SCIENTIFIC COMMITTEE
TENTH REGULAR SESSION

Majuro, Republic of the Marshall Islands
6-14 August 2014

On the potential of identifying fad-association in purse seine catches on the basis of catch sampling

WCPFC-SC10-2014/ST-WP-04 Rev 1

18 July 2014)

S.R. Hare, S.J. Harley and J. Hampton

Oceanic Fisheries Programme (OFP)
Secretariat of the Pacific Community, Noumea, New Caledonia

Rev 1 – 18 July 2014: correction to equation in Table 5.

On the potential of identifying FAD-association in purse seine catches on the basis of catch sampling

S.R. Hare, S. J. Harley and J. Hampton

Oceanic Fisheries Programme, SPC

Abstract

In this study, we investigate the potential of accurately identifying whether individual purse seine sets can be identified as captured in association with a Fish Aggregation Device (FAD) or as an unassociated (FAD-free) set, on the basis of catch sampling. The target tuna catch and length compositions and bycatch amounts were analyzed from more than 50,000 purse seine sets sampled by on board observers who had, in addition to collecting the sampling data, also identified the sets as either "associated" or "unassociated". The tuna data are derived from observer "grab samples" which are, on average, number about 65 fish per purse seine set. Bycatch data are estimated total amounts per set and are not determined by standard sampling.

Methods from the general category of Classification and Regression Tree (CART) modelling were determined most appropriate for the analysis and intended use of results. An attraction of the simplest of the CART methods is that it lends itself to establishing a set of clearly-labelled rules that can be routinely used to estimate whether a sampled purse seine set was likely an associated or unassociated set type. Classification models were developed based on 2007-2011 observer data and tested for misclassification error rates on 2012 data. Models were developed for the full dataset as well as seasonal and regional breakdowns. Two sets of models were developed for each analysis – "tuna-only" and "with bycatch", the difference being the allowance of bycatch species as potential classification variables.

Two types of misclassification errors (MCE) are possible: unassociated sets misidentified as associated (termed false positive or Type I) and associated sets identified as associated sets (false negative or Type-II error). A third error measure, overall MCE, is a weighted average of Type I and Type II error. While all three error types are of interest, the Type II error rate is of most concern in a conservation context. The initial tuna-only CART models had MCE rates of 17-29% with an average of 23%. Inclusion of bycatch lowered error rates by 4-12% to around 14-20% with an average of 16.5%. The appearance of a spatial pattern in the MCE rates, with higher Type I error rates in the west, motivated exploration of MCE rate improvement by analyzing seasonal and regional data subsets. Disaggregating the data by season or region generally yielded modest improvement in classification accuracy, decreasing relative MCE rates 2-10%. An exceptional classification result was achieved in an eastern region bycatch model where MCEs rate of less than 10% were achieved.

An extension to the CART methodology, termed "Bagging Predictors", which employs bootstrap sampling to create multiple classification tree models, was investigated to see if MCE rates could be furthered lowered. The downside to this method is that it is not "field applicable" and requires use

of an interactive computer program. We found that the computer intensive bagging method provided an overall 8-18% decrease in MCE rates, a marginal level of improvement over the much simpler methods. Further, the decrease in MCE rates was not uniform across seasons or regions.

We conducted an analysis on a particular subset of the purse seine data, i.e., sets classified as unassociated during the FAD-closure periods of 2009-2012. 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, MCE rates of unassociated sets during the FAD closure period were found to be equal, or even a bit lower than, MCE rates in the broader analyses.

We conclude with a general discussion of potential operational practices that would help achieve, or make especially challenging, classification results equal or better than those we obtained. Specifically, removal/onboard consumption of bycatch and/or mixing of sets prior to sampling would contaminate individual sets, which formed the basis of our model classification rules. Treatment of bycatch varies across time, fleets, and unloading ports. The classification rules that included bycatch indicated that perhaps as little as a single fish in a set (that might contain over 30 mt of total tuna and bycatch) would be sufficient to have a set assigned as associated. This has implications for any independent sampling scheme.

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)*, yellowfin (*Thunnus albacares*) and bigeye (*Thunnus obesus*) tunas, has grown substantially over the past three decades in the Western and Central Pacific (WCP), increasing from around 100,000 mt in 1980 to nearly 1.8 million mt in 2012 (Harley et al. 2014). 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.

Concerns over the composition of catches associated with FAD-fishing have led to recent calls to regulate FAD-fishing, either via regulatory actions (Fontenau et al. 2013), or educating consumers (WWF 2011[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 general, FAD-associated catches contain 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, August and September, has been instituted annually since 2009 by the Western and Central Pacific Fisheries Commission, the international body responsible for management of the WCP tuna fisheries.

All purse seine vessels operating in territorial waters of nations within the WCP 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. Such a determination might be useful both in retrospective analyses – for example, historical purse seine sets from vessels not carrying an observer and, for future use, in case of observer absence or verification of observer set type determination to satisfy FAD-free adherence concerns. It is this final point that may be of most concern going forward. Fishing on FADs is, generally,

[1] http://awsassets.panda.org/downloads/tuna_fad_position_november_2011_.pdf
[2] e.g., http://iga.com.au/support/about-iga/sustainability/

speaking a more dependable method of locating and catching tuna, however insistence on FAD-free fish, and consumer willingness to pay a premium, has the potential to create an incentive to misreport set type. Additionally, 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.

In this report, we investigate the potential of using observer sampling of purse seine catches to determine set association. The method looks for consistent differences in the relative species composition and mean length of the tuna catches and, optionally, the amount and species of bycatch present in the set. Observer data are used to "train" the models and these models are then applied to test data that were not used in model fitting. 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 was collected during periods when there was little incentive to misreport set association, and the number of observed sets is in the tens of thousands which would tend to override contamination from, presumably, a small number of irregular reports. The initial methodology attempts to develop classification methods that could be deployed "in the field", i.e., utilizes simple rules. We then extend the methodology allowing for more complex models that would require use of an interactive computer program. Finally, we use the methodology to more closely examine purse seine sets during the FAD-closure periods.

Materials and methods

The data used in this analysis come from the Secretariat of the Pacific Community (SPC) maintained observer database which contains observations on purse seine operations dating from 1993 to the present. The database from which these data were extracted represents a filtered, quality-controlled, subset of the total database. Additionally, this analysis is restricted to observed sets with both 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 (Figure 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 as less than 2000 observed sets have been entered into the SPC observer database and the representativeness of this data is unknown. Over the past six years there have been roughly similar numbers of FAD-Free ("unassociated") and FAD-Associated ("associated") observed 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."[3] The proportion of unassociated sets has generally increased since 2010, coinciding with the implementation of the FAD-closure period within the WCP between July and September (with certain exceptions).

The observer data used in the analysis were collected using a method termed "grab sampling" which has been consistently utilized dating to the beginning of onboard purse seine set sampling. In essence, the observer is instructed to randomly collect five tuna from each braille used to empty the purse seine net. Mean grab sample size from each set is 65 fish though variability in sample size

[3] WCPFC. 2013. Conservation and Management Measures for Bigeye, Yellowfin and Skipjack. CMM 2012-01. Available at http://wcpfc.int/system/files/CMM-2012-01-Conservation-and-Management-Measure-BET-YFT-and-SKJ.pdf

is very large, consistent with the nature of purse seine set catch sizes. More recently "spill sampling" has been gradually introduced as a means of reducing potential biases associated with grab sampling. In the assembled dataset, a total of 306 sets contained both grab and spill samples for species and length compositions; only the grab sample data have been utilized. It is anticipated that any sampling scheme devised to classify unassociated and associated sets on the basis of catch sampling will also use grab samples thus the data summaries and classification rules are all based on grab samples. Spill sampling has become of increased importance of late, signified by the 560 spill samples already collected among the 1888 observed and recorded sets in 2013. The potential utility of spill sampling may warrant future investigation as a means of classifying purse seine set type. Bycatch data are not subsampled; observers utilize a variety of means of estimating full set weights of all non-target tuna species.

Potential covariates, or predictor variables, for classifying set type were 1) tuna species composition; 2) various measures of tuna length; and 3) species bycatch per set.

Tuna species composition

In Figure 2, the relative proportions, of the three target tuna species, within associated and unassociated purse seine sets are presented in ternay, or De Finetti (Fonteneau et al. 2010), plots. These plots illustrate that both associated and unassociated sets are most often comprised of 90+% skipjack. However, a couple of 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 sets. 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 a purse seine set.

Tuna length composition

The three target tuna species captured in unassociated sets tend to have a larger size distribution than those in associated sets (Figure 3). Juvenile yellowfin tuna (~ 50cm and smaller), in particular, are not commonly caught in unassociated sets but form the bulk of yellowfin catch in associated sets. We computed mean tuna species length for each set in which any of the three target tuna 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. Figure 4 shows a boxplot of the differences in mean length distribution between set types and these mean lengths are used in the classification analysis. Similar to the naming convention described above for catch composition, these variables are abbreviated SKJ.len, YFT.len, and BET.len where "len" is interpreted as mean length.

Bycatch composition

Bycatch data, estimated total weight per set, was 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). Bycatch species name abbreviations used for naming conventions in the results section are listed in Table 2, followed by "kg", indicating total weight in kg in a set. An examination of the fate of these species indicated that 20-60% of 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. This information is potentially important in devising a port catch sampling scheme. 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 unrealistic for set type determination when port sampling.

Two other potential covariates were also examined – Season (months 1-3, 4-6, 7-9, 10-12) and Region (west of 160°W, 160°W – 180°, east of 180°). Initial tests indicated little predictive power of these covariates across the entire dataset. Subsequently we investigated whether seasonal or regional data subsets yielded different, or improved, classification rules. The breakdown of set association by season and region is illustrated in Table 3.

We extended our analysis to focus on a subset of the observer data specifically related to the recent imposition of "FAD-closure" periods. Beginning with the months of August and September in 2009 and then extended to include July beginning in 2010, fishing on FADs has generally been disallowed, with certain exceptions for archipelagic waters and Pacific Island states. Table 4 illustrates the pronounced shift in unassociated to associated set ratio relative to the overall 2007-2012 period. The question we addressed was whether there is anything "unusual" about the purse seine sets specified as unassociated taken during the FAD-closure period. The premise being that an observer might routinely mark all purse seine sets during the FAD-closure period as unassociated but the fishing vessel might surreptitiously set upon a FAD. Specifically, we wished to examine how these sets would be classified using classification models that excluded the unassociated data from model development. A data subset, hereafter the "FAD-Closure Unassociated Kept aside" (FCUK for short) data, was created. The FCUK data contain all observer sets identified as unassociated from months 7 and 8 in 2009 and months 7, 8, and 9 for the years 2010-12. By subtraction, a FCUK-less data set contains the remainder of the original data... A number of different time periods were used to construct and test the classification rules, and misclassification error rates were compared to overall misclassification rates from the full data analysis.

To determine the appropriate statistical technique for this analysis, a literature review was conducted to ascertain previous work in this area. Specifically regarding identification of purse seine set type association using catch data, there appears to be just a few previous analyses. 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, highly relevant 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.

For this analysis, we settled upon the method of Classification and Regression Tree (CART) modelling. Briefly, CART modelling is a means of variable selection with the attractive feature of clearly illustrating the decisions made to classify a variable among a set of discrete choices. 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 clearly-labelled rules that can be used to estimate whether a sampled purse seine set is 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. All data analyses conducted herein were based on the R Programming language (R Core Team 2013) and model fits used the "rpart" package (Therneau et al. 2013). The two main model fitting control parameters in *rpart* are the "complexity parameter" (cp) and "minimum branch size" (minsplit). For all CART model fits, the settings for these two parameters were: cp=0.01 and minsplit=30.

Following the initial CART modelling, we then utilized a more complex methodology that uses bootstrapping techniques to see if misclassification rates could be improved upon. The downside is that the models are not "field applicable" and require use of a computer to interactively determine set association. The generic name for the resampling methodology is Bagging Predictors (Breiman 1996). A concise summary of the methodology is provided in the abstract to the original paper describing the technique:

> "Bagging predictors is a method for generating multiple versions of a predictor and using these to get an aggregated predictor. The aggregation averages over the versions when predicting a numerical outcome and does a plurality vote when predicting a class. The multiple versions are formed by making bootstrap replicates of the learning set and using these as new learning sets. Tests on real and simulated data sets using classification and regression trees and subset selection in linear regression show that bagging can give substantial gains in accuracy. 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."

A further strength of this method is that it allows for missing data. A related, potentially superior, method also developed by Breiman (2001) called Random Forests is inapplicable to the situation at hand due to the surfeit of missing data in our dataset. We note that there is substantial "missing data" in the sense that classification rules may be based on factors such as average length of a particular tuna species but if no such tuna were captured in a set, then such data are "missing."

The technique of bagging predictors has 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 which provided a balance of complexity and close to best predictive power. Bootstrap samples were of size n out of n with replacement; 30 trees were constructed, minimum branch size was set to 100, and the complexity parameter was set to 0, meaning that any data split which increases overall data fit (subject to other parameter settings) is pursued.

Both the CART, as well as the bagging predictor, models are fitted such that the overall misclassification rate is minimized for the training data set. This overall misclassification error (MCE) rate is a mix of two types of misclassifications, which are referred to as Type I and Type II errors. In this analysis, the interpretation of the two types of errors is as described in Table 5. The overall misclassification error is a weighted average of the two error types, thus always falls between the two. In general, Type II errors are of more interest due to concern over potentially accepting fish caught in association with a FAD but labelled as unassociated. In general, we report only the Type I and Type II MCE rates. It is important to bear in mind that while we report MCE rates for both the training and the test data, ultimately it only the test data MCE rates that illustrate potential predictive utility.

Finally, to perhaps state the obvious: 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 error rate can also be positively viewed as an 80% success rate.

Results

The modeling results are presented in pairs for each of the various datasets, regions, and seasons. The first of each model pair, termed “tuna-only”, uses only tuna species and mean length to develop the classification trees. The second of the model pairs, “with-bycatch”, includes the bycatch species as possible classifying variables.

Classification models without seasonal or regional breakdown.

The CART (hereafter, generally referred to simply as “classification tree”) model developed from fitting to the 2007-2011 data, using only tuna composition and mean length data, is illustrated in Figure 5. As this is one of the simplest of all models, and as a means of illustrating interpretation of how the following models might be used “in the field”, a fuller explanation of this model is presented. 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 percent and sets with composition greater than or equal to 99.8 percent. 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” (Associated) and that 18442 are “incorrect” (in that they are actually “U”) and the total number of sets is 41364 (100% of 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 2683 incorrect classification out of 14111 sets assigned to that node and these sets comprise 34.1% of all sets. On the basis of available predictor variables, there is no rule which 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 nodes has both correctly identified and incorrectly identified set types. The node classified as “A”, has 4460 sets that were misclassified as “A”; these constitute the Type 1 error – 4460 out of 18442 total “U” sets were misclassified. The other two nodes, both classified as “U”, had 592 and 2683 misclassified “A” sets; added together these for the Type II error – 3275 out of 22922 “A” sets. We note that these are the MCE rates for the training data itself, not the test data to which these models are subsequently applied.

Figure 6 shows the classification model for 2007-2011 data when bycatch species are allowed as predictor variables. In this case, the mere presence of rainbow runner (“rru”, 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 528 “U” sets among the 15395 for which this rule was true. To classify the 18442 sets without rainbow runner in the catch as many three classification rules were required to estimate set association. Sets that were almost pure skipjack (SKJ.pct >= 99.4% of catch composition) were classified as "U" while less pure sets where then classified according to mean length of yellowfin and, potentially, percent of bigeye in the set.

One of the classification rules for this model, YFT.len < 72.4 (mean length of yellowfin tuna less than 72.4 cm), presents a special situation that can arise when the dataset contains "NA" values, i.e., not applicable/available. For there to be a non-NA YFT.len value, there had to have been yellowfin tuna in the set. Given the structure of this particular model, it is possible for a set to arrive at this classification branch and contain no yellowfin: for example, if a set had no rainbow runner and was less than 99.4% skipjack but all the non-skipjack catch was bigeye – that set would then contain an "NA" for the YFT.len field. In fact, for the 2007-2001 data, there are 15 cases where a set with YFT.len has a field entry of "NA". The classification algorithm handles such cases in the following manner. The "NA" cases are all assigned the set-association designation that forms the majority of sets at that particular node. Thus, while "A" sets barely outnumber "U" sets, 6665 to 6589, those 15 sets are all classified as "A" and are not subjected to any further classification rules.

Table 6 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 are between 17 and 29% for models based solely on tuna catch, while MCE rates drop substantially, to around 14-20% when bycatch is included in the models. Thus, the inclusion of bycatch reduced Type II MCE rate by 3.6% (in absolute terms), which corresponds to a 20% reduction in relative terms. The overall MCE rate was substantially improved by the addition of bycatch, decreasing to 16.5% from 23.0%, a relative improvement of 28%. For both models, the Type I errors ("U" sets classified as "A" sets) were higher than the Type II error rates. This is not, however, a consistent feature of this type of modelling.

To determine if there were any spatial patterns in the misclassification rates, we aggregated the 2012 sets longitudinally by one degree strips. Within each 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). The results are illustrated in Figure 7. Several interesting features of the analysis are observed in the figure. The lower MCE rates (shown as orange and red colors) for the "with-bycatch" models are consistent across the range of purse seine fishing. The western region, with the highest proportion of purse seine sets, tends to have the highest MCE rates, particularly Type I errors. There are numerous possible explanations for this result, and it would be if interest to determine if this was systemic of the region or peculiar to 2012. This, and other early observations, however led us to attempt some more general regional and seasonal modelling to try and improve overall MCE rates, and they are presented next.

Classification models that include seasonal disaggregation

To improve upon the performance of the models that simply used all 2007-2011 data, without regard to season or region, we next fit models that divided the dataset either into yearly quarters ("Seasonal" models) or large oceanic regions ("Regional" models). Figures 8 and 9 illustrate the Seasonal models fit to the 2007-2011 data disaggregated by quarter, with and without consideration of bycatch, respectively. In the interest of clarity, the details of the number of cases at each node have been eliminated, however overall numbers of correct and incorrect classifications are still listed. Without consideration of bycatch, the first classification rule was always on percentage composition of skipjack in a set, the same initial classification rule for the dataset not broken down by quarter. Subsequent classifications, however, differed considerably among the seasons. Season 2 resulted in the simplest model, Season 3 in the most complex. When bycatch is introduced, once again the presence of rainbow runner in a set is the first classification rule, just as was the case in the season-aggregated dataset. Subsequent classification rules also differ substantially among seasons. Table 6 provides a summary of MCE rates using the same format as Table 6.

Comparison of MCE rates between the Full Year and each of the seasonal rates shows that some seasons outperform and some underperform the full year rate. However, weighted averages of the four seasons, which is the best overall comparison with the Full Year MCE rates, show absolute decreases of 2.6% (tuna-only) and 1.5% (with-bycatch) which are equivalent to relative reductions of about 10%. It thus appears there is some benefit from disaggregating the data in this manner, and using different classification rules for each of the seasons. As is the case for the Full Year models, the bycatch models do consistently outperform the tuna-only models and this holds for all four quarters.

Classification models that include regional disaggregation

We next fit models to the regional datasets, once again without bycatch (Figure 10) and with bycatch (Figure 11). With one strong exception, results basically paralleled those of the seasonal classification. The non-bycatch models split initially on skipjack percentage in the catch. For all three regions the left branch of the models (less pure skipjack sets), the second classification was on yellowfin tuna, either mean length or percentage. For the regional bycatch-included models, Regions 1 and 2 again split initially with regards to presence of rainbow runner. Region 3 is the aforementioned exceptional model result. For Region 3, the initial split was on skipjack composition. The pure skipjack sets (branching to the right) then split on two different bycatch species (rainbow runner and dolphinfish) as well as mean skipjack length. It turns out that, among all models fit, the Region 3 bycatch-included model has the lowest MCE rates and this holds true for both the training dataset as well as the test dataset, as illustrated in Table 8.

Without the inclusion of bycatch information, the regional models do little better than the All Region (same as Full Year in Seasonal table) model, with the weighted regional average only 2% lower. The Region 3 bycatch-free model appears to perform well for the training dataset, particularly with regard to Type II error, but when that model is applied to the Test data, the Type I error rate (41.7%) is the largest of any MCE rates for any of the illustrated model fits. For the bycatch models, there is an an absolute reduction in MCE rate (from 16.5% to 15.2%) between the All Region model and the weighted average of the regional models. This corresponds to a relative reduction in MCE rate of 9%, roughly equivalent to the improvement seen in the seasonal disaggregation for the bycatch models.

Bagging predictors

Bagging predictors were conducted on the same datasets as the simple CART analyses and were similarly based on tuna-only data and then refit with the additional bycatch variables. Because each bagging predictor analysis is based on fitting 30 models (or trees), no attempt is made to illustrate the variables most important to the final results. Similarly, as the MCE error rates differ for each of the individual tree models, we do not report the data fitting MCE error rates, only the test data MCE rates. The Bagging predictors MCE rates for the test data are illustrated in Table 9, and compared to the single tree CART results described above.

As noted earlier, MCE rates for the full data set tended to be in the range of 17-29% for tuna-only models and somewhat lower when bycatch was included, dropping to 14-20%. Seasonal and regional breakdowns decreased absolute MCE rates 1-3% (2-10% in relative terms). The effect of bagging has a range of impacts on predictions. Over all the data, Type I MCE rates of were reduced 15% (tuna-only) and 33% (with bycatch); however Type II MCE rates increased 3% for both types of models. The weighted (overall) MCE rates dropped 8% (tuna-only) and 18% (with bycatch). In terms of the seasonal and regional subsets, differences between the single tree models and bagging predictors were of a highly varied nature, but the overall MCE rate tended to be lower with the bagging models. The magnitude of the overall MCE rate decrease for the seasonal and regional bagging models was roughly similar to that seen when comparing the overall data model and the single tree seasonal and regional models. This improvement does come at a significant cost, both in terms of computing and non-interpretability of the many bagging models used to collectively predict set association.

FAD-closure period

To investigate the possibility of a FAD-closure effect on observer assignation of set association, a total of 10 different year combinations was investigated (Table 10), once again both for tuna-only data as well as with bycatch. Note that in the original full dataset analyses, all observer data between 2007 and 2009 were utilized. However, the FAD-closure period went into effect in 2009, thus there is some question whether the 2007 and 2008 data should be retained in this analysis. The first model fit included the 2007 and 2008 data in predicting set association for the 2009-2011 FCUK-less data; all subsequent model fitting left out the 2007 and 2008 data. Model fits were done for the year combinations of 2009-2011 and 2009-2012, as well as for each year separately. Unlike the CART and Bagging Predictor analyses where the test data (2012) always differed from the training data (2007-2011), the analyses in this section often had matching years; for example the second model in Table 10 predicts 2009-2011 FCUK data on the basis of models fit to 2009-2011 FCUK-less data. In the interests of conciseness, the CART models developed for each dataset are not illustrated, however the model retained variables, listed in order of importance, are included in the results table (Table 10).

Because the FCUK dataset contain only observer-classified unassociated sets, there can only be Type I errors, i.e., sets identified as unassociated but classified by the model as associated. Comparison of Type I MCE rates for the test data (i.e., the FCUK data) are in the same range as those for the training data (i.e., the FCUK-less datasets). In a sense, this is impressive for the following reason: in general, and as illustrated by the earlier results in this paper, test model fits are rarely as good as the training data model fits. This generalization held true for both tuna-only model fits as well as those that included bycatch variables as well.

Discussion

The overarching 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 80% accurate classification. If bycatch data were available, up to 85% accurate classification might be possible. Seasonal and regional breakdowns generally yielded classification rates 2-9% better than the non-disaggregated data. The use of bagging predictors provided a similar level of improvement to set type prediction. The improvement, however, was inconsistent in regards to misclassification error type (I or II) and was highly variable among seasonal and regional variation. It is concluded that the great increase in complexity may not be warranted for the purposes of providing a set of rules to determine set association "in the field".

Regarding the analysis of the FAD-closure datasets, a couple of observations from the model results bear mentioning. First, the important variables for the classification models were very similar among the different year groupings. If bycatch is not used in the tree classification, the percentage of skipjack in the catch was always the most important variable and the second was usually yellowfin percentage or yellowfin mean length. The brief exercise also provided some reassurance in regards to observer classification of purse seine set association during the period of FAD closure. MCE rates of unassociated sets during the FAD closure period are equal, or even a bit lower than, MCE rates during non-FAD closure periods.

We conclude with a discussion of situations where improved, or decreased, set association classification performance might be expected.

Conditions under which catch sampling works best in determining set-type.

One constant across all models is that bycatch, specifically rainbow runner, improves model predictions 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 presence/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.

There is, however, considerable uncertainty as to whether actual bycatch can be confidently assigned to each set. The bycatch may be discarded, consumed by the crew, or mixed between sets. 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 misclassification rates as low as the with-bycatch models for the test data sets.

Generally speaking, analysing data sets disaggregated by either season or region provides modest improvements to model performance – with one notable exception. Region 3, i.e., ocean waters east of 180°, bycatch models achieved MCE rates under 10% for all three error types. Results were more mixed for the Region 3 tuna-only model; Type II MCE rate was less than 5%. However, this came at a high Type I error price as 41.7% of unassociated sets were misclassified as associated. The longitudinal summary (Figure 7) of classification success illustrates that further analyses on other data subsets could be explored.

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). The move towards "spill sampling" will increase both sample size and likely reduce bias, both of which may well increase the precision of models developed to classify set type.

Finally, Harley et al. (2009[4]) 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 associated 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.

Conditions under which catch sampling may be compromised

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 removal of bycatch from sets if bycatch-included models were to be used. "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. Interference with sampling protocol to bias sampling towards tuna bycatch is therefore one 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.

Perhaps the most serious impediment to sampling purse seine sets concerns the issue of mixing sets within a well such that port sampling would not have access to individual sets. The models developed for this analysis use individual sets as "cases" and catch characteristics as predictor variables to classify the cases. If cases instead represented a mix of set types, the analysis becomes complex likely to the point of impossibility. One possibility, not explored here but perhaps worthy of investigation, is application of individual set rules to mixes of sets. The remainder of this paragraph is simple speculation on possible outcomes of such a situation. If bycatch models were used AND accurate bycatch data were available, one almost certain outcome would be a larger misclassification of unassociated sets as associated sets, i.e., an increase in Type I errors. This would occur because the threshold for associated set classification is 1 (one) rainbow runner. Imagine a situation where just one of four combined unassociated sets had a rainbow runner: all four sets would be classified as associated. For rules that involve mean length of either skipjack or yellowfin, it's less clear how mixing of sets would affect set classification. Size samples might not be proportional to set size (on very large sets a max sample size is generally imposed) and a mix of larger associated sets of smaller fish with smaller unassociated sets of larger fish would give an intermediate mean size. From this brief summary, it should be clear that sampling at the level of individual sets, for which 1000s of observer records provide data for model development, would be highly preferable and perhaps the only way forward.

Acknowledgments

We wish to thank Carola Kirchner, Peter Williams, Timothy Lawson, Paul Judd and Alex Tidd for early manuscript reviews. We also thank A. Fonteneau for making the R code with which Figure 2 was produced freely available online.

---

[4] 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)

References

Breiman, L. 1996. Bagging predictors. Machine Learning, 24: 123-140.

Breiman, L. 2001. Random Forests. Machine Learning 45(1): 5-32.

Dagorn, L., Holland, K.N., Restrepo, V., and Moreno, G. 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 and Fisheries 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., pp. 546–555.

Fonteneau, A. Chassot, E., and 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. DOI: 10.1051/alr/2013046

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., and 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(4): 643-668.

Pallarés ,P., Nordstrom, V., Fonteneau, A., Delgado de Molina, A., and 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 Proceedings no. 6, pages 256-263.

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. and Ripley, B. (2013). rpart: Recursive Partitioning. R package version 4.1-3. http://CRAN.R-project.org/package=rpart

Table 1. Summary of observed purse seine set data used in analysis.

Year	Unassociated		Associated		
	Number	Percent of annual sets	Number	Percent of annual sets	Total sets
2007	1133	33.1%	2286	66.9%	3419
2008	1270	36.7%	2186	63.3%	3456
2009	1780	37.5%	2964	62.5%	4744
2010	8137	58.9%	5681	41.1%	13818
2011	6122	38.4%	9805	61.6%	15927
2012	5293	48.8%	5549	51.2%	10842
Total	23735	45.5%	28471	54.5%	52206

Table 2. Mean catch of edible bycatch species in observed purse seine sets. Values are kg/set.

Name	Abbreviation	Unassociated	Associated
barracudas	bar	0.07	2.62
black marlin	blm	6.33	7.61
blue marlin	bum	10.25	16.49
dolphinfish	dol	1.09	29.43
striped marlin	mls	2.20	3.81
rainbow runner	rru	6.98	119.35
sailfish	sfa	0.52	0.45
wahoo	wah	0.35	11.23

Table 3. Number of observed purse seine sets, 2007-2012, disaggregated by set association, season and region. UNA is unassociated and ASS is associated.

Season	UNA	ASS	Total
1	4419	8029	12448
2	5042	9078	14120
3	8345	3572	11917
4	5929	7792	13721
Total	23735	28471	52206

Region	UNA	ASS	Total
1	15034	15523	30557
2	6757	10192	16949
3	1944	2756	4700
Total	23735	28471	52206

Table 4. Number of observed purse seine sets during FAD-closure periods, 2009-2012.

	Unassociated	Associated
2009	796	419
2010	2679	280
2011	1881	711
2012	2260	825
Total	7616	2235

Table 5. Description of the misclassification error (MCE) types and formula 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	$\frac{No. UNA\ sets\ misclassified\ as\ ASS\ sets}{Total\ No. UNA\ sets}$
Type II	False Negative: Associated set misclassified as Unassociated set	$\frac{No. ASS\ sets\ misclassified\ as\ UNA\ sets}{Total\ No. ASS\ sets}$
Overall	Type I + Type II	$\frac{Total\ No. misclassified\ sets}{Total\ No. purse\ seine\ sets}$

Table 6. Comparison of misclassification error rates for model fit to entire 2007-2011 data, without seasonal or regional breakdown.

	Training data (2007-2011)			Test data (2012)		
	Type I	Type II	Overall	Type I	Type II	Overall
Tuna-only	24.2%	14.3%	18.7%	29.1%	17.2%	23.0%
With bycatch	15.4%	11.1%	13.0%	19.6%	13.6%	16.5%

Table 7. Comparison of misclassification error rates for Seasonal models fit to 2007-2011 data. The Avg. values are a weighted average (by no. sets per season) of the seasonal values. The Full Year values are copied from Table 6.

		Training data (2007-2011)			Test data (2012)		
		Type I	Type II	Overall	Type I	Type II	Overall
Tuna-only	S1	32.9%	12.1%	19.1%	30.8%	14.3%	21.2%
	S2	20.8%	13.8%	16.3%	20.4%	15.7%	17.5%
	S3	9.4%	21.6%	13.2%	18.4%	22.7%	19.5%
	S4	17.9%	18.6%	18.3%	35.7%	16.6%	23.9%
	Avg.	18.5%	15.5%	16.8%	24.2%	16.7%	20.4%
	Full Year	24.2%	14.3%	18.7%	29.1%	17.2%	23.0%
With bycatch	S1	19.6%	9.7%	13.1%	14.8%	18.9%	17.2%
	S2	17.7%	9.2%	12.2%	17.4%	9.6%	12.5%
	S3	4.7%	23.3%	10.5%	8.7%	24.7%	13.0%
	S4	15.6%	12.7%	14.0%	28.7%	12.1%	18.4%
	Avg.	13.2%	12.0%	12.5%	15.2%	14.8%	15.0%
	Full Year	15.4%	11.1%	13.0%	19.6%	13.6%	16.5%

Table 8. Comparison of misclassification error rates for Regional models fit to 2007-2011 data. The Avg. values are a weighted average (by no. sets per region) of the seasonal values. The All Region values are taken from Table 6.

		Training data (2007-2011)			Test data (2012)		
		Type I	Type II	Overall	Type I	Type II	Overall
Tuna-only	R1	19.9%	17.8%	18.8%	29.8%	19.6%	24.7%
	R2	10.4%	17.1%	14.6%	16.0%	20.9%	18.6%
	R3	20.4%	3.5%	10.0%	41.7%	4.6%	22.7%
	Avg.	17.3%	16.3%	16.7%	27.0%	18.2%	22.5%
	All Reg.	24.2%	14.3%	18.7%	29.1%	17.2%	23.0%
With bycatch	R1	13.0%	15.2%	14.2%	18.4%	16.4%	17.4%
	R2	10.6%	8.1%	9.1%	15.3%	13.5%	14.4%
	R3	6.9%	7.1%	7.0%	6.2%	9.0%	7.6%
	Avg.	11.9%	11.9%	11.9%	15.9%	14.5%	15.2%
	All Reg.	15.4%	11.1%	13.0%	19.6%	13.6%	16.5%

Table 9. Comparison of misclassification error (MCE) rates between the Bagging Predictor method and the models developed using the simple CART method. The Avg. values are a weighted average (by no. sets per season or region) of the seasonal or regional values. The All Data values are copied from table 6.

		Bagging predictors MCE rate			Change from CART MCE rate		
		Type I	Type II	Overall	Type I	Type II	Overall
Tuna-only	S1	21.5%	19.2%	20.2%	-30%	+34%	-5%
	S2	32.2%	7.6%	16.8%	+58%	-51%	-4%
	S3	17.3%	24.5%	19.2%	-6%	+8%	-1%
	S4	31.8%	16.0%	22.1%	-11%	-3%	-8%
	Avg.	23.7%	15.3%	19.4%	-2%	-8%	-4%
	R1	26.7%	20.5%	23.6%	-10%	+5%	-4%
	R2	19.6%	13.8%	16.5%	+23%	-34%	-11%
	R3	36.1%	6.2%	20.8%	-13%	+35%	-8%
	Avg.	25.7%	16.5%	21.0%	-5%	-10%	-7%
	All Data	24.7%	17.8%	21.2%	-15%	+3%	-8%
With bycatch	S1	10.4%	18.8%	15.3%	-30%	-1%	-11%
	S2	25.3%	7.1%	13.9%	+45%	-26%	+11%
	S3	10.3%	20.8%	13.1%	+18%	-16%	+1%
	S4	21.0%	14.0%	16.7%	-27%	+15%	-9%
	Avg.	15.3%	13.9%	14.6%	+0%	-6%	-3%
	R1	14.6%	16.4%	15.5%	-21%	-0%	-11%
	R2	12.0%	13.1%	12.6%	-22%	-3%	-13%
	R3	18.1%	5.9%	11.9%	+192%	-34%	+56%
	Avg.	14.2%	14.0%	14.1%	-11%	-4%	-7%
	All Data	13.1%	14.0%	13.6%	-33%	+3%	-18%

Table 10. CART model fitting summaries and illustration of misclassification error (MCE) rates when predicting FAD Closure Unassociated Kept aside (FCUK) test data using FCUK-less data (training data) to construct the models.

Tuna-only

			Training data MCE rates					Test data MCE rates		
Training data	n	Variables	Type I	Type II	Overall	Test data	n	Type I	Type II	Overall
2007-2011	36008	SKJ.pct (2), YFT.len	26.0%	13.3%	17.9%	2009-2011	5356	22.8%	NA	22.8%
2009-2011	29133	SKJ.pct, YFT.pct	24.7%	15.3%	18.8%	2009-2011	5356	21.6%	NA	21.6%
2009-2012	37715	SKJ.pct (2), YFT.len	26.1%	14.8%	18.9%	2009-2012	7616	25.5%	NA	25.5%
2009-2011	29133	SKJ.pct, YFT.pct	24.7%	15.3%	18.8%	2012	2260	30.3%	NA	30.3%
2009	3948	SKJ.pct, SKJ.len, YFT.len, BET.pct, SKJ.len	42.4%	4.3%	13.8%	2009	796	29.1%	NA	29.1%
2010	11139	SKJ.pct, YFT.len, YFT.pct, BET.pct	17.1%	20.2%	18.7%	2010	2679	11.8%	NA	11.8%
2011	14046	SKJ.pct, YFT.pct, YFT.len	24.5%	14.9%	17.8%	2011	1881	30.5%	NA	30.5%
2012	8582	SKJ.pct, SKJ.len, YFT.len, SKJ.pct, BET.pct	22.6%	15.7%	18.1%	2012	2260	29.6%	NA	29.6%
2009-2012	37715	SKJ.pct, SKJ.pct, YFT.len	26.1%	14.8%	18.9%	2013	237	20.7%	NA	20.7%
2012	8582	SKJ.pct, SKJ.len, YFT.len, SKJ.pct, BET.pct	22.6%	15.7%	18.1%	2013	237	13.9%	NA	13.9%

With Bycatch

			Training data MCE rates					Test data MCE rates		
Training data	n	Variables	Type I	Type II	Overall	Test data	n	Type I	Type II	Overall
2007-2011	36008	rru.kg, SKJ.pct, YFT.pct, YFT.len, BET.pct	14.4%	12.2%	13.0%	2009-2011	5356	10.4%	NA	10.4%
2009-2011	29133	rru.kg, SKJ.pct, YFT.pct, BET.pct, YFT.len	14.3%	13.2%	13.6%	2009-2011	5356	10.7%	NA	10.7%
2009-2012	37715	rru.kg, SKJ.pct, YFT.len, YFT.pct, BET.pct	15.5%	12.8%	13.8%	2009-2012	7616	15.0%	NA	15.0%
2009-2011	29133	rru.kg, SKJ.pct, YFT.pct, BET.pct, YFT.len	14.3%	13.2%	13.6%	2012	2260	19.9%	NA	19.9%
2009	3948	SKJ.pct, rru.kg, YFT.len, rru.kg, BET.pct	25.8%	7.9%	12.3%	2009	796	17.1%	NA	17.1%
2010	11139	rru.kg, SKJ.pct, YFT.pct, BET.pct, YFT.len	10.2%	15.8%	13.0%	2010	2679	4.7%	NA	4.7%
2011	14046	rru.kg, SKJ.pct, YFT.pct, dol.kg	24.3%	8.6%	13.4%	2011	1881	29.0%	NA	29.0%
2012	8582	rru.kg, BET.pct, YFT.len, YFT.pct	15.9%	14.1%	14.7%	2012	2260	20.4%	NA	20.4%
2009-2012	37715	rruu.kg, SKJ.pct, YFT.len, YFT.pct, BET.pct	15.5%	12.8%	13.8%	2013	237	9.3%	NA	9.3%
2012	8582	rru.kg, BET.pct, YFT.len, YFT.pct	15.9%	14.1%	14.7%	2013	237	10.1%	NA	10.1%

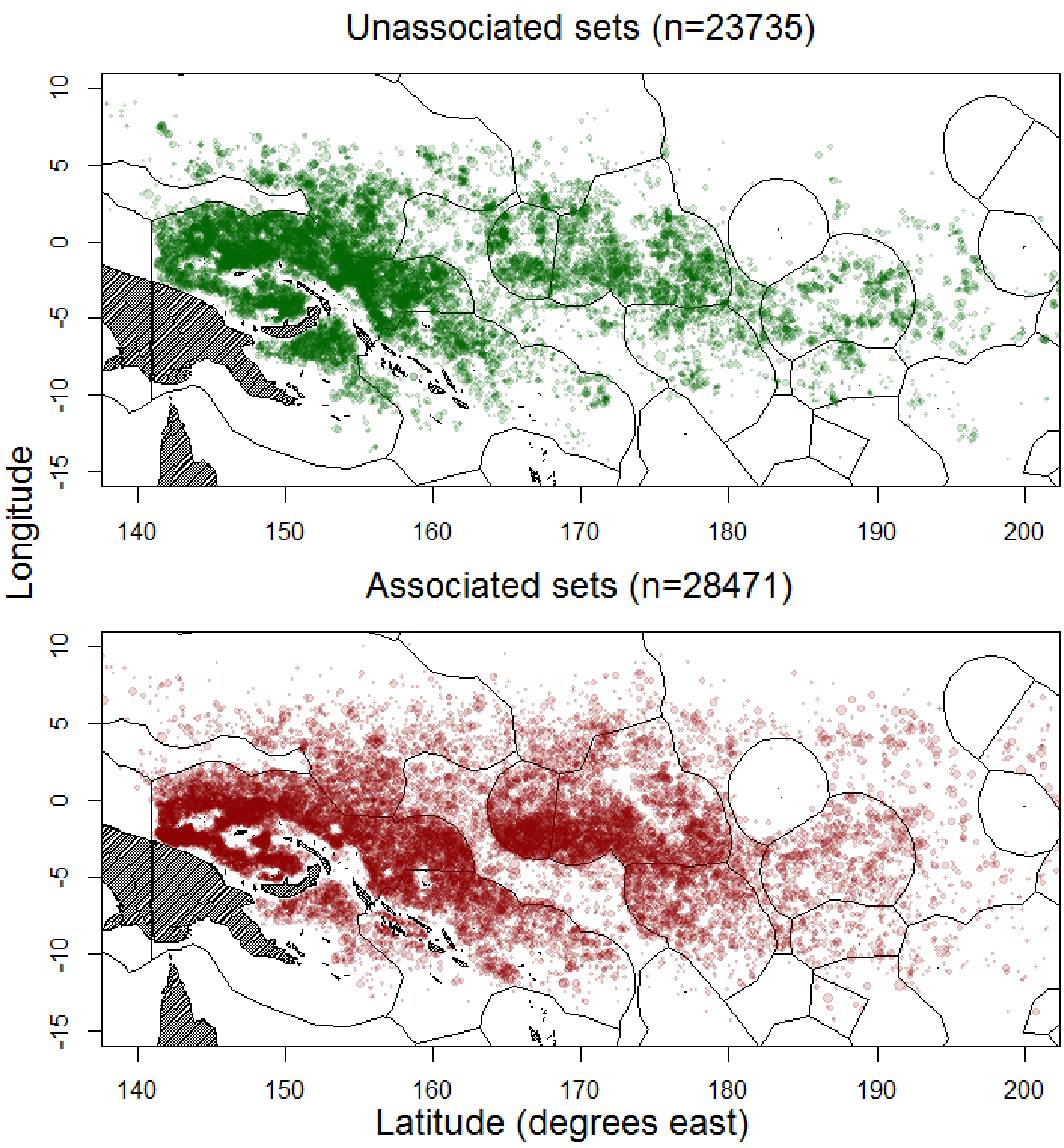


Figure 1. Locations of all observed purse seine sets, separated by set type, for the period 2007-2012. Size of individual circles is proportional to total target tuna catch.

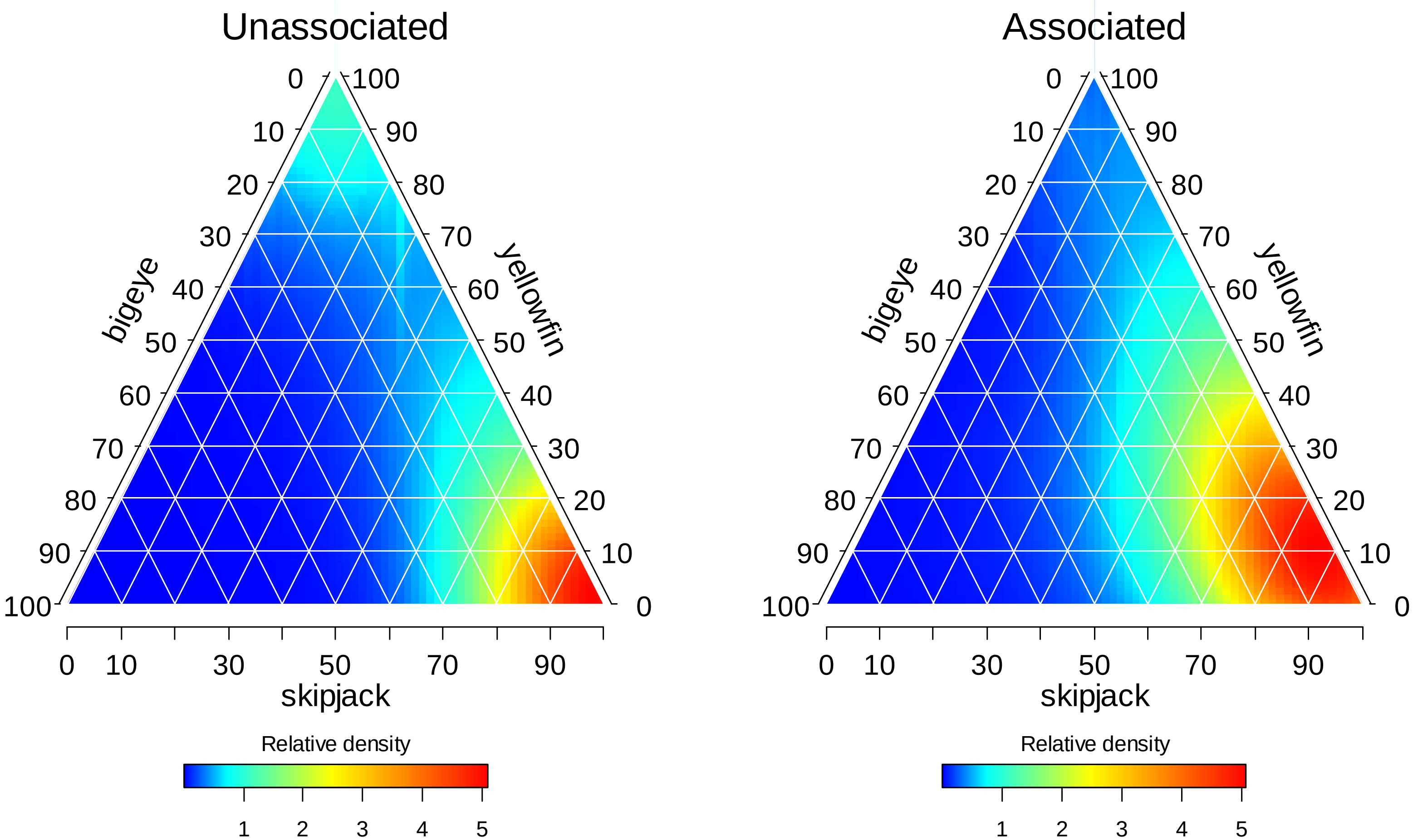


Figure 2. De Finetti (ternay) plots summarizing relative catch composition of the three target tuna species for associated and unassociated purse seine sets. Density indicates relative proportion of total sets having the indicated mix of tuna catch proportions.

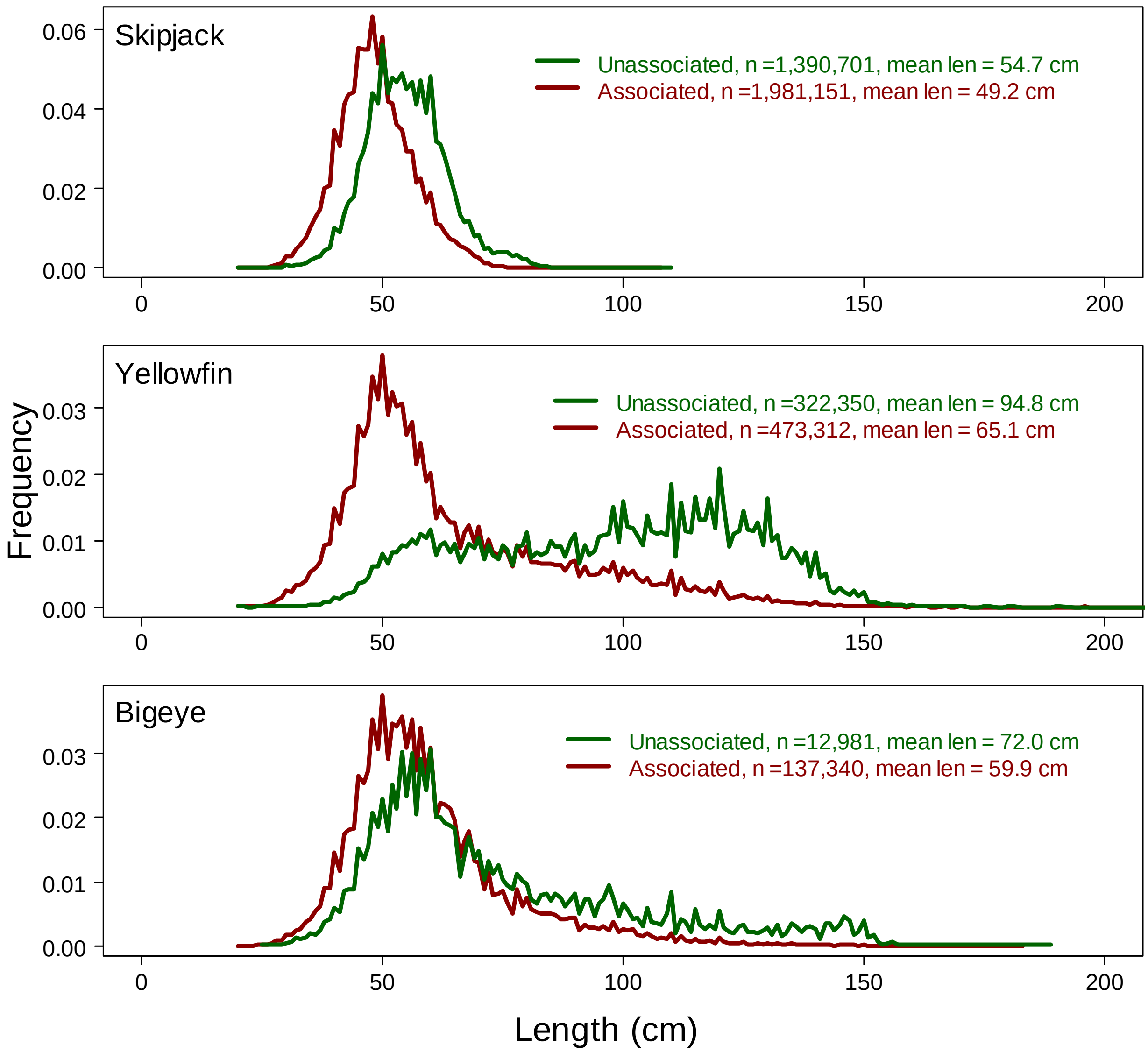


Figure 3. Size distributions for the three target tuna species broken down by set association. The number of measured fish for each distribution is indicated by *n*.

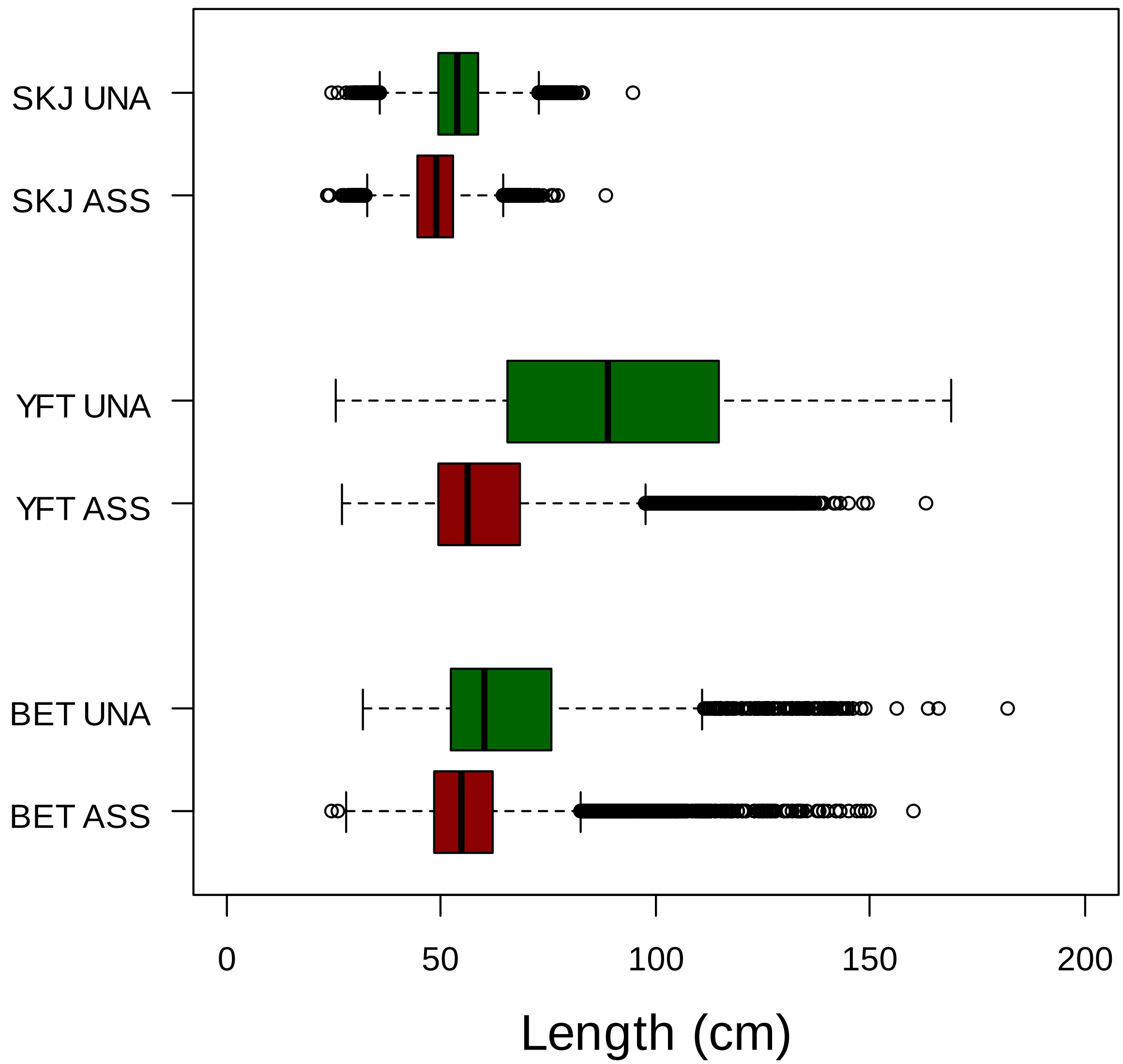
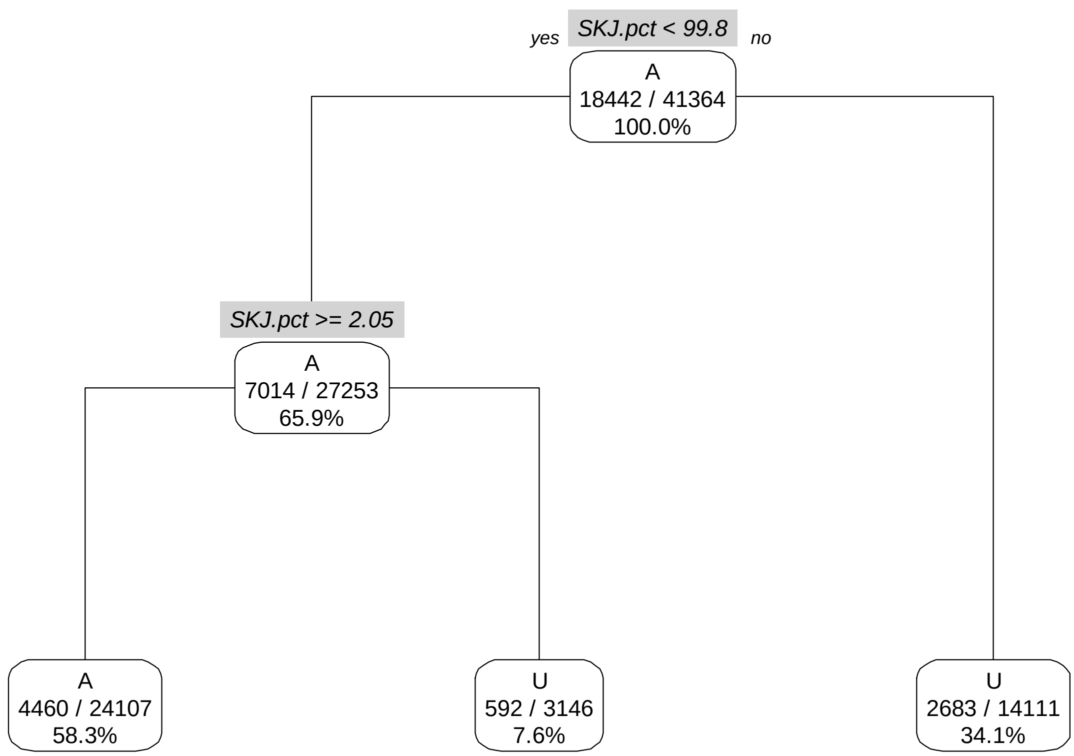
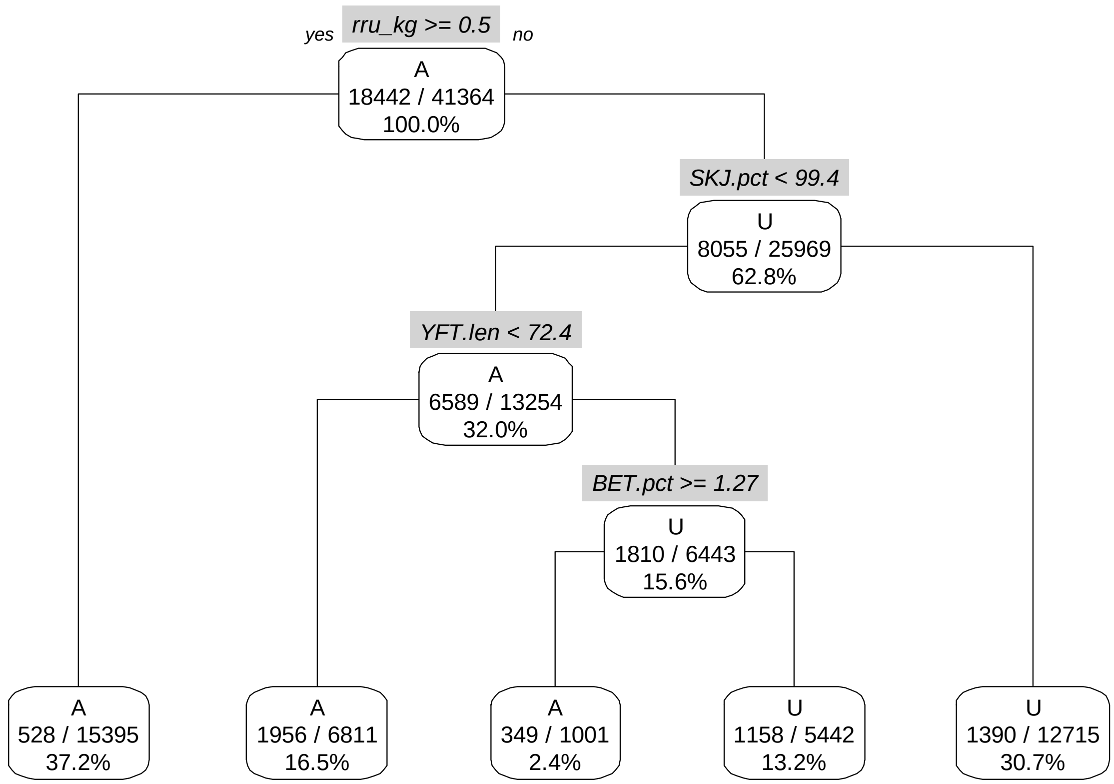


Figure 4. Boxplots of distribution of mean lengths for three target tuna species broken down by set association. The shaded regions show the 25th and 75th quantile while the black bar is the median. Outliers are illustrated by circles, and often represent single measurements, i.e. only one fish caught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



Type I MCE rate, False Positive (U predicted as A) 4460/18442 - 24.2%
Type II MCE rate, False Negative (A predicted as U) 3275/22922 - 14.3%
Overall MCE rate (weighted avg. of Types I and II) 7735/41364 - 18.7%

Figure 5. Classification rules and misclassification error rates developed from 2007-2011 data with no bycatch consideration, and without seasonal or regional disaggregation. See text for interpretation of node statistics.



Type I MCE rate, False Positive (U predicted as A) 2833/18442 - 15.4%
Type II MCE rate, False Negative (A predicted as U) 2548/22922 - 11.1%
Overall MCE rate (weighted avg. of Types I and II) 5381/41364 - 13.0%

Figure 6. Classification rules and misclassification error rates developed from 2007-2011 data, bycatch included, and without seasonal or regional disaggregation. See text for interpretation of node statistics.

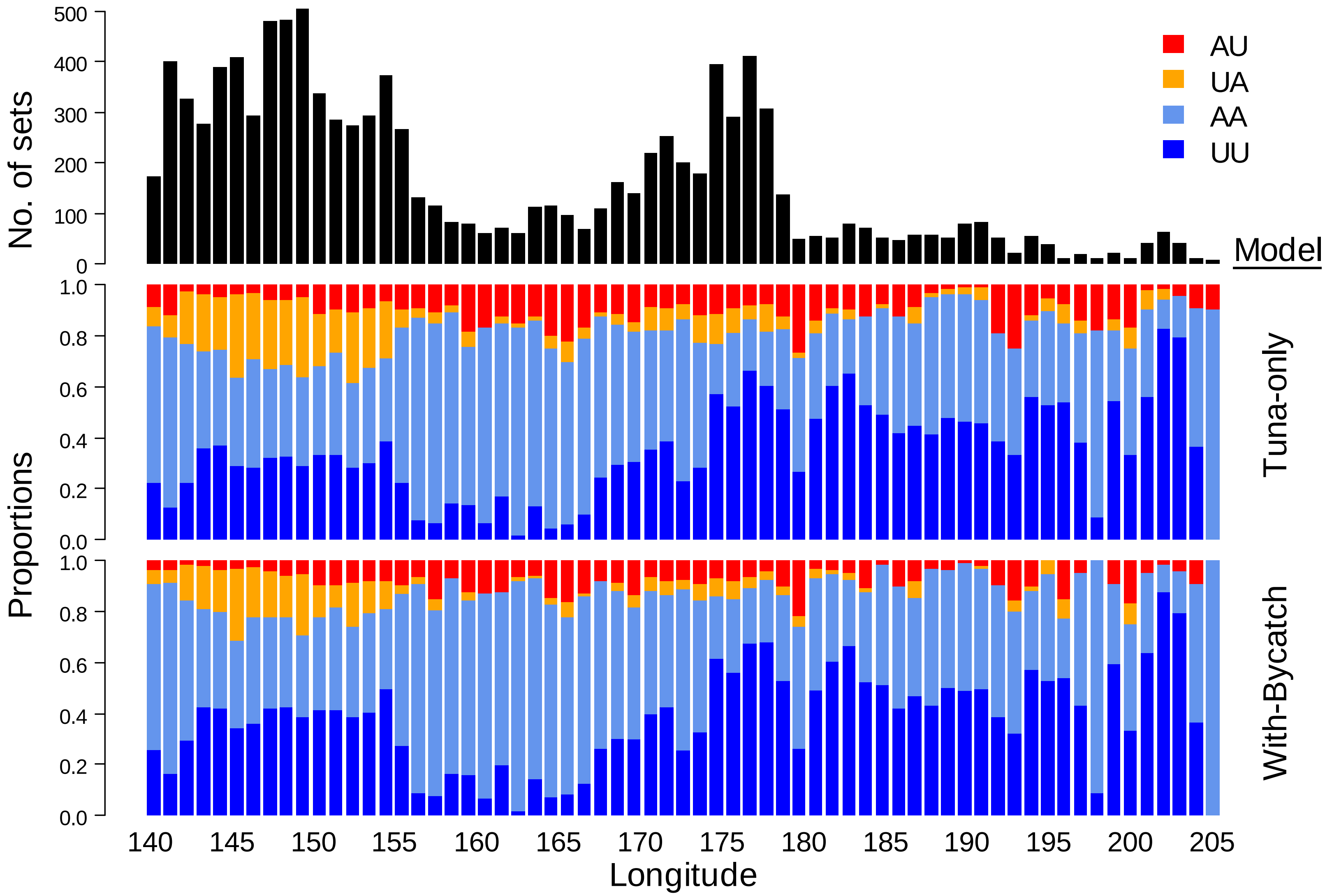


Figure 7. Longitudinal distribution of correct (UU and AA) and incorrect (UA, AU) 2012 purse seine set classifications. The first letter is the observer recorded set type (U indicates unassociated, A indicates associated) and the second letter is the set type predicted by our models.

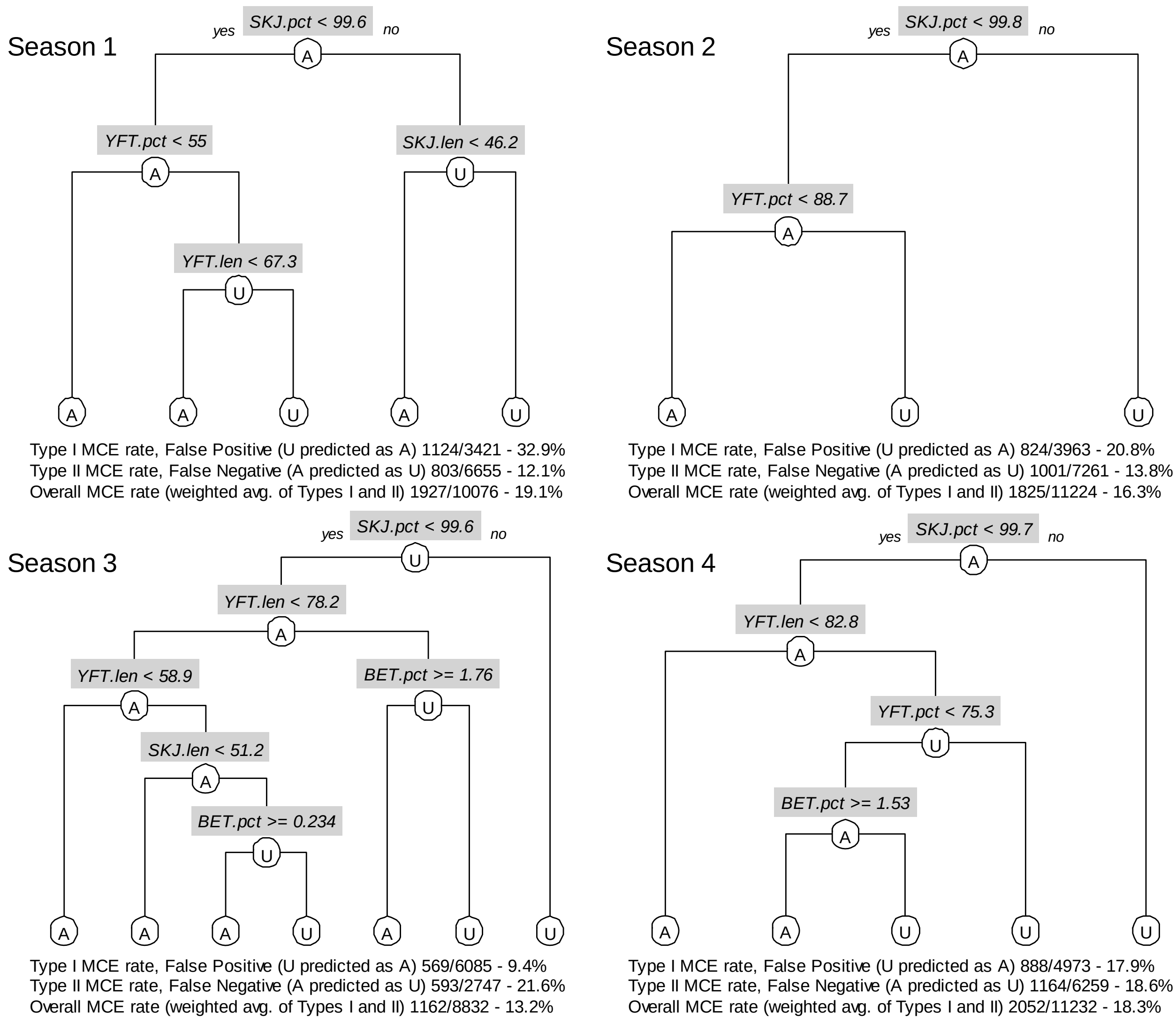


Figure 8. Classification rules and misclassification error rates developed from 2007-2011 data, for seasonal models, without inclusion of bycatch.

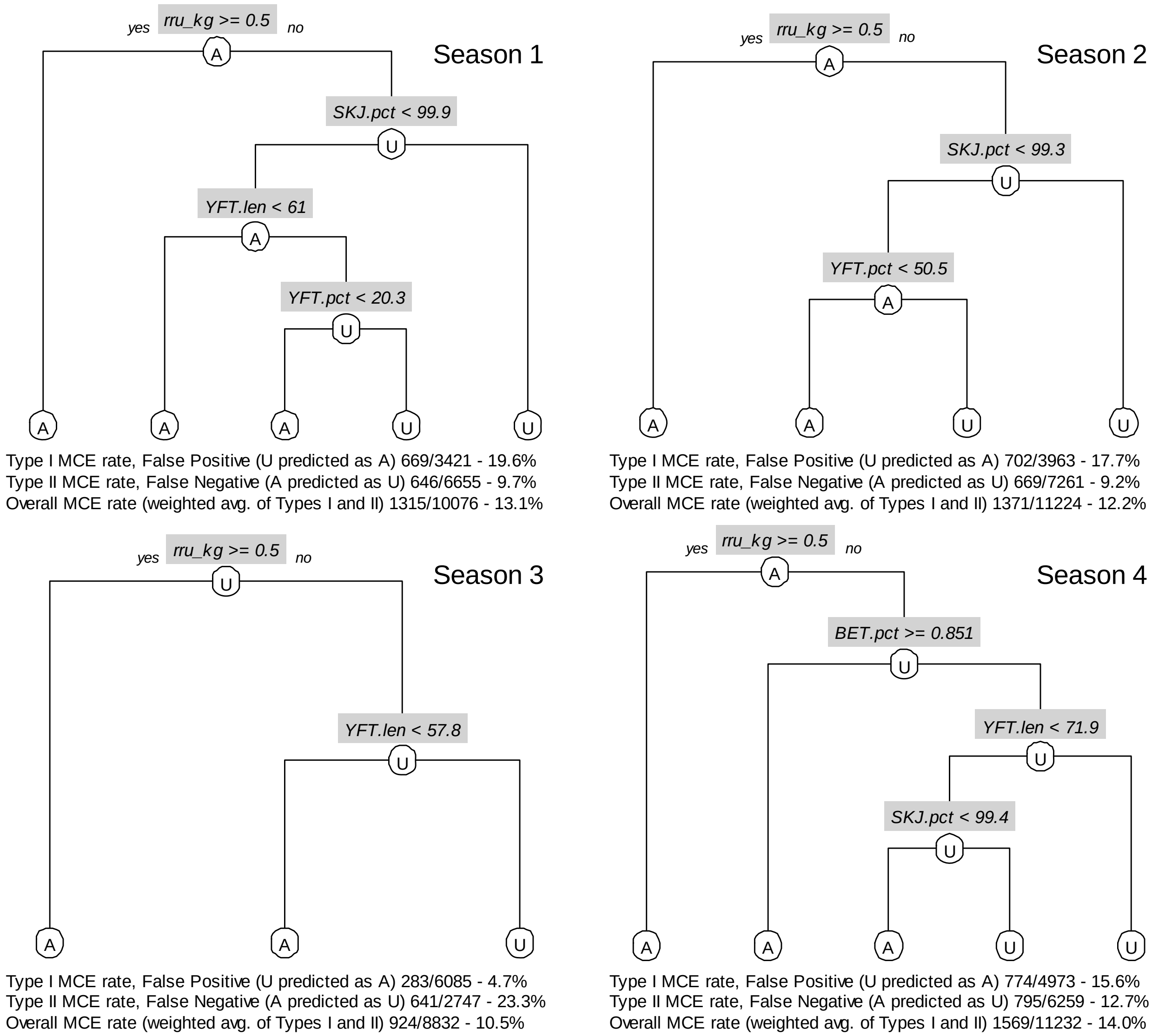
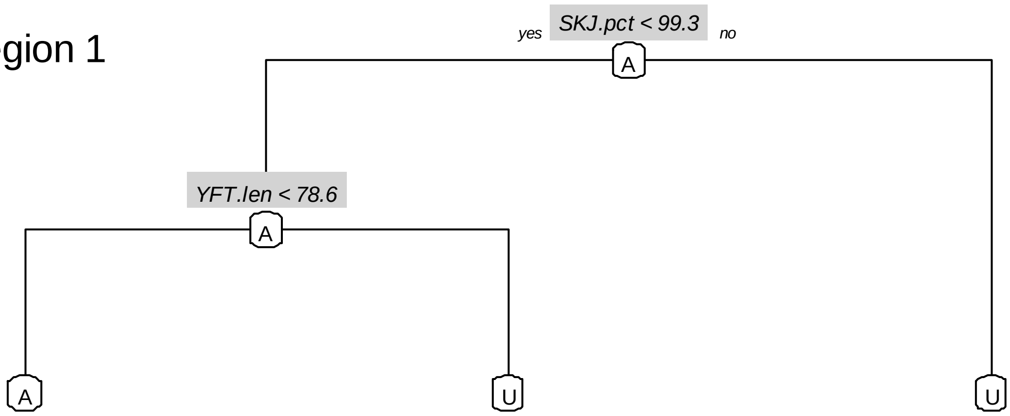


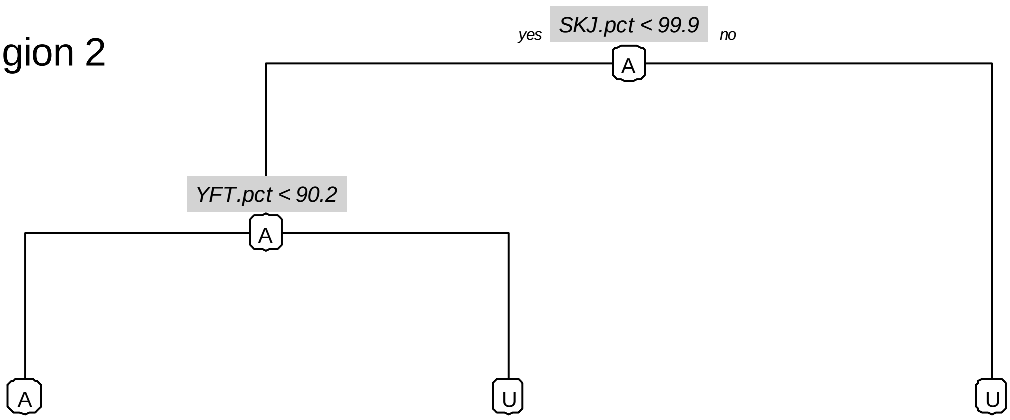
Figure 9. Classification rules and misclassification error rates developed from 2007-2011 data, for seasonal models, bycatch included.

Region 1



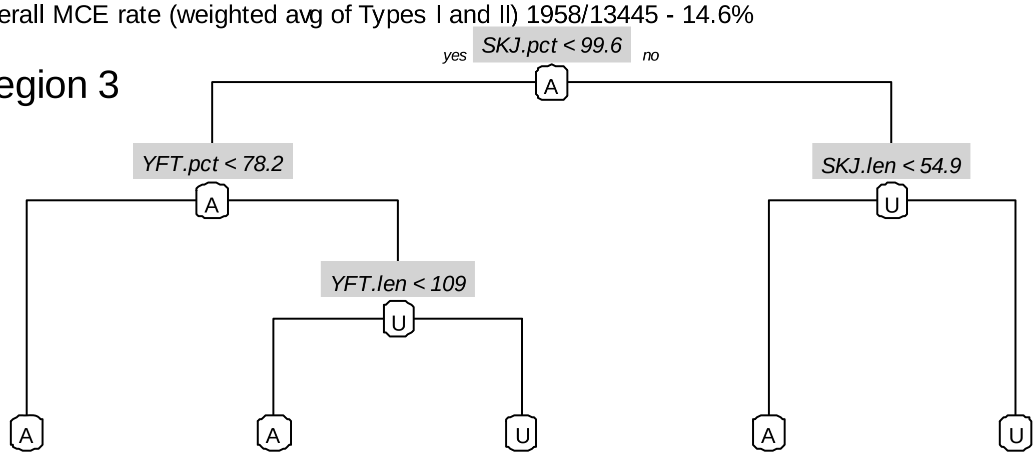
Type I MCE rate, False Positive (U predicted as A) 2396/12033 - 19.9%
Type II MCE rate, False Negative (A predicted as U) 2234/12540 - 17.8%
Overall MCE rate (weighted avg of Types I and II) 4630/24573 - 18.8%

Region 2



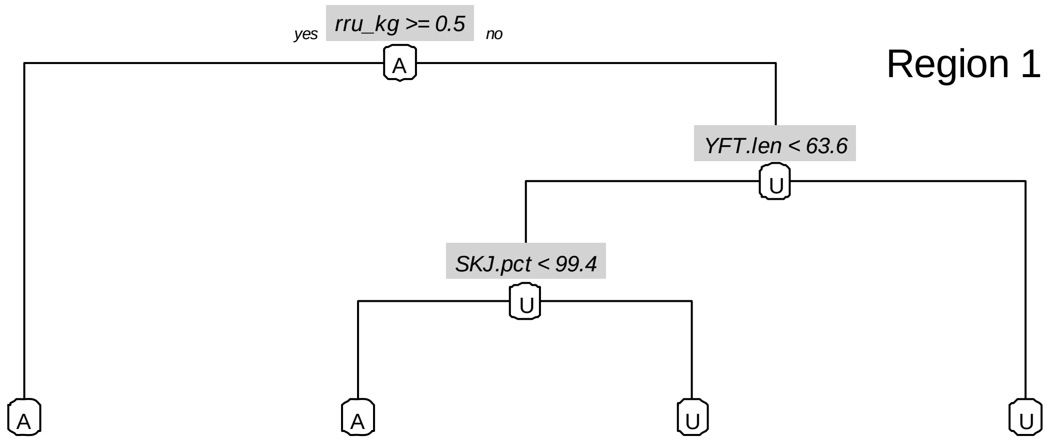
Type I MCE rate, False Positive (U predicted as A) 535/5127 - 10.4%
Type II MCE rate, False Negative (A predicted as U) 1423/8318 - 17.1%
Overall MCE rate (weighted avg of Types I and II) 1958/13445 - 14.6%

Region 3

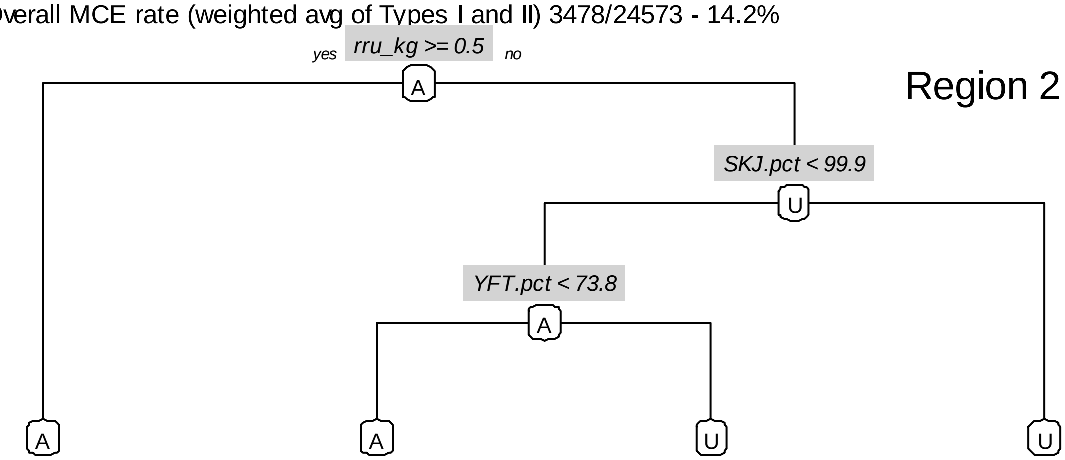


Type I MCE rate, False Positive (U predicted as A) 261/1282 - 20.4%
Type II MCE rate, False Negative (A predicted as U) 72/2064 - 3.5%
Overall MCE rate (weighted avg of Types I and II) 333/3346 - 10.0%

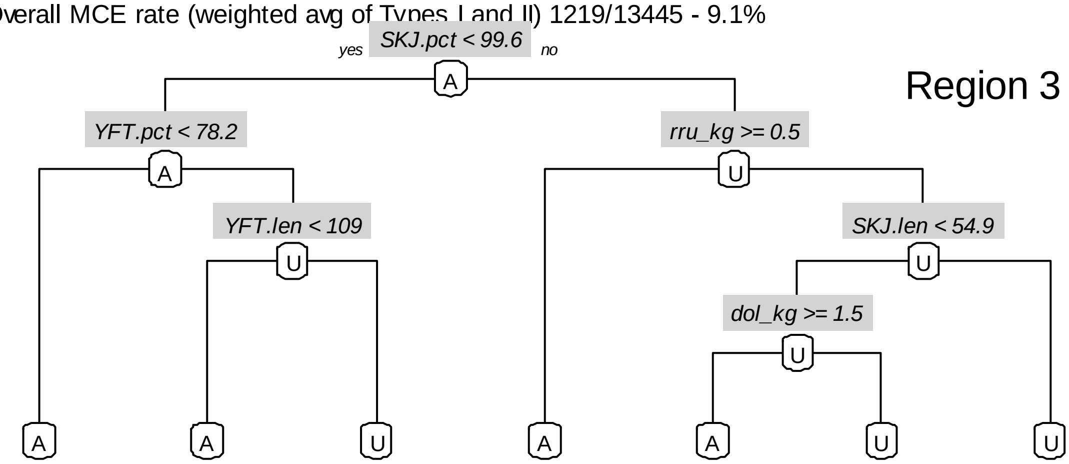
Figure 10. Classification rules and misclassification error rates developed from 2007-2011 data, for regional models, without inclusion of bycatch.



Type I MCE rate, False Positive (U predicted as A) 1566/12033 - 13.0%
Type II MCE rate, False Negative (A predicted as U) 1912/12540 - 15.2%
Overall MCE rate (weighted avg of Types I and II) 3478/24573 - 14.2%



Type I MCE rate, False Positive (U predicted as A) 544/5127 - 10.6%
Type II MCE rate, False Negative (A predicted as U) 675/8318 - 8.1%
Overall MCE rate (weighted avg of Types I and II) 1219/13445 - 9.1%



Type I MCE rate, False Positive (U predicted as A) 88/1282 - 6.9%
Type II MCE rate, False Negative (A predicted as U) 147/2064 - 7.1%
Overall MCE rate (weighted avg of Types I and II) 235/3346 - 7.0%

Figure 11. Classification rules and misclassification error rates developed from 2007-2011 data, for regional models, bycatch included.