

How index selection, compression, and recording schedule impact the description of ecological soundscapes

Becky E. Heath^{1,2}  | Sarab S. Sethi^{1,2,3}  | C. David L. Orme²  | Robert M. Ewers² | Lorenzo Picinali¹ 

¹Dyson School of Design Engineering, Imperial College London, London, UK

²Department of Life Sciences, Imperial College London, London, UK

³Department of Mathematics, Imperial College London, London, UK

Correspondence

Becky E. Heath, Dyson School of Design Engineering, Imperial College London, Exhibition Road, South Kensington, London, UK.

Email: r.heath18@imperial.ac.uk

Funding information

Natural Environment Research Council, Grant/Award Number: NE/P012345/1

Abstract

1. Acoustic indices derived from environmental soundscape recordings are being used to monitor ecosystem health and vocal animal biodiversity. Soundscape data can quickly become very expensive and difficult to manage, so data compression or temporal down-sampling are sometimes employed to reduce data storage and transmission costs. These parameters vary widely between experiments, with the consequences of this variation remaining mostly unknown.
2. We analyse field recordings from North-Eastern Borneo across a gradient of historical land use. We quantify the impact of experimental parameters (MP3 compression, recording length and temporal subsetting) on soundscape descriptors (Analytical Indices and a convolutional neural net derived AudioSet Fingerprint). Both descriptor types were tested for their robustness to parameter alteration and their usability in a soundscape classification task.
3. We find that compression and recording length both drive considerable variation in calculated index values. However, we find that the effects of this variation and temporal subsetting on the performance of classification models is minor: performance is much more strongly determined by acoustic index choice, with Audioset fingerprinting offering substantially greater (12%–16%) levels of classifier accuracy, precision and recall.
4. We advise using the AudioSet Fingerprint in soundscape analysis, finding superior and consistent performance even on small pools of data. If data storage is a bottleneck to a study, we recommend Variable Bit Rate encoded compression (quality = 0) to reduce file size to 23% file size without affecting most Analytical Index values. The AudioSet Fingerprint can be compressed further to a Constant Bit Rate encoding of 64 kb/s (8% file size) without any detectable effect. These recommendations allow the efficient use of restricted data storage whilst permitting comparability of results between different studies.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Animal vocalizations come together with abiotic and human-made sounds to form soundscapes. These soundscapes can be recorded and quantified across large temporal and spatial dimensions to monitor species populations or infer community-level metrics such as biodiversity (Eldridge et al., 2018; Gómez et al., 2018; Roca & Proulx, 2016). Monitoring is crucial to effectively respond to threats such as disease, species loss, and overlogging (Rapport, 1989; Rapport et al., 1998). Previously, the use of in situ expert listeners to monitor species presence and abundance was common (Huff et al., 2000) but is costly and time-consuming; can damage habitats; and is prone to narrow focus and observer bias (Costello et al., 2016; Fitzpatrick et al., 2009). Advances in portable computing now permit remote recording of soundscapes, but produce a volume of data that is very time-consuming to review manually, leading to the development of automated, or semiautomated, methods of analysis (Sethi, Jones, et al., 2020; Towsey et al., 2016).

Soundscape composition is primarily assessed using acoustic indices which describe the soundscape in an abstracted form. Analytical Indices are a type of acoustic index which are summary statistics that describe the distribution of acoustic energy within the recording (Towsey et al., 2014)—over 60 of which have been designed to capture aspects of biodiversity (Buxton et al., 2018; Sueur et al., 2014). These are commonly used in combination to compare the occupancy of acoustic niches, temporal variation, and the general level of acoustic activity (Bradfer-Lawrence et al., 2019) across ecological gradients or in classification tasks (Gómez et al., 2018). These approaches have provided novel insight into ecosystems across the world (Buxton et al., 2018; Eldridge et al., 2018; Fuller et al., 2015; Sueur et al., 2019) but are not foolproof and often have poor transferability (Bohnenstiehl et al., 2018; Mammides et al., 2017). This may result from a lack of standardization: differing index selection, data storage methods, and recording protocols, which all lead to unassessed variation in experimental outputs (Araya-Salas et al., 2019; Bradfer-Lawrence et al., 2019; Sugai et al., 2019).

The output vector from the AudioSet convolutional neural net (CNN; Gemmeke et al., 2017; Hershey et al., 2017) is an attractive replacement for Analytical Indices. This pretrained, general-purpose audio classification algorithm generates a multidimensional acoustic fingerprint of a soundscape which can be used as a more effective suite of acoustic indices (Sethi, Jones, et al., 2020). The AudioSet CNN is trained on two million human-labeled anthropogenic and environmental audio samples, potentially giving it both greater transferability and discrimination than typical ecoacoustic training datasets. Unlike Analytical Indices, however, extra analysis (such as training classifiers/predictive models) is necessary to relate the AudioSet Fingerprint to ecological processes and states.

In ecoacoustics, a continuous uncompressed or lossless recording is generally recommended (Browning et al., 2017; Villanueva-Rivera et al., 2011), but generates huge files. We considered two commonly used approaches to reducing storage requirements

(Towsey, 2018). Firstly, MP3 compression, which is widely used in ecoacoustic studies (e.g., Saito et al., 2015; Sethi, Jones, et al., 2018; Zhang et al., 2016): This lossy encoding removes acoustic information inaudible to *human* listeners (Sterne, 2012) but is suspected of removing ecologically important data (e.g., Sugai et al., 2019; Towsey et al., 2016). Araya-Salas et al. (2019) have recently shown that ecological information is lost under high compression from recordings of isolated animal calls; however, it is not known if this extends to recordings of noisier whole soundscapes.

Secondly, recording schedules also vary in ecoacoustic studies (Sugai et al., 2019). Bradfer-Lawrence et al. (2019) showed that longer and more continuous schedules give more stable Analytical Index values. However, ecoacoustic composition varies with time of day (Bradfer-Lawrence et al., 2019; Fuller et al., 2015; Sethi, Jones, et al., 2020) and so reducing recording periods with temporal subsetting may reduce temporal variation and improve classification (Sugai et al., 2019) even with reduced data. Similarly, index calculation on longer recordings may average away anomalous calls and short-term patterns.

While clear standards are crucial for collaborative research in ecoacoustics, there is uncertainty in the literature on the impacts of the selection of index type, compression level, and recording schedule on the quantification and classification of ecological soundscapes. Here, we:

1. investigated the impact of index selection on the accuracy of a random forest classifier;
2. described the effects of compression, recording length, and temporal subsetting on the values, variance, and classification performance of indices.

In describing how well ecological information is stored in acoustic data under different recording decisions, we identified stronger standards to improve classifier accuracy, precision, and recall and provided a basis for comparison among studies.

2 | METHODS AND MATERIALS

2.1 | Study area

Acoustic samples were collected in Sabah, Malaysian Borneo, at the Stability of Altered Forest Ecosystems (SAFE) project: a large-scale ecological experiment on habitat loss and fragmentation effects on tropical forests (Ewers et al., 2011) which included sites in the Kalabakan Forest Reserve (KFR). Historically, logging within KFR has been heterogeneous, reflecting habitat modifications in the wider area (Struebig et al., 2013), with higher than typical timber extraction rates. This is a diverse forest type from which we have recorded at least 175 species of bird and at least 50 species of amphibian from 26 sites (Sethi, Ewers, et al., 2020). Habitat ranges from areas of grass and low shrub, through logged forest to almost undisturbed primary forest.

2.2 | Soundscape recording

Data were collected from three KFR sites representing a gradient in aboveground biomass (Figure 1a; AGB: Pfeifer et al., 2016): primary forest (AGB = 66.16 t.ha⁻¹), logged forest (AGB = 30.74 t.ha⁻¹), and cleared forest (AGB = 17.37 t.ha⁻¹) (Appendix S1: Supplementary 1). We recorded continuously from a single recorder for a mean of 72 hr at each site (range: 70 to 75) during February and March 2019 (Appendix S1: Supplementary 2a). No rain fell during the recording period, so no recordings were excluded due to confounding geophony (Zhang et al., 2016). In all three sites, we placed individual omnidirectional (Hill et al., 2018) recorders, which were attached to trees (~50 cm diameter and 1–2 m above the ground) and recorded 20-min samples with no break period and stored them as uncompressed files (“raw,” .wav format) at 44.1kHz and 16 bits.

2.3 | Compressing and resizing the raw audio

Continuous 20-min recordings were first split into recordings with a length of 2.5, 5.0, and 10.0 min, using the python package *pydub* (Robert & Webbie, 2018; Figure 1b) resulting in 8, 4, and 2 times as many recordings, respectively. The audio was then converted to lossy MP3 format using the *fre:ac* LAME encoder (Kausch, 2019) under two standard LAME MP3 encoding techniques: constant bit rate (CBR) and variable bit rate (VBR) compressions (Figure 1c). CBR reduces the file size to a specified number of kilobits per second;

VBR varies bitrate per second depending on the analysis of the acoustic content and a quality setting (0, highest quality, larger bitrate; 9 lowest quality, smaller bitrate). Since bitrates are not directly comparable between VBR and CBR—and because storage savings are often the principal driver of compression choices—we used compressed file size as our measure of compression level. We used VBR0 and CBR320, CBR256, CBR128, CBR64, CBR32, CBR16, and CBR8, which resulted in file sizes ranging between 41.6% (CBR320) and 1.04% (CBR8) of the original raw file size and some reductions in Nyquist frequency (Table 1). We do not consider lossless compression, as the storage capacity is much higher and the files are obligatorily the same postdecompression. Previous studies have also found that the lossless compressed audio is largely identical to raw audio (Linke & Deretic, 2020).

2.4 | Quantification of soundscapes using indices

2.4.1 | Analytical indices

We used the *seewave* (ver 2.1.6) (Sueur, Aubin, et al., 2008) and *soundecology* (ver 1.3.3) (Villanueva-Rivera & Pijanowski, 2016) packages in R (ver 3.6.1; R Core Team, 2020) to extract 7 Analytical Indices (Figure 4d): Acoustic Complexity Index (ACI, calculated per minute and averaged), Acoustic Diversity Index (ADI), Acoustic Evenness (AEve), Bioacoustic Index (Bio), Acoustic Entropy (H), Median of the Amplitude Envelope (M), and Normalised Difference Soundscape

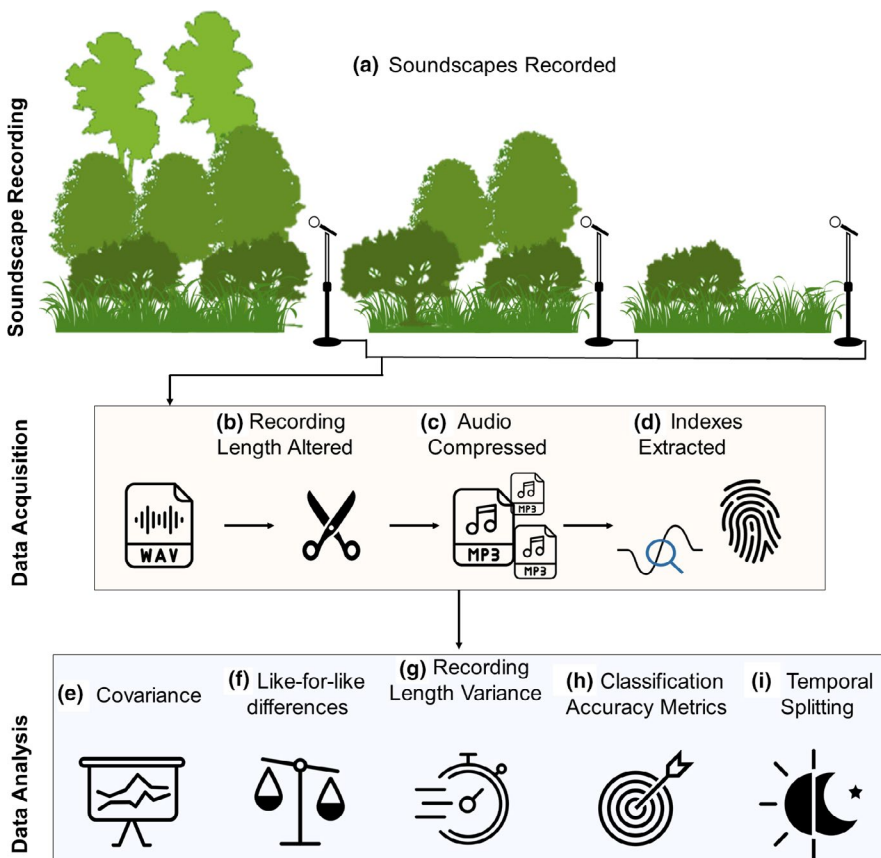


FIGURE 1 Experimental structure. Soundscape Recording: (a) Soundscapes from different forest structures in Malaysian Borneo are recorded. Data Acquisition: (b) Recording length is altered to 20-, 10-, 5-, and 2.5-min chunks; (c) all audio is compressed using nine lossy nine MP3 encoding techniques; (d) Analytical Indices and CNN Derived AudioSet Fingerprint are calculated from audio of all lengths and compressions. Data Analysis: (e) Index covariance is found per index type and correlation with maximum frequency is found; (f) like-for-like differences of indices calculated from compressed versus uncompressed counterparts are found; (g) intragroup variance compared for the recording lengths; (h) the indices of both types, lengths, and compressions are tested with a supervised random forest classification task; (i) the dataset is split into temporal sections and classification accuracy is found

TABLE 1 Bitrate, percentage file size reduction, and maximum encodable frequency for the experimental compression levels

Compression level	Bit storage/s	% File size	Nyquist frequency (kHz)
RAW	Constant: 768 kb	100	22.05
VBRO	Variable: ~ 127–250 kb	Mean = 20.82 Range = 32.64–16.63	22.05
CBR320	Constant: 320 kb	41.6	22.05
CBR256	Constant: 256 kb	33.35	22.05
CBR128	Constant: 128 kb	16.67	22.05
CBR64	Constant: 64 kb	8.33	22.05
CBR32	Constant: 32 kb	4.16	11.025
CBR16	Constant: 16 kb	2.08	8
CBR8	Constant: 8 kb	1.04	4

Index (NDSI; Appendix S1: Supplementary 3). These have been shown to capture diel phases, seasonality, and habitat type (Bradfer-Lawrence et al., 2019). These indices could not be calculated for all recordings due to file reading errors; however, this fault occurred in 0.3% of all recordings (Appendix S1: Supplementary 2b).

2.4.2 | AudioSet fingerprint

The audio was converted to a log-scaled Mel-frequency spectrogram after 16 kHz downsampling and then passed through the “VGG-ish” Convolutional Neural Network (CNN) trained on the AudioSet database (Gemmeke et al., 2017; Hershey et al., 2017; Figure 1d). This generated a 128-dimensional embedding and the 128 values in that embedding described the soundscape of given recording in an abstracted form or fingerprint. Similarly, as in the Analytical Indices, some recordings could not be analyzed by the AudioSet CNN; however, this was only in 0.2% of recordings (Appendix S1: Supplementary 2b).

2.5 | Data analysis

2.5.1 | Impact of index selection: auto-correlation

Analytical Indices often summarize similar features of a soundscape (e.g., dominant frequency and frequency bin occupancy): This overlap may reduce the descriptive scope of the ensemble. We compared the degree of pairwise correlation between the individual Analytical Indices and between the individual values of the AudioSet Fingerprint. We also compared how well each index/feature correlated with the Nyquist frequency (Figure 1e).

2.5.2 | Impact of compression: like-for-like differences

We used an adaption of Bland–Altman plots (Araya-Salas et al., 2019; Vesna, 2009) to visualize the scaled difference (D) between raw (I_{raw})

and compressed (I_{com}) index values, as a percentage of the range of raw values R_{raw} (Figure 1f):

$$D = \frac{I_{\text{com}} - I_{\text{raw}}}{R_{\text{raw}}} \times 100$$

D was not normally distributed (Appendix S1: Supplementary 5a), so median and interquartile ranges were reported. We determined that an index has been altered as a result of compression to be when: (a) the interquartile range of D did not include zero difference or (b) median D was more than $\pm 5\%$ of the R_{raw} . We used Spearman rank correlation to test for a consistent trend in D with increasing compression. To reflect their common use cases, D for Analytical Indices was calculated from the univariate values, while for AudioSet Fingerprints—which is intended as a multidimensional metric— D was calculated separately for each dimension and then given as a mean of all 128 values.

2.5.3 | Impact of recording schedule: recording length

Recordings of longer length may have a reduced variance due to the smoothing of potentially important transient audio anomalies (such as nearby bird or cicada calls). We tested this by comparing the variance of the recording groups at different commonly used recording lengths. The index values are non-normally distributed so we used Levene's test for homogeneity of variance (Figure 1g).

2.5.4 | Impact of parameter alteration on classification task

We used random forest classification models to assess how well the soundscapes were represented by each index type under each different experimental parameter, using the *RandomForest* (ver 4.6-14) (Liaw & Wiener, 2002) package in R (Figure 1h). Models were trained on a 24-hr period of data from each site and tested on the remaining 46 + h of audio. We used 2,000 decision trees to ensure accuracy had stabilized. The model was trained and

tested separately for every combination of index type (Analytical Indices vs. AudioSet Fingerprint), compression level, and recording length. We determined the accuracy, precision, and recall of each combination.

2.5.5 | Impact of temporal subsetting

Soundscapes typically show considerable diel variation in both abiotic and biotic components. To assess the impact of this variance on model performance, we split our recordings into four 6-hr sections centered on the key periods of Dawn (06:00), Noon (12:00), Dusk (18:00), and Midnight (00:00) and then further subdivided these into 3-hr (8 sections) and 2-hr (12 sections) blocks to test how further reductions affected the model (Figure 1i). We trained and tested the random forest model again on each of the temporally subset recordings, with each section used to build models individually, and determined accuracy, precision, and recall as before.

2.5.6 | Modeling the impact of index selection, compression, and recording length on the accuracy metrics

As the accuracy metrics are bound between 0% and 100%, we used a beta regression to model the relationship between each of the experimental parameters and performance metrics (Douma & Weedon, 2019). The model was built using the *betareg* (ver 3.1-3) package in R (Cribari-Neto & Zeileis, 2010). To avoid fitting issues when performance measures are exactly 1, we rescaled all performance measures using $m' = (m(n-1) + 0.5)/n$, where n is the sample size (Smithson & Verkuilen, 2006). The model included pairwise interactions between file size, temporal subsetting, and recording length, and then all interactions of main effects and those pairwise terms with the index selection. We observed that variance in performance measures varied as an interaction of both index choice and a temporal subsetting (Appendix S1: Supplementary 8a), so tested the inclusion of these terms in the precision component of the model. We first treated recording length and temporal subsetting as factors, but also tested a model considering these as continuous variables. We found the Akaike information criterion (AIC) was markedly lower in a beta regression model using factors and including the precision component (Appendix S1: Supplementary 8b).

3 | RESULTS

Although Spearman pairwise correlations of Analytical Indices and Nyquist frequency were low on average (mean = 0.32, IQR = 0.22), we found some strongly correlated sets of indices (Figure 2). ADI, Bio and NDSI all show strong similarities and were closely correlated with maximum recordable frequency; AEve and H were also strongly correlated (Figure 2). Some features of the AudioSet Fingerprint correlated with each other and maximum frequency, but in

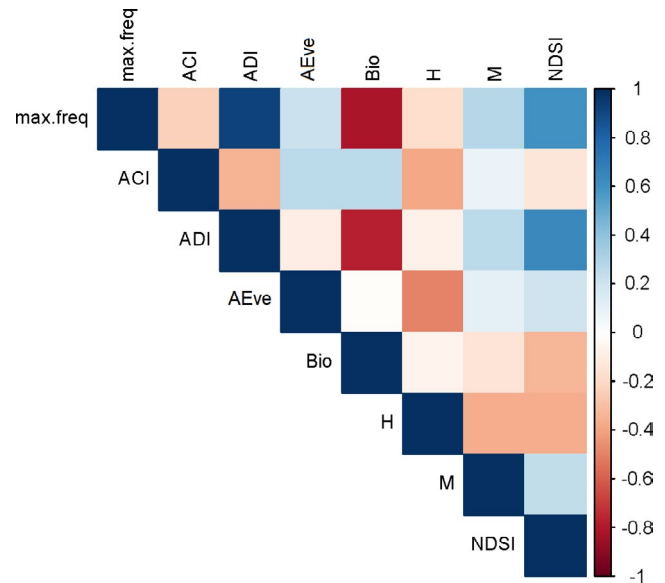


FIGURE 2 Pairwise Spearman correlation matrix for Analytical Indices (all recording lengths and all compressions) and maximum recordable frequency. The color scale shows rho values

general, these features were more weakly correlated (mean = 0.14, IQR = 0.18; figure in Appendix S1: Supplementary 4b).

3.1 | Impact of compression

3.1.1 | Impact of compression: like-for-like differences

Both index types showed both differences under compression and clear trends with increasing compression (Figure 3; confirmed with Spearman's rank correlation, all $p < .001$; Appendix S1: Supplementary 5b). The mode of response showed three broad qualitative patterns, illustrated here using results from the 5-min audio sample (other recording lengths in Appendix S1: Supplementary 5a). (a) Indices which were only affected above a threshold level of compression (AudioSet Fingerprint: CBR16; M: CBR32; and NDSI: CBR8). These indices typically showed low absolute D (median D typically $<15\%$). (b) AEve and H showed the largest differences at an intermediate compression (CBR64) and relatively low absolute differences (median D typically $<30\%$). (c) The remaining indices showed a variety of responses: ADI showed a monotonic response above a threshold, ACI showed changes up to CBR64 and then stabilizes, and Bio showed a stepped pattern of increase. However, all three showed increasing and large changes in absolute D (median D often $>75\%$) with increasing compression.

3.1.2 | Impact of recording schedule: recording length

Three out of seven (43%) of the Analytical Indices (ADI, AEve, and H) and a smaller proportion of the AudioSet Fingerprint

FIGURE 3 Scaled difference in acoustic indices from raw audio with increasing compression in 5-min audio samples (see Appendix S1: Supplementary 4 for 2.5- and 10- and 20-min examples). The horizontal green region shows the $\pm 5\%$ D . Dots and whiskers show the median and interquartile range of D from different indices under increasing levels of compression

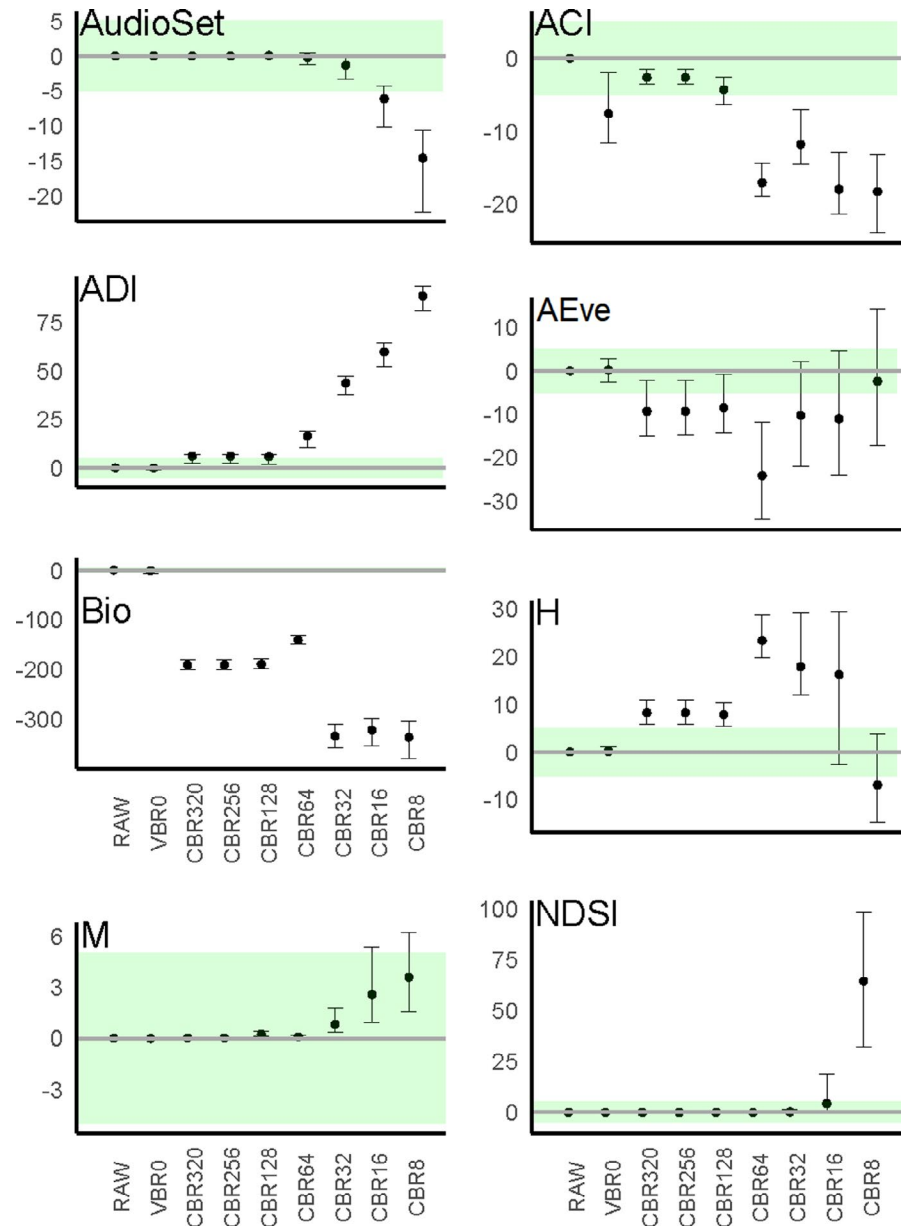


TABLE 2 Confusion matrices from random forest classifiers trained on AudioSet Fingerprint (a, c) and Analytical Indices (b, d) using uncompressed raw audio (a, b) and highly compressed CBR8 audio (c, d)

Observed	AudioSet Fingerprint			Observed	Analytical Indices		
	Predicted				Predicted		
(a) Raw	Cleared	Logged	Primary	(b) Raw	Cleared	Logged	Primary
Cleared	585	9	11	Cleared	484	67	49
Logged	11	508	44	Logged	97	421	46
Primary	17	14	521	Primary	9	61	486
(c) CBR8	Cleared	Logged	Primary	(d) CBR8	Cleared	Logged	Primary
Cleared	585	3	17	Cleared	484	23	98
Logged	2	488	73	Logged	9	379	175
Primary	11	53	488	Primary	9	115	428

values (46 out of 128; 36%) were found to have nonhomogeneous variance in groups of different recording lengths ($p < .05$, Levene's test for homogeneity of variance; Appendix S1: Supplementary 6b).

3.2 | Impact of index selection

Confirming prior findings (Sethi, Jones, et al., 2020), we showed that habitat classifiers derived from 5-min recordings using raw

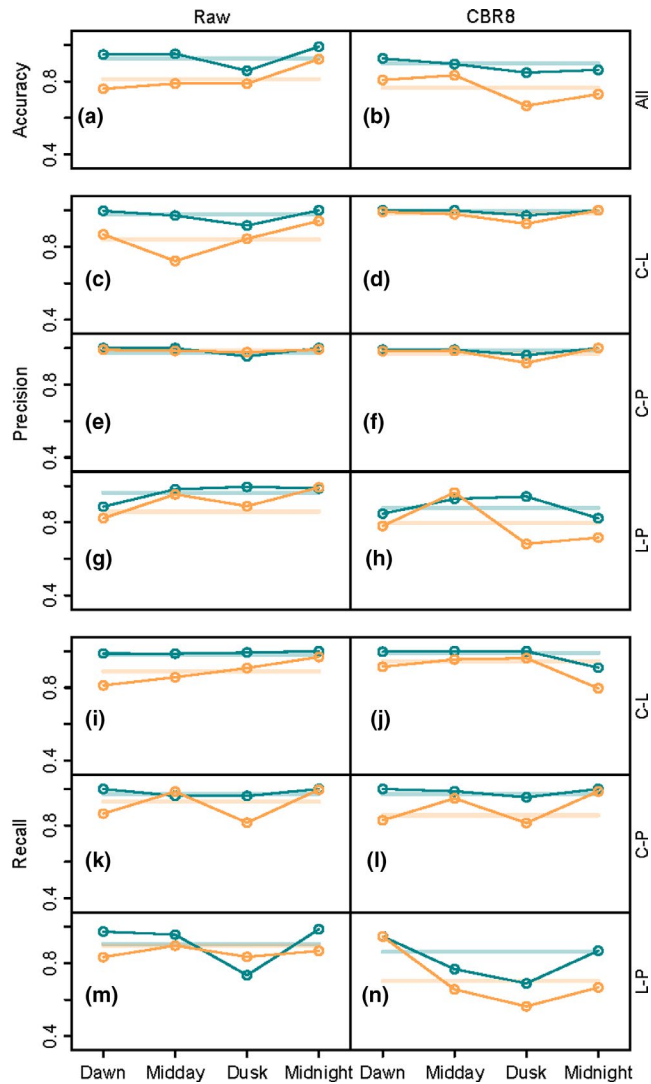


FIGURE 4 Classification model performance as a function of temporal sectioning (x-axis, compression (raw audio, left column; CBR8 compression, right column) and index choice (AudioSet Fingerprint: blue; Analytical Indices: orange). Pale horizontal lines show performance without temporal sectioning. Precision and recall are partitioned into pairwise performance by site (C, cleared forest; L, logged forest; P, primary forest)

audio showed higher accuracy for AudioSet Fingerprint (93.8%) than Analytical Indices (80.9%; Table 2). This advantage held across all recording lengths and performance metrics with performance gains of around 12%–13% in accuracy, precision, and recall (Appendix S1: Supplementary 7b).

Compression decreased accuracy for both AudioSet Fingerprint (CBR8: 90.8%) and Analytical Indices (CBR8: 75.1%; Table 2). Classifiers trained on *compressed* AudioSet Fingerprint, however, still outperformed those trained on *uncompressed* Analytical Indices. For both index types, this reflected a decreased ability to differentiate logged and primary forest. Interestingly, classifiers from both index types showed better discrimination between cleared land and logged forest under strong compression. These patterns were repeated across recording lengths (Appendix S1: Supplementary 5a).

TABLE 3 ANOVA table for the model terms in the beta regression model of the accuracy data (Significance: *** $p < .001$, ** $p < .01$, * $p < .05$. Equivalent tables for precision and recall in Appendix S1: Supplementary 9C)

	df	χ^2
log ₁₀ (File Size)	1	26.2128***
Temporal Subsetting	3	31.6818***
Frame Size	3	15.7820**
Index Type	1	2,985.9825***
log ₁₀ (File Size): Temporal Subsetting	3	18.0278***
log ₁₀ (File Size): Frame Size	3	2.9280
Temporal Subsetting: Frame Size	9	6.3156
log ₁₀ (File Size): Index Type	1	59.0065***
Temporal Subsetting: Index Type	3	7.1061
File Size: Index Type	3	36.2699***
log ₁₀ (File Size): Temporal Subsetting: Index Type	3	13.0715**
log ₁₀ (File Size): Frame Size: Index Type	3	0.8071
Temporal Subsetting: Frame Size: Index Type	9	7.1524

3.2.1 | Impact of temporal subsetting

Temporally subsetting poses a trade-off as when diel variation is reduced, so too are the recording hours available for analysis. Temporally subsetting the day into quarters (Figure 4) yielded a largely unpredictable effect on accuracy, precision, and recall. There were clear differences in discrimination between pairs of sites. Notably comparing cleared and primary forest had the highest precision across each temporal subset, index choice, and compression (Figure 4e,f), but the recall was not markedly different from other pairs (Figure 4 k,l). Temporal windows did not generally help discriminate between logged and primary forest (Table 2, Figure 4g,h,m,n), and the performance difference between AudioSet Fingerprints and Analytical Indices was largely maintained.

3.2.2 | Combined effects of parameter alterations on classification performance

Confirming prior findings (Sethi, Jones, et al., 2020), our model has demonstrated that performance measures were consistently higher when classifiers are trained on the AudioSet Fingerprint, rather than Analytical Indices (accuracy: +16.9% ($z = 10.38_{1799}$ $p < .001$), precision: +15.5% ($z = 9.717_{1799}$ $p < .001$), recall: +16.9% ($z = 10.22_{1799}$ $p < .001$), full model outputs Appendix S1: Supplementary 9C). Index type was by far the largest contributor to model accuracy (Table 3), although there was some effect of temporal subsetting, compression level, and frame size. Despite the considerable impact of compression level on index values, it appeared to have a minor effect on model accuracy (Figure 5, Table 3). The effect of frame size appeared

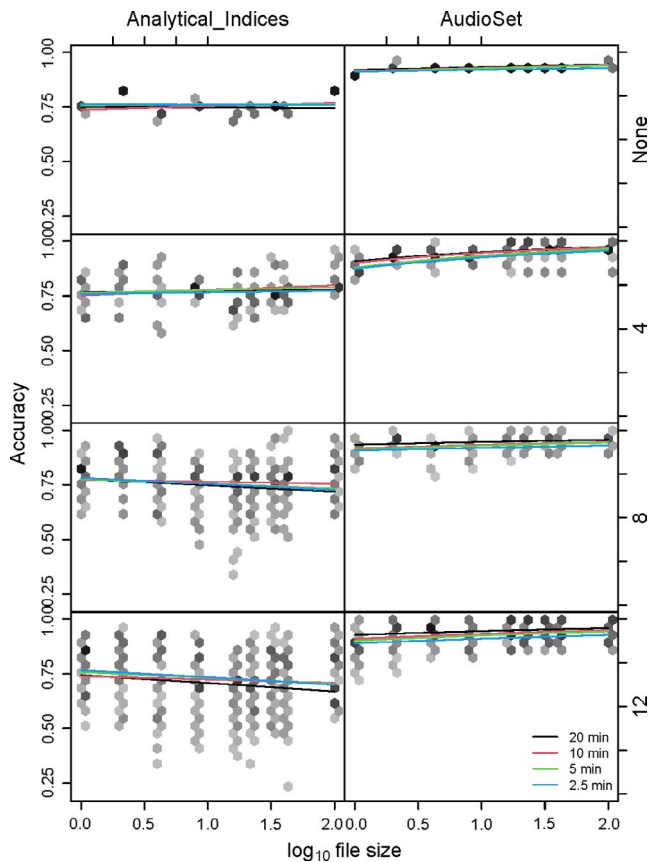


FIGURE 5 Classifier accuracy model predictions as a function of file size (x-axis), index type (columns), temporal subsetting (rows), and frame size (colors, see legend). Hexagon binning is used to show the distribution and density of the underlying data

to increase as the days were cut into smaller temporal subsections; however, this effect was small compared with the contribution of index type (Figure 5). Temporal subsetting appeared to have minimal effect on the accuracy of the AudioSet Fingerprint classifier, which kept consistently high (70%–100%; Figure 5). The classifier trained on Analytical Indices, however, became much more unpredictable when temporal subsetting is used (20%–100%; Figure 5).

4 | DISCUSSION

Ecoacoustics is a new and rapidly expanding field of ecology, with great power to describe ecological systems (e.g., Sethi, Jones, et al., 2020), but methodological choices have proliferated that have poorly known impacts on ecoacoustic analysis. We have shown that the choice of acoustic index is key and confirm (Sethi, Jones, et al., 2020) that a multidimensional generalist classifier (AudioSet Fingerprint) outperforms more traditional Analytical Indices regardless of the levels of audio compression or recording schedule.

Analytical Indices have been constrained to a limited set of features within soundscapes, leading to high degrees of correlation. For example, ADI, AEve, and H indices are all summaries of the evenness of frequency band occupancy (Sueur, Aubin, et al., 2008;

Villanueva-Rivera et al., 2011). This nonindependence can further decrease the dimensionality of suites of Analytical Indices, which are already typically small. Here, we use just the mean values of Analytical Indices, but other studies have incorporated both the mean and standard deviation (Bradfer-Lawrence et al., 2019), which provides further dimensionality. Although the AudioSet Fingerprint clearly benefits from a large number of relatively uncorrelated acoustic features, most Analytical Indices have the advantage of being designed to capture ecologically relevant aspects of the soundscape.

Compression affected the quantification of all indices in both index types (Figure 3) and—although the qualitative patterns were noisy—the groupings seen may reflect the underlying algorithms. The apparent threshold for AudioSet Fingerprint at CBR16 may be due to the obligatory loss in audio quality before samples pass to the CNN used to generate the AudioSet Fingerprint. The audio was downsampled to 16 kHz and then presented as a mel-shifted spectrogram, which increases sensitivity in frequency ranges relevant to human hearing, akin to those frequencies favored in commercial compression. Coupled with its variable quality training set (YouTube Videos), these factors may predispose AudioSet Fingerprint to perform as well with high-quality audio as with intermediate and low-quality MP3s.

The M and NDSI were also largely unaffected by compression until the frequency range is reduced. When mp3 audio is compressed below 32 kb/s the audio swaps from being encoded as MPEG-1 Audio Layer III (which supports max frequency of 16–24 kHz) to MPEG-2 Audio Layer III (max: 8–12 kHz), this change in format results in the removal of signals beyond the cutoff frequency threshold. Further reduction is seen where at CBR8 when encoding changes again to MPEG-2.5 Audio Layer III (max: 4–6 kHz). The M index is explicitly a measure of amplitude (Sueur et al., 2014) and is largely unaffected until downsampling reduces amplitude. Similarly, NDSI measures the proportion of sound in biophonic versus anthropophonic frequency bands: As downsampling progressively eliminates sounds within the frequency range (2–11 kHz) containing most biophony, NDSI is known to increase (Kasten et al., 2012). The ADI index also shows a marked increase in the magnitude of the difference at higher rates of compression (CBR64); however, a small but significant difference can be observed from CBR256. The ADI index measures the spread of frequencies above a certain loudness threshold, the effect of compression on ADI, may therefore suggest that certain high-frequency bands are dominant in this soundscape.

AEve and H, both of which describe the spread and evenness of amplitude over the full range of frequencies, showed a gradual increase in *D* that reversed when the Nyquist frequency reduced. The two measures differ in measuring dominance (AEve: Villanueva-Rivera et al., 2011) and evenness (H: Sueur et al., 2014) across bands but may share a common explanation. In both cases, compression preferentially removed amplitude from some bands, initially decreasing evenness but downsampling removes bands entirely, possibly restoring a more even distribution.

ACI and Bio both shared a dependence on high frequency or quieter sounds and were generally most severely affected by

compression. ACI measures frequency band-dependent changes in amplitude over time (Pieretti et al., 2011) and is reduced when there is minimal variation between time steps. Loss of “masked” sounds under low compression and then 16–24 kHz sound under CBR16 may reflect the loss of ecoacoustic temporal variation: This band includes the calling range of many invertebrates, birds, mammals, and amphibians (Browning et al., 2017). The Bio index similarly quantifies the spread of frequencies in the range 2 kHz–11 kHz, all relative to the quietest 1 kHz band (Boelman et al., 2007): Loss of quiet frequency bands, therefore, make it uniquely sensitive to compression. Despite both of these indices incurring alterations 200% larger than the uncompressed range, the Analytical Indices classifier accuracy still showed robustness to compression, perhaps suggesting these indices are less important for classification than the others. Bradfer-Lawrence et al. (2019) have already shown that the Bio index contributes little additional power when classifying soundscapes, but found that ACI was the strongest individual contributor in this suite of indices (Bradfer-Lawrence et al., 2019). Our findings suggested this ranking may not be consistent across different levels of compression.

Our findings reflect those of an earlier study that explored the effect of mp3 compression (VBR0 and CBR128) on indices describing specific bird calls (Araya-Salas et al., 2019). They found that compression did not cause a systemic deviation in all indices, but rather indices designed to capture extreme frequencies were less precise after compression, particularly with VBR encoded files (Araya-Salas et al., 2019). While some of these principles are present in our findings, the use of a wider range of compressions has allowed us to develop a more complete description of the action of compression on soundscape indices.

We found that even the highest rate of compression caused a comparatively small reduction in the overall accuracy of the classification task (5.8% and 3% for Analytical Indices and the AudioSet Fingerprint, respectively, for the 5-min recordings without temporal subsetting). In both cases, the reduction in accuracy was explained by a higher degree of overlap between primary and logged forests. When audio is compressed, the whole signal is altered but higher frequencies and quieter sounds are more severely altered and reduced than others (Sterne, 2012). Higher and quieter frequencies (akin to specific animal vocalizations) may therefore be more important for separating logged and primary—but less so for discerning cleared from other forest types (which may be more dependent on overall amplitude). These proportionally small differences, while somewhat reassuring, should be considered with caution they may be due to the large differences in habitat structure among our three habitat classes. Combining this with our relatively small sample size, we would like to emphasize that these findings may therefore not be generalizable to areas of more closely related forest.

Both Analytical Indices and AudioSet Fingerprint had similar changes in variance as a result of recording length. Transient vocalizers are therefore likely somewhat important in the determination of the AudioSet Fingerprint and variable importance in some Analytical Indices. The ACI index was not impacted by recording

length despite specifically quantifying how the soundscape changes over time (Pieretti et al., 2011). The ADI, AEve, and H all did incur an alteration in variance as recording length changed; interestingly, these indices do not consider any temporal value but rather just the spread of frequency (Sueur, Pavoine, et al., 2008; Villanueva-Rivera et al., 2011), indicating that transient calls akin to short-term anomalies in frequency are perhaps lost when recording windows are altered.

Finally, we found that subsetting audio data temporally and analyzing them separately had an unpredictable impact on classification accuracy, with the AudioSet Fingerprint classifier staying consistently high while the Analytical Indices classifier was returning accuracies anywhere between 20% and 100%. Temporal subsetting can reduce the impact of diel variation on analyses but poses a trade-off as it reduces the amount of data used to train the classifier. Analytical Indices may perform better over longer recording periods as >120 hr of recordings are required for Analytical Indices to stabilize (Bradfer-Lawrence et al., 2019), yet in our study, we had just 70–75 hr of recordings per site. Overall we found that compression, frame size, and temporal subsetting caused a small decrease in classifier accuracy, with the largest overall contributor being the choice of AudioSet Fingerprinting over Analytical Indices. The AudioSet Fingerprint classifier, temporally sectioned, and trained on just 2 hr of data was able to, on average, outperform the Analytical Indices classifier trained on the full 24 hr.

5 | RECOMMENDATIONS AND CONCLUSION

This study was designed to compare distinct forest types in Malaysian Borneo, and the recording periods used are relatively small. Based on the results of this study, we provide the following four recommendations; however, effort should be made to ensure they are generalizable to the desired area of deployment:

1. We provide additional evidence for the viability and stability of AudioSet Fingerprinting rather than Analytical Indices when classifying soundscapes.
2. Lossless compression is always desirable but if data storage/transmission become a bottleneck to a study, we advise using the VBR (quality = 0) MP3 encoder if using Analytical Indices, which will reduce the file size to roughly 23% of the original while having minimal impact on indices (other than ACI). The AudioSet Fingerprint, however, is more robust to compression and so can tolerate a minimum compression encoding of CBR64 (8% of the original file size) without significant effect.
3. If further compression is a necessity, use indices which describe the general energy of the system rather than those which are dependent on high frequency or quieter sounds, such as ACI.
4. Temporal subsetting may be a useful alternative for capturing soundscape descriptors with AudioSet Fingerprinting when data storage costs are a bottleneck. However, temporal subsetting

should be used with caution when using Analytical Indices owing to the variation in classification accuracy, precision, and recall.

There exists a trade-off between the quality and volume of data that can be stored in ecoacoustics. We have investigated the impact of compression along a gradient of habitat disturbance, providing evidence that compressed audio can be used without severely affecting either of the index type. The ability to use compression may reduce experimental costs, remove bottlenecks in study design, and help remote ecoacoustic recorders reach true autonomy. Moreover, by providing a quantified description of how individual indices, and more broadly grouped index categories, respond to compression, we have enabled comparisons to be drawn between studies of compressed and noncompressed audio. Increasing comparability of studies will become progressively important as global ecoacoustic databases, and recording sites grow and open up novel opportunities to explore datasets across huge temporal and geographic scales.

ACKNOWLEDGMENTS

We firstly thank Dr Henry Bernard at the Sustainability of Altered Forest Ecosystems project in Malaysian Borneo for permitting us to research within their field sites. This project was funded by the Natural Environmental Research Council, UK, within the Quantitative Methods in Ecology and Evolution (QMEE) Centre for Doctoral Training (grant number: NE/P012345/1).

CONFLICT OF INTEREST

No conflict of interest to declare.

AUTHOR CONTRIBUTIONS

Becky E. Heath: Conceptualization (lead); Data curation (lead); Formal analysis (equal); Funding acquisition (equal); Investigation (lead); Methodology (equal); Project administration (lead); Visualization (equal); Writing-original draft (lead); Writing-review & editing (equal). **C. David L. Orme:** Data curation (equal); Formal analysis (equal); Supervision (equal); Visualization (equal); Writing-review & editing (equal). **Sarab S. Sethi:** Conceptualization (supporting); Data curation (equal); Formal analysis (supporting); Methodology (equal); Writing-review & editing (equal). **Robert M. Ewers:** Conceptualization (equal); Methodology (equal); Project administration (equal); Writing-review & editing (equal). **Lorenzo Picinali:** Conceptualization (equal); Formal analysis (equal); Methodology (equal); Project administration (equal); Supervision (equal); Writing-review & editing (equal).

OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at AudioSet/ Analytical Index Data: <https://doi.org/10.5281/zenodo.5153193>. Raw Audio Files: <https://doi.org/10.5281/zenodo.5159914>. Data Analysis Repo:

<https://github.com/BeckyHeath/Experimental-Variation-Ecoacoustics-Analysis-Scripts>.

DATA AVAILABILITY STATEMENT

Acoustic Data: Available at 10.5281/zenodo.5159914. Analytical Indices/AudioSet Fingerprint Data: Available at 10.5281/zenodo.5153193. Analysis Scripts: Available on Github at <https://github.com/BeckyHeath/Experimental-Variation-Ecoacoustics-Analysis-Scripts> (made public after publication).

ORCID

Becky E. Heath  <https://orcid.org/0000-0002-9111-6994>

Sarab S. Sethi  <https://orcid.org/0000-0002-5939-0432>

C. David L. Orme  <https://orcid.org/0000-0002-7005-1394>

Lorenzo Picinali  <https://orcid.org/0000-0001-9297-2613>

REFERENCES

- Araya-Salas, M., Smith-Vidaurre, G., & Webster, M. (2019). Assessing the effect of sound file compression and background noise on measures of acoustic signal structure. *Bioacoustics*, 28(1), 57–73. <https://doi.org/10.1080/09524622.2017.1396498>
- Boelman, N. T., Asner, G. P., Hart, P. J., & Martin, R. E. (2007). Multi-trophic invasion resistance in Hawaii: Bioacoustics, field surveys, and airborne remote sensing. *Ecological Applications*, 17(8), 2137–2144. <https://doi.org/10.1890/07-0004.1>
- Bohnenstiehl, D. W. R., Lyon, R. P., Caretti, O. N., Ricci, S. W., & Eggleston, D. B. (2018). Investigating the utility of ecoacoustic metrics in marine soundscapes. *Journal of Ecoacoustics*, 2(2), 1. <https://doi.org/10.22261/jea.r11561>
- 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(10), 1796–1807. <https://doi.org/10.1111/2041-210x.13254>
- Browning, E., Gibb, R., Glover-Kapfer, P., & Jones, K. (2017). Passive acoustic monitoring in ecology and conservation. *WWF Conservation Technology Series* 1, 2, 1–75. <https://doi.org/10.13140/RG.2.2.18158.46409>
- Buxton, R. T., McKenna, M. F., Clapp, M., Meyer, E., Stabenau, E., Angeloni, L. M., Crooks, K., & Wittemyer, G. (2018). Efficacy of extracting indices from large-scale acoustic recordings to monitor biodiversity. *Conservation Biology*, 32(5), 1174–1184. <https://doi.org/10.1111/cobi.13119>
- Costello, M. J., Beard, K. H., Corlett, R. T., Cumming, G. S., Devictor, V., Loyola, R., Maas, B., Miller-Rushing, A. J., Pakeman, R., & Primack, R. B. (2016). Field work ethics in biological research. *Biological Conservation*, 203, 268–271. <https://doi.org/10.1016/j.biocon.2016.10.008>
- Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34(2), 1–24. <https://doi.org/10.18637/jss.v034.i02>
- Douma, J. C., & Weedon, J. T. (2019). Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods in Ecology and Evolution*, 10(9), 1412–1430. <https://doi.org/10.1111/2041-210X.13234>
- 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>
- Ewers, R. M., Didham, R. K., Fahrig, L., Ferraz, G., Hector, A., Holt, R. D., Kapos, V., Reynolds, G., Sinun, W., Snaddon, J. L., & Turner, E. C.

- (2011). A large-scale forest fragmentation experiment: The stability of altered forest ecosystems project. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1582), 3292–3302. <https://doi.org/10.1098/rstb.2011.0049>
- Fitzpatrick, M. C., Preisser, E. L., Ellison, A. M., & Elkinton, J. S. (2009). Observer bias and the detection of low-density populations. *Ecological Applications*, 19(7), 1673–1679. <https://doi.org/10.1890/09-0265.1>
- Fuller, S., Axel, A. C., Tucker, D., & Gage, S. H. (2015). Connecting soundscape to landscape: Which acoustic index best describes landscape configuration? *Ecological Indicators*, 58, 207–215. <https://doi.org/10.1016/j.ecolind.2015.05.057>
- Gemmeke, J. F., et al. (2017). Audio Set: An ontology and human-labeled dataset for audio events. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (pp. 776–780). <https://doi.org/10.1109/ICASSP.2017.7952261>
- Gómez, W. E., Isaza, C. V., & Daza, J. M. (2018). Identifying disturbed habitats: A new method from acoustic indices. *Ecological Informatics*, 45, 16–25. <https://doi.org/10.1016/j.ecoinf.2018.03.001>
- Hershey, S., et al. (2017). CNN architectures for large-scale audio classification. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (pp. 131–135). <https://doi.org/10.1109/ICASSP.2017.7952132>
- 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 in Ecology and Evolution*, 9(5), 1199–1211. <https://doi.org/10.1111/2041-210X.12955>
- Huff, M. H., Bettinger, K. A., Furguson, H. L., Brown, M. J., & Altman, B. (2000). A habitat-based point-count protocol for terrestrial birds, emphasizing Washington and Oregon. In *General Technical Reports of the US Department of Agriculture, Forest Service (PNW-GTR-501)* (pp. 2–30).
- Kausch, R. (2019). *fre:ac [Windows App]*. <https://www.freac.org>
- Kasten, 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>
- Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, 2(3), 18–22.
- Linke, S., & Deretic, J. A. (2020). Ecoacoustics can detect ecosystem responses to environmental water allocations. *Freshwater Biology*, 65(1), 133–141. <https://doi.org/10.1111/fwb.13249>
- Mammides, C., Goodale, E., Dayananda, S. K., Kang, L., & Chen, J. (2017). Do acoustic indices correlate with bird diversity? Insights from two biodiverse regions in Yunnan Province, south China. *Ecological Indicators*, 82, 470–477. <https://doi.org/10.1016/j.ecolind.2017.07.017>
- Pfeifer, M., Chung, A. C., Turner, E., Lysenko, I., Cusack, J., Kor, L., Khoo, M., Chey, V. K., & Ewers, R. (2016). Mapping the structure of Borneo's tropical forests across a degradation gradient. *Remote Sensing of Environment*, 176, 84–97.
- Pieretti, N., Farina, A., & Morri, D. (2011). A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). *Ecological Indicators*, 11(3), 868–873. <https://doi.org/10.1016/j.ecolind.2010.11.005>
- R Core Team. (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. <https://www.R-project.org/>
- Rapport, D. J. (1989). What constitutes ecosystem health? *Perspectives in Biology and Medicine*, 33(1), 120–132. <https://doi.org/10.1353/pbm.1990.0004>
- Rapport, D. J., Costanza, R., & McMichael, A. J. (1998). Assessing ecosystem health. *Trends in Ecology and Evolution*, 13(10), 397–402. [https://doi.org/10.1016/S0169-5347\(98\)01449-9](https://doi.org/10.1016/S0169-5347(98)01449-9)
- Robert, J., & Webbie, M. (2018). Pydub. GitHub: Retrieved from <http://pydub.com/>
- Roca, I. T., & Proulx, R. (2016). Acoustic assessment of species richness and assembly rules in ensiferan communities from temperate ecosystems. *Ecology*, 97(1), 116–123. <https://doi.org/10.1890/15-0290.1>
- Saito, K., Nakamura, K., Ueta, M., Kurosawa, R., Fujiwara, A., Kobayashi, H. H., Nakayama, M., Toko, A., & Nagahama, K. (2015). Utilizing the Cyberforest live sound system with social media to remotely conduct woodland bird censuses in Central Japan. *Ambio*, 44(4), 572–583. <https://doi.org/10.1007/s13280-015-0708-y>
- Sethi, S. S., Ewers, R. M., Jones, N., Picinali, L., Orme, C. L., Sleutel, J., Shabrani, A., Zulkifli, N., & Bernard, H. (2020). Avifaunal and Herpetofaunal point counts with recorded acoustic data. *Zenodo*. <https://doi.org/10.5281/zenodo.3997172>
- Sethi, S. S., Ewers, R. M., Jones, N. S., Orme, C. D. L., & Picinali, L. (2018). Robust, real-time and autonomous monitoring of ecosystems with an open, low-cost, networked device. *Methods in Ecology and Evolution*, 9, 2383–2387. <https://doi.org/10.1111/2041-210X.13089>
- Sethi, S. S., Jones, N. S., Fulcher, B. D., Picinali, L., Clink, D. J., Klinck, H., Orme, C. D. L., Wrege, P. H., & Ewers, R. M. (2020). Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. *Proceedings of the National Academy of Sciences of the United States of America*, 24, 1–7. <https://doi.org/10.1073/pnas.2004702117>
- Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11(1), 54–71.
- Sterne, J. (2012). *The meaning of a format MP3*. Duke University Press.
- Struebig, M. J., Turner, A., Giles, E., Lasmana, F., Trollington, S., Bernard, H., & Bell, D. (2013). Quantifying the biodiversity value of repeatedly logged rainforests. In *Gradient and comparative approaches from Borneo. Advances in ecological research* (48, 1st ed., 183–224). Elsevier Ltd.
- Sueur, J., Aubin, T., & Simonis, C. (2008). Equipment review: Seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18(2), 213–226. <https://doi.org/10.1080/09524622.2008.9753600>
- Sueur, J., Farina, A., Gasc, A., Pieretti, N., & Pavoine, S. (2014). Acoustic indices for biodiversity assessment and landscape investigation. *Acta Acustica United with Acustica*, 100(4), 772–781. <https://doi.org/10.3813/AAA.918757>
- Sueur, J., Krause, B., & Farina, A. (2019). Climate change is breaking earth's beat. *Trends in Ecology & Evolution*, 34(11), 971–973. <https://doi.org/10.1016/j.tree.2019.07.014>
- Sueur, J., Pavoine, S., Hamerlynck, O., & Duvail, S. (2008). Rapid acoustic survey for biodiversity appraisal. *PLoS One*, 3(12), e4065. <https://doi.org/10.1371/journal.pone.0004065>
- Sugai, L. S. M., Desjonquères, C., Silva, T. S. F., & Llusia, D. (2019). A road-map for survey designs in terrestrial acoustic monitoring. *Remote Sensing in Ecology and Conservation*, 6(3), 220–235. <https://doi.org/10.1002/rse2.131>
- Towsey, M. (2018). The calculation of acoustic indices derived from long-duration recordings of the natural environment. *QUT ePrints* (vol. 110634, pp. 1–12). <https://eprints.qut.edu.au/110634/>
- Towsey, M. W., Truskinger, A. M., & Roe, P. (2016). The navigation and visualisation of environmental audio using zooming spectrograms. In *Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015* (pp. 788–797). <https://doi.org/10.1109/ICDMW.2015.118>
- 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. *Ecological Informatics*, 21(100), 110–119. <https://doi.org/10.1016/j.ecoinf.2013.11.007>
- Vesna, I. (2009). Understanding Bland Altman analysis. *Biochemia Medica*, 19(1), 10–16. <https://doi.org/10.11613/BM.2013.003>
- Villanueva-Rivera, L. J., & Pijanowski, B. C. (2016). *Package "soundecology"*. CRAN, (Package version 1.3.3), pp. 14. <http://ljjvillanueva.github.io/soundecology/>

- Villanueva-Rivera, L. J., Pijanowski, B. C., Doucette, J., & Pekin, B. (2011). A primer of acoustic analysis for landscape ecologists. *Landscape Ecology*, 26(9), 1233–1246. <https://doi.org/10.1007/s10980-011-9636-9>
- Zhang, L., Towsey, M., Zhang, J., & Roe, P. (2016). Classifying and ranking audio clips to support bird species richness surveys. *Ecological Informatics*, 34, 108–116. <https://doi.org/10.1016/j.ecoinf.2016.05.005>

How to cite this article: Heath, B. E., Sethi, S. S., Orme, C. D. L., Ewers, R. M., & Picinali, L. (2021). How index selection, compression, and recording schedule impact the description of ecological soundscapes. *Ecology and Evolution*, 11, 13206–13217. <https://doi.org/10.1002/ece3.8042>

SUPPORTING INFORMATION

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