How Index Selection, Compression and Recording Schedule Impact the description of Ecological Soundscapes.

Becky Heath[1], David Orme[2], Sarab Sethi[1], Robert Ewers[2], and Lorenzo Picinali[1]

[1]Imperial College London
[2]Imperial College London Department of Life Sciences Silwood Park Campus

March 24, 2021

Abstract

1. Environmental soundscapes are increasingly being used as descriptors of 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 landscape classification task. 3. We find that compression and frame size both drive considerable variation in calculated index values. However, we find that the effects of this varaiation 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 substantial (12-16%) increases in all of classifier accuracy, precision and recall. 4. We advise using the AudioSet Fingerprint in soundscape analysis, demonstrating its 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, 23% file size) to reduce file size without affecting most Analytical Index values. The AudioSet Fingerprint can be confidently compressed further to a Constant Bit Rate encoding of 64kb/s (8% file size) without any detectable effect. These recommendations balance the efficient use of restricted data storage against the comparability of results between different studies.

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. Orme[2], Robert M. Ewers[2], Lorenzo Picinali[1]

1. Dyson School of Design Engineering, Imperial College London, London. UK.
2. Department of Life Sciences, Imperial College London, London. UK.
3. Department of Mathematics, Imperial College London, London.UK.

Correspondence Author:

Becky Heath

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

r.heath18@imperial.ac.uk

Introduction:

Animal vocalisations come together with natural and human-made sounds to form soundscapes, which can be used to monitor species populations or infer community-level metrics such as biodiversity (Roca and Proulx, 2016; Eldridge *et al.* , 2018; Gómez, Isaza and Daza, 2018). Such monitoring is crucial to effectively respond to threats (Rapport, 1989; Rapport, Costanza and McMichael, 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 (Fitzpatrick *et al.* , 2009; Costello *et al.* , 2016). Advances in portable computing now permit remote recording of soundscapes, but produce a volume of data that precludes manual review, leading to the development of automated, or semi-automated, methods of analysis (Towsey, Truskinger and Roe, 2016; Sethi *et al.* , 2020).

Soundscape composition is primarily assessed using acoustic indices – summary statistics that describe the distribution of acoustic energy within the recording (Towsey *et al.* , 2014) – and over 60 Analytical Indices which capture aspects of biodiversity have been developed (Sueur *et al.* , 2014; Buxton *et al.* , 2018). 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 (Gomez, Isaza and Daza, 2018). These approaches have provided novel insight into ecosystems across the world (Fuller*et al.* , 2015; Buxton *et al.* , 2016; Eldridge *et al.* , 2018; Sueur, Krause and Farina, 2019) but are not foolproof and often have poor transferability (Mammides *et al.* , 2017; Bohnenstiehl*et al.* , 2018). This may result from a lack of standardisation: differing index selection, data storage methods, and recording protocols, which all lead to unassessed variation in experimental outputs (Araya-Salas, Smith-Vidaurre and Webster, 2019; Bradfer-Lawrence*et al.* , 2019; Sugai *et al.* , 2019).

The AudioSet convolutional neural net (CNN; Gemmeke *et al.* , 2017; Hershey *et al.* , 2017) is an attractive replacement for Analytical Indices. This pre-trained, general-purpose audio classifier generates a multi-dimensional acoustic fingerprint of a soundscape that is a more effective ecological descriptor (Sethi *et al.* , 2020). AudioSet is trained on two million human-labelled anthropogenic and environmental audio samples, potentially giving it both greater transferability and discrimination than typical ecoacoustic training datasets.

In ecoacoustics, a continuous uncompressed or lossless recording is generally recommended (Villanueva-Rivera *et al.* , 2011; Browning*et al.* , 2017), but generates huge files. We consider 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; Zhang *et al.* , 2016; Sethi *et al.* , 2018): this lossy encoding removes acoustic information inaudible to *human* listeners but is suspected of removing ecologically important data (e.g. Towsey, Truskinger and Roe, 2016; Sugai *et al.* , 2019). Araya-Salas, Smith-Vidaurre and Webster (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 (Fuller*et al.* , 2015; Bradfer-Lawrence *et al.* , 2019; Sethi*et al.* , 2020) and so separating recording windows 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. Here, we:

contrast the classification accuracy of index selection choices;

describe the effects of both 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 identify stronger standards to improve both performance and provide a basis for more extensive meta-analysis.

Methods and Materials

Study Area

Acoustic samples were collected in Sabah 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) with 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. Habitat ranges from areas of grass and low shrub, through logged forest to almost undisturbed primary forest.

Soundscape Recording

Data were collected from three KFR sites representing a gradient in above-ground biomass (figure 4a) (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}) (Supplementary 1). We recorded for an average of 72 hours at each site (range: 70 to 75) during February and March 2019 (Supplementary 2a). No rain fell during the recording period, so no recordings were excluded due to confounding geophony (Zhang *et al.* , 2016). In all sites, omnidirectional (AudioMoth, Hill *et al.* , 2018) recorders were attached to trees (~ 50 cm diameter and 1-2 m above the ground) and recorded continuously using 20-minute uncompressed samples ('raw', .wav format) at 44.1kHz and 16 bits.

Compressing and Re-Sizing the Raw Audio

Continuous 20-minute recordings were first split into recordings with a length of 2.5, 5.0 and 10.0 minutes, using the python package*pydub* (Webbie et al., 2018) (Fig. 1b). The audio was then converted to lossy MP3 format using the fre:ac LAME encoder under two standard LAME MP3 encoding techniques: constant bit rate (CBR) and variable bit rate (VBR) compression (Fig. 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 use compressed file size as our measure of compression level. We used VBR0 and CBR320, CBR256, CBR128, CBR64, CBR32, CBR16 and CBR8, resulting in file sizes ranging from 41.6% (CBR320) and 1.04% (CBR8) of the original raw file size and some reductions in maximum coded frequency (Table 1). We do not consider lossless compression, as the storage capacity is much higher and the files are obligatorily the same post decompression. Previous studies have also found that the lossless compressed audio is largely identical to raw audio (Linke and Deretic, 2020).

Compression Level	Bit storage/s	% File Size	Maximum Coded Frequency (kHz)
RAW	Constant: 768kb	100	22.05
VBR0	Variable: ~ 127 – 250kb	mean = 20.82 range = 32.64 – 16.63	22.05
CBR320	Constant: 320kb	41.6	22.05
CBR256	Constant: 256kb	33.35	22.05
CBR128	Constant: 128kb	16.67	22.05
CBR64	Constant: 64kb	8.33	22.05
CBR32	Constant: 32kb	4.16	11.025
CBR16	Constant: 16kb	2.08	8
CBR8	Constant: 8kb	1.04	4

Table 1: Bitrate, percentage file size reduction and maximum encodable frequency for the experimental

compression levels.

Quantification of Soundscapes Using Indices

Analytical Indices

We used the *seewave* (Sueur, Aubin and Simonis, 2008) and*soundecology* (Villanueva-Rivera and Pijanowski, 2016) packages in R (ver 3.6.1; R Core Team, 2020) to extract 7 Analytical Indices (Fig. 4d): Acoustic Complexity Index (ACI), Acoustic Diversity Index (ADI), Acoustic Evenness (AEve), Bioacoustic Index (Bio), Acoustic Entropy (H), Median of the Amplitude Envelope (M), and Normalised Difference Soundscape Index (NDSI) (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 (Supplementary 2b).

AudioSet Fingerprint

The audio was converted to a log-scaled Mel-frequency spectrogram after 16kHz 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) (Fig. 1d). This generates a 128-dimensional embedding and the 128 values in that embedding describe the soundscape of given recording in an abstracted form or fingerprint. Similarly, as in the Analytical Indices, some recordings could not be analysed by the AudioSet CNN, however, this was only in 0.2% of recordings (Supplementary 2b).

Data Analysis

Impact of Index Selection: Auto-Correlation

Analytical Indices often summarise similar features of a soundscape (e.g. dominant frequency and frequency bin occupancy): this overlap may reduce the descriptive scope of the ensemble. We compare the degree of pairwise correlation between the individual Analytical Indices and between the individual features of the AudioSet Fingerprint. We also compare how well each index/feature correlates with the maximum recordable frequency (Fig. 1e).

Impact of Compression: Like-for-Like Differences

We use an adaption of Bland-Altman plots (Vesna, 2009, Araya-Salas, Smith-Vidaurre and Webster, 2019) to visualise 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} (Fig. 1f) :

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

D was not normally distributed (Supplementary 5a), so median and inter-quartile ranges are reported. We determine that an index has been altered as a result of compression to be when: i) the interquartile range of D does not include zero difference or ii) median$D$ is more than +/- 5% of the R_{raw} . We use Spearman rank correlation to test for a consistent trend in $D$with increasing compression. To reflect their common use cases, $D$for Analytical Indices is calculated from the univariate values, while for AudioSet Fingerprints – which is intended as a multidimensional metric – D is calculated separately for each dimension and then averaged.

Impact of Recording Schedule: Recording Length

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

Impact of Parameter Alteration on Classification Task

We use random forest classification models to assess how well the soundscapes are represented by each index type under each different experimental parameter, using the *RandomForest* (Liaw and Wiener 2002) package in R (Fig. 1h). Models were trained on a middle 24 h 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 stabilised. 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 accuracy, precision and recall of each combination.

Impact of Temporal Subsetting

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

Modelling the Impact of all Parameters on 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 and Weedon, 2019). The model was built using the *betareg* package in R (Cribari-Neto and Zeileis, 2010). To avoid fitting issues when performance measures are exactly 1, we rescale all performance measures using m' = (m (n-1) + 0.5) / n, where n is sample size (Smithson & Verkuilen, 2006). The model includes 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 (Supplementary 8a), so tested the inclusion of these terms in the precision component of the model. We first treated frame size 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 (Supplementary 8b).

Results

Although Spearman pairwise correlations of Analytical Indices and maximum recordable frequency were low on average (mean = 0.32, IQR = 0.22), we found some strongly correlated sets of indices (Fig. 2). ADI, Bio and NDSI all show strong similarities and are closely correlated with maximum recordable frequency; AEve and H are also strongly correlated (Fig. 2). Some features of the AudioSet Fingerprint correlate with each other and maximum frequency but in general, these features are more weakly correlated (mean = 0.14, IQR = 0.18, Fig. in Supplementary 4b).

Impact of Compression

Impact of Compression: Like-for-Like Differences

All indices showed both observable differences under compression and clear trends with increasing compression (confirmed with Spearman's rank correlation, all $p < 0.001$, Supplementary 5b). The mode of response showed three broad qualitative patterns, illustrated here using results from the 5-minute audio sample (other recording lengths in Supplementary 5a). (1) 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%). (2) AEve and H showed the biggest differences at an intermediate compression (CBR64) and relatively low absolute differences (median D typically < 30%). (3) The remaining indices showed a variety of responses: ADI showed a monotonic response above a threshold, ACI showed changes up to CBR64 and then stabilises, 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.

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 values (46 out of 128; 36%) were found to have non-homogeneous variance in groups of different recording length ($p < 0.05$, Levene's test for homogeneity of variance, Supplementary 6b).

Impact of Index Selection

Classifiers derived from 5-minute recordings using raw 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 (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 indices, this reflected a decreased ability to differentiate logged and primary forest. Interestingly, both indices showed better discrimination between cleared land and logged forest under strong compression. These patterns were repeated across recording lengths (Supplementary 5a).

	AudioSet Fingerprint	AudioSet Fingerprint	AudioSet Fingerprint		Analytical Indices	Analy
Observed	Predicted	Predicted	Predicted	Observed	Predicted	Predi
a) Raw	Cleared	Logged	Primary	**b) Raw**	Cleared	Logg
Cleared	585	9	11	Cleared	484	67
Logged	11	508	44	Logged	97	421
Primary	17	14	521	Primary	9	61
c) CBR8	Cleared	Logged	Primary	**d) CBR8**	Cleared	Logg
Cleared	585	3	17	Cleared	484	23
Logged	2	488	73	Logged	9	379
Primary	11	53	488	Primary	9	115

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).

Impact of Temporal Subsetting

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

Combined Effects of Parameter Alterations on Classification Performance

Our model has shown that performance measures were consistently higher when classifiers are trained on the AudioSet Fingerprint, rather than Analytical Indices (Accuracy: +16.9% ($z=10.38_{1799}p<0.001$), Precision: +15.5% ($z = 9.717_{1799}p<0.001$), Recall: +16.9% ($z=10.22_{1799}p<0.001$), full model outputs Supplementary 9C). Index type was by far the largest contributor to model accuracy (Table 3), although there was some effect of temporal splitting, compression level and frame size. Despite the considerable impact of compression level on index values, it appeared to have a minor effect of model accuracy (Fig. 5, Table 3). The effect of frame size appeared to increase as the days were cut into smaller temporal subsections, however, this effect was small compared to the contribution of index type (Fig. 5). Temporal subsetting appeared to have minimal effect on the accuracy of the AudioSet Fingerprint classifier, which kept consistently high (70-100%, Fig. 5). The Analytical Indices, however, became much more unpredictable when temporal subsetting is used (20-100%, Fig. 5)

	Df	χ^2
log10(File Size)	1	26.2128 ***
Temporal Subsetting	3	31.6818 ***
Frame Size	3	15.7820 **
Index Type	1	2985.9825 ***
log10(File Size): Temporal Subsetting	3	18.0278 ***
log10(File Size): Frame Size	3	2.9280
Temporal Subsetting: Frame Size	9	6.3156
log10(File Size): Index Type	1	59.0065 ***
Temporal Subsetting: Index Type	3	7.1061
File Size: Index Type	3	36.2699 ***
log10(File Size): Temporal Subsetting: Index Type	3	13.0715 **
log10(File Size): Frame Size: Index Type	3	0.8071
Temporal Subsetting: Frame Size: Index Type	9	7.1524

*Table 3. Anova table for the model terms in the beta regression model of the accuracy data. (Significance: *** p <0.001, ** p<0.01, * p<0.05). Equivalent tables for precision and recall in Supplementary 9C).*

Discussion

Ecoacoustics is a new and rapidly expanding field of ecology, with great power to describe ecological systems (e.g. Sethi *et al.* , 2020), but methodological choices have proliferated that have poorly known impacts on ecoacoustic analysis. We show that the choice of audio index is key and confirm (Sethi *et al.* , 2020) that a multi-dimensional generalist classifier 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 strong non-independence. For example, ADI, AEve and H indices are all summaries of the evenness of frequency band occupancy (Sueur, Aubin and Simonis, 2008; Villanueva-Rivera*et al.* , 2011). This non-independence 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 (Fig. 3) and – although the qualitative patterns are 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 AudioSet CNN. The audio is downsampled to 16kHz and then presented as a mel-shifted spectrogram, which increases sensitivity in frequency ranges relevant to human hearing, akin to those frequencies favoured 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 32kb/s the audio swaps from being encoded as MPEG-1 Audio Layer III (which supports max frequency of 16-24kHz) to MPEG-2 Audio Layer III (max: 8-12kHz), this change in format results in the removal of signals beyond the cut-off frequency threshold. Further reduction is seen where at CBR8 when encoding changes again to MPEG-2.5 Audio Layer III (max: 4-6kHz). 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 vs. 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).

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 maximum coded frequency reduced. The two measures differ in measuring dominance (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 removes amplitude from some bands, initially decreasing evenness but downsampling removes bands entirely, possibly restoring a more even distribution.

ACI and Bio all share a dependence on high frequency or quieter sounds and were generally most severely affected by compression. ACI measures frequency band dependant changes in amplitude over time (Pieretti, Farina and Morri, 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 2kHz- 11kHz, all relative to the quietest 1kHz 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 to classification tasks, but found that ACI was the strongest individual contributor (Bradfer-Lawrence *et al.* , 2019). Our findings suggest 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, Smith-Vidaurre and Webster, 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, Smith-Vidaurre and Webster, 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 the overall accuracy of the classification task (5.8% and 3% for Analytical Indices and the AudioSet Fingerprint respectively, 5minute whole-day). In both cases, the reduction in accuracy was explained by a higher degree of overlap between primary and logged forest. When audio is compressed, the whole signal is altered but higher frequencies and quieter sounds are more severely altered and reduced than others . Higher and quieter frequencies (akin to specific animal vocalisations) 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 level). These proportionally small differences, while somewhat reassuring, should be considered with caution it could due to the large differences among our three habitat classes. Accuracy may not have been conserved so well in areas of more closely related forest.

Both Analytical Indices and AudioSet Fingerprint had similar changes in variance as a result of recording length. Transient vocalisers are therefore likely somewhat important in the determination of the AudioSet Fingerprint and a mixed level of importance in some Analytical Indices. The ACI index was not impacted by recording length despite specifically quantifying how the soundscape changes over time (Pieretti, Farina and Morri, 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 *et al.* , 2008; Villanueva-Rivera *et al.* , 2011), indicating that transient calls akin to short term anomalies in frequency are perhaps lost in when recording windows are altered.

Finally, we found that subsetting audio data temporally and analysing them separately had an unpredictable accuracy on the classification task, with 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. It is recommended that > 120 h of recordings are required for Analytical Indices to stabilise (Bradfer-Lawrence *et al.* , 2019), yet in our study, we had just $70 - 75$ h of recordings per site. Overall we found that compression, frame size and temporal subsetting caused a small decrease 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 hours of data was able to, on average, outperform the Analytical Indices classifier trained on the full 24h.

Recommendations and Conclusion

Based on the results of this study we provide the following four recommendations:

1. When classifying Soundscapes, use AudioSet Fingerprinting rather than Analytical Indices.
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.
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.

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 acoustic index values. 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 non-compressed 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. Such a task can now be cautiously approached using meta-analysis of non-uniform acoustic data, while a simultaneous trajectory towards more standardised practices will enable more rigorous analyses in the future.

Acknowledgements

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.

Author's Contributions

B.E.H., L.P., S.S.S, R.M.E., and C.D.L.O. contributed to the conceptualization and implementation of the study. B.E.H. and R.M.E. led fieldwork and data collection. B.E.H., S.S.S., L.P., and C.D.L.O, designed and ran index extraction pipeline and data analysis. B.E.H and C.D.L.O developed the statistics and figures for the main text and supplementary. B.E.H led the manuscript writing process aided with revisions provided by all authors.

Data Accessibility

Acoustic Data: Will be made available on Zenodo and accessible via permenant DOI

Analytical Indices/ AudioSet Fingerprint Data: Will be available on the SAFE project website, and accessible via permenant DOI.

Analysis Scripts: Available on Github at https://github.com/BeckyHeath/Experimental-Variation-Ecoacoustics-Analysis-Scripts (made public after publication)

References

Araya-Salas, M., Smith-Vidaurre, G. and Webster, M. (2019) 'Assessing the effect of sound file compression and background noise on measures of acoustic signal structure', *Bioacoustics* . Taylor & Francis, 28(1), pp. 57–73. doi: 10.1080/09524622.2017.1396498.

Boelman, N. T. *et al.* (2007) 'Multi-trophic invasion resistance in Hawaii: Bioacoustics, field surveys, and airborne remote sensing',*Ecological Applications* , 17(8), pp. 2137–2144. doi: 10.1890/07-0004.1.

Bohnenstiehl, D. R. *et al.* (2018) 'Investigating the utility of ecoacoustic metrics in marine soundscapes', *Journal of Ecoacoustics* . doi: 10.22261/jea.r1156l.

Bradfer-Lawrence, T. *et al.* (2019) 'Guidelines for the use of acoustic indices in environmental research', *Methods in Ecology and Evolution* , 10(10), pp. 1796–1807. doi: 10.1111/2041-210x.13254.

Browning, E. *et al.* (2017) 'Passive acoustic monitoring in ecology and conservation', *WWF Conservation Technology Series 1* , 2(October), pp. 1–75. doi: 10.13140/RG.2.2.18158.46409.

Buxton, R. T. *et al.* (2018) 'Efficacy of extracting indices from large-scale acoustic recordings to monitor biodiversity',*Conservation Biology* , 32(5), pp. 1174–1184. doi: 10.1111/cobi.13119.

Costello, M. J. *et al.* (2016) 'Field work ethics in biological research', *Biological Conservation* . Elsevier Ltd, 203, pp. 268–271. doi: 10.1016/j.biocon.2016.10.008.

Cribari-Neto, F. and Zeileis, A. (2010) 'Beta regression in R',*Journal of Statistical Software* , 34(2), pp. 1–24. doi: 10.18637/jss.v034.i02.

Douma, J. C. and 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), pp. 1412–1430. doi: 10.1111/2041-210X.13234.

Eldridge, A. *et al.* (2018) 'Sounding out ecoacoustic metrics: Avian species richness is predicted by acoustic indices in temperate but not tropical habitats', *Ecological Indicators* . Elsevier B.V., 95, pp. 939–952. doi: 10.1016/j.ecolind.2018.06.012.

Ewers, R. M. *et al.* (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), pp. 3292–3302. doi: 10.1098/rstb.2011.0049.

Fitzpatrick, M. C. *et al.* (2009) 'Observer bias and the detection of low-density populations', *Ecological Applications* , 19(7), pp. 1673–1679.

Fuller, S. *et al.* (2015) 'Connecting soundscape to landscape: Which acoustic index best describes landscape configuration?',*Ecological Indicators* . Elsevier Ltd, 58, pp. 207–215. doi: 10.1016/j.ecolind.2015.05.057.

Gemmeke, J. F. *et al.* (2017) 'Audio Set: An ontology and human-labeled dataset for audio events', *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* , pp. 776–780. doi: 10.1109/ICASSP.2017.7952261.

Gómez, W. E., Isaza, C. V. and Daza, J. M. (2018) 'Identifying disturbed habitats: A new method from acoustic indices', *Ecological Informatics* , 45. doi: 10.1016/j.ecoinf.2018.03.001.

Hershey, S. *et al.* (2017) 'CNN architectures for large-scale audio classification', *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* , pp. 131–135. doi: 10.1109/ICASSP.2017.7952132.

Huff, M. H. *et al.* (2000) 'A habitat-based point-count protocol for terrestrial birds, emphasizing Washington and Oregon', *General Technical Reports of the US Department of Agriculture, Forest Service* , (PNW-GTR-501), pp. 2–30.

Kasten, E. P. *et al.* (2012) 'The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology', *Ecological Informatics* . Elsevier B.V., 12, pp. 50–67. doi: 10.1016/j.ecoinf.2012.08.001.

Linke, S. and Deretic, J. A. (2020) 'Ecoacoustics can detect ecosystem responses to environmental water allocations', *Freshwater Biology* , 65(1), pp. 133–141. doi: 10.1111/fwb.13249.

Mammides, C. *et al.* (2017) 'Do acoustic indices correlate with bird diversity? Insights from two biodiverse regions in Yunnan Province, south China', *Ecological Indicators* . Elsevier, 82(March), pp. 470–477. doi: 10.1016/j.ecolind.2017.07.017.

Medina-García, A., Araya-Salas, M. and Wright, T. F. (2015) 'Does vocal learning accelerate acoustic diversification? Evolution of contact calls in Neotropical parrots', *Journal of Evolutionary Biology* , 28(10), pp. 1782–1792. doi: 10.1111/jeb.12694.

Pfeifer, M. *et al.* (2016) 'Mapping the structure of Borneo's tropical forests across a degradation gradient', *Remote Sensing of Environment* , (176), pp. 84–97.

Pieretti, N., Farina, A. and Morri, D. (2011) 'A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI)', *Ecological Indicators* . Elsevier Ltd, 11(3), pp. 868–873. doi: 10.1016/j.ecolind.2010.11.005.

Rapport, D. J. (1989) 'What constitutes ecosystem health?',*Perspectives in Biology and Medicine* , 33(1), pp. 120–132. doi: 10.1353/pbm.1990.0004.

Rapport, D. J., Costanza, R. and McMichael, A. J. (1998) 'Assessing ecosystem health', *Trends in Ecology and Evolution* , 13(10), pp. 397–402. doi: 10.1016/S0169-5347(98)01449-9.

Rempel, R. S. *et al.* (2005) 'Bioacoustic monitoring of forest songbirds: interpreter variability and effects of configuration and digital processing methods in the laboratory', *Journal of Field Ornithology* , 76(1), pp. 1–11. doi: 10.1648/0273-8570-76.1.1.

Roca, I. T. and Proulx, R. (2016) 'Acoustic assessment of species richness and assembly rules in ensiferan communities from temperate ecosystems', *Ecology* , 97(1), pp. 116–123. doi: 10.1890/15-0290.1.

Saito, K. *et al.* (2015) 'Utilizing the Cyberforest live sound system with social media to remotely conduct woodland bird censuses in Central Japan', *Ambio* . Springer Netherlands, 44(4), pp. 572–583. doi: 10.1007/s13280-015-0708-y.

Sethi, S. S. *et al.* (2018) 'Robust, real-time and autonomous monitoring of ecosystems with an open, low-cost, networked device',*Methods in Ecology and Evolution* , 2018(December 2017), pp. 1–5. doi: 10.1111/2041-210X.13089.

Sethi, S. S. *et al.* (2020) 'Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set', (24), pp. 1–7. doi: 10.1073/pnas.2004702117.

Struebig, M. J. *et al.* (2013) *Quantifying the Biodiversity Value of Repeatedly Logged Rainforests. Gradient and Comparative Approaches from Borneo* . 1st edn, *Advances in Ecological Research* . 1st edn. Elsevier Ltd. doi: 10.1016/B978-0-12-417199-2.00003-3.

Sueur, J. *et al.* (2008) 'Rapid acoustic survey for biodiversity appraisal', *PLoS ONE* , 3(12). doi: 10.1371/journal.pone.0004065.

Sueur, J. *et al.* (2014) 'Acoustic indices for biodiversity assessment and landscape investigation', *Acta Acustica united with Acustica* , 100(4), pp. 772–781. doi: 10.3813/AAA.918757.

Sueur, J., Aubin, T. and Simonis, C. (2008) 'Equipment review: Seewave, a free modular tool for sound analysis and synthesis',*Bioacoustics* , 18(2), pp. 213–226. doi: 10.1080/09524622.2008.9753600.

Sueur, J., Krause, B. and Farina, A. (2019) 'Climate Change Is Breaking Earth's Beat', *Trends in Ecology & Evolution* , 34(11), pp. 971–973. doi: 10.1016/j.tree.2019.07.014.

Sugai, L. S. M. *et al.* (2019) 'A roadmap for survey designs in terrestrial acoustic monitoring', *Remote Sensing in Ecology and Conservation* , pp. 1–16. doi: 10.1002/rse2.131.

Towsey, M. *et al.* (2014) 'The use of acoustic indices to determine avian species richness in audio-recordings of the environment', *Ecological Informatics* . Elsevier B.V., 21(100), pp. 110–119. doi: 10.1016/j.ecoinf.2013.11.007.

Towsey, M. (2018) 'The calculation of acoustic indices derived from long-duration recordings of the natural environment', 0(August 2017), pp. 0–12.

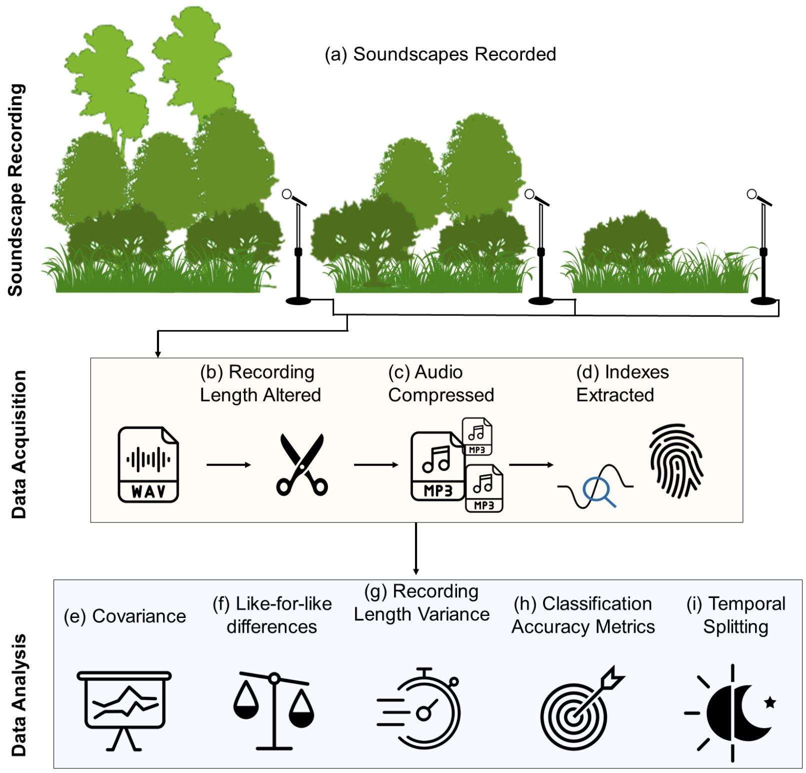
Towsey, M. W., Truskinger, A. M. and Roe, P. (2016) 'The Navigation and Visualisation of Environmental Audio Using Zooming Spectrograms',*Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015* , pp. 788–797. doi: 10.1109/ICDMW.2015.118.

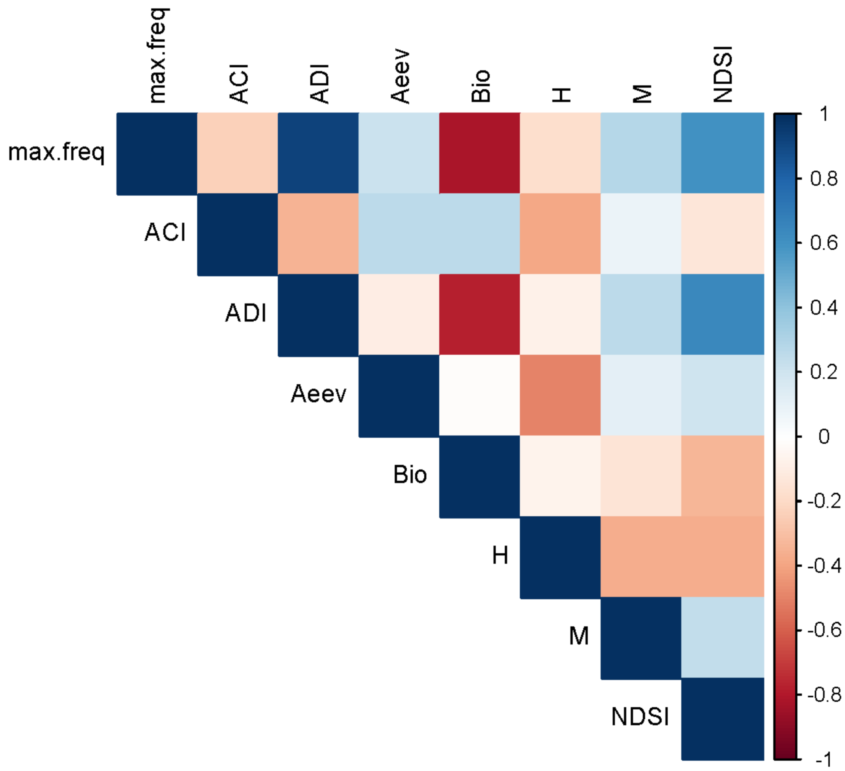
Vesna, I. (2009) 'Understanding Bland Altman Analysis', *Biochemia Medica* , 19(1), pp. 10–16. doi: 10.11613/BM.2013.003.

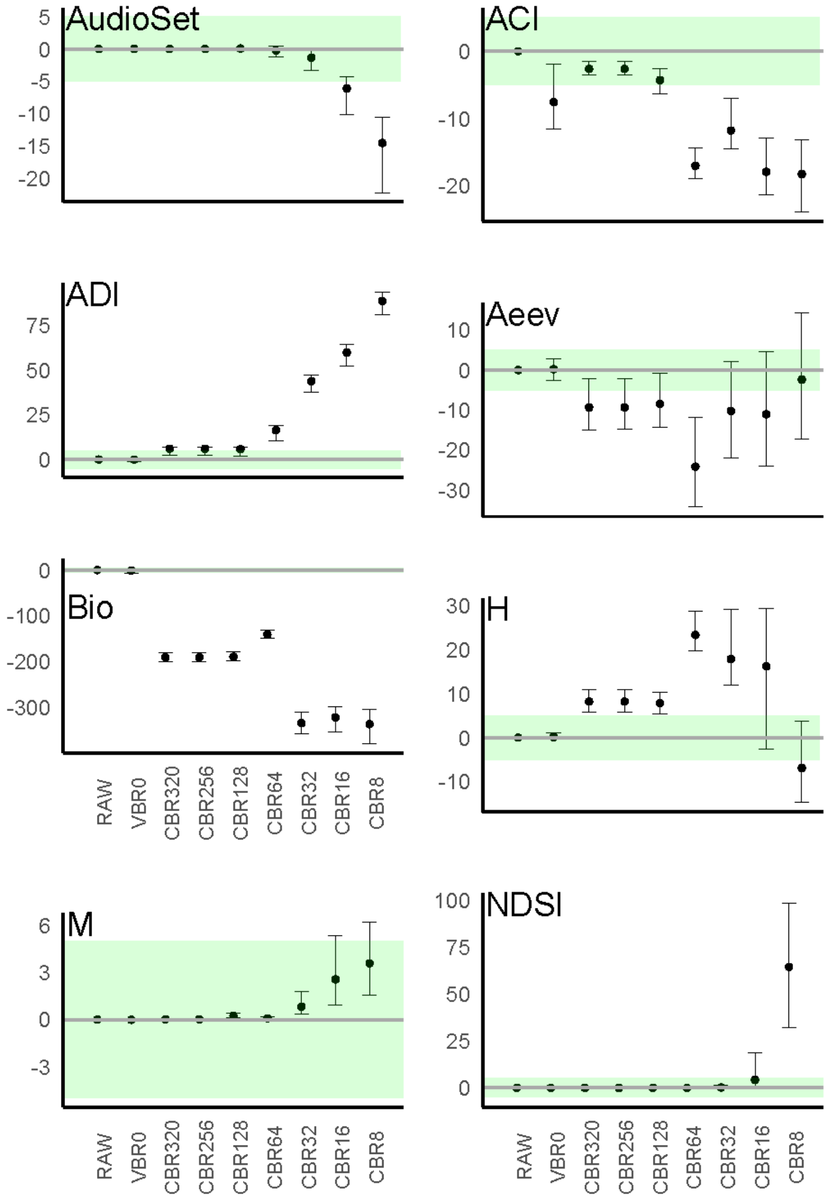
Villanueva-Rivera, L. J. *et al.* (2011) 'A primer of acoustic analysis for landscape ecologists', *Landscape Ecology* , 26(9), pp. 1233–1246. doi: 10.1007/s10980-011-9636-9.

Villanueva-Rivera, L. J. and Pijanowski, B. C. (2016) 'Package "soundecology" ', *http://ljvillanueva.github.io/soundecology/, CRAN* , (Package version 1.3.3), p. 14. Available at: http://ljvillanueva.github.io/soundecology/.

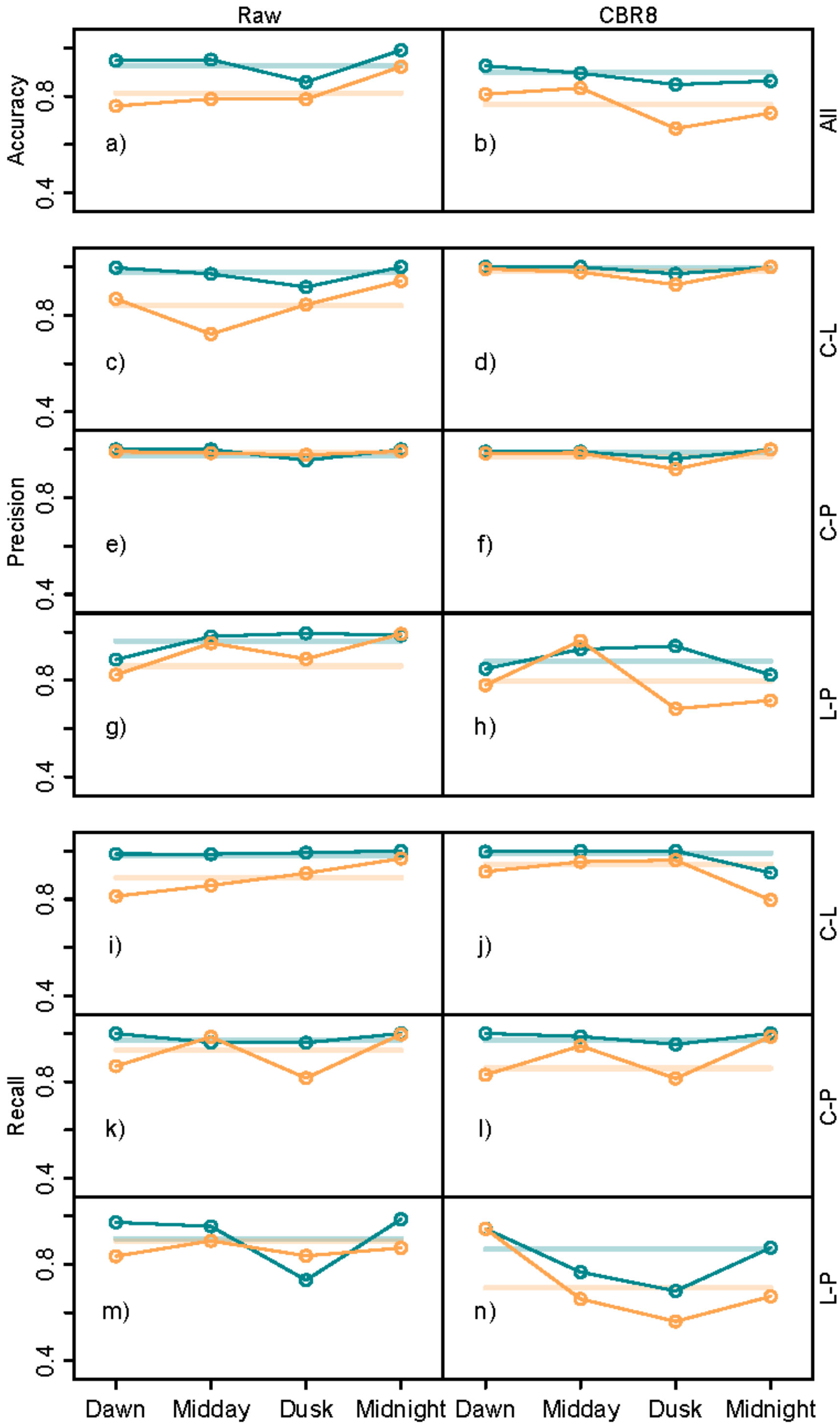
Zhang, L. *et al.* (2016) 'Classifying and ranking audio clips to support bird species richness surveys', *Ecological Informatics* . Elsevier B.V., 34, pp. 108–116. doi: 10.1016/j.ecoinf.2016.05.005.

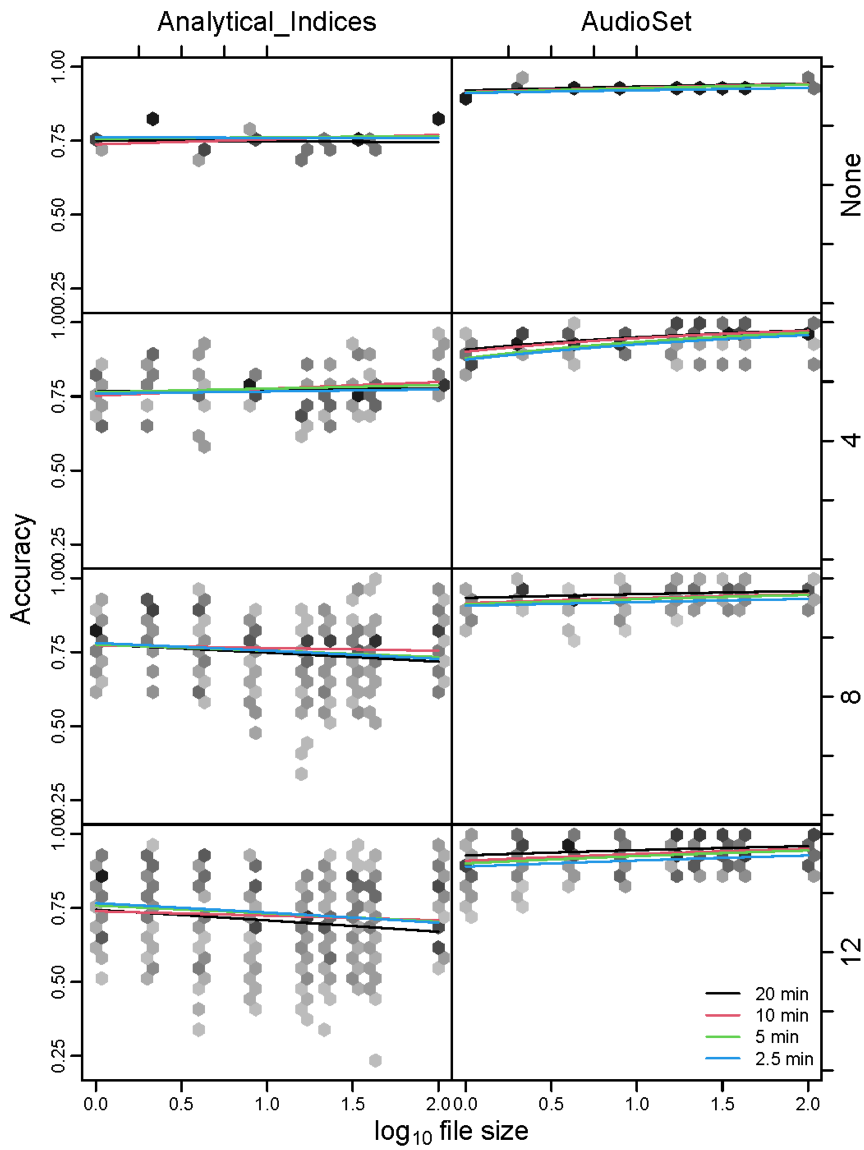

Soundscape Recording
(a) Soundscapes Recorded
Data Acquisition
(b) Recording Length Altered
(c) Audio Compressed
(d) Indexes Extracted
WAV
MP3
Data Analysis
(e) Covariance
(f) Like-for-like differences
(g) Recording Length Variance
(h) Classification Accuracy Metrics
(i) Temporal Splitting


max.freq
ACI
ADI
Aeev
Bio
H
M
NDSI
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1



AudioSet
ACI
ADI
Aeev
Bio
H
M
NDSI
RAW
VBR0
CBR320
CBR256
CBR128
CBR64
CBR32
CBR16
CBR8


Raw
CBR8
Accuracy
Precision
Recall
All
C-L
C-P
L-P
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
m)
n)
0.4
0.8
Dawn
Midday
Dusk
Midnight


Analytical_Indices
AudioSet
None
4
8
12
Accuracy
0.25
0.50
0.75
1.00
0.0
0.5
1.0
1.5
2.0
log10 file size
20 min
10 min
5 min
2.5 min