https://doi.org/10.1016/j.ecolind.2019.105793

hardRain: an R package for quick, automated rainfall detection in ecoacoustic datasets using a threshold-based approach.

Oliver C. Metcalf[1], Alexander C. Lees[1,2], Jos Barlow[3,4], Stuart J. Marsden[1], Christian Devenish[1].

1. Department of Natural Sciences, School of Science and the Environment, Manchester Metropolitan University, Manchester, UK
2. Cornell Lab of Ornithology, Cornell University, Ithaca, NY, USA
3. Lancaster Environment Centre, Lancaster University, Lancaster, Lancashire, UK
4. Departamento de Biologia, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil

Corresponding author: Oliver C. Metcalf email: o.metcalf@mmu.ac.uk

Abstract

The increasing demand for cost-efficient biodiversity data at large spatiotemporal scales has led to an increase in the collection of large ecoacoustic datasets. Whilst the ease of collection and storage of audio data has rapidly increased and costs fallen, methods for robust analysis of the data have not developed so quickly. Identification and classification of audio signals to species level is extremely desirable, but reliability can be highly affected by non-target noise, especially rainfall. Despite this demand, there are few easily applicable pre-processing methods available for rainfall detection for conservation practitioners and ecologists. Here, we use threshold values of two simple measures, Power Spectrum Density (amplitude) and Signal-to-Noise Ratio at two frequency bands, to differentiate between the presence and absence of heavy rainfall. We assess the effect of using different threshold values on Accuracy and Specificity. We apply the method to four datasets from both tropical and temperate regions, and find that it has up to 99% accuracy on tropical datasets (e.g. from the Brazilian Amazon), but performs less well in temperate environments. This is likely due to the intensity of rainfall in tropical forests and its falling on dense, broadleaf vegetation amplifying the sound. We show that by choosing between different threshold values, informed trade-offs can be made between Accuracy and Specificity, thus allowing the exclusion of large amounts of audio data containing rainfall in all locations without the loss of data not containing rain. We assess the impact of using different sample sizes of audio data to set threshold values, and find that 200 15s audio files represents an optimal trade-off between effort, accuracy and specificity in most scenarios. This methodology and accompanying R package 'hardRain' is the first automated rainfall detection tool for pre-processing large acoustic datasets without the need for any additional rain gauge data.

Keywords: Ecoacoustics, Environmental monitoring, Bioacoustics, Soundscape ecology, Rain detection, Acoustic pre-processing

https://doi.org/10.1016/j.ecolind.2019.105793

1.Introduction

Ecological questions are increasingly being answered using large datasets (Hampton et al., 2013; McCallen et al., 2019; Villanueva-Rosales et al., 2014), and faced with an ongoing biodiversity crisis, cost-effective collection of ecological data to address conservation challenges is vital (Gardner et al., 2008). The recent rapid development of cost-effective ecoacoustic sampling methods has facilitated collection of acoustic big data (Burivalova et al., 2019; Deichmann et al., 2018) and catalysed an increase in ecoacoustic monitoring. Despite the cost-effective nature of this sampling method (Deichmann et al., 2018; Hill et al., 2018), there are still significant challenges associated with the analysis of large acoustic datasets,. Automated detection and classification using machine or deep-learning techniques has been widely touted as one answer to this challenge (Priyadarshani et al., 2018). However, large datasets often require initial data cleaning to remove 'noise' (sounds which are not of interest, such as engines, wind and even electrical noises produced by the recorder (Stowell et al., 2016). The presence of hard rainfall (HR) is a significant contributor to noise as it can entirely mask all signals of interest or hinder their identification, and it can be especially problematic in both biodiverse and pluviose ecosystems such as tropical forests where our knowledge of biodiversity is most limited and acoustic data may be most useful. The use of acoustic indices, a common technique for quantifying biodiversity in large datasets without recourse to species level identification (Sueur et al., 2014; Towsey et al., 2014), have also been shown to be biased by the presence of heavy rainfall (Depraetere et al., 2012; Fairbrass et al., 2017; Towsey et al., 2014). Automated detection and excision of audio data at times of high rainfall is therefore often desirable before further analyses are undertaken, especially when using automated classifiers for detection of ecological sounds, as it reduces the potential for false identifications and increases processing time.

Despite the need for effective tools to identify and remove audio segments containing heavy rain, little research currently exists on the topic. Other published methods have different objectives; focussing on detection of rainfall as an objective in its own right (Brown et al., 2019), finding a proxy variable for quantification of total rainfall, or being designed to function in specific geographic areas to study the effect of rainfall within a wider soundscape (Bedoya et al., 2017). This has resulted in prioritising optimisation of accuracy of detection over ease of use and specificity. Other methods, such as the ecoacoustic event detection approach (Farina et al., 2018) allow a holistic approach to identification of all acoustic events, in which rainfall identification becomes a secondary benefit. We argue that many ecologists and conservation practitioners will primarily be interested in quickly identifying the majority of rain files rather than ascertaining the presence or absence of rain, to allow for better classification of ecological sounds and unbiased indices. For these users, the priority will be minimizing effort and maximising specificity –e.g. ensuring that false positive rates are very low so that ecological data are not removed from a dataset to achieve a higher overall accuracy of rainfall detection. Therefore, the most successful reported method of automated rainfall classification Brown et al. (2019), which involves a complex machine-learning approach and an extensive feature set, could be prohibitive for non-specialists. Many users may be willing to trade-off a small amount of accuracy in return for much lower analytical effort and greater ease of comprehension.

A simpler, quicker approach to classification has been proposed by Bedoya et al. (2017). This utilizes two acoustic measures indicative of rainfall taken at a single frequency band to set a decision threshold above which rainfall is determined to be present. However, this method uses minimum values over a period of acoustic data with rain of known intensity (using a rain gauge) to set the decision threshold. Obtaining verified rainfall data may not be possible in many cases, and requires additional cost and effort – especially in closed canopy ecosystems. Additionally the use of minimum values to set thresholds prioritizes accuracy over specificity, potentially leading to avoidably high

https://doi.org/10.1016/j.ecolind.2019.105793

false positive rates for relatively small gains in accuracy and the exclusion of potentially informative audio files. Setting threshold values from the second quartile of the interquartile range (Q2) may give more conservative predictions for the presence of HR, enabling a trade-off between higher specificity scores at the expense of accuracy. Furthermore, the amplitude of rainfall increases most noticeably at two frequency bands, 0.6-1.2 kHz and 4.4-5.6 kHz where the impact of raindrops hitting vegetation is most noticeable. Bedoya measures the indices at 0.6-1.2 kHz as light intensity rainfall is more noticeable, and it contains less biophony than the higher frequencies. However, it is unclear if the use of both of the frequency bands would produce better results when classifying only heavy rain, or in locations with higher levels of anthropophony (man-made noise).

Here we present a user-friendly methodology and associated R package (R Studio Team, 2015) 'hardRain', for automated rainfall detection that maintains high specificity and accuracy for use with new datasets. We build on the thresholding approach of Bedoya, developing a method to remove the need for any additional data from rain gauges to set threshold values. We investigate, at multiple tropical and temperate sites, whether using both 0.6-1.2 kHz and 4.4-5.6 kHz frequency bands provide greater accuracy and specificity than using only the lower frequency band, and assess the optimal number of files containing rainfall to use as training data from which to obtain threshold values. We also explore how differences in location affect classification results, and the trade-offs in accuracy and specificity when using minimum or Q2 values for setting decision thresholds.

2. Methods

2.1. Definition of rainfall

Identifying audio files containing rain without rain gauge data is not straightforward, as light rainfall can be indistinguishable from background noise (Bedoya et al., 2017). However, in these cases, rainfall is less likely to be less disruptive for the automated classification of ecological sounds. Here, we focus on the detection of heavy rainfall, here defined as rainfall that visually masks or significantly degrades other sound events (see Figure 1 for examples). Audio files were manually assigned as either 'Hard Rain (HR)' or 'Clear' through visual inspection of spectrograms in Raven Pro (Cornell Bioacoustics Research Program, 2010). For consistency, a single observer (OM) undertook all manual classifications in this paper.

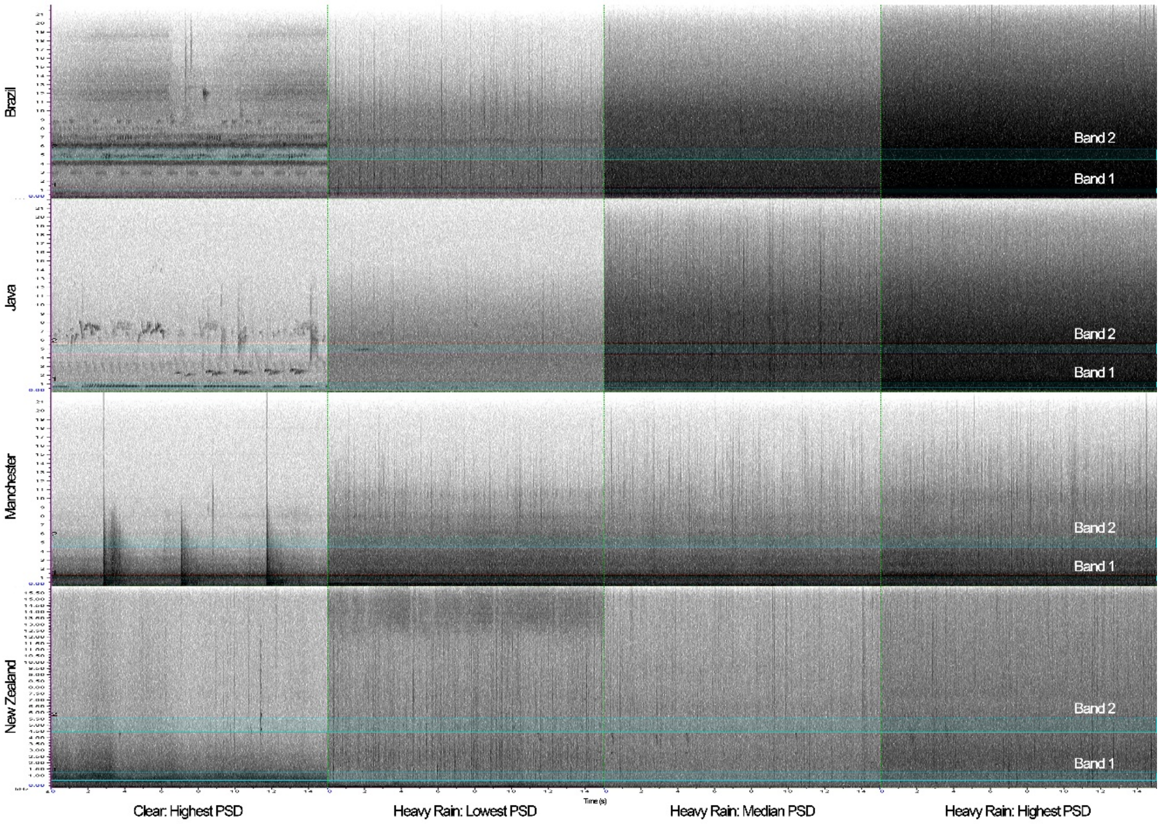

Figure 1. Examples of spectrograms assigned to rainfall present and absent taken from the combined training and test dataset of each country, ranked by power spectral density (PSD).

2.2 Data

This paper uses four primary datasets; two were collected in tropical rain forest; Santarém, Pará state, Brazil (-3.046, -54.947) and West Java, Indonesia (-6.181, 106.827), and two from temperate climates; one from temperate forests in Taranaki, New Zealand (-39.448, 174.414) and one from an urban balcony in Manchester, United Kingdom (53.485, -2.228). All include periods of time when both rainfall and clear weather were prevalent. The Brazil dataset comprises more than 10,000 hrs of data from 29 sites, the Java data set consists of more than 10,000 hours of data from 11 sites in montane forests in West Java with 12 recorders per site, Manchester over 600 hrs from one site and New Zealand over 3,900 hrs from 31 recorders at one site. For further information on data collection locations and durations at each of the sites see supplementary online material (SOM Table 1). Data were collected using Frontier Labs Bioacoustic Audio recorders (Frontier Labs, 2015), with the exception of the New Zealand dataset which used NZ Department of Conservation recorders (see Metcalf et al., 2019 for more information). All audio data were recorded at a sampling rate of 44.1 kHz except the New Zealand data set recorded at 32 kHz. All audio data were subdivided into 15 s sound files.

2.3 Threshold Setting and Optimisation

From each primary dataset, a training and test dataset were selected. The test and training datasets comprised 1000 files each. We manually selected 1,500 files that were then randomly split into 1,000 training files and 500 test files. A further 500 files that had been manually selected as being

https://doi.org/10.1016/j.ecolind.2019.105793

Clear (of Heavy Rainfall) were included in the test dataset, so that both the training and test dataset are composed of 1000 files. The Brazilian training dataset comprised 13 sites including both undisturbed primary and heavily degraded primary forests. The test dataset comprised eight sites and three sites for HR and Clear files respectively. Java training data came from 11 sites, whilst the test dataset used data from eights sites for HR and one site for Clear data. Manchester HR data were collected between 25th-28th April 2019, whilst Clear data was from 4th November 2018. The New Zealand training data were from 18 sites, whilst HR test data came from 16 sites and Clear from 18 sites.

We followed Bedoya et al., (2017) in using power spectral density (PSD) and signal-to-noise ratio (StN) as acoustic indices. The PSD of an acoustic file increases with rainfall intensity, while StN is useful to differentiate files that have high PSD because of continuous rainfall versus those that have high PSD because of non-continuous loud sound sources, such as biophony (e.g. animal vocalisations) or anthropophony. The PSD values in both 0.6-1.2 kHz and 4.4-5.6 kHz frequency bands were calculated for every file with the 'spectro' function from the seewave package in R (Sueur et al., 2008). The window length used to calculate PSD values was set to equal the duration of the audio file (typically 15 s segments – see package documentation; Figure 2 shows these values from the test datasets). We used mean divided by standard deviation of the PSD for the Signal-to-Noise ratio, following Bedoya et al., (2017), although we note a typographical error in point 3 of Algorithm 2.1 as the deviation of the mean is not squared in the standard deviation formula. See SOM Table 2 for all PSD and StN values for all training and test datasets.

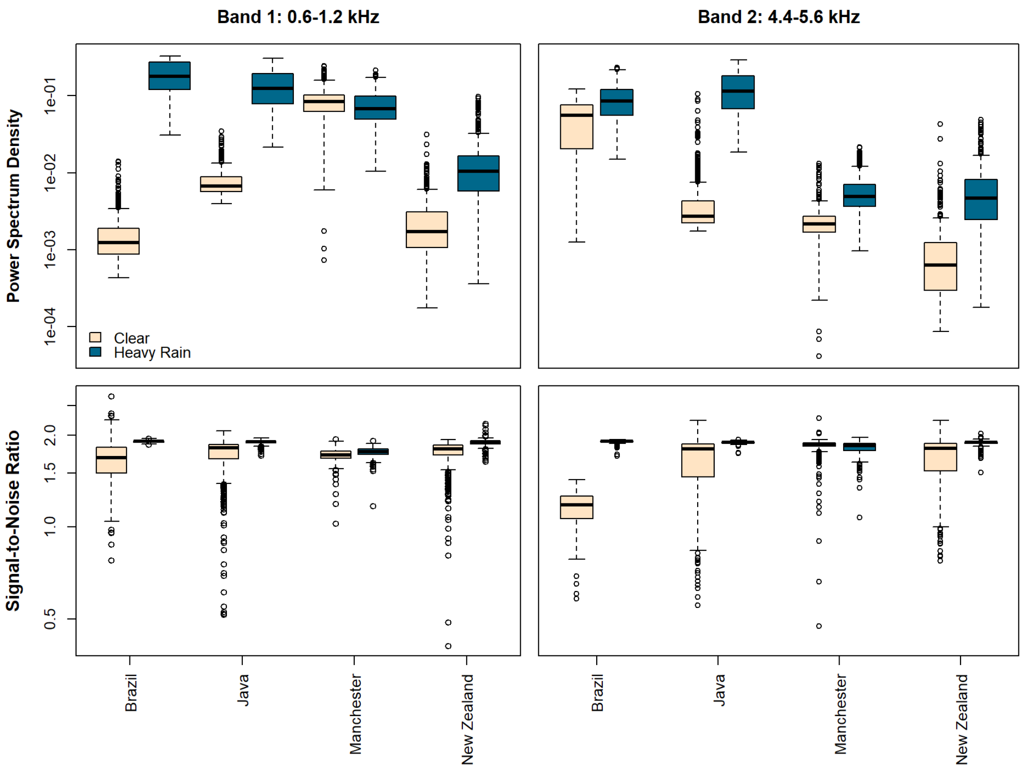

Figure 2 Power Spectral Density and Signal-to-Noise Ratio values for audio files containing heavy rain and clear files from the test datasets. The y-axes are presented on a log scale.

In predicting the presence of heavy rain, we followed Bedoya et al., (2017) in using thresholds for PSD and StN, so that if any of the measured values from an audio file exceed the threshold, they were predicted to contain heavy rain. We used mean balanced accuracy (Accuracy) and specificity

https://doi.org/10.1016/j.ecolind.2019.105793

(Specificity) (Velez et al., 2007) to assess the performance of classifier models. Although accuracy is the primary objective of classification, in some uses the penalty for the rejection of useable data (false-positives) may be far higher than the consequences of keeping files containing rain in the dataset (false-negatives), and specificity is the best measure for that circumstance (Fielding and Bell, 1997).

We tested classification performance using thresholds of PSD and StN from frequency band 1 (e.g. values had to exceed two thresholds to be classified as HR) against classification using PSD and StN from frequency bands 1 and 2 (e.g. values have to exceed four thresholds to be classified as HR) using a paired Wilcoxon rank test. To assess the effect, we took 100 subsamples of n=500 from each of the four countries' training datasets. Minimum and Q2 threshold values were then obtained and used to classify the applicable test dataset. Accuracy and specificity values were calculated by country, threshold choice and the mean of all countries combined.

To optimise the number of training samples required, we assessed the relationship between the number of training samples and accuracy/specificity with the aim of balancing the effort of manually selecting training data and the susceptibility of threshold values to outliers and variation in data sets. For each training dataset, 100 subsamples of size n= 10, 20, 30, 40, 50, 75, 100, then increasing increments of 50 to 1000, were taken and threshold values obtained using both frequency band 1 and 2 and these used to classify the applicable test dataset. Mean accuracy, specificity and their standard deviations were then calculated for each sample size by country and threshold choice. The sample size of n=500 was tested for significant differences in classification Accuracy and Specificity between the countries using Kruskal-Wallis and pairwise Wilcoxon tests, significant at <0.05.

In order to assess if there was overtraining between the test and training datasets, we conducted a case study using the Brazilian primary data. A random sample set of 6,960 files (1 hour from each transect), independent from the test and training data, was taken from the Brazilian primary dataset and manually labelled. A further subsample of 500 files was taken from the Brazilian training dataset to obtain threshold values, and these were used to predict the presence and absence of rainfall in the Brazilian random sample.

3.Results

The results produced by using both frequency bands were on average significantly better than those using just the 0.6-1.2 kHz band across both Specificity and Accuracy, with the exception of Accuracy when using the Q2 threshold, although results varied somewhat by country (Table 1). As Accuracy is not likely to be as important a consideration as Specificity for those choosing to use a Q2 threshold, using two frequency bands was deemed the better choice, and all further results discussed here are for classification with measurements taken from both frequency bands.

Table 1: Accuracy and Specificity scores by country, threshold choice, and number of frequency bands measured. 500 samples were used to set the thresholds.

Country	Mean Accuracy (%)				Mean Specificity (%)			
	Minimum threshold		Q2 Threshold		Minimum threshold		Q2 Threshold	
	1 band	2 bands	1 band	2 bands	1 band	2 bands	1 band	2 bands
Brazil	**99.69±0.00**	**99.67±0.00**	**83.10±0.01**	**69.36±0.01**	100±0.00	100±0.00	100±0.00	100±0.00
Java	99.76±0.00	99.75±0.00	**87.13±0.01**	**71.31±0.01**	**99.80±0.00**	**100±0.00**	100±0.00	100±0.00

Manchester	54.81±0.01	55.73±0.01	79.39±0.01	67.77±0.00	10.15±0.01	12.60±0.01	91.05±0.01	97.39±0.00
New Zealand	51.75±0.03	60.14±0.03	82.65±0.01	72.61±0.01	3.66±0.05	20.49±0.06	98.00±0.00	100±0.00
Mean	76.50±0.01	78.83±0.01	83.07±0.01	70.26±0.01	53.40±0.02	58.27±0.02	97.27±0.00	99.35±0.00

Results with significant differences (corrected p-value <0.05) between one and two bands are in bold. All differences in which two bands performed better than one band are shaded. A table of the p-values can be found in supplementary online material (SOM Table 3).

Detection responses to sample size varied both by country and by the choice of threshold value, but were consistent across Specificity and Accuracy metrics. When using minimum threshold values, Accuracy showed rapid increases until an asymptote at 200 samples for Brazil and Java, but declines for Manchester and New Zealand (Figure 3a). Specificity reaches 100% for all samples sizes in the Brazil and Java datasets, but follows a similar, but steeper trend to Accuracy for Manchester and New Zealand (not shown in Fig 3). Using the Q2 threshold, Specificity is at 100% for all sample sizes for Brazil and Java and New Zealand and around 97% for Manchester (Fig 3b), whilst Accuracy reaches stable scores for all countries between 100 and 200 samples (Fig 3c). Full tables of results are available in SOM Tables 4 and 5.

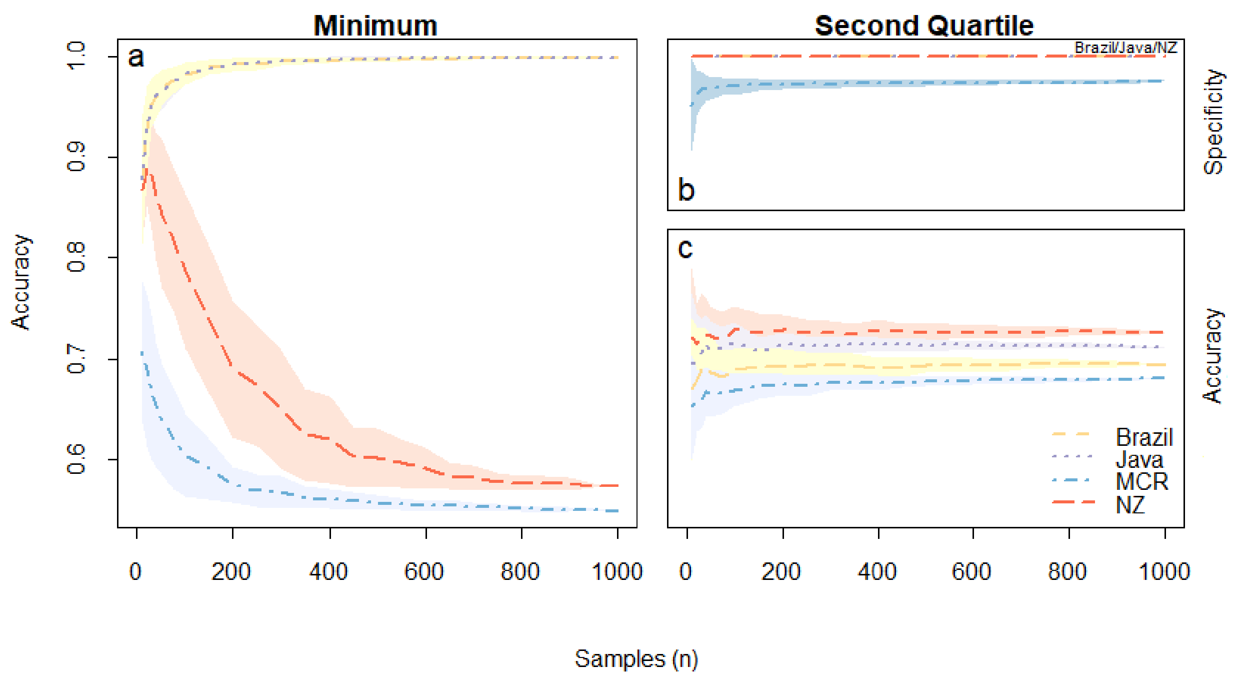

Figure 3 Selected Accuracy and Specificity scores by sample size (n), country and threshold selection method. Specificity scores for minimum threshold method not shown as Specificity=1 for all sample sizes in Brazil and Java data, and below 0.5 for almost all sample sizes in Manchester and New Zealand datasets. The shading represents standard deviation of 100 repetitions. NZ= New Zealand, MCR=Manchester.

Comparison between country scores showed that there were significant pairwise differences between all countries for both threshold choices in Accuracy and Specificity, except where Specificity was at 100% (Table 1; also see SOM Table 3). As expected, there was no clear threshold value choice to maximise both Specificity and Accuracy across all countries. The best Accuracy scores were achieved using Minimum threshold values, >99% for all training sample sizes over 200 for both Brazil and Java but this performed poorly for Manchester and New Zealand (Table 1, Fig 3). This suggests that in some countries, the differentiation is not enough to achieve high levels of Accuracy even when excellent Accuracy scores are achieved with the same method in other locations. Using the Q2 threshold, Accuracy was low for all countries (between 65% and 73%). Despite this, high Specificity scores can be achieved for all countries using the Q2 threshold (Table 1, Fig 3). This highlights that even in datasets where there may be poor distinction between Clear and HR data using PSD and StN indices, 35-50% of all HR files can be identified with loss of less than 5% of data containing no rain. Confusion matrices are provided in Table 2 for the mean scores of a sample size of 500 training files applied to the Manchester and New Zealand test datasets using second quartile thresholds.

Table 2. Confusion matrices with 500 samples of training data using second quartile threshold values.

	Manchester - testing dataset			New Zealand – testing dataset	
	Second Quartile Threshold			Second Quartile Threshold	
	Actual Class				
Predicted Class		TRUE	FALSE	TRUE	FALSE
	TRUE	185	15	230	0
	FALSE	315	485	270	500
	Sensitivity=38.15%, Specificity=97.39%, Accuracy=67.77%			Sensitivity=45.22%, Specificity=100%, Accuracy=72.61%	

The results for classification of the case study using 6,960 files of the Brazilian dataset remained good, although lower than the test scores suggesting a small amount of overtraining between the test and training datasets (Table 3). To read in, measure and classify all 6960 files took 15 min 16 s using a Dell EliteBook laptop with a 4-core Intel Core i7-7600U CPU and 16 GB RAM running Windows 10.

Table 3. Matrix of the Brazilian case study

	Brazil - 6960 randomly selected audio files				
	Minimum Threshold			Second Quartile Threshold	
	Actual Class				
Predicted Class		TRUE	FALSE	TRUE	FALSE
	TRUE	88	14	33	0
	FALSE	22	6836	69	6858
	Sensitivity=86.27%, Specificity=99.68%, Accuracy=92.98%			Sensitivity=32.35%, Specificity=100%, Accuracy=66.18%	

Data are a random sample of the entire audio dataset (n=6960, HR n=102) with threshold values taken from 500 randomly selected audio files from the Brazilian training dataset.

4.Conclusions

We have shown that it is possible to fully automate rainfall identification within audio data from tropical environments using only two simple measurements at two frequency bands, and requiring only a relatively small set of files containing known rainfall to extract threshold values. We also demonstrate that by using different thresholds, minimum and second quartile, the technique can be adjusted for use even in cases where there is poor differentiation between rain presence and absence with a reasonably high level of success. This means that users of hardRain can make informed trade-offs between effort, accuracy and specificity.

The effectiveness of the method is clearly dependent on sample sizes, with standard deviations declining with increasing samples, but divergent impact on Accuracy by site and threshold selection method. Whilst it is possible to devise various stopping rules to optimise the sample number, the optimal solution will vary with the ease of obtaining training files containing rain and the objectives of individual research projects. The standard deviation of Accuracy and Specificity is relatively low for almost all measures at 200 samples (Fig.3, SOM Table 5), with corresponding accuracy and specificity scores close to their maximum for the tropical datasets when using minimum threshold values, and for all datasets when using second quartile values.

Using only PSD and StN as measurements to differentiate between rain presence and absence has clear advantages in minimising effort and ease of understanding. Along with Brown et al., (2019), we did not find StN to be a useful index for classification when we initially analysed our data using the printed formula in Bedoya et al., (2017). However, when we used the standard formula for standard deviation, the use of both PSD and StN was better than just PSD. In some circumstances, even the use of both indices resulted in poor differentiation. This is especially the case for datasets from temperate climates, with Manchester and New Zealand performing worse, presumably due to poorer distinction between PSD scores (Fig 2). This is possibly because rainfall is less intense at these locations, or because rain falling on to predominately concrete (Manchester) and more open temperate forest canopies (New Zealand), results in less amplification than in tropical forests (Java and Brazil). Despite this shortcoming, by using second quartile thresholds between 40-50% of rain data was identified even in Manchester and New Zealand, with no or only a very small percentages of rain-free data misidentified (Table 2).

Although not herein directly compared, our methodology is unlikely to match the AUC scores of the method proposed by Brown et al., (2019) or the accuracy and quantification of Bedoya et al., (2017). For those scholars studying rain through audio data, or requiring extremely precise cleaning, these would be better methods to use. However, our methodology provides a quick and effective classification method that can be applied to audio data, and is especially suited to tropical forests where the need for reliable acoustic data on biodiversity is greatest and rainfall is frequent. For researchers wishing to quickly remove rain files from large datasets prior to classification, this method will often represent the most time-effective way to do so. Additionally for research in which the penalty of false-negatives is far lower than that of false positives, this method of rain detection allows for informed trade-offs between Accuracy and Specificity which previous methods of rain detection do not.

Package description

To facilitate the use of this rain detection method, we have developed the R package ‘hardRain’. The package will i) set thresholds (based on training data consisting of short segments of known rain audio recordings), ii) apply the thresholds to audio data and identify presence of rain in each input file, or subdivisions therein, iii) cut audio segments with rain and save the remaining segments, and optionally, create a label file view in Audacity or Raven software. It can also be used to test the accuracy of the classification using known testing and training data. The package consists of four main functions (Table 4).

Table 4. Functions in the R package 'hardRain'.

Function	Description	Main inputs
getThreshold	This function measures PSD and Signal-to-Noise Ratio on all input training files at two frequency bands (defaults to 0.6-1.2 kHz and 4.4-5.6 kHz) and calculates minimum and 2nd quartile thresholds over these.	wav filenames (and locations where these are stored) of audio segments of known rain, i.e. training data (see above for discussion on how many files are needed), but typically 200 wav files of about 15 s duration
classifyRain	This function takes the testing data, calculates the PSD and Signal-to-Noise Ratio and applies the thresholds produced by getThreshold function and classifies each input file (or subdivision thereof) for the presence / absence of rain. Optionally, if the function is used for accuracy testing, a label can be included denoting which files have presence of rain or not.	wav filenames (and locations) of testing data files may be of short duration already (typically, 15-30 s segments) or may be provided as much longer files (e.g. 2-3 hours) and split into segments within the function, using the t.step argument (division size, in seconds); thresholds from getThreshold()
cutRain	This function takes the output from classifyRain() and cuts out the segments identified as rain in the input wav files and saves the remaining contiguous audio in a new folder and writes a label file for the original length audio file, marking segments with no rain (either or both of these options are available). Optionally, the new start time of each file can be recorded in the filename.	output from classifyRain() -only when longer files are classified in subdivisions; output location for new wav files.
getMetrics	This function does not generally need to be called directly. It is the workhorse function that reads wav files, extracts PSD and Signal-to-Noise for specified frequency bands using seewave function spectro(). This function is called by getThreshold() and classifyRain() which will generally be used directly.	wav filenames (and locations); time division (in seconds) to subdivide wav input files for analysis (optional)

The package can be downloaded from: https://github.com/Cdevenish/hardRain

Before using the classify function it is necessary to decide which threshold values to use. If it is reasonable to make assumptions about the distinction between rain presence and absence, for instance if the data is collected in tropical rain forest, then the threshold can be selected and the results checked after. However, if it is unclear whether there will be a good distinction, accuracy can be tested using the classifyRain function with known testing and training data (i.e. labelled audio segments of heavy rain or clear) and confusion matrices and accuracy metrics produced (see example in vignette).

See vignettes included in the package for further details on functionality.

Acknowledgements: We are very grateful for the insightful comments of Carol Bedoya who kindly provided feedback on the manuscript. We are also grateful to J. Ferreira, E. Berenguer, L. Rossi and F.França for logistical field support in Brazil, R. Junaid and G. C. Apriantoin Java, M. Mcready, M. Rowcliffe, J. Ewen, E. Williams and S. Collinson, New Zealand, and M. Loroño-Leturiondo and G. Abercrombie in the UK. Fieldwork in Brazil was supported by research grants ECOFOR(NE/K016431/1), and AFIRE (NE/P004512/1), PELD-RAS (CNPq/CAPES/PELD 441659/2016-0) and in Java through funding from the Rainforest Trust and Chester Zoo.

References

Bedoya, C., Isaza, C., Daza, J.M., López, J.D., 2017a. Automatic identification of rainfall in acoustic recordings. Ecol. Indic. https://doi.org/10.1016/j.ecolind.2016.12.018

Bedoya, C., Isaza, C., Daza, J.M., López, J.D., 2017b. Automatic identification of rainfall in acoustic recordings. Ecol. Indic. 75, 95–100. https://doi.org/10.1016/J.ECOLIND.2016.12.018

Bioacoustics Research Program, 2010. Raven Pro: Interactive Sound Analysis Software (Version 1.4) [Computer software].

Brown, A., Garg, S., Montgomery, J., 2019. Automatic rain and cicada chorus filtering of bird acoustic data. Appl. Soft Comput. J. 81, 105501. https://doi.org/10.1016/j.asoc.2019.105501

Burivalova, Z., Game, E.T., Butler, R.A., 2019. The sound of a tropical forest. Science (80-.). 363, 28–29. https://doi.org/10.1126/science.aav1902

Deichmann, J.L., Acevedo-Charry, O., Barclay, L., Burivalova, Z., Campos-Cerqueira, M., d'Horta, F., Game, E.T., Gottesman, B.L., Hart, P.J., Kalan, A.K., Linke, S., Nascimento, L. Do, Pijanowski, B., Staaterman, E., Mitchell Aide, T., 2018. It's time to listen: there is much to be learned from the sounds of tropical ecosystems. Biotropica 50, 713–718. https://doi.org/10.1111/btp.12593

Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., Sueur, J., 2012. Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland. Ecol. Indic. 13, 46–54. https://doi.org/10.1016/j.ecolind.2011.05.006

Fairbrass, A.J., Rennett, P., Williams, C., Titheridge, H., Jones, K.E., 2017. Biases of acoustic indices measuring biodiversity in urban areas. Ecol. Indic. 83, 169–177. https://doi.org/10.1016/j.ecolind.2017.07.064

Farina, A., Gage, S.H., Salutari, P., 2018. Testing the ecoacoustics event detection and identification (EEDI) approach on Mediterranean soundscapes. Ecol. Indic. 85. https://doi.org/10.1016/j.ecolind.2017.10.073

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49. https://doi.org/10.1017/S0376892997000088

Frontier Labs, 2015. Bioacoustic Audio Recorder User Guide.

Gardner, T.A., Barlow, J., Araujo, I.S., Cristina Vila-Pires, T., Bonaldo, A.B., Costa, J.E., Esposito, M.C., Ferreira, L. V, Hawes, J., Hernandez, M.I.M., Hoogmoed, M.S., Leite, R.N., Lo-Man-Hung, N.F., Malcolm, J.R., Martins, M.B., Mestre, L.A.M., Miranda-Santos, R., Overal, W.L., Parry, L., Peters, S.L., Antô Nio Ribeiro-Junior, M., Da Silva, M.N.F., Da, C., Motta, S., Peres, C.A., 2008. The cost-effectiveness of biodiversity surveys in tropical forests. Ecol. Lett. 11, 139–150. https://doi.org/10.1111/j.1461-0248.2007.01133.x

Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S., Porter, J.H., 2013. Big data and the future of ecology. Front. Ecol. Environ. https://doi.org/10.1890/120103

Hill, A.P., Prince, P., Piña Covarrubias, E., Doncaster, C.P., Snaddon, J.L., Rogers, A., 2018. AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. Methods Ecol. Evol. https://doi.org/10.1111/2041-210X.12955

McCallen, E., Knott, J., Nunez-Mir, G., Taylor, B., Jo, I., Fei, S., 2019. Trends in ecology: shifts in ecological research themes over the past four decades. Front. Ecol. Environ. https://doi.org/10.1002/fee.1993

Metcalf, O.C., Ewen, J.G., McCready, M., Williams, E.M., Rowcliffe, J.M., 2019. A novel method for using ecoacoustics to monitor post-translocation behaviour in an endangered passerine. Methods Ecol. Evol. https://doi.org/10.1111/2041-210X.13147

Priyadarshani, N., Marsland, S., Castro, I., 2018. Automated birdsong recognition in complex acoustic environments: a review. J. Avian Biol. https://doi.org/10.1111/jav.01447

Stowell, D., Wood, M., Stylianou, Y., Glotin, H., 2016. Bird detection in audio: A survey and a challenge, in: IEEE International Workshop on Machine Learning for Signal Processing, MLSP. https://doi.org/10.1109/MLSP.2016.7738875

Sueur, J., Farina, A., Gasc, A., Pieretti, N., Pavoine, S., 2014. Acoustic indices for biodiversity assessment and landscape investigation. Acta Acust. united with Acust. 100. https://doi.org/10.3813/AAA.918757

Towsey, M., Wimmer, J., Williamson, I., Roe, P., 2014. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. Ecol. Inform. 21, 110–119. https://doi.org/10.1016/j.ecoinf.2013.11.007

Velez, D.R., White, B.C., Motsinger, A.A., Bush, W.S., Ritchie, M.D., Williams, S.M., Moore, J.H., 2007. A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. Genet. Epidemiol. 31, 306–15. https://doi.org/10.1002/gepi.20211

Villanueva-Rosales, N., Peters, D.P.C., Fuentes, O., Cushing, J., Tweedie, C., Havstad, K.M., 2014. Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. Ecosphere. https://doi.org/10.1890/es13-00359.1