Received: 25 October 2020 | Accepted: 10 December 2020

DOI: 10.1111/2041-210X.13548

RESEARCH ARTICLE

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Automated classification of avian vocal activity using acoustic indices in regional and heterogeneous datasets

Daniel A. Yip[1,2] | C. Lisa Mahon[1,2] | Alexander G. MacPhail[2] | Erin M. Bayne[2]

[1]Environment and Climate Change Canada, Whitehorse, YT, Canada

[2]Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada

Correspondence
Daniel A. Yip
Email: dayip@ualberta.ca

Funding information
Alberta Biodiversity Monitoring Institute; Environment and Climate Change Canada

Handling Editor: Robert Freckleton

Abstract

1. Acoustic indices combined with clustering and classification approaches have been increasingly used to automate identification of the presence of vocalizing taxa or acoustic events of interest. While most studies using this approach standardize data collection and study design parameters at the project or study level, recent trends in ecological research are to investigate patterns at regional or continental scales. Large-scale studies often require collaboration between research groups and integration of data from multiple sources to fulfil objectives, which can lead to variation in recording equipment and data collection protocols.
2. Our objectives were to determine how analytical approaches and variation in data collection and processing that is typical of regional acoustic monitoring programmes influences accuracy when identifying vocal activity in breeding birds. We used data from three regional datasets in Northern Alberta, Northern British Columbia, and Southern and Central Yukon, Canada to investigate the effect of analytical framework, sample size, local species richness and data collection variables on classification accuracy.
3. We found supervised classification approaches to be the most effective, with boosted regression trees identifying vocalizations of breeding birds in audio recordings with a 92.0% accuracy and easily able to accommodate variation in data collection and processing parameters. We also provide recommendations on effectively processing large and heterogeneous datasets including sufficient sample size, accommodating potentially confounding variables and selecting suitable model training data.
4. The results presented in this study can help inform decisions in data collection, data processing, and study design and analysis, maximize performance and accuracy during analysis, and efficiently process large, heterogeneous acoustic datasets to answer questions at scales previously difficult to investigate.

KEYWORDS

autonomous recording unit, bioacoustics, bird survey, data integration, ecoacoustics, monitoring, seasonal phenology

1 | INTRODUCTION

Ecoacoustics is frequently used to study the diversity, distribution and behaviour of acoustic animal communities (Sueur & Farina, 2015). In comparison to bioacoustics, which generally views sound as the transmission of information within a signal (Fletcher, 2007), ecoacoustics considers all components of the soundscape, including biophony (animal sounds), geophony (naturally occurring, abiotic sounds) and anthrophony (sounds attributed to human activity; Pijanowski et al., 2011). To quantify patterns within soundscapes, acoustic indices are used to describe the contents of audio recordings by measuring patterns in the distribution and structure of acoustic energy (Phillips et al., 2018; Towsey et al., 2014). These indices can be either singular or vectors of multiple values which quantify different frequency bands of a recording (Towsey, 2017), and are often analogous to metrics used in traditional ecological research such as species diversity or animal abundance (Browning et al., 2017; Gibb et al., 2018; Towsey, Wimmer, et al., 2014), and to date, over 60 acoustic indices have been described in the literature (Buxton et al., 2018).

Traditional bioacoustic monitoring requires the identification of individual species and typically has high effort or processing requirements. Multi-species recognizer and classification approaches are often limited in scope and are specific to certain species or locations, meaning that individual species must often be validated manually (Bradfer-Lawrence et al., 2019). In comparison, the calculation of acoustic indices is efficient, easily automated, and scales well to large acoustic datasets and long-duration recordings (Bradfer-Lawrence et al., 2019, 2020; Buxton et al., 2018). In the past few years, acoustic indices have also been used to identify the presence of vocalizing taxa or acoustic events of interest using clustering and classification approaches (Oliver et al., 2018; Phillips et al., 2018). This approach allows for automated signal processing and classification by relating acoustic indices, or the acoustic state of a soundscape (Phillips et al., 2018), to the presence of vocalizing taxa or acoustic events of interest with relatively little effort. Improvements in storage capacity and ease of use in audio recording technology have significantly increased the volume of bioacoustic data being collected. Manual processing or transcription of such large quantities of data is expensive and time-consuming, and automated processing methods are becoming increasingly important (Buxton et al., 2018).

Clustering and classification approaches are well documented, computationally efficient and have the flexibility to accommodate a variety of covariates, predictor variables and outcome types. Identifying taxa and events using acoustic indices can be done using unsupervised or supervised approaches. With unsupervised approaches, acoustic indices are calculated for audio recordings with unknown contents. Recordings are split into groups based on values of different acoustic indices using clustering techniques such as k-means or hierarchical clustering (Oliver et al., 2018; Phillips et al., 2018; Sankupellay et al., 2015). In the case of supervised approaches, outcomes of interest are identified for each audio recording beforehand, and acoustic indices are modelled using analyses such as linear discriminant analysis, classification trees and support vector machines to predict the contents of each recording (Bellisario et al., 2019; Ellis et al., 2011; Oliver et al., 2018). In both cases, audio recordings require validation to determine the contents of clustered recordings or the accuracy of model predictions.

While this approach to automated signal processing has a great deal of potential, inconsistent methods and a lack of standardized procedures can obscure the effectiveness and accuracy of classification and clustering by introducing systematic biases which can overwhelm the true outcome of interest. A conceptual weakness in this approach has been the uncertainty in linking soundscape components to acoustic indices (Browning et al., 2017; Gibb et al., 2018), although recent work has advanced our understanding of which acoustic indices are appropriate for monitoring wildlife (Bradfer-Lawrence et al., 2020; Eldridge et al., 2018). Several methods for calculating acoustic indices are readily available (Sueur, Aubin & Simonis, 2008; Sueur, Pavoine, et al., 2008; Towsey et al., 2018; Villanueva-Rivera & Pijanowski, 2018) and existing studies use a variety of audio parameters such as different combinations of acoustic indices and types of audio recording units (Oliver et al., 2018; Phillips et al., 2018). Acoustic indices are sensitive to changes in recording parameters (Bradfer-Lawrence et al., 2019), weather conditions (Farina et al., 2011; Sanchez-Giraldo et al., 2020) and background or anthropogenic noise (Fairbrass et al., 2017). How this influences classification performance of audio recordings and soundscape interpretation is not well understood. For example, indicators such as the Acoustic Complexity Index (ACI) which are often used to measure diversity can also be positively correlated to rain and weather events in some studies (Bradfer-Lawrence et al., 2019), but have shown to have relatively little effect in others (Sanchez-Giraldo et al., 2020). Extensive investigation and validation of the relationship between individual indices and ecological community data are important for the effective use of acoustic indices (Harris et al., 2016) and often requires project-specific calibration (Sueur, Pavoine, et al., 2008). As a result, further assessment of the effectiveness of this method to monitor multiple habitats and taxa is needed (Gasc et al., 2013). Early practitioners of ecoacoustics used individual acoustic indices to investigate ecological questions (e.g. Lellouch et al., 2014); however, the most recent consensus is that a suite of acoustic indices is necessary to fully interpret the contents of a soundscape (Phillips et al., 2018; Rychtáriková & Vermeir, 2013; Towsey, Zhang, et al., 2014). Bradfer-Lawrence et al. (2019) present some of the first recommendations on standardized use of acoustic indices for assessment of habitat types and daily diel patterns, and several other recent reviews of the literature provide recommendations on the effective use of bioacoustics for monitoring wildlife (Gibb et al., 2018; Sugai et al., 2019; Sugai et al., 2019). Collaborative monitoring efforts and integration of multiple datasets have emerged as important next steps for the future of bioacoustic monitoring (Gibb et al., 2018). However, additional research and recommendations on best practices for collating data from multiple, unstandardized sources are critical if acoustic indices are to be used to identify seasonal or spatial patterns, such as phenology or habitat associations, at large spatial scales.

Most studies using acoustic indices standardize data collection, calibration and validate results at the project or study level (Browning et al., 2017; Gibb et al., 2018; Sueur, Pavoine, et al., 2008). However, recent trends in ecological research are to investigate patterns at regional scales (Bixler et al., 2016; Buxton et al., 2018; Soulé & Terborgh, 1999). Large-scale studies often require collaboration between research groups and integration of data from multiple sources to fulfil objectives, which can lead to variation in recording equipment and data collection protocols. Additionally, regional monitoring programmes can cover landscapes with a wide range of species assemblages, richness and habitat types, and can include heterogeneous vegetation types and environmental conditions. How variation in recording technology, data collection protocols and landscapes in different geographical areas influences the performance of acoustic indices, specifically for investigating seasonal patterns of acoustic energy through classification and clustering approaches, is unknown.

In this study, we used acoustic indices to identify the presence of avian vocal activity, which we define as the presence of acoustic signals on audio recordings originating from any breeding bird, using three regional datasets of acoustic surveys from the boreal forest biome in Northern Alberta, Northern British Columbia, and Southern and Central Yukon, Canada. Our objectives were to determine how analytical approaches, recording hardware and local species richness influence classification accuracy when identifying periods of vocal activity for breeding birds from audio recordings containing a range of conditions including different weather events and non-target taxa. Specifically, we investigated how classification accuracy for the presence or absence of breeding birds is influenced by (a) the contribution of various acoustic indices used in analysis, (b) analysis using supervised or unsupervised frameworks, (c) site-level species richness used in training and testing datasets, (d) the sample size of audio recordings and (e) individual monitoring programmes and the recording equipment used when recording a survey. Our results will provide guidance to regional monitoring programmes and wildlife managers, and help inform study design, data processing and analysis of large ecoacoustic datasets.

2 | MATERIALS AND METHODS

2.1 | Data collection and processing

We obtained recordings of breeding bird surveys from the Alberta Biodiversity Monitoring Institute's (ABMI) biodiversity monitoring programme (Burton et al., 2014), Canadian Wildlife Service's Yukon Boreal Monitoring (BM) programme (Van Wilgenburg et al., 2020) and the University of Alberta and Canadian Wildlife Service's High Elevation Monitoring (HEM) programme (unpublished results). Recordings were collected from the boreal regions of Northern Alberta, Northern British Columbia, and Southern and Central Yukon Territory between March and August 2016–2019 (Figure 1) using programme-specific standardized protocols (Table 1). Each monitoring programme used different recording

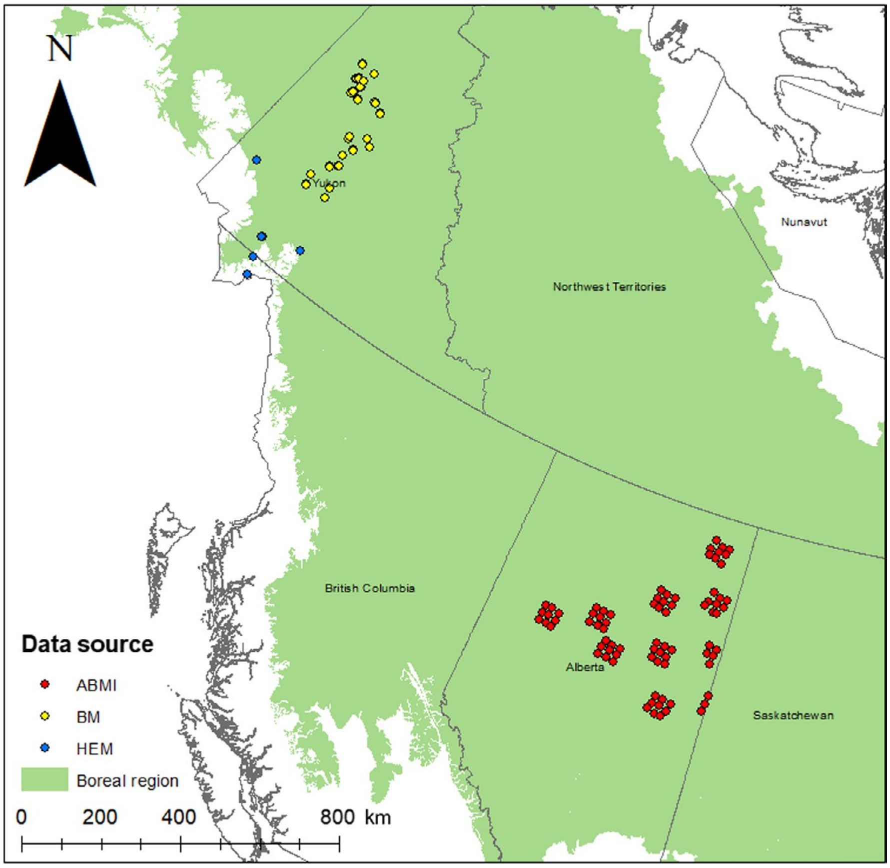

FIGURE 1 Location of acoustic surveys throughout the boreal forest biome of Northern and Western Canada. Red locations are acoustic surveys from the Alberta Biodiversity Monitoring Institute's (ABMI) provincial monitoring programme, yellow locations are acoustic surveys from the Yukon Boreal Monitoring programme (BM) and blue locations are acoustic surveys from the High Elevation Monitoring (HEM) programme. The boreal forest region is highlighted in green

TABLE 1 Project-specific data collection protocols for the Alberta Biodiversity Monitoring Institute (ABMI) provincial monitoring programme, the High Elevation Monitoring Programme (HEM) and the Yukon Boreal Monitoring Programme (BM)

Project	ARU	Audio settings	Number of sites	Deployment dates	Recording schedule	Total recordings
Alberta Biodiversity Monitoring Institute Biodiversity Monitoring Programme	Wildlife Acoustics SM2	Stereo WAV file recorded at 44.1 kHz and 16 bit PCM	80	15 March–25 August 2016	10-min recordings at 12:00 a.m., 2:00 a.m., 1 hr after sunrise, 1.5 hr after sunrise, 12:00 p.m., 3:00 p.m., 1 hr before sunset and 1 hr after sunset (8 recordings per day)	97,920
Canadian Wildlife Service and University of Alberta High Elevation Monitoring Programme	Frontier Labs BAR	Stereo WAV file recorded at 44.1 kHz and 16-bit PCM	11	25 June–7 August 2019	5-min recordings with 15-min intervals between 10:00 p.m. and 8:00 a.m. daily (4 recordings per hour, 40 recordings per day)	18,920
Canadian Wildlife Service Yukon Boreal Monitoring Programme	Wildlife Acoustics SM4	Stereo WAV file recorded at 44.1 kHz and 16-bit PCM	152	1 April–20 June 2018	5-min recordings with 15-min intervals between 10:00 p.m. and 8:00 a.m. daily (4 recordings per hour, 40 recordings per day)	486,400

schedules and different automated recording unit (ARU) models, which can vary in performance due to variables such as signal-to-noise ratio, sensitivity and recording settings (Yip et al., 2017). The ABMI programme deployed SM2 wildlife recorders and the BM programme deployed SM4 wildlife recorders, both from Wildlife Acoustics (Maynard, MA; https://www.wildlifeacoustics.com). The HEM programme deployed Bioacoustic Audio Recorders (BAR) from Frontier Labs (Brisbane, AU; https://frontierlabs.com.au/). While there is a considerable regional and geographical variation between datasets, all recordings were collected in deciduous, coniferous or mixedwood boreal forests. Data collection and recording protocols (i.e. types of recorder, duration of recordings) differed between datasets, but were standardized within projects. We randomly selected a subset of morning recordings from each dataset between 5:00 and 8:00 a.m. with the presence of vocalizing bird species from peak breeding season (June 1–15), along with recordings with little or no vocal activity from pre- and post-breeding periods at similar times so that a variety of weather conditions and non-target sounds would be present when training predictive models and classifiers. We standardized recordings to 3 min and down sampled to a common sampling and bit rate if necessary. We validated recording contents through manual scanning of spectrograms. A subset of SM2 data with the presence of vocalizing, breeding birds (n = 1,868 recordings) were transcribed manually, where human observers visually scanned and listened to audio recordings using standardized spectrogram parameters and sound levels to identify vocalizing individuals to the species level (The Bioacoustic Unit, 2019).

We chose to use summary indices in our analyses as opposed to several vectors of acoustic indices because multiple, scalar indices are easily incorporated into classification and clustering approaches. Previous literature using these analyses has used similar approaches although the indices used vary by study (Oliver et al., 2018; Phillips et al., 2018). We generated 17 acoustic summary indices commonly used to characterize the surrounding soundscape for each 1-min audio segment of all recorded surveys using Ecoacoustics Audio Analysis Software v18.03.0.41 and calculated using default settings (Table 2; Towsey et al., 2018). These were comprised of six indices generated from the signal waveform envelope: Average signal amplitude (dB), Background noise (BGN), Signal-to-noise ratio (SNR), Activity (ACT), Events per second (EVN) and Temporal entropy (ENT); and 11 indices generated from the distribution of spectral energy: Low-frequency cover (LFC), Mid-frequency cover (MFC), High-frequency cover (HFC), Entropy of spectral peaks (EPS), Entropy of the average spectrum (EAS), Entropy of the spectrum of coefficients of variation (ECV), Acoustic complexity index (ACI), Normalized difference soundscape index (NDSI), Cluster count (CLS), three-gram count (3GC) and Spectral peak density (SPD). Brief descriptions of the indices are provided in Table 2, with detailed explanations described in Towsey (2017). An audio segment of 1 min is recommended due to processing efficiency, standardization, recommended to be at the temporal scale of bird vocalizations and to avoid clipping or averaging out of acoustic features (Towsey et al., 2018).

TABLE 2 Descriptions of each of the summary indices included in analyses, calculated using Ecoacoustics Audio Analysis Software v18.03.0.41 with default settings (Towsey et al., 2018). Full descriptions of each acoustic index are described in Towsey (2017)

Acoustic index	Type	Description
Average signal amplitude (dB)	Waveform	Average energy value of the waveform over the duration of an audio segment
Background noise (BGN)	Waveform	The mode of the distribution of waveform energy values. This is roughly equivalent to the amount of acoustic energy that persists through the duration of an audio segment
Signal-to-noise ratio (SNR)	Waveform	Difference between the maximum decibel value and BGN in each frequency bin
Activity (ACT)	Waveform	Proportion of an audio segment that exceeds 3dB in each frequency bin
Events per second (EVN)	Waveform	Rate of acoustic events per second in each frequency bin
Temporal Entropy (ENT)	Waveform	Concentration of acoustic energy in each frequency bin
Low-frequency cover (LFC)	Frequency spectrum	Proportion of spectrogram cells that exceed 3 dB in the 1–1,000 Hz band
Mid-frequency cover (MFC)	Frequency spectrum	Proportion of spectrogram cells that exceed 3dB in the 1,000–8,000 Hz band
High-frequency cover (HFC)	Frequency spectrum	Proportion of spectrogram cells that exceed 3dB in the 8,000–10,982 Hz band
Entropy of spectral peaks (EPS)	Frequency spectrum	Degree of concentration of spectral energy in the spectral-maxima of the mid-frequency band (1,000–8,000 Hz)
Entropy of the average spectrum (EAS)	Frequency spectrum	Degree of concentration of spectral energy in the mean-energy spectrum of the mid-frequency band (1,000–8,000 Hz)
Entropy of the spectrum of coefficient of variation (ECV)	Frequency spectrum	Degree of concentration of spectral energy in the normalized energy-variance spectrum of the mid-frequency band (1,000–8,000 Hz)
Acoustic complexity index (ACI)	Frequency spectrum	Average relative change in acoustic energy in each frequency bin
Normalized difference soundscape index (NDSI)	Frequency spectrum	Ratio of acoustic energy in the 1000–2,000 Hz and 2,000–8,000 Hz frequency band. Commonly used to measure anthropogenic disturbance
Cluster count (CLS)	Frequency spectrum	Number of distinct spectral clusters in the mid-frequency band (1,000–8,000 Hz)
Three-gram count (3GC)	Frequency spectrum	Spectral clusters in the mid-frequency band (1,000–8,000 Hz) that occur more than once
Spectral peak density (SPD)	Frequency spectrum	Number of cells in the mid-frequency band (1,000–8,000 Hz) identified as being a local maxima

2.2 | Statistical analysis

We tested the accuracy of four different analytical approaches, *k*-means clustering, artificial neural networks and self-organizing maps (SOMs), boosted regression trees (BRTs), and a two-step procedure involving SOMs followed by hierarchical clustering, for identifying the presence of bird communities in audio recordings. These approaches are commonly used for similar types of classification and prediction objectives. For this study, breeding birds are defined as vocally active avian species or communities which arrive on the soundscape in the spring or early summer. We used a subset of data that included equal weighting of all three ARUs and datasets (BAR = 2,551 audio segments; SM2 = 3,000 audio segments; SM4 = 2,994 audio segments) and recordings with and without the presence of breeding birds (present = 4,270 audio segments; absent = 4,275 audio segments). First, we tried two supervised approaches where the response variable, the presence of avian vocal activity, is known and the predictors are the suite of acoustic summary indices described above. We predicted the probability of avian vocal activity being present using BRTs and SOMs. We then tested two unsupervised analytical methods (*k*-means and SOMs/hierarchical clustering), where the response variable (the presence of avian vocal activity) is unknown, to cluster recordings with and without vocal activity. For each approach, we split data into 70% training and 30% testing datasets to validate prediction accuracy and confusion matrices to validate model performance. Detailed descriptions of model building and testing for each analytical approach are found in Appendix S1.

For the clustering and SOM approaches, we dropped three acoustic indices (ENT, 3GC and SPD) to reduce multicollinearity in the dataset as suggested by Towsey (2017). This reduction in predictor variables allowed us to reduce multicollinearity while retaining as much information as possible. However, BRTs are robust to multicollinearity (Dormann et al., 2013) so we used the complete set of predictor indices for this approach. BRTs also accept categorical variables; therefore, we included the type of ARU used to record the survey as a predictor variable in addition to each acoustic index value. We reported the relative contribution of predictor indices from BRTs. BRTs were the most effective and accurate analytical approach and were used to investigate the following survey design and analysis parameters.

2.3 | Species richness

To investigate the effect of training models with variation in species composition and richness when classifying audio recordings, we selected recordings from the ABMI dataset (collected using SM2 recorders) containing varying levels of species richness and categorized them into five categories: no species, low richness (2–4 species), medium richness (5–9 species), high richness (≥10 species) and a combined category containing an entire gradient of species richness. ABMI recordings were selected because they had a larger range of diversity across recordings and we categorized and subsampled data to evenly distribute recordings across all levels of species richness because species richness was Poisson distributed and heavily skewed towards certain values. First, we tested the performance of BRTs to predict avian vocal activity at each level of species richness, when models were trained with examples of similar species richness. Second, we tested the performance of BRTs to predict avian vocal activity at different levels of species richness by training on one category of species richness and testing on a different category. Finally, we reported changes in the relative importance of each acoustic index for characterizing low, medium, high and combined richness datasets.

2.4 | Sample size

Collecting training data can be labour and time-intensive, and the effort and sample size required to maximize accuracy in predictive models for this purpose is unknown. We used bootstrapping to determine the minimum number of audio segments required to maximize model performance and achieve a reasonable trade-off between sample size and accuracy. We randomly selected sample sizes ($n = 1{,}000$) between 40 and 6,000 individual recording minutes using a uniform distribution and calculated BRT classification accuracy. We then used the R package ‘SEGMENTED’ (v.1.2; Muggeo, 2017) to perform breakpoint regression to determine the minimum number of 1-min audio segments required for optimal classification accuracy.

2.5 | Variation in monitoring programmes and recording equipment

To investigate the effect of variation between monitoring programmes, such as different recorder models and types, on classification and clustering accuracy, we subset data by recorder type and tested model accuracy with all three ARUs present, each ARU individually and pairs of ARUs in each possible combination. We calculated BRT classification accuracy for each set of data and compared and contrasted results. We also calculated classification accuracy when BRTs were trained on each individual monitoring programme and used to classify data from other monitoring programmes, to determine whether existing models can accurately classify novel datasets.

3 | RESULTS

3.1 | Effectiveness of different classification approaches

Boosted regression trees (BRTs) had the highest classification accuracy of the four analytical approaches tested, followed by supervised SOM (SSOM), unsupervised SOM and hierarchical clustering (USOM), and finally *k*-means clustering. The best performing BRT fitted a final model with 7,700 trees, a learning rate of 0.01, tree complexity of 5, bag fraction of 0.75 and cross validation AUC value of 0.980 ± 0.001 (mean $\pm$ *SE*). All 18 predictors (type of ARU and 17 different acoustic indices) had non-zero influence, although the relative influence of many indices was low (Figure 2). The predictors with the highest relative influence (>2.5%) were MFC (34.4%), NDSI (24.2%), EPS (7.1%), ARU type (6.1%), ACI (4.1%), ECV (3.1%), EAS (3.0%) and dB (2.7%). MFC and NDSI had strong positive influences on the probability of detecting vocal activity while other predictor variables had a more nuanced and complex relationship (Figure 3). Accuracy for predicting the presence of breeding birds on the testing dataset was 92.0%.

We built our self-organizing maps using a hexagonal grid, bubble neighbourhood function and 22×22 grid dimensions. Mean distance to the closest unit in the map for our SSOM was 0.031 and 0.904 for our USOM. The SSOM predicted the presence of avian vocal activity on the testing dataset with an accuracy of 87.4%. For the USOM, we calculated a Hopkin's statistic of 0.17 and three clusters as optimal after calculation of validity indices (suggested by 12/23 indices). Classification accuracy of the USOM for identifying recordings with and without avian vocal activity was 70.9%. For the *k*-means clustering approach, we calculated a Hopkin's statistic of 0.05 and an optimal cluster number of four (based on 8/23 validity indices). Classification accuracy of the *k*-means clustering approach was 66.7%. Confusion matrices for classification and clustering accuracy of all approaches are presented in Table 3.

3.2 | Variance in classification accuracy from survey design parameters

Since BRTs were the most effective method for classifying avian vocal activity, we chose to use this analytical approach when investigating effects of different survey design and data processing parameters that could influence model performance. Model performance varied depending on the species richness used for training and the testing dataset used for classification (Table 4). BRTs had high classification accuracy when trained with medium or high richness datasets, or the combined dataset (>90%), although classification accuracy was slightly lower when classifying low richness data in this scenario. Model performance was low when models were trained with low richness data with the exception of classifying a different test dataset with low richness. In general, classification

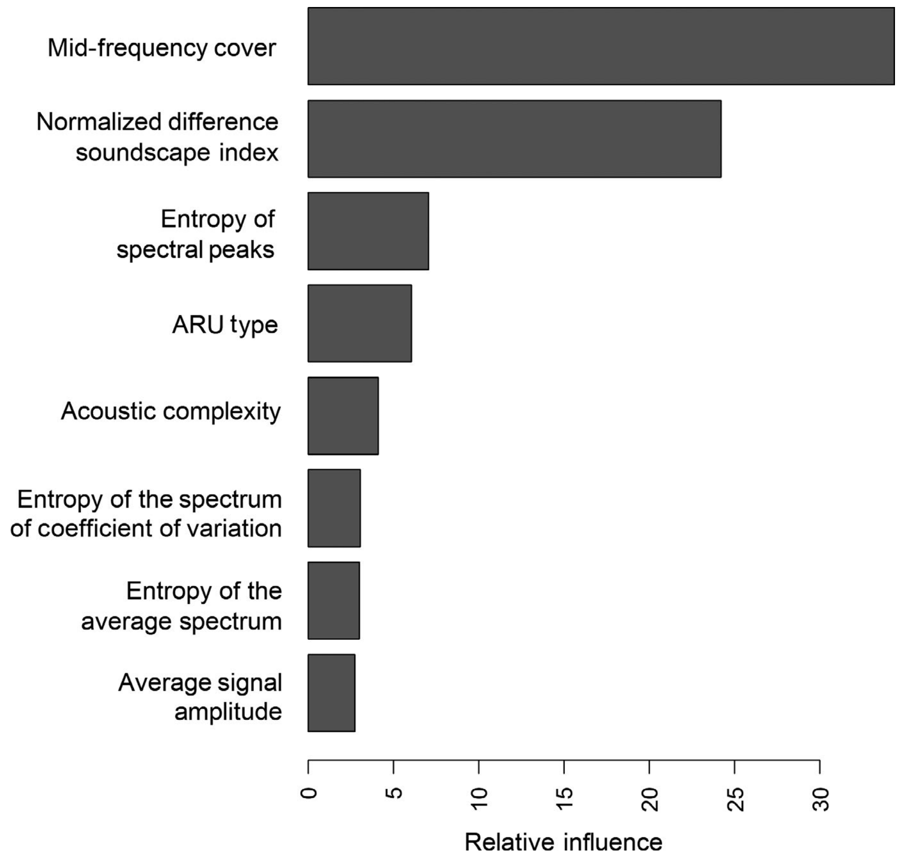

FIGURE 2 Relative influence of the eight most important variables for predicting the presence of birds in audio recordings

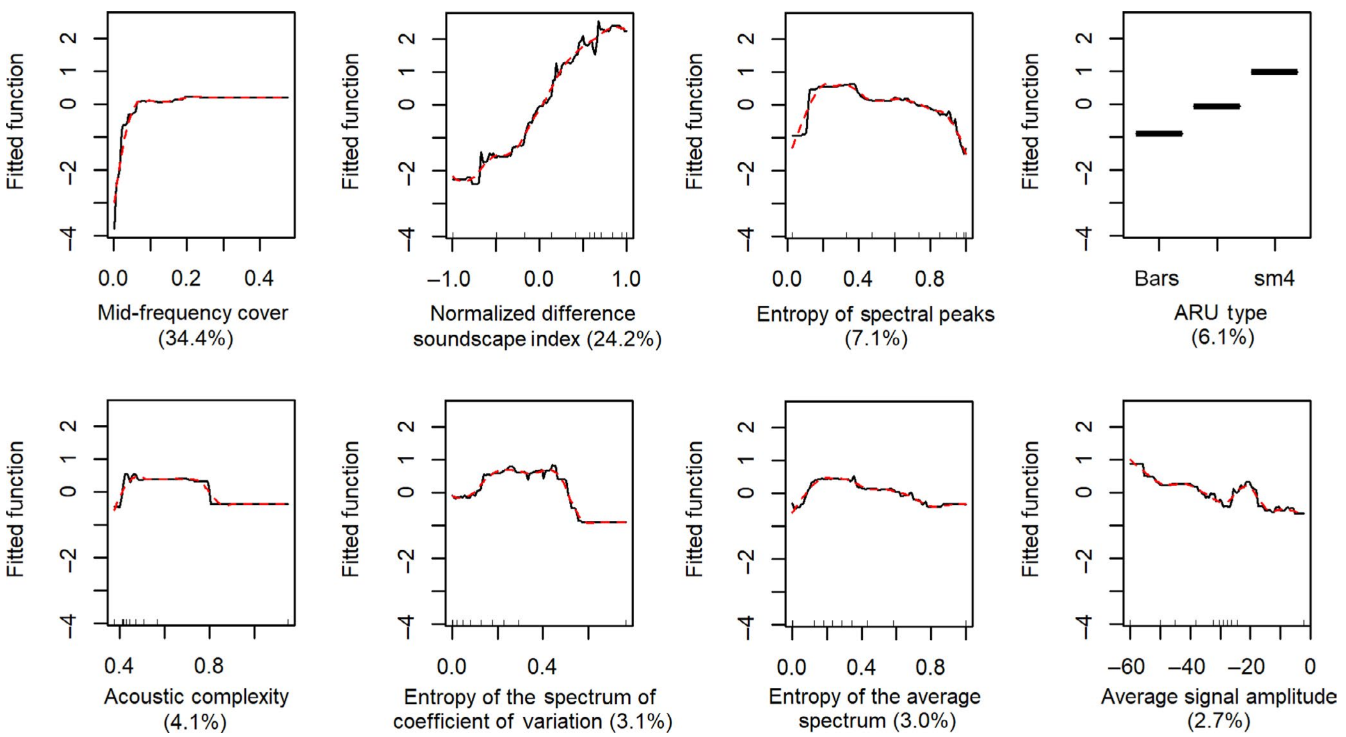

FIGURE 3 Predictor effects for the eight most important predictors in our best performing boosted regression tree model

accuracy was highest when models were trained and tested using data with similar species richness. The relative influence of each acoustic index on classifying vocal activity was similar between models of different species richness with two notable exceptions. The relative influence of EPS increased while MFC decreased, with increasing species richness.

The sample size of processed 1-min audio segments used to train classification models was positively related to the classification accuracy for identifying avian vocal activity (Figure 4). We set starting sample size values at 500 samples and our breakpoint regression model found one breakpoint at 797.83 ± 18.87 samples. Subsets of data from different ARUs resulted in different classification accuracy (Table 5). BARs had much higher accuracy relative to SM2s and SM4s when analysed individually using unsupervised clustering approaches and SM2s had lower accuracy when analysed individually using supervised classification approaches. The number of different ARUs included in the subset of data did not seem to influence accuracy, rather, classification accuracy of multiple ARU models appears to be the mean classification accuracy for those ARUs. BRTs classified avian vocal activity extremely well at >86% for all subsets of ARUs.

Classification accuracy decreased by 5%–20% when BRTs were trained on one individual monitoring programme and used to classify

TABLE 3 Confusion matrices for predicted and observed bird presence for *k*-means clustering (K-M), unsupervised self-organizing maps followed by clustering (USOM), supervised self-organizing maps (SSOM) and boosted regression trees (BRT). True positives (TP) occur when bird presence is correctly predicted, false positives (FP) occur when bird presence is predicted in audio segments without birds, true negatives (TN) occur when an absence of bird presence is correctly predicted and false negatives (FN) occur when an absence of bird presence is predicted in audio segments containing birds. Classification accuracy is calculated as the proportion of audio segments where the presence or absence of birds is correctly predicted: (TP + TN)/(TP + FP + TN + FN)

		Predicted bird presence		
		Present	Absent	Method
Observed bird presence	Present	2,328	1,942	K-M (Unsupervised)
	Absent	907	3,371	
	Present	3,817	461	USOM (Unsupervised)
	Absent	2,320	1,950	
	Present	1,116	160	SSOM (Supervised)
	Absent	156	1,131	
	Present	1,150	111	BRT (Supervised)
	Absent	84	1,220	

TABLE 5 Classification accuracy of boosted regression trees (BRT) when data are collected from multiple ARU types and from single ARUs

ARUs used for data collection	Accuracy
SM2 + SM4 + BAR	92.0
SM4 + BAR	96.3
SM2 + BAR	91.7
SM2 + SM4	90.1
BAR	98.0
SM4	94.4
SM2	86.4

TABLE 4 Classification accuracy of boosted regression trees (BRTs) when training and testing on datasets with varying levels of species richness. Models performed well when trained with medium or high richness data and poorly when trained with low richness data

		Testing data			
		Low richness	Medium richness	High richness	Combined
Training data	Low richness	89.2	78.8	76.1	81.7
	Medium richness	90.3	97.4	93.1	94.5
	High richness	91.4	95.9	98.1	95.5
	Combined	90.4	91.0	89.2	92.9

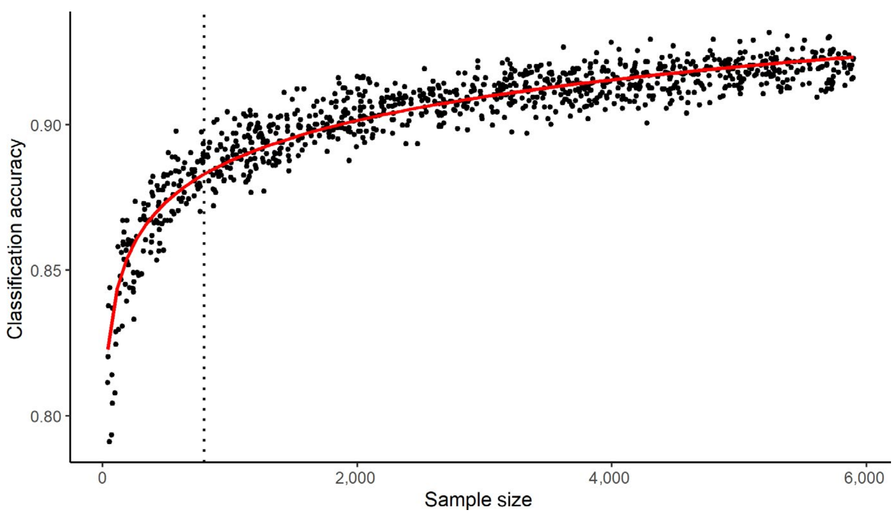

FIGURE 4 An increase in the number of audio segments used to train boosted regression trees is related to an increase in classification accuracy for identifying recordings with breeding birds. The dotted line represents the most efficient sample size for training datasets (n = 798) as indicated by breakpoint regression and the red line represents the line of best fit for the data

TABLE 6 Classification accuracy when boosted regression trees (BRTs) are trained on individual datasets from the Alberta Biodiversity Monitoring Institute's (ABMI) provincial monitoring programme, Yukon Boreal Monitoring program (BM) and High Elevation Monitoring (HEM) programme and used to classify a separate, novel dataset

	Testing data			
	Monitoring programme	ABMI	BM	HEL
Training data	ABMI	—	75.0	81.8
	BM	57.3	—	64.2
	HEL	79.6	72.3	—

data from a separate region. In particular, BRTs trained using data from the BM programme decreased by 15%–20% when used to classify audio from other regions. BRTs trained on the ABMI or HEM programmes decreased by 10%–15% when classifying BM data, and only 5% when BRTs exclusively trained using either of these two datasets were used to classify the other (Table 6).

4 | DISCUSSION

This is the first study to demonstrate that classification of audio recordings containing breeding bird activity using multiple acoustic indices can be done with regional datasets composed of data from a variety of environments and collected using different recording equipment, albeit with some careful considerations. We investigated the effect of analytical approach (unsupervised, supervised), recording equipment, audio processing parameters and training data on classification accuracy using acoustic indices and found that all of these factors require consideration to identify avian vocal activity. We found that supervised analyses significantly outperform unsupervised analyses by up to 40%; in particular, BRTs performed exceptionally well due to their ability to accept categorical predictors which can account for systematic or structural bias in the data. Our only categorical variable, the type of ARU that was used, was shown to contribute 6.1% of relative model influence in BRTs which is unaccounted for with the other approaches. Additional variation in sample environments, species composition and species richness may introduce too much unaccounted variation in the calculation of acoustic indices for unsupervised approaches to be viable in this scenario.

We found a sample size of approximately 800 one-minute audio segments (400 with birds present, 400 without) is sufficient to maximize classification efficiency using BRTs. While this approach does not result in the highest possible classification accuracy, validation of the contents of recordings can be time-consuming and labour-intensive. Our use of breakpoint regression identifies the number of samples required before diminishing returns on classification accuracy occurs, which still results in approximately 90% classification accuracy. We recommend practitioners have an evenly distributed response variable (target vocal activity vs. non-target recordings), a range of possible conditions for non-target recordings (e.g. weather events, non-target taxa, noise), and if additional resources are available, use as many training audio segments as feasible to maximize classification accuracy. Training data can be easily assembled when the target taxa is vocally active in predictable patterns. In our study, we randomly selected target recordings during peak breeding season and dawn chorus and non-target recordings over a range of times and dates where birds would not be vocalizing (i.e. late winter, late summer, non-dawn recordings) so that a variety of weather conditions and non-target sounds would be present. However, purposefully selecting and quantifying abiotic and non-target taxa when creating training data could improve model performance and classification accuracy by ensuring a balanced contribution of different conditions during model training. Acoustic indices could be used to screen audio for abiotic sounds to make this process more efficient (i.e. Metcalf et al., 2020). Finally, if combining data from more than one type of recording unit, we recommend accounting for this using additional predictor variables as systematic differences in technology such as signal-to-noise ratio can influence recording quality (Darras et al., 2020).

Of the eight most important acoustic indices (>2.5% relative influence), four indices were directly related to the intensity and pattern of sound in the mid-frequency band (1–8 kHz). MFC (34.4%), EPS (7.1%), ECV (3.1%) and EAS (3.0%) were all positively related to vocal activity in this frequency band, where bird song is most likely to be found. Additionally, NDSI (24.2%) was an important predictor which measures the ratio of sound in the mid- and low-frequency bands, and is often used to discriminate between biotic and abiotic sounds such as wind or precipitation. The type of ARU used (6.1%) was also an important predictor, but does not appear to be responsible for higher classification accuracy in BRTs relative to other approaches. If this were the case, we would likely see increased accuracy when models are trained on a single dataset containing one ARU due to a reduction in systematic variation in acoustic indices caused by multiple ARU models (Table 5). This could be due to differences in specifications like signal-to-noise ratio or microphone sensitivity compared to the two Wildlife Acoustic recording units, but unfortunately we were unable to separate these two variables which are built into each ARU, although recording quality was standardized among datasets by sampling rate and bit depth. We also tested the accuracy of training BRTs on one dataset and classification on a separate dataset to see if we could use training data from one region to predict the presence of birds in novel data. We found that classification accuracy dropped significantly in this scenario and we recommend that training data representative of what is being classified be used. Finally, since each dataset exclusively used a different ARU model, it is difficult to separate systematic effects of each monitoring programme from the effect of each ARU. For example, BAR units were used in high elevation monitoring, meaning that elevational differences in surrounding environment and weather conditions could influence classification performance. Further study into the individual effect of different ARU models would give additional insight into proper implementation of these classification approaches in the future.

Categorized species richness in both the data used to train BRTs and predict the presence of breeding birds was important for maximizing classification accuracy of the presence of birds on audio recordings. Accuracy was highest when datasets with similar species richness were used for model training and classification. However, models predicted reasonably well on all categories of community complexity when trained using medium richness (5–9 species), high richness (>10 species) and the entire gradient of richness of community complexity (>89%). Models trained with low richness (2–4 species) predicted well on other low richness data (89.2%) but poorly on all other categories (<81.7%). The relative importance of two individual acoustic indices also changed depending on richness, with MFC negatively related and EPS positively related to increased richness, although the reason for this pattern is unclear. We believe acoustic indices from low richness recordings are not distinct enough when compared to recordings with no vocal activity from breeding bird species to effectively classify recordings. Models trained using this data are useful for investigating seasonal patterns and phenology in breeding bird species; however, consideration should be taken when selecting training data depending on research objectives.

Although our BRT approach demonstrated reasonably high classification accuracy, there are several factors not tested in this study that could improve model performance. Species richness was the only measure of species occurrence that we incorporated into our analysis and previous literature indicates it is strongly related to acoustic indices assessing the biophony component of the soundscape (Bradfer-Lawrence et al., 2020; Eldridge et al., 2018; Towsey, Wimmer, et al., 2014). However, other metrics such as animal abundance may have strong associations with certain acoustic indices, particularly those derived from the waveform envelope, and should be explored further. Furthermore, we only investigated the use of BRTs for the survey design and species richness components of our analysis, due to BRTs outperforming other approaches by a large margin. While not as accurate, the other methods tested in this study may show different responses to variation in survey design and species richness and is an area of additional development that could improve this technique in the future. We also demonstrated that classification is most accurate when models are trained with recordings of similar species richness or species assemblage. However, soundscapes can differ between survey locations due to local differences in habitat, which can influence species assemblages and naturally occurring abiotic sound. The audio recordings used in this study were obtained from the boreal forest region of Northern Alberta, Northern British Columbia, and Southern and Central Yukon, Canada. Recordings used in this study were collected from deciduous, coniferous and mixedwood boreal forests. The boreal forest biome is highly heterogeneous containing ecologically diverse ecoregions differing in their geology, topography, vegetation and climate and these features are not included in the datasets we used. Training models with similar species assemblages and habitat types should yield higher classification accuracy. While we were unable to investigate the effect of habitat due to data limitations, classification accuracy was high nonetheless, and habitat classification at a finer scale than what is achieved in this study can be added as an additional predictor to increase model performance. Additionally, classification performance should be tested in other regions and with the presence of other taxonomic groups. Boreal forest soundscapes are characterized by distinct patterns in vocal activity caused by the arrival of breeding birds. Investigation into the utility of this tool when vocal activity is less prominent, as well as to distinguish the presence of birds from other taxonomic groups such as insects or amphibian choruses (Alvarez-Berrios et al., 2016; Campos-Cerqueira & Aide, 2017) should be done in the future to better understand the effectiveness of this approach. Seasonality or time of day could also be added to models to help differentiate between the onset of seasonal vocal activity in birds and other taxa such as amphibian choruses which peak at different times, although caution should be taken to avoid model circularity. Multiple models for different levels of species richness or community assemblages could also be used to investigate finer-scale patterns in seasonal phenology. Finally, while the software we used implements Sueur, Aubin, et al. (2008), Sueur, Pavoine, et al. (2008) formulation for calculating many indices, it is important to note that other formulations of the same indices (i.e. Karsten et al., 2012; Villanueva-Rivera & Pijanowski, 2018) could affect calculated values and model performance. Additionally, we used the recommended 1-min audio segment when calculating indices, but changes to audio segment length could influence the calculation of acoustic indices as well. Standardizing the method to calculate each index, as we have done in this study, is important for maximizing classification accuracy and the effect of other acoustic index formulations requires further investigation.

Parameters for data collection, processing and analysis using automated classification are often inconsistent, and there have been calls to investigate how variation in data collection and analysis can influence results so that standardization can occur. Furthermore, research into acoustic indices is generally a study- or project-specific (Browning et al., 2017; Gibb et al., 2018; Sueur, Pavoine, et al., 2008). In this study, we demonstrate accurate classification on a regional (northern and western boreal forests), heterogeneous dataset from a variety of sources and environments. As with any relatively new tool, guidance for the best practice during implementation can greatly increase utility, performance and accuracy. Thus, our recommendations are as follows:

1. Use supervised approaches when classifying or discriminating a specific component of the soundscape. Models performed significantly better when the target outcome was identified beforehand and validation effort for supervised and unsupervised approaches are similar. Unsupervised approaches require validation of the contents of each audio recording to determine

which category that cluster belongs to and supervised approaches require similar effort to validate the accuracy of classified audio recordings.

2. Use at least 800 training audio segments to build predictive models. Adding additional training data beyond that will improve classification accuracy but a minimum of 800 audio segments will maximize classification efficiency.
3. When building training datasets, ensure audio recordings are representative of the target soundscape (i.e. species richness, region and habitat) and potential non-target soundscape components (i.e. weather events, non-target species).
4. Make sure to include potentially confounding variables (in this study, we included the type of ARU and dataset). BRTs in particular can accept categorical variables.
5. Recommended best practices are to use an ensemble approach by incorporating multiple acoustic indices into models to compensate for weaknesses and downsides in individual indices. Our results suggest that multiple indices play an important role when predicting the presence of breeding birds on audio recordings.

The use of acoustic indices for automated identification of taxa or acoustic events of interest has tremendous potential. This approach has successfully been used to characterize daily diel patterns (Bradfer-Lawrence et al., 2019; Burivalova et al., 2017; Phillips et al., 2018), seasonal phenology (Buxton et al., 2016; Phillips et al., 2018) and vocal activity patterns (Bradfer-Lawrence et al., 2020; Oliver et al., 2018) in birds and other acoustically active species. Acoustic data can be screened to identify biotic activity from geophony or anthrophony prior to manual processing to increase efficiency (Metcalf et al., 2020; Sanchez-Giraldo et al., 2020). Automated classification could also be used to investigate patterns in seasonal phenology at larger scales, using pre-existing ARU datasets or by integrating multiple datasets as we have done here. For example, this method could be used to detect spatial or temporal shifts in habitat and intensity of use patterns associated with habitat quality in changing climate conditions, or continental scale patterns such as the detection of range shifts, extinctions or irruptions associated with changes to climate and weather (Buxton et al., 2016). Many monitoring programmes store large repositories of acoustic data; however, the biggest hurdle is effectively and efficiently interpreting meaningful relationships between the distribution of animals and the environment. Future applications should assess the performance of this approach relative to validated data (i.e. manual processing) to determine how closely classification accuracy of the presence of breeding birds from this approach matches ground-truthed data under a wide range of scenarios. The methods described in this paper present the opportunity to efficiently process large volumes of heterogeneous data to answer questions at scales previously difficult to investigate. Our investigation into the effect of several variables that commonly change within and between studies can help inform decisions in study design and analysis, as well as recommend study parameters that maximize the performance and accuracy of the use of acoustic indices for unsupervised identification of wildlife.

ACKNOWLEDGEMENTS

We would like to thank Logan McLeod, Jean-Francoice Jette, Lena Ware and Zack Davis for assistance with data collection and processing for the BM and HEM projects. We would also like to acknowledge the countless technicians who assisted with collecting and transcribing ABMI's 2016 data. Funding for the BM and EGM projects was provided by Environment and Climate Change Canada.

AUTHORS' CONTRIBUTIONS

D.A.Y., A.G.M. and C.L.M. conceived the ideas and designed methodology; A.G.M. and C.L.M. helped collect and process data; D.A.Y. analysed the data and led writing of the manuscript; A.G.M., C.L.M. and E.M.B. provided feedback and review of multiple drafts of the manuscript. All authors contributed critically to the drafts and gave final approval for publication. The authors have no conflict of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13548.

DATA AVAILABILITY STATEMENT

Data deposited in the Dryad Digital Repository https://doi.org/10.5061/dryad.brv15dv8b (Yip et al., 2020).

ORCID

Daniel A. Yip https://orcid.org/0000-0002-3130-0001

REFERENCES

Alvarez-Berrios, N., Campos-Cerqueira, M., Hernandez-Serna, A., Delgado, C. J. A., Roman-Danobeytia, F., & Aide, T. M. (2016). Impacts of small-scale gold mining on birds and anurans near the Tambopata Nature Reserve, Peru, assessed using passive acoustic monitoring. *Tropical Conservation Science*, *9*, 832–851.

Bellisario, K. M., Broadhead, T., Savage, D., Zhao, Z., Omrani, H., Zhang, S., Springer, J., & Pijanowski, B. C. (2019). Contributions of MIR to soundscape ecology. Part 3: Tagging and classifying audio features using a multi-labelling k-nearest neighbour approach. *Ecological Informatics*, *51*, 103–111.

Bixler, R. P., Johnson, S., Emerson, K., Nabatchi, T., Reuling, M., Curtin, C., Romolini, M., & Grove, J. M. (2016). Networks and landscapes: A framework for setting goals and evaluating performance at the large landscape scale. *Frontiers in Ecology and the Environment*, *14*, 145–153. https://doi.org/10.1002/fee.1250

Bradfer-Lawrence, T., Bunnefeld, N., Gardner, N., Willis, S. G., & Dent, D. H. (2020). Rapid assessment of avian species richness and abundance using acoustic indices. *Ecological Indicators*, *115*, 106400. https://doi.org/10.1016/j.ecolind.2020.106400

Bradfer-Lawrence, T., Gardner, N., Bunnefeld, L., Bunnefeld, N., Willis, S. G., & Dent, D. H. (2019). Guidelines for the use of acoustic indices in environmental research. *Methods in Ecology and Evolution*, *10*, 1796–1807. https://doi.org/10.1111/2041-210X.13254

Browning, E., Gibb, R., Glover-Kapfer, P., & Jones, K. E. (2017). Passive acoustic monitoring in ecology and conservation. *WWF Conservation Technology Series*, *1*(2). WWF-UK, .

Burivalova, Z., Towsey, M., Boucher, T., Truskinger, A., Apelis, C., Roe, P., & Game, E. T. (2017). Using soundscapes to detect variable degrees of human influence on tropical forests in Papua New Guinea. *Conservation Biology*, *32*, 205–215. https://doi.org/10.1111/cobi.12968

Burton, A. C., Huggard, D., Bayne, E. M., Schieck, J., Solymos, P., Muhly, T., Farr, D., & Boutin, S. (2014). A framework for adaptive monitoring of the cumulative effects of human footprint on biodiversity. *Environmental Monitoring and Assessment*, *186*, 3605–3617. https://doi.org/10.1007/s10661-014-3643-7

Buxton, R. T., Brown, E., Sharman, L., Gabriele, C. M., & McKenna, M. F. (2016). Using bioacoustics to examine shifts in songbird phenology. *Ecology and Evolution*, *6*, 4697–4710. https://doi.org/10.1002/ece3.2242

Buxton, R. T., McKenna, M. F., Clapp, M., Meyer, E., Stabenau, E., Angeloni, L. M., Crooks, K., & Wittenmyer, G. (2018). Efficacy of extracting indices from large-scale acoustic recordings to monitor biodiversity. *Conservation Methods*, *32*, 1174–1184. https://doi.org/10.1111/cobi.13119

Campos-Cerqueira, M., & Aide, T. M. (2017). Changes in the acoustic structure and composition along a tropical elevational gradient. *Journal of Ecoacoustics*, *1*, 4. https://doi.org/10.22261/JEA.PNCO7I

Darras, K. F. A., Deppe, F., Fabian, Y., Kartono, A. P., Angulo, A., Kolbrek, B., Mulyani, Y. A., & Prawiradilaga, D. M. (2020). High microphone signal-to-noise ratio enhances acoustic sampling of wildlife. *PeerJ*, *8*, e9955. https://doi.org/10.7717/peerj.9955

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osbourne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Eldridge, A., Guyot, P., Moscoso, P., Johnston, A., Eyre-Walker, Y., & Peck, M. (2018). Sounding out ecoacoustic metrics: Avian species richness is predicted by acoustic indices in temperate but not tropical habitats. *Ecological Indicators*, *95*, 939–952. https://doi.org/10.1016/j.ecolind.2018.06.012

Ellis, D. P. W., Zeng, X., & McDermott, J. H. (2011). Classifying soundtracks with audio texture features. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Prague, pp. 5880–5883.

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

Farina, A., Pieretti, N., & Piccioli, L. (2011). The soundscape methodology for long-term bird monitoring: A Mediterranean Europe case-study. *Ecological Informatics*, *6*, 354–363. https://doi.org/10.1016/j.ecoinf.2011.07.004

Fletcher, N. H. (2007). Animal bioacoustics. In T. D. Rossing (Ed.), *Springer handbook of acoustics*. Springer.AAA

Gasc, A., Sueur, J., Jiguet, F., Devictor, V., Grandcolas, P., Burrow, C., Depraetere, M., & Pavoine, S. (2013). Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetics and functional diversities of bird communities? *Ecological Indicators*, *25*, 279–287.

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2018). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, *10*, 169–185. https://doi.org/10.1111/2041-210X.13101

Harris, S. A., Shears, N. T., & Radford, C. A. (2016). Ecoacoustic indices as proxies for biodiversity on temperate reefs. *Methods in Ecology and Evolution*, *7*, 713–724. https://doi.org/10.1111/2041-210X.12527

Karsten, E. P., Gage, S. H., Fox, J., & Joo, W. (2012). The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology. *Ecological Informatics*, *12*, 50–67. https://doi.org/10.1016/j.ecoinf.2012.08.001

Lellouch, L., Pavoine, S., Jiguet, F., Glotin, H., & Sueur, J. (2014). Monitoring temporal change of bird communities with dissimilarity acoustic indices. *Methods in Ecology and Evolution*, *5*, 495–505. https://doi.org/10.1111/2041-210X.12178

Metcalf, O. C., Lees, A. C., Barlow, J., Marsden, S. J., & Devenish, C. (2020). hardRain: An R package for quick, automated rainfall detection in ecoacoustic datasets using a threshold-based approach. *Ecological Indicators*, *109*, 105793.

Muggeo, V. M. R. (2017). Interval estimation for the breakpoint in segmented regression: A smooth score-based approach. *Australian & New Zealand Journal of Statistics*, *59*, 311–322.

Oliver, R. Y., Ellis, D. P. W., Chmura, H. E., Krause, J. S., Perez, J. H., Sweet, S. K., Gough, L., Wingfield, J. C., & Boelman, N. T. (2018). Eavesdropping on the arctic: Automated bioacoustics reveal dynamics in songbird breeding phenology. *Science Advances*, *4*, eaaq1084. https://doi.org/10.1126/sciadv.aaq1084

Phillips, Y. F., Towsey, M., & Roe, P. (2018). Revealing the ecological content of long-duration audio-recordings of the environment through clustering and visualization. *PLoS ONE*, *13*, e0193345.

Pijanowski, B. C., Villanueva-Rivera, L. J., Dumyahn, S. L., Farina, A., Krause, B. L., Napoletano, B. M., Gage, S. H., & Pieretti, N. (2011). Soundscape ecology: The science of sound in the landscape. *BioScience*, *61*, 203–216. https://doi.org/10.1525/bio.2011.61.3.6

Rychtáriková, M., & Vermeir, G. (2013). Soundscape categorization on the basis of objective acoustical parameters. *Applied Acoustics*, *74*, 240–247. https://doi.org/10.1016/j.apacoust.2011.01.004

Sanchez-Giraldo, C., Bedoya, C. L., Moran-Vasquez, R. A., Isaza, C. V., & Daza, J. M. (2020). Ecoacoustics in the rain: Understanding acoustic indices under the most common geophonic source in tropical rainforests. *Remote Sensing in Ecology and Conservation*, *6*, 248–261. https://doi.org/10.1002/rse2.162

Sankupellay, M., Towsey, M., Truskinger, A., & Roe, P. (2015). Visual fingerprints of the acoustic environment: The use of acoustic indices to characterise natural habitats. *IEEE International Symposium on Big Data Visual Analytics (BDVA)* (pp. 111–118). Hobart, Australia.

Soulé, M. E., & Terborgh, J. (1999). Conserving nature at regional and continental scales – A scientific program for North America. *BioScience*, *49*, 809–817. https://doi.org/10.2307/1313572

Sueur, J., Aubin, T., & Simonis, C. (2008). seewave: A free modular tool for sound analysis and synthesis. *Bioacoustics*, *18*, 213–226. https://doi.org/10.1080/09524622.2008.9753600

Sueur, J., & Farina, A. (2015). Ecoacoustics: The ecological investigation and interpretation of environmental sound. *Biosemiotics*, *8*, 493–502. https://doi.org/10.1007/s12304-015-9248-x

Sueur, J., Pavoine, S., Hamerlynck, O., & Duvail, S. (2008). Rapid acoustic survey for biodiversity appraisal. *PLoS ONE*, *3*, e4065. https://doi.org/10.1371/journal.pone.0004065

Sugai, L. S. M., Desjonqueres, C., Silva, T. S. F., & Llusia, D. (2019). A roadmap for survey designs in terrestrial acoustic monitoring. *Remote Sensing in Ecology and Conservation*, *6*, 220–235. https://doi.org/10.1002/rse2.131

Sugai, L. S. M., Silva, T. S. F., Jr Ribeiro, J. W., & Llusia, D. (2019). Terrestrial passive acoustic monitoring: Reviews and perspectives. *BioScience*, *69*, 15.

The Bioacoustic Unit. (2019). *WildTrax acoustic transcription user guide. Version 2.0*. University of Alberta and Alberta Biodiversity Monitoring Institute.

Towsey, M. W. (2017). *The calculation of acoustic indices derived from long-duration recordings of the natural environment*. QUT ePrints. Retrieved from https://eprints.qut.edu.au/110634/

Towsey, M. W., Truskinger, A., & Roe, P. (2018). *Audio analysis software (version 18.03.0.41)*. QUT Ecoacoustics Research Group.

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

Towsey, M. W., Zhang, L., Cottman-Fields, M., Wimmer, J., Zhang, J., & Roe, P. (2014). Visualization of long-duration acoustic recordings of the environment. *Procedia Computer Science*, *29*, 703–712. https://doi.org/10.1016/j.procs.2014.05.063

Van Wilgenburg, S. L., Mahon, C. L., Campbell, G., McLeod, L., Campbell, M., Evans, D., Easton, W., Francis, C. M., Haché, S., Machtans, C. S., Mader, C., Pankratz, R. F., Russell, R., Smith, A. C., Thomas, P., Toms, J. D., & Tremblay, J. A. (2020). A cost efficient spatially balanced hierarchical sampling design for monitoring boreal birds incorporating access costs and habitat stratification. *PLoS ONE*, *15*, e0234494. https://doi.org/10.1371/journal.pone.0234494

Villanueva-Rivera, L. J., & Pijanowski, B. C. (2018). *soundecology: Soundscape ecology*. R package version 1.3.3. Retrieved from https://cran.r-project.org/package=soundecology

Yip, D. A., Leston, L., Bayne, E. M., Solymos, P., & Grover, A. (2017). Experimentally derived detection distances from audio recordings and human observers enable integrated analysis of point count data. *Avian Conservation and Ecology*, *12*, 11. https://doi.org/10.5751/ACE-00997-120111

Yip, D. A., Mahon, C. L., MacPhail, A. G., & Bayne, E. M. (2020). Data from: Automated classification of avian vocal activity using acoustic indices in regional and heterogeneous datasets. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.brv15dv8b

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Yip DA, Mahon CL, MacPhail AG, Bayne EM. Automated classification of avian vocal activity using acoustic indices in regional and heterogeneous datasets. *Methods Ecol Evol*. 2021;12:707–719. https://doi.org/10.1111/2041-210X.13548