Ecological Indicators 144 (2022) 109437

Contents lists available at ScienceDirect

Ecological Indicators

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

Assessing the effect of human activities on biophony in urban forests using an automated acoustic scene classification model

Zezhou Hao[a], Haisong Zhan[b], Chengyun Zhang[b,*], Nancai Pei[a], Bing Sun[a], Jihong He[a,*], Ruichen Wu[a], Xinhui Xu[c], Cheng Wang[d]

[a] Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou 510520, China
[b] School of Electronics and Communication Engineering, Guangzhou University, Guangzhou 510006, China
[c] Guangdong Eco-Engineering Polytechnic, Guangzhou 510520, China
[d] Research Institute of Forestry, Chinese Academy of Forestry/Key Laboratory of Tree Breeding and Cultivation, National Forestry and Grassland Administration/ Urban Forest Research Center, National Forestry and Grassland Administration, Beijing 100091, China

ARTICLE INFO

Keywords:
Acoustic scenes
Bioacoustic
Biodiversity
Human activity
Machine learning
Urban noise

ABSTRACT

Monitoring biodiversity and assessing the impact of human activities using acoustics is a promising area in the field of urban ecology. Previous studies on urban biodiversity using acoustics are often limited by data continuity and survey scope, making it difficult to answer questions about relationships between bird population dynamics and environmental factors. To some extent, big data methods such as continuous acoustic monitoring have bridged this gap and provided new research paths to address the problem. In this study, we proposed a machine learning (ML) method that uses convolutional neural networks (CNN) and target sound area ratios (TSAR) to quantify the dominance of seven types of acoustic scenes. Acoustic data was recorded at nine sites in three urban forests in Guangzhou, China. Using the site-related sound data, we trained the convolutional neural network and identified seven target soundscape components with an overall F1 score of 0.97, a precision of 0.96, and a recall of 0.97. Spatial patterns of acoustic scenes in urban forests were examined to understand the effective working radius of monitoring equipment and the impacts of differing land use types on the composition of soundscapes. This study indicates significant interactions between human activities and biodiversity using acoustics, demonstrates that vocal organisms respond to environmental changes primarily by changes in their vocal frequencies, and proposes a novel framework for utilizing acoustics to monitor urban biodiversity. Going forth, these analyses help to promote the conservation of biodiversity and the sustainability of urban development.

1. Introduction

Human activities have touched every ecosystem on the planet (Masood, 2018). Biodiversity in urban areas, in particular, is impacted by land use change, habitat loss, and pollution associated with the development and support of high-density human populations (da Silva et al., 2020; Isbell et al., 2017). Given that urban population centers are expanding inexorably across the planet, it is increasingly necessary to improve sustainable development by advancing both our monitoring capabilities and understanding of the impacts of human activities on biodiversity (Soranno and Schimel, 2014). The advent of the big data era may provide new and efficient opportunities for accomplishing these goals (Hampton et al., 2013). For instance, analyzing bird diversity using long-term acoustic monitoring data is a promising new approach that can drastically enhance data collection well beyond the limitations of the human ear and workday, but few studies thus far have successfully employed it (Gibb et al., 2019). The acoustic complexity of communities is predicted based on species richness (Sueur et al., 2008), which makes biophony analysis an effective method for quantifying biophony and assessing species diversity within communities (Depraetere et al., 2012). However, acoustic data collection is difficult in urban forests because of severe noise pollution. While noise reduction is a popularly employed means for separating bird sounds from background urban noise (Farina and Pieretti, 2014), this method may result in the loss of important information about noise pollution and lead to a misinterpretation of the dynamic relationship between it and bird song diversity (Derryberry

* Corresponding authors at: School of Electronics and Communication Engineering, Guangzhou University, Guangzhou 510006, China and Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou 510520, China.
E-mail addresses: cyzhang@gzhu.edu.cn (C. Zhang), jihong.he61@foxmail.com (J. He).

https://doi.org/10.1016/j.ecolind.2022.109437
Received 5 July 2022; Received in revised form 14 August 2022; Accepted 7 September 2022
Available online 15 September 2022

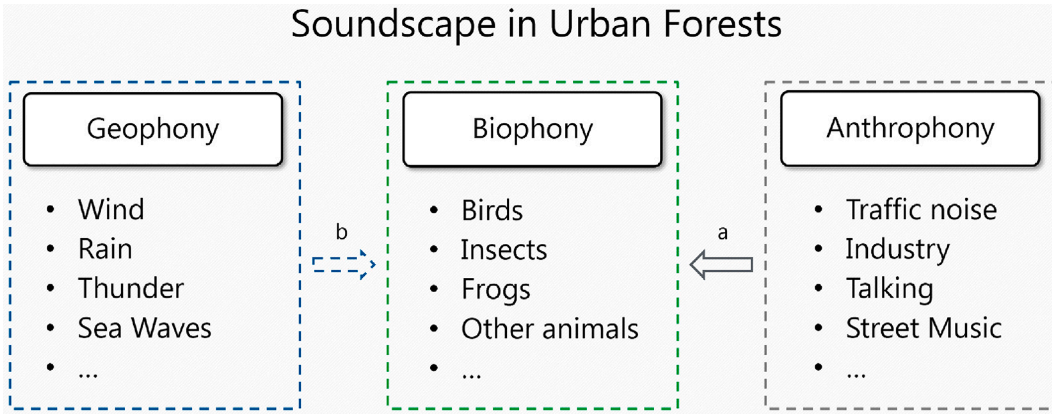

Fig. 1. Three Components of the Soundscape: geophony is the collection of sounds produced by abiotic factors such as wind, waves, rain, lightning, and earthquakes, and is closely related to climatic conditions and weather changes(Mullet et al., 2016); biophony is the collection of sounds produced by non-human living organisms such as birds, insects, amphibians, and other mammals (Krause, 2012); and anthrophony refers to the collection of sounds produced by humans such as from vehicles, industrial production, and audible communication (Joo et al., 2011). Interactions between these components include, a. anthrophony continuously affecting biophony in urban forests and, b. geophony also impacting biophony. Since the interaction between geophony and biophony is mainly due to unpredictable meteorological events, this study focuses on the interaction between anthrophony and biophony.

et al., 2020). On the other hand, a more complete understanding of avian diversity and the impacts of urbanization may be obtained by dividing the soundscape into acoustic scenes and analyzing each without the need for removing information.

Soundscapes can be divided into three interacting categories: geophony, biophony, and anthrophony (Fig. 1). These are composed of the total collections of sounds created by, respectively, abiotic sources, living creatures, and humans. Analyses of anthrophony in soundscape studies, as opposed to large-scale assessments such as remote sensing imagery, can provide a higher resolution and a more direct understanding of the extent of human disturbance on biodiversity. Such impacts include, for instance, the masking of biophony by anthrophony, the changing of the spatial structure of vegetation by habitat fragmentation (Alados et al., 2009), and an overall reshaping of the acoustic environment of forests by urbanization. As a result, these impacts, which are increasing in intensity (Tucker et al., 2014), can reduce the efficiency of inter-organismal communication (Pijanowski et al., 2011). Both prey responses to predator warning sounds and the detection of prey by predators may be impaired (Slabbekoorn and Ripmeester, 2008). Vocalizing in high noise pollution environments requires more time and energy to complete the communication, and yet the continued masking of vocalizations results in reduced reproduction and survival despite the increased effort by species. (Brumm, 2004). Background noise and vocalizations are often at similar levels, and lower background noise is essential to allow the successful transmission of vocalizations for attracting mates, defending territories, and interspecific communication (Slabbekoorn, 2004). By analyzing acoustic scenes and their interactions, we can better quantify species diversity, assess the impacts of human activities on the natural world, and improve the preservation of biodiversity within and near expanding urban centers.

Anthrophony conflicts with the sounds of birds and insects in numerous ways, including in frequency, amplitude, and duration (Aiello et al., 2016). As a result, it interferes with the breeding behavior of birds (Shannon et al., 2016), has significant effects on bird song transmission, domain defense, and attraction to the opposite sex, and ultimately leads to the loss of bird song as species diversity declines (Kociolek et al., 2011). The largest constituent of anthrophony is traffic noise, with a frequency distribution of 500 Hz–2000 Hz and a long duration (Wang et al., 2019). Noise from urban areas significantly affected the chirping of tree crickets. As traffic noise increases, the duration of chirping decreases and the probability of chirping pauses increases (Orci et al., 2016). However, in urban centers with dense populations, green spaces are subject to a variety of other noise scene types, such as loud music from amplified speakers, mechanical sounds from lawn mowers, and human conversations. Whereas anthrophony composed of only traffic noise tends to occupy lower frequencies, the frequency composition of urban noise is more complex. Vocal responses by birds and insects in these more complex urban noise environments are different as well. Hence, the study of the relationship between urban noise and bird song should not be limited to traffic-generated scenes (Nemeth et al., 2013). At the same time, continual improvement of audio recording equipment and storage devices makes sound data collection cheaper, more sensitive, and, as terabytes of data can be saved, longer longitudinal studies are more accessible. Traditional manual classification of sound data is highly limited for processing massive sound datasets, and as a result, there are losses of information about both humans and wildlife (Dawson and Efford, 2009). Machine learning (ML), however, can automatically identify and extract specific sound elements in huge datasets and provide more complete and higher resolution analyses (Cao et al., 2019). ML algorithms are classified as supervised and unsupervised (Sethi et al., 2020), and their employment on big datasets is possible through cloud computing platforms (Bardeli et al., 2010). Using ML and big data, a more comprehensive understanding of the impacts of human activities on bird population dynamics may be obtained. Though ML-based acoustic scene classification models afford a powerful tool for studying bird song dynamics in urban noise environments and the interactions between anthrophony and biophony, they are as yet undeveloped.

In this study, we constructed a convolutional neural network (CNN) model for the classification of sounds produced by birds and insects independently, overlapping sounds produced by birds and insects, birds and humans, insects and humans, and finally silence in urban forest soundscapes. We sampled across urban–rural gradients, variable land-cover types, and other environmental variation to assess the relative impacts of geophony and anthrophony on biophony. In our hypotheses, 1) acoustic scene classification models will help in assessing the impact of urbanization on biodiversity; 2) acoustic monitoring equipment has a limited range of measurement; and 3) individual acoustic scenes and their frequency composition will respond differently to urbanization. We proposed a research path for acoustic scene classification of urban forests that relies on passive monitoring technology. Finally, we aimed to develop a novel analytical framework for examining a massive amount of acoustic monitoring data to understand how human activities impact the biodiversity of urban forests going forth.

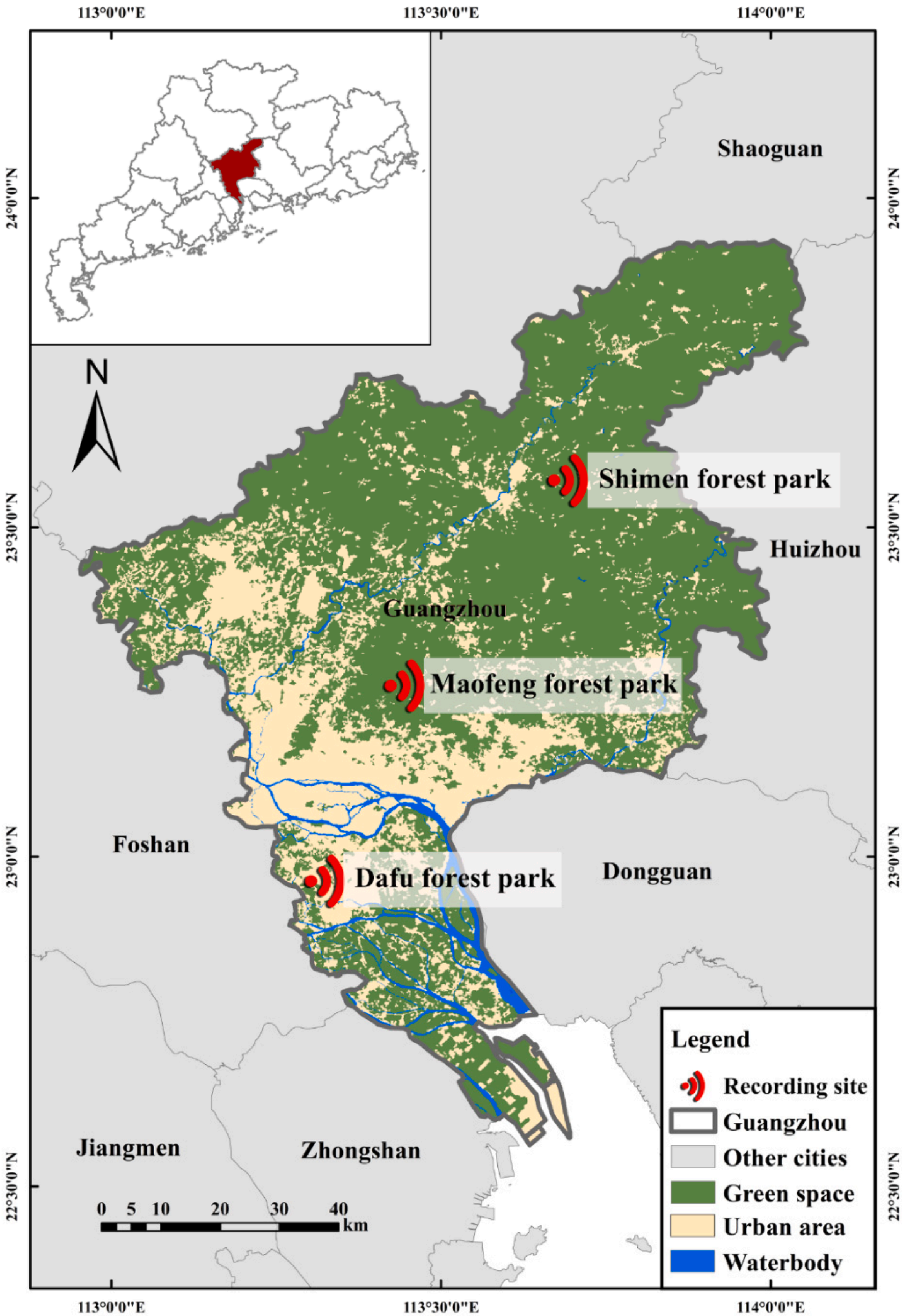

Fig. 2. Study area.

Table 1
Characteristics of the recording point in the study area.

Site	Coordinate	Dominant tree species	Forest coverage	Distance from urban center (KM)	Sound sources
SM01	N 23.628408° E 113.763567°	*Liquidambar formosana*	0.76	75.47	Airplanes Road vehicles Bird Insects Talking Music Natural sound
SM02	N 23.644973° E 113.786411°	*Castanea henryi*	0.76	78.44	
SM03	N 23.622534° E 113.812300°	*Schefflera heptaphylla, Machilus nanmu*	0.69	78.57	
MF01	N 23.304296° E 113.460645°	*Machilus nanmu*	0.78	28.21	
MF02	N 23.299516° E 113.460909°	*Machilus nanmu, Cinnamomum camphora*	0.77	27.85	
MF03	N 23.295443° E 113.464914°	*Machilus nanmu*	0.75	27.79	
DF01	N 22.954151° E 113.315804°	*Elaeocarpus apiculatus*	0.72	19.71	
DF02	N 22.951166° E 113.300258°	*Acacia confusa*	0.75	19.74	
DF03	N 22.942229° E 113.310629°	*Acacia confusa, Eucalyptus robusta*	0.76	20.91	

2. Method

2.1. Study area

The study area was Guangzhou, in the Guangdong Province of southern China, which is the core city of the Guangdong-Hong Kong-Macao Greater Bay Area. Guangzhou is built over a highly variable elevation gradient, with a mountainous region and concentrated forests in the north, smaller hills and basins in the center, and a coastal alluvial plain in the south. The geomorphology of the city results in an urban–rural gradient, with increasing urbanization from north to south as the landscape flattens. With a southern subtropical monsoon climate, the city largely experiences warm temperatures with an average of 21 ℃. Though precipitation is higher during the rainy season from April to June, it is abundant throughout the year. Evergreen broad-leaved forests compose 41.6 % of the city, and overall greenspace covers 45.52 %. The average park space per urban citizen is 18 square meters Meanwhile, improvement of Guangzhou's urban forests in recent years has resulted in an increased diversity of bird communities, and 307 species have been surveyed and recorded. High biodiversity provides ample fodder for addressing the questions about monitoring avian diversity and the impacts of anthrophony on biophony in urban environments.

2.2. Location of recording plots and recording schedule

Sample sites encompass the full urban–rural gradient across Guangzhou, including Shimen National Forest Park (SM), Maofeng Mountain Forest Park (MF), and Dafu Mountain Forest Park (DF) in the north, central, and south, respectively (Fig. 2, Fig. A1). SM is located in the exurban region, MF in the suburbs region, and DF in the urban region. In each urban forest park, three acoustic environment monitoring sites were set up to sample variations in vegetation types and the degree of human interference (Table 1). Three rules were followed in setting the monitoring points: 1) the monitoring points were spaced at a distance of not <200 m to ensure the independence of data; 2) each was on approximately the same slope to control for error caused by topographical differences; 3) each was about 20 m from the forest edge to control the influence of the plant community on noise attenuation.

Song Meter SM4 acoustic recorders (Wildlife Acoustics, Concord, MA) were deployed at each of the nine plots for six consecutive months (October 2021 through April 2022). These were programmed to collect 1-min recordings every ten minutes, amounting to 1296 samples per day. Recordings are obtained in stereo at 16 bits with a 32 kHz sample rate, ensuring coverage of the known biophony and anthrophony. The preamplifier gain was set to 28 dB to meet the signal-to-noise ratio of the learned samples in the deep learning model. On a healthy tree ~10 cm at breast height, audio recording equipment was fixed 1.5 m from the ground to avoid the reverberation of the ground on the sound. Meanwhile, the data from nearby weather stations allowed for the exclusion of periods of rainfall (>10 mm in 24 h) and high wind (average wind speed over 8 m/s) from data analyses to reduce the interference of geophony. In the end, we obtained a total of 132 days, or 190,080 min (3168 h) of data.

2.3. Acoustic scene classification model

2.3.1. Acoustic scene classification scheme

This study aimed to classify biophony, anthrophony, and their hybrid

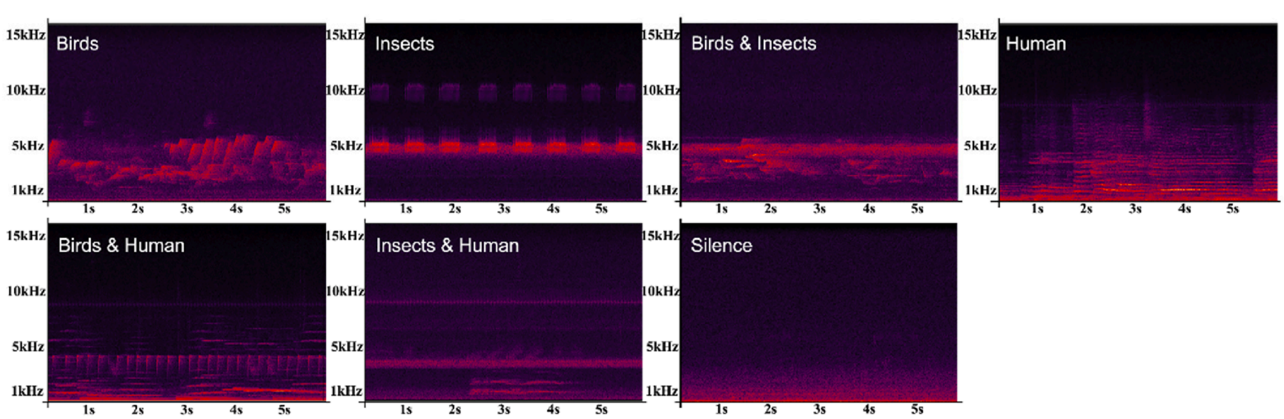

Fig. 3. Schematic of learning samples for different acoustic scenes.

Table 2
Acoustic scene classification and division criteria.

Acoustic scene	Description
Bird sound	A sound segment in which only bird sounds occur is defined as bird sounds, including different species of bird chirps or singings.
Insect sound	Insect sound is a segment of sound that only has insect calls. This includes different kinds of insect calls, but not things like wing beats that aren't calls.
Bird-insect sound	Define the sound segment of bird sound and insect sound as "bird-insect sound."
Bird-human sound	The sound clip that mixes bird sounds and anthrophony is defined as "bird-human sound."
Insect-human sound	Define the sound fragment that mixes insect sound and anthrophony as Bird-human sound.
Human sound	Define human sound as sound clips where only anthrophony occurs, which includes sounds such as car sirens, conversations, music, and construction sounds.
Silence	Defines a sound clip that does not have any sound elements present as "silence"

sounds in complex urban forest soundscapes. Soundscapes were divided into seven types of acoustic scenes: sounds created by birds, insects, or humans independently, and overlapping sounds by birds and insects, birds and humans, and insects and humans, and finally silence (Fig. 3). Specifications for each type of acoustic scene are explained in Table 2. Amphibians are also present in these urban forests, but they are not included in this classification model because the recording sites in this study were not consistently placed near water sources.

2.3.2. Construction of acoustic scene classification model

This project constructed an urban noise scene classification model based on the Log-Mel spectrogram and CNN. The construction of the model consists of three main steps (Fig. 4). First, we prepared training samples for each of the scene definition types. After determining the classification criteria for the seven types of acoustic scene, we used Wavesurfer software to screen the ideal learning samples from the original audio collected by the audio recording equipment (Table A1). To ensure the accuracy and generalizability of the classification model, we collected different sounds of the same type while ensuring a high signal-to-noise ratio (e.g., the bird sound tag collects different types of bird calls). The sampling rate was 22,050 Hz with a duration of 3–5 s. Data from each category was divided into sets with a 9:1 ratio for model training and validation.

Second, training samples of urban noise scenes were transformed into logarithmic Mel spectrogram feature maps. For training, each sample is randomly sampled for 4 s (<4 s is zeroed at the end) and converted into a Mel-spectrogram. The number of FFT points is 1024, the frame shift is 512, the number of Mel filter groups is 128, and the size of the Mel spectrogram obtained for each sample is 128x173. Third, a deep learning model was built based upon features from the Mel spectrograms. The backbone of the model was ResNet18, with the number of output neurons in its last layer modified to the corresponding number of ambient sound categories. The training process uses a cross-entropy loss function and an Adam optimizer. The learning rate was set to 0.0002, the batch size was 32, and the number of training rounds was 200. The highest model validation accuracy is 96.78 %, accuracy rate of each classification scenes is shown in Table A1. Model parameters at this point were saved for subsequent inferences from real environmental data.

2.3.3. Sound analysis

(1) Acoustic scene dominance.

Based on the classification model to obtain information on the total number of seven types of acoustic scenes in the study areas (x1, x2, x3… x7), where i means each sound scene.and calculating the proportion of each in the overall soundscape, the dominance (D) of each scene may be calculated as follows:

$$D_i = \frac{xi}{(x_1 + x_2 + x_3 + x_7)} \tag{1}$$

(2) Calculation method for target sound area ratios (TSAR).

A short-time Fourier transformation is applied to the one-dimensional audio data, with FFT points of 1024 and frame shifts of 512, to generate a two-dimensional spectrogram D. By taking the mode value of D at each pixel and converting it to decibels, *D_db* is obtained. D_{mean} is then calculated by using the mean decibel value of the spectrogram. By setting the scale factor (e.g., 0.8) and traversing the spectrum map, we identified each frequency point in the map as target sound or bottom noise according to this equation (1):

$$D_{process} = \begin{cases} \text{background noise} & D\_db_{i,j} < \alpha^* D_{mean} \\ \text{target sound} & else \end{cases} \tag{2}$$

Band cuts were made in equal proportions at intervals of 1 KHz on the spectrogram. Since the audio sampling rate is 22050 Hz, 11

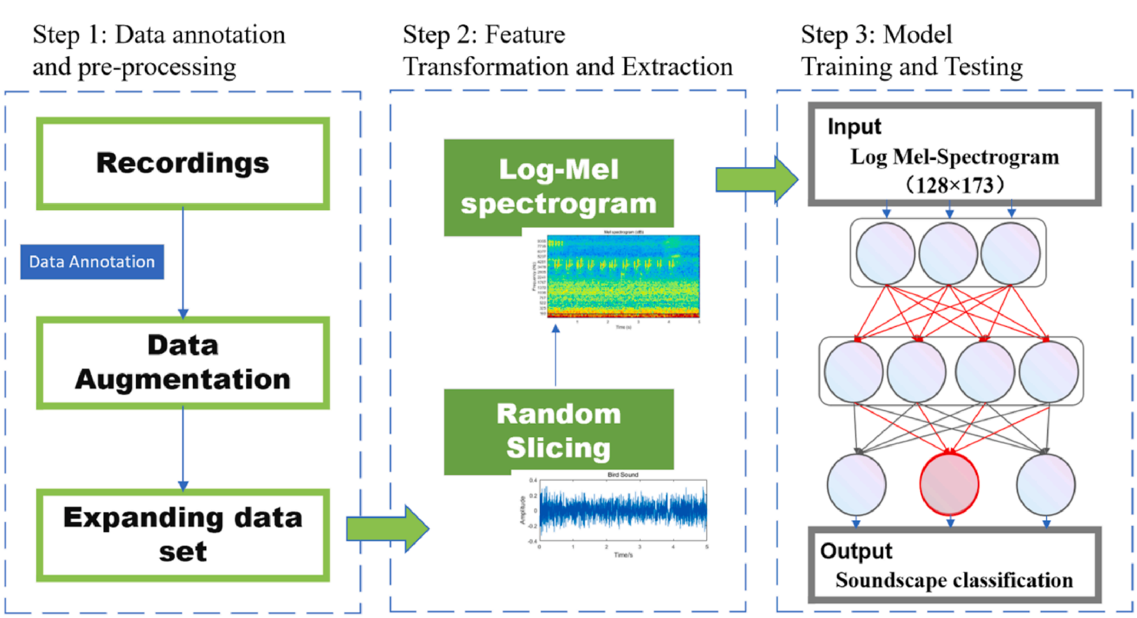

Fig. 4. Steps to build a acoustic scene classification model.

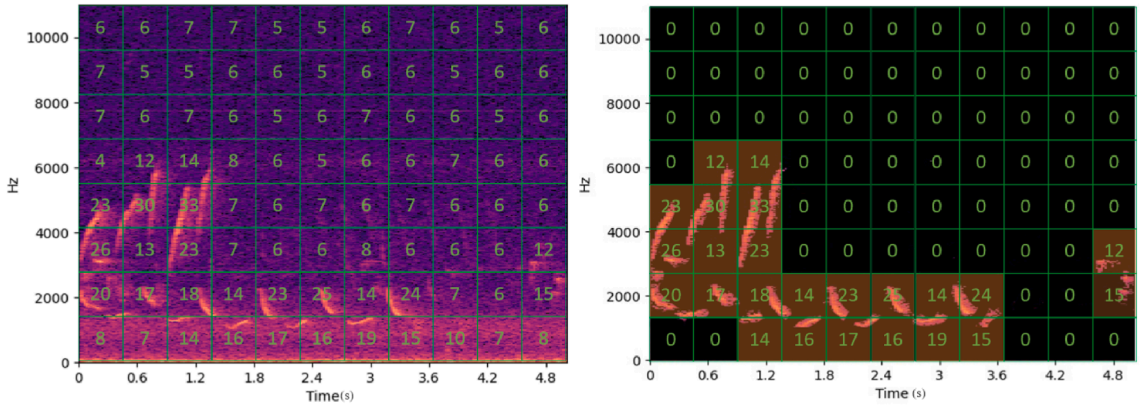

Fig. 5. The left figure shows the original Mel-spectrogram of classification results, and the numbers in the box represent the pixel points, and the target sound area (yellow area) in the right figure is obtained by calculating equation (1). To quantify the intensity and frequency distribution characteristics of target sound, equation (2) calculates the TSAR within each frequency band. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

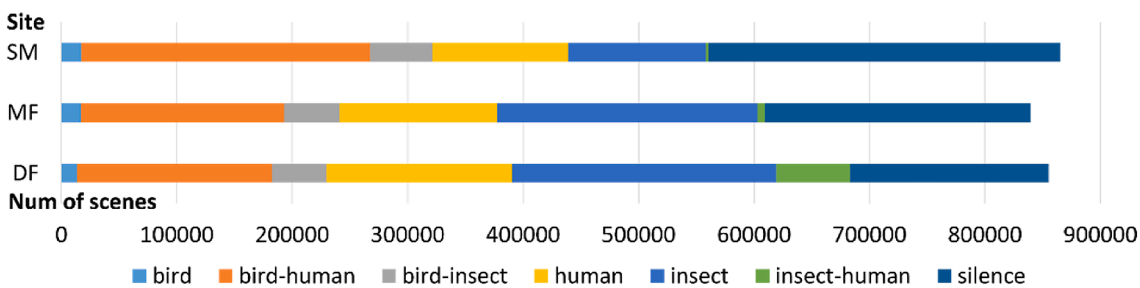

Fig. 6. Composition of different acoustic scenes among the sample sites.

frequency bands were obtained (the width of the last band is 1025 Hz). There are a total of T_num_k pixel points of the target sound for each frequency band, where k is the frequency band subscript. Using the following equation (2), the number of target sound pixels in each band is divided by the total area S of the spectrogram to obtain the proportion of target sound in each band (Fig. 5):

$$r_k = \frac{T\_num_k}{S}, \ k \leqslant 0 < 11 \tag{3}$$

2.4. Assessment of human interference

Land use is an important factor that determines the spatial characteristics of urban forest acoustics and is highly impactful on the soundscapes of urban forests. We used the Google Earth Engine (GEE) platform to obtain information about land use surrounding sample sites, including *urban*, *forest*, *water*, *grass*, and *bare*. The remote sensing image data were selected from Sentinel-2A L1C level data from January 1, 2021 to April 30, 2022, and image de-clouding was performed using the QA60 band. Land-cover type samples were labeled manually using visual inspection, and the labeled samples were divided into training and validation sets at a ratio of 7:3. We employed the random forest (RF) method smilerandomForest() to train the land classification model, and the confusion matrix, OA, and KAPPA coefficients were used to evaluate the classification efficiency (Table A2). After the model training was completed, we calculated land use types within five increasing buffers (100 m, 200 m, 300 m, 400 m, and 500 m) using the *pixelArea*.

2.5. Statistical analysis

We used two-way ANOVAs to examine the differences in dominance of various acoustic scene types across urban–rural gradients. For these, we set the acoustic scene dominance as the dependent variable, and the independent variables were the urban–rural gradient and acoustic scene types. Variables were fitted to normal distributions using a parameter approach. Shapiro-Wilk tests and Q-Q plots were used to test for normality (Fig. A1). Homogeneity of variance and covariance was assessed using Levene's test and Box's M-test, respectively. When necessary, data was log-transformed to meet assumptions of normality and homogeneity. We grouped the data by sites and soundscape events and analyzed the simple main effects on acoustic scene dominance separately. All of these analyses were implemented in R using '*rstatix*' package.

An RF regression model was constructed to explore the explanatory power of environmental factors and acoustic scene dominance at different buffer scales. This model is insensitive to multivariate covariance, is convenient for testing nonlinear effects, and is advantageous in the processing of large data sets because of its computational tractability. For RF models, we used the dominance of each acoustic scene type as the dependent variable, and the quantifications of land use at the increasing buffer scales were independent variables. The explanatory power (R2) and significance (p-value) of the model of environmental factors on the dominance of acoustic scenes were determined using replacement tests (N = 99). The importance of the independent variables for explaining variation in acoustic scene dominance was ranked using a mean decrease in accuracy. A circular function that iteratively calculates the mean value of the model misspecification rate based on OOB data was used to determine the optimal number of mtry high percentage of explained variance. The optimal number of trees collected (ntree) is judged by the plot of model error versus ntree (here set to 1000). These analyses were implemented using the packages '*randomForest*', '*rfUtilities*', and '*rfPermute*' in R.

A redundancy analysis (RDA) was used to identify the influence of landcover on soundscape frequency composition. In these analyses,

Table 3
Random forest model performance sorting.

Soundscape	*p*	R^2	Explained variance %	Soundscape	*p*	R^2	Explained variance %
Bird				**Bird-human**			
Buffer200	0.01	0.17	17.65	Buffer200	0.01	0.55	54.64
Buffer100	0.01	0.14	14.93	Buffer300	0.01	0.48	48.22
Buffer300	0.01	0.12	12.30	Buffer100	0.01	0.48	48.18
Buffer400	0.01	0.11	11.19	Buffer500	0.01	0.42	42.79
Buffer500	0.01	0.11	11.02	Buffer400	0.01	0.42	42.56
Insect				**Insect-human**			
Buffer200	0.01	0.17	17.39	Buffer200	0.01	0.19	19.13
Buffer100	0.01	0.16	15.98	Buffer300	0.01	0.16	15.96
Buffer300	0.02	0.12	12.91	Buffer100	0.01	0.15	14.82
Buffer400	0.02	0.11	11.11	Buffer400	0.01	0.12	11.72
Buffer500	0.02	0.09	9.35	Buffer500	0.01	0.11	11.25
Bird-insect				**Silence**			
Buffer200	0.15	−0.12	−11.75	Buffer200	0.01	0.13	12.58
Buffer100	0.15	−0.13	−11.98	Buffer100	0.02	0.09	9.13
Buffer300	0.15	−0.12	−12.17	Buffer300	0.02	0.09	9.04
Buffer500	0.15	−0.13	−12.23	Buffer400	0.02	0.08	8.41
Buffer400	0.15	−0.12	−12.52	Buffer500	0.02	0.07	6.98
Human							
Buffer200	0.01	0.51	51.15				
Buffer300	0.01	0.45	45.10				
Buffer100	0.01	0.42	42.85				
Buffer400	0.01	0.42	42.80				
Buffer500	0.01	0.39	39.60				

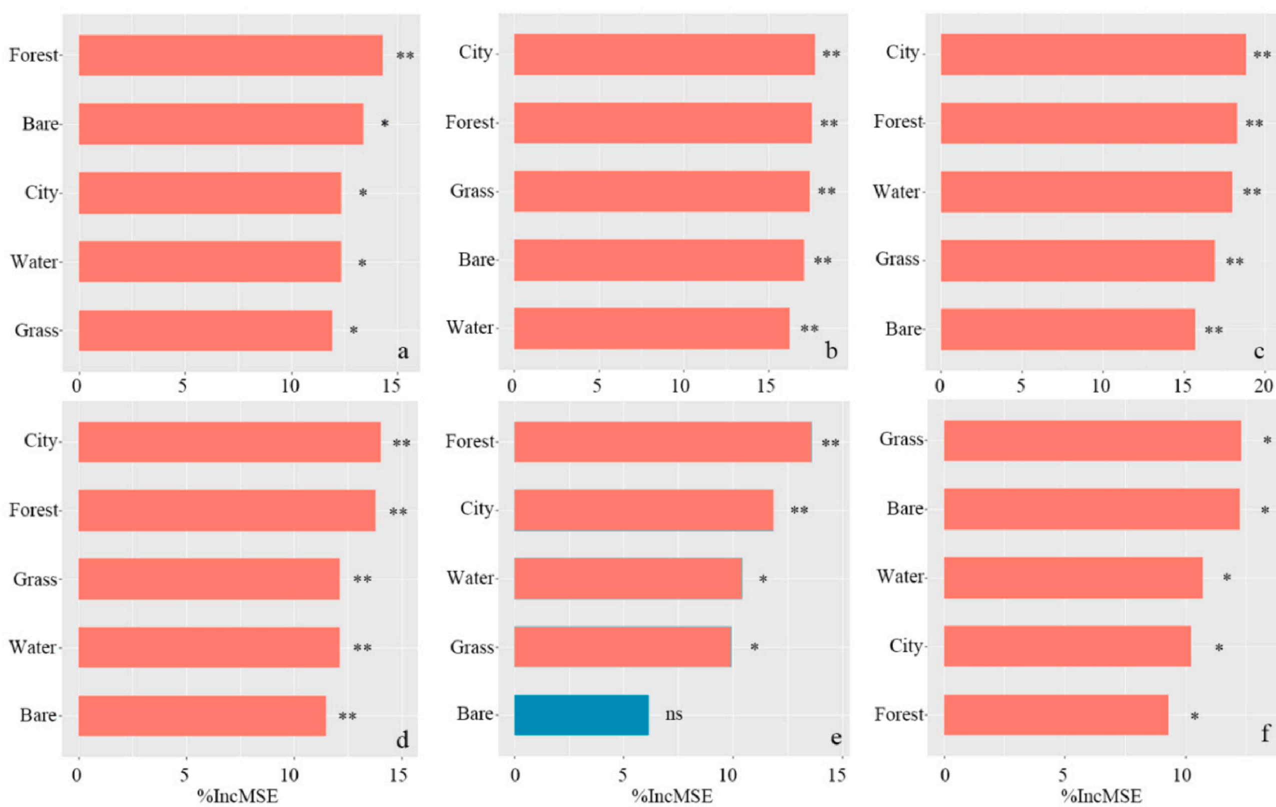

Fig. 7. a) Bird; b) human-bird; c) human; d) insect; e) insect-human; f) silence. Because the model for bird-insect was not significant, the ranking of importance values was not analyzed. **: $p < 0.01$; *: $p < 0.05$; ns: $p > 0.05$.

landcover data was transformed with the Hellinger-transformation (Legendre and Gallagher, 2001), and the results showed that the short gradient length along the first axis was less than three, which meant that RDA was more suitable for this dataset than canonical-correlation analysis (CCA). The band area ratio of each frequency band (N = 10) was treated as the species variable. Since the acoustic scenes occupy different frequencies, it is possible to estimate their intensity using the area of frequency bands (Kwan et al., 2004; Nemeth and Brumm, 2010). Adjusted R^2 was used to determine model performance, and Monte Carlo (MC) permutations (N = 99) were used to determine the significance of the model, axes, and explanatory variables. The explanatory power of landcover types was determined using hierarchical partitioning. These analyses were performed in '*vegan*' and '*rdacca.hp*' package in R.

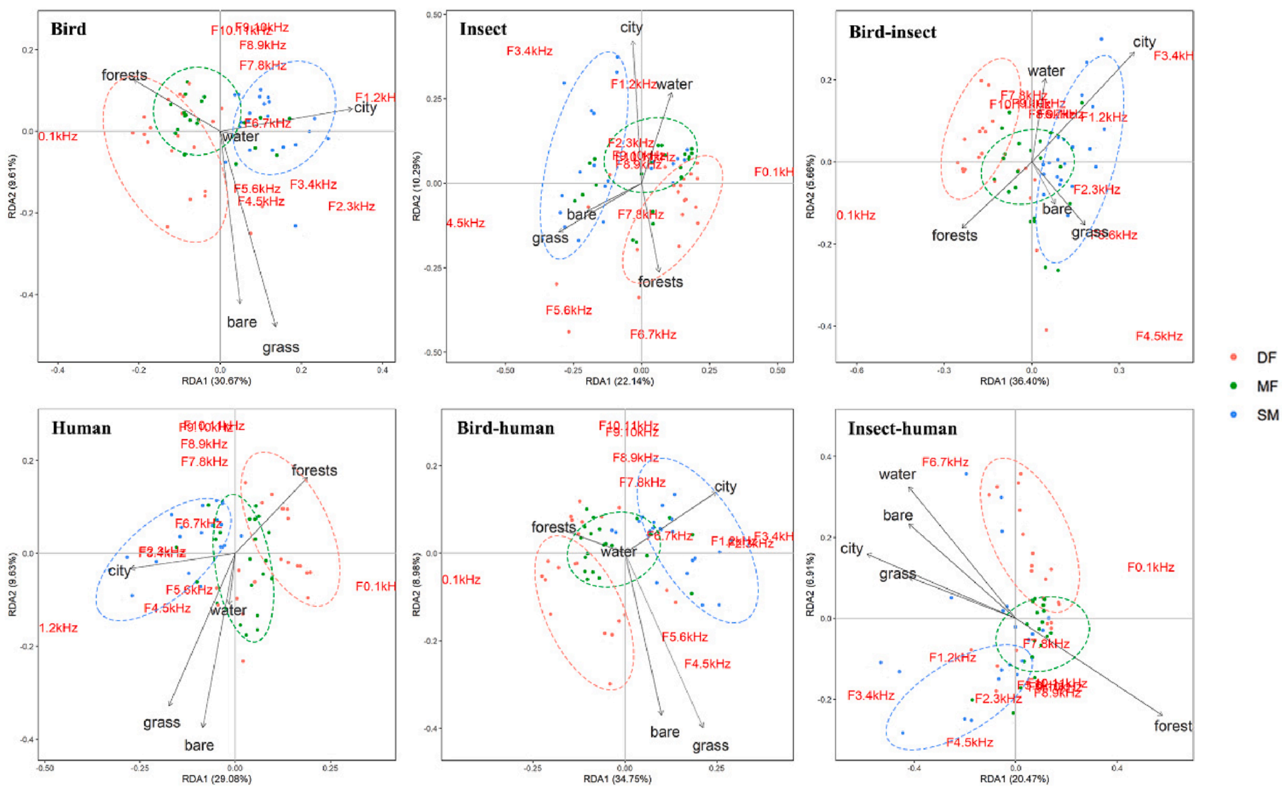

Fig. 8. Redundancy analysis shows the relationship between land use types and ten 1 kHz frequency bands.

3. Results

3.1. Spatial variation across site locations

The model classification results showed that the SM site had the most bird, bird-human and silence scenes; while the DF site had the most human sound scenes (Fig. 6). Two-way ANOVAs showed numerous significant relationships. Acoustic scene dominance was significantly different between sample site locations (F (2, 252) = 2.95, p = 0.003). Significant differences were also discovered between acoustic scene types (F (2, 252) = 156.51, p < 0.001) and interactions in acoustic scene dominance (F (2, 252) = 7.98, p < 0.001). Sample site location was a significant effect for the bird, human, bird-human, and insect-human acoustic scenes, but not for insects and silence (Table A3).

3.2. Relationship between land use type and acoustic scenes

The optimal RF model had the highest percentage of variance explained with the 200 m buffer, followed by 100 m, and 300 m (Table 3). In contrast, urban land-cover types within 400 m and 500 m radii of the sample points had the lowest explanatory power. In terms of model performance for different acoustic scenes, p-values were < 0.05 in RF models for all acoustic scenes except bird-insect. Among them, bird-human and human had the highest R^2 with 0.55 and 0.51, respectively. Silence had the lowest R^2 of 0.13.

Since the 200 m buffer showed the highest explanatory power, the RF model at this scale was used for the importance ranking of urban land-cover types in each acoustic scene (Fig. 7). It was found that the importance values of land-cover types were significant, except for the insect-human model. For each acoustic scene except silence, presence of *forest* and *city* consistent ranked highest and *water* lowest for explaining variation. While RF models containing anthrophony components had *city* in the top two positions, those with biophony components had *forest* in the top two positions. The RF model for silence differed in having *grass* and *bare* land ranked as the top variables.

3.3. Acoustic scenes variation across frequency intervals

RDA detected relationships between the frequency composition of the acoustic scene and urban space factors (Table A4). Urban space had high explanatory power for the frequency composition of acoustic scenes, including >40 % of the variation of bird-insect, human, and bird-human scenes. It was the lowest for the insect-human scene at 24.70 %. The results of decomposing the explained amount of each environmental variable in each RDA model through hierarchical partitioning showed that the three land-cover types of *city*, *forest,* and *grass* performed more prominently in the model. The silence element in an acoustic scene represents the silent sound environment, so the distribution of its frequency interval has no practical significance and is therefore not discussed here.

Fig. 8 illustrates that frequency distributions of different acoustic scenes have an obvious correlation with urban land-cover type, but also some differences. 1–2 kHz is the main interval of urban noise frequency distribution, but all show strong correlation with city elements. Comparing acoustic scenes containing bird sound, it apparent that the low frequency distribution of pure bird sound scene is wider, mainly distributed in 2–6 kHz. The low frequency distribution in bird-insect scene is narrower, only distributed in 4–6 kHz. The low frequency distribution of bird-insect sound scene is discontinuous, but still in the frequency interval of 2–6 kHz. Meanwhile, the low-frequency ranges in the pure bird, bird-human, and bird-insect scenes showed significant positive correlations with *bare* and *grass*, but negative correlations with *forest*. When comparing the scenes containing insect sounds, the high frequency range, 6–8 kHz, showed a significant positive correlation with *forest* and a significantly negative correlation with water elements. In the pure anthrophony scenes, the low frequency range (1–6 kHz) showed a significant positive correlation with *city* and a significant negative correlation with *forest*.

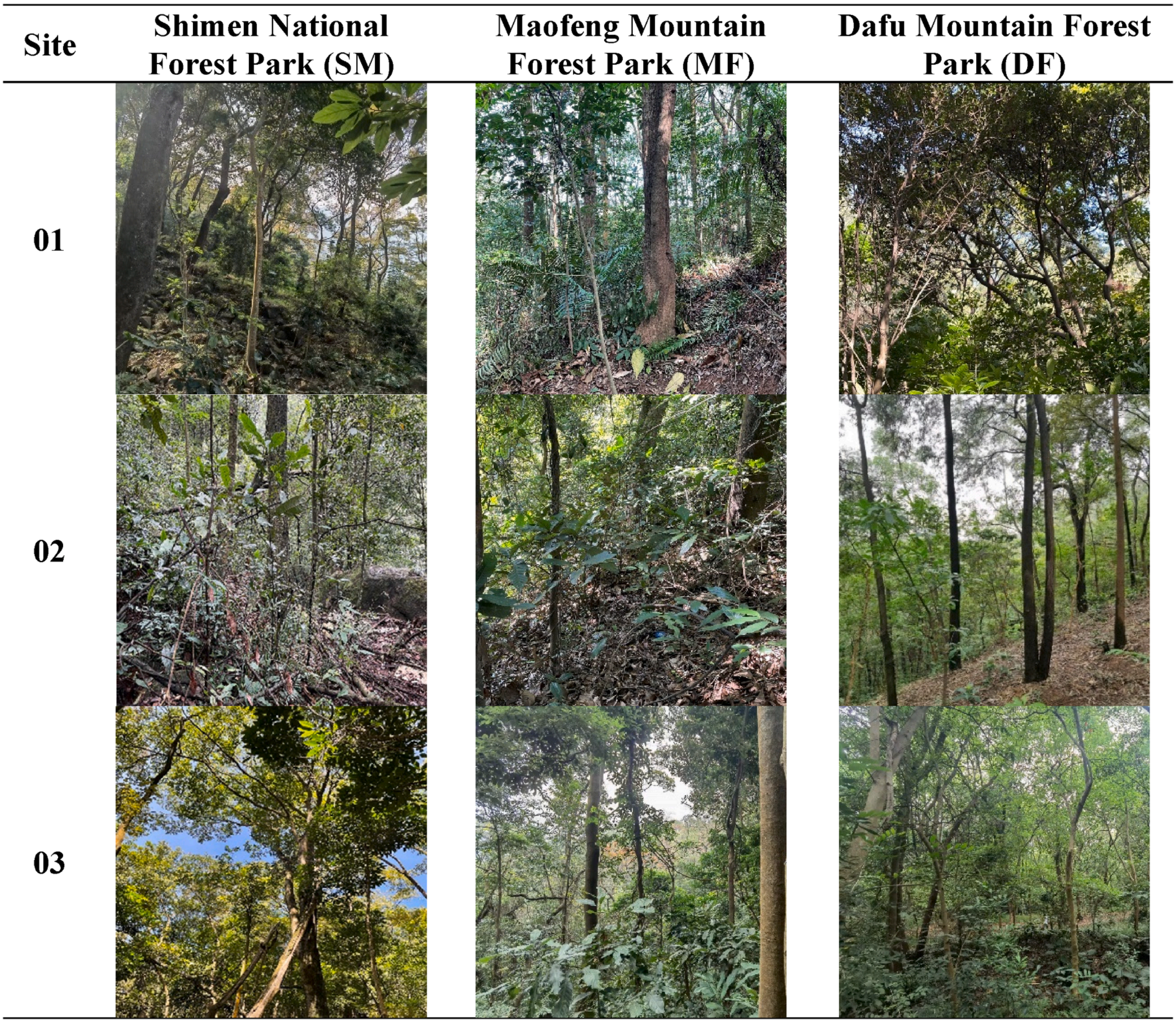

Fig. A1. Photos of the nine sound recording sites.

4. Discussion

4.1. Acoustic scene classification model

The acoustic scene classification model using a deep learning algorithm can automatically analyze massive amounts of raw audio data, and this big data-based framework provides a new research path for answering questions about the relationship between urban biodiversity and human activities. In the process of soundscape research, the evaluation of biodiversity by acoustic indices is often biased by noise and other factors that produce errors (Fairbrass et al., 2017). These are related to the distinct frequency composition characteristics among acoustic scenes (Priyadarshani et al., 2018), as the vast majority of bird sound is distributed between 2 and 6 kHz and common urban noise is between 1 and 4 kHz. Here, we propose a novel methodology for robustly addressing those biases without the loss of data associated with traditional soundscape analyses.The overall classification accuracy of the model obtained in this study was 96.78 %, which is generally consistent with the accuracy of other soundscape analytical models applied in urban environments (Torija et al., 2014). Complex urban noise, diverse types of biological sounds, and the mixed composition of various sounds all affect the accuracy of sound classification models to some extent (Priyadarshani et al., 2018). Collecting as many learning samples of various types of sound scenes related to the study area as possible is essential for improving model accuracy, but this may reduce the generalizability of the model (Li et al., 2018). We illustrate an improvement upon traditional methods in the field of soundscape analysis, such as using field questionnaires or manual labeling (Ulloa et al., 2016), using a deep learning algorithm. Our method not only provides a more complete understanding of the sound elements, their interactions, and their dynamics over time, but can also save time and money because it is automated.

Krause (Krause, 2008) proposed that soundscapes consist of three basic categories: biophony, anthrophony, and geophony. This is an important reference basis for classification in subsequent soundscape studies (Hong et al., 2019), and researchers have adjusted the classes in a targeted manner to meet research needs (Liu et al., 2019). Quinn et al. used biophony, anthrophony, and geophony as classification criteria and obtained an accuracy of 0.94 for the acoustic scene classification model (Quinn et al., 2022). In contrast to the previous classification scenarios, this study was classified urban forest acoustic space into seven categories: bird sound, insect sound, bird-insect sound, anthrophony, bird-human sound, insect-human sound, and silence. Unlike other soundscape studies conducted in urban environments, this study refines the acoustic community sound classification scale and defines bird-insect sound, bird-human sound, and insect-human sound to reveal the use of the urban forest by the acoustic community and its interrelationship with urban disturbance. Our classification model can effectively distinguish pure bird sound, pure insect sound, and biophony disturbed by humans, and the final training results show that the accuracy of the

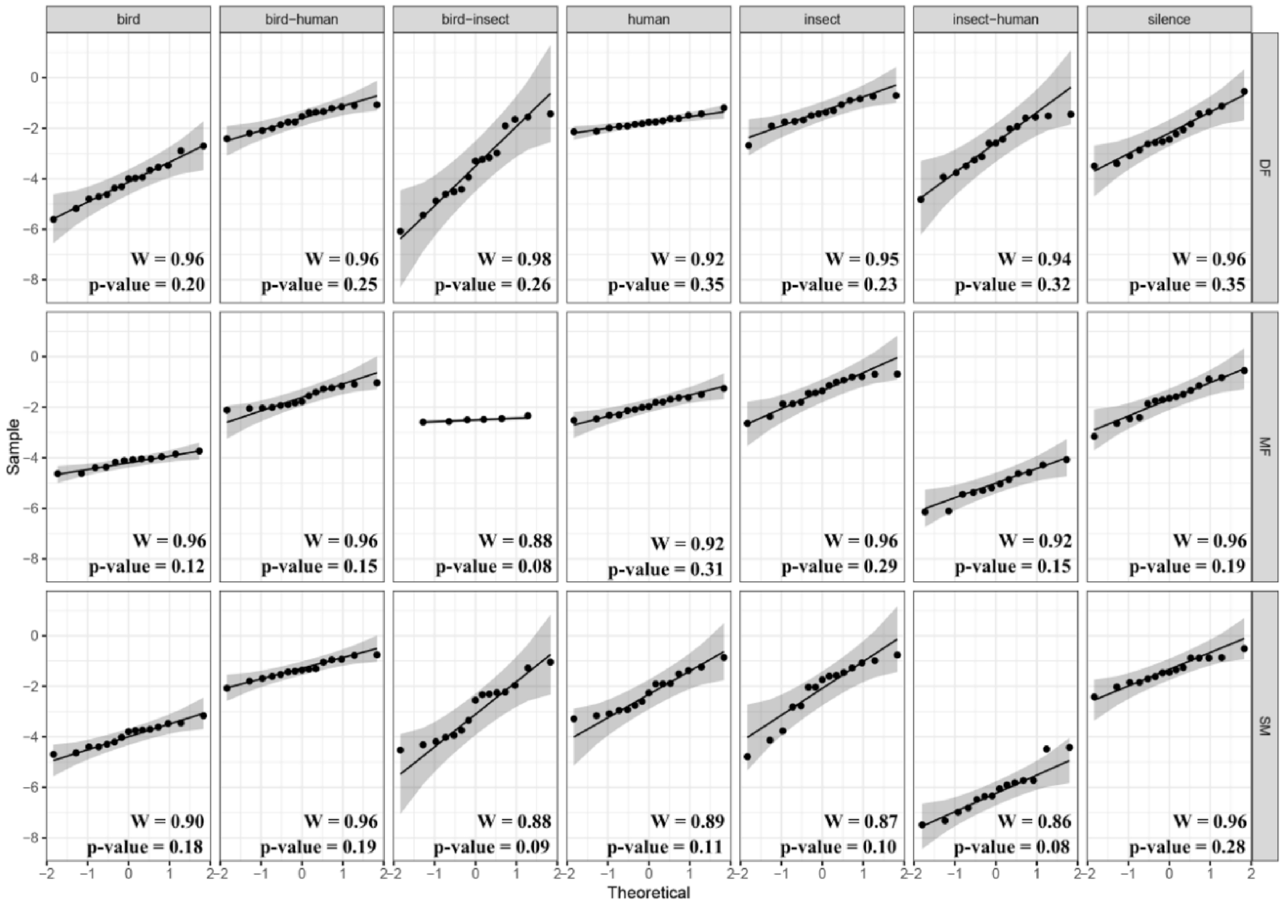

Fig. A2. Normality test (Shapiro-Wilks and Q-Q plot) for acoustic scenes.

Table A1
Number of learning samples by acoustic sceneand the performance of the classification model.

Acoustic scenes	Quantity	Total duration (minutes)	Precision	Recall	F1-score
Bird sound	1271	105.92	0.96	0.95	0.96
Insect sound	465	38.75	0.94	0.99	0.97
Human sound	473	39.42	0.98	0.99	0.99
Bird-insect sound	725	60.42	0.98	0.98	0.98
Bird-human sound	519	43.25	0.96	0.98	0.98
Insect-human sound	512	42.67	0.95	0.94	0.95
Silence	385	32.08	0.95	0.99	0.97

models for bird-human sound and insect-human sound reaches 0.99 and 0.95, respectively. But, geophony elements were not included in the classification model, but it is hoped that long-term monitoring may reveal the impacts of weather variation on the urban forest acoustic community (Metcalf et al., 2020).

Taking bird sounds as an example, based on this classification model, we can compare and study the relationship between pure bird sound scenes and human-disturbed bird sound scenes and spatial factors to clarify the relationship between urban biodiversity and human disturbance. Currently, we are developing a species identification model based on bird sound and anticipate more finely determining which species only vocalize in low noise disturbance environments and which adapt to

Table A3
Effects of sites on the acoustic scenes.

Effect	Acoustic scenes	DFn	DFd	F	*p*
Site	Bird	2	252	110.0	< 0.001
	Insect	2	252	1.76	0.18
	Bird-insect	2	252	13.6	< 0.001
	Human	2	252	212.0	< 0.001
	Bird-human	2	252	57.5	< 0.001
	Insect-human	2	252	29.5	< 0.001
	Silence	2	252	1.33	0.28

Table A2
OA, KAPPA coefficients, and confusion matrix results for the land use classification model.

	Overall accuracy (OA)	KAPPA coefficients	Confusion Matrix					
Land use classification model	0.96	0.95	0	0	0	0	0	0
			0	6151	25	0	15	21
			0	24	1435	4	11	1
			0	3	1	1289	1	0
			0	33	13	0	2750	25
			0	234	5	0	55	1084

Table A4
Permutation tests of variance explained by land use type in RDAs of acoustic scenes.

Soundscape	Variables	Df	F-value	*p*	Explained variation
Bird	**Total**	**5**	**8.22**	**0.001**	**36.80**
	City	1	6.19	0.004	23.51
	Grass	1	2.92	0.049	21.44
	Forest	1	9.31	0.001	19.51
	Bare	1	13.42	0.001	19.48
	Water	1	9.27	0.001	16.09
Insect	**Total**	**5**	**6.95**	**0.001**	**32.40**
	Grass	1	0.78	0.500	21.57
	City	1	7.03	0.002	21.05
	Forest	1	5.15	0.003	20.06
	Bare	1	13.53	0.001	19.26
	Water	1	8.24	0.001	18.09
Bird-insect	**Total**	**5**	**9.65**	**0.001**	**41.10**
	City	1	10.83	0.001	26.23
	Forest	1	6.52	0.001	21.34
	Water	1	11.00	0.001	19.03
	Grass	1	3.90	0.001	17.35
	Bare	1	16.03	0.001	16.06
Human	**Total**	**5**	**10.61**	**0.001**	**43.70**
	Grass	1	9.09	0.002	22.04
	City	1	6.87	0.003	21.17
	Bare	1	17.50	0.001	20.30
	Forest	1	8.10	0.001	18.72
	Water	1	11.46	0.001	17.69
Bird-human	**Total**	**5**	**9.72**	**0.001**	**41.30**
	Grass	1	3.28	0.047	22.11
	City	1	5.21	0.005	21.69
	Bare	1	19.77	0.001	19.27
	Forest	1	8.52	0.001	19.08
	Water	1	11.81	0.001	17.82
Insect-human	**Total**	**5**	**5.06**	**0.001**	**24.70**
	Forest	1	2.65	0.037	25.75
	City	1	13.65	0.001	25.67
	Water	1	4.13	0.006	17.81
	Grass	1	2.34	0.056	16.32
	Bare	1	2.53	0.043	14.37

urban noise. Sound is an effective physical indicator of ecological processes. It is a direct reflection of the behavioral activities of vocalizing species such as birds or insects, and their relationships with their surroundings. Soundscape ecology studies typically employ acoustic indices, which have been found to be closely related to species richness (Eldridge et al., 2018; Pijanowski et al., 2011). Similar to acoustic indices, the TSAR proposed in this study is capable of quantifying dynamic characteristics of acoustic communities. TSAR and the acoustic index need to be further studied to determine their scope of application (Eldridge et al., 2016). It is important to learn from other study (Ren et al., 2022b) and propose standardized protocols for monitoring and evaluating the acoustic properties of urban forests. Acoustic evaluation accuracy will be enhanced by minimizing the effects of differences in recording equipment, processing methods, background noise, and other factors.

4.2. *Effective range of acoustic monitoring equipment*

The effective range of acoustic monitoring equipment has long been a problem in the design of bioacoustic experiments (Mennill et al., 2012; Yip et al., 2017). By comparing the differences in the effects of five land properties on different acoustic scenes, we found that a 200 m buffer can be used as the optimal analysis radius for explaining variation in the soundscapes of the urban forests in Guangzhou. Khanaposhtani and Pieretti also explored the relationship between bird sound activity and road distance in forest communities (Ghadiri Khanaposhtani et al., 2019), and both demonstrated the possibility of examining the impacts of anthropogenic activities on acoustic scenes through spatial distance information. Throughout this study, it was found that urban land properties within a 200 m radius explained the greatest amount of each acoustic scene. As the radius of analysis increased, the explanation decreased. The study also demonstrates the effective monitoring radius of acoustic monitoring equipment, which can be used in the design of bioacoustic experiments. By comparing the results of previous studies (Hao et al., 2021a), we found significant differences in the effective monitoring distances of different monitoring devices. This indicates that it is necessary to limit a reasonable study radius to answer scientific questions when analyzing acoustic scene dominance from a spatial perspective in a complex urban environment. Currently, there are some toolkits for deriving sound-emitting distances based on information such as sound pressure levels, such as '*scikit-maad*' (Ulloa et al., 2021), which can assist in calculating the effective working radius of monitoring devices. Due to the richness of vocalizing species, data on the initial sound pressure levels of a large number of vocalizing species is still missing, and the accuracy of the extrapolation needs further validation.

This study explores the effective range of acoustic monitoring devices. A reasonable deployment of monitoring devices can allow city managers to predict and manage the acoustic environment of urban forests in the future. Using soundscape maps (Hao et al., 2021a), for instance, residents can be encouraged to use and recognize urban green spaces in an effort to increase their green wellbeing (Ren et al., 2022a; Dong et al., 2022). The study was limited to only nine recording devices, and the type of urban forest covered and the spatial scale are still lacking. The cost of acoustic monitoring equipment is one of the reasons why acoustic monitoring cannot be carried out on a large scale. Therefore, this research team is using self-made acoustic monitoring equipment to achieve a wider range of acoustic monitoring equipment with a limited budget, and related research is still underway (Dong et al., 2022).

4.3. *Acoustic scenes and land use type*

In this study, *city* and *forest* were found to be the most important factors influencing acoustic scene dominance. The spatial pattern of the *forest* and *city* landscape is a key factor influencing the acoustic environment of urban forests. While city elements ranked highest in explaining variation in acoustic scenes containing anthrophony, forest elements were highest for those containing biophony. From the perspective of sound sources, urban land use is the main source of noise in this study. Factors such as industrial facilities, road traffic, and buildings shape the urban acoustic space (Guedes et al., 2011) with unique acoustic environmental characteristics (Peckens et al., 2018). The forest land use type, on the other hand, is an important space for vocalizing species to survive in the urban environment. Using an acoustic scene classification model, we show that while bird song scenes are impacted by *forest*, *bare*, and *city*, urban landscapes have the strongest impacts (Dein and Rüdisser, 2020). Consistent with the effect of *bare* on bird song scenes we report, a previous horizontal structure study on bird song diversity found features such as forest canopy openness or forest gaps had significant effects (Pekin et al., 2012). Meanwhile, a comparison of the horizontal structural characteristics of natural and planted forests found that bird song diversity was higher in forest communities with higher vegetation cover (Smith et al., 2013), which is consistent with our finding that forest land use has a significant effect on the dominance of bird song scenes. Insect sound was also most influenced by the *city* and *forest*. Insect sounds were more related to the habitat in which they are found, which may be related to the range and ecological position of insects. In the silence scene, the top types were *grass*, *bare*, and *water*, which are all different forms of open land use and related in sound transmission characteristics. The results of related studies show that sound propagates farther in open environments, while both dense vegetation structures and urban building patterns affect propagation distance. The results of this study show that the dominance of silent scenes is more related to the open environment.

The bird-insect scene was not significant across the different radii of land use types. This suggests the bird-insect scene is not sensitive to variation at the landscape scale, perhaps because insect song frequency is higher. Previous studies show that insect song is more susceptible to the influence of understory vegetation (Hao et al., 2021b), and it may be that the bird-insect sound scene is more influenced by environmental factors at the community scale. On the other hand, it may be related to the monitoring season of this study as insect vocalizations in the southern subtropical region gradually decreases from October to February. At present, we have completed the investigation of forest community structure at the sample plot scale, and the next step will be to study relationships between its characteristics and acoustic scene dominance. We hope this provide further insight on the environmental factors affecting the dominance of bird-insect sound scenes.

RDA analysis is a multivariate statistical method widely used in the field of ecology for explaining relationships between environmental factors and species data (Zhao et al., 2022). In this study, based on the acoustic ecological niche hypothesis and the foundation of related studies (Luther and Wiley, 2009), the frequency intervals of each acoustic scene were used as species data, and the relationship between the frequency composition of each acoustic scene and the urban pattern was effectively explained by RDA analyses. This analysis showed that the frequency intervals of 1–2 kHz were mainly influenced by the urban land-cover type, indicating that low-frequency noise remains a key factor affecting vocalizing species in urban forests. Traffic noise, with its low frequency and long duration (Wang et al., 2019), has a significant effect on birdsong propagation, territory defense, and mate attraction (Des Aunay et al., 2014). Comparing insect and pure bird scenes revealed a narrower range of low frequencies in insect scenes, suggesting that birds actively change song frequency, amplitude, and pitch (Pieretti and Farina, 2013a) or adjust the timing and location of vocalizations (McClure et al., 2013) in response to interference from noise pollution. Meanwhile, comparing pure insect with insect-human scenes revealed that high frequency intervals were positively correlated with *forest* and negatively correlated with city land use. This result suggests that insect and bird communities enhance song frequency to cope with urban noise (Parris and Schneider, 2009). Also, the change of species may be one of the reasons for the change of frequency. Therefore, a future song species identification model will further explore the differences in the composition of vocalizing species in acoustic scenes and reveal individual strategies used for adapting to variably noisy environments. Finally, analyses of pure anthrophony found that the wide frequency interval of 1–6 kHz was correlated with city land use, indicating human-created sounds impact a large range of frequencies beyond just that created by traffic noise. This illustrates the difficulty vocalizing species face in adapting to highly complex and changing urban noise scenes.

5. Conclusion

In urban forests, the acoustic scene classification model and target sound area estimation method can rapidly and accurately analyze the relationship between biodiversity and anthropogenic activities, thus providing potential research directions in urban biodiversity and living environment research. Research based on deep learning and other artificial intelligence methods to analyze sound monitoring data will provide important scientific and technological support to sustainable urban development by helping to identify species sounds, develop sound landscape resources, and conserve soundscapes.

CRediT authorship contribution statement

Zezhou Hao: Methodology, Data curation, Writing – original draft, Writing – review & editing. **Haisong Zhan:** Methodology, Data curation. **Chengyun Zhang:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition. **Nancai Pei:** Conceptualization, Supervision, Writing – review & editing. **Bing Sun:** Supervision, Writing – review & editing. **Jihong He:** Supervision, Investigation, Writing – original draft. **Ruichen Wu:** Methodology, Data curation. **Xinhui Xu:** Methodology, Data curation. **Cheng Wang:** 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.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was funded by Forestry Science and Technology Innovation Fund of Guangdong Province (2021KJCX017), Program of Guangzhou Municipal Science and Technology Bureau (202102080556), GDAS' Special Project of Science and Technology Development (2020GDASYL-20200401001) and Foundation of the Research Project of Education Bureau of Guangzhou (202032882).

Appendix

See Figs. A1 and A2

References

Aiello, L.M., Schifanella, R., Quercia, D., Aletta, F., 2016. Chatty maps: constructing sound maps of urban areas from social media data. R. Soc. Open Sci. 3, 150690.

Alados, C.L., Navarro, T., Komac, B., Pascual, V., Martinez, F., Cabezudo, B., Pueyo, Y., 2009. Do vegetation patch spatial patterns disrupt the spatial organization of plant species? Ecol. Complexity 6, 197–207.

Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K.-H., Frommolt, K.-H., 2010. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. Pattern Recogn. Lett. 31, 1524–1534.

Brumm, H., 2004. The impact of environmental noise on song amplitude in a territorial bird. J. Anim. Ecol. 434–440.

Cao, J., Cao, M., Wang, J., Yin, C., Wang, D., Vidal, P.-P., 2019. Applications 2019, Urban noise recognition with convolutional neural network. Multimedia Tools 78, 29021–29041.

da Silva, P.G., Bogoni, J.A., Heino, J., 2020. Can taxonomic and functional metrics explain variation in the ecological uniqueness of ecologically-associated animal groups in a modified rainforest? Sci. Total Environ. 708, 135171.

Dawson, D.K., Efford, M.G., 2009. Bird population density estimated from acoustic signals. J. Appl. Ecol. 46, 1201–1209.

Dein, J., Rüdisser, J., 2020. Landscape influence on biophony in an urban environment in the European Alps. Landscape Ecol. 35, 1875–1889.

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. Ind. 13, 46–54.

Derryberry, E.P., Phillips, J.N., Derryberry, G.E., Blum, M.J., Luther, D., 2020. Singing in a silent spring: Birds respond to a half-century soundscape reversion during the COVID-19 shutdown. Science (New York, N.Y.) 370, 575–579.

Des Aunay, G.H., Slabbekoorn, H., Nagle, L., Passas, F., Nicolas, P., Draganoiu, T.I., 2014. Urban noise undermines female sexual preferences for low-frequency songs in domestic canaries. Anim. Behav. 87, 67–75.

Dong, Y., Ren, Z., Fu, Y., Hu, N., Guo, Y., Jia, G., He, X., 2022. Decrease in the residents' accessibility of summer cooling services due to green space loss in Chinese cities. Environ. Int. 158, 107002.

Eldridge, A., Casey, M., Moscoso, P., Peck, M., 2016. A new method for ecoacoustics? Toward the extraction and evaluation of ecologically-meaningful soundscape components using sparse coding methods. PeerJ 4, e2108.

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. Ind. 95, 939–952.

Fairbrass, A.J., Rennert, P., Williams, C., Titheridge, H., Jones, K.E., 2017. Biases of acoustic indices measuring biodiversity in urban areas. Ecol. Ind. 83, 169–177.

Farina, A., Pieretti, N., 2014. Sonic environment and vegetation structure: A methodological approach for a soundscape analysis of a Mediterranean maqui. Ecol. Inf. 21, 120–132.

Ghadiri Khanaposhtani, M., Gasc, A., Francomano, D., Villanueva-Rivera, L., Jung, J., Mossman, M., Pijanowski, B., 2019. Effects of highways on bird distribution and soundscape diversity around Aldo Leopold's shack in Baraboo, Wisconsin, USA. Landscape Urban Plann. 192, 103666.

Gibb, R., Browning, E., Glover-Kapfer, P., Jones, K.E., 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol. Evol. 10, 169–185.

Guedes, I.C.M., Bertoli, S.R., Zannin, P.H., 2011. Influence of urban shapes on environmental noise: a case study in Aracaju—Brazil. Sci. Total Environ. 412, 66–76.

Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A. L., Duke, C.S., Porter, J.H., 2013. Big data and the future of ecology. Front. Ecol. Environ. 11, 156–162.

Hao, Z., Wang, C., Sun, Z., van den Bosch, C.K., Zhao, D., Sun, B., Xu, X., Bian, Q., Bai, Z., Wei, K., 2021a. Soundscape mapping for spatial-temporal estimate on bird activities in urban forests. Urban For. Urban Green. 57, 126822.

Hao, Z., Wang, C., Sun, Z., Zhao, D., Sun, B., Wang, H., van den Bosch, C.K., 2021b. Vegetation structure and temporality influence the dominance, diversity, and composition of forest acoustic communities. Forest Ecol. Manage. 482, 118871.

Hong, X., Liu, J., Wang, G., Jiang, Y., Wu, S., Lan, S., 2019. Factors influencing the harmonious degree of soundscapes in urban forests: A comparison of broad-leaved and coniferous forests. Urban For. Urban Green. 39, 18–25.

Isbell, F., Gonzalez, A., Loreau, M., Cowles, J., Díaz, S., Hector, A., Mace, G.M., Wardle, D.A., O'Connor, M.I., Duffy, J.E., Turnbull, L.A., Thompson, P.L., Larigauderie, A., 2017. Linking the influence and dependence of people on biodiversity across scales. Nature 546, 65–72.

Joo, W., Gage, S.H., Kasten, E.P., 2011. Analysis and interpretation of variability in soundscapes along an urban–rural gradient. Landscape Urban Plann. 103, 259–276.

Kociolek, A., Clevenger, A., St. Clair, C., Proppe, D., 2011. Effects of road networks on bird populations. Conserv. Biol. 25, 241–249.

Krause, B., 2008. Anatomy of the soundscape: evolving perspectives. J. Audio Eng. Soc. 56, 73–80.

Krause, B., 2012. The great animal orchestra: finding the origins of music in the world's wild places. Hachette UK.

Kwan, C., Mei, G., Zhao, X., Ren, Z., Xu, R., Stanford, V., Rochet, C., Aube, J., Ho, K.C., 2004. Bird classification algorithms: theory and experimental results, 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. V-289.

Li, Z., Zhang, L., Du, S., Liu, W., 2018. Acoustic scene classification based on binaural deep scattering spectra with CNN and LSTM. The 2018 Workshop on Detection and Classification of Acoustic Scenes and Events.

Liu, J., Wang, Y., Zimmer, C., Kang, J., Yu, T., 2019. Factors associated with soundscape experiences in urban green spaces: A case study in Rostock, Germany. Urban For. Urban Greening 37, 135–146.

Luther, D.A., Wiley, R.H., 2009. Production and perception of communicatory signals in a noisy environment. Biol. Lett. 5, 183–187.

Masood, E., 2018. Battle over biodiversity. Nature 560, 423–425.

McClure, C.J., Ware, H.E., Carlisle, J., Kaltenecker, G., Barber, J.R., 2013. An experimental investigation into the effects of traffic noise on distributions of birds: avoiding the phantom road. Proc. R. Soc. B: Biol. Sci. 280, 20132290.

Mennill, D.J., Battiston, M., Wilson, D.R., Foote, J.R., Doucet, S.M., 2012. Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour. Meth. Ecol. Evol. 3, 704–712.

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. Ind. 109, 105793.

Mullet, T.C., Gage, S.H., Morton, J.M., Huettmann, F., 2016. Temporal and spatial variation of a winter soundscape in south-central Alaska. Landscape Ecol. 31, 1117–1137.

Nemeth, E., Brumm, H., 2010. Birds and anthropogenic noise: are urban songs adaptive? Am. Nat. 176, 465–475.

Nemeth, E., Pieretti, N., Zollinger, S.A., Geberzahn, N., Partecke, J., Miranda, A.C., Brumm, H., 2013. Bird song and anthropogenic noise: vocal constraints may explain why birds sing higher-frequency songs in cities. Proc. R. Soc. B-Biol. Sci. 280, 20122798.

Orci, K.M., Petróczki, K., Barta, Z., 2016. Instantaneous song modification in response to fluctuating traffic noise in the tree cricket Oecanthus pellucens. Anim. Behav. 112, 187–194.

Parris, K.M., Schneider, A., 2009. Impacts of traffic noise and traffic volume on birds of roadside habitats. Ecol. Soc. 14.

Peckens, C., Porter, C., Rink, T., 2018. Wireless sensor networks for long-term monitoring of urban noise. Sensors 18, 3161.

Pekin, B.K., Jung, J., Villanueva-Rivera, L.J., Pijanowski, B.C., Ahumada, J.A., 2012. Modeling acoustic diversity using soundscape recordings and LIDAR-derived metrics of vertical forest structure in a neotropical rainforest. Landscape Ecol. 27, 1513–1522.

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.

Pijanowski, B.C., Villanueva-Rivera, L.J., Dumyahn, S.L., Farina, A., Krause, B.L., Napoletano, B.M., Gage, S.H., Pieretti, N., 2011. Soundscape ecology: the science of sound in the landscape. Bioscience 61, 203–216.

Priyadarshani, N., Marsland, S., Castro, I., 2018. Automated birdsong recognition in complex acoustic environments: a review. J. Avian Biol. 49, jav-01447.

Quinn, C.A., Burns, P., Gill, G., Baligar, S., Snyder, R.L., Salas, L., Goetz, S.J., Clark, M.L., 2022. Soundscape classification with convolutional neural networks reveals temporal and geographic patterns in ecoacoustic data. Ecol. Ind. 138, 108831.

Ren, Z., Fu, Y., Dong, Y., Zhang, P., He, X., 2022a. Rapid urbanization and climate change significantly contribute to worsening urban human thermal comfort: A national 183-city, 26-year study in China. Urban Clim. 43, 101154.

Ren, Z., Zhao, H., Fu, Y., Xiao, L., Dong, Y., 2022b. Effects of urban street trees on human thermal comfort and physiological indices: a case study in Changchun city, China. J. For. Res. 33, 911–922.

Sethi, S.S., Jones, N.S., Fulcher, B., 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. PNAS 117, 17049–17055.

Shannon, G., McKenna, M.F., Angeloni, L.M., Crooks, K.R., Fristrup, K.M., Brown, E., Warner, K.A., Nelson, M.D., White, C., Briggs, J., 2016. A synthesis of two decades of research documenting the effects of noise on wildlife. Biol. Rev. 91, 982–1005.

Slabbekoorn, H., 2004. Habitat-dependent ambient noise: consistent spectral profiles in two African forest types. J. Acoust. Soc. Am. 116, 3727–3733.

Slabbekoorn, H., Ripmeester, E.A.P., 2008. Birdsong and anthropogenic noise: implications and applications for conservation. Mol. Ecol. 17, 72–83.

Smith, T.B., Harrigan, R.J., Kirschel, A.N., Buermann, W., Saatchi, S., Blumstein, D.T., de Kort, S.R., Slabbekoorn, H., 2013. Predicting bird song from space. Evol. Appl. 6, 865–874.

Soranno, P.A., Schimel, D.S., 2014. Macrosystems ecology: Big data, big ecology. Front. Ecol. Environ. 12, 3.

Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., 2008. Rapid acoustic survey for biodiversity appraisal. PLoS ONE 3, e4065.

Torija, A.J., Ruiz, D.P., Ramos-Ridao, Á.F., 2014. A tool for urban soundscape evaluation applying support vector machines for developing a soundscape classification model. Sci. Total Environ. 482, 440–451.

Tucker, D., Gage, S.H., Williamson, I., Fuller, S., 2014. Linking ecological condition and the soundscape in fragmented Australian forests. Landscape Ecol. 29, 745–758.

Ulloa, J.S., Gasc, A., Gaucher, P., Aubin, T., Réjou-Méchain, M., Sueur, J., 2016. Screening large audio datasets to determine the time and space distribution of Screaming Piha birds in a tropical forest. Ecol. Inf. 31, 91–99.

Ulloa, J., Haupert, S., Latorre Gil, J., Aubin, T., Sueur, J., 2021. scikit-maad: An open-source and modular toolbox for quantitative soundscape analysis in Python. Methods Ecol. Evol. 12, 1–7.

Wang, H., Cai, M., Cui, H., 2019. Simulation and analysis of road traffic noise among urban buildings using spatial subdivision-based beam tracing method. Int. J. Environ. Res. Public Health 16, 2491.

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. The Condor: Ornithol. Appl. 119, 73–84.

Zhao, Y., Sheppard, S., Sun, Z., Hao, Z., Jin, J., Bai, Z., Bian, Q., Wang, C., 2022. Soundscapes of urban parks: An innovative approach for ecosystem monitoring and adaptive management. Urban For. Urban Greening 71, 127555.