

Towards the derivation of fisheries-independent abundance indices for tropical tuna: Report on biomass estimates obtained from a multi-frequency echosounder buoy model (M3I+)

A. Diallo¹, Y. Baidai^{1,2}, L. Mannocci¹ and M. Capello^{1*}

¹ MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Sète, France.

² Centre de Recherches Océanologiques (CRO), Abidjan, Côte d'Ivoire.

* manuela.capello@ird.fr (corresponding author)

1. Introduction

For several decades, the industrial tropical tuna purse-seiners have employed drifting Fish Aggregating Devices (FAD) worldwide to increase their chances of locating tuna aggregations and catching them (Fonteneau et al, 2013, Dagorn et al, 2013). Drifting FADs were first equipped with radio beacons and in later years with GPS buoys, to remotely locate them. During the end of 2000's, these buoys were further improved with the addition of echosounders. Nowadays, thanks to the undisputed benefits of such buoys for the fishermen and their affordable costs, all drifting FADs deployed by the industrial purse seiners are equipped with echosounder buoys that can remotely provide, in near real-time, estimates of the amount of tuna at FADs, together with their GPS position. The buoys manufacturers constantly invest in the development of novel hardware and software, in order to improve the accuracy of the biomass estimates provided by the buoys. In the last years, multi-frequency buoys were developed, aiming at improving the biomass estimates and the discrimination of tuna species at the FADs.

Recent studies have highlighted the potential of the echosounder buoys data for studying various aspects of the ecology and behaviour of fish associated with floating objects, including the derivation of fisheries independent abundance indices (Lopez et al., 2017; Moreno et al., 2016; Capello et al. 2016, Orue et al., 2019, Baidai et al. 2017, Baidai et al 2018, Santiago et al. 2019). Such data is non-conventional, since the low-cost echosounders integrated into the buoys operate at a much lower resolutions (both in time and space) than the scientific echosounders used in scientific campaigns. Also, the data transmitted by the buoys are compressed and simplified, due to satellite communications constraints. Finally, depending on the buoy models and brands, the technical characteristics of the buoys and the format of the data can differ. In order to exploit the acoustic data provided by the buoys for scientific purposes, dedicated studies are required to evaluate the accuracy of the biomass estimates obtained by these buoys and how these estimates depend on the buoy models and brands. The exploitation of this data by the scientific community is still recent. Two main brands are established within the echosounder buoys' market: Marine Instruments and Satlink. The works of Lopez et al. (2016, 2017) and Orue et al. (2019) focused on Satlink buoys whereas two recent papers of Baidai et al. (2017, 2018) worked on M3I buoys from Marine Instruments. Since 2016, Marine Instruments developed a new buoy model (M3I+), working with two frequencies (50 kHz and 200 kHz) and presenting an improved hardware and software with respect to the previous models of the same brand. Following its introduction, this new buoy model has been gradually substituting the older M3I model (Table 1).

2. Objectives

The general objective of this work is to exploit the novel data obtained from the M3I+ buoys in the Indian ocean for deriving novel abundance indices for tropical tuna.

The specific objectives of this study are:

- to obtain biomass estimates from the M3I+ buoy model and assess their accuracy.
- to compare the accuracy of M3I+ with other buoy models (mainly M3I).
- to compare different metrics (e.g., lifetime of the aggregation and colonization times) obtained from different buoy models (M3I+ and M3I) instrumenting FADs located in the same spatio-temporal strata.

3. Structure of the document

In the following, the document is structured along the following sections:

Data description (section 4): The acoustic, logbook and observers databases used in the analysis are described and a detailed description of the acoustic data obtained from the M3I+ buoys is provided.

Biomass estimate (section 5): The results obtained using different algorithms (GLM, Random Forest and Deep Learning) for estimating the tuna biomass are provided, together with the algorithm characteristics.

Comparison of metrics with M3I buoys (section 6): Different metrics (e.g., lifetime of the aggregation and colonization times) obtained from two different buoy models (M3I+ and M3I) are compared.

Discussion (section 7): Results obtained for the different algorithms and frequencies are compared on the light of the current knowledge on tuna behavior at FADs and the perspectives of this work in terms of abundance estimates are discussed.

4. Data description

4.1 Acoustic and activity databases

We used echosounder buoys, logbook and observers databases recorded between 2016 and 2018 in the Indian ocean, hosted by Ob7 (Observatory of Tropical Exploited Pelagic Ecosystems, IRD).

The echosounder database contains the acoustic data transmitted by all M3I+ buoys (Marine Instruments) deployed by the French fleet. Table 1 shows the number of M3I+ buoys constituting the database for the Indian ocean, since 2016, when the M3I+ was first introduced. The number of buoys for the other Marine Instruments buoys models (M3I, M4I and MSI) is also reported, revealing that the number of M3I+ models rapidly increased after its introduction in the market.

Buoy model	2016	2017	2018
M3I+	70	2709	3570
M3I	10827	11489	11187
M4I	818	274	68
MSI	14	7	0
Total	11729	14479	14825

Table 1: Number of Marine Instruments buoys, for each model and year, in the Indian ocean on a yearly basis since 2016, constituting the Ob7 buoys database.

From the logbook and observers databases, we considered the three following activities on FADs: sets, visits (without catch) and deployments. For all activities, we considered the reported GPS position and datetime, as well as the unique identification number of the buoy (named below *buoy ID*) instrumenting the FADs. The latter information was not present for all sets reported, and was maximized by combining the data present in both the observers and logbook databases. When the information was present in both databases, the logbook data were chosen. For the activities corresponding to a set, the amount of tuna catches (in tons) for the major tuna species: yellowfin tuna (YFT, *Thunnus albacares*), skipjack tuna (SKJ, *Katsuwonus pelamis*), bigeye tuna (BET, *Thunnus obesus*) was considered.

4.2 M3I+ buoys characteristics and data

The M3I+ echosounder buoys exploits two different frequencies, at 50 and 200 kHz with beam angles of respectively 36° and 8°. They are powered by Solar Panel rechargeable batteries and dispose of a pack of alkaline batteries as a backup. Similarly to the previous M3I models (Baidai et al. 2018), the M3I+ buoys sample the water column between 0 and 150 m, with a vertical resolution of 3 m. The buoy performs acoustic samplings every 1 min and provides, in its default mode, 12 acoustic samples per day (i.e., one sample for each time bin of 2 hours), plus 4 acoustic samples at dawn. Such acoustic samples correspond, for each frequency, to the highest acoustic energy recorded across the full water column during a given time bin.

For both frequencies and all time-bins, the acoustic samples are expressed as a suite of discrete acoustic scores comprised between 0 and 15, recorded for each 3-meters layer comprised between 0 and 150 meters (50 layers). The first layer (depth 0-3 m) corresponds to the blind angle of the acoustic beam, and its acoustic energy is set by default to a fixed value. The acoustic samples are visualized in real-time by the fishermen through a dedicated software (MSB) provided by the constructor, where the acoustic scores are represented through a color code (see Figure 1).

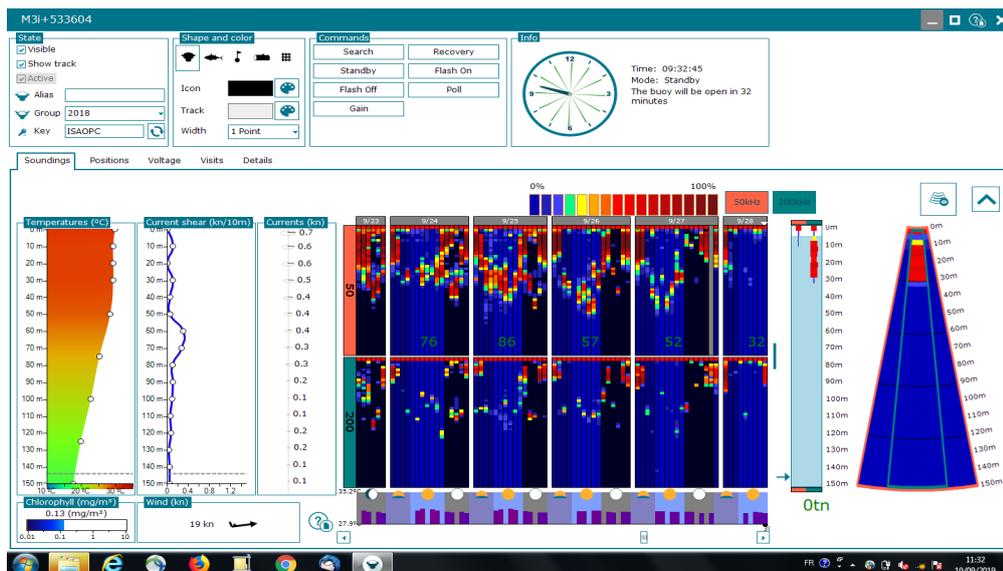


Figure 1: Overview of the acoustic data obtained through the M3I+ buoys visualized through the MSB software (Marine Instruments). In the center of the figure, the two rows correspond to the acoustic samples obtained respectively at 50 and 200 kHz. The color scale on the top represents the 15 acoustic scores.

4.3 Matched echosounder and activity databases

Based on the unique *buoy ID* number, the acoustic data obtained from the M3I+ echosounder buoys were matched with the activity data reported in observers and logbook databases. The matched database was first filtered by eliminating all buoy IDs showing inconsistent GPS positions between the reported activity and the buoy data. To this purpose, each time the distance between the reported activity and the position of the buoy recorded 24 hours before the activity was larger than 180 Km (corresponding to a maximum daily drifting speed of 4 knots for the buoy), the data was discarded.

Then, similarly to the procedure applied by Baidai et al. (2018), the echosounder data recorded 24 hours before the visit (without a set) of a given buoy were associated to tuna absence (i.e., 0 tons of tuna at the FAD), considering that, if tuna aggregations were present at the FAD, they would have resulted in a set. Similarly, the acoustic data recorded 5 days after a deployment (during 24 h) were associated to tuna absence. The rationale of this choice, already presented in Baidai et al. (2018), was to incorporate the acoustic signature of the bycatch species, that generally colonize the FADs early after their deployment, as a tuna absence. In order to minimize the possible bias related to an early colonization of tuna, for the visits and deployment activities, only the buoys that did not correspond to a set during the 7 days after the visit/deployment were considered. Finally, for the buoy ID corresponding to a set, the echosounder data recorded during 24 h on the day before the set were associated to the corresponding tuna catches. Table 2 reports the final database used in the analysis.

<i>Activities</i>	<i>Years</i>		
	2016	2017	2018
<i>Deployments</i>	5	904	886
<i>Sets</i>	3	325	387
<i>Visits</i>	1	60	103
<i>TOTAL</i>	9	1289	1376

Table 2: Total number of activities by year considered in the analysis, resulting from the combination of the M3I+ acoustic database and the logbook/observers databases.

The distribution of the tuna catches (for the three major tuna species) per set and the catch proportion of each tuna species constituting the final database are shown in Figure 2 and 3.

The average acoustic scores obtained at 50 kHz and 200 kHz corresponding to tuna absences, tuna catches of less than 25 tons and tuna catches of more than 25 tons are shown in Figures 4 and 5 respectively.

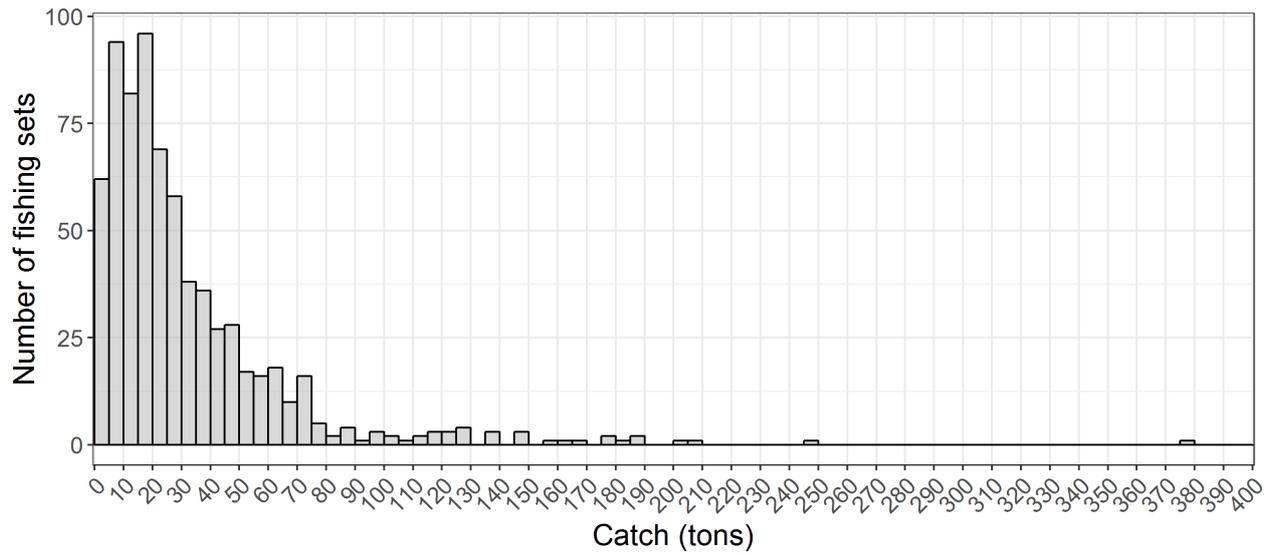


Figure 2: Distribution of the catch per-set for the major tuna species (Skipjack, Bigeye and Yellowfin) constituting the final database used in the analysis.

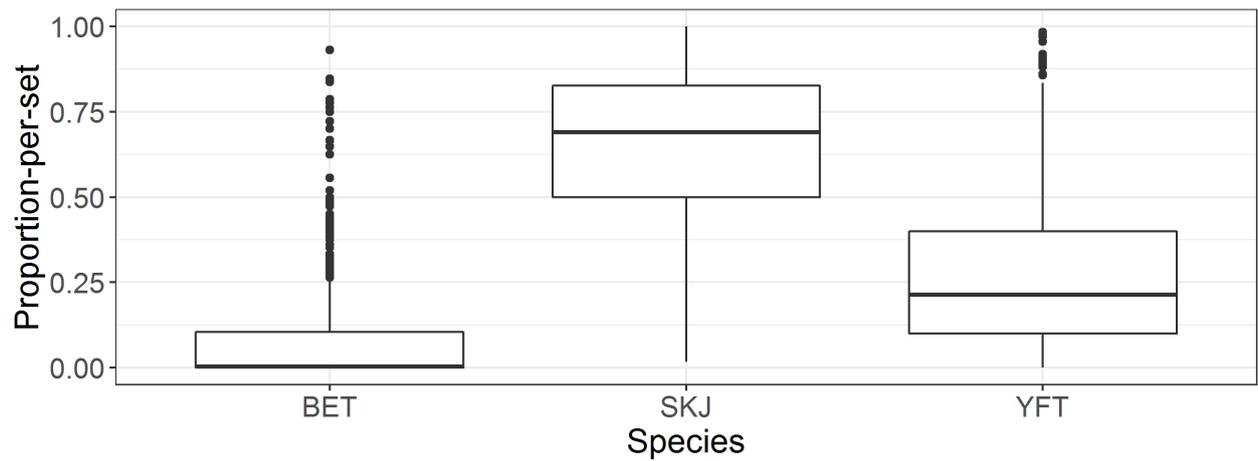


Figure 3: Boxplot of the proportion per set of the three major tuna species. BET stands for bigeye tuna, SKJ for skipjack and YFT for yellowfin tuna.

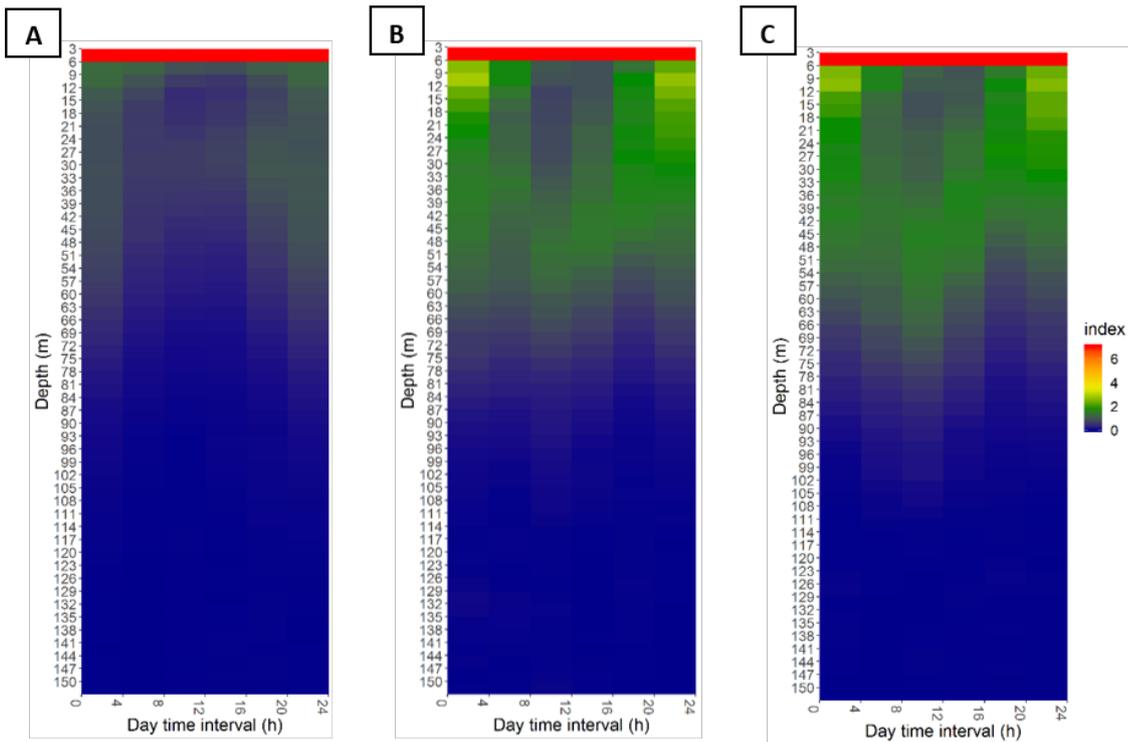


Figure 4: Average acoustic scores by depth layer and time interval for 50 kHz. Panel A correspond to tuna absence, panel B corresponds to catches of less than 25 tons and panel C to catches of more than 25 tons.

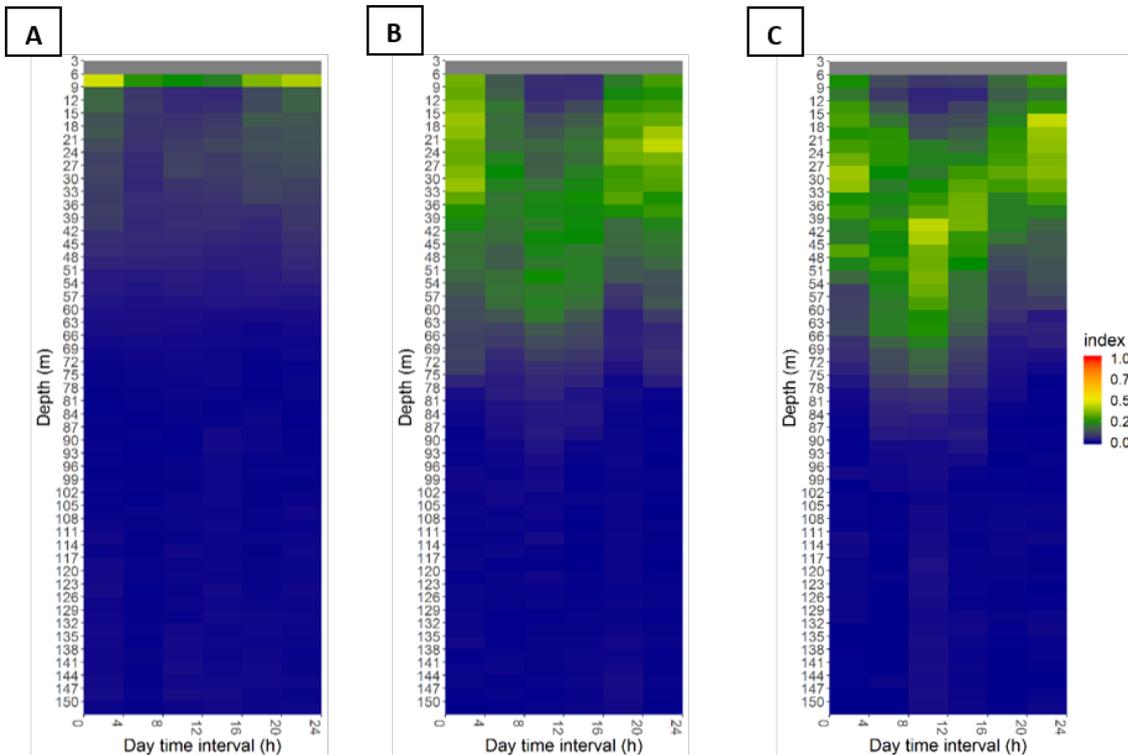


Figure 5: Average acoustic scores by depth layer and time interval for 200 kHz. Panel A correspond tuna absence, panel B corresponds to catches of less than 25 tons, and panel C to catches of more than 25 tons.

5. Biomass estimates

5.1 Random Forest algorithm

5.1.1 Algorithm description

We considered the supervised learning approach developed by Baidai et al. (2018). This approach is based on a preliminary processing of the acoustic data recorded during a full day (24 hours) of sampling, followed by a classification based on the random forest algorithm. Data preliminary processing consisted, for each frequency, in clustering the acoustic data sampled by the buoy over 6 temporal bins of 4-hours and 6 aggregated-depth layers, which summarized the daily acoustic information into 6×6 matrices referred to as "daily acoustic matrices" (Figure 6). First, in order to aggregate the acoustic samples over time, the 24-hours data were aggregated over 6 slices of 4 hours each, by taking the highest acoustic sample over 4-hours intervals, i.e., the highest sum of acoustic scores recorded over all layers. This resulted in a matrix of 6 columns (one for each time bin), and 50 rows for the different depth layers. Secondly, clustering methods identified 6 homogeneous groups of depth layers. The cluster analysis was based on a dissimilarity matrix computed from Euclidean distance and Ward's method for merging clusters, using the R package "cluster" (Murtagh and Legendre 2014).

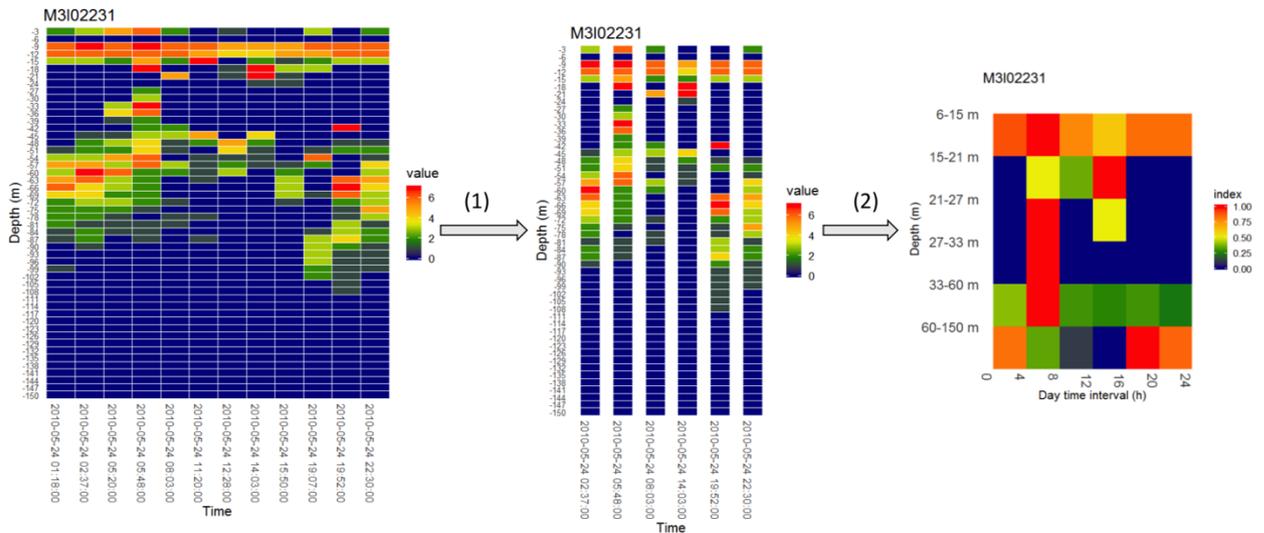


Figure 6: (From Baidai et al. 2018). Acoustic data pre-processing. (1) Temporal resolution reduction, selecting the highest acoustic sample over a 4 hours period. (2) Layers aggregation gathering the 50 vertical layers into 6 layers based on cluster analysis. The final output is a 6×6 matrix (daily acoustic matrix) summarizing the aggregation on a full sampling day.

The estimates of tuna biomass were then carried out using a random forest algorithm (R package "random forest" (Liaw and Wiener 2014)) trained on the combined acoustic and activity data. First, a binary classification algorithm was run, considering the data recorded on FAD deployments and visits without fishing sets as tuna absence and positive fishing sets as tuna presence. Then, following Baidai et al. (2018), the performances of the algorithm to estimate tuna biomass were studied considering 4 classes ("No tuna", "less than 10", "between 10 and 25 tons" and "more than 25 tons"). Finally, the algorithm performances were assessed for 3 classes ("No tuna", "less than 25 tons" and "more than 25 tons").

Three thousand trees were grown for each classification. For each classification model, the number of variables randomly sampled as candidates at each split (denoted as “mtry” in the R package “random forest”) was assessed through a grid-search strategy implemented with the R package caret (Kuhn, 2008). The optimal "mtry" was then defined as the value generating the lowest classification error rate. A stratified down-sampling procedure was also applied, for dealing with the imbalanced number of observations in the different size categories. This sampling procedure consisted in resampling the dominant size category to make their frequencies closer to the rarest size category (Kuhn and Johnson, 2013). The importance of the predictors in the classification process in each ocean was assessed through the analysis of the mean decreased accuracy in the random forest model (i.e., the increase of prediction error after permuting each variable while all others are left unchanged during the tree construction; Breiman, 2001). Model training and evaluation were performed through a hold-out validation method repeated 10 times. In each of the 10 replica, the original dataset was divided into two subsets: the training set and the validation dataset (75% and 25% of the initial data, respectively). The latter was used to calculate the model accuracy (proportion of correct predictions), the kappa coefficient (Cohen, 1968) and to build confusion matrices.

To the purpose of obtaining biomass estimates from the data obtained both at 50 kHz and 200 kHz, multiple combinations of frequencies were tested. First, the algorithm was run on the acoustic data provided by either the 50 kHz, or the 200 kHz frequency taken separately, using the Baidai et al. (2018) procedure explained above. Then, the algorithm was run using the acoustic scores recorded at both frequencies, following the two following approaches. In the first approach (labelled “50 & 200 kHz” in Table 3), the acoustic scores of the two frequencies were pre-processed separately, leading to two 6x6 daily acoustic matrices, using the the Baidai et al. (2018) procedure. Then, a 12x6 acoustic matrix was constructed (with 6 columns for 50 kHz and 6 columns for 200 kHz), merging the daily acoustic matrices obtained for the two frequencies. The random forest algorithm was then run on this augmented daily acoustic matrix. In the second approach (labelled “50 x 200 kHz” in Table 3), a 6x6 daily acoustic matrix was obtained from the multiplication of the daily acoustic matrices obtained at 50 kHz and 200 kHz. The random forest algorithm was then run on this 6x6 daily acoustic matrix.

Combination of frequencies	Daily Acoustic matrix size	Daily Acoustic matrix elements
50 kHz	6x6	Baidai et al. 2018
200 kHz	6x6	Baidai et al. 2018
50 & 200 kHz	12x6	First 6 columns corresponding to the daily acoustic matrix of 50 kHz, last 6 columns of 200 kHz
50 x 200 kHz	6x6	Matrix elements obtained multiplying the elements of the daily acoustic matrices of 50 kHz and 200 kHz

Table 3. Description of the different combinations of frequencies used to estimate the biomass.

5.1.2 Results for the Random Forest algorithm

5.1.2.1 Clustering of depth layers

The cluster analysis was run for all combinations of frequencies described in Table 3. For all combinations, the clusters analysis provided the same 6 groups of layers, see Figure 7.

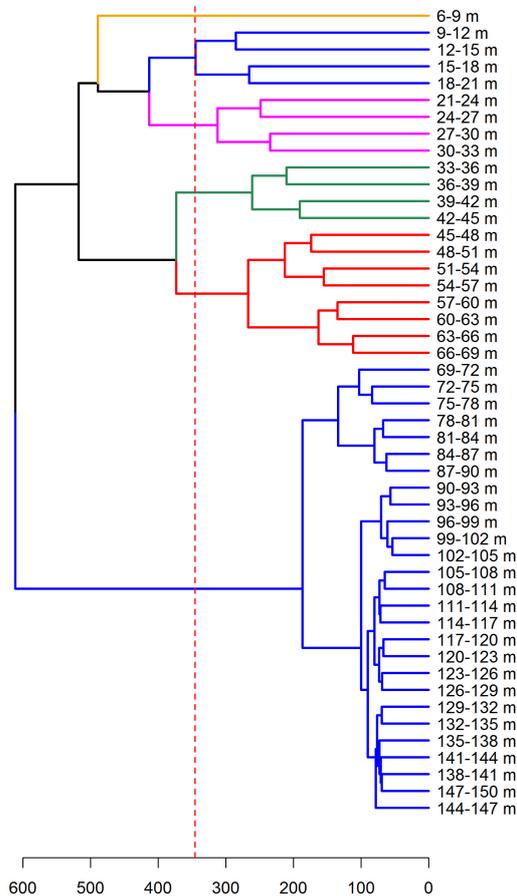


Figure 7: Dendrogram resulting from the cluster analysis of the raw acoustic data. The red horizontal line indicates the height at which the dendrogram was sliced to create the 6 groups of layers. Colors identify the different groups of depth layers constituting the clusters.

5.1.2.2 Binary classification : Presence/Absence

Random Forest results derived from the different combination of frequencies (Table 3) showed similar performances from one combination of frequencies to another (Table 4) with a mean accuracy value of 0.86 ± 0.014 , and a mean kappa of 0.73 ± 0.027 . The good performances of the random forests for Presence/Absence are also revealed by the corresponding confusion matrices (Figure 8) which show a predominance of diagonal elements (true positives and true negatives) for all frequencies combinations. Variables importance analysis (Figure 9) revealed that deep layers (>35m) are the most important for tuna detection. Diel patterns, with tuna being more present during the daytime, appeared clearly for 50 kHz and 50x200 kHz. Nevertheless, these patterns were less clear for 200 kHz (where the shallowest depths are also important at nighttime) and completely reversed in 50 & 200 kHz. For the 50 & 200 kHz, a deeper inspection of the random forest outputs for different values of “mtry” (see 5.1.1) revealed the presence of another solution, with the same accuracy and kappa, having the same variable importance found for 50 kHz and

200 kHz taken separately (as in Figure 9A and 9B). Supplementary tables providing all performance metrics can be found in Appendix 1.

Performance metrics	50 kHz	200 kHz	50 & 200 kHz	50 x 200 kHz
Accuracy	0.88 ± 0.01	0.85 ± 0.02	0.86 ± 0.02	0.85 ± 0.02
Cohen's Kappa	0.76 ± 0.03	0.71 ± 0.04	0.73 ± 0.03	0.70 ± 0.04

Table 4: Summary of performance metrics for the various combination of frequencies tested for binary classification.

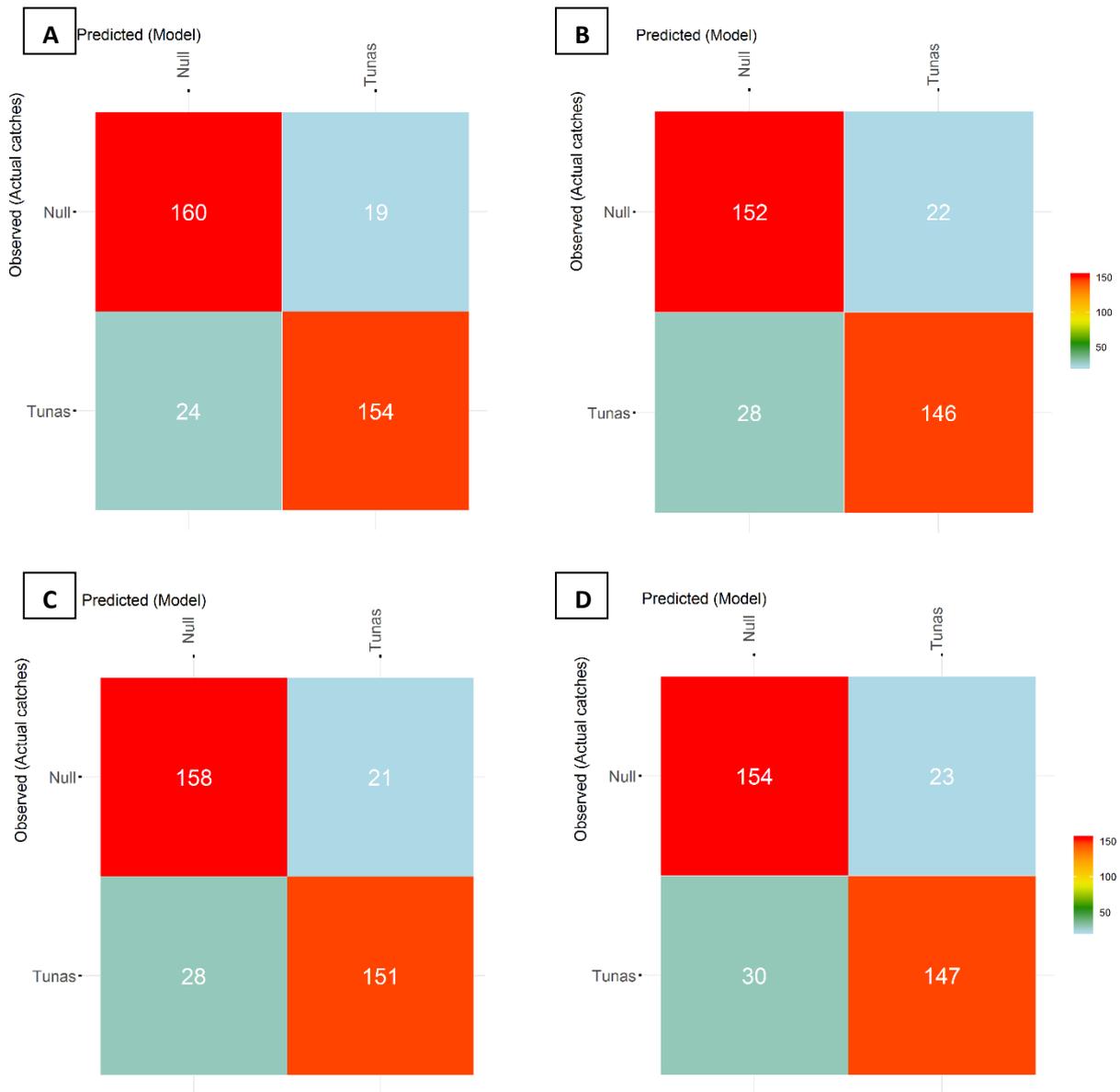


Figure 8: Confusion matrices for binary classification random forests: 50 kHz (panel A), 200 kHz (panel B), 50 & 200 kHz (panel C) and 50 x 200 kHz (panel D). The color scale is the same for all matrices.

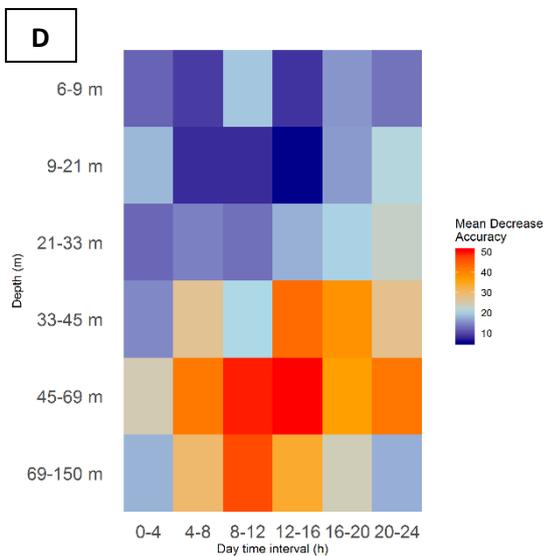
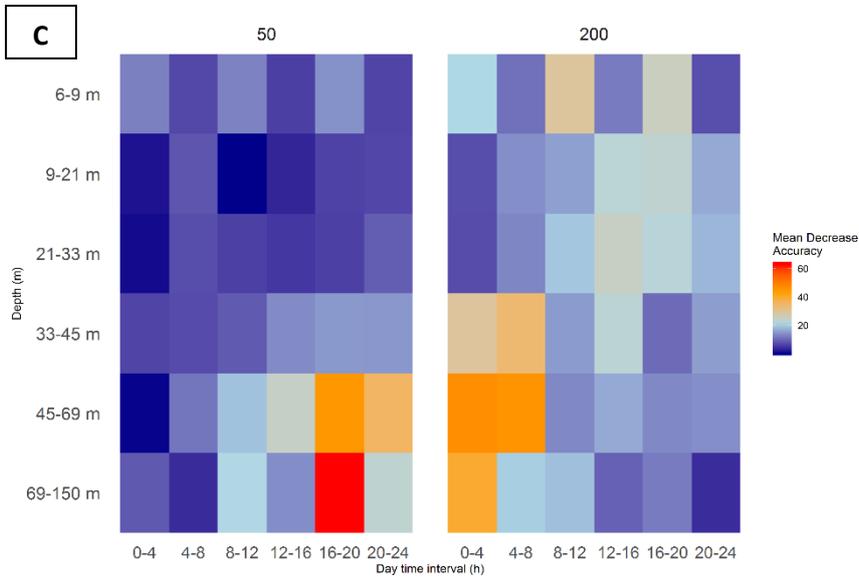
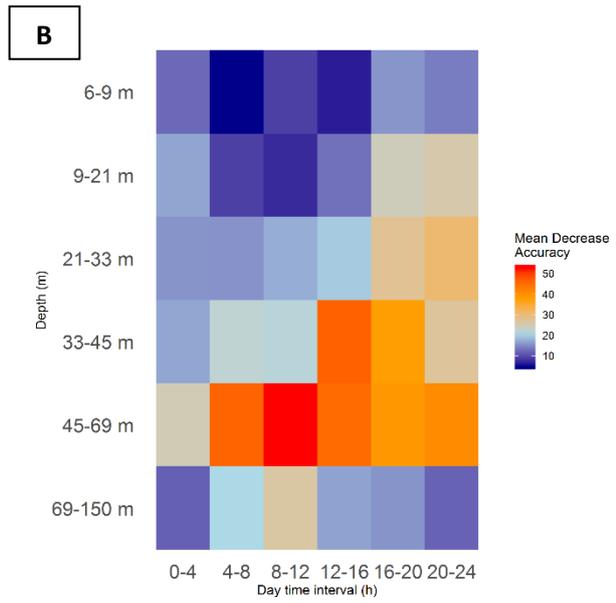
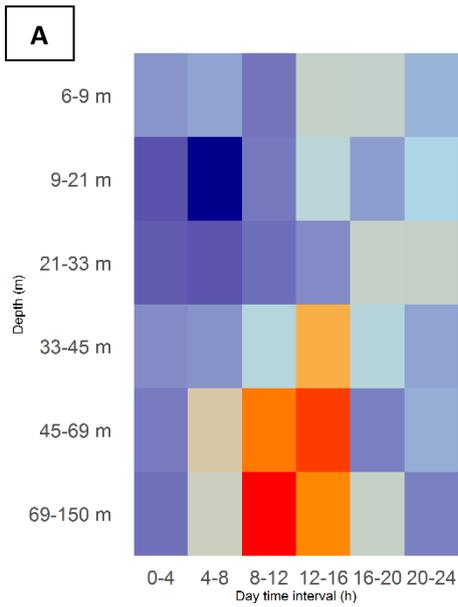


Figure 9: Figure 5: Importance of layers for binary classification random forests : 50 kHz (panel A), 200 kHz (panel B), 50 & 200 kHz (panel C) and 50 x 200 kHz (panel D).

5.1.2.3 Multiclass random forests : size of the aggregation

4-classes classification

With a mean accuracy value of 0.47 ± 0.017 , and a mean kappa of 0.30 ± 0.022 , the different combination of frequencies (Table 3) showed similar performances from one combination of frequencies to another (Table 5). Overall, multiclass models performances were lower than the binary classification. Corresponding confusion matrices (Figure 10) showed good results for extreme classes (Null and Higher 25), but medium classes had poorer performances. The variable importance analysis (Figure 11) revealed similar patterns to the binary classification during the daytime, accompanied by higher weights on the acoustic scores recorded at both shallow and deep layers during the nighttime.

Performance metrics	50 kHz	200 kHz	50 & 200 kHz	50 x 200 kHz
Accuracy	0.48 ± 0.04	0.45 ± 0.03	0.47 ± 0.03	0.49 ± 0.03
Cohen's Kappa	0.31 ± 0.05	0.27 ± 0.04	0.30 ± 0.04	0.32 ± 0.04

Table 5: Summary of performance metrics for the various combination of frequencies tested for 4-classes random forests.

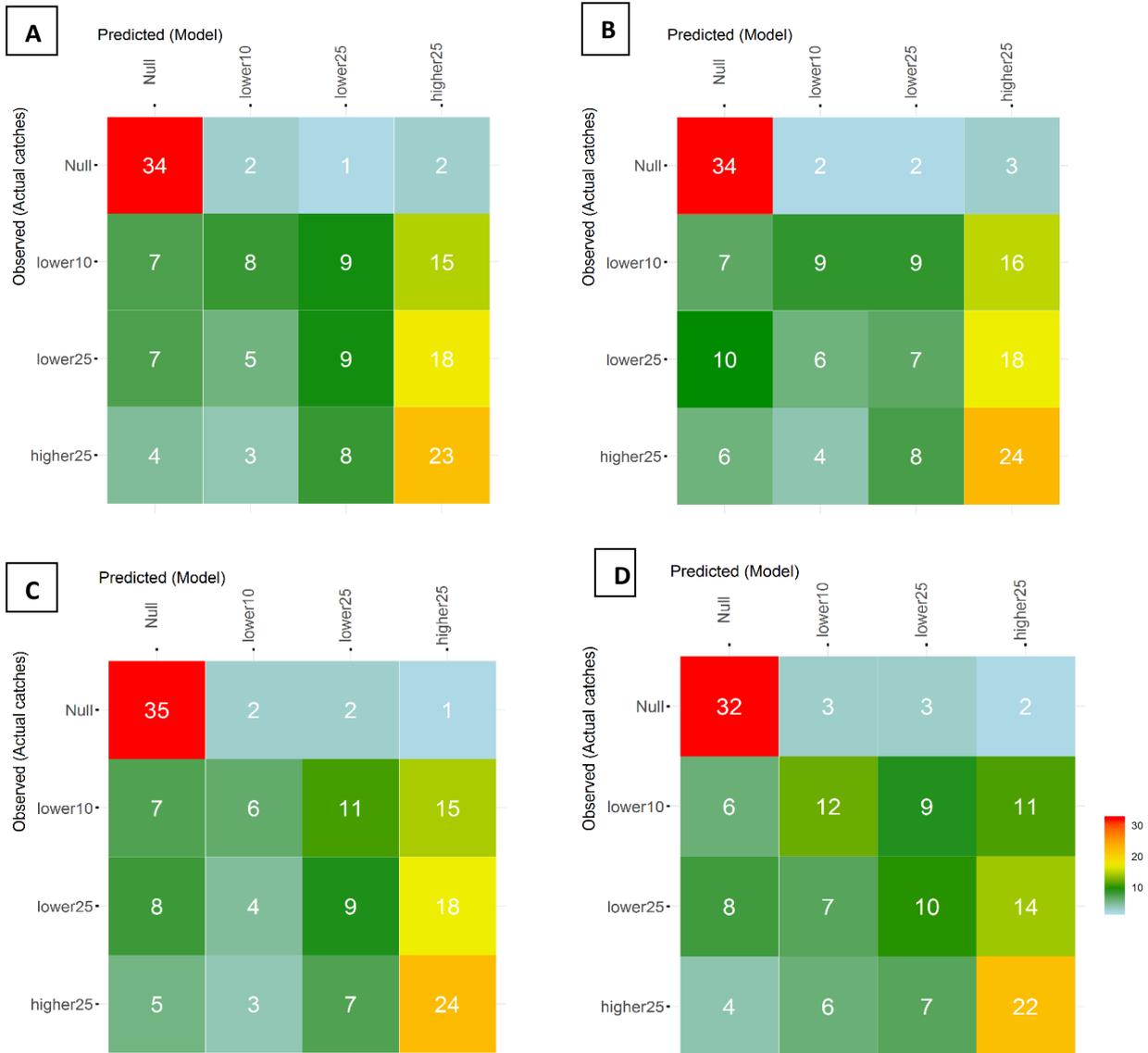


Figure 10: Confusion matrices for 4-classes multiclass random forests. 50 kHz (panel A), 200 kHz (panel B), 50 & 200 kHz (panel C) and 50 x 200 kHz (panel D). The color scale is the same for all confusion matrices.

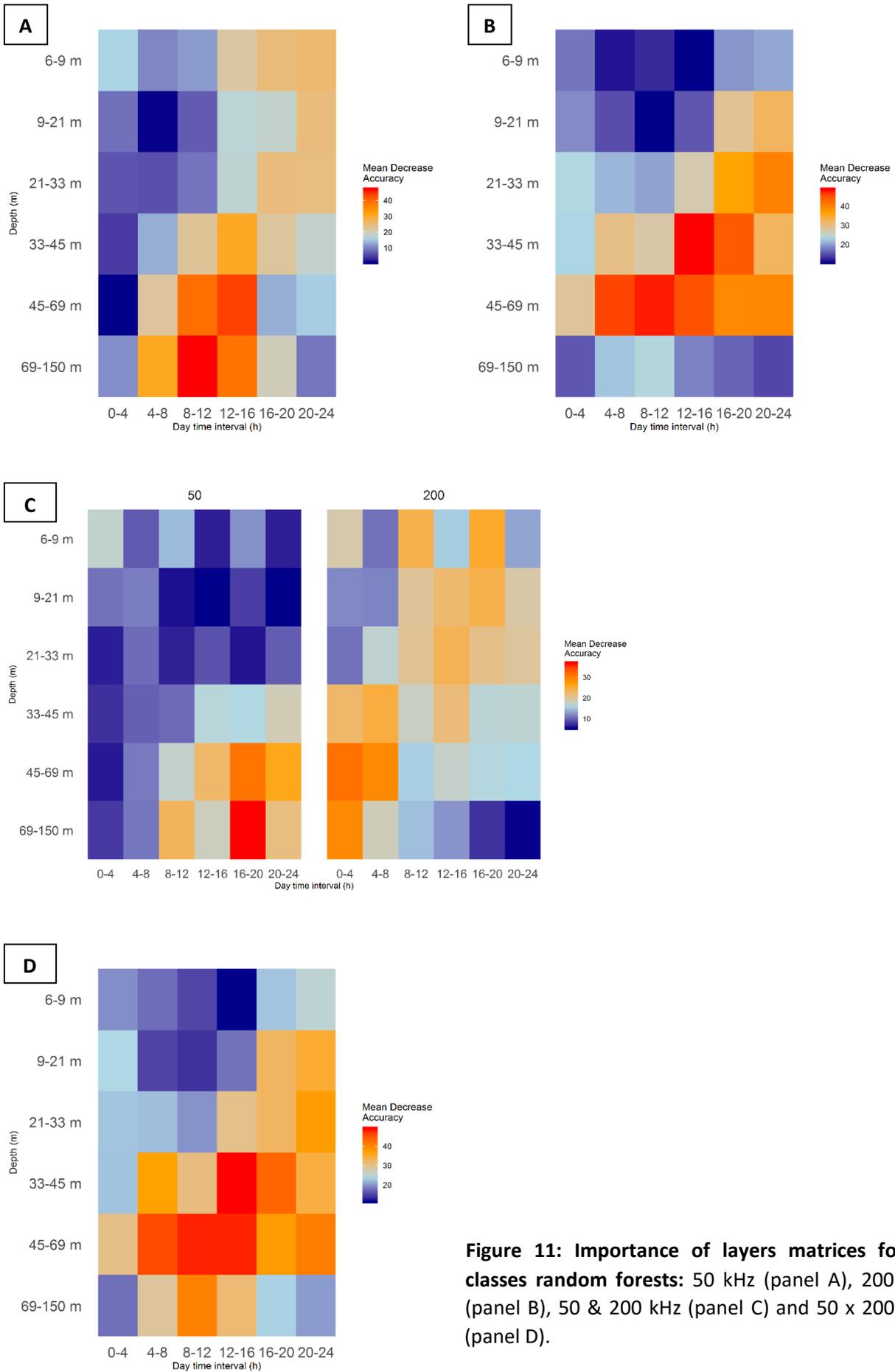


Figure 11: Importance of layers matrices for 4-classes random forests: 50 kHz (panel A), 200 kHz (panel B), 50 & 200 kHz (panel C) and 50 x 200 kHz (panel D).

3-classes classification

Accuracy and Kappa values of 3-classes random forests were improved, compared to the 4-classes, with mean values of respectively 0.62 ± 0.018 and 0.43 ± 0.026 . The differences among performances were very small from one combination of frequencies to another (Table 6). Confusion matrices (Figure 12) showed good results on the diagonal elements, even though the class 'NULL' was the best represented.

Performance metrics	50 kHz	200 kHz	50 & 200 kHz	50 x 200 kHz
Accuracy	0.63 ± 0.04	0.61 ± 0.01	0.64 ± 0.03	0.60 ± 0.03
Cohen's Kappa	0.44 ± 0.06	0.42 ± 0.02	0.46 ± 0.04	0.40 ± 0.04

Table 6: Summary of main performance metrics for the various combination of frequencies tested for 3-classes random forests.

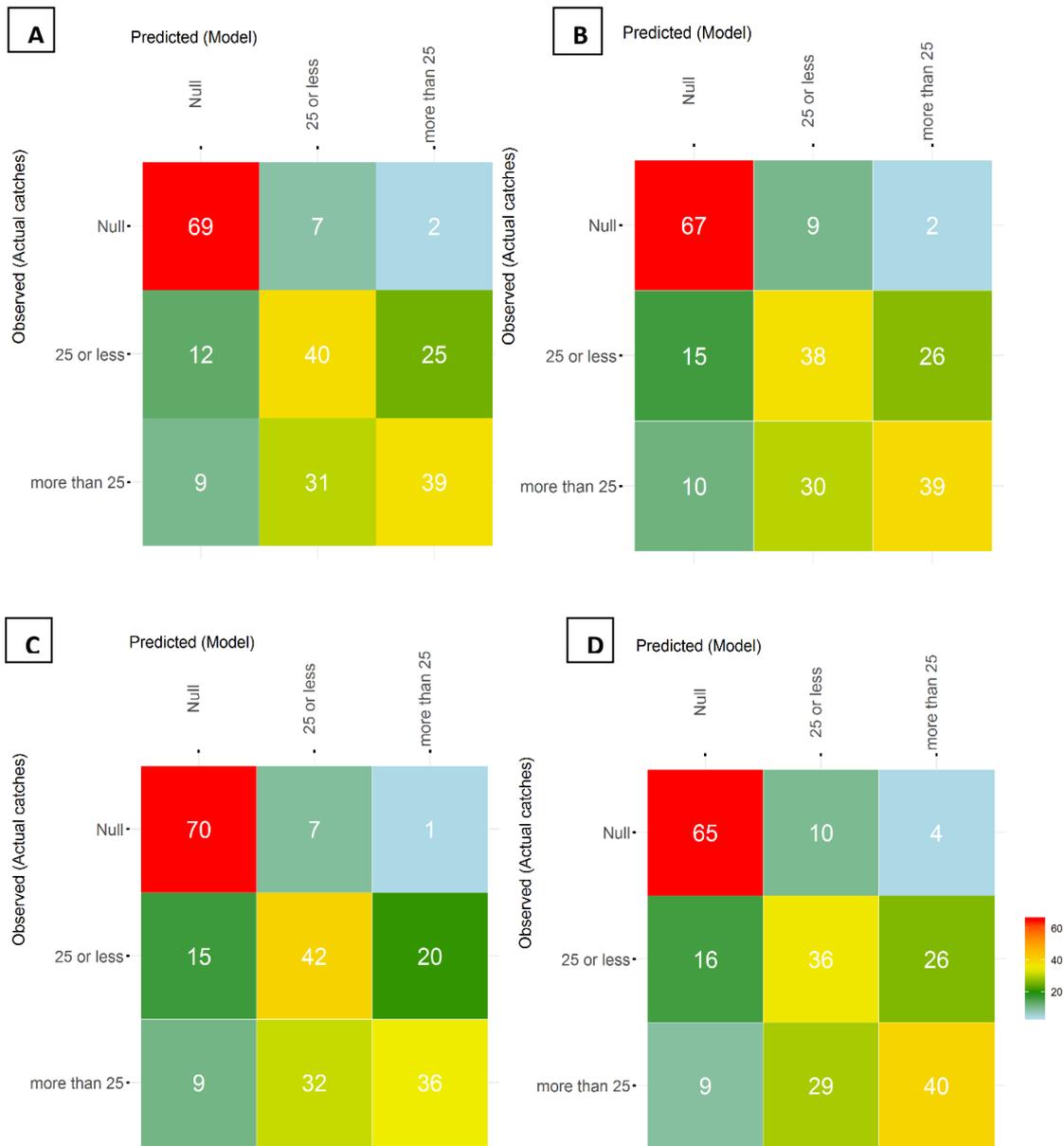


Figure 12: Confusion matrices for 3-classes random forests for 50 kHz panel A, 200 kHz panel B, 50 and 200 kHz combined panel C and their multiplication panel D performed with all tuna species dataset. The color scale is the same for four confusion matrices.

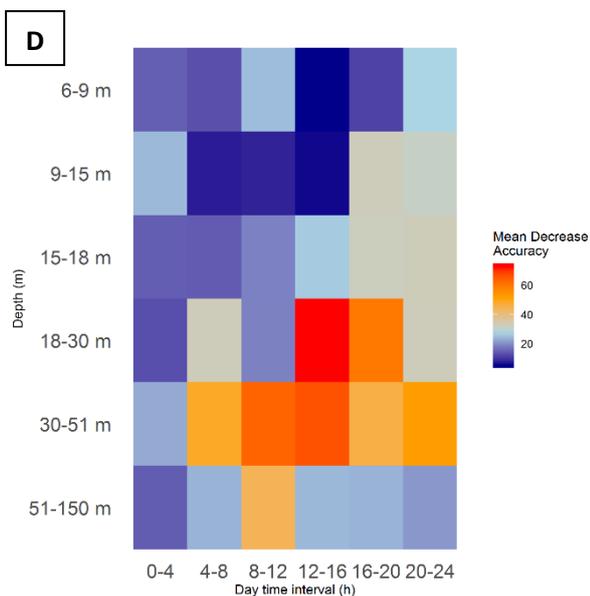
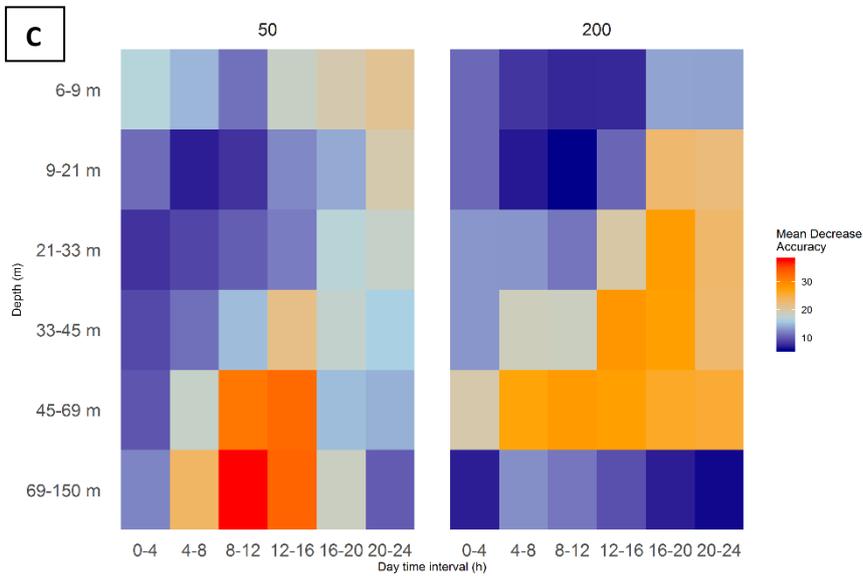
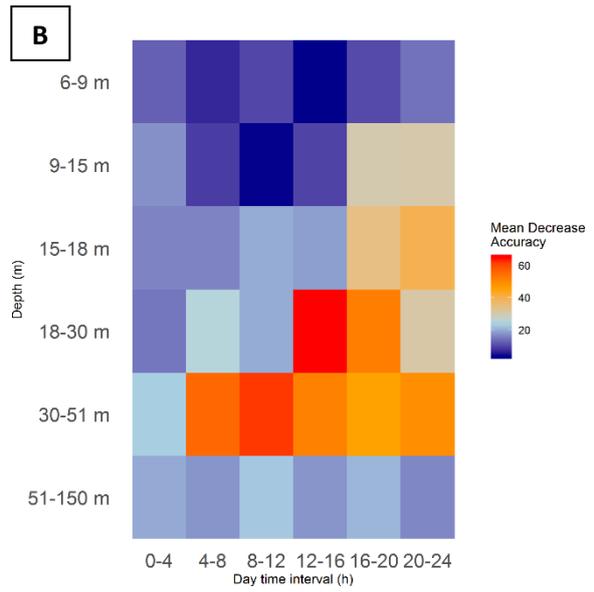
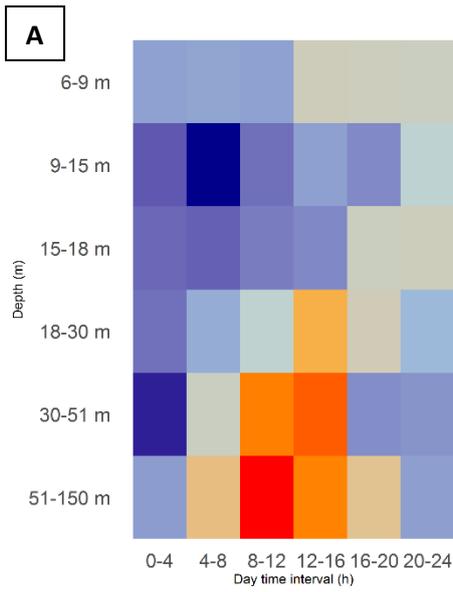


Figure 13: Importance of layers matrices for 3-classes random forests : 50 kHz (panel A), 200 kHz (panel B), 50 & 200 kHz (panel C) and 50 x 200 kHz (panel D).

5.2 GLM

Generalized Linear Models (GLMs) were run considering as dependent variables the tuna catches (including zeros for visits and deployments) and as explanatory variables the acoustic scores of the daily 6x6 acoustic matrices defined in 5.1.

The GLMs were run using the `glm` function of the R package “stat”. A gaussian family with an identity link function was initially used. Since the catch data exhibited overdispersion, quasi-Poisson and tweedie distribution families with log link functions were also used.

The predictive capability of each model was then measured by computing the coefficient of determination (r^2) between the predicted and actual catches. All GLMs were built considering separately the two sampling frequencies.

Coefficients of determinations (Table 7) were overall low for all GLMs and frequencies, indicating low predictive capability. The highest coefficients of determination were obtained for Gaussian GLMs. However, these models did not account for overdispersion and predicted negative catches (see reliability boxplots in Figure 14). Despite their lower predictive capabilities, quasi-Poisson and Tweedie GLMs predicted positive catches by definition (Figures 15, 16). We note extreme high catches predicted by Tweedie GLMs (Figure 16). Across all GLMs and frequencies, acoustic variables corresponding to depth layers below 30 m and time of day between 4 and 12 am, were significant, indicating some consistency across the models.

Frequency (kHz)	Gaussian GLMs	Quasi-Poisson GLMs	Tweedie GLMs
50	0.18	0.09	0.04
200	0.16	0.10	0.06

Table 7: Coefficients of determination of GLMs (r^2).

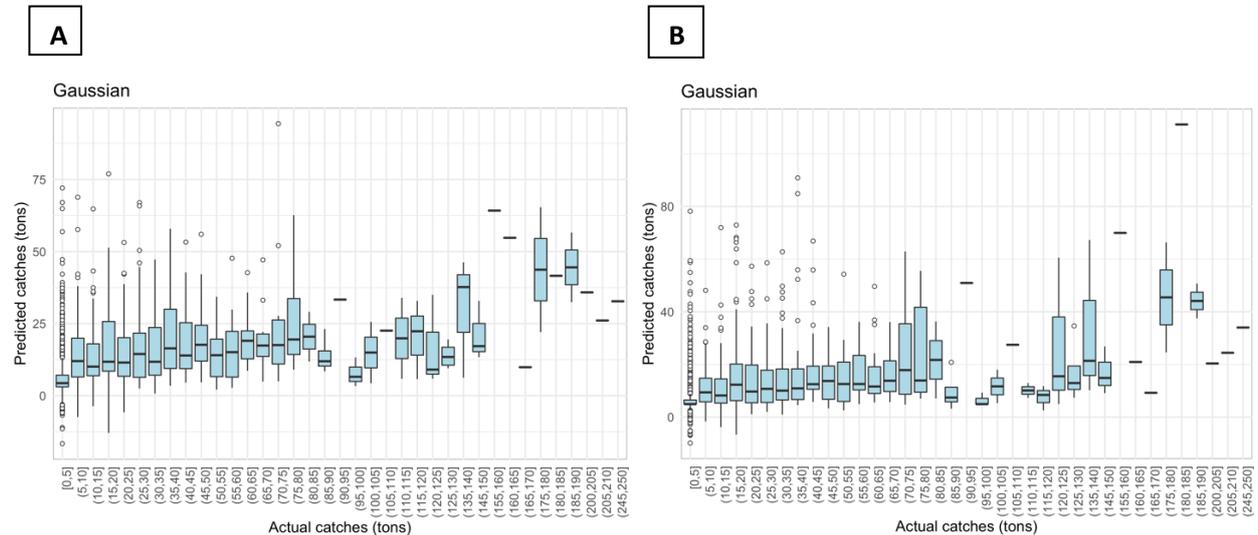


Figure 14: Reliability boxplots for Gaussian GLMs. Actual catches versus predicted catches from acoustic samples obtained at 50 kHz (panel A) and 200 kHz (panel B).

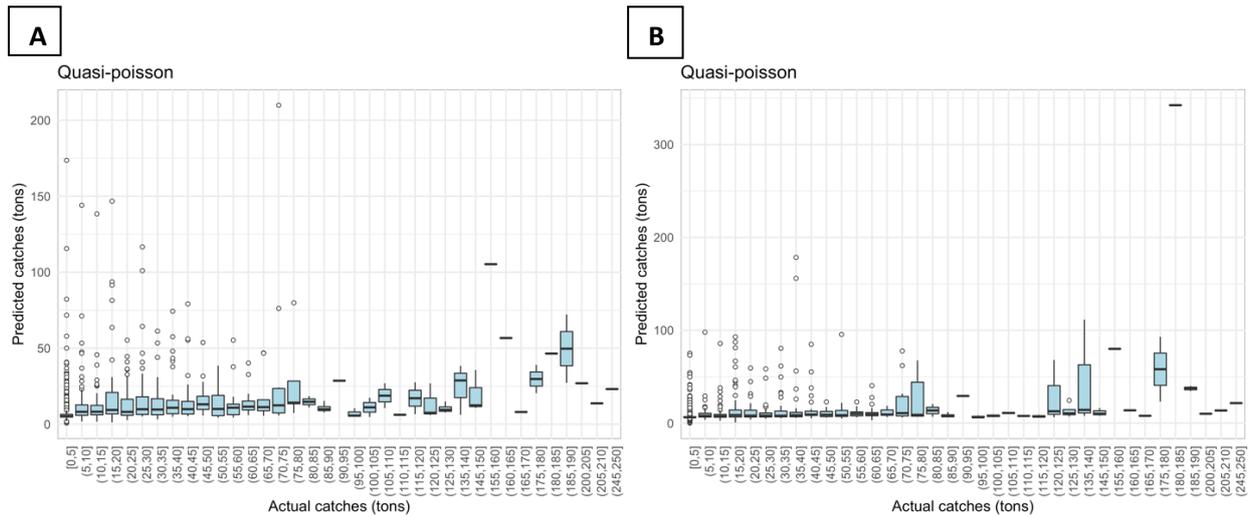


Figure 15: Reliability boxplots for quasi-Poisson GLMs. Actual catches *versus* predicted catches from acoustic samples obtained at 50 kHz (panel A) and 200 kHz (panel B).

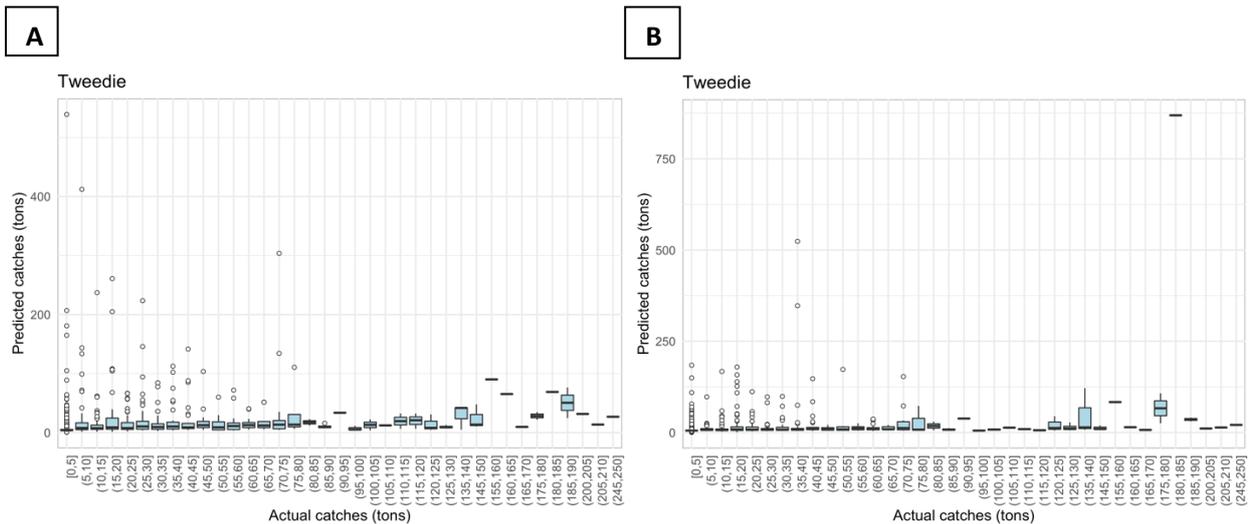


Figure 16: Reliability boxplots for Tweedie GLMs. Actual catches *versus* predicted catches from acoustic samples obtained at 50 kHz (panel A) and 200 kHz (panel B).

5.3 Deep Learning

5.3.1 Algorithm description

We assessed the performances of biomass estimates obtained using the Deep Learning algorithms of the R package “keras”. In order to compare the performances of the deep learning algorithms with the results obtained with Random Forest and GLM, the algorithms were run on the 6x6 daily acoustic matrices defined in 5.1, for 50 kHz and 200 kHz. We used a sequential neural network model, constituted by four linear stacks of layers. The first layer was constituted by 36 neurons (the so-called input shape), equivalent to the number of explanatory variables (i.e., the elements of the 6x6 matrix). This layer was followed by two hidden layers of respectively 256 neurons and 128 neurons, activated through the “relu” activation function. The output layer, for the binary model, was constituted by 2 neurons (presence and absence). The output layer was

connected to the second hidden layer through the “softmax” activation function. A dropout rate of 0.4 was applied on the two hidden layers to reduce the number of parameters. The network was trained using the RMSprop optimization algorithm with the loss function “binary_crossentropy”. The training dataset corresponded to 80% of the total sample and the remaining 20% constituted the test dataset. The model was first trained on 50 epochs. The final model parameters were chosen considering 20 epochs. This value, chosen upon the trend of the validation metrics (see e.g. Figures 17), provided the optimum values of the loss and accuracy that avoided overfitting.

5.3.2 Results

5.3.2.1 Binary classification : Presence/Absence

The evolution of the loss and accuracy for the model run as a function of the training epochs for the 50 kHz and 200 kHz data is shown in Figure 17. The best model run on the test database, showed performances slightly lower than the Random Forest, with an accuracy of respectively 0.84 and 0.81 for 50 kHz and 200 kHz.

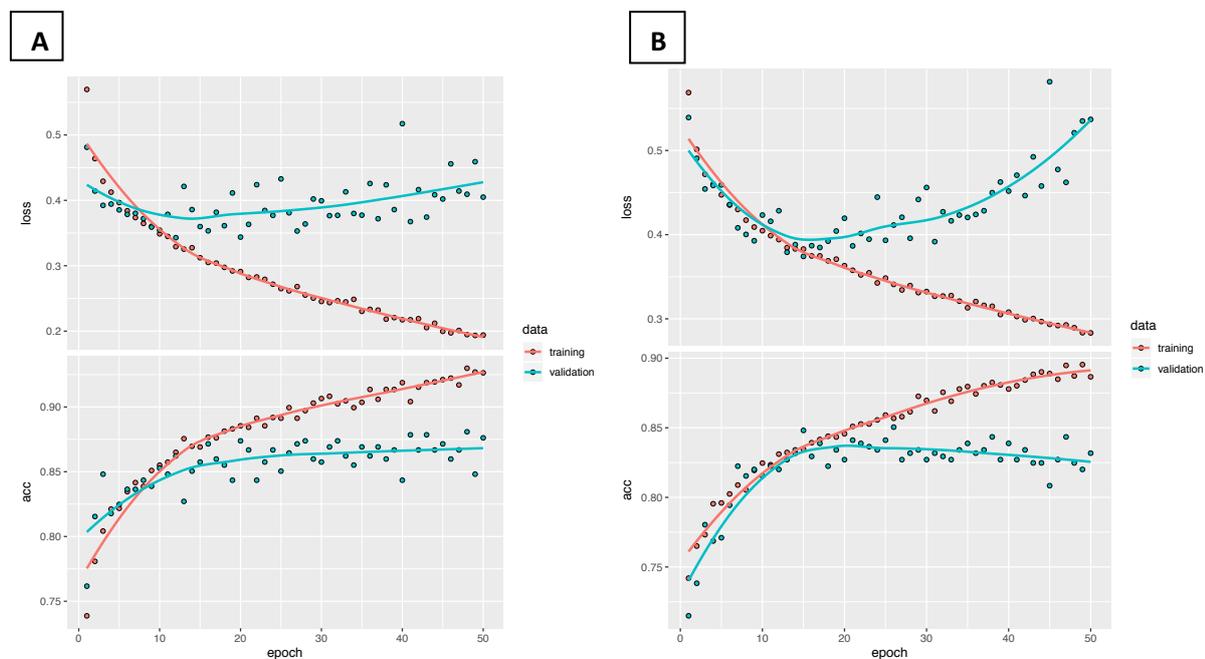


Figure 17: Loss function (top) and accuracy (bottom) of the estimation of tuna presence/absence obtained through the deep learning algorithm as a function of the training epochs. The red line corresponds to the training dataset and the blue line to the validation dataset. Panel A shows the results for 50 kHz and panel B for 200 kHz.

6. Comparison of Metrics

6.1 Metrics

We compared different metrics (colonization time, time between two aggregations and lifetime of an aggregation) obtained from the acoustic data recorded by two different buoy models: M3I+ and M3I. For the M3I+, the choice of the classification algorithm was based on the results obtained in Section 5. Accordingly, the random forests binary classification algorithm based on the 50 kHz acoustic data was chosen, since it demonstrated a high accuracy (0.88) for the assessment of the presence/absence of tuna. For the M3I model, we used the random forests binary classification algorithm developed by Baidai et al. (2018). In order to estimate the three

metrics, we considered the acoustic data corresponding to newly-deployed FADs instrumented by M3I and M3I+ buoys in the Indian ocean in the period 2016-2018. Only buoys with a soaking time of at least 30 days were considered in the analysis. Table 8 provides the number of buoys considered.

Buoy model	Number of buoys
M3I	5748
M3I+	1368

Table 8: Number of newly-deployed FAD-buoy pairs used for the calculation of the metrics, for each buoy model.

The M3I and M3I+ random forest binary classification models were applied to the trajectories of all newly-deployed FAD-buoy pairs, providing the time series of tuna presence/absence for each pair. Despite the high classification accuracy of the random forest models, it was important to consider that the algorithms could still erroneously classify the presence/absence of tuna over a subset of days. In order to avoid this bias, we post-processed the presence/absence time series considering that, if a presence (or absence) event lasted only 1 day, this was due to a classification error. Therefore, isolated presence/absence events (lasting only 1 day) were removed from the time series.

As a proxy of colonization times, we compared the first day of tuna detection recorded by the two buoy models (Figure 18A). For both buoy models, we found a high variability. However, the first day of tuna detection for M3I (average=23.0 days, SD= 22 days) were generally longer than M3I+ (average=17.7 days, SD=18 days). The Mann-Whitney test of comparison confirmed the differences between the two buoy models (p-value = 7.9 e-13). Moreover, only 8.55% of the M3I+ buoy models were never colonized within the monitoring period, in contrast with 17.66% for the M3I buoys. In order to explain these differences among buoy models, we considered the fraction of buoys available in the dataset as a function of the number of days after deployment. However, both buoy model datasets showed similar soaking times, see Figure 18B, revealing that this factor could not affect our results.

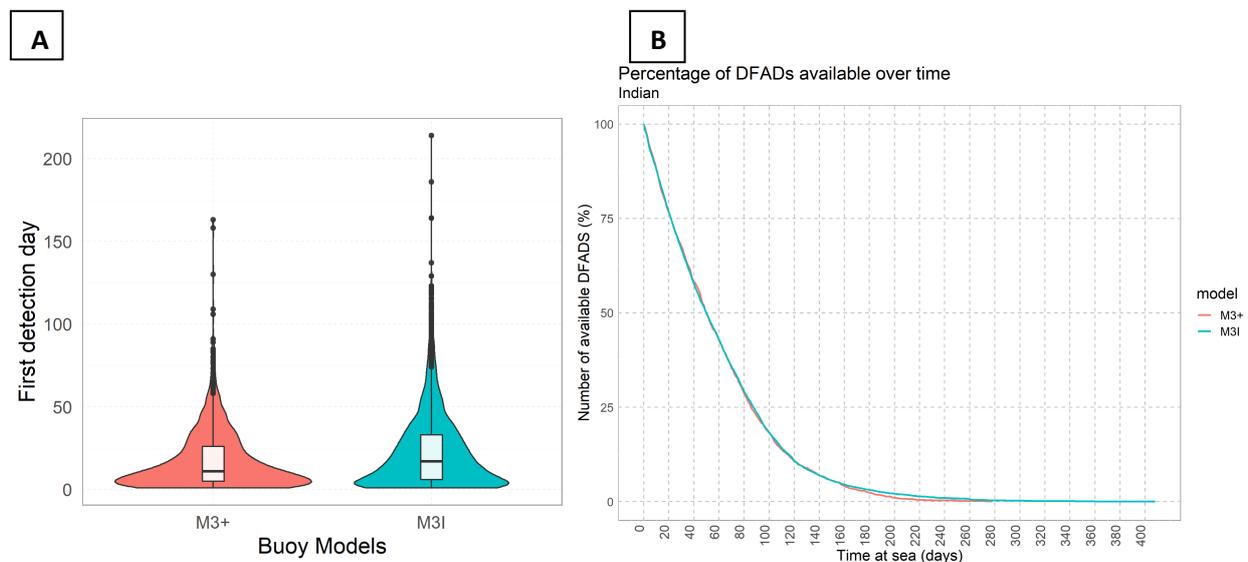


Figure 18. Panel A: First day of tuna detection for M3I+ and M3I buoys. Panel B: Percentage of buoys in the water as a function of the number of days after deployment.

The lifetime of a tuna aggregation (named herein aCRT, defined as the total number of consecutive days of tuna presence at a FAD), and the time between two aggregations (named aCAT, defined as the total number of consecutive days between two tuna aggregations at a FAD) are shown in Figure 19. Both quantities also demonstrated a high variability.

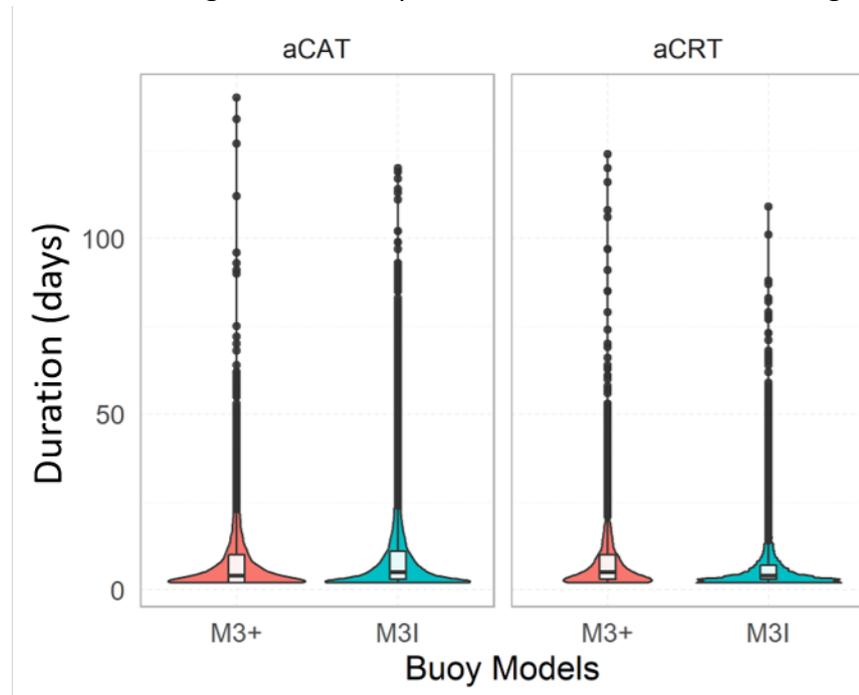


Figure 19: Lifetime of tuna aggregations (aCRT, left) and time between two aggregations (aCAT, right) for M3I+ and M3I buoys.

However, the M3I buoys showed smaller average aCRTs (average= 6 days, SD=6.6 days) relative to the M3I+ (average=8.4 days, SD=10.9 days). The Mann-Whitney test of comparison confirmed the differences among buoy models (p -value = 2.57×10^{-58}). Reversely, average aCATs were higher for M3I (average=9.7 days, SD=12 days) than for M3I+ (average=8.3 days, SD=9.9 days). The Mann-Whitney test of comparison confirmed the differences among buoy models (p -value = 1.38×10^{-9}).

Finally, Figure 20 provides the kernel density of the positions recorded for the two buoy models, demonstrating that the M3I and M3I+ buoys covered a similar region.

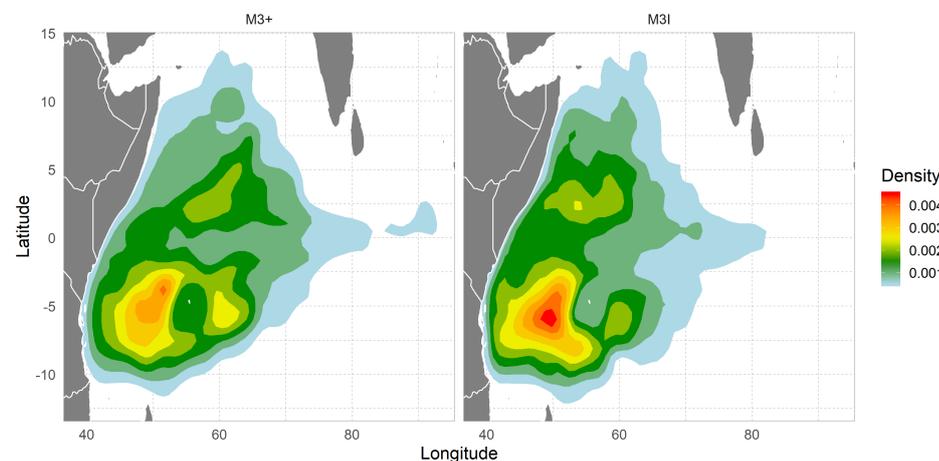


Figure 20: Kernel density of the positions of the M3I and M3I+ buoys considered in the analysis.

6.2 Accuracy of presence/absence classification algorithms as a function of the aggregation size for M3I and M3I+

We tested whether the dissimilarity between the metrics obtained in 6.1 for the M3I and M3I+ buoys could be explained in terms of the higher sensibility of the M3I+ buoys to detect small tuna aggregations relative to the M3I model. To this purpose, we evaluated the predictive performances of the M3I and M3I+ binary classification models on a new dataset, constituted by acoustic and catch data recorded in the Indian ocean in 2019. This dataset corresponded to 326 sets conducted on M3I buoys and 155 sets on M3I+ buoys. This choice ensured a high number of new catch data independent from those used for training the algorithms. Catch and acoustic data corresponding to the same buoy IDs were combined and filtered using the same procedure described in section 4.3. The random forest classification algorithms optimized for the M3I+ (binary classification on 50 kHz, see section 5) and M3I (Baidai et al. 2018) were then run to classify the acoustic data provided by each buoy one day prior to the set into tuna presence/absence. As such, the catch data, considered as a proxy of the size of the tuna aggregation, were used to ground-truth the performance of the algorithms to assess tuna presence for different aggregation sizes. Figure 21 provides the percentage of correctly classified tuna presences as a function of the tuna aggregation size for the two buoy models. For tuna aggregations >25 tons, the performances of the two buoys were the highest and showed similar values (respectively 90% and 92% of correctly classified tuna presence for M3I and M3I+). However, for smaller aggregation the performances of the M3I buoys were lower than the M3I+ (particularly for size classes between 5 and 25 tons).

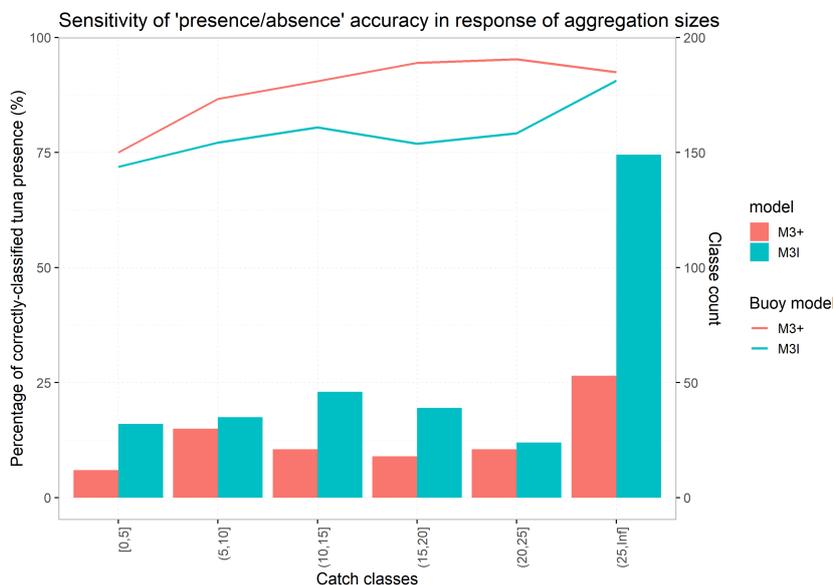


Figure 21. Percentage of correctly-classified tuna presence for each buoy model as a function of the tuna aggregation size (lines, left axis) and number of sets used for each class (bars, right axis).

7. Discussion

The use of multi-frequency echosounder buoys has been rapidly increasing in recent years. Since the acoustic response of different tuna species depends on the frequency (Boyra et al, 2018, Moreno et al. 2019), multi-frequency buoys could offer a promising tool to discriminate the tuna biomass and tuna species more accurately. However, no scientific studies so far evaluated the accuracy of these buoys for estimating the tuna biomass at FADs.

The specific objective of this study was to assess the accuracy for biomass estimates conducted by M3I+ buoys, one of the main buoy models deployed by the industrial purse seiners on drifting FADs. Thanks to the large echosounder and activity dataset at our disposal, that allowed matching the acoustic data with field observations of tuna biomass (including tuna absence) for nearly 2,700 activity events, we could evaluate the performances of different algorithms to estimate the accuracies of this buoy model at two frequencies (50 and 200 kHz).

A first inspection of the data based on the simple averaged acoustic scores recorded at different layers and times of the day, revealed clear differences between the FADs without tuna and those with tuna, both for the 50 and 200 kHz frequencies (Figure 4 and 5). Reversely, the average of the acoustic data recorded for tuna catches of less than 25 tons were very close to those found for more than 25 tons. For tuna catches of less than 25 tons and more than 25 tons, similar temporal and vertical patterns were found among frequencies. However, clear differences appeared in the strength of the acoustic scores depending on the frequency, with lower values for 200 kHz relative to those recorded at 50 kHz. These differences may be explained in terms of the smaller acoustic cone of the 200 kHz frequency (8°) relative to the cone of 50 kHz (36°), i.e., in terms of a smaller amount of acoustic energy received by the former and the more reduced chances of detecting the tuna school for a smaller acoustic cone.

Remarkably, the results of all algorithms appeared to be independent on the frequency used. This implies that, despite the lower average magnitude, the acoustic scores recorded at 200 kHz are as effective as those recorded at 50 kHz to identify tuna presence and biomass. In order to verify if larger differences among frequencies would appear for a specific tuna species, we have also run the Random Forest algorithms considering the catches of skipjack only (see the Appendix 2). Indeed, skipjack, which is the major tuna species caught at FADs, shows a higher acoustic response at higher frequencies (Boyra et al. 2019). However, our results showed that, even for skipjack, both frequencies provided the same accuracies.

The random forest algorithms run for assessing presence/absence of tuna at the FADs showed the highest accuracies for both frequencies. All performance metrics of the random forest algorithm found for M3I+ (see Appendix 1 for all metrics) were very close to those found for M3I in previous studies (Baidai et al. 2018).

The importance of variables for the M3I+ buoys (Figures 9, 11, 13) demonstrated high capabilities for detecting tuna in the deeper layers. This result may be a consequence of the improved software and hardware characteristics for the M3I+ buoy model. Nevertheless, the accuracies of the multiclass classification were only slightly higher than those found for the M3I buoys in previous studies (Baidai et al. 2018). Similarly to M3I buoys, the M3I+ can best discriminate tuna absences and aggregations larger than 25 tons (Figure 10). Reversely, this buoy cannot clearly discriminate the size of small aggregations (10 tons and less than 25 tons). The use of 3 classes (0 tons, < 25 tons and > 25 tons), provided a better accuracy but the confusion matrices (Figure 12) still demonstrated that it is very hard to discriminate the size of a tuna aggregation from the acoustic signal provided by the buoys. One possible factor explaining the low algorithm performances for assessing the size of the tuna aggregations may rely in the limited number of observations of tuna catches constituting the different classes. In this respect, larger catch

databases may improve the algorithm performances for estimating the tuna aggregation sizes. On the other hand, it is important to notice that the acoustic cone of the buoy can only sample a subportion of the aggregation and this data could be intrinsically limited. Considering a tuna school located between 30 m and 70 m depth (Forget et al. 2015), the diameter of the acoustic cone would range between 20-45m at 50 kHz and 4-10m at 200 kHz, respectively. These ranges, together with the fact that tuna is a circumnatanant species (Castro et al 2002) located in the proximity of the FAD but not necessarily right underneath, suggest that it is likely that the buoys detect only a subportion of the whole tuna aggregation.

Beside the Random Forest, other algorithms were tested to evaluate the accuracy of the M3I+ echosounder buoys. The GLMs demonstrated poor correlations between the estimated biomass and the actual catch. Reversely, the Deep Learning algorithm provided, for the presence-absence classification at 50 kHz, similar accuracies to those found for the Random Forest. Preliminary runs of the Deep Learning algorithms were also conducted on larger acoustic daily matrices (12x50), where the acoustic data recorded each depth layer (3m) and time bin (2 hours) were provided as an input variables. However, the performances of the Deep Learning algorithm on this 12x50 matrix were lower than those found using a 6x6 daily acoustic matrix. This result may be explained in terms of model parsimony, i.e., an algorithm based on a reduced number of parameters can provide higher performances than an overparametrized algorithm. It also implies that the preprocessing of the 12x50 acoustic matrix into a 6x6 matrix is a key step to attain higher accuracies for all algorithms.

This study constitutes a milestone for the derivation of novel abundance indices of tropical tuna based on the echosounder buoys data. In this framework, evaluating the accuracy of biomass estimates is a key step prior to any scientific study that exploits this data.

We found that the most accurate information provided by the buoys is tuna presence/absence at the FADs. However, the comparison of the aggregation metrics based on presence/absence data (colonization times, lifetime of an aggregation, time between two aggregations) revealed a sensitivity of the results with respect of the buoy model. Overall, the M3I+ buoys provided smaller colonization times, longer lifetimes of the aggregation and shorter times between two aggregations with respect to those found for the M3I. Our results are consistent with the hypothesis that the two buoy models have a different sensitivity to detect tuna aggregations. Namely, the M3I+ buoys can detect the presence of small tuna aggregations (< 25 tons) better than the M3I. These findings highlight the importance of assessing the accuracy of the biomass estimates for each buoy model. They also underline the need to carefully interpret the results obtained for any buoy-related index across different years, since the proportion of each buoy model used by the purse-seiners evolves in time. Further studies, accounting for the different sensitivity of the buoys to detect tuna aggregations into the aggregation metrics, should address the issue of how to merge the data obtained from different buoy models to construct novel fisheries-independent indicators of abundance for tropical tuna.

Acknowledgements

We thank Laurent Dagorn for the interesting discussions. This project was funded by the Food and Agriculture Organization (FAO). The work of Y.B was also funded by the ANR project BLUEMED

(ANR-14-ACHN-0002). The authors are grateful to Orthongel and its contracting parties (CFTO, SAPMER, SAUPIQUET) for providing the echosounder buoys data. The authors also thank all the skippers who gave time to share their experience and knowledge on the echosounder buoys. The authors sincerely thank the contribution of the staff of the Ob7 on the databases of the echosounder buoys and observers' data. We are also grateful to the buoy manufacturers for their useful advice and information on the echosounder buoys.

References

Baidai Y, Capello M, Amandé J, Billet N, Floch L, Simier M, Sabarros, P and Dagorn L. (2017) Towards the derivation of fisheries-independent abundance indices for tropical tunas : progress in the echosounders buoys data analysis. IOTC-2017-WPTT19-22

Baidai, Y., Amande, M.J., Gaertner, D., Dagorn, L., Capello, M., (2018). Recent advances on the use of supervised learning algorithms for detecting tuna aggregations under fads from echosounder buoys data, IOTC-2018-WPTT20-25_Rev1. Mahé.

Boyra, G., Moreno, G., Sobradillo, B., Pérez-Arjona, I., Sancristobal, I., Demer, D.A., Ratilal, He.P., 2018. Target strength of skipjack tuna (*Katsuwonus pelamis*) associated with fish aggregating devices (FADs). *ICES J. Mar. Sci.* 75, 1790–1802.

Capello, M., et al. "Population assessment of tropical tuna based on their associative behavior around floating objects." *Sci Rep.* 6 (2016): 36415.

Cohen, J., 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol. Bull.* 70, 213–220. <https://doi.org/10.1037/h0026256>

Dagorn, L., Holland, K.N., Restrepo, V., Moreno, G., 2013. 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, 391–415.

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.

Forget, F.G., Capello, M., Filmlalter, J.D., Govinden, R., Soria, M., Cowley, P.D., Dagorn, L., 2015. Behaviour and vulnerability of target and non-target species at drifting fish aggregating devices (FADs) in the tropical tuna purse seine fishery determined by acoustic telemetry. *Can. J. Fish. Aquat. Sci.* 72, 1398–1405. <https://doi.org/10.1139/cjfas-2014-0458>

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R news* 2, 18–22. <https://doi.org/10.1177/154405910408300516>

Lopez, J., et al. "A model based on data from echosounder buoys to estimate biomass of fish species associated with fish aggregating devices." *Fishery Bulletin.* 114.2 (2016): 166–178.

Lopez, J., Moreno, G., Ibaibarriaga, L., Dagorn, L., 2017a. Diel behaviour of tuna and non-tuna species at drifting fish aggregating devices (DFADs) in the Western Indian Ocean, determined by fishers' echo-sounder buoys. *Mar. Biol.* 164, 44. <https://doi.org/10.1007/s00227-017-3075-3>

Moreno, G., et al. "Fish aggregating devices (FADs) as scientific platforms." *Fish Res.* 178 (2016): 122–129.

Moreno, G., Boyra, G., Sancristobal, I., Itano, D., Restrepo, V., 2019. Towards acoustic discrimination of tropical tuna associated with Fish Aggregating Devices. *PLoS One* 14, e0216353. <https://doi.org/10.1371/journal.pone.0216353>

Murtagh, F., Legendre, P., 2014. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J. Classif.* 31, 274–295. <https://doi.org/10.1007/s00357-014-9161-z>

Orue B, Lopez J, Moreno G, Santiago J, Soto M, Murua H. Fishers' echo-sounder buoys to estimate biomass of fish species associated with fish aggregating devices in the Indian Ocean IOTC–2016–WPTT18–28

Orue, B., Lopez, J., Moreno, G., Santiago, J., Boyra, G., Soto, M., Murua, H., 2019. Using fishers' echo-sounder buoys to estimate biomass of fish species associated with drifting fish aggregating devices in the Indian Ocean. *Rev. Investig. Mar. AZTI* 26, 1–13.

Santiago, J., Uranga, J., Quincoces, I., Orue, B., Merino, G., Murua, H., Boyra, G., 2019. A novel index of abundance of juvenile yellowfin tuna in the Atlantic Ocean derived from echosounder buoys. 2nd Jt. t-RFMO FAD Work. Gr. Meet. 24.

APPENDIX 1: Detailed performance metrics obtained for the Random Forest algorithm

1.1. Binary classification : Presence/Absence

Performance metrics	50 kHz	200 kHz	50 & 200 kHz	50 x 200 kHz
<i>Accuracy</i>	0.88 ± 0.01	0.85 ± 0.02	0.86 ± 0.02	0.85 ± 0.02
<i>Cohen's Kappa</i>	0.76 ± 0.03	0.71 ± 0.04	0.73 ± 0.03	0.70 ± 0.04
<i>Sensitivity</i>	0.87 ± 0.02	0.82 ± 0.03	0.84 ± 0.01	0.83 ± 0.03
<i>Specificity</i>	0.90 ± 0.02	0.88 ± 0.03	0.88 ± 0.03	0.87 ± 0.02
<i>Precision</i>	0.89 ± 0.02	0.87 ± 0.02	0.88 ± 0.03	0.87 ± 0.02
<i>F₁ score</i>	0.88 ± 0.01	0.85 ± 0.02	0.86 ± 0.01	0.85 ± 0.02

Table S1: Summary of performance metrics for all four combinations of frequencies tested for binary classification.

1.2. Multiclass classification

1.2.1. 4-classes classification

Performance metrics	50 kHz				
	No tuna	<10 tons	[10-25]tons	> 25 tons	Average
<i>Sensitivity</i>	0.88 ± 0.06	0.23 ± 0.07	0.20 ± 0.09	0.60 ± 0.09	0.48 ± 0.32
<i>Specificity</i>	0.84 ± 0.03	0.84 ± 0.02	0.92 ± 0.04	0.70 ± 0.04	0.83 ± 0.09
<i>Precision</i>	0.65 ± 0.05	0.33 ± 0.05	0.44 ± 0.09	0.40 ± 0.06	0.46 ± 0.14
<i>F₁ score</i>	0.75 ± 0.04	0.27 ± 0.07	0.27 ± 0.08	0.48 ± 0.07	0.44 ± 0.23
<i>Accuracy</i>					0.48 ± 0.04
<i>Cohen's Kappa</i>					0.31 ± 0.05

Table S2: Performance metrics for 50 kHz frequency model tested for 4-classes classification.

Performance metrics	200 kHz				
	No tuna	<10 tons	[10-25]tons	> 25 tons	Average
<i>Sensitivity</i>	0.82 ± 0.08	0.17 ± 0.06	0.23 ± 0.05	0.58 ± 0.07	0.45 ± 0.31
<i>Specificity</i>	0.81 ± 0.04	0.84 ± 0.02	0.91 ± 0.05	0.70 ± 0.03	0.82 ± 0.09
<i>Precision</i>	0.60 ± 0.04	0.27 ± 0.11	0.45 ± 0.06	0.40 ± 0.02	0.43 ± 0.14
<i>F₁ score</i>	0.69 ± 0.03	0.21 ± 0.07	0.30 ± 0.04	0.47 ± 0.04	0.42 ± 0.21
<i>Accuracy</i>					0.45 ± 0.03
<i>Cohen's Kappa</i>					0.27 ± 0.04

Table S3: Performance metrics for 200 kHz frequency model tested for 4-classes classification.

50 & 200 kHz

Performance metrics	No tuna	<10 tons	[10-25]tons	> 25 tons	Average
<i>Sensitivity</i>	0.88 ± 0.05	0.17 ± 0.04	0.23 ± 0.09	0.62 ± 0.05	0.48 ± 0.34
<i>Specificity</i>	0.82 ± 0.02	0.93 ± 0.03	0.84 ± 0.02	0.71 ± 0.03	0.83 ± 0.09
<i>Precision</i>	0.63 ± 0.03	0.46 ± 0.15	0.32 ± 0.09	0.41 ± 0.02	0.46 ± 0.13
<i>F₁ score</i>	0.73 ± 0.02	0.24 ± 0.06	0.27 ± 0.09	0.49 ± 0.02	0.43 ± 0.23
<i>Accuracy</i>					0.47 ± 0.02
<i>Cohen's Kappa</i>					0.30 ± 0.03

Table S4: Performance metrics for 50 & 200 kHz model tested for 4-classes classification.

50 x 200

Performance metrics	No tuna	<10 tons	[10-25]tons	> 25 tons	Average
<i>Sensitivity</i>	0.82 ± 0.06	0.25 ± 0.04	0.32 ± 0.06	0.56 ± 0.08	0.49 ± 0.26
<i>Specificity</i>	0.84 ± 0.04	0.84 ± 0.04	0.87 ± 0.04	0.77 ± 0.03	0.83 ± 0.04
<i>Precision</i>	0.64 ± 0.05	0.35 ± 0.05	0.45 ± 0.09	0.44 ± 0.02	0.47 ± 0.12
<i>F₁ score</i>	0.71 ± 0.04	0.29 ± 0.04	0.37 ± 0.06	0.49 ± 0.04	0.47 ± 0.18
<i>Accuracy</i>					0.47 ± 0.03
<i>Cohen's Kappa</i>					0.30 ± 0.04

Table S5: Performance metrics for 50 x 200 kHz model tested for 4-classes classification.**1.2.1. 3-classes classification**

50 kHz

Performance metrics	No tuna	< 25 tons	> 25 tons	Average
<i>Sensitivity</i>	0.88 ± 0.03	0.51 ± 0.06	0.49 ± 0.06	0.63 ± 0.22
<i>Specificity</i>	0.86 ± 0.03	0.76 ± 0.04	0.82 ± 0.02	0.81 ± 0.05
<i>Precision</i>	0.76 ± 0.05	0.51 ± 0.06	0.58 ± 0.04	0.62 ± 0.13
<i>F₁ score</i>	0.82 ± 0.03	0.51 ± 0.06	0.53 ± 0.04	0.62 ± 0.17
<i>Accuracy</i>				0.63 ± 0.04
<i>Cohen's Kappa</i>				0.44 ± 0.06

Table S6: Performance metrics for 50 kHz model tested for 3-classes classification.

<i>200 kHz</i>				
Performance metrics	No tuna	< 25 tons	> 25 tons	Average
<i>Sensitivity</i>	0.86 ± 0.04	0.48 ± 0.02	0.50 ± 0.05	0.61 ± 0.21
<i>Specificity</i>	0.84 ± 0.02	0.75 ± 0.02	0.83 ± 0.02	0.81 ± 0.05
<i>Precision</i>	0.73 ± 0.03	0.49 ± 0.02	0.59 ± 0.03	0.60 ± 0.12
<i>F₁ score</i>	0.79 ± 0.03	0.49 ± 0.02	0.54 ± 0.04	0.61 ± 0.16
<i>Accuracy</i>	0.61 ± 0.01			
<i>Cohen's Kappa</i>	0.42 ± 0.02			

Table S7: Performance metrics for 200 kHz model tested for 3-classes classification.

<i>50 & 200 kHz</i>				
Performance metrics	No tuna	< 25 tons	> 25 tons	Average
<i>Sensitivity</i>	0.90 ± 0.03	0.55 ± 0.03	0.46 ± 0.07	0.64 ± 0.23
<i>Specificity</i>	0.85 ± 0.02	0.75 ± 0.05	0.87 ± 0.02	0.82 ± 0.06
<i>Precision</i>	0.75 ± 0.03	0.52 ± 0.04	0.63 ± 0.03	0.63 ± 0.12
<i>F₁ score</i>	0.82 ± 0.03	0.54 ± 0.03	0.53 ± 0.03	0.63 ± 0.16
<i>Accuracy</i>	0.64 ± 0.03			
<i>Cohen's Kappa</i>	0.46 ± 0.04			

Table S8: Performance metrics for 50 & 200 kHz model tested for 3-classes classification.

<i>50 x 200 kHz</i>				
Performance metrics	No tuna	< 25 tons	> 25 tons	Average
<i>Sensitivity</i>	0.82 ± 0.03	0.46 ± 0.06	0.51 ± 0.05	0.60 ± 0.20
<i>Specificity</i>	0.84 ± 0.03	0.75 ± 0.04	0.81 ± 0.04	0.80 ± 0.05
<i>Precision</i>	0.72 ± 0.04	0.48 ± 0.04	0.58 ± 0.05	0.59 ± 0.12
<i>F₁ score</i>	0.77 ± 0.03	0.47 ± 0.04	0.54 ± 0.03	0.59 ± 0.16
<i>Accuracy</i>	0.60 ± 0.03			
<i>Cohen's Kappa</i>	0.40 ± 0.04			

Table S9: Performance metrics for 50 x 200 kHz model tested for 3-classes classification.

APPENDIX 2: Random Forest algorithm for estimating the biomass of skipjack

We applied the same Random Forest algorithm described in 5.1 to estimate the biomass of skipjack tuna. To this purpose, we applied the same procedure described in 5.1, considering only the catches of skipjack tuna.

2.1 Binary classification : Presence/Absence

Performance metrics	50 kHz	200 kHz	50 & 200 kHz	50 x 200 kHz
Accuracy	0.88 ± 0.01	0.86 ± 0.01	0.88 ± 0.02	0.85 ± 0.02
Cohen's Kappa	0.77 ± 0.02	0.72 ± 0.03	0.76 ± 0.03	0.70 ± 0.03

Table S10: Summary of performance metrics for the various combination of frequencies tested for binary classification of SKJ presence/absence.

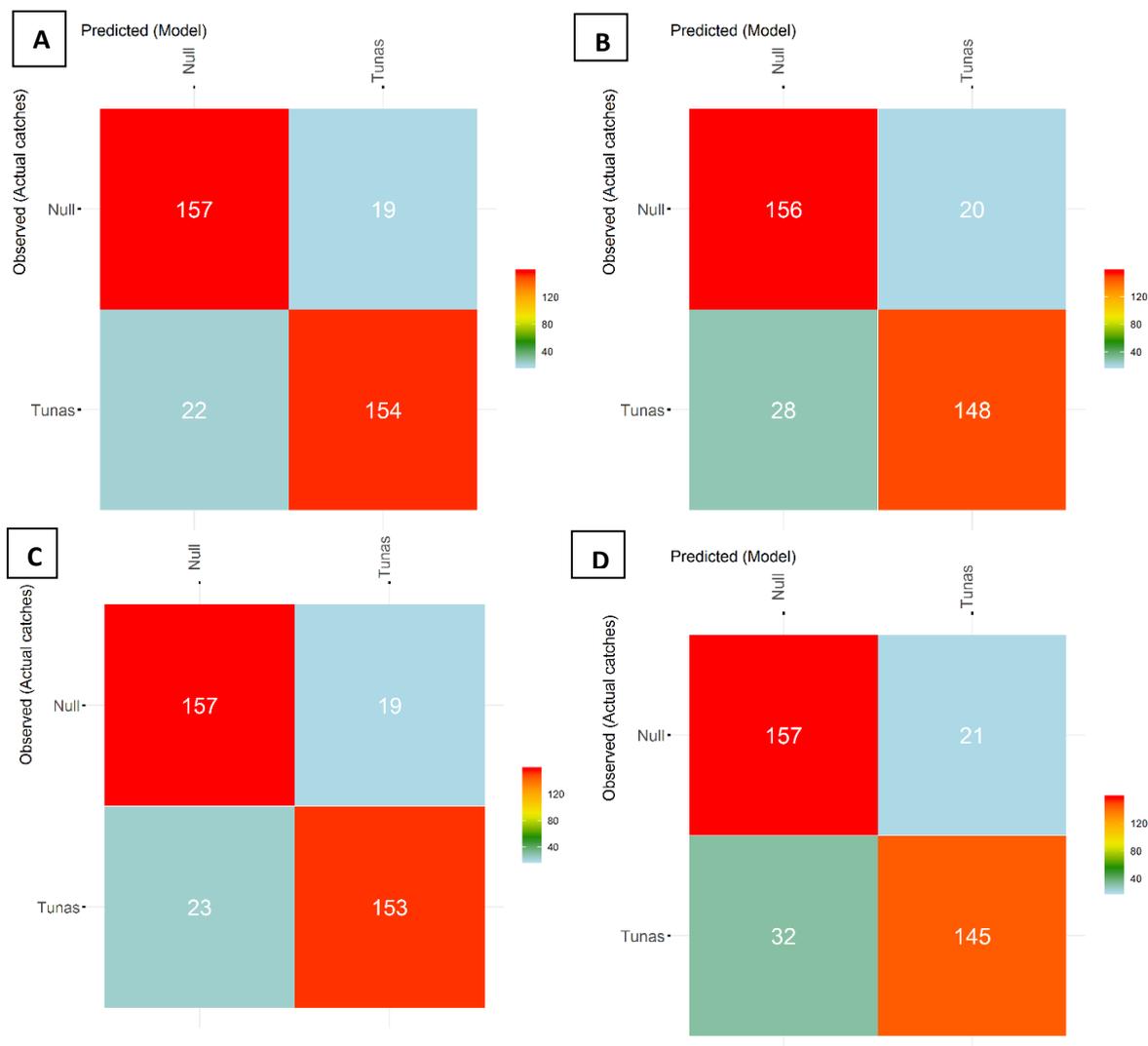


Figure S1: Confusion matrices for binary classification random forests for skipjack: 50 kHz (panel A), 200 kHz (panel B), 50 & 200 kHz (panel C) and 50 x 200 kHz (panel D).

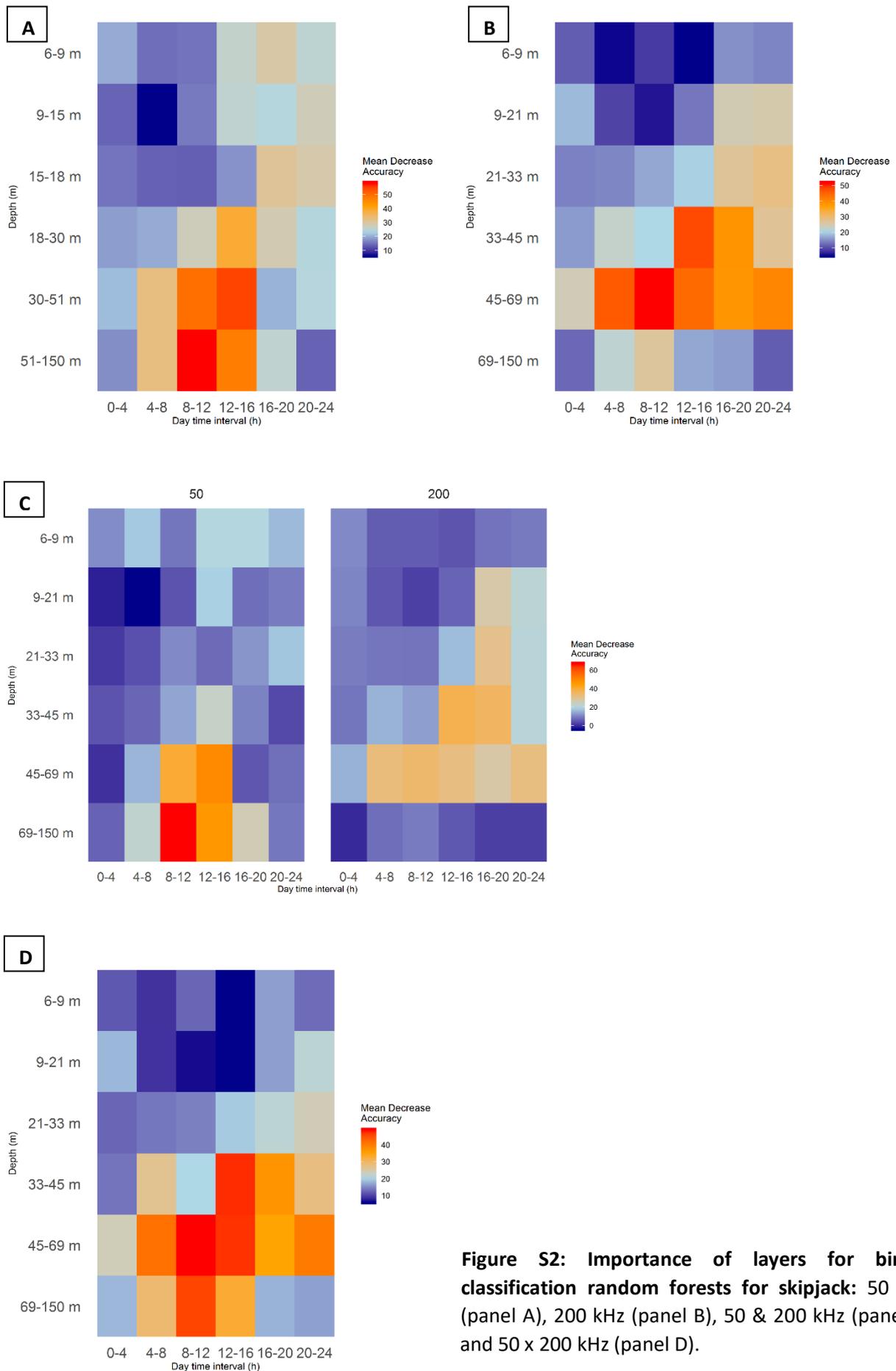


Figure S2: Importance of layers for binary classification random forests for skipjack: 50 kHz (panel A), 200 kHz (panel B), 50 & 200 kHz (panel C) and 50 x 200 kHz (panel D).

2.2. Multiclass Random Forests on SKJ: size of the aggregation

4-classes

Performance metrics	50 kHz	200 kHz	50 & 200 kHz	50 x 200 kHz
Accuracy	0.48 ± 0.02	0.45 ± 0.04	0.47 ± 0.04	0.44 ± 0.03
Cohen's Kappa	0.30 ± 0.03	0.27 ± 0.05	0.30 ± 0.05	0.25 ± 0.05

Table S11: Summary of performance metrics for the various combination of frequencies tested for 4-classes classification.

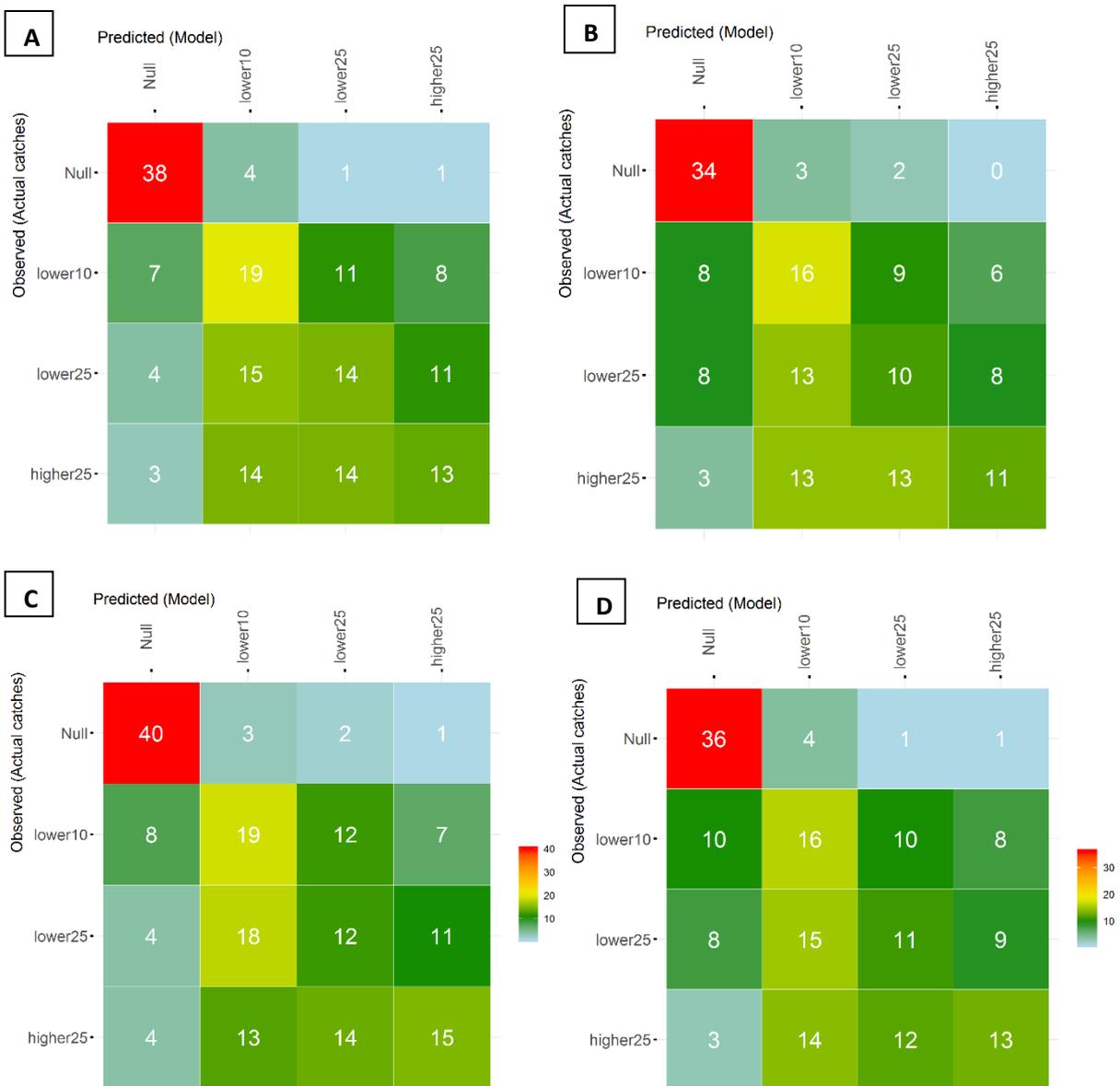


Figure S3: Confusion matrices for 4-classes multiclass random forests for skipjack. 50 kHz (panel A), 200 kHz (panel B), 50 & 200 kHz (panel C) and 50 x 200 kHz (panel D). A, B and D confusion matrices share the same color scale.

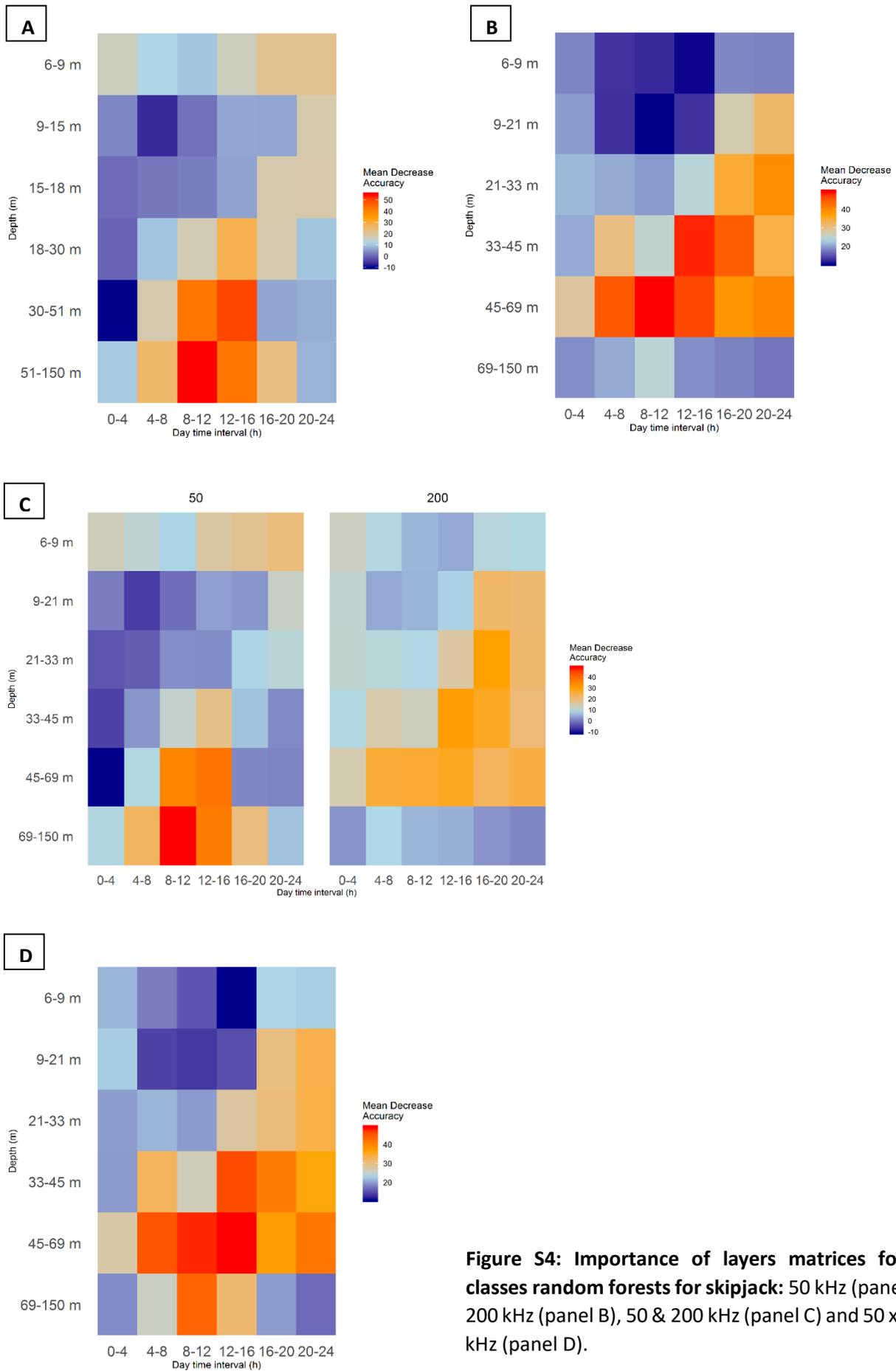


Figure S4: Importance of layers matrices for 4-classes random forests for skipjack: 50 kHz (panel A), 200 kHz (panel B), 50 & 200 kHz (panel C) and 50 x 200 kHz (panel D).

2.3. Detailed performance metrics obtained for the Random Forest algorithm

2.3.1. Binary classification : Presence/Absence metrics

Performance metrics	50 kHz	200 kHz	50 & 200 kHz	50 x 200 kHz
<i>Accuracy</i>	0.88 ± 0.01	0.86 ± 0.01	0.88 ± 0.02	0.85 ± 0.02
<i>Cohen's Kappa</i>	0.77 ± 0.02	0.72 ± 0.03	0.76 ± 0.03	0.70 ± 0.03
<i>Sensitivity</i>	0.87 ± 0.03	0.84 ± 0.02	0.87 ± 0.03	0.82 ± 0.03
<i>Specificity</i>	0.89 ± 0.02	0.89 ± 0.03	0.89 ± 0.02	0.88 ± 0.02
<i>Precision</i>	0.89 ± 0.02	0.88 ± 0.02	0.89 ± 0.02	0.87 ± 0.02
<i>F₁ score</i>	0.88 ± 0.01	0.86 ± 0.01	0.88 ± 0.03	0.84 ± 0.02

Table S13: Detail of performance metrics for the various combination of frequencies tested for binary classification.

2.3.2. Multiclass random forests : size of the aggregation

4-classes classification

Performance metrics	50 kHz				Average
	No tuna	<10 tons	[10-25]tons	> 25 tons	
<i>Sensitivity</i>	0.86 ± 0.06	0.32 ± 0.05	0.42 ± 0.08	0.30 ± 0.06	0.48 ± 0.26
<i>Specificity</i>	0.89 ± 0.02	0.80 ± 0.03	0.76 ± 0.02	0.85 ± 0.02	0.83 ± 0.06
<i>Precision</i>	0.76 ± 0.05	0.35 ± 0.04	0.36 ± 0.05	0.41 ± 0.05	0.47 ± 0.20
<i>F₁ score</i>	0.79 ± 0.05	0.33 ± 0.04	0.39 ± 0.07	0.34 ± 0.06	0.46 ± 0.22
<i>Accuracy</i>					0.48 ± 0.02
<i>Cohen's Kappa</i>					0.30 ± 0.03

Table S14: Detail of performance metrics for 50 kHz model tested for 4-classes classification.

200 kHz

Performance metrics	No tuna	<10 tons	[10-25]tons	> 25 tons	Average
<i>Sensitivity</i>	0.87 ± 0.08	0.24 ± 0.08	0.41 ± 0.06	0.28 ± 0.07	0.45 ± 0.29
<i>Specificity</i>	0.84 ± 0.03	0.80 ± 0.06	0.75 ± 0.04	0.87 ± 0.03	0.82 ± 0.05
<i>Precision</i>	0.64 ± 0.05	0.29 ± 0.07	0.36 ± 0.06	0.43 ± 0.07	0.43 ± 0.15
<i>F₁ score</i>	0.74 ± 0.05	0.26 ± 0.07	0.36 ± 0.05	0.34 ± 0.07	0.43 ± 0.21
<i>Accuracy</i>					0.45 ± 0.04
<i>Cohen's Kappa</i>					0.27 ± 0.05

Table S15: Detail of performance metrics for 200 kHz model tested for 4-classes classification.

50 & 200 kHz

Performance metrics	No tuna	<10 tons	[10-25]tons	> 25 tons	Average
<i>Sensitivity</i>	0.88 ± 0.04	0.42 ± 0.10	0.26 ± 0.06	0.33 ± 0.08	0.47 ± 0.28
<i>Specificity</i>	0.89 ± 0.02	0.75 ± 0.04	0.80 ± 0.06	0.86 ± 0.05	0.83 ± 0.06
<i>Precision</i>	0.72 ± 0.04	0.36 ± 0.07	0.31 ± 0.04	0.45 ± 0.08	0.46 ± 0.18
<i>F₁ score</i>	0.79 ± 0.03	0.38 ± 0.08	0.28 ± 0.04	0.37 ± 0.07	0.46 ± 0.20
<i>Accuracy</i>					0.47 ± 0.04
<i>Cohen's Kappa</i>					0.30 ± 0.05

Table S16: Detail of performance metrics for 50 & 200 kHz model tested for 4-classes classification.

50 x 200

Performance metrics	No tuna	<10 tons	[10-25]tons	> 25 tons	Average
<i>Sensitivity</i>	0.84 ± 0.06	0.25 ± 0.06	0.36 ± 0.06	0.31 ± 0.05	0.44 ± 0.27
<i>Specificity</i>	0.83 ± 0.03	0.82 ± 0.03	0.74 ± 0.04	0.86 ± 0.03	0.81 ± 0.05
<i>Precision</i>	0.63 ± 0.06	0.31 ± 0.05	0.32 ± 0.07	0.42 ± 0.06	0.42 ± 0.15
<i>F₁ score</i>	0.72 ± 0.05	0.28 ± 0.05	0.34 ± 0.06	0.36 ± 0.05	0.43 ± 0.20
<i>Accuracy</i>					0.44 ± 0.03
<i>Cohen's Kappa</i>					0.25 ± 0.05

Table S17: Detail of performance metrics for 50 x 200 kHz model tested for 4-classes classification.