tunas in the catch, and set-level estimates of total bycatch, also identified the sets as either "associated" or "unassociated". Classification tree (CT) models were developed based on 2007–2011 observer data and tested for misclassification error rates on 2012 data. Two types of model misclassification errors (MCE) are possible: unassociated sets misclassified as associated (termed false positive or Type I) and associated sets misclassified as unassociated sets (false negative or Type II error). A third error measure, overall MCE, is a weighted average of Type I and Type II errors. The classification rules developed on the basis of observer catch sampling tended to be nearly presence/absence, e.g. greater than 99% skipjack composition or presence of 0.5 kg rainbow runner, likely keyed by the modest observer sample sizes. Overall MCE rates were 21.8% for the initial tuna-only CT model and 14.4% for the bycatch-included model. The improvement in overall classification for the bycatch models derived principally from a reduction in Type I errors. The addition of auxiliary non-sampling variables (e.g., longitude, month) and use of more complex resampling extensions to CT modelling led to little to no improvement in MCE rates. We employed our methodology to analyze a particular subset of the purse seine data, i.e., sets from the FAD-closure periods of 2009–12. The intent was to determine if MCE rates of these particular sets were greater than the MCE rates found in the more general analysis. Reassuringly, the MCE rates of sets during the FAD closure period were found to be equal, or even a bit lower than MCE rates in the broader analyses based on our best performing model.

1. Introduction

Purse seine catches are generally categorized as either "unassociated" or "associated" with fish aggregation devices (FADs). Purse seine fishing, specifically targeting skipjack (*Katsuwonus pelamis*) and yellowfin (*Thunnus albacares*) tunas, but also taking small amounts of bigeye (*Thunnus obesus*) tuna, has grown substantially over the past three decades in the Western and Central Pacific Ocean (WCPO), increasing from around 100,000 mt in 1980 to nearly 1.8 million mt in 2012 (Harley et al., 2012). Unassociated, or free-school, fishing accounted for the majority of purse seine catches up until the mid-1990s; since that time catches have been

near evenly split between unassociated and associated sets. In this context, the term "... Fish Aggregation Device (FAD) means any man-made device, or natural floating object, whether anchored or not, that is capable of aggregating fish." (WCPFC, 2009)

Concerns over the composition of catches associated with FAD-fishing have led to recent calls to regulate FAD-fishing, either via regulatory actions (Fonteneau et al., 2013), or educating consumers.[1] As part of the increasing consumer scrutiny related to seafood sustainability, increasing numbers of sea food purchasers seek tuna that have been certified to be free school caught.[2] In

---

* Corresponding author.
  E-mail address: stevenh@spc.int (S.R. Hare).

Table 1
Summary of observed purse seine set data used in analysis. Abbreviations are as follows: FAD – fish aggegation device; UNA – unassociated; ASS – associated.

	Non-FAD closure sets			FAD closure sets			Total sets		
	UNA	ASS	Total	UNA	ASS	Total	UNA	ASS	Total
2007	1133	2286	3419				1133	2286	3419
2008	1270	2186	3456				1270	2186	3456
2009	984	2545	3529	796	419	1215	1780	2964	4744
2010	5458	5401	10859	2679	280	2959	8137	5681	13818
2011	4241	9094	13335	1881	711	2592	6122	9805	15927
2012	3033	4724	7757	2260	825	3085	5293	5549	10842
Total	16119	26236	42355	7616	2235	9851	23735	28471	52206

general, FAD-associated catches contain a higher proportion of bigeye tuna, a greater array of bycatch species and typically smaller sized fish than unassociated schools (Dagorn et al., 2012). A FAD-closure period, covering the months of July (since 2010), August and September (since 2009), has been instituted annually since 2009 by the Western and Central Pacific Fisheries Commission (WCPFC), the international body responsible for management of the WCPO tuna fisheries.

The WCPFC mandated 100% observer coverage starting with the 2010 fishing year (WCPFC, 2013). Prior to 2010, observer coverage of purse seine catches was in the vicinity of 20%, and actual observer coverage levels since 2010 have been around 65% (Williams, 2014). All purse seine vessels operating in waters of nations within the WCPO are required to complete vessel logs for every set, including classifying sets as unassociated or associated. Observers also routinely record set association for every set while aboard a vessel. Despite this duplicate recording of set type, there remains demand for an independent determination of set type (Harley et al., 2009). For example, such a determination might be useful in retrospective analyses such as examining performance of new observers. Observer determination of set type might be either purposefully or inadvertently incorrect, e.g., the observer might be unaware that a set is associated with a FAD, given that FADs can be objects as small as pieces of rope or floating garbage bags. Historically, whale sharks and, possibly, turtles or marine mammals, have also served as natural FADs. We note, however, that setting of purse seines on incidental or natural FADs has become extremely rare given the vast deployment of satellite, and often sonar-equipped, FADs which are believed to number in the tens of thousands in the WCPO.[3]

We fit a sequence of classification models of increasing complexity to predict purse seine set association. We start with the use of observer-collected sampling data and then explore the utility of additional non-sampling ("auxiliary") predictors. The simplest models look for consistent differences in observer-collected sampling data, specifically relative species composition and length compositions of the tuna catches and, optionally, the amount and species of bycatch present in the set. More complex models make use of bootstrap techniques applied to the observer sampling data, and are then fitted with the auxiliary predictors to attempt to improve set association prediction. For each model and dataset, one subset of the data is used to "train" the models and these models are then applied to another subset, the test data, which were not used in model fitting. Our approach here is to initially develop relatively simple and robust classification rules and we then extend the methodology to determine if more complex models and auxiliary predictors increase purse seine set association prediction.

For the purposes of model development, historical observer set type classification is taken as "truth". We feel this to be valid both because much of the data were collected during periods when there was little incentive on the part of vessel operators to misreport set association (and possibly pressure observers to do likewise). However, concerns over potential "contamination", i.e., intentional mis-labeling, did lead us to isolate purse seine sets from the FAD closure periods during 2009–2012. While there are spatial and national exceptions by which setting on FADs is still allowed during the FAD closure period, broad regions are open only to free school purse seining. Possibly clandestine FAD setting and/or routine labeling of sets as non-FAD associated without observer verification constitute potential avenues of set type intentional mis-labelling. The classification models developed in this analysis were trained and tested on datasets without the FAD closure data; we conclude the analysis with an examination of the performance of the preferred classification models on prediction of FAD association for the FAD closure dataset.

2. Materials and methods

The data used in this analysis come from the Secretariat of the Pacific Community (SPC) maintained observer database that contains observations on purse seine operations from 1993 to the present. The database from which these data were extracted represents a filtered, quality-controlled, subset of the total database (for example, we excluded data for the first trip by a new observer). Additionally, this analysis is restricted to observed sets with both
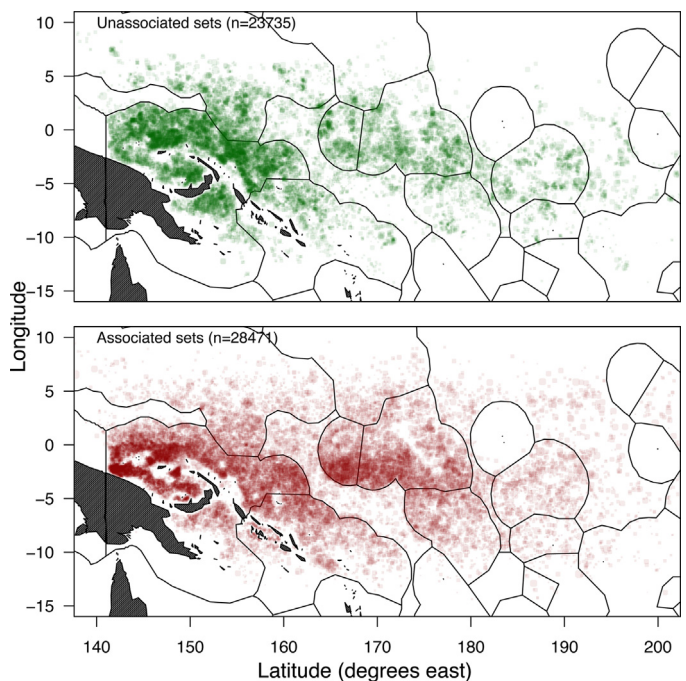

Fig. 1. Locations of all observed purse seine sets, separated by set type, for the period 2007–2012. Size of individual circles is proportional to total target tuna catch. Exclusive Economic Zone boundaries are informal and purely illustrative.

---

[3] Pew Environment Group. 2014. Estimating the use of drifting Fish Aggregation Devices (FADs) around the globe. Discussion Paper available at http://www.pewtrusts.org/ ~/media/legacy/uploadedfiles/FADReport1212pdf.pdf.
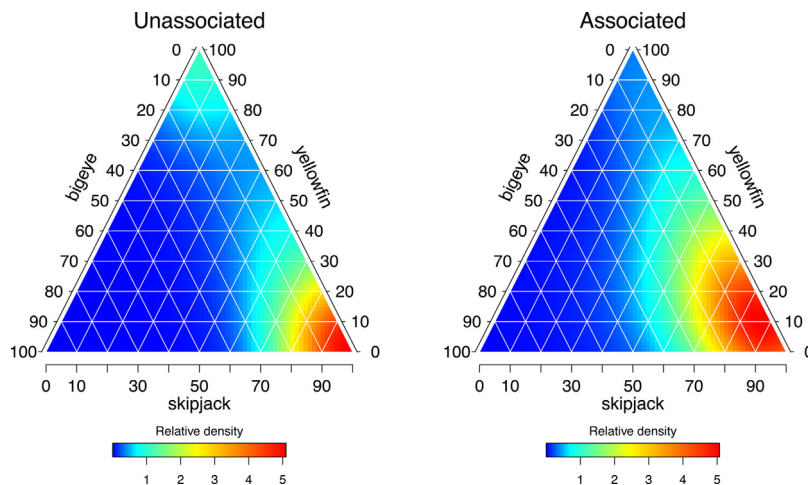
Fig. 2. De Finetti (ternary) plots summarizing relative catch composition of the three target tuna species for associated and unassociated purse seine sets. An example of how to determine the percentages of the three tuna species at a particular location in a graph is provided. Density indicates relative proportion of total sets having the indicated mix of tuna catch proportions.

recorded target tuna catch as well as recorded tuna lengths. For purposes of data summaries and model fitting, we limited the dataset to the 2007–2012 time frame. Table 1 lists the number of observer-classified purse seine sets, with associated sets comprising 54.5% of all sets over the 2007–2012 time frame. The spatial distribution of the sets shows essentially complete overlap between the two set type associations (Fig. 1).

The 52,206 sets comprise 73.2% of all observed sets in the filtered database. Years earlier than 2007 represent a time period that is likely less relevant to more recent years in terms of fishing methods, areas, or catch composition. Data for 2013 are at present very incomplete and not representative with respect to areas and seasons so were not included. Over the past six years there have been roughly similar numbers of FAD-Free ("unassociated") and FAD-Associated ("associated") observed sets. The proportion of unassociated sets has generally increased since 2010, coinciding with both increasing total purse seine effort and implementation of the FAD-closure periods within the WCPO. A total of 9851 sets are from the FAD closure periods of 2007–2012. These sets were not included in the development or testing of the classification models, but were held aside for subsequent testing with the preferred classification models. To summarize, the training data consisted of 34,598 sets (non-FAD closure data, 2007–2011), the test data consisted of 7757 sets (non-FAD closure data, 2012); and the 9851 FAD-closure sets were examined for evidence of potential intentional mis-labeling.

The observer data used in the analysis were collected using a method termed "grab sampling" which has been consistently utilized dating to the start of onboard purse seine set sampling. In essence, the observer is instructed to randomly collect five tuna

Table 2
Mean catch and frequency of occurrence of edible bycatch species in observed purse seine sets.

Name	Abbreviation	Unassociated		Associated	
		kg/set	occurrence	kg/set	occurrence
Barracudas	bar	0.07	0.4%	2.62	2.9%
Black marlin	blm	6.33	3.6%	7.61	4.4%
Blue marlin	bum	10.25	6.6%	16.49	8.4%
Dolphinfish	dol	1.09	1.6%	29.43	26.0%
Striped marlin	mls	2.20	1.5%	3.81	2.5%
Rainbow runner	rru	6.98	2.6%	119.35	63.5%
Sailfish	sfa	0.52	1.0%	0.45	0.8%
Wahoo	wah	0.35	0.9%	11.23	16.9%

from each brail used to empty the purse seine net. Mean grab sample size from each set is 65 fish though variability in sample size is very large, consistent with the nature of purse seine set catch sizes. To put this "low intensity" sampling rate into perspective, a 100% skipjack set typically averages 30 mt; at an average weight of 3 kg, the observer sample of 65 skipjack represents 0.65% (195/30,000). Bycatch data are not subsampled; observers utilize a variety of means of estimating full set weights of all non-target species. While quantifying bycatch is considered a "secondary" observer priority, experienced observers are considered capable of enumerating bycatch, particularly the "edible" species that are often separated and retained from the set. The perceived importance of bycatch enumeration, both for stock assessment and management purposes, has also led to increased training in bycatch estimation in recent years.[4]

Potential covariates, or predictor variables, for classifying set type were (1) tuna species composition; (2) various measures of tuna length; (3) species bycatch per set; and (4) non-sampling variables related to the time and place of fishing. These categories are discussed in turn.

2.1. Tuna species composition

In Fig. 2, the relative proportions of the three target tuna species, within associated and unassociated purse seine sets, are presented in ternary, or De Finetti, plots (Fonteneau et al., 2010). These plots illustrate that both associated and unassociated sets are most often comprised of 90+% skipjack. However, at least two differences in relative catch composition between the two set types are also evident. Unassociated sets targeted on skipjack tend to be purer, and there are occasional sets that are nearly 100% pure yellowfin. Associated sets most frequently contain 10–20% yellowfin and/or bigeye tuna. In the results section, these three variables are abbreviated as SKJ.pct, YFT.pct, and BET.pct, representing the percentage of skipjack, yellowfin, and bigeye tuna, respectively in the grab samples for the set.

2.2. Tuna length composition

The three target tuna species captured in unassociated sets tend to have a larger size distribution than those in associated sets

---

[4] Pers. comm., Peter Sharples, Observer and Port Sampler Coordinator, SPC.
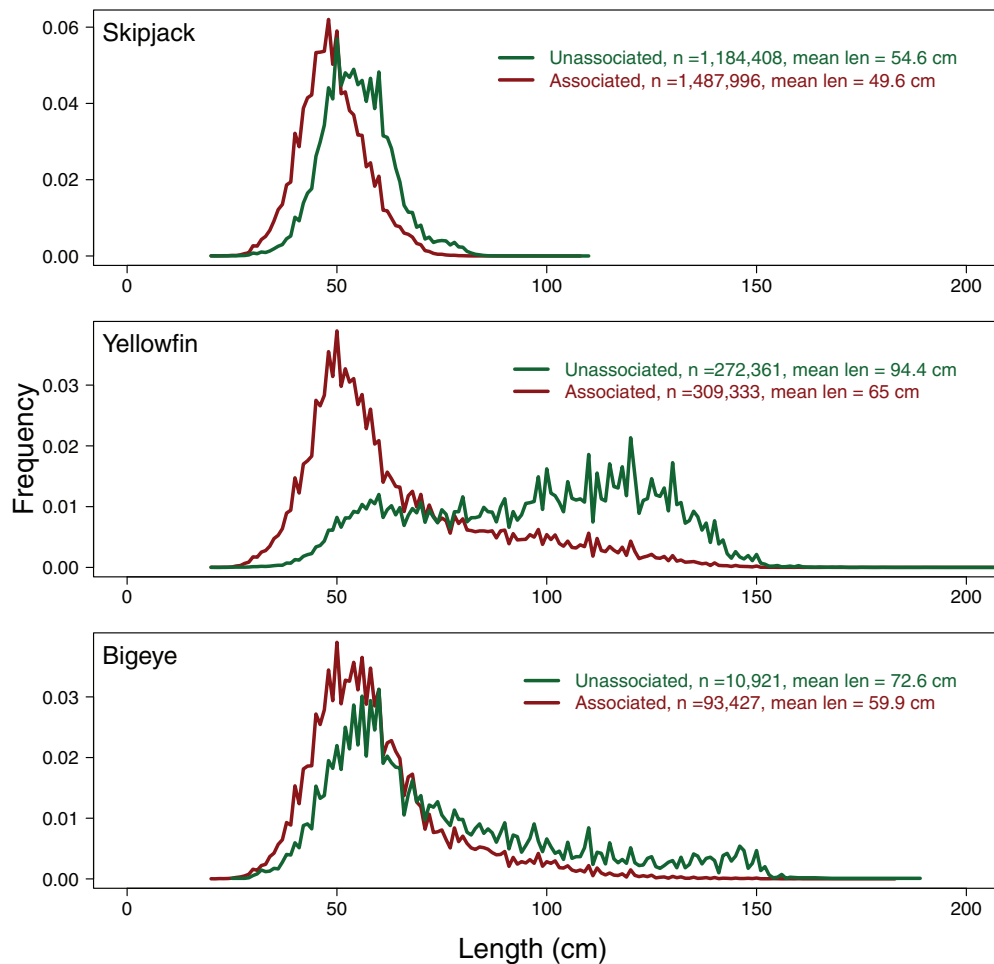
Fig. 3. Size distributions for the three target tuna species, measured by observers between 2007 and 2012, broken down by set association. The number of measured fish for each distribution is indicated by *n*.

(Fig. 3). Small yellowfin tuna (≤50 cm), in particular, are not commonly caught in unassociated sets, but form the bulk of yellowfin catch in associated sets. We computed the mean lengths of skipjack, yellowfin and bigeye tuna for each set in which any of these species were captured. The 25th, 50th and 75th length quantiles were also computed, but early analyses showed no improvement over use of simple mean length, and they were dropped from the analysis. Fig. 4 shows a boxplot of the differences in mean length distribution between set types. These variables are abbreviated SKJ.len, YFT.len, and BET.len where "len" is interpreted as mean length in the set.

2.3. Bycatch composition

Bycatch data, estimated total weight per set, were limited to the eight most common "edible species" – barracudas (*Sphyraena* spp.), black marlin (*Istiompax indica*), blue marlin (*Makaira mazara*), dolphinfish (*Coryphaena hippurus*), striped marlin (*Kajikia audax*), rainbow runner (*Elagatis bipinnulata*), sailfish (*Istiophorus platypterus*), and wahoo (*Acanthocybium solandri*)). With the exception of rainbow runner, dolphinfish and wahoo in associated sets, the bycatch rates of the eight most common bycatch species are very low (Table 2), but we chose the edible species as these were most likely to be retained and thus their presence more easily observable. Bycatch species name abbreviations used for naming conventions in the results section are listed in Table 2, followed by mean total weight and frequency of occurrence per set. An examination of the fate of these species indicated that 20–60% of
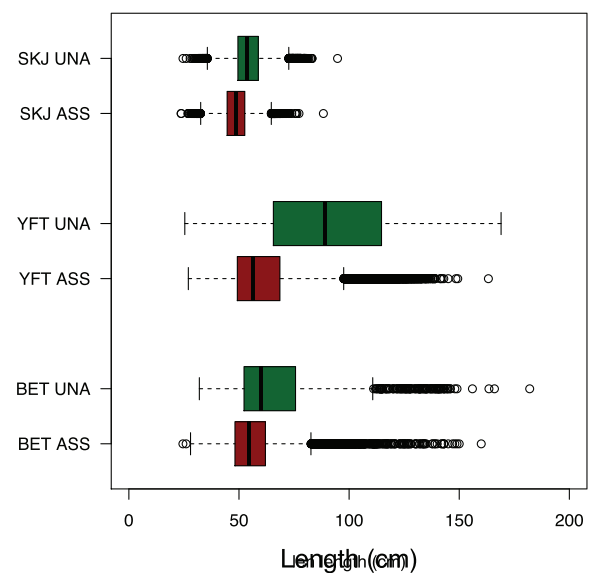

Fig. 4. Boxplots of distribution of mean lengths for three target tuna species broken down by set association. The shaded regions show the 25th and 75th quantile while the black bar is the median. Outliers are illustrated by circles, and often represent single measurements, i.e. only one fish caught in a set. Tuna species abbreviations are: skipjack (SKJ), yellowfin (YFT), bigeye (BET); UNA indicates unassociated sets and ASS indicates associated sets.

fish might be retained (varying by the flag of the vessel and unloading port) and much of the retained fish is consumed onboard by the crew. We note that while sharks are a common bycatch in associated sets, a strict no-retention policy for certain species makes use of shark bycatch data less suitable for set type determination.

2.4. Non-sampling variables

We complemented the observer-collected sampling data with a set of variables relating to the temporal, spatial and environmental characteristics of each purse seine set.

1. Temporal – year (2007–2012), month (1–12).
2. Spatial – latitude (∼30°S–30°N), longitude (∼135–205°E).
3. Environmental – sea surface temperature (SST), monthly 1° × 1° mean value where purse seine catch was taken, as indicated from the Reynolds–Smith Optimally Interpolated Version 2 dataset (Reynolds et al., 2002).
4. Associated – total purse seine set weight, vessel flag and Exclusive Economic Zone (EEZ). EEZ and vessel flag are country-specific referencing national waters where fishing took place and the nationality of the fishing vessel, respectively.

There has been relatively little work done to date on the subject of predicting FAD-association from catch-related data. Pallarés et al. (2003) used two variables – an average sample weight and a catch diversity index – to assign unobserved catches as either unassociated or associated. Their analysis, however, was based on very small sample sizes and the intent was to classify sets for historical purposes and no cross-validation was conducted. In a more recent study, Lennert-Cody et al. (2013) used a classification technique known as "random forests" to determine set association for the purposes of estimating dolphin mortality associated with purse seine fishing.

Random forests is a technique within the more general set of methods collectively referred to as Classification and regression tree-based methods. Tree-based models have found widespread application in the fields of decision-making and prediction. Our analysis is comprised solely of classification tree-based models as they are used to predict factors (purse seine set association in our analysis); regression trees are used to predict values. The predictor, or classification, variables can be either categorical or continuous. Each step of the decision is conditioned on a "branch" of the decision tree, each branch of which is determined through a recursive estimation process. This method lends itself to establishing a set of simple rules that can be used to estimate whether a sampled purse seine set is likely to be FAD-unassociated or FAD-associated. Models are developed by sequentially identifying variables that best separate the data into similar categories, continuing until the decreased improvement in classification does not warrant addition of more predictor variables. Variable importance for CT models is computed for each variable based on the decrease in the Gini impurity index (Breiman et al., 1984).

The simplest method, hereafter referred to as the CT model, produces a single prediction tree, with the model utilizing the full set of predictor variables and values. All data analyses conducted herein were based on the R Programming language (R Core Team, 2013) and the CT model fits used the "rpart" package (Therneau et al., 2013). The two main CT model fitting control parameters in rpart are the "complexity parameter" (cp) and "minimum branch size" (minsplit). For all CT model fits, the settings for these two parameters were: cp = 0.01 and minsplit = 30. A strength of CT models is the allowance of missing values; a hierarchical decision process ensures that a classification can always be made. The importance of this feature is further discussed under random forests and the

implementation methodology illustrated by example in the Results section.

Following the initial CT modelling, we then utilized more complex methodologies that use bootstrapping techniques to see if misclassification rates could be improved upon. The method of bagging predictors ("Bootstrap aggregating", Breiman, 1996) – hereafter, BP model – is based on building CT models for a number of datasets, each a bootstrap replicate of the original dataset. The BP model prediction is a plurality (majority in the case of a binary response variable) vote among the bootstrapped CT models. Per the original paper describing the method (Breiman, 1996): "the vital element is the instability of the prediction method. If perturbing the learning set can cause significant changes in the predictor constructed, then bagging can improve accuracy."

BP models have several adjustable parameters, e.g., the size of bootstrap samples, the number of trees to construct, the minimum branch size, etc. For this analysis, we explored a number of settings. Some of the settings can result both in data overfitting and substantial increases in computing time, possibly with little increase in predictive power. We ultimately chose the following settings that provided a balance of complexity and close to best predictive power. Bootstrap samples were of size n out of n with replacement (where n is the number of sets in the training sample); 30 trees were constructed, minimum branch size was set to 100 (i.e., no branches with fewer than 100 observations are split off), and the complexity parameter was set to 0, meaning that any data split that increases overall model fit (subject to other parameter settings) is pursued. All BP models were fitted using the "ipred" package (Peters et al., 2015). Also, as BP models are, at their core, an ensemble of CT models, they also allow for missing data. The decrease in the Gini impurity index for each variable is summed across the models and plotted to illustrate relative variable importance.

The third set of models we used were random forests (Breiman, 2001), hereafter RF models. Widely used in other disciplines, RF models are relatively new to ecology (Cutler et al., 2007). RF models have a number of attractive features, including being nonparametric, efficient in handling large data sets, and being fairly robust to overfitting issues (Yi, 2012). The general idea behind RF models is to extend bagging by constructing trees from subsets of the prediction variables. In this way, a "random forest" of trees, each built from subsets of the data and using a subset of variables, is built to produce more robust predictors. Typically, two-thirds of the training data set is used to develop the forest and the remaining "out of bag" values are used to test the predictors. We fit RF models using the randomForest package (Liaw and Wiener, 2015).

RF models have two basic parameters, the number of variables used at each node and the number of trees in the forest, and the model tends not to be very sensitive to their values (Liaw and Wiener, 2002). For all analyses, we used the randomForest default values of sqrt(p), where p is the number of variables and 500 trees, respectively. A third parameter, class weight, can be adjusted to balance prediction error. We develop two sets of RF models, one that minimizes total prediction error (unequal error rates) and a second that minimizes total error subject to error types being of equal magnitude (balanced error rates). RF models can handle missing-observation data, typically by imputing values based on proximity measures. However, our dataset contains substantial "not applicable" (NA) data that cannot be imputed. The three average length variables take on values only when there are catches of those tuna species in the purse seine set. Imputing average length values for sets with no catches of a tuna species would be inappropriate. The RF models were fitted without the average length variables because the incidence of zero catch for any of the three species in any given set, hence no average length values, is quite high and RF models would require deletion of all sets with N/A average length values. While alternative measures to quantify the relative importance

Table 3
Description of the misclassification error (MCE) types and formulas for computing MCE rates. Abbreviations are UNA (for unassociated) and ASS (for associated).

Error type	Description	MCE rate calculation
Type I	False positive: unassociated set misclassified as associated set	$\frac{No.\,UNA\,sets\,misclassified\,as\,ASS\,sets}{Total\,No.\,of\,UNA\,sets}$
Type II	False negative: associated set misclassified as Unassociated set	$\frac{No.\,ASS\,sets\,misclassified\,as\,UNA\,sets}{Total\,No.\,of\,ASS\,sets}$
Overall	Type I + Type II	$\frac{No.\,misclassified\,sets}{Total\,No.\,of\,purse\,seine\,sets}$

of variables across the forest of models are available (Nicodemus, 2011), we report only the Gini Impurity Index, consistent with output reported from the other methods.

We fitted and compared five model/dataset combinations as follows:

1 CT1, sampling data only.
2 CT2, sampling and auxiliary data.
3 BP, sampling and auxiliary data.
4 RF1, sampling data (minus average length variables) and auxiliary data, unequal error rates.
5 RF2, sampling data (minus average length variables) and auxiliary data, balanced error rates.

All five models are fitted such that an overall misclassification error (MCE) rate is minimized for the training data set. This overall MCE rate is a mix of two types of misclassifications, which are referred to as Type I and Type II errors. Type I, or false positive, errors are unassociated sets misclassified by the model as associated set; Type II, or false negative, errors are associated sets misclassified by the model as unassociated sets. The calculation of the two types of errors is as described in Table 3. The overall MCE rate is a weighted average of the two error types, and thus always falls between the two. For the second of the two RF models, we adjust the class weight parameter such that Type I and Type II error rates are equal; this has the effect of increasing the overall MCE relative to the RF1 model. It is important to bear in mind that while we report MCE rates for both the training and the test data, ultimately it is only the test data MCE rates that illustrate actual predictive utility. We also note that the CT models yield easily interpreted, easily applied, classification rules; the other models are of the "black box" variety requiring use of a computer to generate classifications and interpret classification rules. We use the measure of MCE rate to illustrate how often our models fail to correctly predict set association. The success rate of the models is simply 100 minus the MCE rate thus a 20% MCE rate can also be positively viewed as an 80% success rate.

To compare and contrast the relative importance of the predictor variables across models, we rescaled the Gini Importance Index values for each model such that the most important variable had a value of 1.0 and all other variables were computed as a proportion of the maximum Gini value for that model.

Following testing of the five models on the 2012 dataset, we applied the preferred model to the FAD-closure dataset. Concern over possible misclassification, either involuntary or deliberate, was viewed as reason to set these data aside during model development. We applied the classification models (using both tuna-only and with-bycatch variables) to the FAD closure data to determine if MCE rates were greater than for the non-FAD closure data.

3. Results

The modeling results are presented in pairs for each of the model types. The first of each model pair, termed "tuna-only", uses only tuna species composition and mean lengths to develop the classification models. The second of the model pairs, "with-bycatch", includes the bycatch species as possible classifying variables. Auxiliary predictors are included in all pairs of models except the initial CT1 models.

3.1. Classification tree model (CT1) using just observer-sampling data

The CT model developed from fitting to the 2007–2011 data, using only tuna composition and mean length data, is illustrated in Fig. 5 (left panel). This is the simplest of all models, but as all other models are generalizations or extensions of this basic model, we provide a fuller explanation of model interpretation. The first classification rule (SKJ.pct <99.8) divides the initial data set into two halves: sets with a skipjack composition of less than 99.8% and sets with composition greater than or equal to 99.8%. This implies that among the predictor variables, this partition point provided the highest initial rate of correct separation into "U" (unassociated) and "A" (associated) sets. Of course, there are instances of both set types above and below the classification rule, and additional rules are then added to attempt to better classify the two groups. The structure of the classification tree is such that any sets for which the answer to the condition is "yes" proceed to the left while those for which the answer is "no" proceed to the right.

Three lines of information are contained in each node. The first line is which set type has the majority of observations. The second line lists the number of "incorrect" observations over the total number of observations in that node. The third line lists the percentage of the total observations described in that node. Thus, the top node shows that a majority of the sets are "A" and that 13,086 are "incorrect" (in that they are actually "U") and the total number of sets is 34,598 (100% of non-FAD closure sets for 2007–2011). Sets for which the answer to the first condition was "no", are split off to the right and form a terminal node. This node is classified as "U"; there are 2561 incorrect classification out of 10,652 sets assigned to that node and these sets comprise 30.8% of all sets. On the basis of available predictor variables, there is no rule that can further refine those sets, subject to the complexity parameter and minimum branch size settings. Sets for which the answer to the first condition was "yes" are split to the left, where they are subjected to a second classification. This condition asks whether skipjack percentage in the catch is greater than or equal to 2.05%. If the answer is yes, those sets are sent to the left where they form a terminal node, classified as "A". Those sets for which skipjack percent was less than 2.05% proceed to the right and form a terminal node classified as "U". Multiple use of the same variable (such as SKJ.pct in this case) is not uncommon as more branches are developed in refinement of the classifications.

Each of the three terminal nodes has both correctly identified and incorrectly identified set types. The node classified as "A", has 3307 sets that were misclassified as "A"; these constitute the Type 1 error – 3307 out of 13,086 total "U" sets were misclassified (25.3%). The other two nodes, both classified as "U", had 502 and 2561 misclassified "A" sets; added together these for the Type II error – 3063 out of 21,512 "A" sets (14.2%). The overall error is then computed as the sum of all misclassified (3307 + 502 + 2,561) divided by total sets (34,598), for a MCE of 18.4%. We note that these are the MCE rates for the training data itself, not the test data to which these models are subsequently applied.

Fig. 5 (right panel) shows the classification model for 2007–2011 data when bycatch species are allowed as predictor variables. In
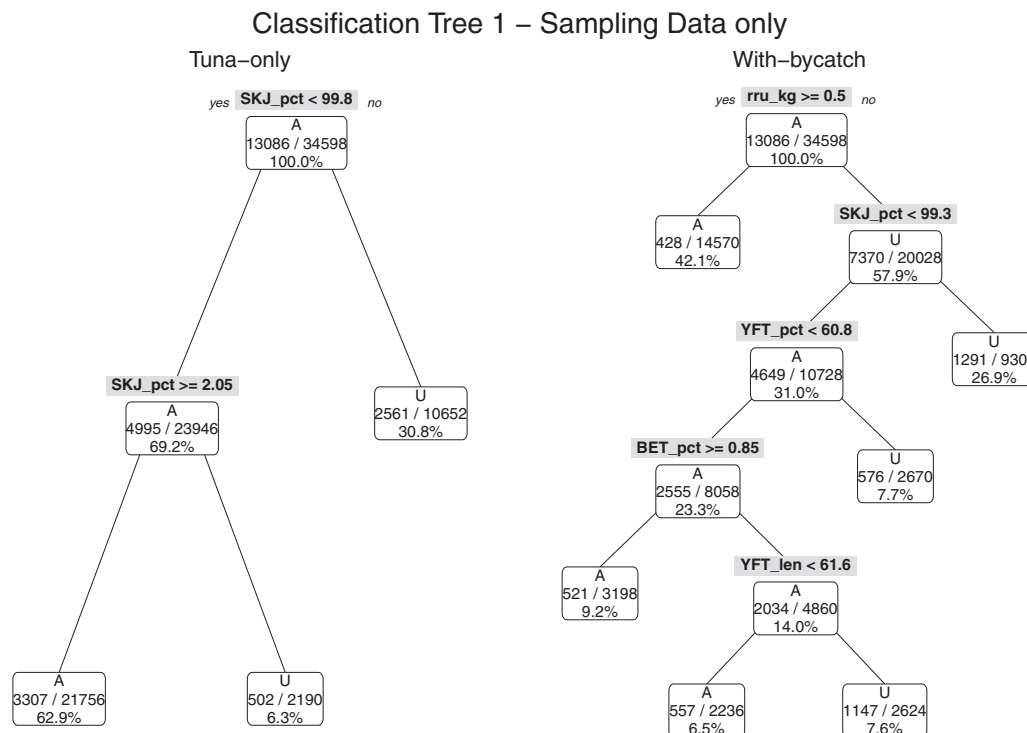
Classification Tree 1 – Sampling Data only

Tuna–only

With–bycatch

Fig. 5. Classification rules and node statistics for the two classification tree models developed from observer-sampling data for the period 2,007,011. Misclassification error rates are reported in Table 5.

Table 4
Comparison of misclassification error rates for model CT1 fit to 2007–2011 non-FAD closure data.

	Training data (2007–2011)			Test data (2012)		
	Type I	Type II	Overall	Type I	Type II	Overall
Tuna-only	25.3%	14.2%	18.4%	28.1%	17.8%	21.8%
With bycatch	11.5%	14.0%	13.1%	12.1%	16.0%	14.4%

this case, the mere presence of rainbow runner ("rru_kg", greater than or equal to 0.5 kg in a set) was the first classification rule. Sets for which this was true formed a terminal node with all sets all classified as "A". As can be seen in the node statistics, this is a powerful rule as there were only 428 "U" sets among the 14,570 for which this rule was true. To classify the 20,028 sets without rainbow runner in the catch as many four classification rules were required to predict set association. Sets that were almost pure skipjack (SKJ.pct >= 99.3% of catch composition) were classified as "U" while less pure sets where then classified according to percent of yellowfin, percent of bigeye and mean length of yellowfin, in the set.

Table 4 reports the MCE rates for the two models described above. We list the MCE rates for the training data, i.e., how well the model performed on the data used to fit the model, and then the error rates when the model fits are applied to the test data. Type I and II MCE rates for the test data are between 18 and 28% for models based solely on tuna catch, while MCE rates drop to around 12–16% when bycatch is included in the models. The inclusion of bycatch was especially effective in lowering Type I errors, reducing the MCE rate by an absolute amount of 16.0%, which corresponds to a 57% reduction in relative terms. Type II errors, however, increased by 1.8% (in absolute terms), which corresponds to an 11% increase in relative terms. The overall MCE rate decreased to 14.4% from 21.8%, a relative improvement of 34%. The tuna-only model has higher Type I and Type II errors; the with-bycatch model was opposite. The relative mix of error types is determined by the data

and which set type can be more readily predicted by the variables; the model seeks to minimize the overall MCE. We also note that there is no interaction between the two sets of models. Inclusion of bycatch can result in use or non-use of classification variables from the tuna-only model. A visual illustration of one form of variability in MCE rates is illustrated in Fig. 6. Within each 1° longitude strip we computed, for both the tuna-only and with-bycatch models, the proportion of correctly classified sets (unassociated classified as unassociated, associated classified as associated) and misclassified sets (Type 1 – unassociated misclassified as associated and Type 2 – associated misclassified as unassociated). Several features of the analysis are observed in the figure. The tuna-only model is characterized by high Type 1 (solid red line) MCE rates in the west and very low Type I MCE rates east of 155°E. Type II errors for the tuna-only model (solid yellow line) are much more consistent across the fishing range. There are multiple possible explanations for this result: for example, anchored FADs (as opposed to drifting FADs) are more commonly fished in this region and tuna stock assessments treat this region separately due to differing size compositions, thus the mix and size of tunas may contrast with the more generally derived classification rules. Differences in bycatch might assist in better defining FAD-association. The with-bycatch model greatly reduces the level of Type I errors (dashed red line) with little consistent effect on Type II errors. Even with the addition of bycatch as a predictor, there remains some spatial structure, in the latitudinal distribution of Type I MCE rates, which suggests that inclusion of a spatial variable may further lower the MCE rate. Besides the spatial structure, set type misclassification patterns might exist temporally, nationally, by EEZ, etc.

3.2. Classification tree model (CT2) with auxiliary predictors

A second set of CT models, which include the auxiliary variables, was fitted to the same training and test datasets and the results are illustrated in Fig. 7. MCE rates, for all five sets of models, for both
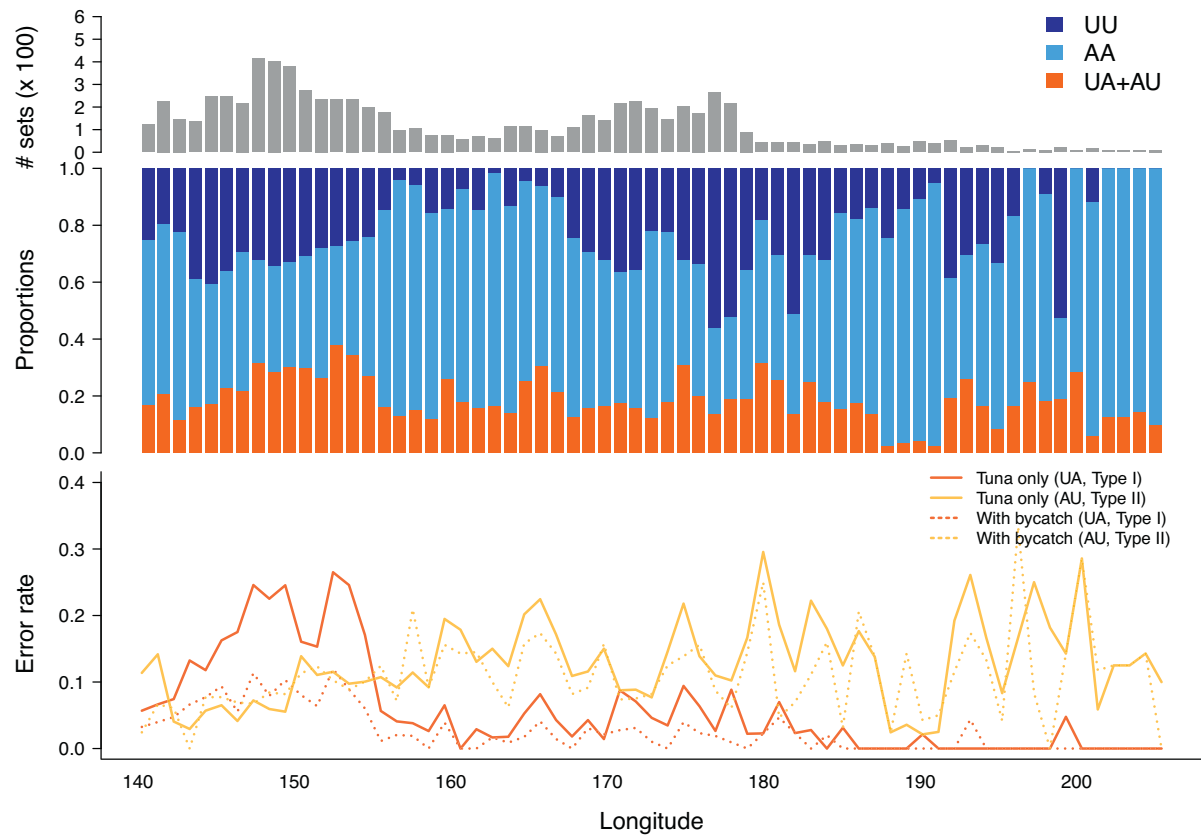
Fig. 6. Longitudinal distribution (summed within 1° longitude strips) of 2012 purse seine sets and classification results from CT1 models. Top panel illustrates total sets. Middle panel illustrates proportions of correct (UU and AA) and incorrect (UA + AU) classifications for the tuna-only model. The first letter is the observer recorded set type (U indicates unassociated, A indicates associated) and the second letter is the set type predicted by our model. The bottom plot shows the misclassification error (MCE) rates for Type I (UA) and Type II (AU) errors for the tuna-only (solid line) and with-bycatch (dashed line) models. The two solid lines sum to the height of the incorrect (UA + AU, in red) bars in the middle panel. See text for definition of MCE rate calculations.
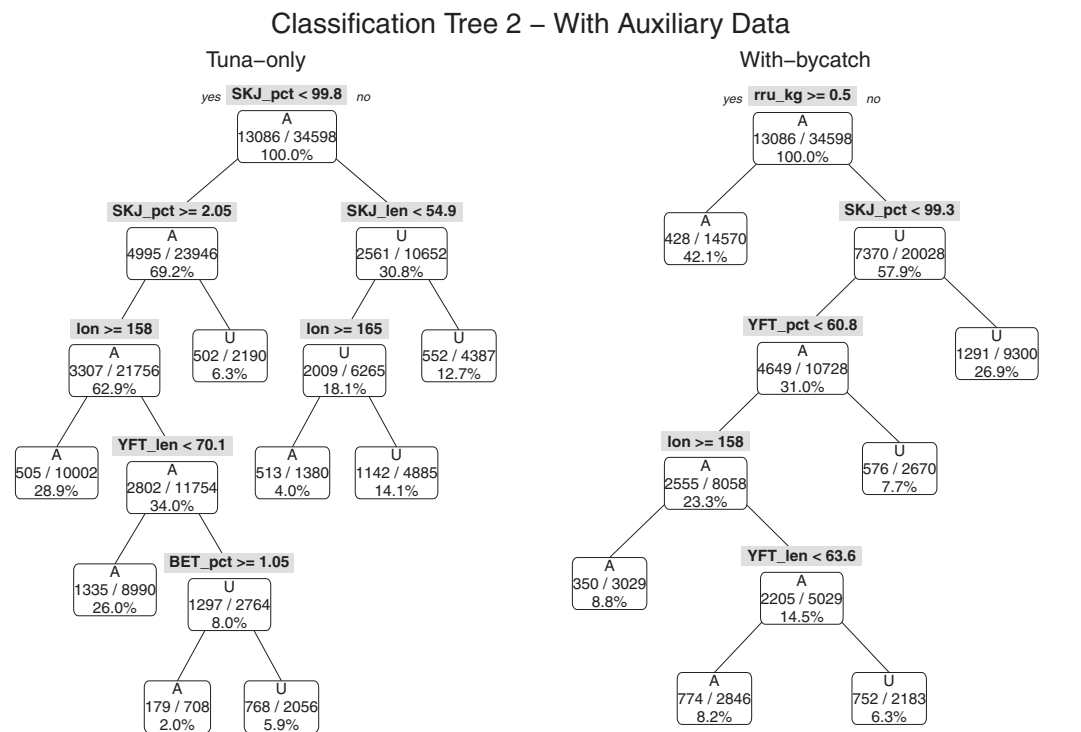

Fig. 7. Classification rules and node statistics for the two classification tree models developed from observer-sampling and auxiliary data for the period 2007–2011. Misclassification error rates are reported in Table 5.

Table 5
Comparison of misclassification error (MCE) rates among the five models.

Tuna-only models

	Training data (2007–2011)			Test data (2012)		
	Type I	Type II	Overall	Type I	Type II	Overall
CT1	25.3%	14.2%	18.4%	28.1%	17.8%	21.8%
CT2	19.3%	13.8%	15.9%	25.4%	14.4%	18.7%
BP	17.1%	9.7%	12.5%	38.3%	8.4%	20.1%
RF1	16.3%	9.6%	12.1%	48.4%	9.5%	24.8%
RF2	12.4%	12.3%	12.3%	41.2%	13.0%	24.0%

With-bycatch models

	Training data (2007–2011)			Test data (2012)		
	Type I	Type II	Overall	Type I	Type II	Overall
CT1	11.5%	14.0%	13.1%	12.1%	16.0%	14.4%
CT2	11.9%	12.2%	12.1%	12.9%	14.7%	14.0%
BP	11.9%	8.3%	9.7%	28.6%	8.1%	16.1%
RF1	10.4%	8.0%	8.9%	35.9%	8.8%	19.4%
RF2	8.9%	8.8%	8.9%	31.4%	10.1%	18.4%

Note: abbreviations are: CT1– classification tree with observer sampling data; CT2 – classification tree with sampling and auxiliary data; BP – bagging predictors with sampling and auxiliary data; RF1 – random forests with sampling and auxiliary data, unequal error rates; RF2 – random forests with sampling and auxiliary data, balanced error rates.

the training data and test data model fits are listed in Table 5. For the tuna-only data, there was an improvement of CT2 model fits over CT1 model fits, measured by overall MCE rate, for both the training and test data; overall MCE was lowered to 15.9% (from 18.4%) for the training data and from 21.8 to 18.7% for the test data. In both cases, both the Type I and Type II errors decreased. While the overall MCE rates for the with-bycatch model decreased for both the training (13.1–12.1%) and test data (14.4–14.0%), this was accomplished by lowering Type II errors, at the expense of slightly increasing Type I errors. The overall reduction in MCE was accompanied by the addition of longitude variables for both models and an increase in complexity for the tuna-only model. For the tuna-only model, the two rules from CT1 were retained, but a split on skipjack length (at 54.9 cm) joined the split on skipjack percentage (at 2.05). The skipjack length split was added to accommodate the subsequent longitude split (at 165°); it is only with the longitude data that the model finds support for the skipjack length split. Similar new branches for both models were built as a result of adding the auxiliary variables. Ultimately, the CT2 models performed the best among the five model types at classifying the test data.

3.3. Bagging predictors (BP)

Both BP models (tuna-only and with-bycatch) were fitted using the observer-sampling and auxiliary data predictors. The models showed a marked improvement in fitting the training data, relative to the CT1 and CT2 models, but did a poorer job than the CT2 models of predicting the test data (Table 5). The BP models produced near balanced Type I/Type II MCE rates for the training data but much more unbalanced MCE rates for the test data. Because the BP models generate bootstrapped datasets every time when run, there is a small amount of variance in both the training and test data MCE rates. This variance, while slightly larger for the test data MCE rates, was in all cases less than 0.1%. For the bootstrapped models, variable importance is illustrated using a rescaling of the Gini index (Fig. 8). This figure allows a visual and quantitative comparison of variable importance both across model types (simple classification models CT1/CT2 and bootstrapped models BP/RF1/RF2) as well as the tuna-only and with-bycatch variants. The most important variables in the BP models are generally the same as those seen in the two CT models, namely SKJ.pct, YFT.pct, YFT.len, and BET.pct and, in the case of the with-bycatch model, the presence of rainbow runner

Table 6
Comparison of model CT2 misclassification error rates (MCE) of purse seine sets during the FAD closure periods.

	Tuna-only			With bycatch		
	Type I	Type II	Overall	Type I	Type II	Overall
2009	34.3%	9.3%	25.7%	13.9%	11.5%	13.1%
2010	14.0%	35.0%	16.0%	7.1%	40.7%	10.3%
2011	19.2%	21.5%	19.9%	11.8%	23.1%	14.9%
2012	42.6%	16.8%	35.7%	15.9%	19.6%	16.9%
All years	25.9%	19.2%	24.4%	11.6%	21.8%	13.9%

(rru_kg). Longitude is the most important auxiliary variable for both models as well.

3.4. Random forest models, RF1 (minimum error rates) and RF2 (equal error rates)

The two RF models were fitted using the same data as the BP and CT2 models with the exception that the three average length variables were not included. Despite this exclusion, the RF1 models provided the best fit, in terms of overall MCE rate, to the training data, with rates of 12.1% and 8.9% for the tuna-only and with-bycatch models, respectively. The RF2 model, with case weights adjusted to give equal Type I and Type II error rates, performed only slightly worse than the RF1 model, in terms of training data MCE rates. However, when applied to the test data, the two sets of RF models performed more poorly than the three other model types, in that they had the highest overall MCE rates. In terms of predicting the test data, the best performing model was the with-bycatch RF2 model with an overall MCE rate of 18.4%; the worst performing model was the tuna-only RF1 model with an overall MCE rate of 24.8%. The Type I and Type II MCE rates for the test data sets were extremely unbalanced despite being nearly balanced (in the RF1 models) and evenly balanced (in the RF2 models). Perhaps as a result of not having access to mean length data, the ranking of important variables for the RF models was quite different than for the CT and BP models (Fig. 8).

3.5. Prediction of FAD closure sets

The model fits discussed above were restricted to purse seine sets made outside the FAD closure period, first instituted in 2009. On the basis of the results presented above, we have selected model CT2 as our preferred model for classifying purse seine sets in our dataset. We next apply this model to the FAD-closure dataset to determine if the model predicts these data as accurately, less accurately, or differently than it classifies the non-FAD closure data. We apply the model, using both the tuna-only and the with-bycatch variants, to each year of the FAD-closure data as well as to the four years (2009–12) collectively. In general, the results (Table 6) do not indicate a substantially higher MCE rate for the FAD-closure data. The overall MCE for the tuna-only model is 33% higher than for the non-FAD closure data, but there is high interannual variability; the with-bycatch MCE rate – for the four years combined – is actually slightly lower than for the non-FAD closure data. Interannual variability in MCE rates is considerably lower for the with-bycatch model.

4. Discussion

The original goal of this analysis was to determine what level of correct purse seine set association could be achieved with access to sampling of individual sets. Our results suggest that, given access to observer type sampling of sets, simple classification models, based only on tuna catch, could provide up to 78% accurate classification.
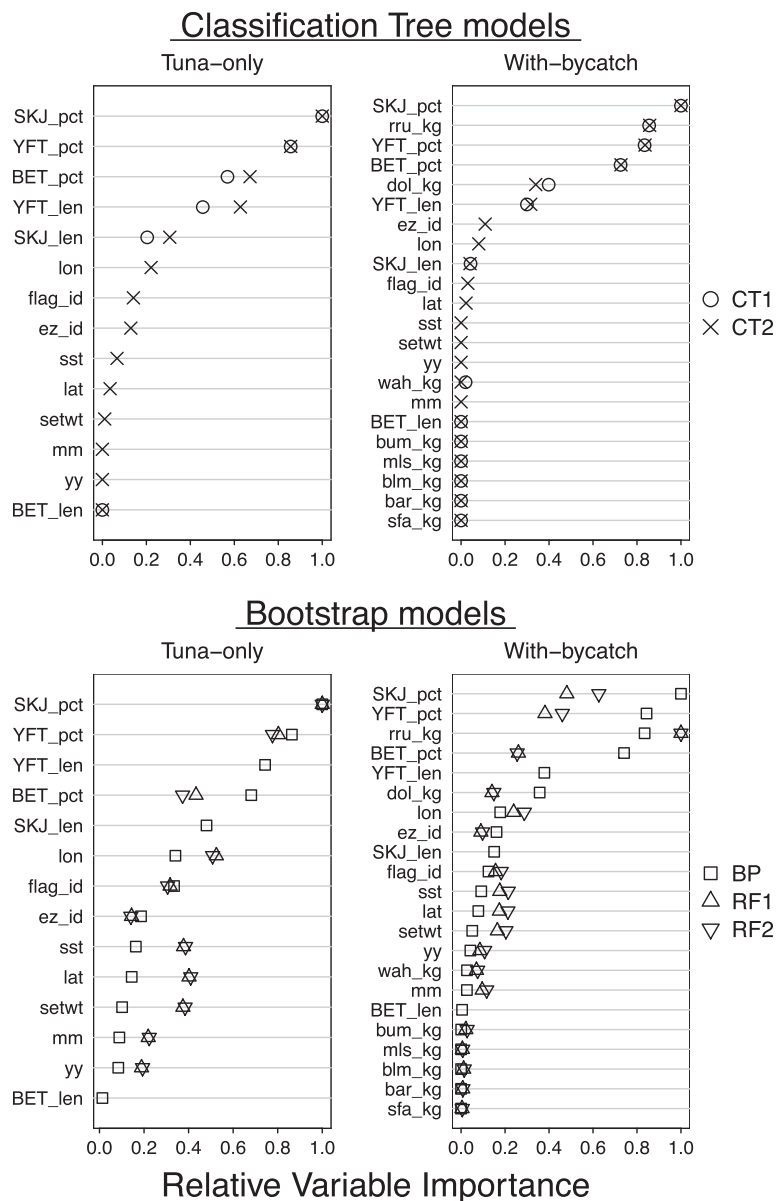
Fig. 8. Ranking of importance of classification variables, measured as sum of all decreases in Gini impurity, for each of the five sets of models. The two classification tree (CT) models are plotted in the upper panels, the three bootstrapping models (BP, RF1, RF2) are plotted in the lower panels.

If bycatch data were available, up to 85% accurate classification might be possible. Additional auxiliary variables to the CT models marginally reduced MCE rates for the tuna-Only from 21.8 to 18.7%, while with-bycatch model MCE rates decreased from 14.4 to 14.0%. The use of bagging predictors produced lower MCE rates for the training data, but performed about as well as the CT1 model, and worse than the CT2 model, on classifying the test data. The two RF models, which did not utilize the mean length predictors, performed very well on the training datasets; when applied to the test data however, the predictions were poorer than the CT and BP models. On the basis of these results we conclude that, for the types of data tested, the use of auxiliary predictors provides marginal improvement and resampling techniques provide no improvement over the simpler (to fit and interpret) CT models based solely on observer sampling data.

Inclusion of bycatch, specifically rainbow runner, as a predictor variable greatly improves model classification of set type. This is true not only for training data sets, but also for the test data sets. The classification rule for rainbow runner is literally pres-

ence/absence. As the smallest possible recorded amount of rainbow runner bycatch for any set in the database is 1.0 kg (data are recoded as metric tons to three decimal places), and the classification rule for rainbow runner is always set at 0.5 kg, the models are using presence of rainbow runner as the strongest indicator of FAD-association. Once a set has been classified as associated on the basis of rainbow runner presence, no models include additional steps to further separate those sets indicating none of the other variables contain predictive power. While other bycatch species show similar levels of discrepancy in mean catch per set type, rainbow runner appears more than twice as frequently in sets as the second most common bycatch species (dolphin fish) and several times more frequently than any of the others. Within unassociated sets, two other bycatch species – blue marlin and black marlin – occurred more frequently than rainbow runner.

Another advantage that derives from the use of bycatch in the classification models is the sharp reduction in Type I errors, i.e., misclassification of unassociated sets as associated. One type of set that is typically misclassified in the tuna-only models is a free school set

containing a large fraction of small yellowfin or bigeye tuna. The lack of bycatch (specifically rainbow runner) in such sets prevents these sets from being classified as associated.

With the exception of the bootstrapping model RF2, our models seek to minimize overall MCE rate on the training data set. A potential criticism of this approach is that the cost of misclassification errors is unequal depending on one's perspective. From a conservation, or sustainable seafood purchasing, perspective it is undesirable to have high Type II errors: the situation when a FAD-associated set is misclassified (accepted) as being unassociated. Similarly, from a fisher's perspective, a Type I error: when a FAD-free set is misclassified as being FAD-associated, is highly undesirable. In this analysis, we do not consider the societal or conservation costs of one error type over the other, opting only to determine how well we can predict set association at different levels of data quantity and model complexity. As noted, we did attempt to equalize the two error types (model RF2) as this is an option for random forest models. As was the case with the BP and RF1 models, the classification of the training data was quite good – substantially superior to the two CT models. However, when the RF2 model was applied to the test data, it performed substantially worse than the CT models. In fact, all three bootstrap models had much more unbalanced error type ratios than the CT models, both for the Tuna-only and With-bycatch datasets. This may occur from model overfitting and/or the non-availability of the mean length data to the RF models. Constructing a forest of models, all based on various subsets of the data and predictor variables increases the "importance" of several variables that have no predictive value for the two CT models (see Fig. 8). The CT models maintained relatively balanced error rate types for both the training and test data sets. Given its minimum data requirement, Tuna-only CT1 model (i.e., without auxiliary data) which included bycatch was very nearly the best overall model, with both error types within 2% (in absolute terms) of the overall MCE. The addition of auxiliary variables (i.e., model CT2), did lower overall MCE, and slightly lower the discrepancy in error types, but the improvement was minor compared to the effect of including bycatch.

A key question is then how accurately can observers estimate bycatch. The bycatch may be discarded or quickly set aside for consumption by the crew. We have chosen 'edible' bycatch species here as we believe that they are most easy to observe, and most likely to relate to school association. Without exception, all the bycatch-free models had skipjack percentage as the first-order classification rule, with pure sets (SKJ.pct >99.5%) classified as unassociated sets. However, none of these bycatch-free models had MCE rates as low as the with-bycatch models for the test data sets.

Regarding the analysis of the FAD-closure dataset, some observations from the model results bear mentioning. Using the CT2 model as our best performing and, therefore "preferred" model, the overall MCE rates for the years of 2009–12 combined were quite similar in magnitude to the MCE rates of sets from the non-FAD closure period. In fact, the with-bycatch model MCE rates were a bit lower. There was interannual variability in all three error type but it was not biased toward either Type I or Type II error. This exercise provides some reassurance in regards to observer classification of purse seine set association during the period of FAD closure.

One other point bears mentioning in regards to possible increased confidence in sampling purse seine catches to identify set type. The vast majority (>99%) of all sampled sets were sampled using the "grab sample" method. Essentially, an observer is instructed to "grab" a sample of fish, striving for representativeness, for each set. The observer grab sample is generally 100 fish or less: just 18% of the observed sets were sampled for more than 100 fish, less than 1.5% were sampled for more than 300 fish. Mean grab sample size across all sets is 65 fish. Both species composition and mean length estimates are based on these samples. Thus,

catch composition – especially for the rarer yellowfin and bigeye species – is only roughly estimated (this is likely less an issue with estimated mean lengths) at the set level. Given the low intensity sampling of purse seine catches, catch purity is an issue: sometimes the presence of a single non-skipjack tuna or bycatch species, is sufficient for sets to be classified as associated. The move toward "spill sampling" (Lawson, 2011), where a smaller number of larger samples are taken from a purse seine set, is one potential improvement in this regard. Spill samples are 'spilled' into a bin rather than hand selected and are designed to overcome fish selection bias. The overall larger sample sizes and reduced bias may well increase the precision of models developed to classify set type.

Finally, while our focus was limited to consideration of catch sampling data, there are a number of non-sampling characteristics that might potentially improve set classification. e.g., vessel flag, EEZ, and set time. Harley et al. (2009) demonstrated that time of day is a possibly important distinguishing characteristic between set types. Historically, associated sets occurred pre-dawn and unassociated sets occurred throughout the day. It is generally believed that unassociated sets cannot occur during darkness (light is needed to find and encircle the fish), but associated sets theoretically could occur at any time of day. Therefore, time of day is probably best for excluding pre-dawn unassociated sets rather than assisting in classifying unassociated sets. We did not explicitly consider time of day in this analysis, but intend to further pursue this factor in future work on this subject.

There are several operational activities that could serve to make classification rates reported in this analysis unreliable and overly optimistic. The most significant would be the failure to accurately record bycatch if bycatch-included models were applied. "Clean" skipjack sets are, almost without exception, classified as unassociated sets. Sets that are not "pure" skipjack but which have very high levels of either yellowfin or bigeye (in essence, a different form of a "clean" set) are also typically classified as unassociated. Accurate observer recording of observed bycatch is a high priority duty of observers; the concern here is intentional hiding of bycatch by vessel personnel. Further, interference with sampling protocol to bias sampling toward particular species of tuna is another means of influencing determination of set type. A second-order possibility, if classification rules were "known", would be to manipulate mean size, particularly of yellowfin tuna: large yellowfin almost always come from unassociated sets while small yellowfin can come from either set type. Finally, we note that observers might be unaware of a purse seine set on an incidental (piece of rope, plastic bag) or natural (whale shark) FAD, thus misclassifying such sets as unassociated. The effect of this form of misclassification results in an increased Type I (as well as overall) error rate. In this regard, our error rates would be reduced if such misclassifications were corrected.

Acknowledgments

References

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth and Brooks/Cole, Monterey.
Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88 (11), 2783–2792.

Dagorn, L., Holland, K.N., Restrepo, V., Moreno, G., 2012. Is it good or bad to fish with FADs? What are the real impacts of the use of drifting FADs on pelagic marine ecosystems? Fish Fish. 14 (3), 391–415.

Fonteneau, A., Chassot, E., Ortega-Garcia, S., Delgado de Molina, A., Bez, N., 2010. On the use of the De Finetti ternary diagrams to show the species composition of free and FAD associated tuna schools in the Atlantic and Indian oceans, in: tropical tunas. Coll. Vol. Sci. Pap., 546–555.

Fonteneau, A., Chassot, E., Bodin, N., 2013. Global spatio-temporal patterns in tropical tuna purse seine fisheries on drifting fish aggregating devices (DFADs): taking a historical perspective to inform current challenges. Aquat. Living Resour. 26, 37–48, http://dx.doi.org/10.1051/alr/2013046

Harley, S., Williams, P., Hampton, J., 2009. Analysis of purse seine set times for different school associations: a further tool to assist in compliance with FAD closures? WCPFC-SC5-2009/ST- WP-07 (available at http://www.wcpfc.int/node/2126).

Harley, S., Williams, P., Nicol, S., and Hampton, J., 2014. The western and central pacific tuna fishery: 2012 overview and status of stocks. Tuna fisheries Assessment Report No. 13, Secretariat of the Pacific Community, Oceanic Fisheries Programme. 31 pp. Available at http://www.spc.int/OceanFish/en/publications/doc_download/1205-tuna-fisheries-assessment-report-no-13).

Lennert-Cody, C.E., Rusin, J.D., Maunder, M.N., Everett, E.H., Largacha Delgado, E.D., Tomlinson, P.K., 2013. Studying small purse-seine vessel fishing behavior with tuna catch data: implications for eastern Pacific Ocean dolphin conservation. Mar. Mamm. Sci. 29, 643–668.

Lawson, T., 2011. Purse-seine length frequencies corrected for selectivity bias in grab samples collected by observers. WCPFC-SC7-2011/ST-IP-02. Available at http://www.spc.int/DigitalLibrary/Doc/FAME/Meetings/WCPFC/SC7/ST-IP-02pdf

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news, The newsletter of the R project, Vol. 2/3: 18–22. Available at: http://cran.r-project.org/doc/Rnews/Rnews_2002-3pdf

Liaw, A., Wiener, M., 2015. randomForest: Breiman and Cutler's random forests for classification and regression. R package version 4. 6–10. http://cran.r-project.org/web/packages/randomForest/index.html

Nicodemus, K.K., 2011. Letter to the editor: on the stability and ranking of predictors from random forest variable importance measures. Brief Bioinform. 12 (July (4)), 369–373.

Pallarés, P., Nordstrom, V., Fonteneau, A., Delgado de Molina, A., Ariz, J., 2003. Definition of criteria to identify FAD and free school sets based on the species composition and average weight of the samples from the Indian Ocean European fleet of purse seiners. IOTC Proc. 6, 256–263.

Peters, A., Hothorn, T., Ripley, B.D., Therneau, T., Atkinson, B., 2015. ipred: improved Predictors. R package version 0. 9–4. http://cran.r-project.org/web/packages/ipred/index.html

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., Wang, W., 2002. An improved in situ and satellite SST analysis for climate. J. Clim. 15, 1609–1625.

R Core Team, 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Therneau, T., Atkinson, B., Ripley, B. (2013). rpart: Recursive Partitioning. R package version 4. 1–3. http://CRAN.R-project.org/package=rpart

WCPFC, 2009. Conservation and Management Measures for Bigeye and Yellowfin Tuna in the Western and Central Pacific Ocean. CMM 2008-01. Available at http://www.wcpfc.int/doc/cmm-2008-01/conservation-and-management-measure-bigeye-and-yellowfin-tuna-western-and-central

WCPFC., 2013. Conservation and Management Measures for Bigeye, Yellowfin and Skipjack.CMM 2012-01. Available at http://www.wcpfc.int/doc/cmm-2012-01/conservation-and-management-measure-bigeye-yellowfin-and-skipjack

Williams, P., 2014. Scientific data available to the Western and Pacific Fisheries Commission.WCPFC-SC10—2014/ST WP-1. Available at: http://www.wcpfc.int/node/18878

Yi, Q., 2012. Random forest for Bioinformatics. Ensemble Machine Learning: Methods and Applications. Springer-Verlag, New York, pp. 307–323 http://dx.doi.org/10.1007/978-1-4419-9326-7