OIKOS

Research

Soundscapes predict species occurrence in tropical forests

Sarab S. Sethi, Robert M. Ewers, Nick S. Jones, Jani Sleutel, Adi Shabrani, Nursyamin Zulkifli and Lorenzo Picinali

S. S. Sethi (https://orcid.org/0000-0002-5939-0432) ✉ (s.sethi16@imperial.ac.uk), Norwegian Inst. for Nature Research, Trondheim, Norway. – SSS and N. S. Jones (https://orcid.org/0000-0002-4083-972X), Dept of Mathematics, Imperial College London, London, UK. – R. M. Ewers, Dept of Life Sciences, Imperial College London, Ascot, UK. – J. Sleutel and N. Zulkifli, Southeast Asia Rainforest Research Partnership, Lahad Datu, Malaysia. JS also at: Dept of Biology, Vrije Univ. Brussel, Brussels, Belgium. – A. Shabrani, WWF-Malaysia, Sabah Office, Kota Kinabalu, Malaysia. – L. Picinali (https://orcid.org/0000-0001-9297-2613), Dyson School of Design Engineering, Imperial College London, London, UK.

Oikos
00: 1–9, 2021
doi: 10.1111/oik.08525

Subject Editor: Silke Bauer
Editor-in-Chief: Dries Bonte
Accepted 4 November 2021

Accurate occurrence data is necessary for the conservation of keystone or endangered species, but acquiring it is usually slow, laborious and costly. Automated acoustic monitoring offers a scalable alternative to manual surveys but identifying species vocalisations requires large manually annotated training datasets, and is not always possible (e.g. for lesser studied or silent species). A new approach is needed that rapidly predicts species occurrence using smaller and more coarsely labelled audio datasets. We investigated whether local soundscapes could be used to infer the presence of 32 avifaunal and seven herpetofaunal species in 20 min recordings across a tropical forest degradation gradient in Sabah, Malaysia. Using acoustic features derived from a convolutional neural network (CNN), we characterised species indicative soundscapes by training our models on a temporally coarse labelled point-count dataset. Soundscapes successfully predicted the occurrence of 34 out of the 39 species across the two taxonomic groups, with area under the curve (AUC) metrics from 0.53 up to 0.87. The highest accuracies were achieved for species with strong temporal occurrence patterns. Soundscapes were a better predictor of species occurrence than above-ground carbon density – a metric often used to quantify habitat quality across forest degradation gradients. Our results demonstrate that soundscapes can be used to efficiently predict the occurrence of a wide variety of species and provide a new direction for data driven large-scale assessments of habitat suitability.

Keywords: bioacoustics, machine learning, soundscape, species occurrence, tropical forest

Introduction

Ecosystems are being subjected to increasing external pressures from land-use change and global warming (Walther et al. 2002, Lambin and Meyfroidt 2011). These pressures have resulted in global biodiversity declines, as the natural habitats required to support many species shrink and disappear (Newbold et al. 2015). Efforts to slow this decline often aim to protect areas of high conservation value that may support

NSO
NORDIC SOCIETY OIKOS

www.oikosjournal.org

populations of endangered or keystone species (Mills et al. 1993). This leads to the key question: how can we identify such locations rapidly, accurately and on a large scale?

An established solution is to carry out manual surveys of the region of interest (Brown et al. 2013). Common approaches include actively searching for species of interest, deploying traps to capture them or looking for features that may indicate their presence (e.g. nests). However, manual surveys are expensive, labour intensive, and do not scale well temporally or spatially (Gijzen 2013). In contrast, automated passive acoustic monitoring has shown promise as a route to gaining scalable insight into ecological systems (Gibb et al. 2019). Audio data can be recorded and analysed inexpensively, in real-time and over extended durations, making it an increasingly powerful tool for ecologists and conservationists (Pijanowski et al. 2011, Sueur and Farina 2015).

Species occurrence data can be extracted from audio recordings automatically by detecting vocalisations. Using a large training dataset of annotated examples a machine learning model can be trained to identify calls made by a target species (Aide et al. 2013, Stowell et al. 2016, Wrege et al. 2017). For example, BirdNET leveraged citizen science to train a model on over 226 000 high quality recordings of vocal species found in historically well-studied regions of the world (Kahl et al. 2021). Detecting vocalisations, however, relies upon three key assumptions; 1) the species has at least one unique vocalisation, 2) the species is active and audible during the recording and 3) there exists a large, labelled dataset of the species' vocalisations (or the resources to collate such training data from scratch). These barriers are particularly difficult to overcome when searching for lesser studied species in highly biodiverse and noisy environments such as tropical forests (Stowell et al. 2019, Gibb et al. 2019), or for species that are largely silent. Novel approaches are required that can utilise acoustic monitoring data in complementary ways to survey a broader array of species and in a more data efficient manner.

Analysing soundscapes in their entirety provides an alternate route to the automated analysis of eco-acoustic data (Pijanowski et al. 2011). In this approach, features of the audio signal are used to infer habitat quality, without the need for species specific training data (Sueur et al. 2008, Pieretti et al. 2011, Sethi et al. 2020b). Whilst soundscape features have been shown to correlate with summary metrics of biodiversity (e.g. alpha or beta diversity of a whole community), they are not normally used to provide direct evidence for how suitable a habitat is for a single given species.

In this study we demonstrate that an environment's overall soundscape fingerprint can be used as a powerful indicator of species occurrence. We built upon prior work looking at community level structure in soundscapes, and employed acoustic features derived from a convolutional neural network (Sethi et al. 2020). Using these learned high-dimensional features allowed us to characterise soundscapes in finer detail than would be possible with algorithmically derived traditional soundscape indices. Rather than focussing on species-specific vocalisations, our model learned acoustic features which indicated species presence using only coarsely-labelled point count data from across a gradient of tropical forest degradation in Sabah, Malaysia. By performing a fully cross-validated classification task, we were able to predict occurrence accurately for a number of avifaunal and herpetofaunal species without the need for large, precisely annotated training datasets. Additionally, we showed that soundscapes are a more accurate predictor of species occurrence than above-ground carbon density – the standing density of live or dead woody matter – a metric often used to quantify habitat quality across tropical forest degradation (Jucker et al. 2018). Our findings indicate a promising new route for audio data to be used for the conservation of species on a large scale, and across a wide range of taxa, without many of the limitations of vocalisation detection-based approaches.

Material and methods

Study location and estimates of habitat quality

This work was undertaken at the Stability of Altered Forest Ecosystems (SAFE) Project in Sabah, Malaysia (Ewers et al. 2011) between March 2018 and February 2020. The month of sampling wasn't controlled for since the region shows no clear seasonality and there were no unusual climactic events (e.g. El Niños) during the period covered by our study (Walsh and Newbery 1999). We surveyed eleven sites across a land-use intensity gradient (Fig. 1): two sites in oil palm plantations, two sites in salvage logged forest (last logged in the early 2010s), five sites in selectively twice-logged forest (logged in the 1970s and early 2000s), and two sites in forest inside a protected area (where small amounts of illegal logging activity had occurred). The minimum distance between sites was 583 m (mean pairwise separation = 7.6 km), and as such the soundscapes could be considered as independent.

In November 2014 airborne LiDAR data was acquired of the SAFE landscape (Jucker et al. 2018). This flight followed the most recent round of logging, and therefore the measured canopy structure would not have changed significantly by 2018–2020 (when our point counts were performed). The raw LiDAR point cloud was used to produce a pitfree canopy height model at 1 m resolution. Above-ground carbon density (ACD) was calculated at 1 ha resolution using top of canopy height and gap fraction (Swinfield et al. 2020). ACD is closely correlated with above-ground biomass (AGB) which has been used as a metric of forest intactness at SAFE Project regularly, as decreases in AGB are primarily driven by historical and current anthropogenic pressures (Brant et al. 2016, Luke et al. 2017, Riutta et al. 2018, Williamson et al. 2021). We averaged ACD values within a 100 m radius of each of our sampling sites (mean samples per site = 3, range = 2–4), for use as a quantitative measure of habitat quality. Integrating habitat quality over a larger area provides a more appropriate metric than a single point estimate as most of the species surveyed move and interact with the environment

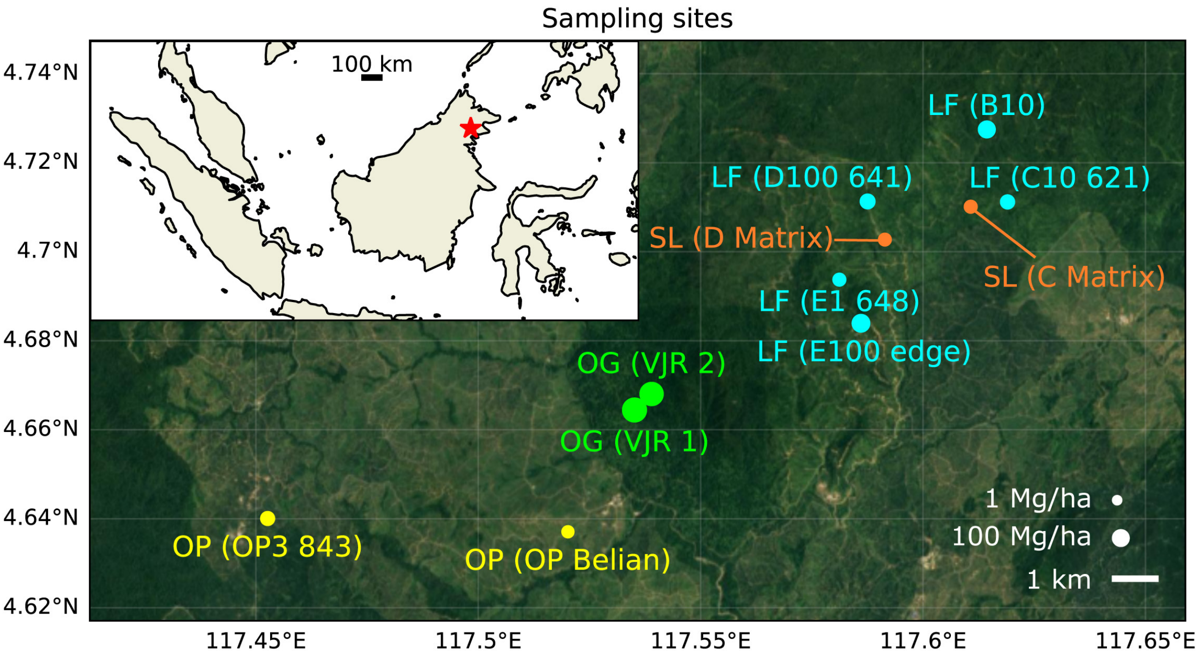

Figure 1. Map of the study location. We collected data from 11 sites at the Stability of Altered Forest Ecosystems (SAFE) experiment in Sabah, Malaysian Borneo. Sites were distributed across old-growth forest (OG), logged forest (LF), salvage logged forest (SL) and oil palm plantations (OP) to capture a gradient of above-ground carbon density – as denoted by size of markers (range 1.8–110 Mg ha^{-1}).

beyond their immediate surroundings. We also tested the effect of averaging ACD over 250 m and 500 m radiuses from each site.

Avifaunal and herpetofaunal point counts

Across the 11 sampling sites, we carried out 790 avifaunal and 771 herpetofaunal point counts (of which 483 were undertaken simultaneously). Each point count lasted 20 min and surveys were distributed evenly throughout the 24 hrs of the day (Supporting information). There was a mean of 2.99 point counts per hour per site for avifaunal species ($\sigma = 1.17$), and 2.92 point counts per hour per site for herpetofaunal species ($\sigma = 1.13$). During point counts, we recorded all visual or aural encounters of avifaunal or herpetofaunal species within a 10 m radius of the sampling site. Species were cross-referenced with the Global Biodiversity Information Facility (GBIF) backbone taxonomy to validate taxonomic classifications (GBIF Secretariat 2020).

Occurrence data (presence/absence) was thus acquired for 175 avifaunal and 53 herpetofaunal species. To ensure we had sufficient data to train our models, species present in fewer than 50 point counts were removed from the dataset. With a reduced threshold we expect to have reduced accuracy, but species of conservation concern tend to have low abundance and represent an important use case. We therefore set a lower threshold of 15 point counts for a separate group of species classified as vulnerable or critically endangered by the IUCN Red List (Baillie et al. 2004). In total this gave us a set of 32 avifaunal and seven herpetofaunal species (Supporting information). Five of the 32 avifaunal species were listed as vulnerable or critically endangered, but none of the seven herpetofaunal species were.

Audio data and acoustic feature extraction

During each point count a simultaneous 20-min audio recording was made using a Tascam DR-05 recorder mounted at chest height (nominal input level −20 dBV, frequency range 20 Hz–22 kHz). Raw audio data was recorded to a single channel at 44.1 kHz in the WAV format.

We calculated learned acoustic features of the audio using a pretrained convolutional neural network (CNN), 'VGGish', developed by Hershey et al. (2017). The CNN was trained to perform a general-purpose audio classification task using an extremely large annotated dataset (Gemmeke et al. 2017), resulting in a general 128-dimensional acoustic feature embedding. For full details on the CNN-based feature extraction process and other applications of these features in an ecological context (Sethi et al. 2020b).

The VGGish CNN takes a 16 kHz log-scaled Mel-frequency spectrogram as an input – as used by Hershey et al. when training the network. First, a spectrogram is computed using the magnitudes of the short-time Fourier transform with a window size of 25 ms, a window hop of 10 ms and a periodic Hann window. The frequencies of the spectrogram are mapped to 64 mel-frequency bins covering the range 125–7500 Hz, and the magnitude values are offset by 0.01 before taking their logarithm. From the resulting time-frequency representation of the audio (of dimensions 96 × 64), the CNN outputs one 128-dimensional feature vector per 0.96 s of audio.

Since our raw audio data was recorded at a higher sample rate, we pre-processed it (as recommended by the VGGish documentation) by down-sampling to 16 kHz using a Kaiser window filter to avoid aliasing. During the analysis we also investigated how averaging consecutive CNN-derived acoustic features over the following longer time periods affected our results: 1.92, 2.88, 3.84, 4.80, 5.76, 6.72, 7.68, 8.64, 9.60, 29.76, 59.52 and 299.52 s.

Predictions of species occurrence

For each species we split point counts in the training dataset into two groups; one where the target species was present (pres) and the other where it was absent (abs). We fit a Dirichlet-process Gaussian mixture model (DP-GMM) to acoustic features from each group to obtain the probability density functions (PDFs) p_{pres} and p_{abs} (Blei and Jordan 2006), using an upper bound of 500 components and diagonal covariance matrices. Other hyperparameters were left as default using the scikit-learn *BayesianGaussianMixture* implementation.

The PDFs p_{pres} and p_{abs} allow us to identify which regions of acoustic feature space (and hence, which sounds) are associated with the species being present or absent in a given recording. Intuitively, if an audio feature is within a high probability region of p_{pres} and a low probability of p_{abs}, this indicates that the species is likely to be present (and vice versa). A formal description of how we derive classification scores from these PDFs is given below.

For each 20-min audio recording, we first split the audio into N non-overlapping 0.96 s segments. We defined the set S of CNN-derived acoustic feature vectors corresponding to each segment as, $S=\{X_1, X_2, \dots X_N\}$. To calculate features representing longer timescales than 0.96 s, we averaged consecutive members of S using non-overlapping windows. For each feature X_i we calculated a likelihood ratio, $L_i=\log(p_{\text{pres}}(X_i)) - \log(p_{\text{abs}}(X_i))$, allowing us to define a new set, $S_L=\{L_1, L_2, \dots L_N\}$. To obtain an overall classification confidence indicating the probability of the species being present in the full 20-min recording, we looked at four properties of S_L; 1) $\lambda_1=\max(S_L)$, 2) $\lambda_2=\min(S_L)$, 3) $\lambda_3=\text{mean}(S_L)$ and 4) $\lambda_4=\text{P}_{\%}(S_L)$ (for percentiles 10, 20, 30, 40, 50, 60, 70, 80 and 90). We found that the 60th percentile metric, $\lambda_4=\text{P}_{60}(S_L)$, provided the most accurate predictions, and therefore report results only for this definition of classification confidence (Supporting information). Henceforth λ will be used to refer to λ_4.

To assess the extent to which soundscapes can predict species occurrence we performed an eleven-fold cross-validation classification task for each species. There were only two sites for three of the land-use categories (OG, SL, OP). Therefore, to ensure a representative test we were unable to train on fewer and test on multiple sites at once. In each fold, data from ten sites were used as a training set (to fit p_{pres} and p_{abs}), and data from the remaining eleventh site was used as a test set to assess the model's accuracy. In this way we ensured that we did not report artificially high accuracies by overfitting to site specific soundscapes but learned generalisable acoustic characteristics that indicated species presence in previously unseen locations. We measured the ability of λ to classify a species as present in a point count using the area under the receiver operating characteristic curve (AUC) metric, which allowed us to avoid setting an arbitrary decision threshold on λ. An AUC of 0.5 represents chance predictions of occurrence, and an AUC equal to 1 is the case where perfect predictions are made. Mean AUC was calculated for each species across all 11 folds.

For each species we generated null distributions of AUC values (those that should be expected if there was no link between soundscape features and species occurrence) to calculate statistical significance of predictions. We used acoustic features at the 2.88 s timescale, as these features maximised mean AUC across all species (Supporting information). We randomly shuffled classification confidence scores (λ) 100 times within each of the 11 folds, and measured AUC using the unshuffled occurrence labels. 100 null mean AUC values were obtained by averaging across the 11 folds, and we used a threshold of $p \leq 0.05$ to determine statistical significance.

We performed a similar eleven-fold cross-validation classification task using above-ground carbon density (ACD) data, to compare the predictive power of soundscapes versus ACD, a more traditional metric of habitat quality. In each fold, we identified the site in the training set with ACD most similar (least difference) to the site in the test set. Then, to predict species occurrence in each 20-min point count, we used the mean species occurrence from point counts at the same time of day from the previously identified similar site.

Analysis of performance across species

To quantify how temporally structured occurrence patterns were for each species, we formulated a contingency table from the ground truth point count data with species occurrence as one variable (averaged if multiple point counts were performed at the same hour at any given site) and hour of day as the other. On this contingency table we calculated a χ^2 statistic. We then calculated Spearman's correlation coefficient, ρ, between the χ^2 statistic and AUC across all 39 species to test whether accuracy of our predictions was correlated with how temporally structured each species' occurrence patterns were. We also calculated Spearman's correlation coefficient between the total number of point counts in which each species was found and AUC to investigate whether rarity of species had an effect on accuracy of predictions. Finally, we calculated Spearman's correlation coefficient between the AUC of species occurrence predictions and ACD of the site at which these predictions were made. In all cases p-values were obtained analytically.

Results

Soundscapes are highly indicative of species occurrence

We were able to predict species occurrence from soundscape recordings for four of the seven non-threatened herpetofaunal species, all 27 non-threatened avifaunal species, and three of the five threatened avifaunal species ($p \leq 0.05$, Fig. 2a). Mean

AUC across all species was highest using features corresponding to 2.88 s segments of audio (Supporting information), although the most accurate classifications for a single species was found for the bold-striped tit-babbler *Macronus bornensis* when using 0.96 s per feature (0.87 AUC). Variation in AUC between species was larger than the variation for a given species across different timescales of features. Even with features averaged over almost five minutes, we were able to predict species occurrence from soundscapes with AUCs of up to 0.82 (sooty-capped babbler *Malacopteron affine*). Spectrograms (Supporting information) confirm the intuition that we did not learn to identify species vocalisations, but rather the model learned indicative characteristics of the soundscape that played out over longer timescales than any single species call (i.e. minutes rather than seconds). From herein we will only consider results using acoustic features at the optimal 2.88 s timescale.

Performance of soundscape based predictions was worse for the five Red List threatened avifaunal species compared to the other 27 species (T-test on AUCs; $p < 0.001$). Nevertheless, occurrence was still predicted with accuracies better than chance ($p \leq 0.05$) for three threatened avifaunal species; the black hornbill *Anthracoceros malayanus* (0.69 AUC, n = 15), the rhinoceros hornbill *Buceros rhinoceros* (0.69 AUC, n = 34) and the short-toed coucal *Centropus rectunguis* (0.75 AUC, n = 23). Both across all species and within each of the three groups of avifaunal, herpetofaunal and threatened avifaunal species, we found AUC was not significantly correlated with total number of encounters (Spearman correlation; $p > 0.05$, Supporting information).

We found that higher AUCs were attained when species were consistently encountered at the same hours of the day (Fig. 2b, Spearman correlation; $\rho = 0.64$, $p < 0.001$). Non-threatened avifaunal species had more temporally structured occurrence patterns than non-threatened herpetofaunal species (T-test on χ^2 statistics; $p = 0.04$), explaining the difference in AUCs between the taxonomic groups (T-test on AUCs; $p < 0.001$). Nevertheless, AUCs for four of the seven herpetofaunal species were still better than would be expected by chance, and reached up to 0.86 for the tree hole frog *Metaphrynella sundana* – possibly due to its unusually high χ^2 statistic when compared to the other herpetofaunal species (Fig. 2).

There was a close relationship between predicted occurrence from soundscape data and the pattern of true occurrence across habitat types and time of day (Fig. 3, Supporting information shows similar visualisations for all 39 species). We found that soundscape classification confidence was higher at the true times at which a species would be present, whether the species was diurnal (Fig. 3a, yellow-vented bulbul *Pycnonotus goiavier*), nocturnal (Fig. 3c, tree hole frog) or found only during very specific hours (Fig. 3b, sooty-capped babbler). We also found that soundscape predictions reflected true observations of species habitat niches. For example, the sooty-capped babbler (Fig. 3b) and tree hole frog (Fig. 3c) were commonly found in forest habitats – either logged or inside protected areas – whereas the yellow-vented bulbul was found more often in heavily disturbed habitats (salvage logged forest and oil palm). In all three cases, classification confidence derived from soundscape data reflected these habitat partitioning patterns. There was no significant relationship between the accuracy of occurrence predictions and the AGB of the site at which predictions were made (Supporting information).

Soundscapes predict occurrence more accurately than above-ground carbon density

We found that soundscape features predicted species occurrence more accurately than a comparison model based on

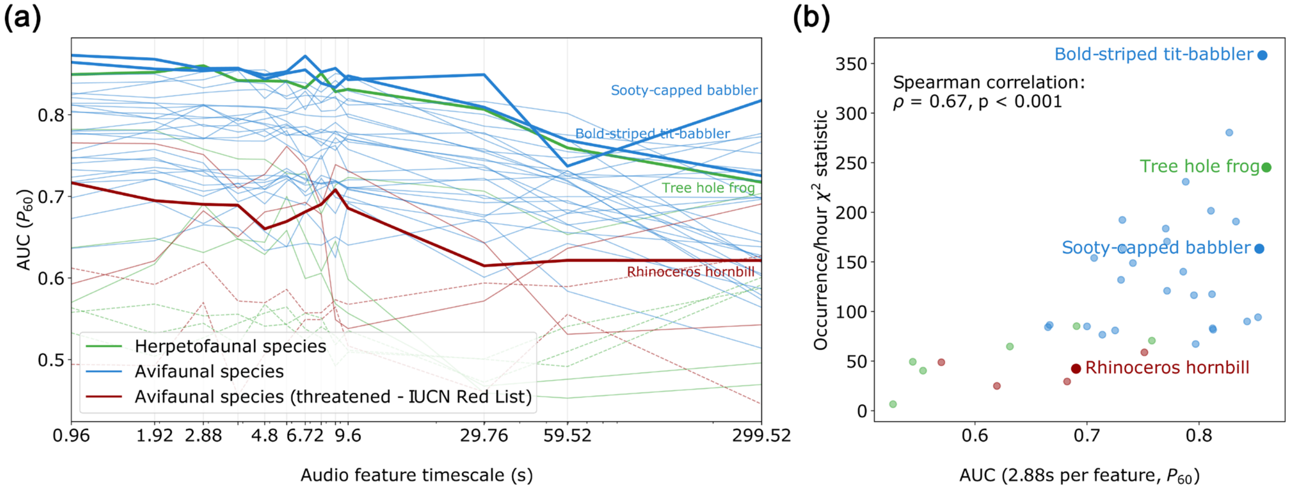

Figure 2. Soundscape features reliably predict species occurrence. We measured how predictive soundscapes were of species presence across 27 non-threatened avifaunal species (blue), five threatened avifaunal species (brown) and seven non-threatened herpetofaunal species (green). (a) We found soundscapes features across a wide range of timescales were predictive of species occurrence for 34 species (dotted lines indicate species for which $p > 0.05$) – using area under the receiver operating characteristic curve (AUC) as a metric of model accuracy. (b) The accuracy of occurrence predictions was significantly correlated with a χ^2 statistic measuring how correlated hour of day was with species occurrence ($p < 0.001$). In both panels we highlight results from four indicative taxa chosen to reflect the variety of species included in this study.

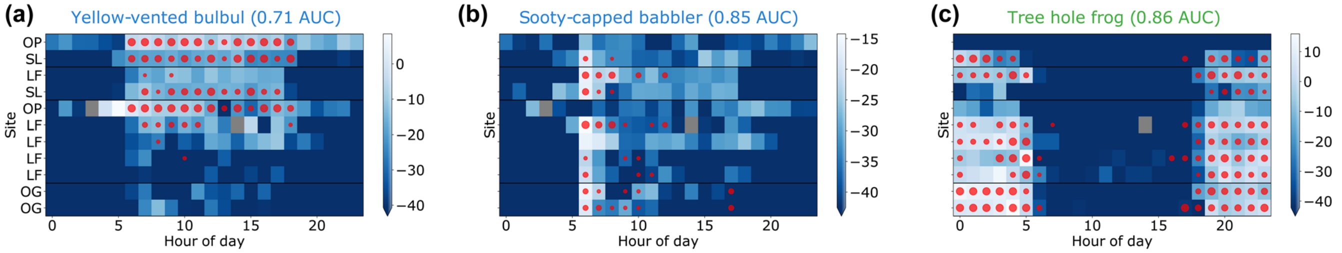

Figure 3. Soundscapes predict occurrence for species with varying habitat and temporal niches. Median classification scores, λ, from dark blue (low) to white (high) are shown for occurrence predictions from soundscapes for three species: (a) yellow-vented bulbul, (b) sooty-capped babbler and (c) tree hole frog. Overlaid in red is true occurrence data, where circle sizes indicate how often the species was found at that site and hour during the manual point counts. Sites are ordered by land-use type with low-quality habitats at the top and high-quality habitats at the bottom (the Supporting information provides the same visualisation for all 39 species).

above-ground carbon density (ACD) data, a metric often used as a proxy for tropical forest habitat quality (Fig. 4, paired T-test on AUCs; $p < 0.001$). The soundscape-based model produced increased AUCs for 34 of the 39 species surveyed, including for all five threatened avifaunal species. Mean accuracy of occurrence predictions for the non-threatened avifaunal group was increased by 0.08 AUC, for the threatened avifaunal group by 0.11 AUC and for the non-threatened herpetofaunal group by 0.04 AUC. This followed the trends noted earlier, as avifaunal species which exhibited strong temporal occurrence patterns benefited the most from the soundscape based approach. Per species there was a mean percentage increase in AUC of 12% across all 39 species surveyed. The figures reported are using ACD averaged over a 100 m radius from each sampling point, but soundscapes also performed better than ACD when averaging over a 250 m ($p < 0.001$) and 500 m ($p < 0.001$) radius from each point.

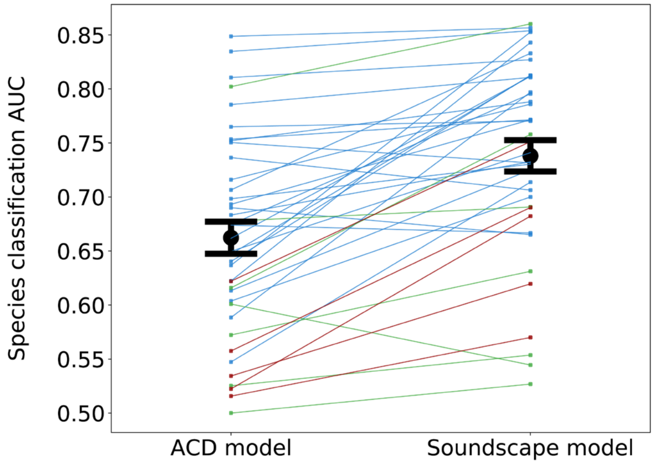

Figure 4. Soundscape features are a better indicator of species occurrence than above-ground carbon density (ACD). We compared occurrence predictions using soundscapes to a comparison model using ACD data. Lines connect AUC metrics for the same species, with threatened avifaunal species in brown, non-threatened avifaunal species in blue and non-threatened herpetofaunal species in green. In black is the mean and standard error for AUC across all 39 species for each model.

Discussion

We investigated whether soundscapes could indicate the occurrence patterns of 39 species across two taxonomic groups. Our results demonstrate this is indeed feasible, and that the most accurate indications could be obtained for species with strong temporal occurrence patterns. We found no significant correlation between rarity of species and accuracy of predictions and were even able to predict occurrence for the black hornbill *Anthracoceros malayanus* (0.68 AUC) with just 15 observations across 790 point counts. Performance was lower for species listed as vulnerable or endangered by the IUCN Red List, but these are not the only ones of conservation interest. Species which are particularly good indicators of habitat quality, those that have a disproportionate ecological impact on their environment, or those that fulfil important economic functions are often referred to as 'keystone species' (Mills et al. 1993). Whilst these species are sometimes also endangered, this is not always the case. For example, within the 'non-threatened' species, we had the rough guardian frog *Limnonectes finchi*, a species only ever found close to suitable water sources (Inger and Voris 1988). We also had the white crowned shama *Copsychus stricklandii* which due to their unique singing ability is threatened by a high rate of unsustainable trade in Southeast Asia (Leupen et al. 2018). We were able to predict occurrence for both species accurately with AUCs of 0.76 and 0.77, respectively. Furthermore, we found our approach worked equally well across the whole degradation gradient – demonstrating that soundscapes can be used to accurately predict species occurrence across a variety of habitat types typically found in and around tropical rainforests.

We found that our model was most accurate when using shorter timescale acoustic features – with features representing individual seconds rather than averaged over minutes of audio. This may simply be a matter of resolution – with longer timescale features the details of how soundscapes move between different modes are lost. The average of shorter features over these long time periods will therefore only provide a crude overview of the overall soundscape, leading to less accurate predictions of occurrence. Nonetheless, there was still significant predictive information contained within long timescale features, indicating that a coarse acoustic overview is often all that is required.

Our model learned to identify soundscape features that were uniquely found when the species of interest was present. We investigated a few point counts and found that high classification confidences did not correlate with vocalisations – in contrast with how audio data is typically used to infer species occurrence (Diwakar et al. 2007). Instead, soundscapes typical to the habitat type or time of day that the species was likely to be found led to predictions of species presence. We believe the model learned to pick up on these broad soundscape level fluctuations in acoustic features due to the relatively low number of vocalisations present in the dataset, together with the high overall temporal and spatial variability of soundscapes across all of our audio recordings. In many ways, this is analogous to considering the soundscape as an indirect metric of habitat suitability, much as landscape elevation models or predator prey networks have been used as indirect measures of habitat suitability (Store and Jokimäki 2003, Hirzel and Lay 2008). Whilst all species surveyed in this study produce vocalisations, basing predictions on soundscapes as a whole means that our approach may allow audio data to be used to predict the occurrence of completely silent species.

Equally tantalisingly, there is a possibility that with a less heterogenous, larger dataset a similar approach to ours may enable automated discovery of species vocalisations. Whilst it wasn't the case in our study, this situation would occur if the predominant distinguishing acoustic features between present and absent samples was the sound of the species vocalising. In this case, acoustic features with the highest classification confidences would correspond to the species' vocalisations. Automatically extracting vocalisations from passive recordings made in the wild may even allow us to discover calls and behaviours that cannot be reproduced with the same species in a more controlled environment.

Other types of data, beyond audio, can be used to predict species occurrence at a given place and time. Measuring the standing density or biomass of woody matter has been used extensively as a habitat quality indicator at the field site we surveyed (Brant et al. 2016, Luke et al. 2017, Riutta et al. 2018, Williamson et al. 2021). In this study, however, we showed that soundscapes were in fact better predictors of species occurrence for 31 of the 39 species surveyed than a simple model based on above-ground carbon density. Furthermore, acquiring high resolution airborne LiDAR data from planes or satellites (as used to derive ACD in this study) can be prohibitively expensive, is sampled infrequently, and is not a viable option in every field site (Lefsky et al. 2002, Popescu et al. 2011). By contrast, our audio recording protocol only involved using an inexpensive handheld recorder deployed to gather a 24 h acoustic record per site. Recordings of this type could be made rapidly and sampled regularly from a large number of sites, providing wide coverage with minimal capital outlay.

The link between habitat suitability and species occurrence data is clear – species are more likely to be found in habitats that are able to sustainably support their needs (Hirzel et al. 2006). Thus, by showing that occurrence for a wide range of species can be accurately predicted by soundscapes, this opens up a new avenue for assessing habitat suitability from audio data. One use-case may be in assisting the identification of areas of high conservation value within agricultural landscapes, as required by certification agencies such as the roundtable for sustainable palm oil (Brown et al. 2013). Additionally, as collaborative eco-acoustic datasets continue to grow (Baker et al. 2015), we may be able to harness soundscape data to produce large-scale habitat suitability maps, and identify those species that are most at risk from mounting global pressures (Walther et al. 2002).

Conclusion

In this study we have demonstrated that soundscapes can be used to predict species occurrence across a wide range of species in tropical forests. We found that the most accurate predictions could be made for species with strong temporal occurrence patterns, including for species of specific conservation concern, and that soundscape-based predictions outperformed those based on a more traditional metric of habitat quality, ACD. Future work could scale our approach to global datasets to produce models which infer species occurrence data with a higher accuracy and across diverse biomes. Our findings indicate a new route for audio data to be used as an impactful, scalable and widely applied conservation tool.

Acknowledgements – Thanks to Sahil Loomba and Henry Bernard for their contributions to the study, and to the staff at the SAFE Project and SEARRP for logistical support in data collection.
Funding – This project was supported by funding from WWF (Biome Health Project), Sime Darby Foundation (SAFE Project), NERC (NE/K007270/1 to NSJ), and EPSRC (EP/R511547/1 to SSS, EP/N014529/1 to NSJ). Data was collected from Malaysia under an SaBC permit granted to SSS (JKM/MBS.1000-2/2 JLD.8 (63)).

Author contributions

Sarab S. Sethi: Conceptualization (lead); Data curation (equal); Formal analysis (lead); Investigation (lead); Methodology (lead); Project administration (lead); Writing – original draft (lead); Writing – review and editing (equal). **Robert M. Ewers**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Supervision (equal); Writing – review and editing (equal). **Nick S. Jones**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Supervision (equal); Writing – review and editing (equal). **Jani Sleutel**: Data curation (equal); Investigation (equal); Writing – review and editing (equal). **Adi Shabrani**: Data curation (equal); Investigation (equal); Writing – review and editing (equal). **Nursyamin Zulkifli**: Data curation (equal); Investigation (equal); Writing – review and editing (equal). **Lorenzo Picinali**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (lead); Investigation (equal); Methodology (equal); Supervision (equal); Writing – review and editing (equal).

Data availability statement

Raw data from the point counts can be found at <https://zenodo.org/record/3997172#.YZfCS7rTVGw> (Sethi et al. 2019). Processed audio and field data are stored at <https://zenodo.org/record/4048019#.YZfCubrTVGx>, and code with instructions to reproduce results and figures is at <https://github.com/sarabsethi/sscape_spec_occ_preds_sethi2020>.

Supporting information

The supporting information associated with this article is available from the online version.

References

Aide, T. M. et al. 2013. Real-time bioacoustics monitoring and automated species identification. – PeerJ 1: e103.

Baillie, J. et al. 2004. 2004 IUCN Red List of threatened species: a global species assessment. – IUCN.

Baker, E. et al. 2015. BioAcoustica: a free and open repository and analysis platform for bioacoustics. – Database 2015: bav054.

Blei, D. M. and Jordan, M. I. 2006. Variational inference for Dirichlet process mixtures. – Bayesian Anal. 1: 121–143.

Brant, H. L. et al. 2016. Vertical stratification of adult mosquitoes (Diptera: Culicidae) within a tropical rainforest in Sabah, Malaysia. – Malar. J. 15: 370.

Brown, E. et al. 2013. Common guidance for the identification of high conservation values. – HCV Resour. Netw.

Diwakar, S. et al. 2007. Psychoacoustic sampling as a reliable, non-invasive method to monitor orthopteran species diversity in tropical forests. – Biodivers. Conserv. 16: 4081–4093.

Ewers, R. M. et al. 2011. A large-scale forest fragmentation experiment: the Stability of Altered Forest Ecosystems Project. – Phil. Trans. R. Soc. B 366: 3292–3302.

GBIF Secretariat 2020. GBIF backbone taxonomy. – Checkl. Dataset Accessed March 2020. <www.gbif.org/>.

Gemmeke, J. F. et al. 2017. Audio Set: an ontology and human-labeled dataset for audio events [WWW Document]. – Google AI, <https://ai.google/research/pubs/pub45857>, accessed 24 July 2018.

Gibb, R. et al. 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. – Methods Ecol. Evol. 10: 169–185.

Gijzen, H. 2013. Big data for a sustainable future. – Nature 502: 38–38.

Hershey, S. et al. 2017. CNN architectures for large-scale audio classification. – In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 131–135.

Hirzel, A. H. and Lay, G. L. 2008. Habitat suitability modelling and niche theory. – J. Appl. Ecol. 45: 1372–1381.

Hirzel, A. H. et al. 2006. Evaluating the ability of habitat suitability models to predict species presences. – Ecol. Model. Predict. Species Distrib. 199: 142–152.

Inger, R. F. and Voris, H. K. 1988. Taxonomic status and reproductive biology of bornean tadpolecarrying frogs. – Copeia 1988: 1060–1061.

Jucker, T. et al. 2018. Canopy structure and topography jointly constrain the microclimate of human-modified tropical landscapes. – Global Change Biol. 24: 5243–5258.

Kahl, S. et al. 2021. BirdNET: a deep learning solution for avian diversity monitoring. – Ecol. Inform. 61: 101236.

Lambin, E. F. and Meyfroidt, P. 2011. Global land use change, economic globalization and the looming land scarcity. – Proc. Natl Acad. Sci. USA 108: 3465–3472.

Lefsky, M. A. et al. 2002. Lidar remote sensing of above-ground biomass in three biomes. – Global Ecol. Biogeogr. 11: 393–399.

Leupen, B. T. et al. 2018. Trade in white-rumped shamas *Kittacincla malabarica* demands strong national and international responses. – Forktail J. Asian Ornithol. 34: 1–8.

Luke, S. H. et al. 2017. The impacts of habitat disturbance on adult and larval dragonflies (Odonata) in rainforest streams in Sabah, Malaysian Borneo. – Freshwater Biol. 62: 491–506.

Mills, L. S. et al. 1993. The keystone-species concept in ecology and conservation: management and policy must explicitly consider the complexity of interactions in natural systems. – BioScience 43: 219–224.

Newbold, T. et al. 2015. Global effects of land use on local terrestrial biodiversity. – Nature 520: 45–50.

Pieretti, N. et al. 2011. A new methodology to infer the singing activity of an avian community: the acoustic complexity index (ACI). – Ecol. Indic. 11: 868–873.

Pijanowski, B. C. et al. 2011. Soundscape ecology: the science of sound in the landscape. – BioScience 61: 203–216.

Popescu, S. C. et al. 2011. Satellite lidar vs small footprint airborne lidar: comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. – Remote Sens. Environ. DESDynI VEG-3D 115: 2786–2797.

Riutta, T. et al. 2018. Logging disturbance shifts net primary productivity and its allocation in Bornean tropical forests. – Global Change Biol. 24: 2913–2928.

Sethi, S. S. et al. 2020. Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. – Proc. Natl Acad. Sci. USA 117: 17049–17055.

Store, R. and Jokimäki, J. 2003. A GIS-based multi-scale approach to habitat suitability modeling. – Ecol. Model. 169: 1–15.

Stowell, D. et al. 2016. Bird detection in audio: a survey and a challenge. 2016 IEEE 26th Int. Workshop on Machine Learning for Signal Processing (MLSP). doi: 10.1109/MLSP.2016.7738875. <https://ieeexplore.ieee.org/document/7738875>.

Stowell, D. et al. 2019. – Automatic acoustic identification of individuals in multiple species: improving identification across recording conditions. – J. R. Soc. Interface 16 20180940 20180940 <https://royalsocietypublishing.org/doi/10.1098/rsif.2018.0940>.

Sueur, J. and Farina, A. 2015. Ecoacoustics: the ecological investigation and interpretation of environmental sound. – Biosemiotics 8: 493–502.

Sueur, J. et al. 2008. Rapid acoustic survey for biodiversity appraisal. – PLoS One 3: e4065.

Swinfield, T. et al. 2020. LiDAR canopy structure 2014. doi: 10.5281/zenodo.4020697

Walsh, R. P. D. and Newbery, D. M. 1999. The ecoclimatology of Danum, Sabah, in the context of the world's rainforest regions, with particular reference to dry periods and their impact. – Phil. Trans. R. Soc. B 354: 1869–1883.

Walther, G.-R. et al. 2002. Ecological responses to recent climate change. – Nature 416: 389–395.

Williamson, J. et al. 2021. Riparian buffers act as microclimatic refugia in oil palm landscapes. – J. Appl. Ecol. 58: 431–442.

Wrege, P. H. et al. 2017. Acoustic monitoring for conservation in tropical forests: examples from forest elephants. – Methods Ecol. Evol. 8: 1292–1301.