Dolphin species classification

Technical report

Vyacheslav Lyubchich et al.

2025-08-20

Table of contents

1 Packages

Load R packages used in this report, including dplyr (Wickham et al. 2023), tidyr (Wickham et al. 2024b), ggplot2 (Wickham et al. 2024a), patchwork (Pedersen 2024), banter (Archer and Sakai 2023), pdp (Greenwell 2022), and rfPermute (Archer 2023).

2 General parameters

Significance level $\alpha = 5\%$.

3 Data summaries

PAMGuard and ROCCA whistle summaries were additionally processed using `code/dataprocess.R` comprising the following steps.

1. For the T1C data, use the column `UTC` to update ROCCA-generated event IDs:
    - Sort the data chronologically and identify uninterrupted sequences of event IDs from the ROCCA output;
    - Number the uninterrupted sequences of event IDs;
    - Update the event IDs by appending the ROCCA-generated event ID and the assigned number;
    - Split the new events such that the time gap between consecutive whistles is no more than 5 minutes, update the event IDs again.
2. For the NOAA data, correct the dates, using the original column `Source`.
3. For the Watkins data, correct the dates, using the original column `Source` and metadata.
4. For the Brazil data, use the column `UTC` to split the data in 5-minute events and update the ROCCA-generated single event ID.
5. From each dataset, select the columns `species`, ID columns (`Source`, `event.id`, `call.id`, `Year`, `Month`, and `UTC`), and potential predictors.
6. Combine the datasets.
7. Remove columns in which all values are the same.

Eventually, assign the response variable: species.

Predictors for whistle classification (50 predictors) in alphabetic order: duration, freqAbsSlopeMean, freqBeg, freqBegDwn, freqBegEndRatio, freqBegSweep, freqBegUp, freqCenter, freqCOFM, freqEnd, freqEndDwn, freqEndSweep, freqEndUp, freqMax, freqMaxMinRatio, freqMean, freqMedian, freqMin, freqNegSlopeMean, freqNumSteps, freqPosSlopeMean, freqQuarter1, freqQuarter2, freqQuarter3, freqRange, freqRelBW, freqSlopeMean, freqSlopeRatio, freqSpread, freqStdDev, freqStepDown, freqStepUp, freqSweepDwnPercent, freqSweepFlatPercent, freqSweepUpPercent, infldur, inflMaxDelta, inflMaxMinDelta, inflMeanDelta, inflMedianDelta, inflMinDelta, inflStdDevDelta, numInflections, numSweepsDwnFlat, numSweepsDwnUp, numSweepsFlatDwn, numSweepsFlatUp, numSweepsUpDwn, numSweepsUpFlat, stepdur.

Total number of whistles is 14822, including 8375 whistles of bottlenose dolphins and 6447 whistles of common dolphins (see details in Table 1).

Table 1: Number of whistles available from each source and year along with the event counts.

Source	Dolphin species	1958	1975	1987	2014	2016	2017	2018	Whistle, count	Whistle, %	Event, count	Event, %
AMAPPS	Common	0	0	0	0	2637	0	0	2637	17.791	9	0.696
T1C	Bottlenose	0	0	0	0	5251	2049	1075	8375	56.504	1225	94.668
UFRJ	Common	0	0	0	3580	0	0	0	3580	24.153	29	2.241
Watkins	Common	131	97	2	0	0	0	0	230	1.552	31	2.396
Total	–	131	97	2	3580	7888	2049	1075	14822	100	1294	100.001

From Table 2 and Figure 1, bottlenose dolphin encounters and common dolphin encounters recorded in the Watkins dataset contain much fewer whistles than common dolphin encounters from the UFRJ and AMAPPS datasets.

Table 2: Statistical summaries of the number of whistles per event.

Source	Dolphin species	n	mean	sd	median	min	max	skew	kurtosis	se
AMAPPS	Common	9	293.00	316.82	127	24	1023	1.27	0.27	105.61
T1C	Common	1225	6.84	10.21	3	1	101	3.94	22.67	0.29
UFRJ	Bottlenose	29	123.45	143.29	51	1	495	1.21	0.20	26.61
Watkins	Common	31	7.42	9.01	4	1	32	1.79	1.82	1.62

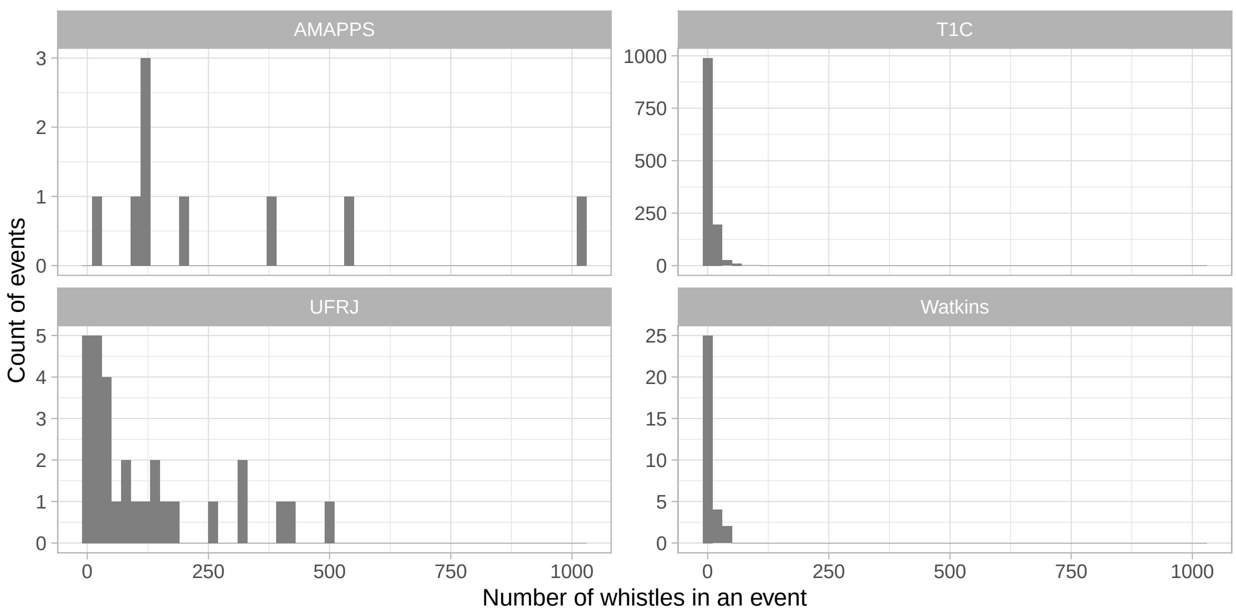


Figure 1: Distributions of events by the number of whistles in each. The bin width is 20.

4 Results

4.1 Compare models

See Figure 2 and Table 3 with results of 5-fold cross-validation applied 5 times (each boxplot corresponds to 25 values). The folds were created considering the underlying species proportions, hence each testing set (validation fold) contained about 245 events with bottlenose dolphins and about 14 events with common dolphins.

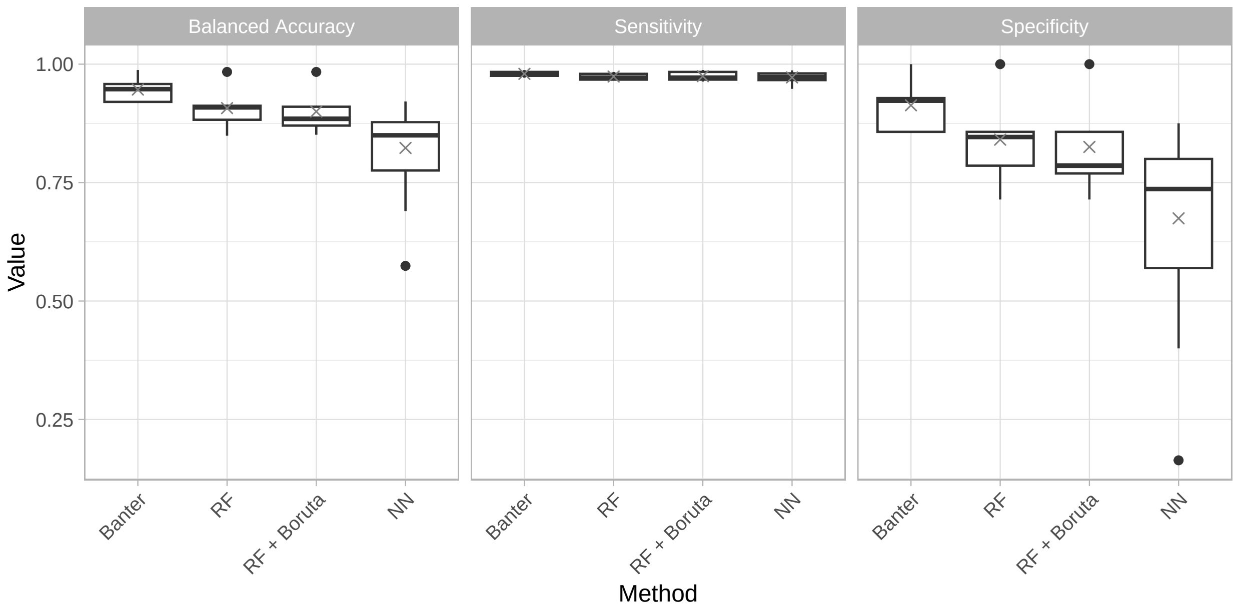


Figure 2: Boxplots of classification performance metrics from cross-validation.

Table 3: Average performance of different methods in cross-validation.

Method	Sensitivity	Specificity	Balanced Accuracy
Banter	0.98	0.91	0.95
NN	0.97	0.67	0.82
RF	0.97	0.84	0.91
RF + Boruta	0.97	0.83	0.90

Figure 3 shows high variability of accuracy of classification based on the number of whistles in an event. Figure 4 focuses on the events with fewer whistles and shows that the accuracy is generally increasing with the number of whistles in an event, and accuracy of classifying events with just one whistle is not very different from accuracy obtained for events with more whistles.

The best-performing model is the BANTER model (stacked random forests) retrained to classify between two species of dolphins.

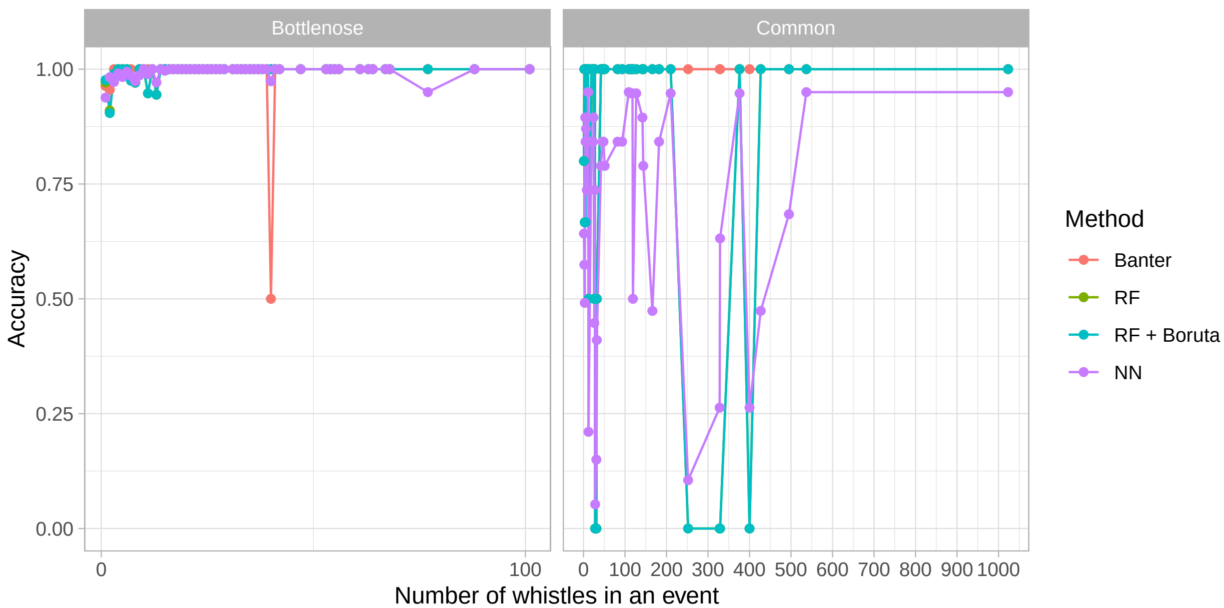


Figure 3: Cross-validation accuracy of classifying events for each species, based on the applied method and whistle count per event.

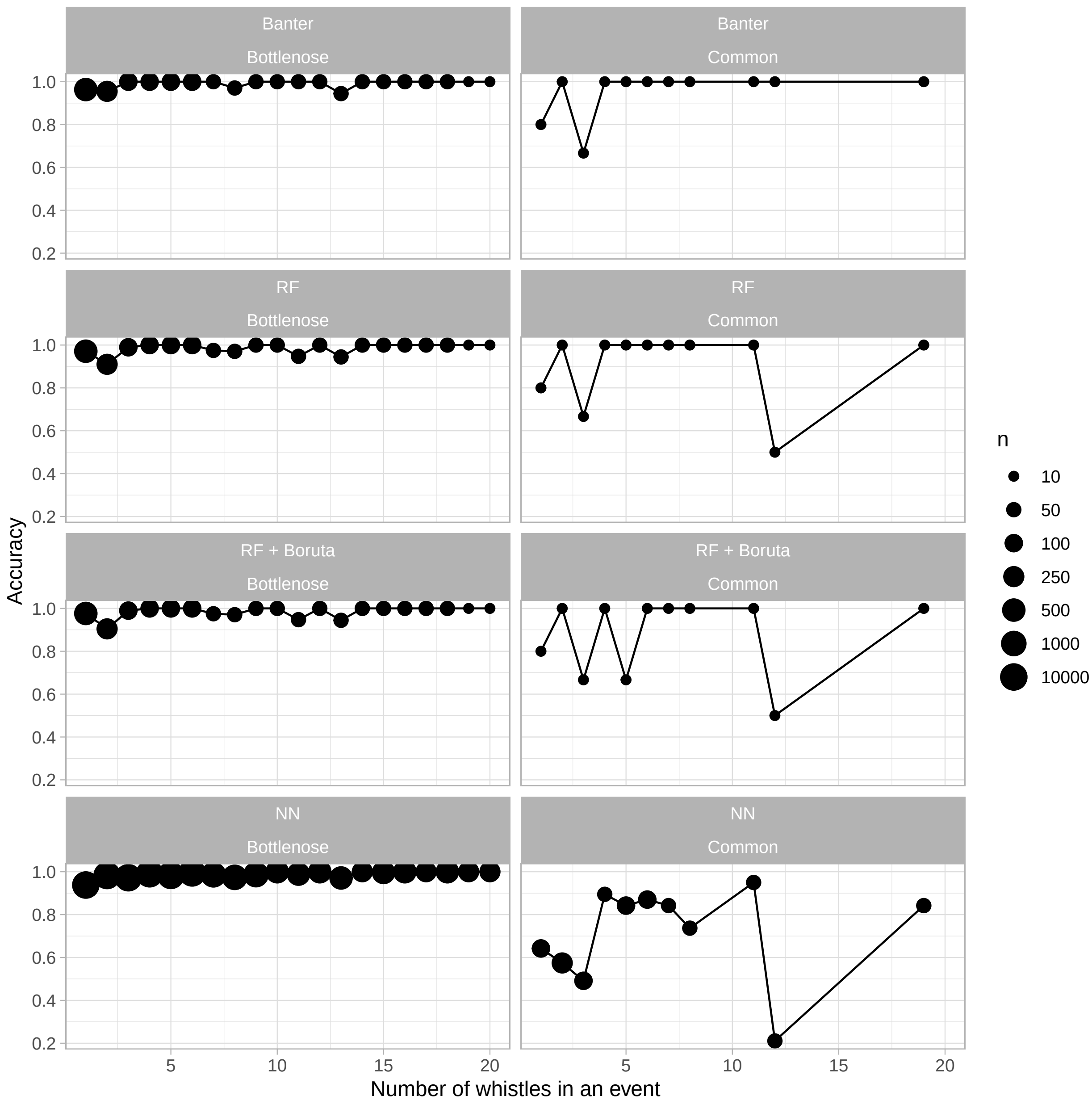


Figure 4: Cross-validation accuracy of classifying events for each species (for events with up to 20 whistles), based on the applied method and whistle count per event. The point sizes correspond to the number of times the events of given length appeared in cross-validation.

4.2 Analysis of the selected model

Refit the model on the whole dataset.

4.2.1 Level 1: Whistles

Summary of the detector-level model from the package BANTER

```
#> Model run times:
#>    start                 stop                  run.time
#> dw "2025-07-08 13:42:50" "2025-07-08 13:44:01" "1.19 mins"
#>
#> Number of events and model classification rate:
#>      species num.events    dw
#> 1 Bottlenose       1225 91.52
#> 2     Common         69 82.64
#> 3    Overall       1294 87.66
```

Another summary

```
#>
#> Call:
#>  randomForest(x = params$predictors, y = params$response, ntree = params$ntree,      replace = FALSE, sampsize =
#>               Type of random forest: classification
#>                     Number of trees: 1000
#> No. of variables tried at each split: 7
#>
#>         OOB estimate of  error rate: 12.34%
#> Confusion matrix:
#>            Bottlenose Common class.error
#> Bottlenose       7665    710  0.08477612
#> Common           1119   5328  0.17356910
```

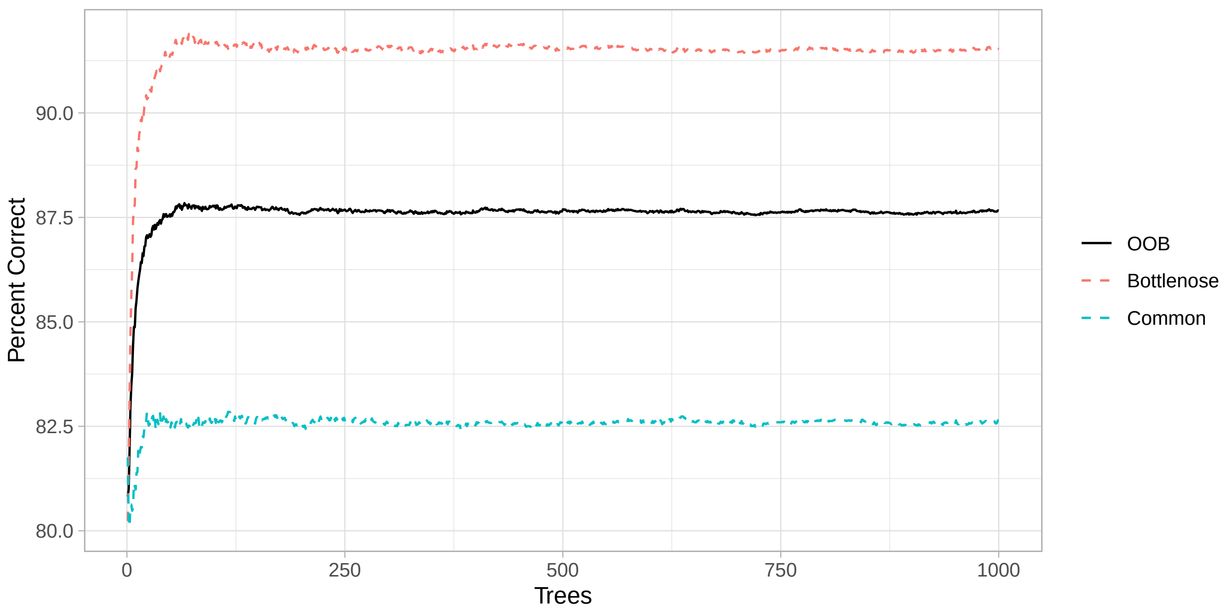


Figure 5: Detector-level model performance on out-of-bag (OOB) data.

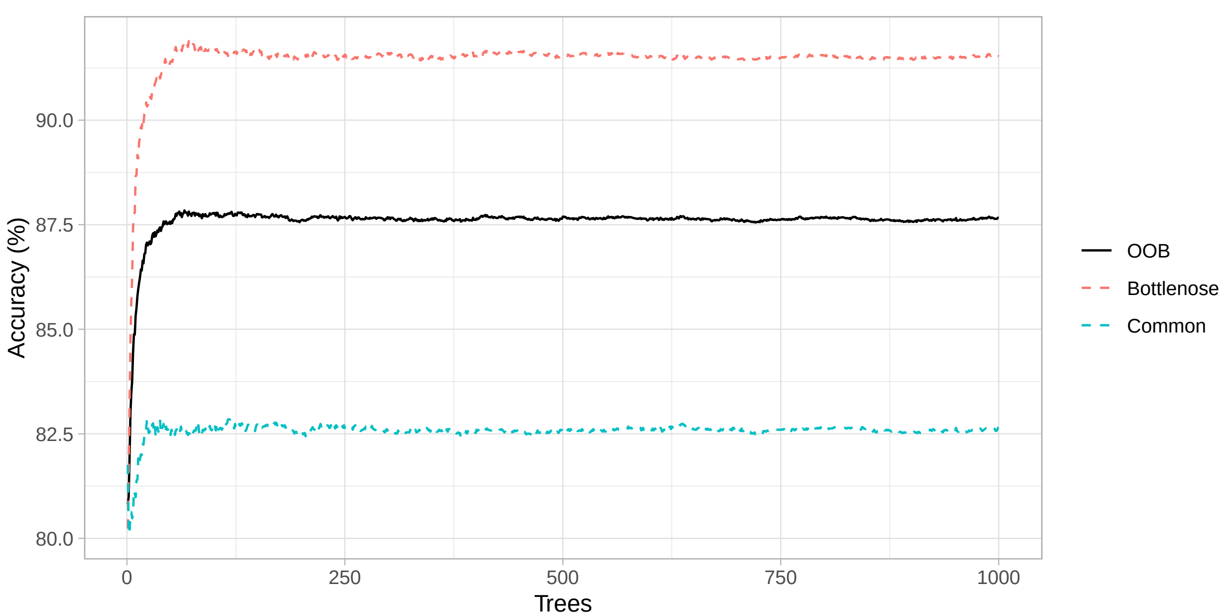


Figure 6: Detector-level model performance on out-of-bag (OOB) data.

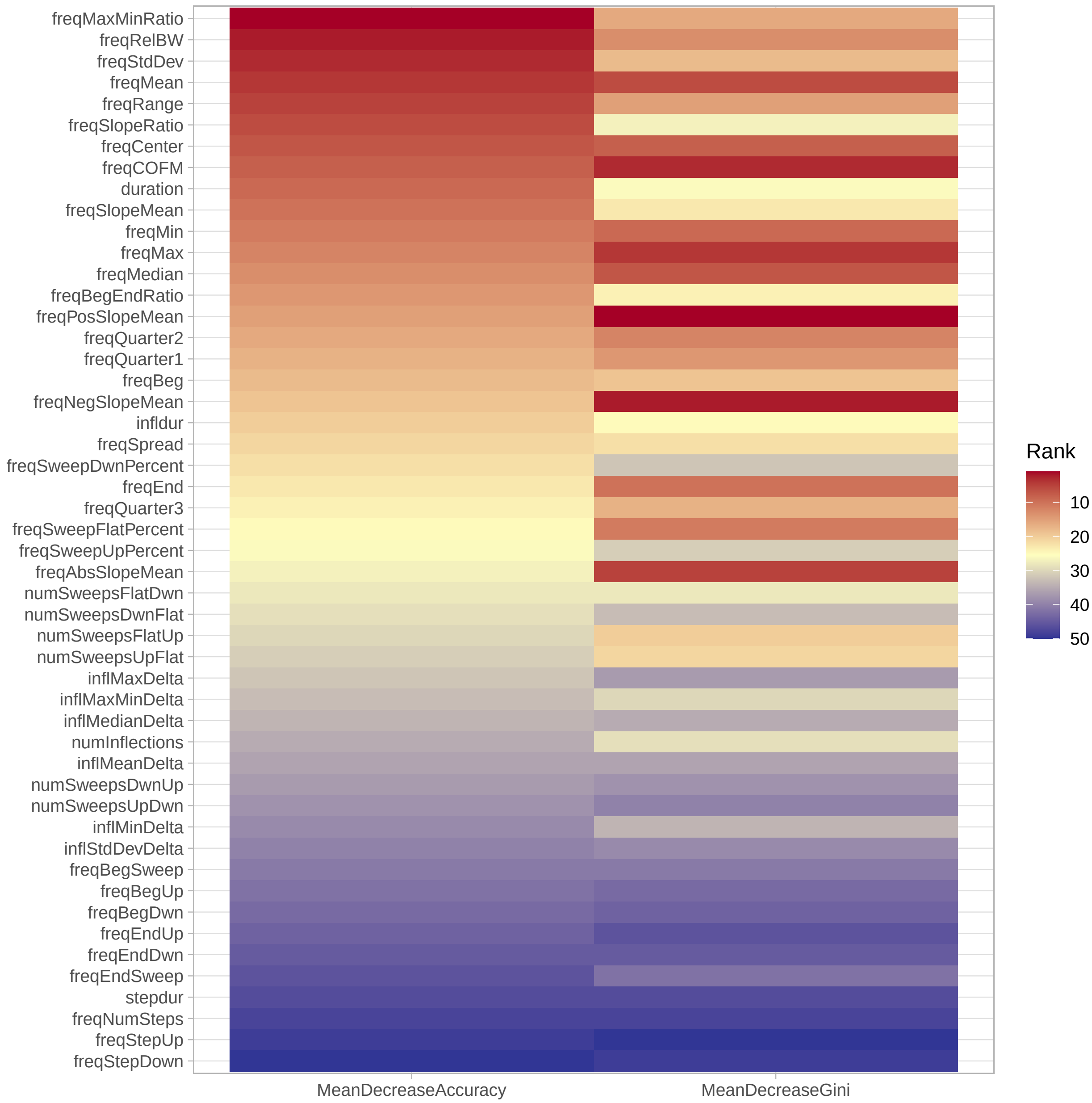


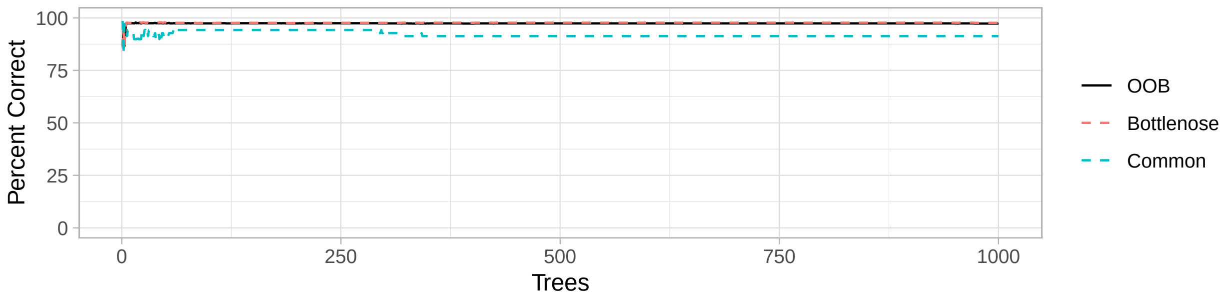
Figure 7: Variables in the detector-level model ranked by their importance (mean decrease in classification accuracy), with the most important on top.

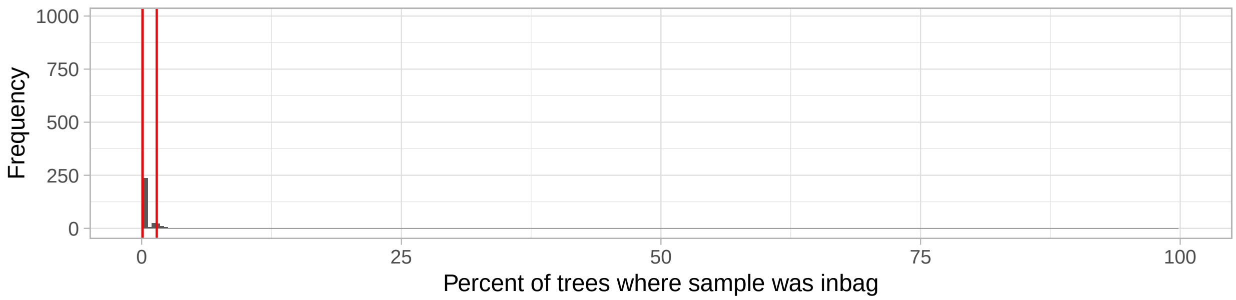
4.2.2 Level 2: Events

Summary of the detector-level model from the package BANTER

```
#> Model run times:
#>       start                 stop                  run.time
#> dw    "2025-07-08 13:42:50" "2025-07-08 13:44:01" "1.19 mins"
#> event "2025-07-08 13:44:01" "2025-07-08 13:44:04" "3.76 secs"
#>
#> Number of events and model classification rate:
#>      species num.events    dw event
#> 1 Bottlenose       1225 91.52 97.71
#> 2     Common         69 82.64 91.30
#> 3    Overall       1294 87.66 97.37
```

```
#>
#> Number of trees: 1000
#>
#> Sample sizes:
#> Bottlenose     Common
#>          1          1
#>
#> Distribution of percent correctly classified overall in last 500 (50%) trees:
#>    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#>   97.30   97.37   97.37   97.37   97.37   97.37
#>
#> Sample inbag proportion distribution:
#>          Min. 1st Qu. Median Mean 3rd Qu. Max.
#> expected  0.1     0.4    0.8  0.8     1.1  1.4
#> observed  0.0     0.0    0.1  0.2     0.1  2.4
#>
#> Confusion matrix:
#>            Bottlenose Common pct.correct LCI_0.95 UCI_0.95
#> Bottlenose       1197     28    97.71429 96.71335 98.47591
#> Common              6     63    91.30435 82.02793 96.74171
#> Overall            NA     NA    97.37249 96.34754 98.17366
```





Another summary

```
#>
#> Call:
#>  randomForest(formula = species ~ ., data = x@model.data, ntree = ntree,      sampsize = sampsize, replace = FAL
#>                Type of random forest: classification
#>                      Number of trees: 1000
#> No. of variables tried at each split: 1
#>
#>         OOB estimate of  error rate: 2.63%
#> Confusion matrix:
#>            Bottlenose Common class.error
#> Bottlenose       1197     28  0.02285714
#> Common              6     63  0.08695652
```

Overall performance is adequate, with high overall accuracy.

Table 4 shows BANTER summary with the percent of each species correctly classified for each detector model and the event model. It is a summary of the diagonal values from the confusion matrices for all models.

Table 4: Percent of correct classifications by different levels of the model.

species	dw	event
Bottlenose	91.522	97.714
Common	82.643	91.304
Overall	87.660	97.372

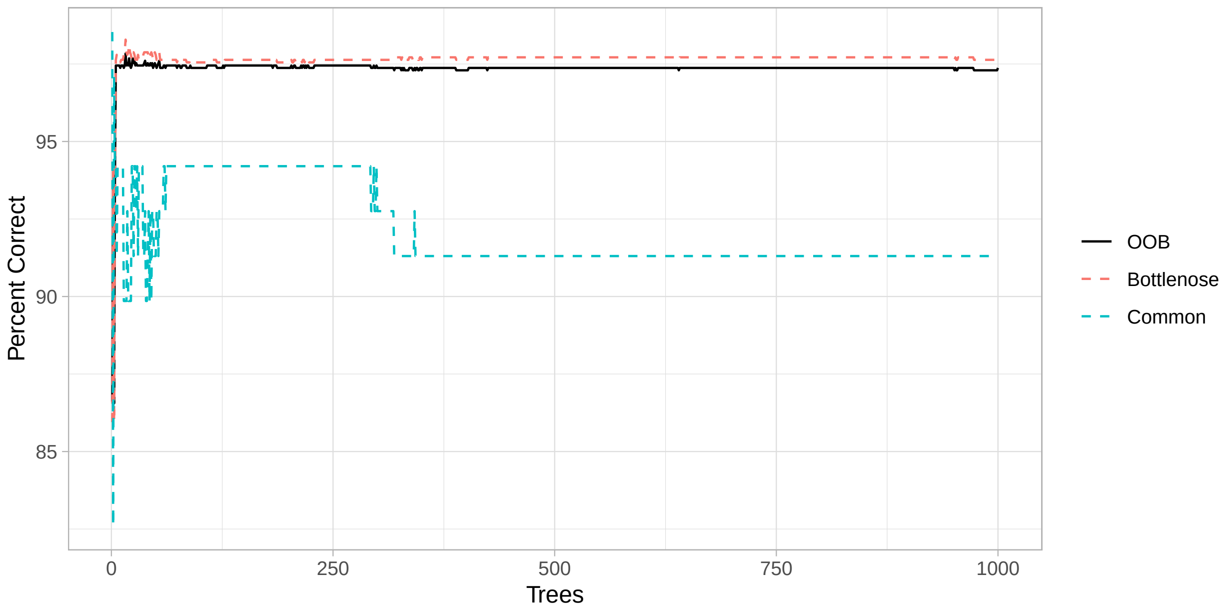


Figure 8: Event-level model performance on out-of-bag (OOB) data.

Table 5: Percent of correct classifications given different thresholds for the proportion of trees that need to vote for that species.

	Bottlenose	Common	Overall
pct.correct_0.1	97.7	91.3	97.4
pct.correct_0.2	97.7	91.3	97.4
pct.correct_0.3	97.7	91.3	97.4
pct.correct_0.4	97.7	91.3	97.4
pct.correct_0.5	97.7	91.3	97.4
pct.correct_0.6	97.3	89.9	96.9
pct.correct_0.7	96.2	88.4	95.8
pct.correct_0.8	94.7	84.1	94.1
pct.correct_0.9	89.6	75.4	88.9
pct.correct_0.95	84.9	72.5	84.2

From https://taikisan21.github.io/PAMpal/banterGuide.html#UNDER_DEVELOPMENT: "These values will always decrease as the percent of trees threshold increases. That is because as stringency is decreased (lower thresholds), more samples are likely to be correctly classified. These values give an indication of the fraction of events that can be classified with high certainty."

In our analysis, there are many cases which were classified with high certainty (see Table 5). The certainty is higher for bottlenose dolphins (Table 5). However, some of the events that were classified as common dolphins with high certainty (> 90%) were bottlenose dolphin occurrences (right plot in Figure 11).

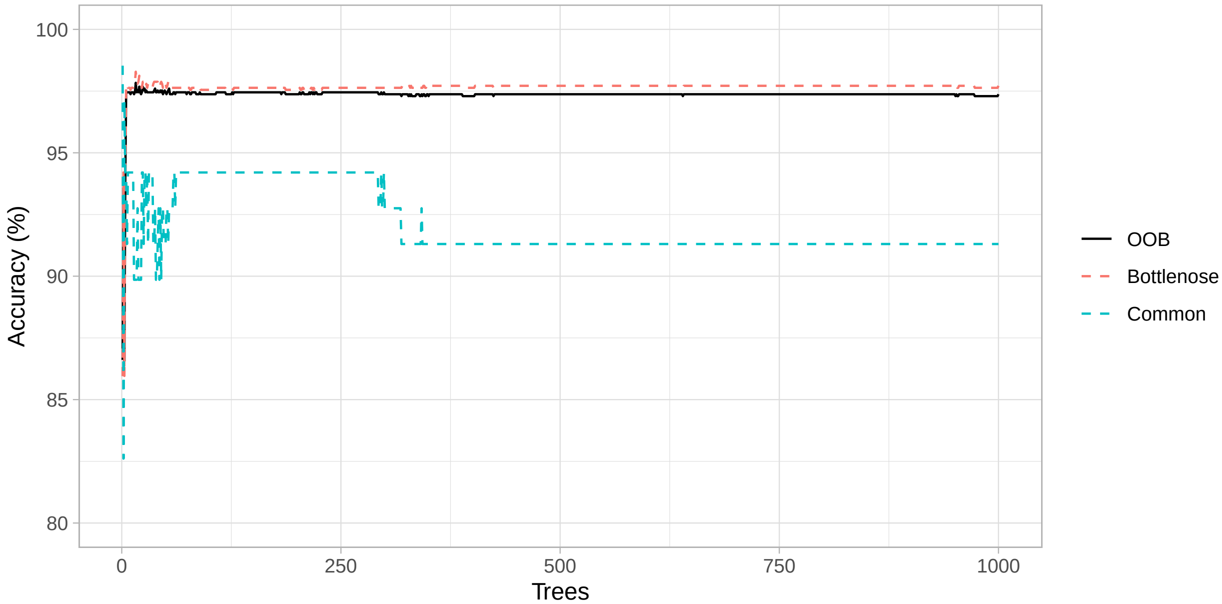


Figure 9: Event-level model performance on out-of-bag (OOB) data.

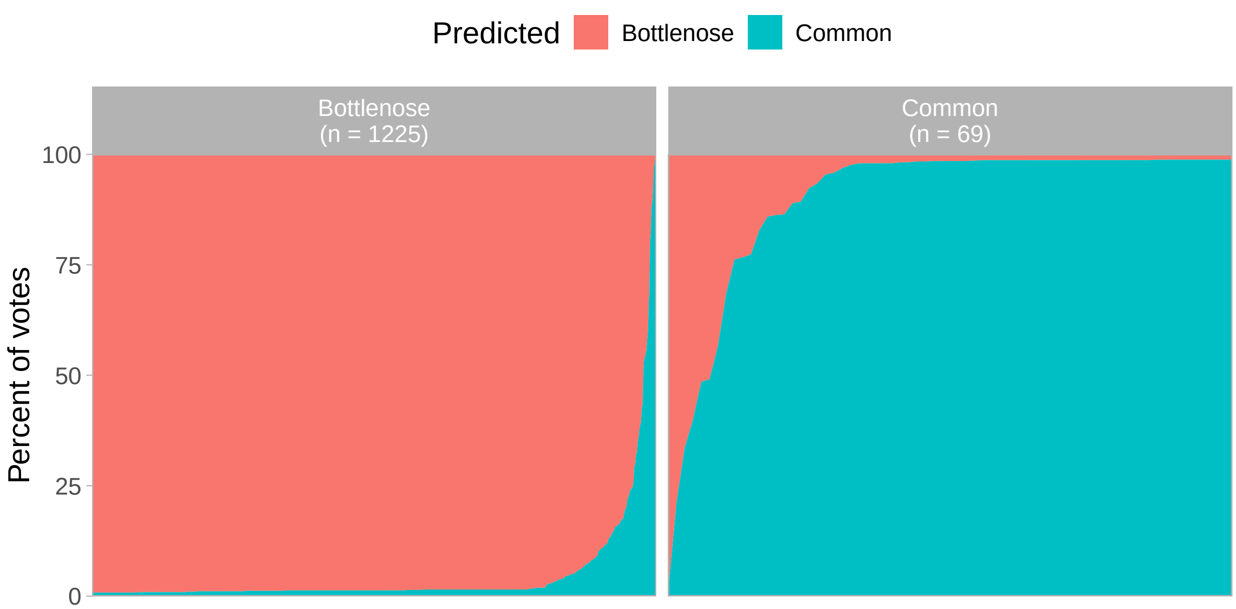


Figure 10: Number of trees that predicted each species (i.e., votes from each tree).

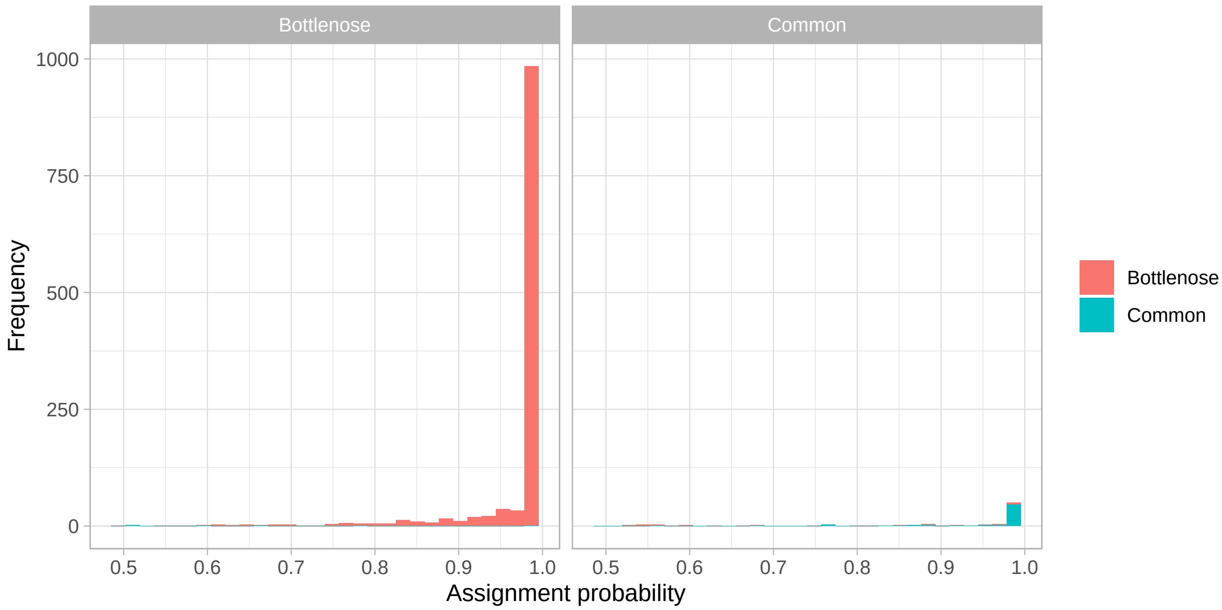


Figure 11: Distribution of votes (assignment probabilities) for each of the predicted species. The true species are coded by colors.

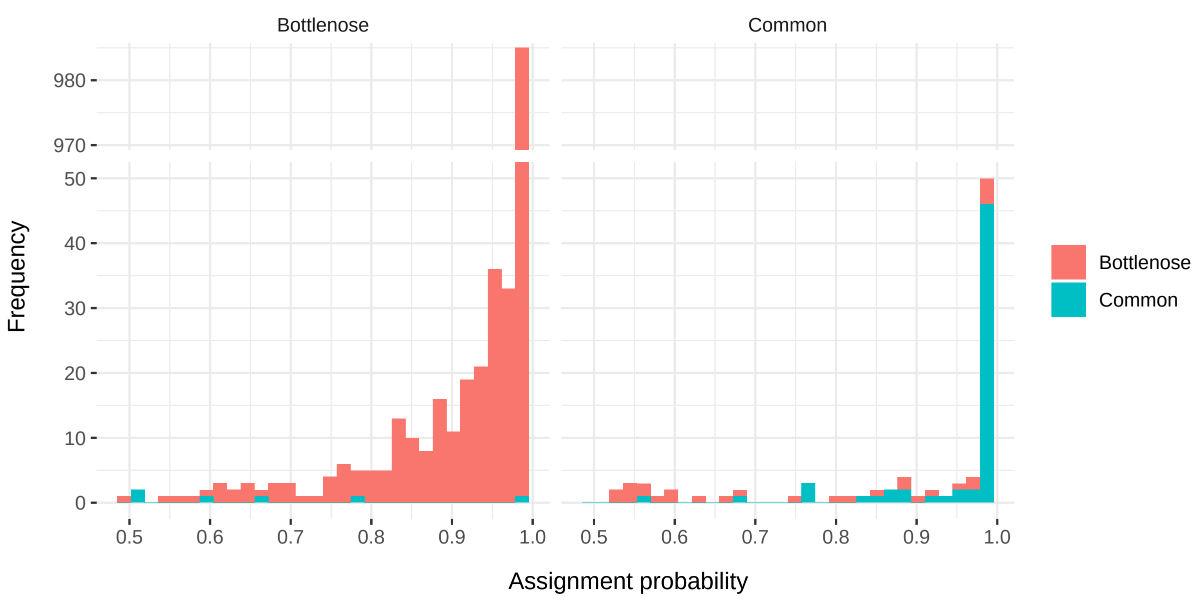


Figure 12: Distribution of votes (assignment probabilities) for each of the predicted species. The true species are coded by colors.

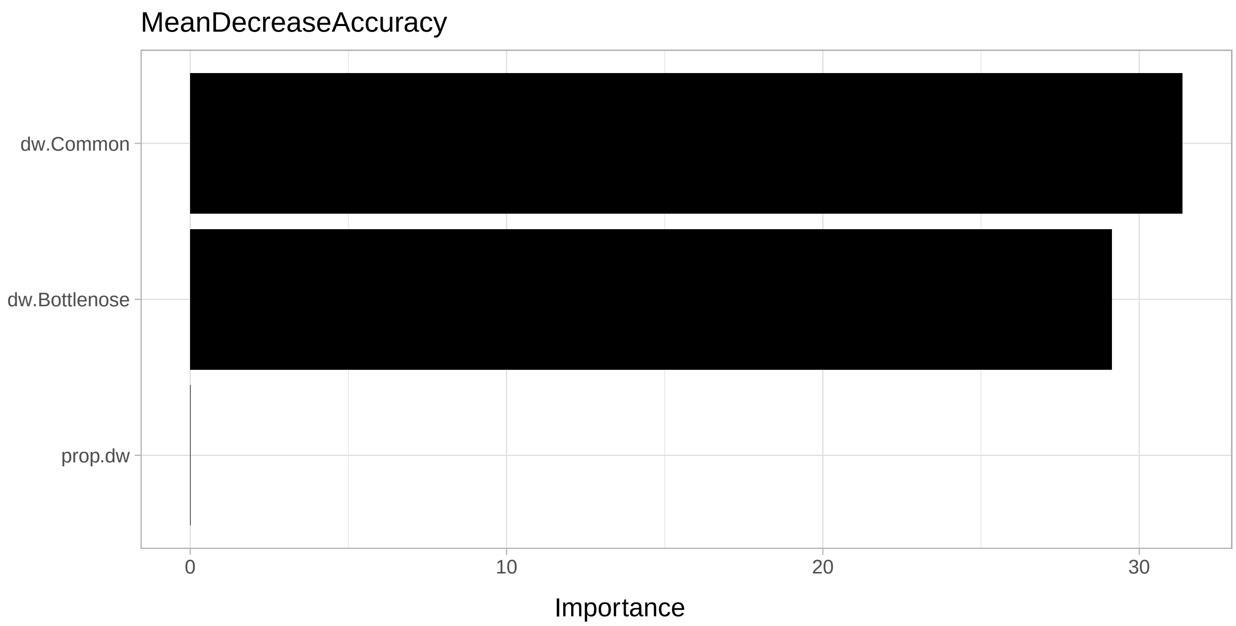


Figure 13: Variables in the event-level model ranked by their importance (mean decrease in classification accuracy), with the most important on top.

4.2.3 Misclassified events

There are 34 misclassified events (Table 6).

Table 6: List of misclassified events with the model vote proportions for each species.

id	original	predicted	Bottlenose	Common
Bra_B13h55m12s08may2014_17	Common	Bottlenose	0.99	0.01
Bra_B13h55m12s08may2014_8	Common	Bottlenose	0.60	0.40
T1C_2016MarApr_AutoEvent75_0225	Bottlenose	Common	0.37	0.63
T1C_2016MayJun_AutoEvent17_0553	Bottlenose	Common	0.11	0.89
T1C_2017AprSep_AutoEvent178_0679	Bottlenose	Common	0.44	0.56
T1C_2017AprSep_AutoEvent190_0687	Bottlenose	Common	0.47	0.53
T1C_2017AprSep_AutoEvent47_0593	Bottlenose	Common	0.45	0.55
T1C_2017AprSep_AutoEvent89_0626	Bottlenose	Common	0.19	0.81
T1C_2017OctDec_AutoEvent118_0776	Bottlenose	Common	0.12	0.88
T1C_2017OctDec_AutoEvent121_0778	Bottlenose	Common	0.10	0.90
T1C_2017OctDec_AutoEvent147_0786	Bottlenose	Common	0.05	0.95
T1C_2017OctDec_AutoEvent199_1013	Bottlenose	Common	0.43	0.57
T1C_2017OctDec_AutoEvent217_0792	Bottlenose	Common	0.33	0.67
T1C_2017OctDec_AutoEvent221_1023	Bottlenose	Common	0.46	0.54
T1C_2017OctDec_AutoEvent335_1036	Bottlenose	Common	0.18	0.82
T1C_2017OctDec_AutoEvent356_0946	Bottlenose	Common	0.25	0.75
T1C_2017OctDec_AutoEvent435_0959	Bottlenose	Common	0.41	0.59
T1C_2017OctDec_AutoEvent47_0906	Bottlenose	Common	0.08	0.92
T1C_2017OctDec_AutoEvent52_0909	Bottlenose	Common	0.45	0.55
T1C_2017OctDec_AutoEvent550_1054	Bottlenose	Common	0.32	0.68
T1C_2017OctDec_AutoEvent701_0886	Bottlenose	Common	0.02	0.98
T1C_2017OctDec_AutoEvent794_0891	Bottlenose	Common	0.01	0.99
T1C_2018JulAug_AutoEvent10_1247	Bottlenose	Common	0.03	0.97
T1C_2018JulAug_AutoEvent12_1272	Bottlenose	Common	0.15	0.85
T1C_2018JulAug_AutoEvent14_1149	Bottlenose	Common	0.01	0.99
T1C_2018JulAug_AutoEvent29_1072	Bottlenose	Common	0.01	0.99
T1C_2018JulAug_AutoEvent30_1136	Bottlenose	Common	0.01	0.99
T1C_2018JulAug_AutoEvent34_1140	Bottlenose	Common	0.40	0.60
T1C_2018JulAug_AutoEvent3_1193	Bottlenose	Common	0.46	0.54
T1C_2018JulAug_AutoEvent9_1214	Bottlenose	Common	0.46	0.54
Wat_AutoEvent10	Common	Bottlenose	0.51	0.49
Wat_AutoEvent6	Common	Bottlenose	0.79	0.21
Wat_AutoEvent8	Common	Bottlenose	0.51	0.49
Wat_AutoEvent9	Common	Bottlenose	0.66	0.34

4.3 Use the selected model

4.3.1 Chesapeake Bay bottlenose

The 22 unlabeled events were mostly classified as bottlenose dolphins (Table 7).

Table 7: Summary of the predicted species.

	Number of events	Percent
Bottlenose	22	100

Show which events (if any) were classified as common dolphins

```
#>  [1] event.id      n             UTC           Year          Month         Source        prop.dw       dw.Bottle
#>  [9] dw.Common     predicted
#> <0 rows> (or 0-length row.names)
```

Check number of whistles per event (Figure 14):

```
#>
#>  1  2  3  5  8 15 16 21
#>  5  6  5  2  1  1  1  1
```

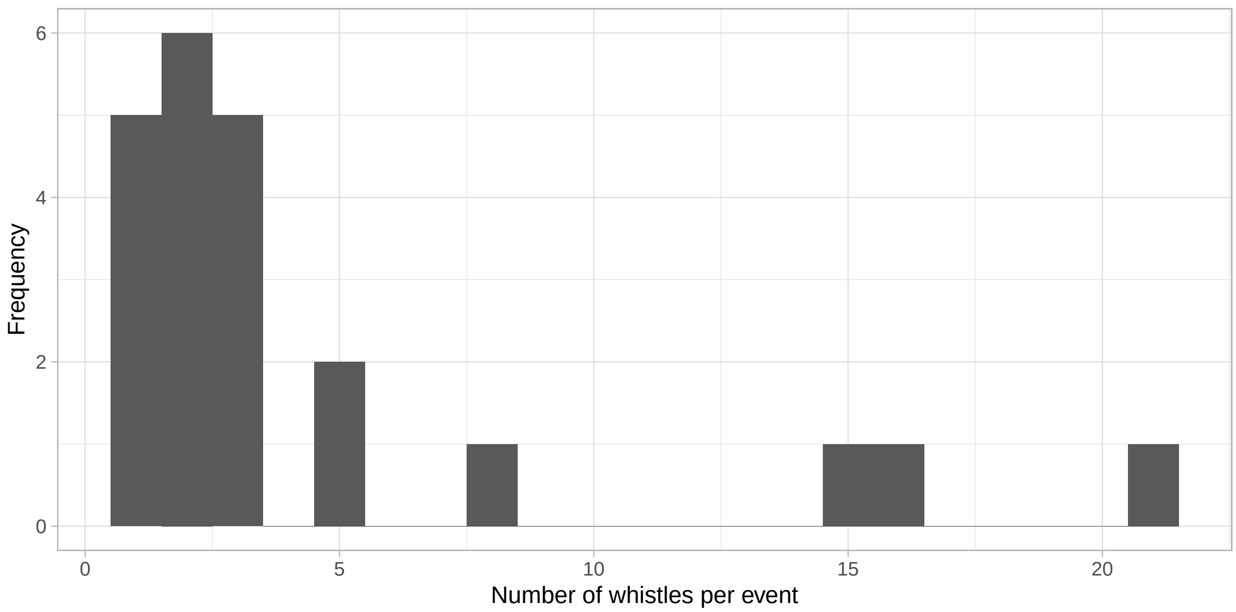


Figure 14: number of whistles per event in the Chesapeake Bay dataset.

4.3.2 Unlabeled data

The 2458 unlabeled events were mostly classified as bottlenose dolphins (Table 8).

Table 8: Summary of the predicted species.

	Number of events	Percent
Bottlenose	2339	95.16
Common	119	4.84

```
#>        predicted
#> Source  Bottlenose Common
#>   A5C         2117     94
#>   MBuoy        222     25
```

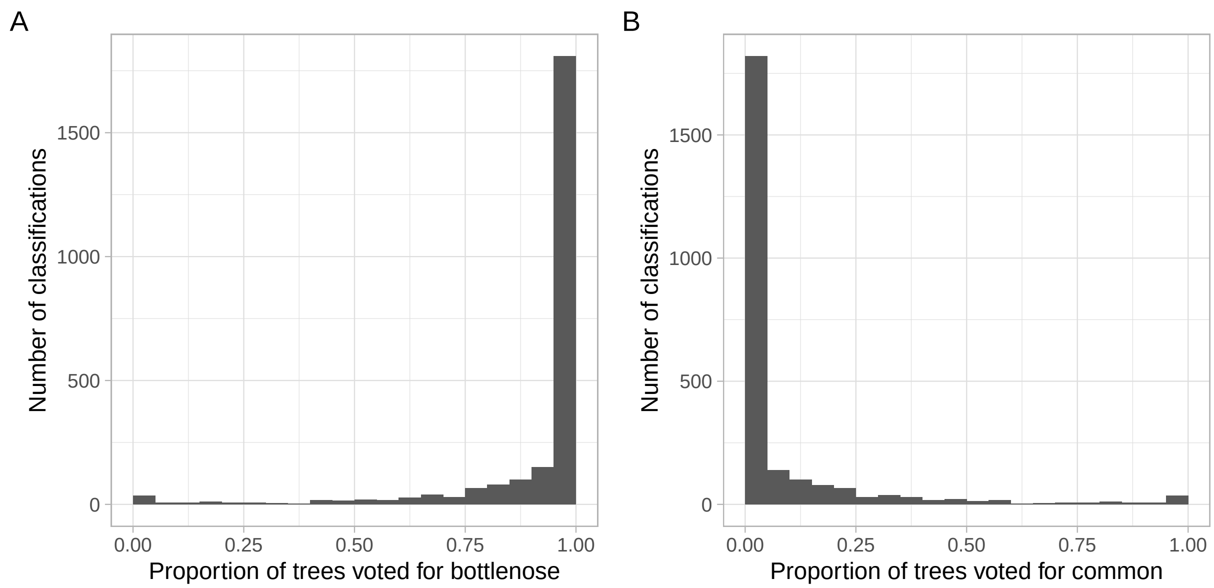


Figure 15: Assignment probabilities for each of the species (since this is a binary classification, the plots are mirrow images of each other).

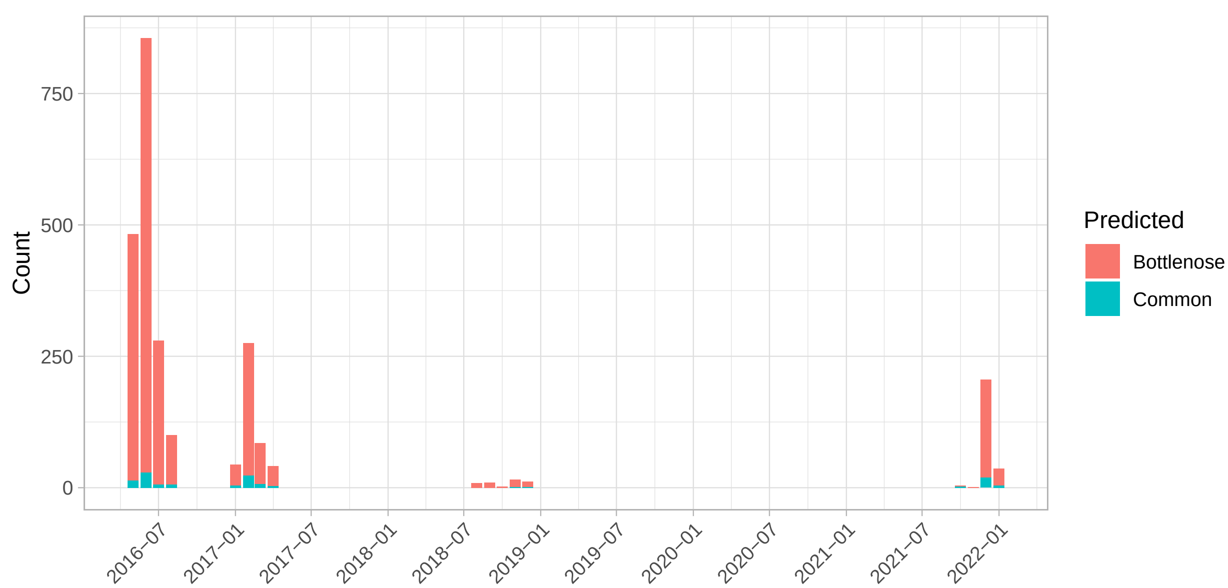


Figure 16: Number of events and their predicted species by month.

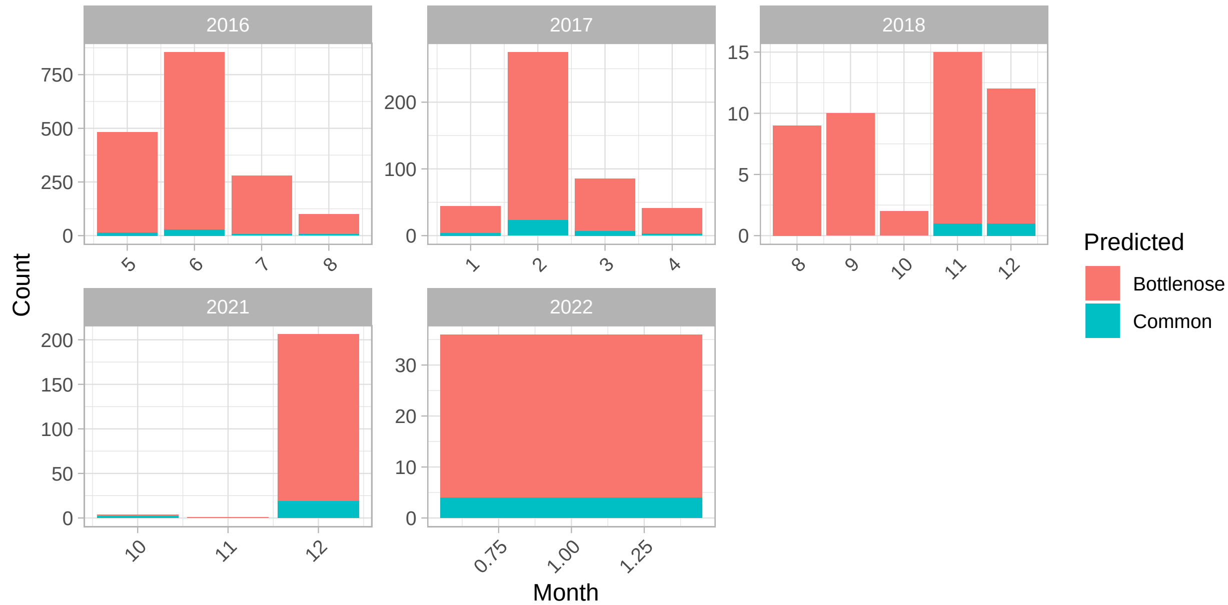


Figure 17: Number of events and their predicted species grouped by year. Note different scales across the years.

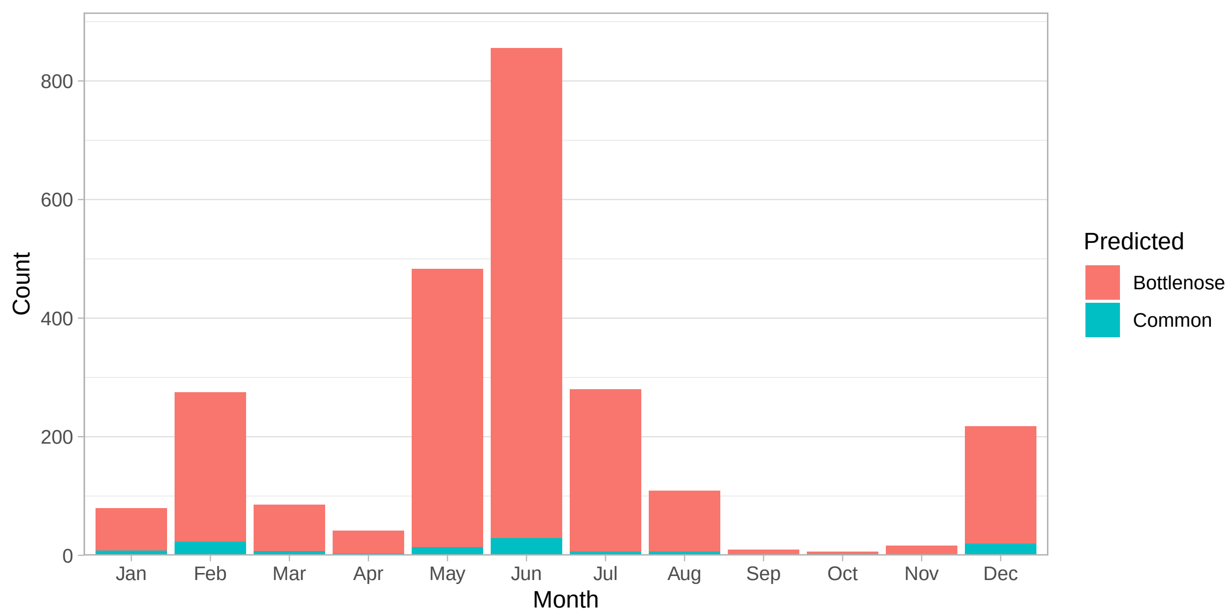


Figure 18: Number of events and their predicted species grouped by month of detections (all years combined).

References

Archer E (2023) rfPermute: Estimate permutation p-values for random forest importance metrics. R package version 2.5.2, https://github.com/EricArcher/rfPermute

Archer E, Sakai T (2023) Banter: BioAcoustic eveNT classifiER. R package version 0.9.6, https://CRAN.R-project.org/package=banter

Greenwell BM (2022) Pdp: Partial dependence plots. R package version 0.8.1, https://github.com/bgreenwell/pdp

Pedersen TL (2024) Patchwork: The composer of plots. R package version 1.2.0, https://patchwork.data-imaginist.com

Wickham H, Chang W, Henry L, et al (2024a) ggplot2: Create elegant data visualisations using the grammar of graphics. R package version 3.5.1, https://ggplot2.tidyverse.org

Wickham H, François R, Henry L, et al (2023) Dplyr: A grammar of data manipulation. R package version 1.1.4, https://dplyr.tidyverse.org

Wickham H, Vaughan D, Girlich M (2024b) Tidyr: Tidy messy data. R package version 1.3.1, https://tidyr.tidyverse.org

S1 Appendix

S1.1 Whistle characteristics by species

Summaries of important whistle characteristics by species:

```
#> 
#>  Descriptive statistics by group 
#> group: Bottlenose
#>                 vars    n     mean      sd   median       min      max    range  skew kurtosis    se
#> freqMaxMinRatio    1 8375     1.11    0.11     1.07      1.01     2.27     1.26  2.91    13.31  0.00
#> freqRelBW          2 8375     0.10    0.09     0.07      0.01     0.78     0.77  2.17     6.49  0.00
#> freqStdDev         3 8375   393.89  290.55   300.15     65.18  2705.87  2640.69  1.86     5.09  3.17
#> freqMean           4 8375 13992.09 3922.10 12415.76   7347.90 23874.02 16526.13  1.12     0.26 42.86
#> freqRange          5 8375  1253.45  872.75   984.38    187.50  7921.88  7734.38  1.82     4.56  9.54
#> freqSlopeRatio     6 8375    -1.15    0.69    -1.02    -10.34     0.00    10.34 -2.43    13.51  0.01
#> freqCenter         7 8375 13992.94 3909.31 12398.44   8226.56 23812.50 15585.94  1.13     0.26 42.72
#> freqCOFM           8 8375     0.12    0.08     0.10      0.00     0.70     0.70  1.83     5.50  0.00
#> duration           9 8375     0.20    0.05     0.18      0.15     0.72     0.58  2.30     8.02  0.00
#> freqSlopeMean     10 8375  2002.03 5697.00  1526.99 -26311.55 34804.40 61115.95  0.15     1.55 62.25
#> freqBeg           11 8375 13741.96 4089.48 12281.25   5156.25 23953.12 18796.88  1.01     0.16 44.69
#> ------------------------------------------------------------------------------------------- 
#> group: Common
#>                 vars    n     mean      sd   median       min      max     range  skew kurtosis     se
#> freqMaxMinRatio    1 6447     1.19    0.18     1.14      1.02     2.86      1.84  2.31     8.22   0.00
#> freqRelBW          2 6447     0.16    0.13     0.13      0.02     0.96      0.95  1.51     2.77   0.00
#> freqStdDev         3 6447   622.27  511.32   490.12     79.90  4686.51   4606.61  1.87     5.12   6.37
#> freqMean           4 6447 15089.43 3614.75 14519.53   7346.59 26094.23  18747.64  0.36    -0.70  45.02
#> freqRange          5 6447  2196.97 1692.38  1828.12    281.25 12187.50  11906.25  1.53     2.91  21.08
#> freqSlopeRatio     6 6447    -1.07    0.55    -1.00     -6.00     0.00      6.00 -2.09     9.06   0.01
#> freqCenter         7 6447 15084.02 3568.88 14578.12   7889.06 25593.75  17704.69  0.36    -0.74  44.45
#> freqCOFM           8 6447     0.30    0.26     0.22      0.00     2.29      2.28  1.64     3.82   0.00
#> duration           9 6447     0.23    0.10     0.20      0.15     1.02      0.87  2.50     8.71   0.00
#> freqSlopeMean     10 6447  2469.39 9841.61  2313.31 -60646.19 62068.97 122715.15 -0.21     3.44 122.57
#> freqBeg           11 6447 14753.64 3748.58 14250.00   4621.88 26625.00  22003.12  0.26    -0.68  46.69
```

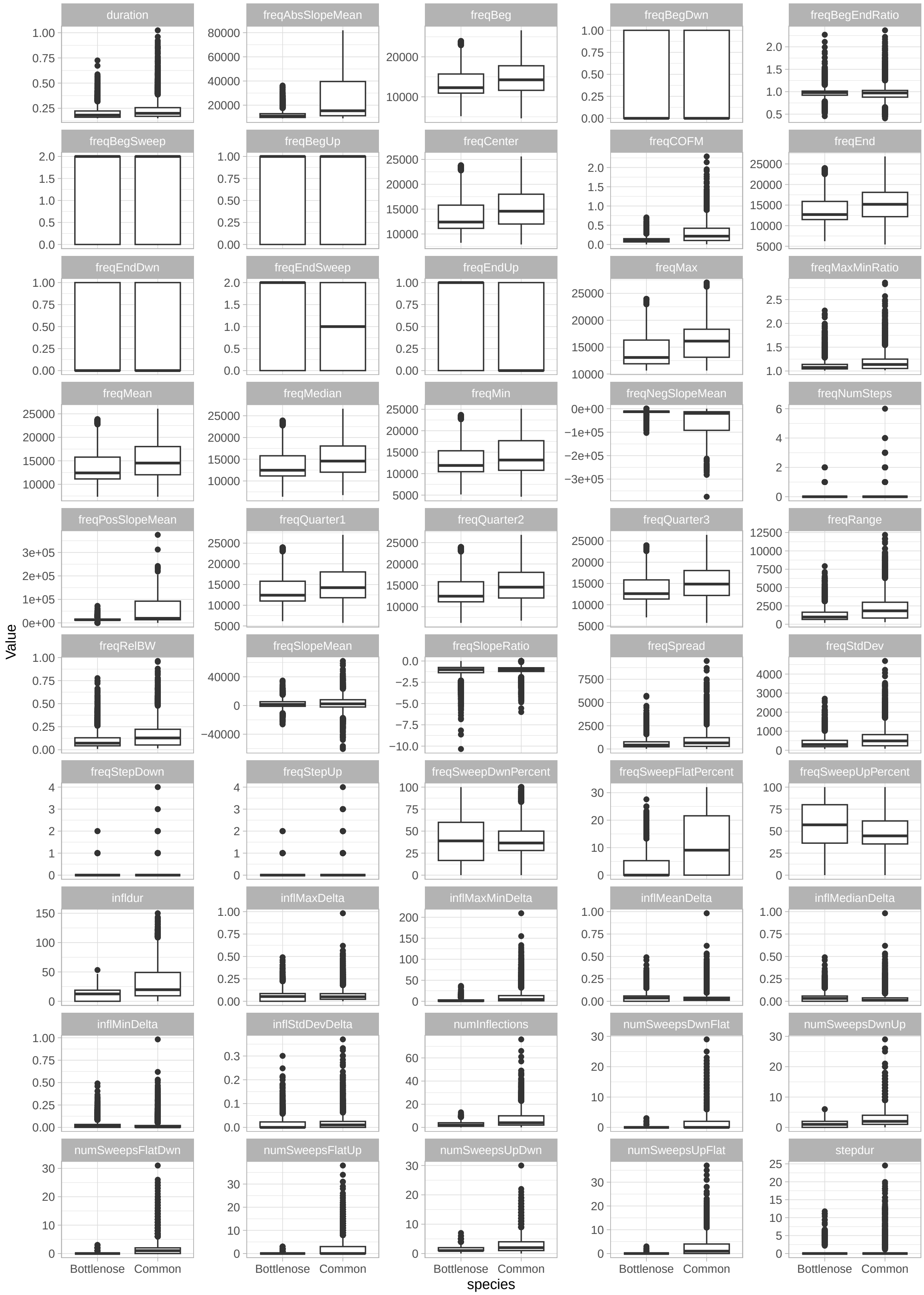


Figure S1: Boxplots of predictors for whistle classification, by species.

S1.2 Whistle characteristics by data source

Summaries of important whistle characteristics by source of the data:

```
#>
#>  Descriptive statistics by group
#> group: AMAPPS
#>                 vars    n     mean       sd   median       min      max     range  skew kurtosis     se
#> freqMaxMinRatio    1 2637     1.28     0.21     1.22      1.02     2.86      1.84  1.98     5.87   0.00
#> freqRelBW          2 2637     0.23     0.14     0.20      0.02     0.96      0.95  1.26     1.75   0.00
#> freqStdDev         3 2637   793.80   561.46   632.90     83.98  4686.51   4602.53  1.86     4.84  10.93
#> freqMean           4 2637 13473.15  3409.23 12581.90   7346.59 26094.23  18747.64  1.43     1.84  66.39
#> freqRange          5 2637  3017.21  1787.84  2625.00    375.00 12187.50  11812.50  1.32     2.07  34.82
#> freqSlopeRatio     6 2637    -1.02     0.26    -1.00     -4.11     0.00      4.11 -1.37    11.95   0.01
#> freqCenter         7 2637 13475.54  3344.78 12562.50   8250.00 25593.75  17343.75  1.42     1.78  65.13
#> freqCOFM           8 2637     0.49     0.27     0.45      0.04     2.29      2.25  1.31     3.01   0.01
#> duration           9 2637     0.23     0.09     0.20      0.15     0.86      0.71  2.20     6.71   0.00
#> freqSlopeMean     10 2637  3619.21 12880.54  5042.99 -60646.19 62068.97 122715.15 -0.38     1.84 250.83
#> freqBeg           11 2637 13025.95  3514.78 12187.50   5812.50 26625.00  20812.50  1.21     1.55  68.45
#> ------------------------------------------------------------------------------------------
#> group: T1C
#>                 vars    n     mean      sd   median       min      max    range  skew kurtosis    se
#> freqMaxMinRatio    1 8375     1.11    0.11     1.07      1.01     2.27     1.26  2.91    13.31  0.00
#> freqRelBW          2 8375     0.10    0.09     0.07      0.01     0.78     0.77  2.17     6.49  0.00
#> freqStdDev         3 8375   393.89  290.55   300.15     65.18  2705.87  2640.69  1.86     5.09  3.17
#> freqMean           4 8375 13992.09 3922.10 12415.76   7347.90 23874.02 16526.13  1.12     0.26 42.86
#> freqRange          5 8375  1253.45  872.75   984.38    187.50  7921.88  7734.38  1.82     4.56  9.54
#> freqSlopeRatio     6 8375    -1.15    0.69    -1.02    -10.34     0.00    10.34 -2.43    13.51  0.01
#> freqCenter         7 8375 13992.94 3909.31 12398.44   8226.56 23812.50 15585.94  1.13     0.26 42.72
#> freqCOFM           8 8375     0.12    0.08     0.10      0.00     0.70     0.70  1.83     5.50  0.00
#> duration           9 8375     0.20    0.05     0.18      0.15     0.72     0.58  2.30     8.02  0.00
#> freqSlopeMean     10 8375  2002.03 5697.00  1526.99 -26311.55 34804.40 61115.95  0.15     1.55 62.25
#> freqBeg           11 8375 13741.96 4089.48 12281.25   5156.25 23953.12 18796.88  1.01     0.16 44.69
#> ------------------------------------------------------------------------------------------
#> group: UFRJ
#>                 vars    n     mean      sd   median       min      max    range  skew kurtosis     se
#> freqMaxMinRatio    1 3580     1.11    0.12     1.07      1.02     2.49     1.48  2.71    13.01   0.00
#> freqRelBW          2 3580     0.10    0.09     0.07      0.02     0.85     0.84  1.92     5.13   0.00
#> freqStdDev         3 3580   479.42  410.80   326.63     79.90  3165.63  3085.74  1.80     4.38   6.87
#> freqMean           4 3580 16406.20 3248.64 17963.27   8545.31 23674.41 15129.10 -0.24    -0.59  54.30
#> freqRange          5 3580  1545.36 1269.20  1078.12    281.25 11671.88 11390.62  1.95     5.51  21.21
#> freqSlopeRatio     6 3580    -1.12    0.68    -1.00     -6.00     0.00     6.00 -1.67     5.17   0.01
#> freqCenter         7 3580 16390.47 3220.10 17929.69   9046.88 23671.88 14625.00 -0.24    -0.61  53.82
#> freqCOFM           8 3580     0.16    0.13     0.12      0.00     1.95     1.95  2.80    16.43   0.00
#> duration           9 3580     0.23    0.10     0.19      0.15     1.02     0.87  2.61     9.42   0.00
#> freqSlopeMean     10 3580  1691.33 6485.03  1573.15 -32856.26 25142.05 57998.30 -0.19     1.13 108.39
#> freqBeg           11 3580 16141.22 3353.66 17625.00   6703.12 23953.12 17250.00 -0.27    -0.55  56.05
#> ------------------------------------------------------------------------------------------
#> group: Watkins
#>                 vars   n     mean      sd   median      min      max    range  skew kurtosis     se
#> freqMaxMinRatio    1 230     1.28    0.24     1.21     1.03     2.46     1.44  2.07     5.56   0.02
#> freqRelBW          2 230     0.23    0.16     0.19     0.03     0.84     0.82  1.30     1.76   0.01
#> freqStdDev         3 230   879.26  607.46   688.81   116.79  3428.53  3311.73  1.03     0.68  40.05
#> freqMean           4 230 13124.69 2829.04 12717.09  7950.09 21488.50 13538.40  0.71     0.11 186.54
#> freqRange          5 230  2935.28 1928.54  2426.37   355.08  9243.75  8888.67  0.97     0.35 127.16
#> freqSlopeRatio     6 230    -1.06    0.75    -0.88    -4.63     0.00     4.63 -1.69     4.47   0.05
#> freqCenter         7 230 13190.45 2759.01 12841.99  7889.06 21316.41 13427.34  0.78     0.06 181.92
#> freqCOFM           8 230     0.30    0.21     0.24     0.04     1.30     1.26  1.42     2.42   0.01
```

```
#> duration            9 230     0.28     0.14     0.23      0.15     0.92     0.77  2.02     4.62   0.01
#> freqSlopeMean      10 230  1397.15 11554.53  2537.15 -41254.56 42737.23 83991.79 -0.09     1.08 761.88
#> freqBeg            11 230 12964.04  3141.67 12726.56   4621.88 21754.69 17132.81  0.53     0.35 207.16
```

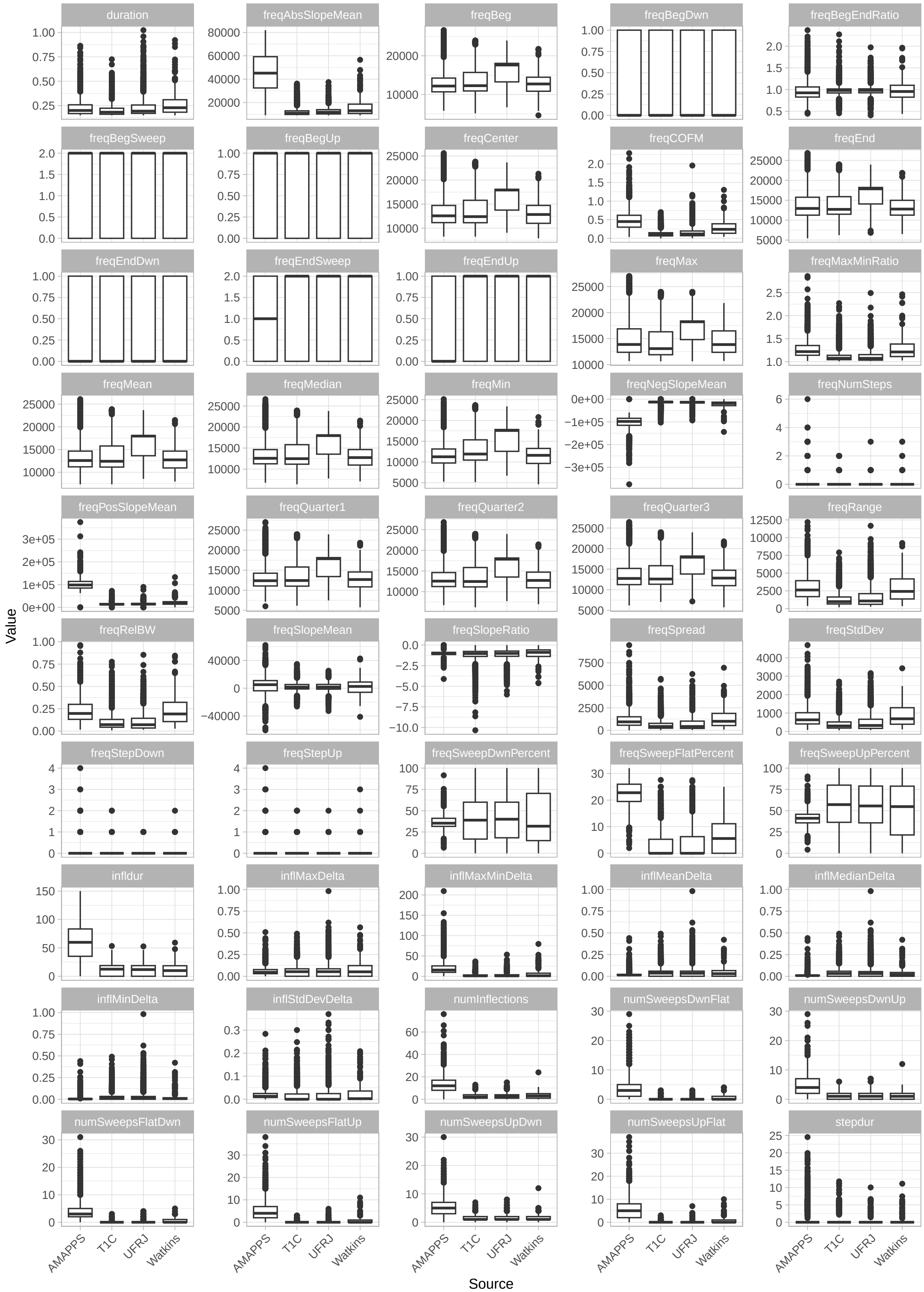


Figure S2: Boxplots of predictors for whistle classification, by source.

S1.3 Whistle characteristics for the manuscript summary table

Table S1: Summary of Whistle Characteristics by Source

AMAPPS	Watkins	UFRJ	Aggregated (UFRJ, AMAPPS, Watkins)	T1C
1.28 (0.21)	1.28 (0.24)	1.11 (0.12)	1.19 (0.18)	1.11 (0.11)
0.23 (0.14)	0.23 (0.16)	0.10 (0.09)	0.16 (0.13)	0.10 (0.09)
793.80 (561.46)	879.26 (607.46)	479.42 (410.80)	622.27 (511.32)	393.89 (290.55)
13473.15 (3409.23)	13124.69 (2829.04)	16406.20 (3248.64)	15089.43 (3614.75)	13992.09 (3922.10)
3017.21 (1787.84)	2935.28 (1928.54)	1545.36 (1269.20)	2196.97 (1692.38)	1253.45 (872.75)
-1.02 (0.26)	-1.06 (0.75)	-1.12 (0.68)	-1.07 (0.55)	-1.15 (0.69)
13475.54 (3344.78)	13190.45 (2759.01)	16390.47 (3220.10)	15084.02 (3568.88)	13992.94 (3909.31)
0.49 (0.27)	0.30 (0.21)	0.16 (0.13)	0.30 (0.26)	0.12 (0.08)
0.23 (0.09)	0.28 (0.14)	0.23 (0.10)	0.23 (0.10)	0.20 (0.05)
3619.21 (12880.54)	1397.15 (11554.53)	1691.33 (6485.03)	2469.39 (9841.61)	2002.03 (5697.00)
13025.95 (3514.78)	12964.04 (3141.67)	16141.22 (3353.66)	14753.64 (3748.58)	13741.96 (4089.48)

S1.4 Partial dependence plots from the selected model

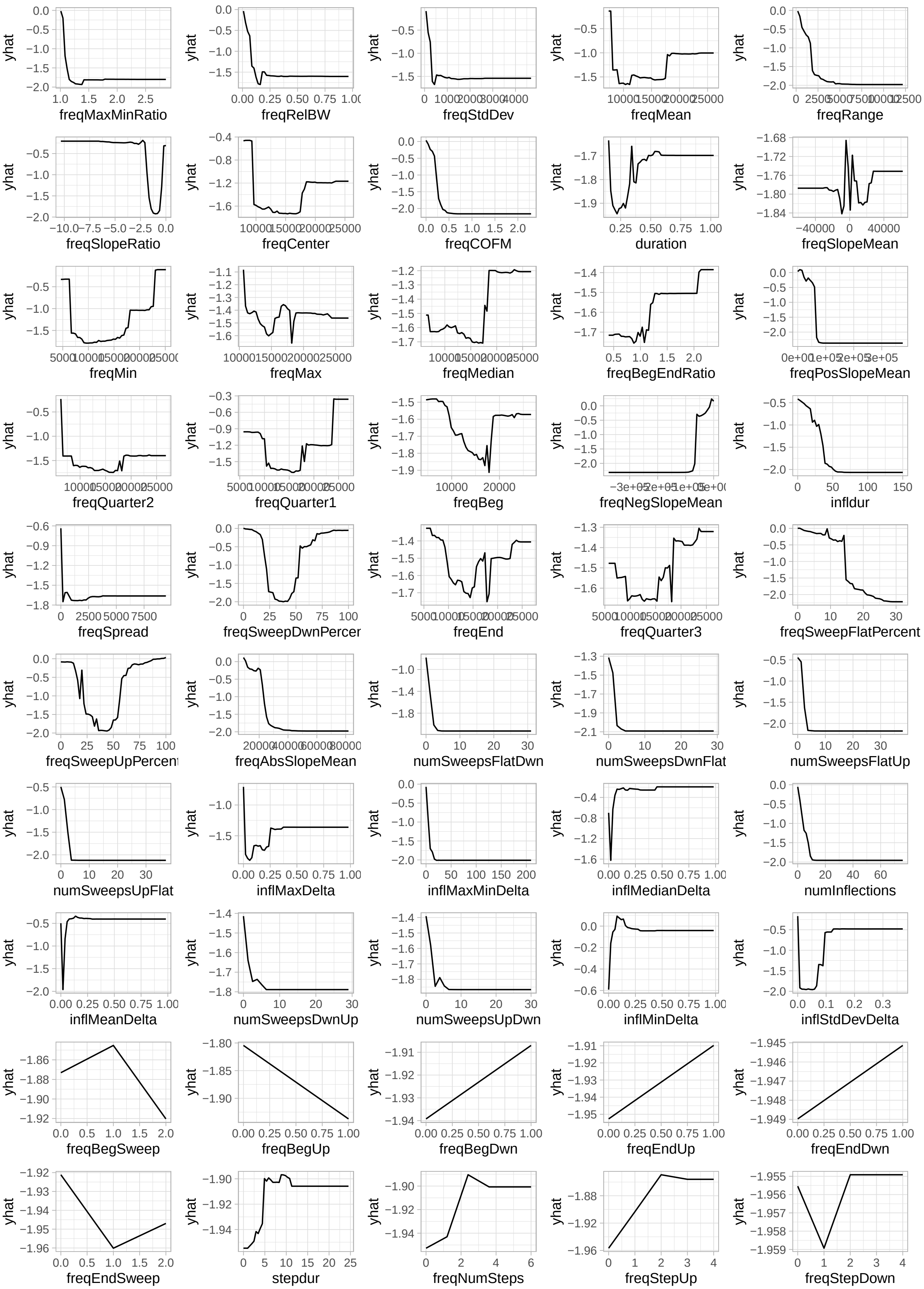


Figure S3: Partial dependence plots from the detector-level classification model, from the most to the least important predictor.