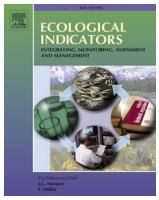
Ecological Indicators 138 (2022) 108831

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind



Soundscape classification with convolutional neural networks reveals temporal and geographic patterns in ecoacoustic data

Colin A. Quinn [a,*], Patrick Burns [a], Gurman Gill [b], Shrishail Baligar [c], Rose L. Snyder [d], Leonardo Salas [d], Scott J. Goetz [a], Matthew L. Clark [e]

[a] School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA
[b] Department of Computer Science, Sonoma State University, Rohnert Park, CA, USA
[c] Electrical Engineering and Computer Science, University of California, Merced, CA, USA
[d] Point Blue Conservation Science, Petaluma, CA, USA
[e] Center for Interdisciplinary Geospatial Analysis, Geography, Environment and Planning, Sonoma State University, Rohnert Park, CA, USA

ARTICLE INFO

Keywords:
Machine learning
Convolutional neural network (CNN)
Ecoacoustics
Anthropophony
Biophony
Naturally quiet landscapes
Soundscape ecology

ABSTRACT

Interest in ecoacoustics has resulted in an influx of acoustic data and novel methodologies to classify and relate landscape sound activity to biodiversity and ecosystem health. However, indicators used to summarize sound and quantify the effects of disturbances on biodiversity can be inconsistent when applied across ecological gradients. This study used an acoustic dataset of 487,148 min from 746 sites collected over 4 years across Sonoma County, California, USA, by citizen scientists. We built a custom labeled dataset of soundscape components and applied a deep learning framework to test our ability to predict these soundscape components: human noise (Anthropophony), wildlife vocalizations (Biophony), weather phenomena (Geophony), Quiet periods, and microphone Interference. These soundscape components allowed us to balance predicting variation in environmental recordings and relative time to build a custom labeled dataset. We used these data to quantify soundscape patterns across space and time that could be useful for environmental planning, ecosystem conservation and restoration, and biodiversity monitoring. We describe a pre-trained convolutional neural network, fine-tuned with our sound reference data, with classification achieving an overall F0.75-score of 0.88, precision of 0.94, and recall of 0.80 across the five target soundscape components. We deployed the model to predict soundscape components for all acoustic data and assess their hourly patterns. We noted an increase in Biophony in the early morning and evening, coinciding with peak animal community vocalization (e.g., dawn chorus). Anthropophony increased during morning/daylight hours and was lowest in the evenings, coinciding with diurnal patterns in human activity. Further, we examined soundscape patterns related to geographic properties at recording sites. Anthropophony decreased with increasing distance to major roads, while Quiet increased. Biophony and Quiet were comparable to Anthropophony at more urban/developed and agriculture/barren sites, while Biophony and Quiet were significantly higher than Anthropophony at less-developed shrubland, oak woodland, and conifer forest sites. These results demonstrate that acoustic classification of broad soundscape components is possible with small datasets, and classifications can be applied to a large acoustic dataset to gain ecological knowledge.

1. Introduction

The value of different sounds across the landscape has long been recognized as socially valuable (Schafer, 1993; Southworth, 1969), and acoustic data are becoming more economical and efficient to collect, permitting characterization of spatial and temporal patterns of biodiversity, human activity, and other sounds (Depraetere et al., 2012; Shonfield and Bayne, 2017). The acoustic quality of habitats is also recognized as a vital dimension of conservation (Dumyahn and Pijanowski, 2011; Schafer, 1993), as increasingly excessive human noise can have a range of direct deleterious effects on biodiversity (e.g., acoustic masking from overlapping communication frequency ranges) (Doser et al., 2019; Francis et al., 2017). Identifying naturally quiet landscapes and relating patterns in anthropogenic and biotic noise is essential in understanding the effects of changing human activity on biodiversity and noise reduction on conservation and management efforts of

* Corresponding author at: School of Informatics, Computing, and Cyber Systems, Northern Arizona University, 1295 Knoles Dr, Flagstaff, AZ 86011, USA.
E-mail address: cq73@nau.edu (C.A. Quinn).

https://doi.org/10.1016/j.ecolind.2022.108831
Received 3 November 2021; Received in revised form 23 March 2022; Accepted 31 March 2022
Available online 5 April 2022

protected areas (Newport et al., 2014; Rice et al., 2020).

The unique assemblage of sounds across a landscape is collectively referred to as a soundscape (Krause, 2002; Pijanowski et al., 2011) and is treated as an ecological characterization of landscapes (Pavan, 2017, Sethi et al., 2020). Recorded soundscapes consist of anthropogenic activity (anthropophony), wildlife vocalizations (biophony), and weather-related phenomena (geophony; Pijanowski et al., 2011), along with quiet (the ambient sound or lack of acoustic events) at a given timeframe. Because soundscapes integrate the acoustic dynamics of an ecosystem, they can be considered as "community phenotypes," integrating vocalizing species, anthropogenic noise, and natural phenomena (Lellouch et al., 2014). Capturing soundscape snapshots provide meaningful insights related to biodiversity and human impacts, highlighting changes such as degraded habitats (Bush et al., 2018; Dumyahn and Pijanowski, 2011; Sueur et al., 2008). Recent work has related soundscape activity to geographic characteristics such as land-use change (Eldridge et al., 2018; Sethi et al., 2020), as well as gradients in vegetation and forest structure (Boelman et al., 2007; Dröge et al., 2021; Farina and Pieretti, 2014). Human impact has been linked to soundscape variation, including increased anthropogenic noise near high traffic roads (Doser et al., 2019), complex interactions between biophony and anthropophony in urban soundscapes (e.g., overlapping frequencies) (Fairbrass et al., 2017), soundscape similarity in oil palm production with forested soundscapes (Furumo and Aide, 2019), and impacts of snowmobile activity on winter quiet (Mullet et al., 2017b). Additional research linking patterns in sounds across landscapes and time will improve the utility of ecoacoustic methods for informing conservation and land management.

Two standard methods of assessing the information in soundscapes include deriving acoustic indices and manual identification of acoustic events (i.e., done by a human, not an algorithm). Acoustic indices describe the acoustic energy in amplitude, time, and frequency space (Sueur et al., 2014), while manual identification provides specific time and frequency data (Grant and Samways, 2016; Rose et al., 2018). Both methods summarize sound into meaningful, interpretable ecological indicators (e.g., species diversity or acoustic complexity). However, acoustic indices vary in their ability to convey meaningful biodiversity information, making comparisons between studies or geographic regions non-trivial (Bradfer-Lawrence et al., 2019; Metcalf et al., 2021). Poor performance of acoustic indices has been attributed to the presence of confounding, background sounds in recordings, making identification and filtering of non-biophonic noise a necessary step for indices to be applied for consistent biodiversity and human impact monitoring (Depraetere et al., 2012; Eldridge et al., 2018; Fairbrass et al., 2017). Likewise, manual identification of sound sources is highly time-intensive and requires detailed knowledge of target animal vocalizations, thus, limiting this approach to smaller datasets (Pérez-Granados and Traba, 2021; Shonfield and Bayne, 2017).

Due to recent computational innovations, soundscape dynamics and their associated patterns of biodiversity can be derived with less effort and time using deep learning identification (Christin et al., 2019; Fairbrass et al., 2019), source separation (Lin and Tsao, 2020), and unsupervised classification (Sethi et al., 2020). Efforts in environmental sound classification established the effectiveness of convolutional neural networks (CNN), a type of deep learning architecture (Lecun et al., 2015), for soundscape classification (Piczak 2015; Salamon and Bello, 2017). In ecoacoustics, CNNs have been applied in species-specific vocalization (Christin et al., 2019; Kahl et al., 2018; Ruff et al., 2021) and targeted sound classification (e.g., gunshots or rain) (Metcalf et al., 2020; Sánchez-Giraldo et al., 2020). Alternatively, source separation methods have successfully identified specific types of sound by separating sound mixtures into individual sources (Eldridge et al., 2018; Lin and Tsao, 2020). Few ecoacoustic studies have attempted to classify entire soundscapes using broadly defined sound categories such as anthropophony, biophony, geophony, and quiet. One study that classified broader soundscape components developed two CNN classifiers that modeled biophony and anthropophony in urban London (Fairbrass et al., 2019). They found that CNN models trained on a limited amount of expertly annotated sound samples outperformed multiple acoustic indices in representing human and animal activity patterns in urban soundscapes (Fairbrass et al., 2019). Other efforts have successfully modeled acoustic-environmental relationships of anthropophony, biophony, and geophony, but this required manual identification of sounds in almost 60,000 recordings (Mullet et al., 2016).

Ecoacoustic analyses typically focus on one or a few target sounds, but to characterize these sounds with confidence, other, confounding sounds need to be identified and, in some cases, removed. Some methods to account for unwanted sound include avoiding sites close to roads (Duchac et al., 2020), using pre-programmed amplitude or frequency audio filters (Bedoya et al., 2017; Duchac et al., 2020; Towsey, 2013), relating meteorological data to recordings to filter weather events (Desjonquères et al., 2018; Gasc et al., 2018), or manual identification of noises (Bradfer-Lawrence et al., 2019; Gordon et al., 2018). Comparatively, deep learning solutions can efficiently and accurately extract important data features, allowing for noise from signal separation. Products from deep learning can address many of the above issues in accounting for confounding noises and classifying soundscapes in one modeling framework (Christin et al., 2019). For example, these modeling efforts can negate the need for acoustic filtering, which is broadly applied but is ineffective in specific environments (e.g., frequency filtering in urban landscapes) (Fairbrass et al., 2017). Deep learning models can also be more broadly applicable than traditional acoustic indicators because they can be updated when presented with new acoustic data (Fairbrass et al., 2019; Ruff et al., 2021).

Here, we present methods to classify soundscape components using a deep learning approach in Sonoma County, California, USA, an area that includes a gradient of natural and anthropogenic ecosystems. We aim to demonstrate the ability to classify broadly inclusive soundscape components: Anthropophony, Biophony, Geophony, Quiet, and physical or electronic recorder Interference (ABGQI) while accounting for modeling uncertainty using deep-learning practices and accuracy on par with current modeling efforts (Christin et al., 2019; Ruff et al., 2021). We investigate the relationship between these soundscape indicators with land use/land cover and road distance to understand soundscape variation across human impact and geographic gradients. Soundscape classification allows for (a) automated identification of unwanted sounds in large amounts of data, (b) modeling the effects and interactions of different sounds, and (c) use of modeling products themselves (e.g., classified acoustic samples) to understand spatio-temporal patterns in sound activity.

2. Methods

2.1. Study region and acoustic data collection

The study area was Sonoma County, California, USA (38.51°N, 122.93°W), covering 4,152 km^2 north of the San Francisco metro area (Fig. 1). The county has a Mediterranean climate with average annual precipitation of 1,040 mm (Supplementary Materials S.1). Non-native annual grasses dominate grasslands and can be unmanaged or support beef and dairy cows. Urban areas and agriculture (primarily vineyards) are concentrated in valleys (Fig. 1). The county's common vocalizing animals include multiple bird species, amphibians, and invertebrates (e.g., crickets, katydids, cicadas).

Autonomous recording units (ARUs) were deployed across Sonoma County in 2017–2020 as part of the Soundscapes to Landscapes (S2L; soundscapes2landscapes.org) citizen science project. Sites were selected across the county based on a topographic lowland and highland stratification, then broad land use/land cover (LULC) types, such as forest, shrubland, herbaceous, urban areas, and agriculture (sonomavegmap.org; Supplementary Materials S.1). This stratification scheme provided field sites with heterogeneous vegetation types, various vocalizing

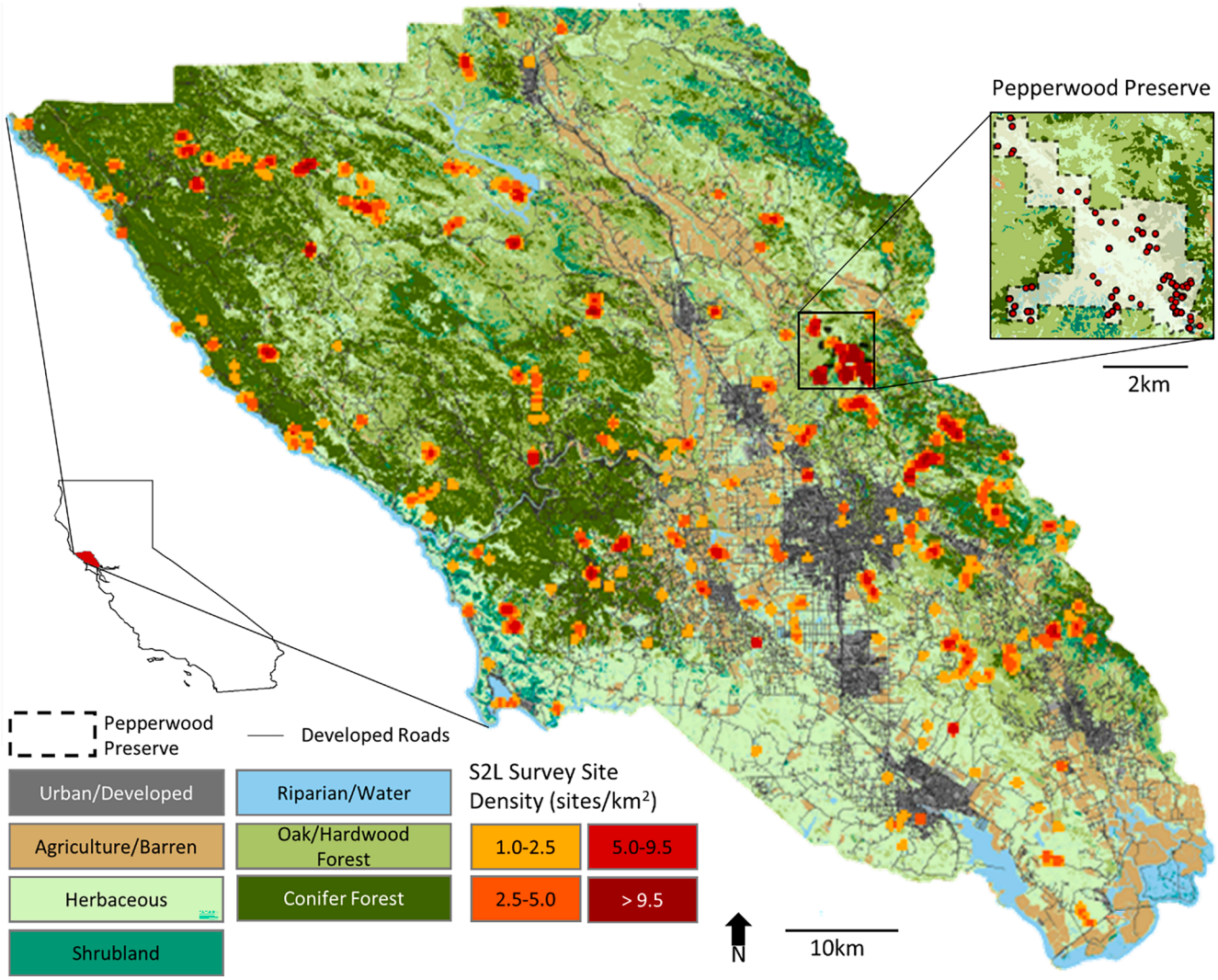


Fig. 1. Sonoma County, California, USA land use/land cover classes derived from Sonoma County Fine-scale Vegetation and Habitat Map. Inset of Pepperwood Preserve sites. We show Soundscapes to Landscapes site location densities from 2017 to 2020 (n = 746).

species, and a range of human impacts to capture diverse acoustic settings.

At each site, citizen scientists deployed a single ARU, either an Android-LG smartphone with an attached microphone in a waterproof case or an AudioMoth recording device (Hill et al., 2018) in a vinyl protective pouch. ARUs were deployed for at least 72 h, with programming to record 60-s every 10 min, resulting in six minutes per hour and 144 min per 24-hour period. Each minute recording was saved in a waveform audio file format (.wav) with 16-bit digitization depth and 44.1 kHz or 48 kHz sampling rate for LG ARUs and AudioMoth ARUs, respectively. This sampling rate allowed for an effective upper frequency of 22.05 kHz for LG ARUs and 24 kHz for AudioMoth ARUs.

We collected 487,148 min (~8,000 h) of sound data during bird breeding seasons (March-July) across 746 sites with an average of 642 ± 370 min per site (Fig. 1). The 746 sites were distributed across years unevenly: 2017 (n = 122), 2018 (n = 89), 2019 (n = 345), and 2020 (n = 190) and ARUs: LG (n = 201) and AM (n = 545). Sampling effort was not consistent across years due to an early project prototyping phase with less data (2017) and COVID-19 lockdown (2020), which delayed recorder deployment. Furthermore, sites were not evenly distributed across LULC types: agriculture/barren (n = 11), conifer forest (n = 222), oak/hardwood forest (n = 283), herbaceous (n = 147), riparian/wetland (n = 21), shrubland (n = 47), and urban/developed (n = 15) (Supplementary Materials S.2). Uneven distribution of sites across LULC types is primarily due to land accessibility in Sonoma.

2.2. Deep learning classification of soundscape components

Deep learning, a branch of machine learning, uses multiple hidden, non-linear transformations to automatically derive abstracted representations of raw data (e.g., images or words). Comparatively, non-deep learning modeling techniques are limited in their ability to interpret raw data without prior feature extraction by a human (e.g., amplitude or frequency) (Christin et al., 2019; Lecun et al., 2015). Here, we are interested in identifying specific types of sounds in audio recordings, represented as spectrogram images (Fig. 2). We used a CNN to which we supplied a training dataset composed of acoustic images labeled present or absent for a given class (e.g., is human noise present?). Training the model (learning) begins with extracting basic "features" (e.g., shape and texture) from the image set. These extraction steps become abstracted within the model until the final step, which attempts to classify the image by relating learned features to the known image label. After training, we used another set of labeled images withheld during model training (test set) for model selection and accuracy evaluation. We then predicted the presence and absence of sounds in all recordings with the best performing model.

To identify soundscape components across Sonoma County, we used a CNN to classify five broad target classes: Anthropophony, Biophony,

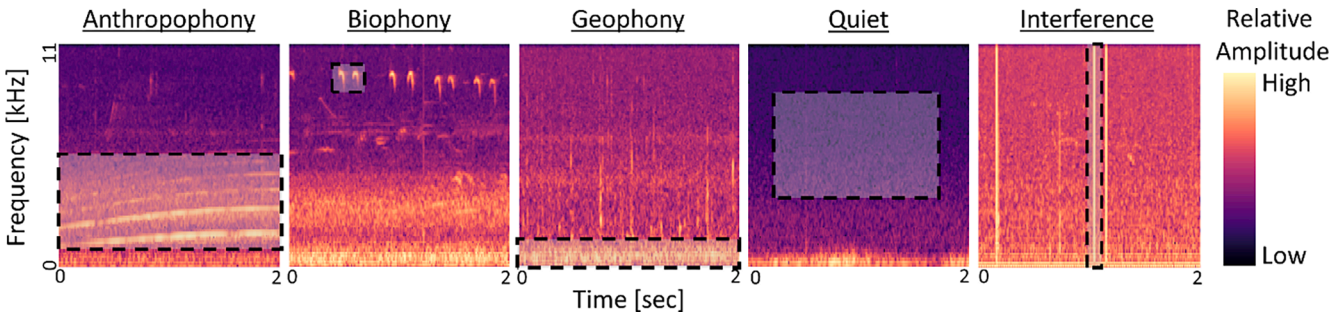


Fig. 2. The five broad soundscape classes shown here are examples of the acoustic activity represented in Mel spectrograms. The Mel spectrograms are 2-s long (x-axis), span from 0 kHz to 11.025 kHz (y-axis), and display relative amplitude (z-axis). Manually annotated regions of interest (ROIs) are represented as black hashed and shaded boxes. See Section 2.2.3 for Mel spectrogram methods.

Geophony, Quiet, and Interference (ABGQI) (Fig. 2). The Quiet class represented periods without other soundscape components below 11 kHz (i.e., no discernible acoustic events), resulting in periods with minimal change in sound from baseline ARU self-noise levels (Fig. 2 – Quiet; Supplementary Materials S.5). Interference is defined here as broadband, short duration, electronic or physical microphone interference events, for example, when a branch hits a recorder during a strong wind gust. We classified Interference to represent recording error, while ABGQ relay ecologically meaningful information. Converting raw acoustic data to spectrograms allowed us to use the well-established pre-trained MobileNetV2 architecture, a lightweight CNN trained on Google search imagery (ImageNet; github.com/tensorflow/models/tree/master/research/slim/nets/mobilenet) for transfer learning (Christin et al., 2019; Yosinski et al., 2014). Other ecoacoustic recognition tasks have successfully demonstrated the effectiveness of spectrograms and CNNs (Fairbrass et al., 2019; Sethi et al., 2020).

2.2.1. *Labeled dataset collection*

To create labeled data for ABGQI classification, we randomly sampled audio files from the entire S2L dataset, first stratified by LULC type and then by site. An upper limit of 350 audio files was sampled from each LULC type amounting to 2,367 sampled recordings. Sites within each LULC type were sampled equally (i.e., if urban/developed land cover contained 15 sites, 23 audio files were randomly selected from each site for 345 total audio files). We verified that random sampling within LULC and sites resulted in an even sampling across 24 h (Supplementary Materials S.3).

Citizen scientists and team members were randomly assigned unique recordings to identify ABGQI sounds from the audio file subset by loading the recording in Raven Lite 2 audio software program (Cornell Lab of Ornithology, Cornell University, Ithaca, NY; ravensoundsoftware.com). Audio files were then annotated with regions of interest (ROIs; Figs. 2 and 3), or discrete sound events, by visualizing spectrograms with a standard frequency range of 0–12 kHz and spectrogram window size of 512 samples. We chose a cutoff duration of 5-s because some sound events are not temporally discrete (e.g., constant wind or vehicle traffic) and limited the length of ROIs to no more than 5 ROIs of the same class per audio file to account for oversampling and biasing individual sound files. C.Q. then verified that every ROI was a true presence. This process resulted in 5,396 ROI annotations in 1,194 of the 2,367 subset

Table 1
The number of regions of interest (ROIs) from S2L and Freesound data, the number of 2-s Mel spectrogram (Mel spec) samples created from all ROIs, and the final 2-s training set size for each soundscape class.

Sound Class	S2L ROIs	Freesound ROIs	Mean ROI length (s)	Total 2-s Mel spec	Training set size
Anthropophony	916	883	1.67	2,170	1,920
Biophony	2,372	556	1.88	3,336	3,086
Geophony	957	801	1.60	1,955	1,705
Quiet	765	0	3.06	1,023	773
Interference	386	0	2.05	430	330
TOTAL	**5,396**	**2,240**	–	**8,914**	**7,814**

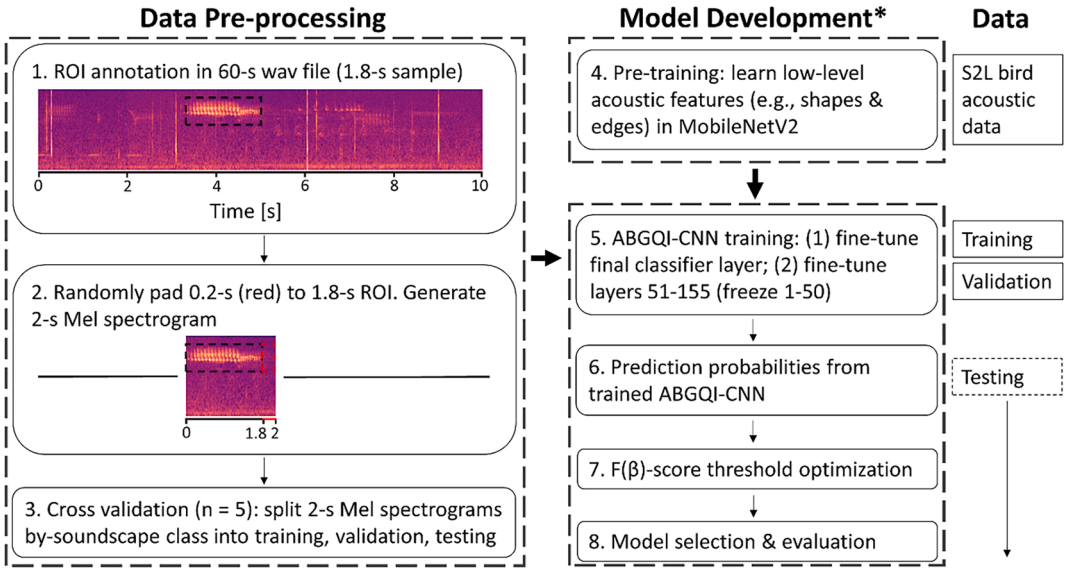


Fig. 3. Our workflow for data preprocessing (1–3) and model development (4–8). *(5–8) performed on each cross-validation data split.

recordings with an average count of 2.77 ± 3.56 ROIs per recording with a presence (Table 1). A list of the target sounds is in Supplementary Table S.4.1 and ROI collection methods at (https://doi.org/10.5281/zenodo.6027024).

2.2.2. Auxiliary Freesound data

CNN classification training typically requires thousands to millions of images or hundreds of hours of acoustic data (Christin et al., 2019; Knight et al., 2017). We added ROI samples using open-access Freesound data because there were fewer than 1,000 ROIs per class in our ABGQI ROI data (freesound.org; file data at https://doi.org/10.5281/zenodo.6027024). C.Q. listened to each Freesound file for quality (e.g., clarity and presence of single, unmixed sound) because recordings were long and variable (60.25 ± 75.13-s) and were sometimes poorly labeled. We used the autodetec function in the package warbleR (Araya-Salas and Smith-Vidaurre, 2017) in the statistical software program R (R Core Team, 2020) to isolate sound events in recordings. If an event was evident in the spectrogram (e.g., consistent with the general characteristics of the soundscape component), it was included in the ROI dataset resulting in the addition of 2,240 Freesound ROIs (Table 1).

2.2.3. Spectrogram generation and cross-validation

We chose to use 2-s Mel spectrogram representations of ROIs to balance short-duration acoustic events (e.g., bird chirps) with longer-duration events (e.g., vehicle traffic) (Zhong et al., 2020). We clipped and combined all ROIs into a single synthetic acoustic file for each soundscape component. ROIs > 2-s were directly added to these recordings, whereas ROIs < 2-s were padded randomly around the ROI center time for a total of 2-s (Fig. 3; e.g., if an ROI was 1.5-s, 0.5-s from the audio file would be randomly added before or after the ROI). This process meant short ROIs had additional surrounding context, and padding could introduce unknown sounds, reflecting the model's real-world application. Because we directly added ROIs > 2-s to each soundscape component's synthetic recording, some 2-s spectrogram segments contained multiple unique ROIs. Synthetic soundscape component recordings were spliced into 2-s segments and converted into Mel spectrograms using python 3.7.7′s librosa 0.6.3 library (McFee et al., 2015; Python Software Foundation, 2016). Although ARUs sampled at a rate of 44.1 kHz or higher (effective reproduction of frequencies below 22.05 kHz), spectrograms were generated from 0 kHz to 11.025 kHz (half the Nyquist frequency), regardless of ROI frequency range. The upper-frequency limit overlapped with most acoustic indices and allowed for frequency shifts above the 0–8 kHz frequency window where most anthropogenic and low-frequency animal vocalizations occur (Boelman et al., 2007; Kasten et al., 2012; Sueur et al., 2014). The resulting image was 432x432 pixels in red–green–blue color bands representing amplitude levels (Fig. 2). Spectrograms were down-sampled to 224x224 pixel, 3-banded images for CNN training, testing, and deployment; the default input size for the pre-trained CNN.

In CNN modeling practices, training data are used for learning data features, while validation data are used for model assessment and tuning hyper-parameters during training. In contrast, testing data are withheld to perform an independent model performance evaluation and generate accuracy estimates. We randomly sampled spectrograms from each sound class into validation (n = 200), testing (n = 50), and training (all remaining component spectrograms; Table 1) datasets five times for a five-fold cross-validation training approach (Fig. 3). Cross-validation provided a measure of model performance to strengthen model evaluation given the small dataset size. We decreased the number of Interference validation samples to n = 50 because of the class's low sample count. We had 850 total validation samples and 250 total testing samples. The model with the highest class-specific accuracy in cross-fold validation was selected for evaluation and prediction to the county-wide dataset. We recognized the small size of the test set but prioritized training performance, provided evaluation metrics in a cross-fold validation framework, and provided an independent accuracy test on a soundscape dataset in the Supplementary Materials (S.12).

2.2.4. CNN transfer learning

We developed the CNN in Python (v.3.7.7, Python Software Foundation, 2016) using the TensorFlow v.2.3.0 framework (Abadi et al., 2015). The fully trained CNN model and related code are available at https://doi.org/10.5281/zenodo.6112713. CNN training is most effective when sizeable labeled training datasets are available (Christin et al., 2019). However, when large labeled datasets are unavailable, or we desire faster training time, features developed in existing CNNs can be "transferred" to the new dataset through transfer learning (Yosinski et al., 2014). We tested pitch shift augmentation but ultimately did not include any form of data augmentation in the final CNN as it required increased computation with minimal change in accuracy (Salamon and Bello, 2017; Piczak, 2015) (Supplementary Materials S.7).

We applied concepts from transfer learning using the pre-trained MobileNetV2 architecture. First, we pre-trained the ImageNet weighted MobileNetV2 using existing labeled acoustic data of calls from 54 bird species from the S2L project, some recordings overlapping ABGQI data (Fig. 3). We used S2L bird data ROIs to learn low-level spectrogram features from an existing, large dataset similar to the domain of interest (Supplementary Materials S.8). Following acoustic feature pre-training, we trained the S2L-MobileNetV2 CNN with ABGQI data (Supplementary Materials S.9). This modeling framework leveraged previously learned low-level spectrogram features from S2L bird data while modifying weights to learn high-level, ABGQI-specific features. The fully trained CNN produced a vector of five probabilities using a binary cross-entropy classifier with sigmoid activation (one value for ABGQI each). We refer to the fully-trained model as the ABGQI-CNN.

2.3. Classification threshold optimization

During model training, we used a custom threshold process, and testing data to (1) create binary classified values (present = 1 / absent = 0) for all classes in each 2-s spectrogram and (2) assess model performance. Threshold value choices should be made on a study-to-study basis depending on the research question (Knight et al., 2017). Each class was optimized individually using the 50 presences and 200 absences from the testing data by setting threshold values from 0 to 1 in 0.0001 increments, forcing probabilities below the threshold to zero/absent and probabilities above the threshold to one/present, and selecting the threshold where the F(β)-score was maximized. F(β)-score (Eq.1):

$$F(\beta) = \frac{\left(1+\beta^2\right)*(precision*recall)}{\left(\beta^2*precision\right)+recall} \quad (1)$$

where precision = TP/(TP + FP) and recall = TP/(TP + FN); TP is true positive, FP is false positive, and FN is false negative.

Generally, prioritizing precision is favored to minimize false positives (Kahl et al., 2021; Knight et al., 2017; LeBien et al., 2020), and automated audio classification tasks can result in excess false positives (Balantic and Donovan, 2020). Prioritizing precision over recall means that classifications are conservative estimates for each class; i.e., there are fewer false positives and many true positives, but at the cost of increased false negatives. We used soundscape component CNN predictions and threshold values based on maximum F(β)-scores to produce binary classifications (Fig. 5). This framework allowed each 2-s spectrogram to have more than one class present, resulting in some 2-s spectrograms without a present soundscape prediction. In these cases, we created a sixth post-classification category, "Unidentified," which occurred when ABGQI were absent in a spectrogram. We tested three β values: 0.50, 0.75, and 1.00. β values below 1.00 weight precision higher than recall while 1.00 weights recall equal to precision (Kahl, 2020). Following CNN threshold selection, we used changes in evaluation

metrics (F(β)-score, precision, recall) and the rate of Unidentified samples in the test dataset to select the optimal F(β)-score model.

2.4. ABGQI-CNN model deployment

We deployed the modeling framework for inference on the entire S2L acoustic dataset to classify ABGQI and the Unidentified class (Fig. 4). Recordings used in the ABGQI-CNN training process (n = 1,195 recordings) constituted 0.25% of the entire S2L audio dataset. In our subsequent analyses, we removed these recordings to account for any bias in model deployment. The model deployment included the generation of 2-s spectrograms, ABGQI-CNN probability predictions, and classification with optimized prediction thresholds (Fig. 4). Deployment resulted in binary ABGQI classifications for every 2-s spectrogram (n = 14,578,590), indicating the presence or absence of each of the five classes or unidentified. We calculated percent time present, the number of predicted 2-s class presences relative to total 2-s spectrograms, using presence/absence values aggregated hourly and by-site for a rate of sound present. For example, if a site with 100 2-s clips had five 2-s presence, the rate was 0.05 or 5%.

2.5. Sound pattern statistical analyses

2.5.1. Covariate data

We used ABGQI percent time present to examine temporal variation in soundscape recording site characteristics. The goal of these analyses was to examine how classified sounds reflect (1) expected patterns across the landscape, (2) site characteristics (e.g., does Anthropophony vary with distance from road), and (3) temporal sound variation (e.g., diel sound patterns). Covariates in these analyses included the majority LULC type within a 50-m radius of recorder location, distance to the nearest road, temporally relevant measures of recording (i.e., diel, monthly, and annual patterns), and meteorological (MET) station data. We extracted LULC data using aggregated classes from the Sonoma County Fine-scale Vegetation and Habitat Map (sonomavegmap.org). Road distance was calculated as the linear distance from a site to the nearest improved road (i.e., paved or high-use roads; gis-sonomacounty.hub.arcgis.com). We used meteorological data from Pepperwood Preserve's five MET stations (Ferrell et al., 2021a, 2021b) with 15-min resolution data co-occurring near 56 S2L sites at Pepperwood (n = 39,552 recordings).

2.5.2. Comparative analyses

We used hourly estimates of ABGQI to analyze diurnal patterns in rates of sounds (24-hour and night vs. day) among LULC types. We used the by-site estimates of ABGQI daytime data (5 a.m. to 8 p.m.), a temporal subset that focused on animal and human activity co-occurrence, to compare ABGQI percent time present with overlapping deployment dates (day of the year), LULC, and road distance. Overlapping deployment dates (each site's first day of recording) were used to analyze whether there were annual differences in soundscape activity. This aspect is essential because, for example, annual start dates may overlap each year differently with events such as bird breeding season and therefore capture different amounts of Biophony. Comparisons among LULC types helped discern any diagnostic ABGQI patterns within these ecosystems. We further explored how these relationships are modulated by distance from the nearest road. Road distance was treated as a grouped variable to examine patterns in sound in 100-m intervals (e.g., 0–99 m, 100–199 m) from 0 to 1000 m and > 1000 m. We took hourly averages of wind speed data from Pepperwood MET stations to the nearest recording site and compared these data to hourly amounts of soundscape components using linear regression (i.e., does Geophony co-vary with wind speed).

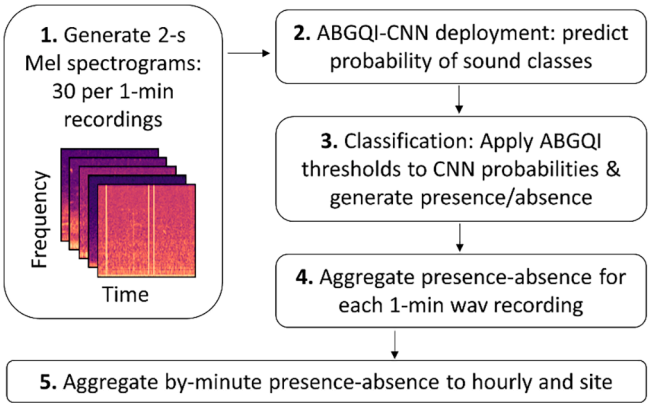


Fig. 4. Workflow chart for the deployment of ABGQI-CNN on the entire S2L recording dataset.

Because sample sizes and variances among treatment categories were not homogeneous, we used a Kruskal-Wallis test to evaluate effects for tests with more than two samples. If a Kruskal-Wallis test was significant (large χ^2 statistic and z-values and $p < 0.05$ indicate considerable differences are present), we then applied Dunn's simultaneous multiple-comparison test with a Bonferroni correction to account for the inflated error rate to identify which pairwise relationships were significantly different (when p-adj < 0.05). We used Mann-Whitney's *U* test for two-sample tests (i.e., day and night soundscape activity). We reported all pairwise test results in the Supplementary Materials (S.15).

2.5.3. Modeling of soundscape components

We used main-effects-only linear regression models to find underlying factors and drivers of the amount of predicted soundscape components. Descriptive covariates included road distance (log-transformed), number of total 2-s samples at the site (log-transformed), LULC types, ARU type, year, and type of sound (i.e., ABQGI). The year was included to capture year-unique effects, such as deployment date or changes in human activity. The response, amount of sound, was the percent of positively predicted 2-s samples for each ABGQI as a logit score and included all soundscape data. For example, if Biophony was predicted in 80 of 100 2-s samples, this would result in a 0.8 positive rate (80%), converted to a logit value of 1.39. The initial model included all candidate variables, quadratic and cubic road distances, and interactions between road distance and soundscape type and LULC. We used the function stepAIC from the package MASS in R (Venables and Ripley, 2002) to select the optimal model (Supplementary Materials S.10).

3. Results

3.1. Model performance

We used the highest performing ABGQI model that included pre-training with bird vocalization and auxiliary Freesound data (F0.75-score = 0.883 ± 0.035) for evaluation and deployment (optimal model results in Table 2). Performance was determined from cross-validation

Table 2
Optimal model performance was based on independent test data for each soundscape class using recommended evaluation metrics. We assessed model classification performance with the F0.75-score. We also report F1-score and area under the curve (AUC) for comparison with other studies.

Soundscape Class	Precision	Recall	F0.75-score	F1-score	AUC
Anthropophony	0.975	0.780	0.894	0.867	0.939
Biophony	0.913	0.840	0.885	0.875	0.984
Geophony	0.923	0.720	0.838	0.809	0.959
Quiet	0.891	0.820	0.864	0.854	0.984
Interference	0.977	0.860	0.932	0.915	0.980
AVERAGE	**0.936**	**0.804**	**0.883**	**0.865**	**0.969**

results. Cross-validation had consistent accuracy across iterations (F0.75-score = 0.838 ± 0.03; precision = 0.902 ± 0.03; recall = 0.754 ± 0.03) indicating that our sample sizes were sufficient to learn spectrogram features (i.e., consistent performance across folds). Pre-training with S2L bird vocalization data resulted in the default MobileNetV2 learning acoustic-specific features and increased average F0.75-score from 0.659 to 0.756, average precision from 0.643 to 0.790, and a decrease in average recall from 0.712 to 0.704. Adding Freesound data to ABGQI data training, the default MobileNetV2 resulted in an overall increase in average F0.75-score from 0.659 to 0.760, average precision increased from 0.643 to 0.847, and average recall decreased from 0.712 to 0.672. Combined, pre-training with S2L bird vocalization and auxiliary Freesound data increased the average F0.75-score from 0.659 to 0.883. The optimal F(β)-score value of β = 0.75 was chosen for threshold values (Fig. 5) based on a balance between prioritizing precision over recall and accounting for the number of unidentified 2-s spectrograms (see Supplementary Materials S.11 for cross-validation threshold evaluation). We have provided an additional measure of model accuracy and generalizability in the Supplementary Materials (S.12).

3.2. Statistical analyses of soundscape components

The final ABGQI-CNN predicted the highest site average hourly amount of Quiet (μ = 24.6%, σ = 27.7%), followed by Biophony (μ = 24.2%, σ = 24.8%), Interference (μ = 9.8%, σ = 17.4%), Geophony (μ = 8.5%, σ = 12.9%), and Anthropophony (μ = 5.5%, σ = 9.8%), while Unidentified averaged 29.3% ± 17.5%. See Supplementary Materials (S.13) for a summary of sounds in Unidentified samples.

3.2.1. Diurnal LULC patterns

Mann-Whitney's U tests revealed significant differences in the amount of Anthropophony, Biophony, Geophony, and Quiet during nighttime hours (8 p.m. to 5 a.m.) compared to daytime hours (5 a.m. to 8 p.m.; sample size and test results in Supplementary Tables S.15.1 and S.15.2). Anthropophony was lowest during nighttime hours on average and was approximately two times higher during the day (μ = 6.79%; U = 2.364x10^7, p < 0.001). Similar doubling patterns in activity between night and day were observed for Biophony (night = 15.4% and day = 29.5%; U = 1.917x10^7, p < 0.001), while Quiet had the opposite pattern (night = 37.3% and day = 16.6%; U = 5.032x10^7, p < 0.001). These patterns were consistent when visualizing data partitioned by LULC, but we did not test for within-LULC differences.

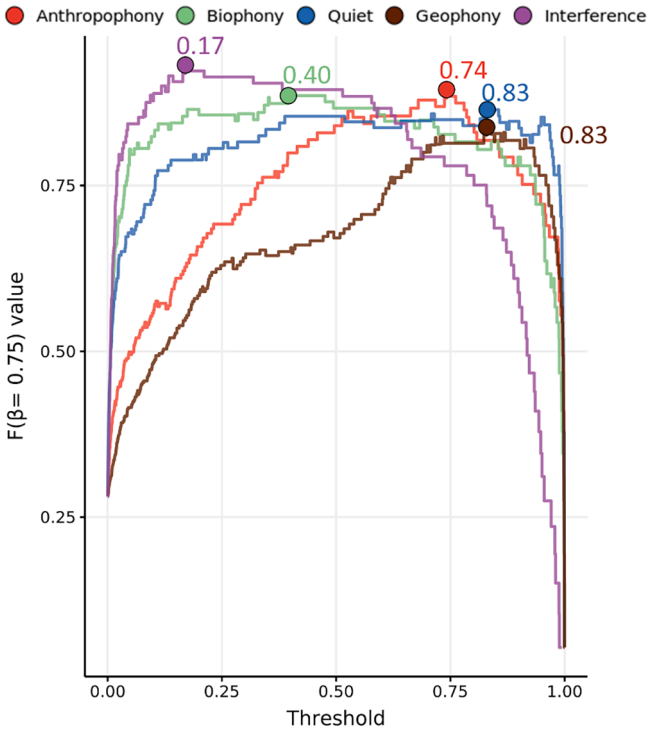


Fig. 5. Threshold optimization values for maximum F0.75-score (circles), optimized independently for each soundscape class.

Biophony peaked between 5 and 8 a.m., depending on LULC, and gradually decreased until a local daytime minimum at 3–4 p.m. for all LULCs (Fig. 6). There was a slight rise in activity at 8 p.m. for all LULC types except oak/hardwood and conifer forests, and lower activity throughout all nighttime hours to a global minimum at 4 a.m. for all LULCs. Quiet was highest from 3 to 4 a.m. and lowest (i.e., there was the highest soundscape activity) at 10 a.m., opposite Biophony and Anthropophony activity. Maximum Geophony generally occurred in the afternoon between 12 and 3 p.m., while the minimum occurred during early-morning (5–7 a.m.) and evening (8–10 p.m.). Interference was highest between 11 a.m. and 3 p.m., implying more broad frequency spikes occurred later in the day regardless of LULC type.

3.2.2. Annual and deployment date differences

Kruskal-Wallis and Dunn tests revealed significant differences (p-adj < 0.05) in Anthropophony (χ^2 = 35.049, p < 0.001), Biophony (χ^2 = 29.582, p < 0.001), and Quiet (χ^2 = 75.744, p < 0.001) among years when tested on data from only the overlapping range of annual deployment dates (May-01 to July-05; Supplementary Tables S.15.3 - S.15.5). These differences reveal year effects in our results. Limiting data to overlapping dates of deployment resulted in significant loss of observations when data were stratified by LULC or road distance (annual data decreased by: 2017 = 46.7%, 2018 = 48.3%, 2019 = 44.0%, 2020 = 17.9%; total sites = 460 / 746). We note this inter-annual variance and acknowledge that differences in start and end dates of survey campaigns among years may add to this variance. Therefore, we use the entire datasets only for analyses that do not require accounting for a year effect (i.e., sections 3.2.3–3.2.5 and 3.3).

3.2.3. Daytime LULC stratification

Kruskal-Wallis tests using daytime recordings (5 a.m. to 8 p.m.) revealed significant differences among LULC types for Anthropophony (χ^2 = 97.798, p < 0.001), Biophony (χ^2 = 18.891, p = 0.004), and Quiet (χ^2 = 109.92, p < 0.001), but not Geophony (Table S.15.6). Dunn test results showed which LULC types were significantly different within soundscape components (Fig. 7; Supplementary Table S.15.7).

There was a weak but significant difference in Biophony between herbaceous and oak/hardwood forest sites (z value = 30.745, p-adj = 0.044). Overall, the most Biophony occurred at riparian/wetland sites and the least at agriculture/barren sites. Other LULC types had comparable Biophony to each other.

The amount of Anthropophony in urban/developed sites was greater and differed significantly from herbaceous (z value = -3.163, p-adj = 0.0328), shrubland (z value = -4.506, p-adj < 0.001), oak/hardwood forest (z value = -4.993, p-adj < 0.001), and conifer forest (z value = -5.931, p-adj < 0.001) Anthropophony. We observed more Anthropophony in herbaceous and riparian/wetland sites compared to both oak/hardwood (z value = -4.579, p-adj < 0.001 and z value = -4.380, p-adj < 0.001, respectively) and conifer forests (z value = -6.818, p-adj < 0.001 and z value = -5.475, p-adj < 0.001, respectively). Agriculture/barren Anthropophony was higher than conifer forests as well (z value = -3.783, p-adj = 0.003). Hourly patterns in Anthropophony showed the most magnitude change (i.e., the largest difference between the minimum and maximum amount present) in urban/developed sites while riparian sites had a similar, but less pronounced pattern (Fig. 6). Other less human-impacted LULC types, such as oak/hardwood forest, conifer forest, shrubland, and herbaceous, showed similar patterns, but with lower overall amounts of Anthropophony, while agriculture/barren sites showed no discernable diel patterns.

Quiet was lowest and co-occurred with Anthropophony in more

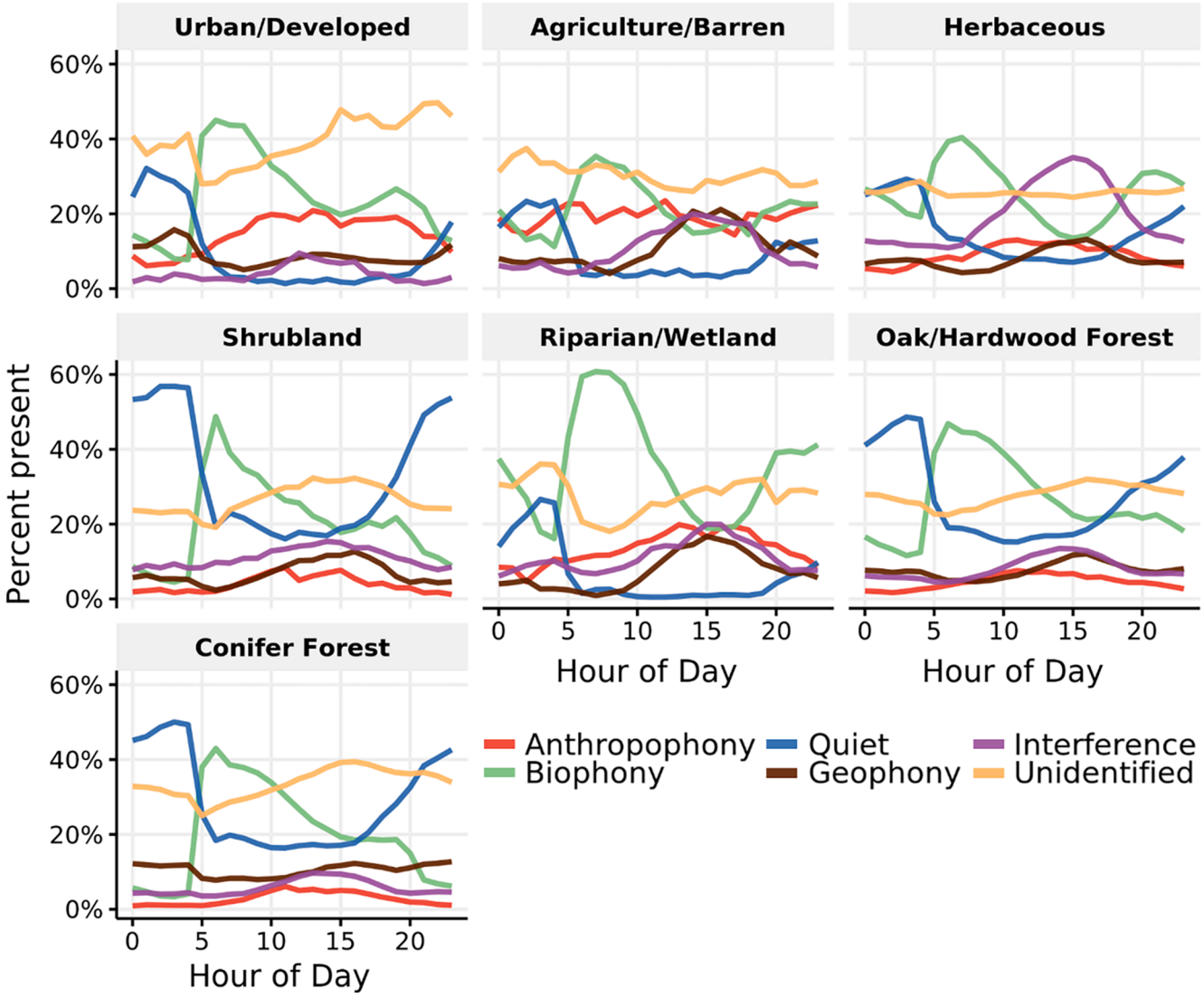


Fig. 6. Soundscape component classifications aggregated by the hour for all 746 recording sites. Sites were stratified by LULC type, and lines represent the average percent of predicted ABGQI and Unidentified over each hour interval.

human-impacted urban/developed, agriculture/barren, and riparian/wetland sites compared to overall higher amounts of Quiet in shrubland, oak/hardwood forest, and conifer forest sites, while herbaceous sites fell between these two groups (Fig. 7). When spikes of Biophony co-occurred with persistent anthropogenic noise, Quiet correspondingly decreased to near-zero (urban/developed, agriculture/barren, and riparian/wetland). Quiet never rose above 35% presence at these more human-impacted sites, whereas at less human-impacted sites, Quiet reached over 50% in the evenings and rarely dropped below 15–20% during the day.

3.2.4. Road distance

Kruskal-Wallis tests using daytime recordings (5 a.m. to 8 p.m.) were significant among road-distance classes for Anthropophony (χ^2 = 48.543, $p < 0.001$) and Quiet (χ^2 = 46.521, $p < 0.001$) and not significant for Geophony and Biophony (Fig. 8; sample size and test results in Tables S.15.8 and S.15.9). Anthropophony was highest at sites closest to roads than sites farther from roads (e.g., 0–99 m and >1000 m [z value = -5.064, p-adj < 0.001]; 100–199 m and >1000 m [z value = -3.707, p-adj = 0.012]). Similar trends existed for Quiet where sites closer to roads had less Quiet than sites farther from roads (e.g., 0–99 m and >900–999 m [z value = 3.725, p-adj = 0.011]; 0–99 m and >1000 m [z value = 4.757, p-adj < 0.001]; Supplementary Table S.15.10). We observed a higher ratio of Biophony to Anthropophony with increased road distance (Fig. 8).

3.2.5. Effect of wind speed on soundscapes

Wind speed was positively related to the amount of Geophony (R^2 = 0.032, t-value = 6.32, p = 0.072), Anthropophony (R^2 = 0.007, t-value = 2.677, p = 0.008), and Interference (R^2 = 0.020, t-value = 5.065, p < 0.001). Quiet was negatively related to wind speed (R^2 = 0.007, t-value = -3.093, p = 0.002) while Biophony did not have a significant relationship (p = 0.2336). Geophony rose throughout the afternoon, which reflects wind activity that also peaks, on average, later in the day. However, Interference also follows this pattern of rising throughout the day until an afternoon peak, and in some LULCs exceeds Geophony.

3.3. Factors affecting the amount of soundscape components

Stepwise variable selection and likelihood ratio tests resulted in a regression model containing recording year, the number of 2-s Mel spectrograms per site, soundscape component (ABGQI), LULC, road distance, and the interaction between soundscape class and road distance (Table 3; adjusted R^2 = 0.25, F(19, 3691) = 66.36, p < 0.001). The

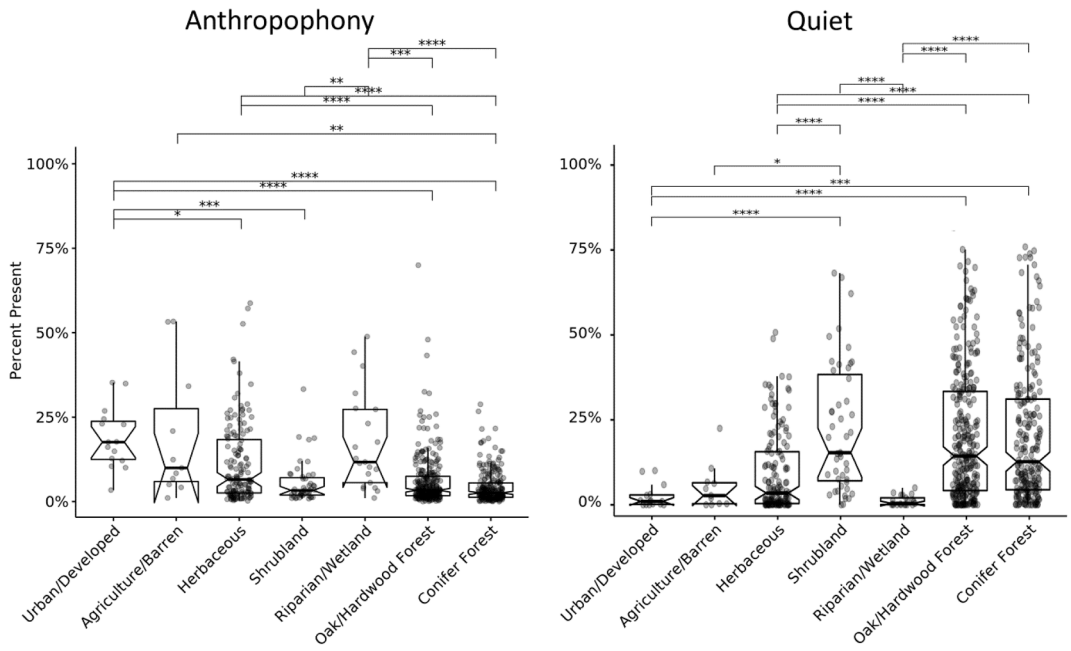


Fig. 7. Acoustic dissimilarity patterns of sites grouped by LULC for daytime recordings (5 a.m. to 8 p.m.). Significance notation of pairwise Dunn test results on the overhanging brackets, non-significant pairs not annotated, as follows: * p-adj < 0.05; ** p-adj < 0.005; *** p-adj < 0.0005; **** p-adj < 0.00005.

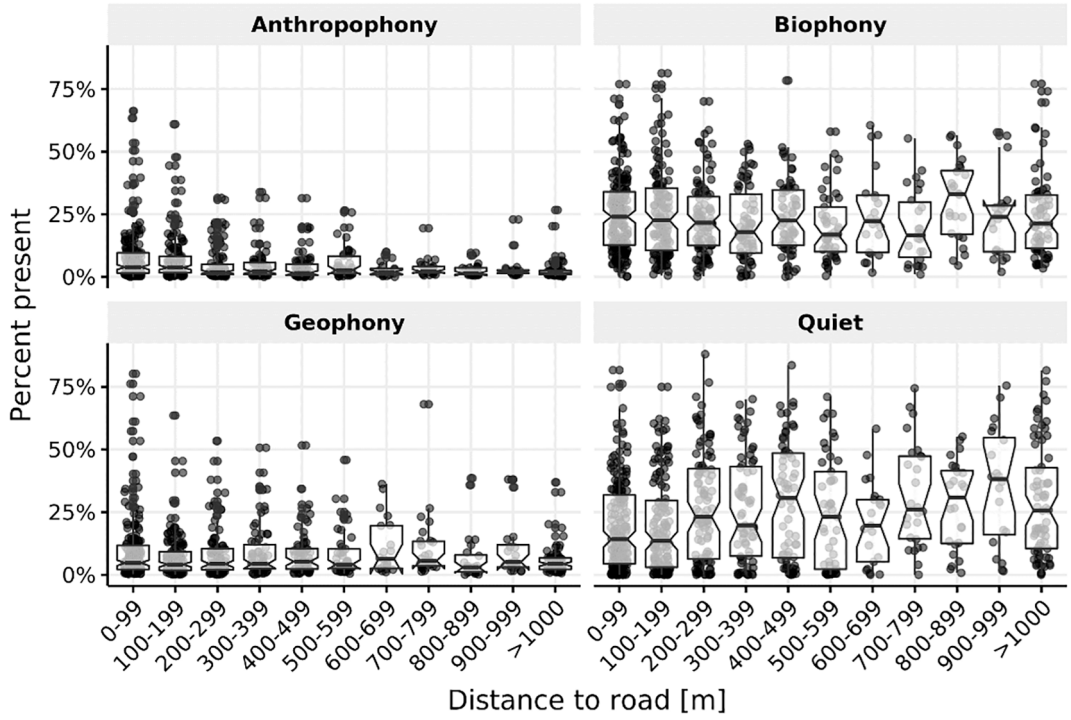


Fig. 8. Soundscape components grouped by distance from roads for daytime recordings (5 a.m. to 8p.m.). Notches reflect confidence intervals ($\pm 1.58 * IQR/\sqrt{n}$) around the median.

amount of sound increased with each progressive field season and was weakly related to the number of recordings. Anthropophony decreased the most with distance from roads, Biophony and Geophony decreased to a lesser degree, and Interference and Quiet increased (Fig. 8). LULC was not significant in determining the amount of sound, but including LULC resulted in a model with an AIC score decreased by 15.

4. Discussion

We optimized the development of a soundscape classifier with comparable accuracy to other ecoacoustic, deep learning tasks (Supplementary Materials S.17; Fairbrass et al., 2019; Ruff et al., 2021) and expanded these previous efforts to include novel soundscape classes (i.e., Geophony, Quiet, ARU Interference). Additionally, we demonstrated that broadly classified soundscape components reveal systematic acoustic patterns about time (e.g., hourly and annual) and geographic (i.

Table 3
Estimated variable coefficients (Estimate), standard errors (SE), t-values, and p-values for all parameters in the final regression model for the amount of sound. The amount of sound is the logit rate of 2-s files predicted present for each soundscape component. The reference level is the Year 2017, Anthropophony, and urban/developed. Statistically significant factors in bold (p-value < 0.05).

Variable	Estimate	SE	t-value	p-value
(Intercept)	**−3.646**	**0.684**	**−5.332**	**< 0.001**
Year18	0.049	0.098	0.502	0.615
Year19	**0.246**	**0.074**	**3.342**	**< 0.001**
Year20	**0.308**	**0.080**	**3.844**	**< 0.001**
log(number Mel spectrograms)	**0.123**	**0.062**	**1.967**	**0.049**
Biophony	**1.039**	**0.384**	**2.707**	**0.007**
Geophony	**−0.886**	**0.384**	**−2.309**	**0.021**
Interference	**−1.624**	**0.384**	**−4.232**	**< 0.001**
Quiet	**−1.104**	**0.386**	**−2.855**	**0.004**
LULC-Agric./Barren	0.090	0.273	0.329	0.742
Herbaceous	0.32	0.190	1.695	0.090
Shrubland	0.2612	0.207	1.259	0.208
Riparian/Wetland	0.203	0.239	0.846	0.398
Oak/Hardwood Forest	0.166	0.187	0.889	0.374
Conifer Forest	−0.036	0.187	−0.190	0.849
log(Road Distance)	**−0.265**	**0.049**	**−5.392**	**< 0.001**
Biophony:log(Road Distance)	**0.202**	**0.069**	**2.940**	**0.003**
Geophony:log(Road Distance)	**0.266**	**0.069**	**3.867**	**< 0.001**
Interference:log(Road Distance)	**0.350**	**0.069**	**5.090**	**< 0.001**
Quiet:log(Road Distance)	**0.528**	**0.069**	**7.624**	**< 0.001**

e., road distance, LULC, and wind) properties. We chose to first focus on understanding and communicating model assumptions to contribute to the responsible use of deep learning methods in ecoacoustics (Wearn et al., 2019). Our model can be re-trained, and methods readily extended to other environments (https://doi.org/10.5281/zenodo.6112713).

4.1. Deep learning model implementation

Creating a high-performance classification model with a limited dataset is a vital step in expanding the application of CNNs for ecology and conservation where large labeled datasets or computing resources are not available (Salamon et al., 2017; Wearn et al., 2019). We provide evidence that an existing CNN (MobileNetV2) can be pre-trained and fine-tuned with soundscape data from our study site and applied to ecoacoustic problems using a small training dataset (e.g., target classes with approximately 1,000 samples (Çoban et al., 2020)). Other studies that have classified similar soundscape components achieved similar precision and recall for Biophony and Anthropophony (Supplementary Materials S.17) but have lower accuracy or do not classify Geophony, Interference, or Quiet (e.g., Fairbrass et al., 2019; Mullet et al., 2016). We used a custom classification and threshold optimization approach to understand how predictions differed based on the choice of F-score threshold values. Using F(β)-scores with $\beta < 1.00$ allowed us to prioritize precision, which can prevent underfitting the model (Abdi and Hashemi, 2016), a known issue with small datasets and is a preferred approach in soundscape classification tasks (LeBien et al., 2020). Elusive or rare species classification would comparatively benefit by maximizing recall to avoid missing infrequent vocalizations (MacLaren et al., 2018; Shiu et al., 2020).

Ecoacoustic classification may be subject to spatial and temporal autocorrelation sources in model assessment (Ploton et al., 2020), which has not been thoroughly assessed. Recent work demonstrated spatial (Holgate et al., 2021; Shaw et al., 2021) and temporal (Scarpelli et al., 2021) autocorrelation patterns in acoustic recordings and acoustic indices. However, no work to our knowledge has investigated sources of autocorrelation in a deep learning model bias context. Future work would benefit from an investigation into sources of autocorrelation related to optimistically biased accuracy metrics in model assessment.

4.1.1. Class imbalance in training deep learning models

Small training datasets can result in bias when detecting underrepresented classes (Christin et al., 2019; Wearn et al., 2019). A class with low training samples results in poorly constrained model parameters and, therefore, poor class generalization (Wearn et al., 2019). In our case, we had low membership for Interference: n = 430 or 12.9% of the majority class size (Biophony), meaning our ABGQI-CNN may not have fully captured Interference features with the same confidence as other classes. We attempted to address the issue using a minority class oversampling method to bring the class membership of smaller classes up to the maximum membership class (Mohammed et al., 2020). This method resulted in higher performance for Interference at the cost of lower precision for larger classes, such as Biophony and Anthropophony. Because a priority of this classification task was to understand ecologically meaningful soundscape classes, we chose not to use balanced classes in training. We recommend that similar studies using CNN classification be aware of how a class imbalance in small datasets influences results, even after observing high evaluation metrics. Comparatively, if classes of interest have significantly low membership and are important (i.e., rare or cryptic species), techniques to augment or increase class membership are highly advised to reduce model bias.

4.1.2. Accounting for instrument error and sound detection

We observed non-trivial amounts of Interference (approximately 10% of recordings) and suggest acoustic studies be aware of ARU deployment techniques and instrument sensitivity to these events (e.g., strapping the unit to a tree to prevent banging in the wind). We believe rapid broad frequency interference should be appropriately accounted for and can negatively influence automated methods, for instance, that compare background noise to changes in acoustic energy (e.g., the acoustic complexity index) (Pieretti et al., 2011).

In terms of Unidentified sounds, our model performed worse in urban/developed sites with an increase late in the day and when Biophony decreased (Fig. 6). The inability of our CNN to detect all forms of sounds is partly due to our training dataset containing single-label data, not multi-label mixed-signal spectrograms, and limited samples of some Anthropophony sounds (i.e., composed of 74% vehicle traffic and machinery; Supplementary Materials S.13). Further, Anthropophony had the most Freesound samples relative to S2L samples in our training set (i.e., Freesound = 883 ROIs: S2L = 912 ROIs), possibly resulting in underfitting. The use of data from multiple ARUs and training with auxiliary Freesound data may reflect the lower performance observed in our soundscape validation (Supplementary Materials S.12) and more Unidentified samples in the S2L dataset (29.3%) relative to the test dataset (9.2%). Increasing training data specifically from the S2L dataset would decrease potential generalizability to other datasets but increase our performance. Without these data, we can lower the number of Unidentified samples and generate less conservative predictions using a more lenient threshold optimization metric like F1-score.

Variation from ARU models was considered and found insignificant in the multivariate regression of the amount of sound when we included the effects of other site characteristics. However, we observed a consistent pattern of less pixel variation in spectrograms in LG versus AudioMoth ARUs, resulting in more Quiet in 2017 and 2018, years with LG deployments (note: the CNN was trained with ROIs from both devices). We did not apply a correction but recognized that some variation in soundscape components is due to underlying variation in ARU sound detectability. To properly investigate where the effects of ARU variation originate (e.g., signal to noise ratio, sensitivity, sampling rate), a paired ARU deployment would control for landscape characteristics and possibly allow for a future correction to acoustic data from varying ARUs. We could not investigate these differences with the current dataset as ARUs were deployed at different sites with different recording lengths. Future research on soundscape differences among LULC types with variable structure and topography may provide insights into differences in sound detection among ARUs (Rappaport et al., 2020).

4.2. Soundscape patterns

4.2.1. Annual variation and recording length

The lowest Anthropophony was observed in 2020, which may reflect the reduction in human activity at a landscape scale during the onset of social lockdowns due to COVID-19, agreeing with reduced urban noise levels at a local scale (Aletta et al., 2020). Even though Biophony was highest in 2020, trends based on prior years make it difficult to say if this is related to the COVID-19 impact. Sampling effort was positively related to the amount of sound detected (Table 3), implying that extended recording times result in a higher rate of detections. This finding aligns with recent work that recommends extended ARU deployment to compute acoustic indices (Bradfer-Lawrence et al., 2019).

4.2.2. Anthropophony

Increases in Anthropophony throughout the day coincided with higher human activity (e.g., car and airplane traffic) and decreased activity in the evenings (Francis et al., 2017), reflecting our expectation that urban areas have the most human activity in Sonoma County. However, Anthropophony decreased in urban areas at nighttime to slightly above other, less human-impacted areas (e.g., riparian/wetlands and herbaceous). Decreased nighttime Anthropophony is likely because Sonoma County urban areas are mostly suburban, surrounded by rural areas, and is peripheral to more densely populated areas in the San Francisco Bay Area. Higher Anthropophony in riparian/wetland areas may be due to the location of these sites in the Laguna de Santa Rosa area of the county, which is crossed by east–west road corridors. Agriculture/barren sites, primarily vineyards in Sonoma, most likely have a large amount of anthropogenic noise associated with machinery (Lie et al., 2016), little impact from diurnal human activity patterns, and less vegetation structural and compositional complexity, which could negatively affect vocalizing animal species resulting in less Biophony (Burns et al., 2020; Dröge et al., 2021). For example, tropical ecoacoustic studies found that more structurally-complex vegetation can maintain soundscape diversity in agricultural landscapes (Dröge et al., 2021; Gage et al., 2015; Villanueva-Rivera et al., 2011). We observed persistent anthropogenic noise in agriculture/barren areas, which may be why we observe the least Biophony, particularly in the dawn chorus; however, urban areas still experience a large Biophony dawn chorus when Anthropophony is lower. We can increase Anthropophony's recall to capture more anthropogenic noise to improve our understanding of these patterns.

4.2.3. Biophony

The highest Biophony reflects the dawn chorus when increased avian and mammal activity occurs (Krause and Farina, 2016) and the dusk chorus when insect and amphibian activity occurs (Gasc et al., 2018; Krause and Farina, 2016; Naguib and Riebel, 2014). Dusk chorus Biophony is most pronounced in herbaceous and riparian/wetland sites, indicating increased insect and amphibian activity at these LULCs. Even though Anthropophony was lower in conifer forest, oak/hardwood forest, and shrubland sites, we did not observe more Biophony relative to sites with high Anthropophony, an inconsistent finding compared with prior ecoacoustic studies (Doser et al., 2019; Francis et al., 2017). However, lower Biophony in forests may be a product of detectability in denser structural landscapes (Rappaport et al., 2020).

Although Biophony did not increase with decreased human impact, we observed a shift in the relative amounts of Biophony to Anthropophony with less human impact. Most of the day, agriculture/barren and urban/developed sites were the only LULC types with comparable or higher amounts of Anthropophony relative to Biophony. Lack of variation in Biophony across LULCs indicates it is not as sensitive to LULC type as Anthropophony but can still capture the systematic dawn and dusk choruses and reflects high activity in riparian/wetland areas. Not capturing a decrease in Biophony in human-impacted areas may be due to our data not including many sites from urban centers in the county, being more oriented towards forested and lower-impacted landscapes. Notably, Biophony does not differentiate activity between, for example, a repeatedly vocalizing frog compared and a complex dawn chorus with multiple animal species vocalizing. Future work can utilize Biophony as a pre-filter for more refined taxonomic group classification or species with sufficiently labeled data.

4.2.4. Quiet

Naturally quiet landscapes are spaces where anthropogenic noise is absent, and invasive biophonic species are minimal (Dumyahn and Pijanowski, 2011; Pavan, 2017). Quiet places have a range of benefits, such as increased wildlife reproductive success and human well-being, highlighting the need to conserve and maintain naturally quiet landscapes (Buxton et al., 2019; Dumyahn and Pijanowski, 2011). We are limited in making inferences regarding naturally quiet landscapes because we chose to set the upper frequency of analyses to 11 kHz. This choice was made to extend the ABGQI-CNN to other ecoacoustic work, such as examining the relationships between soundscape components and acoustic indices. Many acoustic indices are limited in their upper-frequency range from 8 to 11 kHz (e.g., Boelman et al., 2007; Kasten et al., 2012). This frequency range allows indices to capture numerous wildlife vocalizations (e.g., frogs, crickets, and birds [0.2–8 kHz]; Villanueva-Rivera et al., 2011) and anthropogenic noise (typically 0–2 kHz; Joo et al., 2011) while prior work has shown a significant amount of environmental acoustic activity occurs below 9–12 kHz (Metcalf et al., 2021; Pavan, 2017; Towsey, 2013). There may be other signals above 11 kHz (i.e., insects and bats) that our modeling efforts did not reflect, which would result in lower amounts of Quiet and higher Biophony. However, Anthropophony, Geophony, Interference, and a significant amount of Biophony occurred below this frequency threshold; therefore, we interpret times of Quiet to reflect periods with generally lower acoustic activity.

Quiet is most usefully interpreted alongside Biophony and Anthropophony, the former contributing to naturally quiet landscapes while the latter deteriorates these landscapes. Quiet was highest at night when human, weather, and most biotic activity were lowest (Mullet et al., 2017b). Conversely, the loudest time of day coincided with the dawn bird chorus and times of persistent anthropogenic activity (e.g., commuter and air traffic). Nevertheless, even though Biophony is negatively related to Quiet, biotic noises tend to be more culturally valuable and have positive effects on human well-being than human noise intrusion (Dumyahn and Pijanowski, 2011; Krause, 2002). Although Biophony was consistent across LULC types, increases in Anthropophony in currently less-impacted LULC types could negatively affect wildlife communities by increasing fitness costs for species (Francis and Barber, 2013; Mullet et al., 2017a), which could have cascading adverse effects on ecosystems. Deep learning classifiers with finer taxonomic groups or species could measure animal community composition not captured by our general Biophony class and potentially reveal differences between naturally quiet and noisier landscapes impacted by human activity and highlight species that may be robust against noise pollution (Slabbekoorn and Ripmeester, 2008). Future data collection can be aided by identifying times or spaces with increased Quiet to ensure ARUs are deployed with a higher likelihood to record ecologically meaningful signals (e.g., after 5 a.m. in conifer forests).

4.2.5. Geophony and Interference

We observed systematic afternoon increases in Geophony in agriculture/barren and riparian/wetlands with less pronounced patterns in other LULC types. In general, we observed relatively low Geophony, and our data may not have captured rain events as recording seasons begin at the onset of Sonoma County's dry season in May. Geophony predominantly reflects wind patterns, with low confidence classifying streams or rain. We believe Interference may serve as a proxy for wind-related Geophony activity based on (1) a weak but correlated pattern with

Geophony from the Pepperwood Preserve analysis and (2) evidence from training set creation where we frequently observed Interference events co-occurring with gusts of wind. If we follow this hypothesis, combining Geophony and Interference signals reflects afternoon peaks in sound activity related to meteorological patterns in the wind (Fig. 6). This additive pattern is evident in less structurally-complex LULCs (agriculture/barren, herbaceous, shrubland, riparian/wetland), most notably in herbaceous sites. In herbaceous sites, microphones were attached to small temporary poles and were subject to wind more than other LULC types, where ARUs were generally mounted on larger tree stems. At sites where the deployment of ARUs results in minimal physical shielding, wind-based Interference can be non-trivial and lead to numerous occurrences of Interference noise events. Interference requires proper identification and consideration in future acoustic analyses (e.g., acoustic indices), so events are not treated as ecologically meaningful signals and can be used to confirm the occurrence of wind. The ABGQI-CNN can reliably identify these recording errors.

4.3. Soundscape patterns related to roadway proximity

Roads are fixed sources of anthropogenic noise, and extensive research has documented the deleterious effects of traffic on animal communities (Barber et al., 2011; Buxton et al., 2019; Doser et al., 2019; Ware et al., 2015). Ecoacoustic work has examined patterns in sound relative to traffic noise and found that biotic sound activity is inversely correlated with traffic intensity and distance to roads (Doser et al., 2019; Pieretti and Farina, 2013). However, these studies used either acoustic indices (Pieretti and Farina, 2013), approximated traffic noise from geographic road use data (Barber et al., 2011; Doser et al., 2019), or required intensive manual categorization of spectrograms (Buxton et al., 2019) to quantify traffic levels. Instead, we were able to quantify distinct patterns in anthropogenic noise related to distance from roads among LULCs, demonstrating the benefits of deep learning classification coupled with geospatial analysis.

Our linear regression modeling results support general findings that anthropogenic noise decreases with distance from roads (Buxton et al., 2019; Mullet et al., 2016), but we do not find there is a significant inverse pattern with Biophony as other studies note (Doser et al., 2019), except when we look at LULC types individually (shrubland and oak/hardwood forest areas). Sites are rarely more than 1,500 m from major roads ($n = 26$, 3.5% of all sites), and there appears to be a heavy influence on patterns from these distant sites on analytical trends. At these distances, topographic position (e.g., on a hilltop or in a valley) may have a more significant effect on soundscapes than sites closer to roads that do not experience high levels of sound attenuation (Lyon, 1973; Yip et al., 2017). Anthropophony in riparian/wetland sites was similar to urban/developed and agriculture/barren sites; however, site distance from roads is significantly higher for riparian (492 ± 392 m, $n = 21$) than urban/developed (40 ± 20 m, $n = 15$) or any other LULC type. Higher Anthropophony in these riparian/wetland sites may be due to proximity to principal road corridors in the Laguna de Santa Rosa area of the county where sound can travel farther along less vegetative and morphologically complex riparian/wetland corridors (Wiley and Richards, 1978).

These findings indicate that we can detect general soundscape patterns relative to distance from roads. However, to better explain the causation of these patterns and relate soundscape classes to biodiversity, we need to incorporate other spatial characteristics of the landscape, such as topography, forest structure, climate, and other human impact layers in a more holistic modeling framework. Urban and environmental planning related to road construction can benefit from quantifying meaningful soundscape components (e.g., ABQ) with these or similar methods prior to, during, and following development to measure the effects of increased human activity on naturally quiet landscapes.

5. Conclusion

Autonomous sound recording is becoming a more prominent tool for ecological monitoring. However, the abundance of acoustic data requires analytical approaches to account for error and non-biotic sounds. The ability to rapidly train and deploy a classifier to accurately identify almost 70% of a large acoustic dataset with 93% precision, as we found here, enables ecoacoustic researchers to study broad patterns and interactions of sounds within a soundscape. Identified soundscape components can serve as promising ecoacoustic indicators that: (1) reflect temporal and environmental factors that can be used to limit noisy human activities to times when the impact on wildlife is minimized, (2) aid in conservation and management efforts to prioritize at-risk landscapes, and (3) optimize recorder deployment to capture ecologically-meaningful acoustic signals. Furthermore, the ABGQI-CNN and these data can help filter wanted or unwanted sounds to optimize sound monitoring and discriminate meaningful acoustic events when applying acoustic indices, which can be impacted by the presence of Anthropophony, Geophony, or Interference, reducing their measurement value. In summary, we have shown that it is possible to identify ARU error with Interference, quantify areas rich in vocalizing animal activity with Biophony, understand variations in human noise using Anthropophony, and identify quieter landscapes with data products generated from our ABGQI-CNN modeling approach.

CRediT authorship contribution statement

Colin A. Quinn: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. **Patrick Burns:** Data curation, Methodology, Writing – review & editing. **Gurman Gill:** Formal analysis, Methodology, Writing – review & editing. **Shrishail Baligar:** Data curation, Methodology, Software, Writing – review & editing. **Rose L. Snyder:** Data curation, Project administration, Writing – review & editing. **Leonardo Salas:** Data curation, Funding acquisition, Methodology, Project administration, Writing – review & editing. **Scott J. Goetz:** Funding acquisition, Resources, Supervision, Writing – review & editing. **Matthew L. Clark:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The Soundscapes to Landscapes project was funded by NASA's Citizen Science for Earth Systems Program 16-CSESP 2016-0009 under cooperative agreement 80NSSC18M0107. We are grateful to the hundreds of citizen scientists who participated in Soundscapes to Landscapes between 2017-2021. The following citizen scientists are recognized by name for each of their expert contributions of over 100 volunteer hours spent working on the project: Wendy Schackwitz (777 hours), David Leland (702 hours), Taylour Stephens (279 hours), Jade Spector (208 hours), Tiffany Erickson (190 hours), Teresa Tuffli (140 hours), Miles Tuffli (129 hours), Katie Clas (121 hours), Bob Hasenick (119 hours). This work benefited from the Intellichirp Northern Arizona University capstone team, who helped develop the S2L ABGQI ROI dataset (Steven Enriquez, Josh Kruse, Michael Ewers, Zhenyu Lei) and the Sonoma State University Koret team who helped with early model development (Alex Dewey, Antone Silveria, Jonathan Calderon Chavez, Vincent Valenzuela). We want to thank Pepperwood Preserve for the generous site access across multiple field seasons and for providing

meteorological data. We ran computational analyses on Northern Arizona University's Monsoon computing cluster, funded by Arizona's Technology and Research Initiative Fund.

Data and Materials Availability

Code for the ABGQI-CNN and to reproduce results: https://doi.org/10.5281/zenodo.6112713. Supplementary Materials and associated data: https://doi.org/10.5281/zenodo.6027024.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2022.108831.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Rafal Jozefowicz, Y.J., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Schuster, M., Monga, R., Moore, S., Murray, D., Olah, C., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vijay Vasudevan, F.V., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Abdi, L., Hashemi, S., 2016. To combat multi-class imbalanced problems by means of over-sampling techniques. IEEE Trans. Knowl. Data Eng. 28, 238–251. https://doi.org/10.1109/TKDE.2015.2458858.

Aletta, F., Oberman, T., Mitchell, A., Tong, H., Kang, J., 2020. Assessing the changing urban sound environment during the COVID-19 lockdown period using short-term acoustic measurements. Noise Mapp. 7, 123–134. https://doi.org/10.1515/noise-2020-0011.

Araya-Salas, M., Smith-Vidaurre, G., 2017. warbleR: an r package to streamline analysis of animal acoustic signals. Methods Ecol. Evol. 8, 184–191.

Balantic, C.M., Donovan, T.M., 2020. Statistical learning mitigation of false positives from template-detected data in automated acoustic wildlife monitoring. Bioacoustics 29, 296–321. https://doi.org/10.1080/09524622.2019.1605309.

Barber, J.R., Burdett, C.L., Reed, S.E., Warner, K.A., Formichella, C., Crooks, K.R., Theobald, D.M., Fristrup, K.M., 2011. Anthropogenic noise exposure in protected natural areas: Estimating the scale of ecological consequences. Landsc. Ecol. 26, 1281–1295. https://doi.org/10.1007/s10980-011-9646-7.

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

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. Ecol. Appl. 17, 2137–2144. https://doi.org/10.1890/07-0004.1.

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 Ecol. Evol. 00, 1–12. https://doi.org/10.1111/2041-210X.13254.

Burns, P., Clark, M., Salas, L., Hancock, S., Leland, D., Jantz, P., Dubayah, R., Goetz, S.J., 2020. Incorporating canopy structure from simulated GEDI lidar into bird species distribution models. Environ. Res. Lett. 15 https://doi.org/10.1088/1748-9326/ab80ee.

Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., Martius, C., Zlinszky, A., Calvignac-Spencer, S., Cobbold, C.A., Dawson, T.P., Emerson, B.C., Ferrier, S., Gilbert, M., Thomas, P., Herold, M., Jones, L., Leendertz, F.H., Matthews, L., Millington, J.D.A., Olson, J.R., Ovaskainen, O., Raffaelli, D., Reeve, R., Rödel, M.-O., Rodgers, T.W., Snape, S., Visseren-Hamakers, I., Vogler, A.P., White, P.C.L., Wooster, M.J., Yu, D.W., 2018. The Promise and Practice of Connecting Earth Observation to Biodiversity and Ecosystem Services. Nat. Ecol. Evol. doi: 10.1038/s41559-017-0176.

Buxton, R.T., McKenna, M.F., Mennitt, D., Brown, E., Fristrup, K., Crooks, K.R., Angeloni, L.M., Wittemyer, G., 2019. Anthropogenic noise in US national parks – sources and spatial extent. Front. Ecol. Environ. 559–564 https://doi.org/10.1002/fee.2112.

Çoban, E. B., Pir, D., So, R., & Mandel, M. I. (2020, May). Transfer Learning from Youtube Soundtracks to Tag Arctic Ecoacoustic Recordings. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 726-730). IEEE.

Christin, S., Hervet, É., Lecomte, N., 2019. Applications for deep learning in ecology. Methods Ecol. Evol. 10, 1632–1644. https://doi.org/10.1111/2041-210X.13256.

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

Desjonquères, C., Rybak, F., Castella, E., Llusia, D., Sueur, J., 2018. Acoustic communities reflects lateral hydrological connectivity in riverine floodplain similarly to macroinvertebrate communities. Sci. Rep. 8, 1–11. https://doi.org/10.1038/s41598-018-31798-4.

Doser, J.W., Hannam, K.M., Finley, A.O., 2019. Characterizing functional relationships between technophony and biophony: A western New York soundscape case study. Landsc. Ecol. 2 https://doi.org/10.1007/s10980-020-00973-2.

Dröge, S., Martin, D.A., Andriafanomezantsoa, R., Burivalova, Z., Fulgence, T.R., Osen, K., Rakotomalala, E., Schwab, D., Wurz, A., Richter, T., Kreft, H., 2021. Listening to a changing landscape: Acoustic indices reflect bird species richness and plot-scale vegetation structure across different land-use types in north-eastern Madagascar. Ecol. Indic. 120 https://doi.org/10.1016/j.ecolind.2020.106929.

Duchac, L.S., Lesmeister, D.B., Dugger, K.M., Ruff, Z.J., Davis, R.J., 2020. Passive acoustic monitoring effectively detects Northern Spotted Owls and Barred Owls over a range of forest conditions. Condor 122, 1–22. https://doi.org/10.1093/condor/duaa017.

Dumyahn, S.L., Pijanowski, B.C., 2011. Soundscape conservation. Landsc. Ecol. 26, 1327–1344. https://doi.org/10.1007/s10980-011-9635-x.

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

Fairbrass, A.J., Firman, M., Williams, C., Brostow, G.J., Titheridge, H., Jones, K.E., 2019. CityNet—Deep learning tools for urban ecoacoustic assessment. Methods Ecol. Evol. 10, 186–197. https://doi.org/10.1111/2041-210X.13114.

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

Farina, A., Pieretti, N., 2014. Sonic environment and vegetation structure: a methodological approach for a soundscape analysis of a Mediterranean maqui. Ecol. Inform. 21, 120–132. https://doi.org/10.1016/j.ecoinf.2013.10.008.

Ferrell, R.M., Comendant, T., Micheli, E., Dodge, C., Stern, M., Flint, L., Flint, A., Neville., J.A., 2021a. Pepperwood Long-Term Soil and MET Data – Oak and Grass Stations. Environmental Data Initiative. Retrieved from https://pasta.lternet.edu/package/eml/edi/943/1.

Ferrell, R.M., Comendant, T., Micheli, E., Neville, J.A., 2021b. Pepperwood MET soil moisture sites 2019 - 2021. Environmental Data Initiative. Retrieved from https://pasta.lternet.edu/package/eml/edi/865/1.

Francis, C.D., Barber, J.R., 2013. A framework for understanding noise impacts on wildlife: an urgent conservation priority. Front. Ecol. Environ. 11, 305–313. https://doi.org/10.1890/120183.

Francis, C.D., Newman, P., Taff, B.D., White, C., Monz, C.A., Levenhagen, M., Petrelli, A. R., Abbott, L.C., Newton, J., Burson, S., Cooper, C.B., Fristrup, K.M., McClure, C.J. W., Mennitt, D., Giamellaro, M., Barber, J.R., 2017. Acoustic environments matter: synergistic benefits to humans and ecological communities. J. Environ. Manage. 203, 245–254. https://doi.org/10.1016/j.jenvman.2017.07.041.

Furumo, P.R., Aide, M.T., 2019. Using soundscapes to assess biodiversity in Neotropical oil palm landscapes. Landsc. Ecol. 34, 911–923. https://doi.org/10.1007/s10980-019-00815-w.

Gage, S.H., Joo, W., Kasten, E.P., Fox, J., Biswas, S., 2015. Acoustic observations in agricultural landscapes. Ecol. Agric. landscapes long-term Res. path to Sustain. 360–377.

Gasc, A., Gottesman, B.L., Francomano, D., Jung, J., Durham, M., Mateljak, J., Pijanowski, B.C., 2018. Soundscapes reveal disturbance impacts: biophonic response to wildfire in the Sonoran Desert Sky Islands. Landsc. Ecol. 33, 1399–1415. https://doi.org/10.1007/s10980-018-0675-3.

Gordon, T.A.C., Harding, H.R., Wong, K.E., Merchant, N.D., Meekan, M.G., McCormick, M.I., Radford, A.N., Simpson, S.D., 2018. Habitat degradation negatively affects auditory settlement behavior of coral reef fishes. Proc. Natl. Acad. Sci. U. S. A. 115, 5193–5198. https://doi.org/10.1073/pnas.1719291115.

Grant, P.B.C., Samways, M.J., 2016. Use of ecoacoustics to determine biodiversity patterns across ecological gradients. Conserv. Biol. 30, 1320–1329. https://doi.org/10.1111/cobi.12748.

Hill, A.P., Prince, P., Piña Covarrubias, E., Doncaster, C.P., Snaddon, J.L., Rogers, A., 2018. AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. Methods Ecol. Evol. 9, 1199–1211. https://doi.org/10.1111/2041-210X.12955.

Holgate, B., Maggini, R., Fuller, S., 2021. Mapping ecoacoustic hot spots and moments of biodiversity to inform conservation and urban planning. Ecol. Indic. 126 https://doi.org/10.1016/j.ecolind.2021.107627.

Joo, W., Gage, S.H., Kasten, E.P., 2011. Analysis and interpretation of variability in soundscapes along an urban-rural gradient. Landsc. Urban Plan. 103, 259–276. https://doi.org/10.1016/j.landurbplan.2011.08.001.

Kahl, S., 2020. Identifying Birds by Sound: Large-scale Acoustic Event Recognition for Avian Activity Monitoring. Technische Universitat Chemnitz.

Kahl, S., Wilhelm-Stein, T., Klinck, H., Kowerko, D., Eibl, M., 2018. Recognizing Birds from Sound - The 2018 BirdCLEF Baseline System. arXiv preprint arXiv:1804.07177.

Kahl, S., Wood, C.M., Eibl, M., Klinck, H., 2021. BirdNET: a deep learning solution for avian diversity monitoring. Ecol. Inform. 61, 101236 https://doi.org/10.1016/j.ecoinf.2021.101236.

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. Ecol. Inform. 12, 50–67. https://doi.org/10.1016/j.ecoinf.2012.08.001.

Knight, E.C., Hannah, K.C., Foley, G.J., Scott, C.D., Brigham, R.M., Bayne, E., 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. Avian Conserv. Ecol. 12, art14. https://doi.org/10.5751/ACE-01114-120214.

Krause, B., 2002. The loss of natural soundscapes. Earth Isl. J. 17, 27–29.

Krause, B., Farina, A., 2016. Using ecoacoustic methods to survey the impacts of climate change on biodiversity. Biol. Conserv. 195, 245–254. https://doi.org/10.1016/j.biocon.2016.01.013.

LeBien, J., Zhong, M., Campos-Cerqueira, M., Velev, J.P., Dodhia, R., Ferres, J.L., Aide, T. M., 2020. A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. Ecol. Inform. 59, 101113 https://doi.org/10.1016/j.ecoinf.2020.101113.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Lellouch, L., Pavoine, S., Jiguet, F., Glotin, H., Sueur, J., 2014. Monitoring temporal change of bird communities with dissimilarity acoustic indices. Methods Ecol. Evol. 5, 495–505. https://doi.org/10.1111/2041-210X.12178.

Lie, A., Skogstad, M., Johannessen, H.A., Tynes, T., 2016. Occupational noise exposure and hearing: a systematic review. Int. Arch. Occup. Environ. Health 89, 351–372. https://doi.org/10.1007/s00420-015-1083-5.

Lin, T.H., Tsao, Y., 2020. Source separation in ecoacoustics: a roadmap towards versatile soundscape information retrieval. Remote Sens. Ecol. Conserv. 6, 236–247. https://doi.org/10.1002/rse2.141.

Lyon, R.H., 1973. Propagation of Environmental Noise: More theoretical and experimental work could permit the prediction and subsequent control of environmental noise. Science 179 (4078), 1083–1090.

MacLaren, A.R., McCracken, S.F., Forstner, M.R.J., 2018. Development and Validation of Automated Detection Tools for Vocalizations of Rare and Endangered Anurans 9, 144–154. doi: 10.3996/052017-JFWM-047.

McFee, B., Raffel, C., Liang, D., Ellis, D.P.W., McVicar, M., Battenberg, E., Nieto, O., 2015. librosa: Audio and music signal analysis in python. In Proceedings of the 14th python in science conference, 18-25.

Metcalf, O.C., Barlow, J., Devenish, C., Marsden, S., Berenguer, E., Lees, A.C., 2021. Acoustic indices perform better when applied at ecologically meaningful time and frequency scales. Methods Ecol. Evol. 12, 421–431. https://doi.org/10.1111/2041-210X.13521.

Metcalf, O.C., Lees, A.C., Barlow, J., Marsden, S.J., Devenish, C., 2020. hardRain: An R package for quick, automated rainfall detection in ecoacoustic datasets using a threshold-based approach. Ecol. Indic. 109, 105793 https://doi.org/10.1016/j.ecolind.2019.105793.

Mohammed, R., Rawashdeh, J., Abdullah, M., 2020. Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. 2020 11th Int. Conf. Inf. Commun. Syst. ICICS 2020 243–248. doi: 10.1109/ICICS49469.2020.239556.

Mullet, T.C., Farina, A., Gage, S.H., 2017a. The acoustic habitat hypothesis: an ecoacoustics perspective on species habitat selection. Biosemiotics 10, 319–336. https://doi.org/10.1007/s12304-017-9288-5.

Mullet, T.C., Gage, S.H., Morton, J.M., Huettmann, F., 2016. Temporal and spatial variation of a winter soundscape in south-central Alaska. Landsc. Ecol. 31, 1117–1137. https://doi.org/10.1007/s10980-015-0323-0.

Mullet, T.C., Morton, J.M., Gage, S.H., Huettmann, F., 2017b. Acoustic Footprint of Snowmobile Noise and Natural Quiet Refugia in an Alaskan Wilderness. Nat. Areas J. 37, 332–349. https://doi.org/10.3375/043.037.0308.

Naguib, M., Riebel, K., 2014. Singing in space and time: the biology of birdsong. doi: 10.1007/978-94-007-7414-8.

Newport, J., Shorthouse, D.J., Manning, A.D., 2014. The effects of light and noise from urban development on biodiversity: Implications for protected areas in Australia. Ecol. Manag. Restor. 15, 204–214. https://doi.org/10.1111/emr.12120.

Pavan, G., 2017. Fundamentals of Soundscape Conservation. Ecoacoustics Ecol. Role Sounds 235–258. https://doi.org/10.1002/9781119230724.ch14.

Pérez-Granados, C., Traba, J., 2021. Estimating bird density using passive acoustic monitoring: a review of methods and suggestions for further research. Ibis (Lond. 1859). 1–19. https://doi.org/10.1111/ibi.12944.

Piczak, K.J., 2015. Environmental sound classification with convolutional neural networks. IEEE Int. Work. Mach. Learn. Signal Process. MLSP 2015-November. doi: 10.1109/MLSP.2015.7324337.

Pieretti, N., Farina, A., 2013. Application of a recently introduced index for acoustic complexity to an avian soundscape with traffic noise. J. Acoust. Soc. Am. 134, 891–900. https://doi.org/10.1121/1.4807812.

Pieretti, N., Farina, A., Morri, D., 2011. A new methodology to infer the singing activity of an avian community: the Acoustic Complexity Index (ACI). Ecol. Indic. 11, 868–873. https://doi.org/10.1016/j.ecolind.2010.11.005.

Pijanowski, B.C., Farina, A., Gage, S.H., Dumyahn, S.L., Krause, B.L., 2011. What is soundscape ecology? An introduction and overview of an emerging new science. Landsc. Ecol. 26, 1213–1232. https://doi.org/10.1007/s10980-011-9600-8.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. 11, 1–11. https://doi.org/10.1038/s41467-020-18321-y.

Python Software Foundation. (2016). Python Language Reference. Retrieved from http://www.python.org.

Rappaport, D.I., Royle, J.A., Morton, D.C., 2020. Acoustic space occupancy: Combining ecoacoustics and lidar to model biodiversity variation and detection bias across heterogeneous landscapes. Ecol. Indic. 113, 106172 https://doi.org/10.1016/j.ecolind.2020.106172.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Rice, W.L., Newman, P., Miller, Z.D., Taff, B.D., 2020. Protected areas and noise abatement: a spatial approach. Landsc. Urban Plan. 194, 103701 https://doi.org/10.1016/j.landurbplan.2019.103701.

Rose, S.J., Allen, D., Noble, D., Clarke, J.A., 2018. Quantitative analysis of vocalizations of captive Sumatran tigers (Panthera tigris sumatrae). Bioacoustics 27, 13–26. https://doi.org/10.1080/09524622.2016.1272003.

Ruff, Z.J., Lesmeister, D.B., Appel, C.L., Sullivan, C.M., 2021. Workflow and convolutional neural network for automated identification of animal sounds. Ecol. Indic. 124, 107419 https://doi.org/10.1016/j.ecolind.2021.107419.

Salamon, J., Bello, J.P., 2017. Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Signal Process. Lett. 24, 279–283. https://doi.org/10.1109/LSP.2017.2657381.

Salamon, J.J., Bello, J.P., Farnsworth, A., Kelling, S., 2017. Fusing shallow and deep learning for bioacoustic bird species classification. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 141–145.

Sánchez-Giraldo, C., Bedoya, C.L., Morán-Vásquez, R.A., Isaza, C.V., Daza, J.M., 2020. Ecoacoustics in the rain: understanding acoustic indices under the most common geophonic source in tropical rainforests. Remote Sens. Ecol. Conserv. 6, 248–261. https://doi.org/10.1002/rse2.162.

Scarpelli, M.D.A., Liquet, B., Tucker, D., Fuller, S., Roe, P., 2021. Multi-index ecoacoustics analysis for terrestrial soundscapes: a new semi-automated approach using time-series motif discovery and random forest classification. Front. Ecol. Evol. 9, 1–14. https://doi.org/10.3389/fevo.2021.738537.

Schafer, R.M., 1993. The Soundscape: Our Sonic Environment and the Tuning of the World. Simon and Schuster.

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. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.2004702117.

Shaw, T., Hedes, R., Sandstrom, A., Ruete, A., Hiron, M., Hedblom, M., Eggers, S., Mikusiński, G., 2021. Hybrid bioacoustic and ecoacoustic analyses provide new links between bird assemblages and habitat quality in a winter boreal forest. Environ. Sustain. Indic. 11 https://doi.org/10.1016/j.indic.2021.100141.

Shiu, Y., Palmer, K.J., Roch, M.A., Fleishman, E., Liu, X., Nosal, E., 2020. Use of deep neural networks for automated detection of marine mammal species 1–29. doi: 10.1038/s41598-020-57549-y.

Shonfield, J., Bayne, E.M., 2017. Autonomous recording units in avian ecological research: current use and future applications. Avian Conserv. Ecol. 12 https://doi.org/10.5751/ace-00974-120114.

Slabbekoorn, H., Ripmeester, E.A.P., 2008. Birdsong and anthropogenic noise: implications and applications for conservation. Mol. Ecol. 17, 72–83. https://doi.org/10.1111/j.1365-294X.2007.03487.x.

Southworth, M., 1969. The sonic environment of cities. Environ. Behav. 1, 22.

Sueur, J., Farina, A., Gasc, A., Pieretti, N., Pavoine, S., 2014. Acoustic indices for biodiversity assessment and landscape investigation. Acta Acust. United with Acust. 100, 772–781. https://doi.org/10.3813/AAA.918757.

Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., 2008. Rapid acoustic survey for biodiversity appraisal. PLoS One 3, 1–10. https://doi.org/10.1371/journal.pone.0004065.

Towsey, M., 2013. Noise removal from waveforms and spectrograms derived from natural recordings of the environment.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, fourth ed. Springer, New York.

Villanueva-Rivera, L.J., Pijanowski, B.C., Doucette, J., Pekin, B., 2011. A primer of acoustic analysis for landscape ecologists. Landsc. Ecol. 26, 1233–1246. https://doi.org/10.1007/s10980-011-9636-9.

Ware, H.E., McClure, C.J.W., Carlisle, J.D., Barber, J.R., Daily, G.C., 2015. A phantom road experiment reveals traffic noise is an invisible source of habitat degradation. Proc. Natl. Acad. Sci. U. S. A. 112, 12105–12109. https://doi.org/10.1073/pnas.1504710112.

Wearn, O.R., Freeman, R., Jacoby, D.M.P., 2019. Responsible AI for conservation. Nat. Mach. Intell. 1, 72–73. https://doi.org/10.1038/s42256-019-0022-7.

Wiley, R.H., Richards, D.G., 1978. Physical constraints on acoustic communication in the atmosphere: implications for the evolution of animal vocalizations. Behav. Ecol. Sociobiol. 3, 69–94.

Yip, D.A., Bayne, E.M., Sólymos, P., Campbell, J., Proppe, D., 2017. Sound attenuation in forest and roadside environments: Implications for avian point-count surveys. Condor 119, 73–84. https://doi.org/10.1650/CONDOR-16-93.1.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? Adv. Neural Inf. Process. Syst. 4, 3320–3328 arXiv preprint arXiv:1411.1792.

Zhong, M., LeBien, J., Campos-Cerqueira, M., Dodhia, R., Lavista Ferres, J., Velev, J.P., Aide, T.M., 2020. Multispecies bioacoustic classification using transfer learning of deep convolutional neural networks with pseudo-labeling. Appl. Acoust. 166, 107375 https://doi.org/10.1016/j.apacoust.2020.107375.