Ecological Informatics 95 (2026) 103739

Contents lists available at ScienceDirect

Ecological Informatics

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

Studying domain shifts in bioacoustic recording style between focal and soundscape recordings with acoustic indices

Sean Perry [a,*], Tianqi Zhang [a], Siya Kamboj [a], Anu Jajodia [b], Dhruv Tomar [c], Ryan Kastner [a]

[a] University of California, San Diego, 9500 Gilman Drive, La Jolla, 92093, CA, United States of America
[b] Carleton College, Northfield, 55057, MN, United States of America
[c] University of Illinois at Urbana-Champaign, 901 W Illinois St, Urbana, 61801, IL, United States of America

ARTICLE INFO

Dataset link: github.com/UCSD-E4E/egci_bioacoustic_shifts

Keywords:
Bioacoustics
Domain shift
Machine learning
Focal recordings
Soundscape recordings
Passive acoustic monitoring

ABSTRACT

A major challenge in machine learning (ML) for studying passive acoustic monitoring is the domain shift between directional, handheld, well-labeled focal audio data and omnidirectional, passively recorded, under-labeled soundscapes. ML models trained on the vastly more labeled focal recording datasets struggle to generalize when tested on soundscape recordings. Research in this area has typically focused on the methodology behind training ML models, both to study and alleviate the domain shift. However, less work has been done in quantifying the domain shift hypothesis beyond model performance and species density. This potentially limits our understanding of the factors behind domain shift. We find that acoustics indices like the Ecoacoustic Global Complexity Index (EGCI), an acoustic complexity index rooted in information theory, can effectively describe the domain shift and assist in productive research of this domain shift. We demonstrate in two experiments over BirdSet how EGCI, Acoustic Complexity Index, Acoustic Diversity Index, and the Bioacoustic Index can be used to quantify domain shift. We find that an SVM can be trained to classify successfully between soundscape and focal recordings using these quantifiers as features (achieving, for instance, an average accuracy of 83.33 % with EGCI values) and that each index has significant divergence with Kolmogorov–Smirnov (KS) testing between focal and soundscape recordings. To understand how these domain shift quantifiers affect model loss, we use correlation, feature importance in random forest, and linear regression to find relationships between these quantifiers of domain shift and model loss.

1. Introduction

Audio recordings are increasingly used in ecological and conservation sciences (Pijanowski et al., 2011; Farina et al., 2024). Passive acoustic monitoring, the process of using microphones to collect field recordings of ecologically significant data, has been on the rise as cheaper hardware, large benchmarking datasets, and developments in machine learning (ML) make large-scale bioacoustic surveys increasingly feasible (Dufourq et al., 2022; Stowell, 2018; Mutanu et al., 2022; Clark et al., 2023; Kahl et al., 2021b). These factors have enhanced the ability to monitor endangered species and ecosystems and improve the study of species behavior (Wood et al., 2023, 2024; Kramer et al., 2024; Teixeira et al., 2019; Penar et al., 2020).

Despite this progress, a critical challenge limiting the use of bioacoustics is the covariate shift between large public bioacoustic datasets and recordings collected during field deployments. In machine learning, covariate or domain shifts occur when the training data differ from the intended use case, causing models to struggle with generalization (Quinonero-Candela et al., 2022; Y et al., 2019). In bioacoustics, one such domain shift typically occurs because the training data are focal recordings, whereas the testing data are soundscapes from audio recordings in the same region of the world.

Focal recordings are audio captured from directional microphones held by human recordists (Rauch et al., 2025a). As a result, the recordings are fairly clean and contain little noise, and thus have a high signal-to-noise ratio (SNR). Focal recordings make up a significant portion of the large, labeled, and publicly available scientific audio recordings from sources such as xeno-canto or iNaturalist have aided in the development of bioacoustic machine learning. However, the use case for machine learning models in passive acoustic monitoring is for *soundscape recordings*, which are recorded from omnidirectional microphones using some automated recording process. By nature of being passively recorded, they indirectly record the entire environment around capture the soundscape around them, resulting in stronger distortions due to audio propagation and recordings (Rauch et al., 2025a).

* Corresponding author.
E-mail addresses: shperry@ucsd.edu (S. Perry), kastner@ucsd.edu (R. Kastner).

https://doi.org/10.1016/j.ecoinf.2026.103739
Received 11 September 2025; Received in revised form 22 March 2026; Accepted 22 March 2026
Available online 5 April 2026

Given the large amount of labeled focal recordings and the small amount of labeled soundscapes, many models are trained on focal recordings and tested for generalization on soundscapes. As found by and large in the field, such models tend to perform better on holdout sets of focal recordings than holdout soundscape datasets. It is hypothesized that the difference between the active and passive recording styles creates this domain shift and results in lower model performance (Boudiaf et al., 2023; Goëau et al., 2018; Kahl et al., 2021a; Liang et al., 2024; Ghani et al., 2025; Chasmai et al., 2024; Rauch et al., 2025a; Denton et al., 2021).

However, the descriptions of the two recording styles lack specificity. There is also limited research into characterizing the domain shift, begging the question of what specific properties/metrics distinguish focal from soundscape. Previous work largely focused on technical solutions that improve model performance, such as data augmentation and domain adaptation techniques (Rauch et al., 2025a; Boudiaf et al., 2023; Kahl et al., 2021a). This potentially means more understanding into the domain shift can be gained by changing our measurement of domain shift. In this paper, we explore ways to formally quantify the domain shift between focal and soundscape recordings in bioacoustics (RQ1).

There has also been work comparing the soundscapes of different sites and regions using complexity indices, a collection of heuristics aimed at identifying the amount of biophony, i.e., sounds generated by non-human organisms in a given region (Budka et al., 2023; Bradfer-Lawrence et al., 2023). In particular, the *Ecoacoustic Global Complexity Index (EGCI)* uses methods from information theory and ecology, in particular, entropy and complexity derived from the raw audio data to characterize soundscape recordings (Colonna et al., 2020). The idea was that sites with greater biodiversity have more varied acoustic activity, resulting exhibit more varied acoustic activity, resulting in higher entropy and greater audio recording characterize the differences in the audio data itself between various recording sites from a given deployment.

We propose using acoustic complexity indices to measure differences in recording styles that underlie the domain shift between focal and soundscape recordings, rather than using them to measure the differences between sites. Given that the hypotheses surrounding the bioacoustic domain shift center on noise, an unsupervised method that measures entropy and complexity can separate these domains using information-theoretic quantifiers. Furthermore, since soundscapes with greater ecoacoustic activity are more complex (Colonna et al., 2020), it is possible that focal recordings, which provide clearer, more direct information, may also be more complex than soundscapes. If this is true, it could provide a meaningful divergence from which to study domain shift across various audio recording styles.

We propose that acoustic indices can be used to better quantify domain shift between clear focal recordings and noisy soundscape recordings, without the need to quantify the original species call in the audio data. We demonstrate using the bioacoustic domain shift benchmarking dataset BirdSet (Rauch et al., 2025a) that EGCI and other acoustic complexity metrics can meaningfully separate focal from soundscape recordings and in particular entropy can be a meaningful predictor of model performance in current SOTA models. We provide all code and precomputed EGCI metrics of BirdSet data on GitHub so others may use these techniques for their datasets at https://anonymous.4open.science/r/egci_bioacoustic_shifts-7021/

Research Questions

- RQ1: How does one measure the domain shift between focal and soundscape recordings in bioacoustics?
- RQ2: Can acoustic indices meaningfully separate focal and soundscape recordings?
- RQ3: What can acoustic indices determine about the recording styles between focal and soundscape recordings?
- RQ4: How do acoustic indices correlate with model performance on soundscapes?

2. Related work

2.1. Bioacoustic domain shift

Domain shifts between soundscape and focal recordings were initially noted in the early BirdCLEF competitions, particularly around 2016 and 2017, when organizers observed that test focal recordings had significantly higher performance than soundscape recordings from various regions worldwide (Goëau et al., 2016, 2017). Here is where the initial hypotheses of focal to soundscape domain shift arose: the differences in SNR, number of species, and so on. However, the original BirdNET paper is more commonly cited in the literature as the basis for domain shift (Kahl et al., 2021b). BirdNET was created in the space of the early BirdCLEF competitions, and thus, the ideas regarding domain shift likely arrived from these early issues in the BirdCLEF competitions.

Focal-to-soundscape domain shift is characterized primarily as a difference in devices and recording methods, collectively called "recording style". Focal recordings are described as being hand recorded by listening and directing a microphone towards a "focal" species. In contrast, soundscape recordings are passively recorded using microphones that are omnidirectional (Boudiaf et al., 2023; Goëau et al., 2018; Kahl et al., 2021a; Liang et al., 2024; Ghani et al., 2025; Chasmai et al., 2024; Rauch et al., 2025a; Denton et al., 2021). The result is demonstrated in Fig. 2, where the focal recording can contain a much stronger signal than the soundscape.

Beyond domain shifts in BirdSet, other changes can occur in bioacoustics. Among these are regional domain shifts, not simply with different environments but as well as weather and regional dialects (Bidarouni and Abeßer, 2024; Heinrich et al., 2025). To further complicate matters, other shifts that BirdSet also intends to address, which are not the focus of this paper, are label uncertainty (error in the label creation process) and task shift (multiclass vs. multilabel classification) (Rauch et al., 2025a). Other competitions beyond BirdCLEF, e.g., DCASE, have also identified microphone quality as a potential domain shift (Dohi et al., 2022).

That motivated the next few years of work on domain adaptation for bioacoustics, as researchers sought to improve the aforementioned hypotheses. Techniques such as reverb, adding pink noise, and MiXiT have been used to improve the domain shift (Rauch et al., 2025a; Somervuo et al., 2023; Boudiaf et al., 2023; Kahl et al., 2021a). The techniques have been helpful, providing additional evidence to support the original hypothesis. For example, using reverb from real-world impulses suggests that distance is a problem for soundscape recordings (Somervuo et al., 2023). Beyond augmentation, domain-specialized domain adaptation techniques have also domain-specific adaptation techniques have also attracted significant interest in the field, such as pseudo-labeling and sampling methods improvements with these techniques are still limited, work has continued to from these techniques are still limited, work has continued today through ongoing of a dataset from the BirdCLEF competitions for easier testing on domain shift (Rauch et al., 2025a).

2.2. Analyzing the hypothesis of domain shift

While bioacoustic domain shift research is primarily focused on mitigating its effects, there is a notable lack of study on the datasets underlying the domain shift. The best work in this direction focuses on SNR: signal-to-noise ratio. Focal recordings, akin to the ones in BirdSet are described as having high SNR whereas soundscapes, due to sound attenuation, has lower SNR [citation needed]. Work has additionally shown that at different distances, as SNR declines so too does birdnet model performance as BirdNet was primarily trained on high-SNR focal recordings (Pérez-Granados, 2023; Michaud et al., 2025). The key limiting factor of SNR is it requires a labeled signal to calculated. The stated goal of this work is to identify methods that can work without annotated data such that the methods maybe usable with unannotated

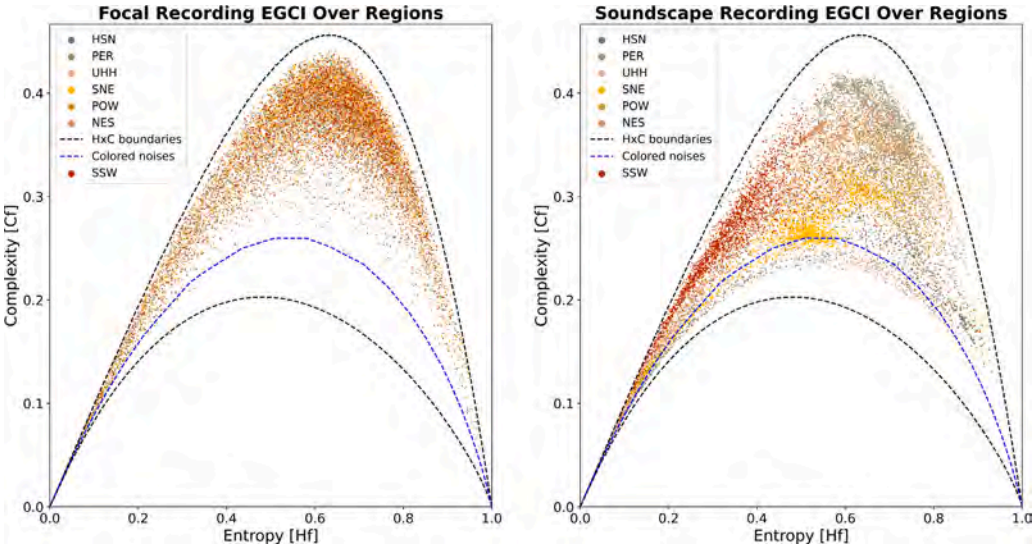

Fig. 1. The differences between focal and soundscape recordings using EGCI over the regions of interest created from the visualization code in Colonna et al. (2020). The seven different regions of BirdSet are plotted in different colors. Note that EGCI values exist in a non-linear space defined by the curves. Theoretical colored noise plots are plotted for clarity. Visually, the focal and soundscape recordings show differences, both overall and with respect to the seven regions. Focal recordings appear to typically be higher in complexity and entropy than their soundscapes' counterparts, with the exception of PER which has high complexity, potentially due to the fact the recordings were done during the dusk-dawn chorus with high species activity (Rauch et al., 2025a). Yet, for all regions, focal EGCI is incredibly homogeneous in its distribution compared to the soundscapes, which is the graphic one might expect from audio recordings from various parts of the world, making focal recordings highly unusual. For an look of this plot by individual region, see Appendix D. This graphic is one of the motivating forces behind this paper.

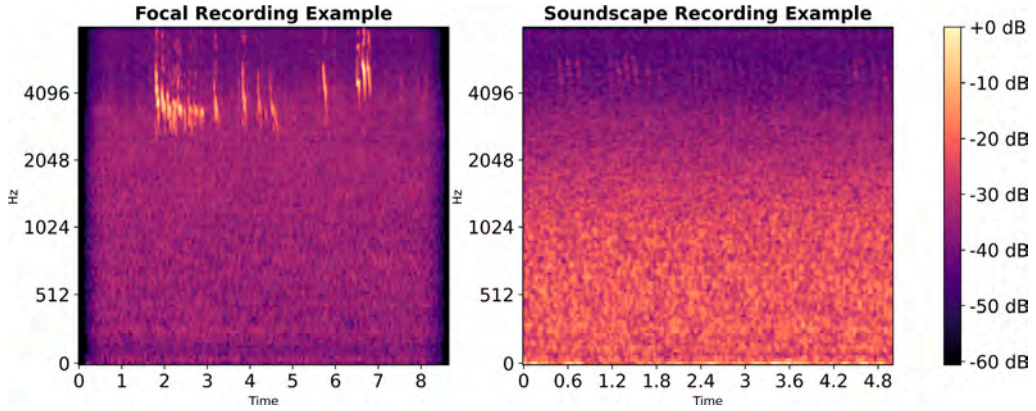

Fig. 2. Example of a focal and soundscape recording of (Buff-bellied Pipit *Anthus rubescens*) from the HSN dataset split of BirdSet. Observe how the signal is much clearer in the focal recording whereas the soundscape recording is distorted and faded into the background. The focal recording is from Xeno-Canto (XC358862) by the recordist (Marvin, 2017). The soundscape recording is from the file HSN_005_20150709_063105_245_250.ogg (Rauch et al., 2025a). Visualizations are based on the librosa visualization code.

soundscapes or focal recordings. Therefore, we study this work without SNR in this paper.

One common method studies differences in metadata. A common finding from metadata differences is that focal recordings have different distributions of species counts and spatial distributions than their soundscape counterparts due to biases from actively recording data; citizen scientists tend to record data in areas that are easier to access and easier to label (Mair and Ruete, 2016; van Merriënboer et al., 2024b). This shows that the data-generating processes behind focal and soundscape datasets differ. However, metadata analysis does not address the noise or features found in the audio data.

Regarding the measurement of differences between audio recordings, a similar study was conducted by Bidarouni and Abeßer, who investigated not focal-to-omnidirectional domain shift, but rather shifts between soundscape locations (Bidarouni and Abeßer, 2024). The paper focused on quantifying properties in the spectrogram distributions to predict model loss on a target domain. They conduct similar analyses to those we propose, namely, finding a metric to describe the domain shift. Experiment 2 (Section 4.2) builds upon their methodology to predict loss and use it to interpret what is occurring within the domain shift.

2.3. Acoustic complexity indices

Before the use of machine learning for bioacoustic research, significant time was spent in ecology attempting to develop a set of heuristics to describe the content of audio recordings without the need for an intelligent method. The resulting techniques are referred to as Acoustic Complexity Indices. These methods are named after the first index in the field, the Acoustic Complexity Index (ACI), which uses an algorithm that examines the intensity of spectrogram cells relative to nearby cells to determine correlations with ecologically relevant sound data (Pieretti et al., 2011; Farina et al., 2021). Since then, numerous ACI variants have been created, each attempting to address

various weaknesses associated with ACI as a proxy for bioacoustic activity (Bradfer-Lawrence et al., 2019). These variant ACIs could serve as proxies for the health of a given environment. In these cases, ACI can be interpreted as a method for studying acoustic differences across audio datasets.

If the goal is to better understand the differences in information between audio datasets, one ACI stands out: the *Ecoacoustic Global Complexity Index (EGCI)* (Colonna et al., 2020). Rather than using a spectrogram, EGCI computes an autocorrelation matrix from the signal, which is used to calculate Von Neumann Entropy and Statistical Complexity. *Statistical Complexity* is conceptually similar to the complexity of physical systems, where complexity is considered high if it is neither random noise (high entropy) nor orderly (low entropy) (Shiner et al., 1999; Rosso et al., 2007; López-Ruiz et al., 1995). Since complexity is defined in terms of entropy, plotting entropy against complexity yields two concave curves that bound the entropy and complexity values (López-Ruiz et al., 1995). Colonna et al. found that higher complexity and higher entropy potentially imply higher ecological activity (Colonna et al., 2020).

Previous work with ACIs, including EGCI, has focused on describing ecoacoustic activity within an environment or studying differences between various soundscapes within a region (Gasc et al., 2013; Colonna et al., 2020). We propose that these indices be further applied as a data exploration tool to study the audio content differences in recording styles between soundscapes and focal recordings.

3. Methodology

3.1. Dataset

BirdSet (Rauch et al., 2025a) is a collection of various datasets used in previous BirdCLEF competitions, which serve as the basis for much of the work done in domain shift studies (Boudiaf et al., 2023). The dataset is split into separate regions; our analysis uses the High Sierra Nevada (HSN), Madre De Dios Peru (PER), Powdermill Forest (POW), Sierra Nevada (SNE), Columbia and Costa Rica soundscapes (NES), Sapsucker Woods (SSW), and Hawaii (UHH) dataset splits. Each region's data is sampled at 32 kHz and contains focal recordings and soundscapes datasets, referred to as *train* and *test_5s*, respectively. Note that the train split of these datasets is from xeno-canto and consists only of focal recordings (Rauch et al., 2025a,b). For focal recordings, a random 5-second segment is extracted from each sampled clip. For each region, 2000 random focal clips are sampled, along with 2000 of the test soundscape clips, which are already cut into 5-second segments. It should be noted that this is a simple random sample that is not stratified by class, as we wish to demonstrate how this can be applied on unlabeled datasets that may or may not be balanced. In practice, soundscape and focal recordings datasets do not share the same distribution of species populations (Rauch et al., 2025a; van Merriënboer et al., 2024a), highlighting a part of the domain shift problem in bioacoustics. All data is loaded via the librosa library.

3.2. EGCI setup, implementation, and exploration

We use an *nlag* of 256, which specifies the number of steps ahead used to compute the autocorrelation matrix, as in the original EGCI paper (Colonna et al., 2020). However, we make a few changes to the EGCI algorithm's implementation. Namely, when computing the Eigenvalues of the autocorrelation matrix via singular value decomposition, we use NumPy's SVD implementation rather than scikit-learn, which was slower to compute as scikit-learn does not provide parallelization for the singular value decomposition computation.

3.3. Other indices tested

We will also experiment with other acoustic features to gauge their ability to address domain shift conflicts. These include the following:

- **Number of Species Present (NumSpecies)**: As noted in previous literature, species density can be a strong indicator of model performance (Kahl et al., 2021b,a). Therefore we will include this in our experiments to understand the effect it may have when addressing domain shift experiments.
- **Acoustic Complexity Index (ACI)** (Pieretti et al., 2011): Directly targeted at the presence of measuring the amount of bird vocalizations in audio recordings. This index relies on a algorithm that computes intensity values across the frequency of the spectrogram (Pieretti et al., 2011).
- **Acoustic Diversity Index (ADI)** (Villanueva-Rivera et al., 2011): Whereas EGCI Entropy computes Von Neumann entropy based on the autocorrelation of a audio signal, this computes the Shannon entropy of the frequency bands in a spectrogram (Colonna et al., 2020; Bradfer-Lawrence et al., 2023)
- **Bioacoustic Index (BI)** (Boelman et al., 2007): This index, to quote Budka et al., "measures the area under the log spectrum curve" (Budka et al., 2023) to find correlation to biocoustic activity.

These indices were chosen as they are frequently used, often times together, in literature (Budka et al., 2023; Bradfer-Lawrence et al., 2023). It should be noted this is not an exhaustive list, and this analysis could be potentially used over a myriad of indices. Use the below experiments as guides for how one can approach studying domain shift with hand crafted features. The implementation of these acoustic indices were used by Haupert et al. (2025)

3.4. Last general note regarding experiments

Each experiment will be run twice with 2 conditions: once sampling from an region's soundscape recordings, and again this time sampling from a dataset of only the soundscapes of a given region that contain only 1 labeled bird. This is done to address two different perspectives. The first again is to learn how to separate unlabeled bioacoustics datasets were all that is known is focal vs. soundscape recordings. A metric here maybe successful if it can be used without concern for labeling. The second reason is there that if you do not account for multiple species, it becomes more difficult to evaluate metrics for domain shift if the differences may also be caused by multiple species. Thus both approaches are demonstrated throughout the paper.

4. Experiments

4.1. Experiment 1: Predicting soundscape vs. Focal recordings

4.1.1. Experiment design

The first experiment addresses RQ1 and RQ2 by demonstrating that complexity indices can reveal differences in soundscapes and focal recordings within a given region. We conduct a permutation test in which we randomize the predictions of a binary classifier trained to classify between focal and soundscape recordings using EGCI and the other acoustic indices as features. We also perform a KS test between the empirical distribution of index metrics for focal and soundscape recordings. Finally, we test in which direction the average metric score changes when moving between focal and soundscapes recordings using a permutation test. Fig. 3 shows the experimental workflow.

We train SVM classifiers with an RBF kernel for each region on the entropy and complexity of 2000 random focal clips and 2000 random soundscape clips (recall this is done twice, once for any possible soundscape and again for soundscapes of only 1 annotation, see Section 3.4).

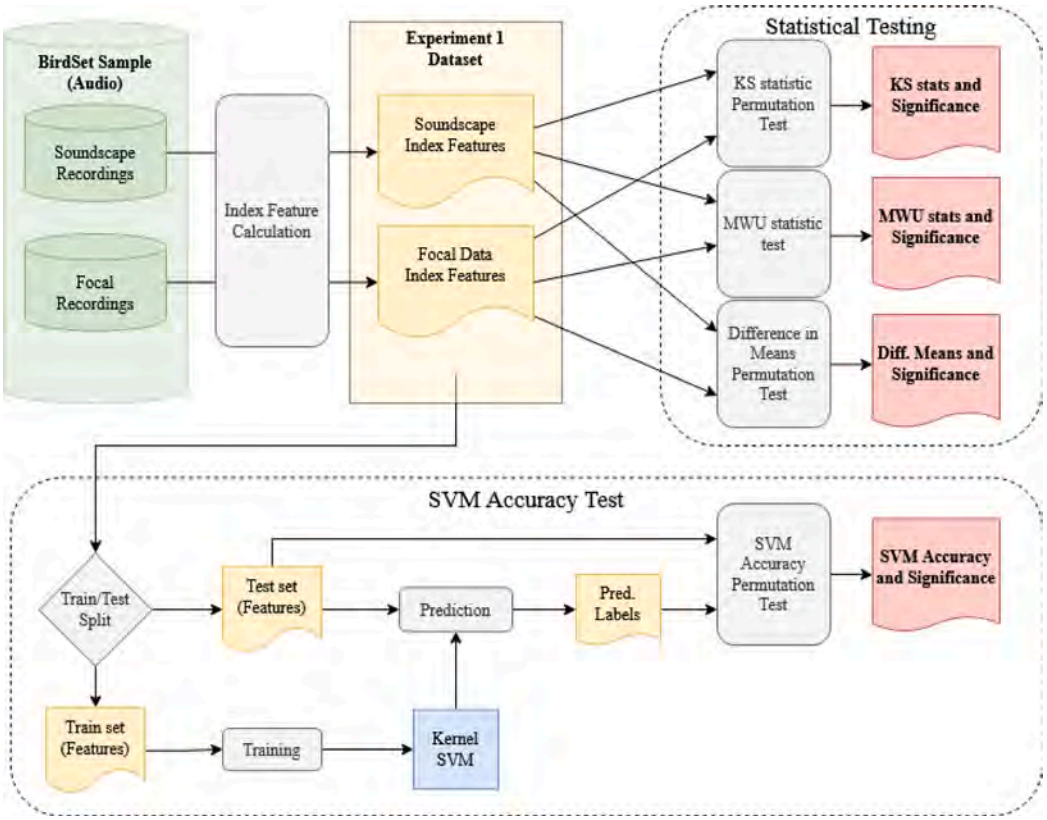

Fig. 3. Experimental workflow for Experiment 1 on a single region. BirdSet Sample contains 2000 sampled soundscape and focal audio clips from a single region.

We perform an 80/20 train/test split and record accuracy on the test set. For the permutation test over 1000 trials, we randomize the test true labels and compare against the model output. The null hypothesis is that EGCI and other acoustic indices are unable to distinguish between focal and soundscape conditions, so the models would perform at random chance. We will examine the proportion of randomized-label trials that achieved an accuracy greater than our observed accuracy in each region, at the significance level $\alpha = 0.05$. In this case, accuracy is defined as the number of correct predictions divided by the total number of examples. This experiment is conducted for each pair of possible acoustic indices (Such as entropy and complexity, ACI and BI, ACI and entropy, etc.). Our goal for studying pairs of acoustic indices comes from the inspiration for this work, namely EGCIs, as they were just a pair of entropy and complexity (Colonna et al., 2020). If just two features can identify a soundscape or focal recording well, then that maybe more informative than if a feature vector of a dozen or so features are able to classify it.

As a secondary test, we will also check for a significant difference in the empirical distributions of each index between soundscape and focal recordings using the same 2000 random pairs (see Section 3.4). We perform a KS test and Mann–Whitney U test (via scipy and asymptotic method for computing Virtanen et al., 2020) for difference in distributions at a significance level of $\alpha = 0.05$ (Massey, 1951; Mann and Whitney, 1947) per acoustic index.

Finally, to demonstrate how these indices can be used to understand the domain shift, we perform a permutation test to determine whether the mean of a given index distribution in soundscapes is lower than the mean in focal recordings. To do this, we take the same 2000 random clips (see Section 3.4) and compute a difference in means as our test statistic. Then, we permute the labels of the focal and soundscape recording 1000 times and calculate the difference in means for each permuted trial. To compute the p-value, we compute the proportion of simulated test statistics that are equally or more extreme than the observed test statistic with a significance level of $\alpha = 0.05$.

Table 1
The sample sizes per region given we filtered if there were at least one labeled bird in a soundscape.

Region	Focal NumSamples	Soundscape NumSamples	Bird-only soundscape subset NumSamples
HSN	2000	2000	892
PER	2000	2000	1789
UHH	2000	2000	1420
SNE	2000	2000	1383
POW	2000	2000	1945
NES	2000	2000	787
SSW	2000	2000	639

For the KS test, Mann–Whitney U test, and Difference in Means permutation test, we repeated the experiment when filtering the initial 2000 samples in soundscapes (see Section 3.4) for birds only and reported the scores to determine the impact unlabeled audio had versus labeled audio. This means there are fewer samples in the bird-only soundscape data. Table 1 reports the number of samples for the bird-only experiments.

4.1.2. Results

SVM performance. With the SVM performance, most features showed a significant signal for classifying between focal versus soundscape recordings, and models achieved above chance performance (p-values were close to 0). Fig. 4 shows the performance for the HSN scores. Scores for other regions (and for the 1 species selected test runs) can be found in Appendix A. As shown in the HSN section, overwhelming from Fig. 4, the Bioacoustic Index score is well above 90% regardless of which feature the index was paired with. Most features scored reasonably (80%–90%) when paired with only one other feature, implying that even a small number of ecoacoustic features contain enough information to successfully distinguish between the two classes.

Fig. 4. HSN Scores over each pair of features. Along the diagonal were SVM trained from a single variable. For the other regions, please refer to the Appendix. All p-values were 0 (better than chance at detecting focal vs. soundscape recordings).

Table 2
KS test results for each index per region. KS statistics are shown alongside p-value for each index and region. Note there is significant difference ($p < 0.001$) in all indices.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
HSN stat	0.254500	0.658000	0.847000	0.562500	0.654000	0.554000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
PER stat	0.335000	0.199000	0.621000	0.631500	0.535000	0.707000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
UHH stat	0.381000	0.616000	0.423000	0.319500	0.524500	0.290000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
SNE stat	0.410500	0.732000	0.871000	0.738000	0.794500	0.315500
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
POW stat	0.215000	0.430500	0.365500	0.646000	0.362000	0.827000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
NES stat	0.431000	0.324500	0.645500	0.584500	0.445000	0.606500
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
SSW stat	0.684000	0.640000	0.561500	0.248500	0.693500	0.680500
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Table 3
KS test results for each index per region filtered for bird only. KS statistics are shown alongside the p-value for each region. In this test, audio clips with no bird call were filtered out. There is again a significant difference ($p < 0.001$) in all metrics.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
HSN stat	0.260287	0.648493	0.784029	0.540794	0.656000	0.153587
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
PER stat	0.340454	0.219617	0.621737	0.623544	0.533849	0.790386
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
UHH stat	0.348113	0.564831	0.400704	0.313683	0.512415	0.377465
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
SNE stat	0.402342	0.737739	0.853440	0.757887	0.796377	0.456255
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
POW stat	0.217670	0.427418	0.355424	0.642815	0.361335	0.850386
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
NES stat	0.415444	0.374077	0.590444	0.505939	0.454940	0.156290
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
SSW stat	0.726779	0.661064	0.584537	0.186166	0.692740	0.283255
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

KS test. Table 2 shows the KS Test results. The focal and soundscape recordings across all features and regions exhibit statistically significant differences in distribution (at $\alpha = 0.05$). Likewise, when filtering soundscapes for recordings that contain labeled birds and the separate split of single 1 labeled annotations in soundscapes, we see a similar performance, as shown in Tables 3 and 4.

Mann–Whitney U tests. Likewise to the KS-test, the Mann–Whitney U test results are found in Tables 5, 6, and 7 and like the KS-test, the results were statistically significant in all tests.

Permutation in a difference of means. Finally, we observe that for the difference of means test in all 2000 sampled clips in Table 8, the filtered

Table 4
KS test results for the secondary test against 1 bird samples from soundscapes to focal recordings, similar to previous test, results are consistent, other than for NumSpecies which as expected is not relevant for determining focal or soundscape information with single species in both distributions.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
SSW stat	0.702000	0.651000	0.557500	0.215500	0.684500	0.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
HSN stat	0.221000	0.646500	0.786000	0.573500	0.646500	0.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
PER stat	0.407000	0.157000	0.697500	0.673000	0.531000	0.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
UHH stat	0.377500	0.657500	0.441500	0.360500	0.529000	0.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
SNE stat	0.495325	0.739395	0.879973	0.750452	0.802404	0.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
POW stat	0.096000	0.488000	0.447500	0.660000	0.383000	0.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
NES stat	0.419000	0.349500	0.589000	0.506500	0.439500	0.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000

Table 5
Mann–Whitney U test for all soundscape vs. focal for each feature.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
HSN stat	2 327 710.000000	3 569 287.000000	3 838 392.000000	877 855.000000	2 707 892.000000	2 971 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
PER stat	1 319 566.000000	1 416 031.000000	3 540 090.000000	499 501.000000	1 607 981.000000	797 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
UHH stat	2 827 877.000000	3 444 658.000000	3 065 323.000000	1 397 570.000000	2 449 056.000000	2 044 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.126299
SNE stat	2 959 677.000000	3 470 035.000000	3 866 855.000000	383 939.000000	3 345 341.000000	1 986 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.639375
POW stat	1 472 013.000000	2 884 247.000000	2 978 828.000000	454 947.000000	1 544 542.000000	401 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
NES stat	2 993 724.000000	2 806 778.000000	3 596 421.000000	706 312.000000	1 680 979.000000	3 090 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
SSW stat	3 467 041.000000	3 434 311.000000	3 218 431.000000	1 646 184.000000	2 849 442.000000	3 180 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Table 6
Mann–Whitney U test for soundscape of only samples with non-zero amount of labeled species vs. focal for each feature.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
HSN stat	1 041 556.000000	1 582 110.000000	1 674 489.000000	399 924.000000	1 207 995.000000	755 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
PER stat	1 192 028.000000	1 208 638.000000	3 172 313.000000	455 636.000000	1 431 847.000000	375 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
UHH stat	1 958 085.000000	2 381 244.000000	2 142 596.000000	964 389.000000	1 734 237.000000	884 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
SNE stat	2 009 289.000000	2 402 011.000000	2 663 154.000000	223 668.000000	2 310 652.000000	752 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
POW stat	1 421 795.000000	2 795 384.000000	2 875 229.000000	448 901.000000	1 496 715.000000	291 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
NES stat	1 173 469.000000	1 140 579.000000	1 373 575.000000	334 745.000000	635 523.000000	664 000.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
SSW stat	1 134 341.000000	1 111 877.000000	1 049 367.000000	559 155.000000	910 241.000000	458 000.000000
pvalue	0.000000	0.000000	0.000000	0.000002	0.000000	0.000000

soundscape test for only birds in Table 9, and the secondary test run against the only 1 annotation soundscapes Table 10. The difference in means between soundscapes and focal tends to be negative, implying that focal recordings tend to have higher values in that feature. Note that there are plenty of exceptions in this work. For example, the acoustic diversity index does not have a statistically significant negative difference in means. In fact, it would seem that it may actually have a significantly positive difference. The same is true for species density. For the other features and regions, there are a handful of exceptions as well. PER and POW are major counterexamples as they tend to have features that fail to reject the null hypothesis. Interestingly, the significance of *NumSpecies* also flipped after filtering out no-bird clips, suggesting that species density is more similar when there are species present, and the difference before was due to a higher presence of "dead air" clips with no bird audio in passively-recorded datasets.

4.1.3. Discussion

In the SVM, Mann–Whitney U and KS Tests, we addressed RQ2, "Can Acoustic Indices meaningfully separate focal and soundscape recordings?" We find strong evidence that these features can be very useful as features. The SVM experiments successfully *classify* between focal and soundscape with accuracy greater than chance. Likewise, both the Mann–Whitney U and KS-Tests provide some evidence that the distributions of each acoustic index feature differ significantly between focal and soundscape recordings. These patterns are generally consistent across regions. In both cases, these features exhibit properties that an "ideal" metric for describing domain shift should have,

Table 7
Mann–Whitney U test for soundscapes of only labeled species vs. focal for each feature.

	Entropy	Complexity	ACI	ADI	BI
SSW stat	3 523 259.000000	3 462 820.000000	3 276 305.000000	1 695 750.000000	2 826 041.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000
HSN stat	2 255 428.000000	3 514 699.000000	3 788 111.000000	861 878.000000	2 661 256.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000
PER stat	1 108 183.000000	1 686 212.000000	3 647 376.000000	387 629.000000	1 654 727.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000
UHH stat	2 785 736.000000	3 545 477.000000	3 117 044.000000	1 351 242.000000	2 503 664.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000
SNE stat	3 098 092.000000	3 510 055.000000	3 873 911.000000	357 070.500000	3 398 814.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000
POW stat	1 932 675.000000	2 994 108.000000	3 052 066.000000	392 876.000000	1 698 061.000000
pvalue	0.065252	0.000000	0.000000	0.000000	0.000000
NES stat	2 973 923.000000	2 867 008.000000	3 478 761.000000	817 743.000000	1 666 410.000000
pvalue	0.000000	0.000000	0.000000	0.000000	0.000000

Table 8
Test statistic (difference in means) and p-values for each index over soundscapes and focal records from a permutation test. Significant results ($p < 0.001$) are highlighted in gray. Results are more mixed here as we start getting region specific variation in the domain shift. This can be an example for what can be done with these quantifiers and domain shift to further study the problem.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
HSN stat	−0.106090	−0.129710	−40.239461	0.182644	−153.745499	−0.480000
p-value	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000
PER stat	0.105977	0.033745	−25.415208	0.328571	64.588294	1.418000
p-value	1.000000	1.000000	0.000000	1.000000	1.000000	1.000000
UHH stat	−0.197293	−0.132805	−28.495483	0.014607	−187.564789	0.053500
p-value	0.000000	0.000000	0.000000	0.783000	0.000000	0.995000
SNE stat	−0.116062	−0.083304	−39.517153	0.372099	−874.527733	0.137000
p-value	0.000000	0.000000	0.000000	1.000000	0.000000	1.000000
POW stat	0.071987	−0.025604	−22.457386	0.414861	366.780375	1.846000
p-value	1.000000	0.000000	0.000000	1.000000	1.000000	1.000000
NES stat	−0.135649	−0.036706	−34.585739	0.266427	436.252051	−0.539000
p-value	0.000000	0.000000	0.000000	1.000000	1.000000	0.000000
SSW stat	−0.243566	−0.085780	−26.013406	0.002811	−432.283341	−0.562500
p-value	0.000000	0.000000	0.000000	0.575000	0.000000	0.000000

Table 9
Test statistic (difference in means) and p-values for each index over soundscapes and focal records from a permutation test. Filtering for birds only, with significant results ($p < 0.001$) highlighted in gray. Results are once again mixed. Note that the significance of *NumSpecies* for HSN, NES, and SSW has flipped.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
HSN stat	−0.112678	−0.127500	−38.305398	0.194526	−160.491097	0.165919
p-value	0.000000	0.000000	0.000000	1.000000	0.000000	1.000000
PER stat	0.104738	0.036046	−25.515280	0.327944	75.384138	1.703186
p-value	1.000000	1.000000	0.000000	1.000000	1.000000	1.000000
UHH stat	−0.175623	−0.118870	−27.227271	0.064665	−173.976900	0.483803
p-value	0.000000	0.000000	0.000000	1.000000	0.000000	1.000000
SNE stat	−0.105358	−0.082630	−39.008817	0.394995	−873.387007	0.644252
p-value	0.000000	0.000000	0.000000	1.000000	0.000000	1.000000
POW stat	0.073099	−0.025065	−22.094621	0.414634	366.918752	1.926478
p-value	1.000000	0.000000	0.000000	1.000000	1.000000	1.000000
NES stat	−0.135812	−0.039214	−32.486096	0.242825	441.858664	0.171537
p-value	0.000000	0.000000	0.000000	1.000000	1.000000	1.000000
SSW stat	−0.259432	−0.087335	−27.510816	0.023280	−426.315428	0.369327
p-value	0.000000	0.000000	0.000000	0.893000	0.000000	1.000000

demonstrating that each metric contains information on the differences between focal and soundscape audio.

A possible limitation of this work is that the nature of acoustic indices has previously been shown to distinguish acoustic differences between sites (Colonna et al., 2020; Budka et al., 2023; Pieretti et al., 2011; Bradfer-Lawrence et al., 2023). Our experiment does not currently control for differences between the sites where focal recordings have been taken and those where the soundscapes may have been recorded. By attempting to do this work using BirdSet and separating soundscapes from focal from sites around the world, ideally this work can better aligns to discussions of domain shift rather than site specific differences. As will be discussed in the discussion, future work should consider joint focal and soundscape recording deployments for better comparisons of domain shift.

To investigate RQ3, we tested for a significant negative difference in the mean acoustic index features between focal and soundscape recordings. Our results show that most features have negative differences, with the largest exception being the Acoustic Diversity Index overwhelmingly have positive differences regardless of data split. Two notable regions fail to reject the null: PER and POW. The test statistics for these regions are positive, indicating that their soundscapes had higher values in the feature than the focal recordings, contradicting our initial hypothesis. One possible hypothesis is that both POW and PER soundscape recordings were taken during the dusk-dawn chorus,

Table 10
Test statistic (difference in means) and p-values for each index over soundscapes and focal records from a permutation test. Filtering for soundscapes with a single species only, with significant results ($p < 0.001$) highlighted in gray. Results are mixed but strong.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
SSW stat	−0.255341	−0.086683	−28.110576	0.026188	−417.193980	0.000000
p-value	0.000000	0.000000	0.000000	0.955000	0.000000	0.000000
HSN stat	−0.093970	−0.121067	−38.693120	0.181919	−152.163349	0.000000
p-value	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000
PER stat	0.132551	0.023524	−26.294263	0.346644	14.968354	0.000000
p-value	1.000000	1.000000	0.000000	1.000000	0.896000	0.000000
UHH stat	−0.192650	−0.140472	−29.979621	0.010709	−224.527221	0.000000
p-value	0.000000	0.000000	0.000000	0.713000	0.000000	0.000000
SNE stat	−0.129820	−0.088408	−40.743803	0.356824	−884.996590	0.000000
p-value	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000
POW stat	0.020273	−0.032603	−22.833291	0.412329	379.420721	0.000000
p-value	1.000000	0.000000	0.000000	1.000000	1.000000	0.000000
NES stat	−0.130184	−0.035336	−32.009845	0.238115	432.560225	0.000000
p-value	0.000000	0.000000	0.000000	1.000000	1.000000	0.000000

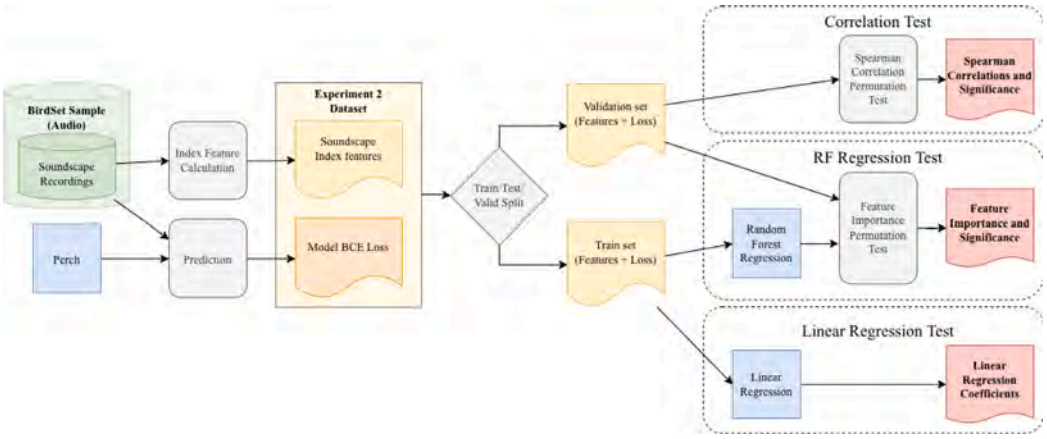

Fig. 5. Experiment workflow for Experiment 2. Here, BirdSet Sample contains 2000 sampled soundscape audio clips from a single region. Test split of the dataset was omitted as it was created but not used.

the time during sunset and sunrise. This may lead to having higher information complexity and bioacoustic index measurements as birds tend to vocalize more at those times of day, leading to overlapping sounds (Rauch et al., 2025a; Bustamante and Garitano-Zavala, 2024; Gil and Llusia, 2020).

Acoustic diversity Index being primarily positive also makes sense as larger ADI values implies greater amounts of noise (Budka et al., 2023). Typically ADI is used to measure the presence of species calls, yet this demonstrates how an acoustic index can also be used to compare distributions of audio from different recording styles.

What these results shows is that our method can identify domain shift quantifiers in different regions around the world to classify between focal and soundscape recordings then use those metrics to learn about the different properties behind domain shift. Future work could use this to identify a metric for a new region of the world and study how focal recordings may different with audio from soundscapes from said area. Yet one may wonder how the change in these metrics could impact a new model over these unlabeled soundscape recordings. This framework has not yet discusses model performance and how different properties of domain shift may impact models. This leads us into our next set of experiments and a key question: How do these features relate model performance?

4.2. Experiment 2: Using indices for predicting loss

4.2.1. Experiment design

It is well established in the literature that domain shift leads to poor model performance (Rauch et al., 2025a). From Experiment 1 (Section 4.1), we find that acoustic index metrics reliably encode features that distinguish focal from soundscape, quantifying the domain shift. Thus, we hypothesize that these metrics can reliably predict which recordings will result in higher model loss. We conduct experiments to this end, examining correlations, feature importance in random forest regression, and coefficients in a linear model to study the relationship between acoustic index features and model loss. These experiments are inspired by the work of Bidarouni and Abeßer, who also used linear regression to correlate measures of domain shift between deployment sites (Bidarouni and Abeßer, 2024). Fig. 5 shows the experimental workflow.

For these experiments, we sample 2000 random soundscape clips for each of the given regions. As a reminder about Section 3.4, this is done twice, once sampled from all soundscapes and then again for soundscapes with only 1 annotation. After processing the data for loss and each of the filters, we remove rows with any null values. The sample is split into a train and test split (80/20% split). The train is then further divided into a train/validation split (80/20%), yielding a 64/16/20% train/validation/test split. The purpose of the splits was to allow potential downstream hyperparameter tuning on the validation data as part of the experiments. However, none of the experiments we ran ended up not requiring hyperparameter tuning, so we did not use the test split in any experiment. Future work could therefore use our data to find improved models for studying soundscape recordings with using all 3 splits.

To estimate model loss, we compute the binary cross-entropy (BCE) loss on Perch. Binary cross-entropy (BCE) loss estimates uncertainty in a prediction, so even if the model makes the right prediction, an out-of-distribution data point may still have high loss. Perch is a state-of-the-art (SOTA) bioacoustic classification model created by Google,

primarily trained on xeno-canto recordings (Ghani et al., 2023). We chose Perch over BirdNET, another popular SOTA model for bird classification, due to concerns about data contamination. BirdNET has been trained on the soundscapes from BirdSet to help improve its domain shift performance (Rauch et al., 2025a), whereas Perch has been trained only on the focal recordings (Wood et al., 2024; Ghani et al., 2023). This experiment calculates BCE loss from Perch for each region. The loss is computed with respect to that region's species. Perch supports thousands of species, but we filter the logits for only the species of interest in each region, similar to experiments in the BirdSet paper (Rauch et al., 2025a).

The first subexperiment examines Spearman correlations between the features of a given soundscape recording and its associated BCE loss to identify the direction in which each metric relates to the loss. From there, we can compare those results with the permutation-in-mean-diffs experiment from experiment set 1 to infer how domain shift may impact model performance.

The second experiment examines the permutation importance of each feature for predicting model loss in random forest regressor models. These models are implemented in scikit-learn (sklearn) with default settings for all regions (Pedregosa et al., 2011b) and trained over the train split described above. Each feature is standardized, then used to predict model loss; afterward, we will assess each feature's importance. Importance measures the average change in mean squared error (MSE) after including a feature (Pedregosa et al., 2011b,a). Positive Importance scores indicate a large increase in MSE, indicating greater relevance for correct predictions. We perform a 1-sample 100-trial permutation test for if each feature is not important to the model (H_0 : $\text{importance}_{x_i} = 0$) or is very important for the model (H_a : $\text{importance}_{x_i} > 0$), with a level of significance $\alpha = 0.05$.

Finally, we use linear regression to help measure how effective each metric was at predicting model loss. These loss, entropy, and complexity tuples are collected and split into a test and train split. We train a simple linear regression model and a robust linear regression model using Huber Loss, using each of the standard normalized features to predict model loss. We evaluate the five linear regression assumptions for each dataset (Casson and Farmer, 2014). Namely, we will also report the condition number for multicollinearity, the Jarque–Bera test for Normality, the Durbin–Watson test for autocorrelation, the Goldfeld–Quandt test for Heteroskedasticity, and the Harvey–Collier test for Linearity. These functions come from the statsmodels library (Perktold et al., 2024). If these metrics indicate that we meet the assumptions of linear regression, we will then report each feature's coefficients and p-values as reported by the statsmodels library (Perktold et al., 2024).

There are two confounding factors in all of these experiments: The number of species vocalizing simultaneously and the lack of species vocalizing. Models have been observed to perform worse in regions with a greater number of concurrent bird vocalizations in a soundscape (Kahl et al., 2021a). It is possible that more vocalizations could increase the complexity or entropy of a data point, thereby changing model performance without necessarily altering the style in which the data is recorded (see Table B.24 in the Appendix for more information). Therefore, we add the number of labeled vocalizations in a 5-second clip as a confounding factor. Coincidentally, we observed that the loss of recordings with no labeled birds is a constant value of 0.693147 or close to $\ln(2)$ which is the theoretical loss of random guessing (note we average BCE Loss per class, thus in a case of a model randomly guessing for all classes, we expect a value of $\ln(2)$) (Bhagoj et al., 2021). Thus, we conduct two versions of each experiment: one with all the data and one with only birds, since we expect the regression models to predict $\ln(2)$ more frequently with non bird audio recordings. This is filtered from the same sample, thus our number of samples per region can differ in the bird-only example, as shown in Table 11.

Table 11
Number of samples in all regions (after null filtering), and number of samples in the bird only samples after filtering for recording with at least one label.

Region	Count of all samples	Count of bird only samples
PER	2000	1768
UHH	1985	1408
SNE	2000	1370
POW	2000	1927
NES	2000	775
HSN	1997	923
SSW	2000	691

4.2.2. Results

Correlations. Starting with the correlation experiments using all 2000 soundscape examples shown in Table 12, we find that the number of species is highly correlated with poor model performance in the all data model. We find very little agreement between regions on which features have statistically significant relationships with model performance. For example, PER is perhaps the only model that shows a statistically significant relationship in correlation across all features. Every index has at least two regions where it fails to reject the null, and it is not consistent which region it fails to reject the null (as opposed to the case in which we failed to reject the null regarding POW and PER in Experiment 1 across all the features).

Even when filtering the soundscapes for only labeled birds, as shown in Table 13, we see many of the same patterns of inconsistent results. These results indicate that the clearest feature for model performance is the number of species in a given soundscape sample.

Importance. Results for importance can be found for all sampled soundscape clips at Table 15, the subset with labeled birds at Table 16, and finally the secondary sample soundscapes with only 1 bird in Table 14. Again we see that "dead air" limits the results in Table 15. If we account for the presence of birds in the chunk, as is the case in Tables 14 and 15, the indices have higher relevant to predicting model loss in random forest trees. Consistently across all runs, bioacoustic index does not have strong importance in all runs and results are mixed which index has the most important feature other than the number of species in a chunk, even when only looking at a single species (see Table 17).

Linear regression. The Linear Regression experiments fail several of the assumptions. In particular, we fail the assumptions regarding normality for all regions (see the Jarque–Bera test results in Table 18). Since we fail some assumptions for linear regression, we are unlikely to be able to draw statistically strong conclusions from the results.

Barring the assumption checks above, Table 19 are the coefficients and p-values for each region and feature for all soundscape data.

To account for NumSpecies dominating the prediction due to what happens to the loss with no birds, here are the results from filtering for only birds in the 2000 sample as shown in the assumption checks in Table 20 and the coefficient results in Table 21

Finally we check the second sample of only 1 bird label soundscape clips as shown in Tables 23 and 22.

4.2.3. Discussion

In comparison to Experiment 1 (Section 4.1), these results are far more varied, as some of the techniques are better at predicting loss in some regions and not others, particularly when using the number of species as a feature. Even then, the correlations to loss are weak, with many of the statistically significant correlations having a small magnitude. We have not met the assumptions to draw conclusions from linear regression and that, with the correlation results, makes interpreting the results from evaluating model importance difficult. What the result do show is that the number of species in a given audio sample has a drastic impact on model loss with strong correlation and model importance,

Table 12
Spearman's correlations between each index and model loss over all soundscape data. Significant results ($p < 0.05$) are highlighted in gray. Significance is mixed for these experiments, as is the direction of the correlation for each region.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
PER corr	−0.146930	0.269118	0.126109	−0.120434	0.248311	0.913601
p-value	0.008479	0.000001	0.024065	0.031255	0.000007	0.000000
UHH corr	0.077050	0.187153	0.042857	0.087337	0.159244	0.769939
p-value	0.170490	0.000797	0.446309	0.120120	0.004417	0.000000
SNE corr	−0.015086	−0.048858	0.102993	0.060257	0.096891	0.798210
p-value	0.788064	0.383701	0.065755	0.282523	0.083536	0.000000
POW corr	0.192535	0.073689	0.099159	−0.030178	0.184297	0.891859
p-value	0.000534	0.188569	0.076521	0.590684	0.000925	0.000000
NES corr	0.082787	0.074317	0.066208	0.046729	0.131955	0.700704
p-value	0.139494	0.184823	0.237594	0.404794	0.018197	0.000000
HSN corr	−0.174508	−0.079409	0.074390	−0.106627	0.086667	0.769493
p-value	0.001727	0.156425	0.184391	0.056731	0.121819	0.000000
SSW corr	−0.091155	0.004171	−0.083230	−0.015593	0.052011	0.840803
p-value	0.103603	0.940755	0.137381	0.781114	0.353728	0.000000

Table 13
Spearman's correlations between each index and model loss over all soundscape data with filtering for bird only. Significant results ($p < 0.05$) are highlighted in gray. Like Table 12, significance is mixed.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
PER corr	−0.162219	0.316713	0.178513	−0.079298	0.244986	0.896265
p-value	0.005879	0.000000	0.002402	0.180361	0.000027	0.000000
UHH corr	−0.093218	0.059498	0.047742	0.051759	0.270184	0.733232
p-value	0.161584	0.372249	0.474151	0.437726	0.000037	0.000000
SNE corr	−0.134446	−0.142343	0.037130	0.018276	0.069197	0.648046
p-value	0.045887	0.034441	0.582982	0.787027	0.305795	0.000000
POW corr	0.202375	−0.100966	0.044245	0.015955	0.116105	0.847339
p-value	0.000351	0.076851	0.439096	0.780336	0.041727	0.000000
NES corr	0.100082	0.049916	0.044151	0.245259	0.094897	0.432232
p-value	0.272719	0.585064	0.629184	0.006473	0.298466	0.000001
HSN corr	−0.431070	−0.367105	−0.089268	−0.155548	0.197821	0.440884
p-value	0.000000	0.000009	0.296003	0.067480	0.019578	0.000000
SSW corr	−0.109818	−0.127135	0.155133	−0.038947	−0.168251	0.673785
p-value	0.274293	0.205177	0.121367	0.698990	0.092595	0.000000

Table 14
Spearman's correlations between each index and model loss over soundscape data with 1 annotation only. Significant results ($p < 0.05$) are highlighted in gray.

	Entropy	Complexity	ACI	ADI	BI
SSW corr	−0.097556	−0.015436	−0.043744	−0.000500	−0.004934
pvalue	0.081429	0.783264	0.435497	0.992889	0.929940
HSN corr	−0.423171	−0.297254	−0.148140	−0.227512	0.174335
pvalue	0.000000	0.000000	0.007947	0.000040	0.001746
PER corr	0.012429	0.001560	0.144055	0.058204	0.051994
pvalue	0.824720	0.977828	0.009871	0.299274	0.353883
UHH corr	−0.101657	−0.072261	−0.030536	−0.043967	0.090263
pvalue	0.070240	0.198719	0.587468	0.434601	0.108148
SNE corr	−0.117728	−0.112127	−0.096623	−0.041789	−0.024497
pvalue	0.035285	0.045042	0.084397	0.456310	0.662424
POW corr	0.026968	−0.483739	0.072961	0.398873	−0.154015
pvalue	0.630791	0.000000	0.192986	0.000000	0.005766
NES corr	0.038412	0.029460	−0.018085	0.038436	−0.016764
pvalue	0.493540	0.599555	0.747244	0.493262	0.765139

which we hypothesized given previous research (Kahl et al., 2021a). Our initial thinking was by including this into the model, this could account for the impact of number of species allowing domain shift to be more accounted for by the indices. However, it seemed that the number of species ended up dominating the regression.

During our secondary runs, we factor this in better by only sampling from cases of a single species being present, but it does not significantly impact the results. The assumption checks still fail but are better met, model importance weighs the indices more (as number of species no longer is a useful predictor). Correlations are however, less clear and for most regions lack any statistically significant correlations. Only HSN and POW have any useful metrics. Untimely from these two experiment runs, it appears that the quantifiers identified in experiment 1 have inconsistent relationships to model loss if at all.

In many ways the results being inconsistent across regions of the world can make sense if we assume based on Fig. 1 that the distribution of these metrics differ greatly depending on the soundscape. Rather than the focal recordings, a given soundscape differs greatly than another. We attempted to use Perch for these experiments, a model trained on focal recordings from around the globe. If domain shift is different per region of the world, then it follows the change to loss in relation to a measure of domain shift will be different per region.

The fact we see weak evidence of loss being impacted by these domain shift quantifiers lends itself to a new hypothesis: Perch may be more resilient to domain shift than we expected. After all, the only thing the model seemed to consistently struggle with was not metrics that separate domains, but rather the number of species in a given clip, implying that the model may be able to adapt to the different regions of the world better than expected. It did, after all, recently achieve SOTA

Table 15
Importance of each feature as measured by importance permutation for random forest regressor models over all soundscapes. Significant results ($p < 0.05$) are highlighted in gray. Results are dominated by NumSpecies. Note that negative importance scores are from random noise, and may be interpreted as 0.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
PER mean	0.009067	0.023564	0.008130	0.004902	0.004974	1.574628
PER p-value	0.100000	0.000000	0.070000	0.170000	0.060000	0.000000
UHH mean	0.045546	0.018358	0.056589	0.013371	0.064686	1.195869
UHH p-value	0.000000	0.040000	0.000000	0.130000	0.000000	0.000000
SNE mean	0.178855	0.060663	0.095593	0.055517	0.030454	1.311060
SNE p-value	0.000000	0.010000	0.010000	0.040000	0.010000	0.000000
POW mean	0.026333	0.049115	0.018216	0.035618	0.030013	1.639225
POW p-value	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
NES mean	−0.000265	−0.009782	0.032597	0.162014	0.067828	0.523934
NES p-value	0.460000	0.670000	0.320000	0.000000	0.080000	0.000000
HSN mean	0.313562	−0.010150	0.047587	0.005425	−0.007760	0.414179
HSN p-value	0.000000	0.580000	0.110000	0.420000	0.590000	0.000000
SSW mean	0.031046	−0.016818	−0.044544	−0.018968	0.054085	1.016321
SSW p-value	0.230000	0.740000	0.910000	0.690000	0.100000	0.000000

Table 16
Importance of each feature as measured by importance permutation for random forest regressor models over all soundscapes. Significant results ($p < 0.05$) are highlighted in gray. Note that, compared to Table 15, the other features become much more important now that NumSpecies is more varied. EGCI, in particular, is more important in many regions than the other features. Note the mean is the average over the distribution of possible importance via permutation trial, so values with significant p-values have a major of their MSE increase above 0.

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
PER mean	0.011600	0.016938	0.008238	0.010533	0.010460	1.677344
PER p-value	0.000000	0.000000	0.000000	0.040000	0.000000	0.000000
UHH mean	0.015550	0.019200	0.070033	0.019936	0.058992	1.521206
UHH p-value	0.040000	0.090000	0.000000	0.030000	0.010000	0.000000
SNE mean	0.087186	0.032563	0.054693	0.043380	−0.003192	1.441193
SNE p-value	0.000000	0.020000	0.020000	0.020000	0.690000	0.000000
POW mean	0.019772	0.041393	0.004326	0.029229	0.021511	1.886577
POW p-value	0.000000	0.000000	0.090000	0.000000	0.000000	0.000000
NES mean	0.008591	0.030098	0.091060	0.068946	0.027138	1.313475
NES p-value	0.210000	0.020000	0.000000	0.000000	0.100000	0.000000
HSN mean	0.192378	0.062136	0.079676	0.002860	−0.012736	1.437171
HSN p-value	0.000000	0.000000	0.000000	0.420000	0.730000	0.000000
SSW mean	−0.012975	−0.010575	−0.056376	−0.004187	−0.035312	1.325118
SSW p-value	0.770000	0.800000	1.000000	0.600000	0.910000	0.000000

Table 17
Importance of each feature with the one annotation soundscape run.

	Entropy	Complexity	ACI	ADI	BI
SSW mean	0.128658	0.011975	0.080662	0.076547	0.060563
SSW p-value	0.000000	0.230000	0.000000	0.000000	0.030000
HSN mean	0.515735	0.243675	0.238571	0.126380	0.063726
HSN p-value	0.000000	0.000000	0.000000	0.000000	0.000000
PER mean	0.283323	0.229406	0.145114	0.431127	0.081671
PER p-value	0.000000	0.000000	0.000000	0.000000	0.000000
UHH mean	0.020474	0.017151	0.262665	0.054517	0.107054
UHH p-value	0.060000	0.310000	0.000000	0.000000	0.000000
SNE mean	0.227538	0.090867	0.183407	0.206336	0.080270
SNE p-value	0.000000	0.000000	0.000000	0.000000	0.000000
POW mean	0.233137	0.541327	0.167056	0.587305	0.342548
POW p-value	0.000000	0.000000	0.000000	0.000000	0.000000
NES mean	0.129252	0.298772	0.184891	0.419490	0.128498
NES p-value	0.000000	0.000000	0.000000	0.000000	0.000000

performance over BirdSet thus making it by being trained primarily with focal recording theoretically good at domain shift (van Merriënboer et al., 2024b). Future work may consider using the experiment framework created here to test a range of models and the impact these metrics have on the loss. Models have been improving over the years and thus resiliency to domain shift maybe be measured over time with this approach.

5. Discussion: Linking Experiment 1 and 2

The goal of this work was two fold: identify domain shift quantifiers without species-classification models and species-labels and demonstrate the impact those domain shift quantifiers may have on model performance. Experiment 1 helped to identity some metrics that can describe the relationship between focal recordings and the soundscape recordings of a given region. This allowed us to create Experiment 2 to test how these metrics that seem to have relationships to the distribution shifts between focal and soundscapes recordings may impact a model's loss over soundscapes.

We believe that this framework allows for greater exploration of domain shift issues in bioacoustics to better define what domain shift is and explore its impact on models. For example, a good next step of this work would be to take a focal dataset and introduce a method that can dramatically change the distribution, like introducing MixUp augmentations with some random collection of soundscape recordings which is known to have impacts on domain shift in bioacoustics (Zhang et al., 2018; Jordal, 2025), and observing how that impacts the distribution of focal versus augmented focal versus the original soundscape using experiment 1. Then test the impact domain shift quantifiers have in experiment 2 with a model trained on the focal recordings and augmented focal recordings testing against the same soundscape recordings. This kind of experimentation would not only demonstrate what techniques could improve soundscape performance but also why it works in relation to domain shift.

Table 18
Diagnostic information for the OLS predicting loss of Perch for each region, including Kurtosis (kurt), Jarque–Bera (JB), Durbin–Watson (DW), Condition Number (CN), Goldfeld–Quandt test (GQ) and its p-value (GQ-P), Harvey–Collier test (HC) and its p-value (HC-P), and R^2 as conducted in Experiment 2. This is over all soundscape examples.

	PER	UHH	SNE	POW	NES	HSN	SSW
Omnibus:	11.580	375.776	553.204	26.616	188.909	887.535	667.043
Prob(Omnibus):	0.003	0.000	0.000	0.000	0.000	0.000	0.000
Skew:	0.025	1.065	1.640	0.160	0.527	3.039	1.665
Kurtosis:	3.565	11.299	14.189	3.852	7.487	18.507	27.834
Durbin–Watson:	1.904	1.998	2.066	2.080	1.996	2.084	1.997
Jarque–Bera (JB):	17.133	3878.632	7250.321	44.226	1131.218	14 761.122	33 483.418
Prob(JB):	0.000190	0.00	0.00	2.49e−10	2.29e−246	0.00	0.00
Cond. No.	2.48	3.79	5.26	3.06	4.79	2.83	2.83
R-Sqaure	0.858000	0.693000	0.648000	0.787000	0.558000	0.552000	0.641000
HC	0.649074	−2.228864	−1.123191	0.647772	−1.274923	0.952575	1.139342
Prob(HC)	0.516407	0.025999	0.261568	0.517249	0.202570	0.340987	0.254775
QG	1.063774	0.808002	0.793019	1.057952	0.840379	1.242645	0.899016
Prob(QG)	0.436951	0.007686	0.003580	0.478725	0.028986	0.006415	0.180792

Table 19
Coefficients of OLS model trained on soundscape train dataset. Significant coefficients ($p < 0.05$) are highlighted in gray.

Index	Const	Entropy	Complexity	ACI	ADI	BI	NumSpecies
PER coef	0.8373	−0.0047	0.0015	0.0014	−0.0002	−0.0004	0.0943
p-value	0.000	0.001	0.286	0.231	0.878	0.759	0.000
UHH coef	1.0031	−0.0025	−0.0219	−0.0023	0.0157	0.0249	0.2867
p-value	0.000	0.794	0.026	0.709	0.039	0.000	0.000
SNE coef	0.8259	−0.0221	−0.0046	−0.0012	0.0157	−0.0021	0.1234
p-value	0.000	0.000	0.463	0.672	0.000	0.462	0.000
POW coef	0.9726	−0.0122	−0.0307	−0.0046	0.0086	−0.0179	0.1873
p-value	0.000	0.000	0.000	0.264	0.002	0.000	0.000
NES coef	0.7169	0.0036	−0.0007	−0.0051	0.0021	−0.0023	0.0366
p-value	0.000	0.079	0.711	0.000	0.039	0.039	0.000
HSN coef	0.9546	−0.0890	−0.0166	−0.0397	0.0410	0.0200	0.3359
p-value	0.000	0.000	0.182	0.000	0.001	0.032	0.000
SSW coef	0.7223	−0.0027	−0.0026	0.0015	0.0019	−0.0015	0.0503
p-value	0.000	0.088	0.057	0.268	0.165	0.209	0.000

Table 20
Checking the assumptions of bird only soundscape regression OLS models. Again note the relatively high JB scores.

	PER	UHH	SNE	POW	NES	HSN	SSW
Omnibus:	7.133	172.178	304.759	31.286	12.254	187.604	92.505
Prob(Omnibus):	0.028	0.000	0.000	0.000	0.002	0.000	0.000
Skew:	−0.039	0.822	1.401	0.245	0.345	1.650	0.885
Kurtosis:	3.434	7.121	9.955	3.816	3.394	6.617	7.950
Durbin–Watson:	2.033	2.046	1.988	1.969	2.022	2.035	2.002
Jarque–Bera (JB):	9.279	742.203	2066.389	46.500	12.769	552.455	460.670
Prob(JB):	0.00966	6.80e−162	0.00	7.99e−11	0.00169	1.09e−120	9.27e−101
Cond. No.	2.60	3.55	4.69	3.00	5.06	3.25	3.51
R-Sqaure	0.811000	0.567000	0.480000	0.787000	0.310000	0.334000	0.349000
HC	1.589686	1.959079	0.214741	0.471191	0.806430	−0.928780	−0.133402
Prob(HC)	0.112183	0.050413	0.830019	0.637589	0.420396	0.353414	0.893944
QG	1.055157	1.026354	0.895598	0.829729	0.932568	0.917278	0.943205
Prob(QG)	0.523257	0.783829	0.251138	0.021418	0.592218	0.478908	0.685035

Table 21
OLS model trained on only bird examples and their coefficients. Significant coefficients ($p < 0.05$) are highlighted in gray.

Index	Const	Entropy	Complexity	ACI	ADI	BI	NumSpecies
PER coef	0.8504	−0.0044	0.0027	−0.0009	−0.0016	−0.0013	0.0851
p-value	0.000	0.010	0.092	0.516	0.273	0.344	0.000
UHH coef	1.1380	0.0125	−0.0377	−0.0126	0.0176	0.0355	0.2540
p-value	0.000	0.336	0.003	0.136	0.087	0.000	0.000
SNE coef	0.8966	−0.0294	−0.0097	0.0062	0.0185	−0.0028	0.1069
p-value	0.000	0.001	0.266	0.169	0.000	0.518	0.000
POW coef	0.9776	−0.0091	−0.0318	−0.0101	0.0040	−0.0205	0.1829
p-value	0.000	0.003	0.000	0.012	0.165	0.000	0.000
NES coef	0.7558	0.0151	−0.0082	−0.0178	0.0054	−0.0122	0.0324
p-value	0.000	0.007	0.103	0.000	0.041	0.000	0.000
HSN coef	1.3221	−0.2104	0.0129	−0.0666	0.0406	0.0308	0.2421
p-value	0.000	0.000	0.668	0.001	0.103	0.118	0.000
SSW coef	0.7888	−0.0053	−0.0104	−0.0031	0.0061	−0.0015	0.0493
p-value	0.000	0.389	0.071	0.449	0.160	0.691	0.000

Table 22
OLS model checks using only 1 annotated examples.

	SSW	HSN	PER	UHH	SNE	POW	NES
Omnibus:	145.503	305.237	74.610	106.716	209.303	13.255	13.541
Prob(Omnibus):	0.000	0.000	0.000	0.000	0.000	0.001	0.001
Skew:	0.608	1.359	−0.391	0.468	0.916	0.223	0.228
Kurtosis:	5.423	5.157	2.382	5.052	5.359	3.236	2.790
Durbin–Watson:	1.907	1.994	2.056	1.963	1.981	2.006	1.967
Jarque–Bera (JB):	392.061	640.173	53.022	269.544	475.718	13.598	13.394
Prob(JB):	7.33e−86	9.73e−140	3.06e−12	2.95e−59	5.00e−104	0.00111	0.00123
Cond. No.	3.63	3.15	2.44	3.41	4.54	3.23	4.90
R-Sqaure	0.016000	0.185000	0.050000	0.044000	0.060000	0.289000	0.045000
HC	1.926355	−0.835784	−100	−0.758284	1.750654	0.436739	0.943862
Prob(HC)	0.054282	0.403434	−100	0.448422	0.080246	0.662375	0.345419
QG	1.078506	1.146652	1.001338	1.100121	1.210711	0.975925	1.148250
Prob(QG)	0.341605	0.085654	0.986578	0.231580	0.016196	0.759089	0.082034

Table 23
OLS model trained on only 1 annotation soundscape examples and their coefficients. Significant coefficients ($p < 0.05$) are highlighted in gray.

Index	Const	Entropy	Complexity	ACI	ADI	BI
SSW coef	0.7576	−0.0037	−0.0013	0.0004	0.0061	−0.0052
p-value	0.000	0.223	0.620	0.851	0.004	0.004
HSN coef	1.2088	−0.1621	−0.0194	−0.0157	0.0206	0.0084
p-value	0.000	0.000	0.214	0.123	0.129	0.439
PER coef	0.7481	−0.0080	−0.0015	0.0047	0.0061	−0.0005
p-value	0.000	0.000	0.142	0.000	0.000	0.630
UHH coef	0.9509	0.0088	−0.0392	0.0061	0.0131	0.0354
p-value	0.000	0.332	0.000	0.319	0.093	0.000
SNE coef	0.8317	−0.0067	−0.0195	−0.0140	0.0247	−0.0011
p-value	0.000	0.298	0.003	0.000	0.000	0.746
POW coef	0.7223	−0.0076	−0.0314	−0.0207	0.0186	−0.0192
p-value	0.000	0.001	0.000	0.000	0.000	0.000
NES coef	0.7442	0.0017	0.0028	−0.0112	0.0064	−0.0088
p-value	0.000	0.639	0.371	0.000	0.000	0.000

6. Conclusion

We present experiments that demonstrate how acoustic indices can be used to study domain shifts in bioacoustic data. We showed how acoustic indices can easily identify differences between focal and soundscape recordings without more complex methods, and how regression can be used to learn how these indices describe how current models interact with soundscape recordings. Notably, we found evidence that all regions suffer from species density with soundscape recordings and suggested that perch may have more resiliency to domain shift than we hypothesized.

A key limitation with this work and a matter for future work, however, is whether these differences are site-dependent. Acoustic indices have remarkable success at identifying differences between sites (Bradfer-Lawrence et al., 2019; Colonna et al., 2020). While we attempted to show that soundscape and focal recordings can be separated across regions, one may still argue that these methods do not show differences in recording style but rather site information. What is really needed in this study of domain shift is explicit experiments testing different recording styles at the same site to see how they affect which data microphones (handheld, passively recorded, etc.) collect. That would be a far stronger dataset than BirdSet for understanding domain shift, by directly testing the recording-style hypothesis.

Future work may also consider expanding this kind of analysis beyond bioacoustics. Unlike other acoustic indices and despite being “Ecoacoustic” in name, EGCI metrics make no assumptions about directly measuring bioacoustic data. The only assumption made to compute the EGCI features is that one has audio recordings. This means that EGCI could be used to measure the domain shift of other acoustic-related tasks. For example, EGCI may be able to quantify shifts in noise pollution across different areas of a city, as well as variations in tone of voice, etc. In such cases, this paper may offer additional ideas on how to employ EGCI as a diagnostic tool and an exploratory data analysis method for any audio dataset.

Domain shift in bioacoustics does still remains a significant challenge for applying machine learning to classify large amounts of audio data. The limited amount of data available for bioacoustics limits use of semi-supervised techniques (van Merriënboer et al., 2024b) making it important to get as much out of the limited focal recording datasets as possible for model performance. With greater study of the factors behind domain shift, more work can be do to understand which techniques may help improve bioacoustic’s domains shifts between recording styles around the world.

CRediT authorship contribution statement

Sean Perry: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tianqi Zhang:** Writing – review & editing, Methodology, Conceptualization. **Siya Kamboj:** Writing – review & editing, Writing – original draft, Software. **Anu Jajodia:** Writing – review & editing, Writing – original draft, Software. **Dhruv Tomar:** Writing – review & editing. **Ryan Kastner:** Writing – review & editing, Supervision, Resources, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

We the authors used Grammarly for spell check and grammar as well as a Gemini Thinking Model to do a final review of our paper for spelling, logical fallacies, consistency, and other small improvements to the writing. Claude Sonnet 4.6 was used for creating tools to automate

formatting tables from a latex format outputted by pandas dataframes and also for formatting latex. Authors reviewed all suggestions and made edits manually and therefore take full responsibility for the paper.

Funding sources

This work was supported by the National Science Foundation Research Experience for Undergraduates Award 2244123, CloudBank fund "REU Site: Engineers for Exploration" (Award NAIRR240173). This material is based upon work supported by the Google Cloud Platform Award.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sean Perry reports financial support and administrative support were provided by National Science Foundation. Tianqi Zhang reports financial support was provided by National Science Foundation. Siya Kamboj reports financial support was provided by National Science Foundation. Anu Jajodia reports financial support was provided by National Science Foundation. Dhruv Tomar reports financial support was provided by National Science Foundation. Ryan Kastner reports equipment, drugs, or supplies was provided by Google LLC. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. All region SVM results

See Figs. A.6–A.16.

A.1. One bird species SVM accuracies

See Figs. A.17 and A.18.

Appendix B. Species density vs. indicies

See Table B.24.

Appendix C. Huber regression results

A secondary check was done with Huber Regression to see the effect of L2 normalization on the regression. The adjusted results can be found here. They are in the appendix because this work was done without assumption checking and thus should be considered with extreme caution (see Tables C.26 and C.27).

Appendix D. Focal vs. soundscape visualizations split by region

See Figs. D.19–D.25.

Data availability

The processed data (computed EGCI metrics) and code used to conduct this work are hosted at github.com/UCSD-E4E/egci_bioacoustic_shifts. The original dataset used for this work is the bioacoustic benchmark BirdSet is published under (Rauch et al., 2025a).

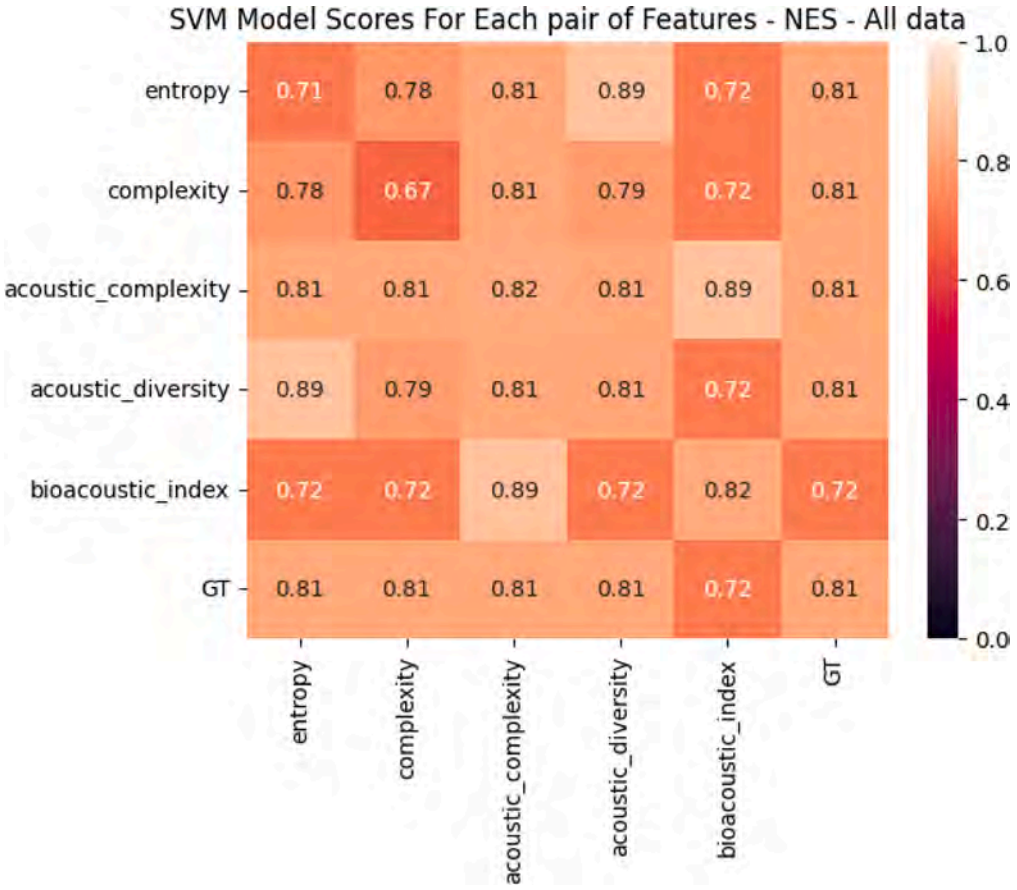

Fig. A.6. SVM accuracy scores for NES. All p-values were 0.

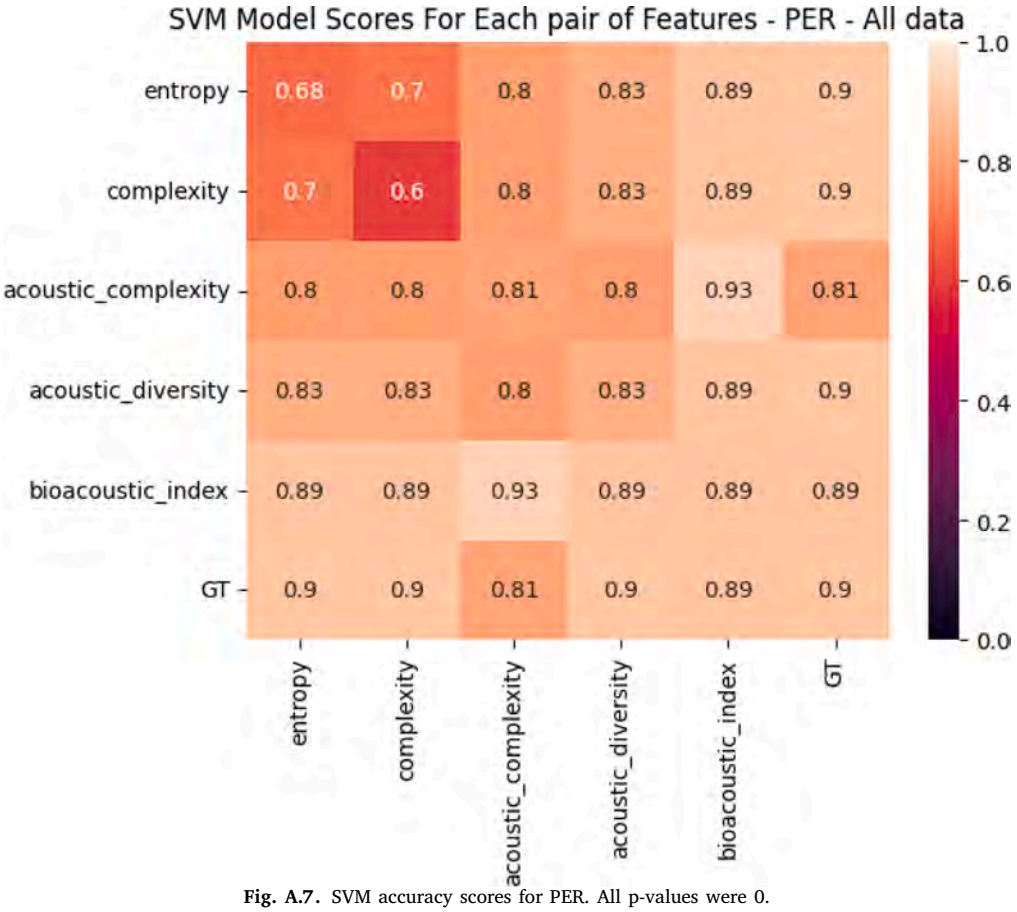

Fig. A.7. SVM accuracy scores for PER. All p-values were 0.

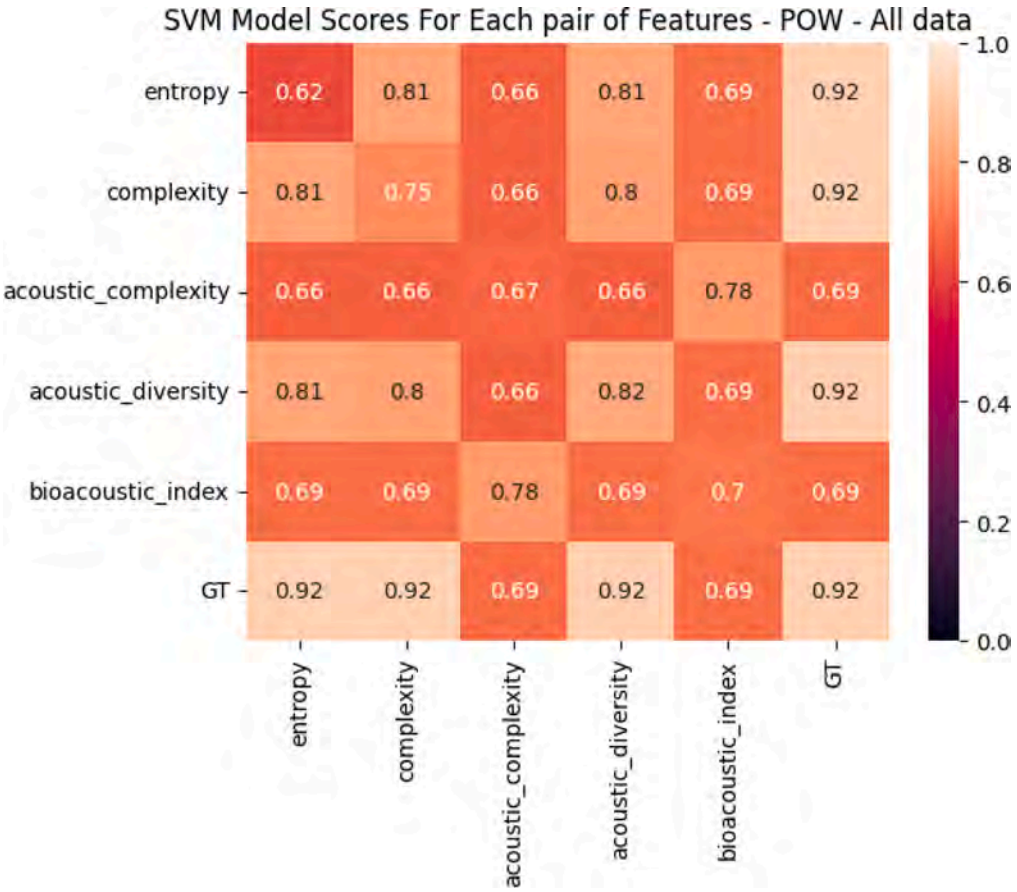

Fig. A.8. SVM accuracy scores for POW. All p-values were 0.

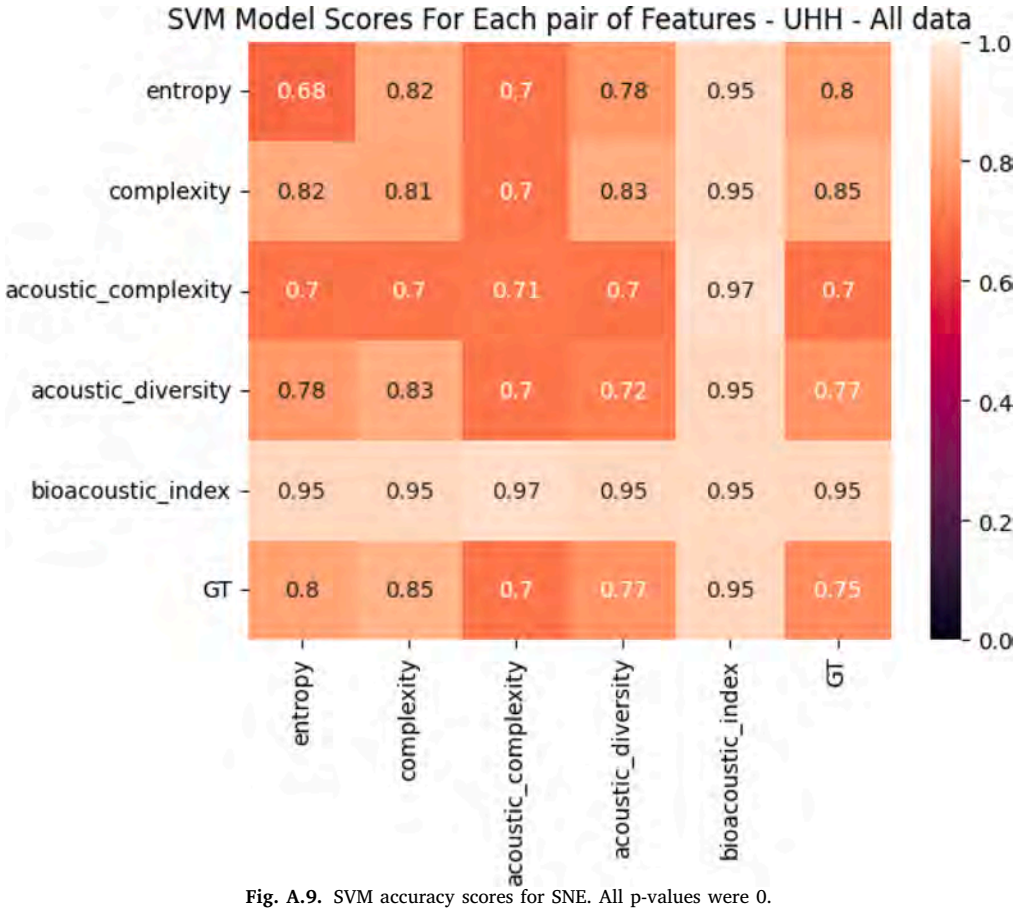

Fig. A.9. SVM accuracy scores for SNE. All p-values were 0.

Fig. A.10. SVM accuracy scores for SSW. All p-values were 0.

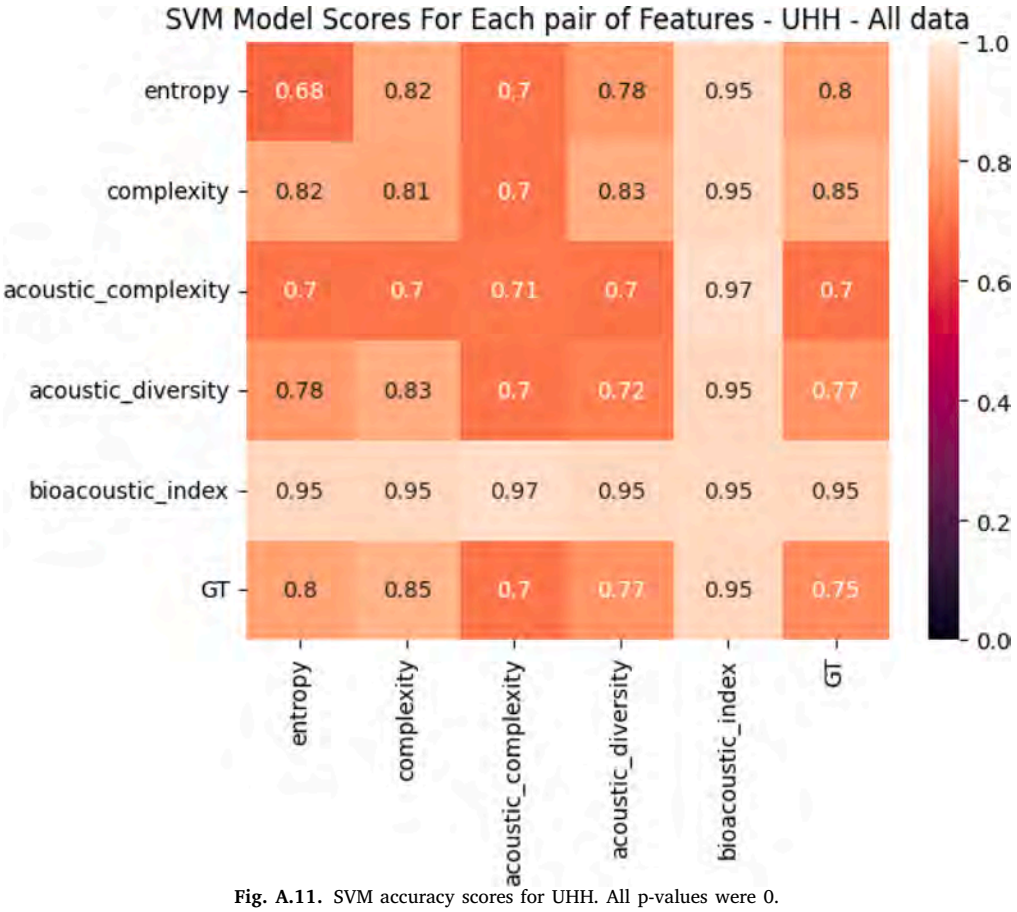

Fig. A.11. SVM accuracy scores for UHH. All p-values were 0.

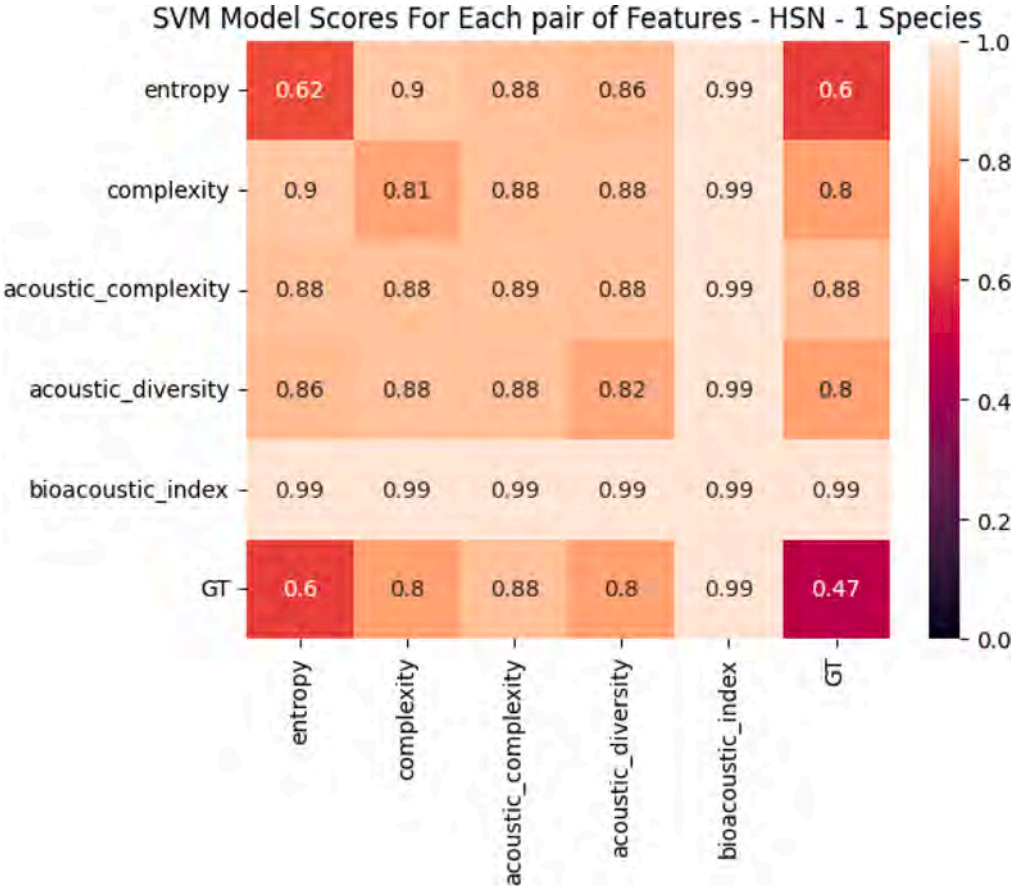

Fig. A.12. SVM accuracy scores for HSN for only 1 bird species. All p-values were 0.

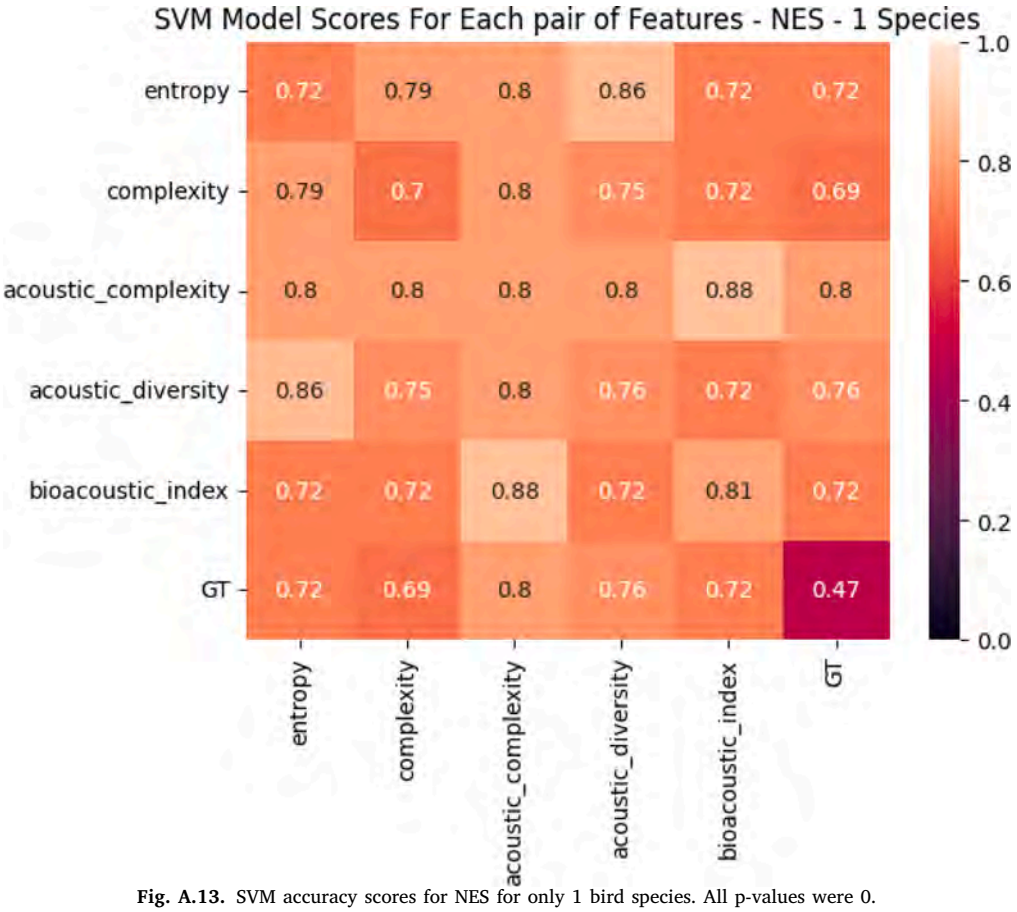

Fig. A.13. SVM accuracy scores for NES for only 1 bird species. All p-values were 0.

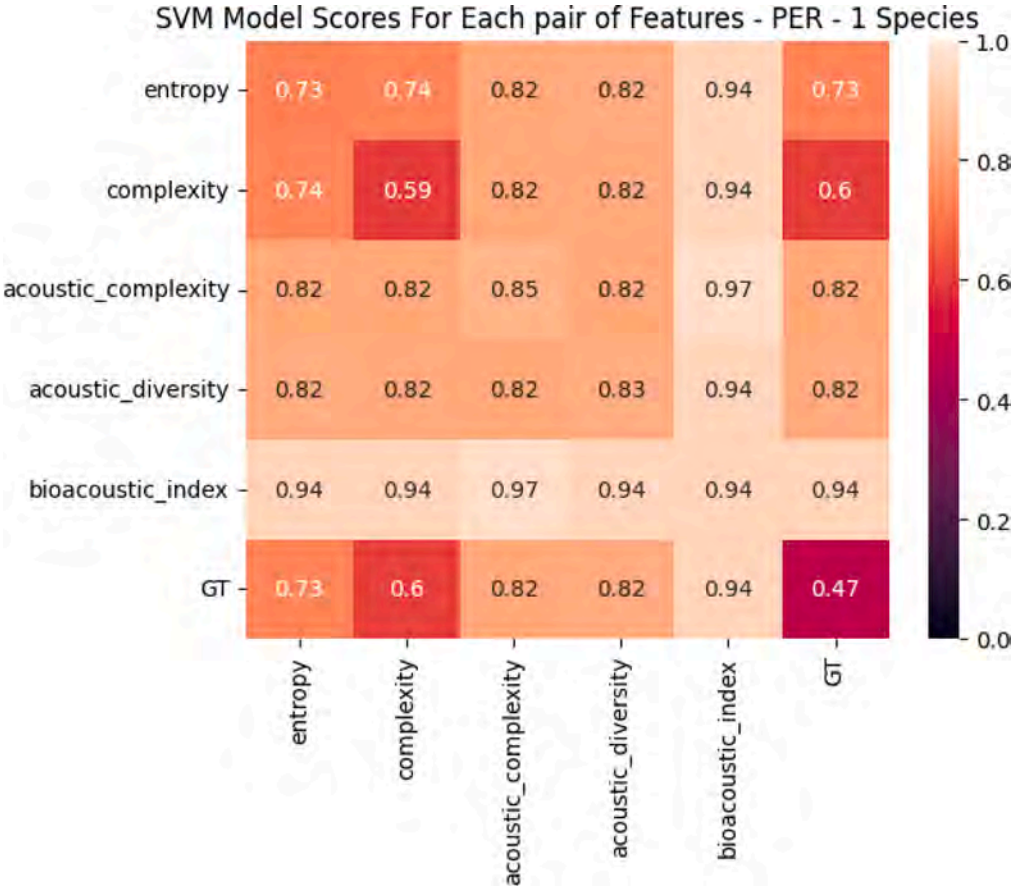

Fig. A.14. SVM accuracy scores for PER for only 1 bird species. All p-values were 0.

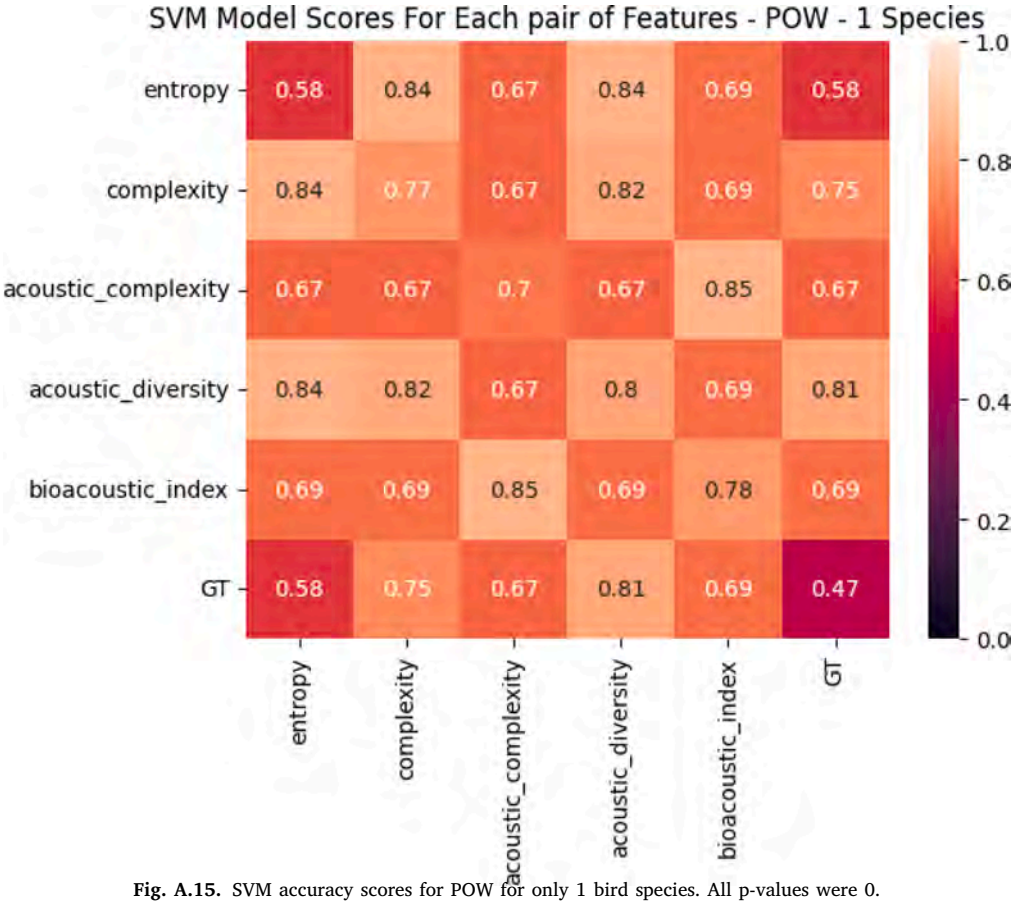

Fig. A.15. SVM accuracy scores for POW for only 1 bird species. All p-values were 0.

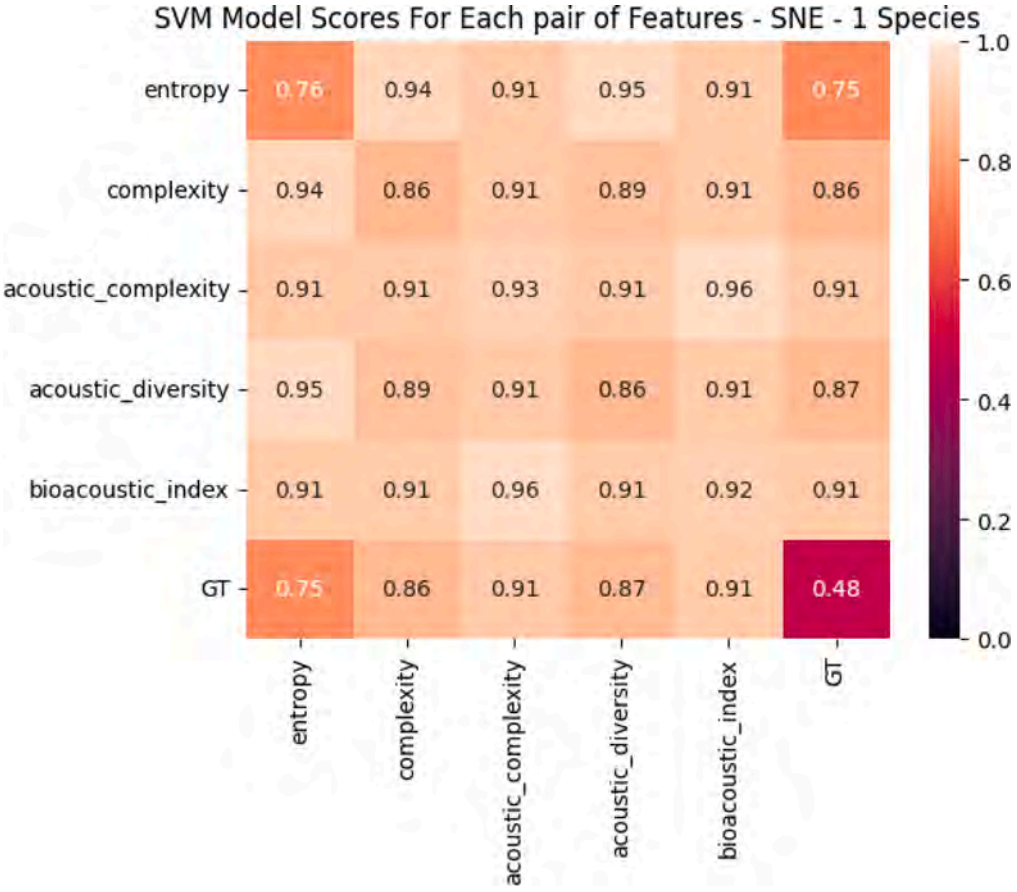

Fig. A.16. SVM accuracy scores for SNE for only 1 bird species. All p-values were 0.

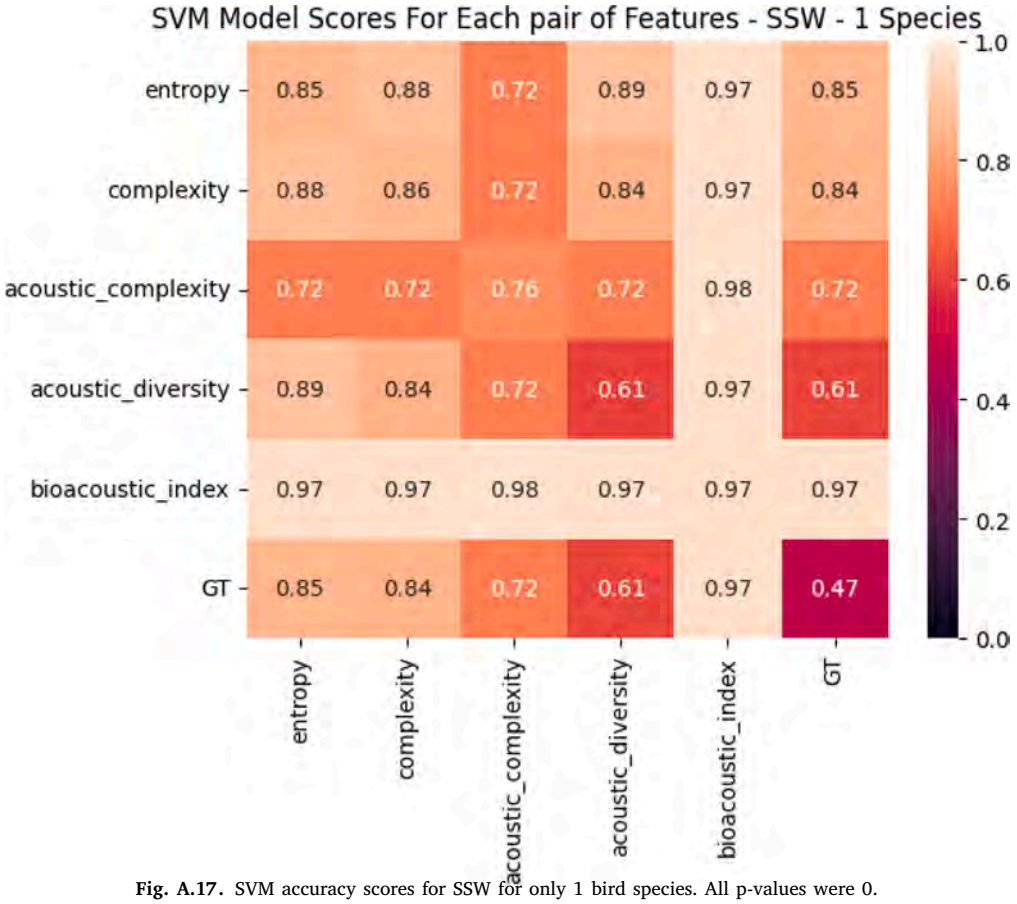

Fig. A.17. SVM accuracy scores for SSW for only 1 bird species. All p-values were 0.

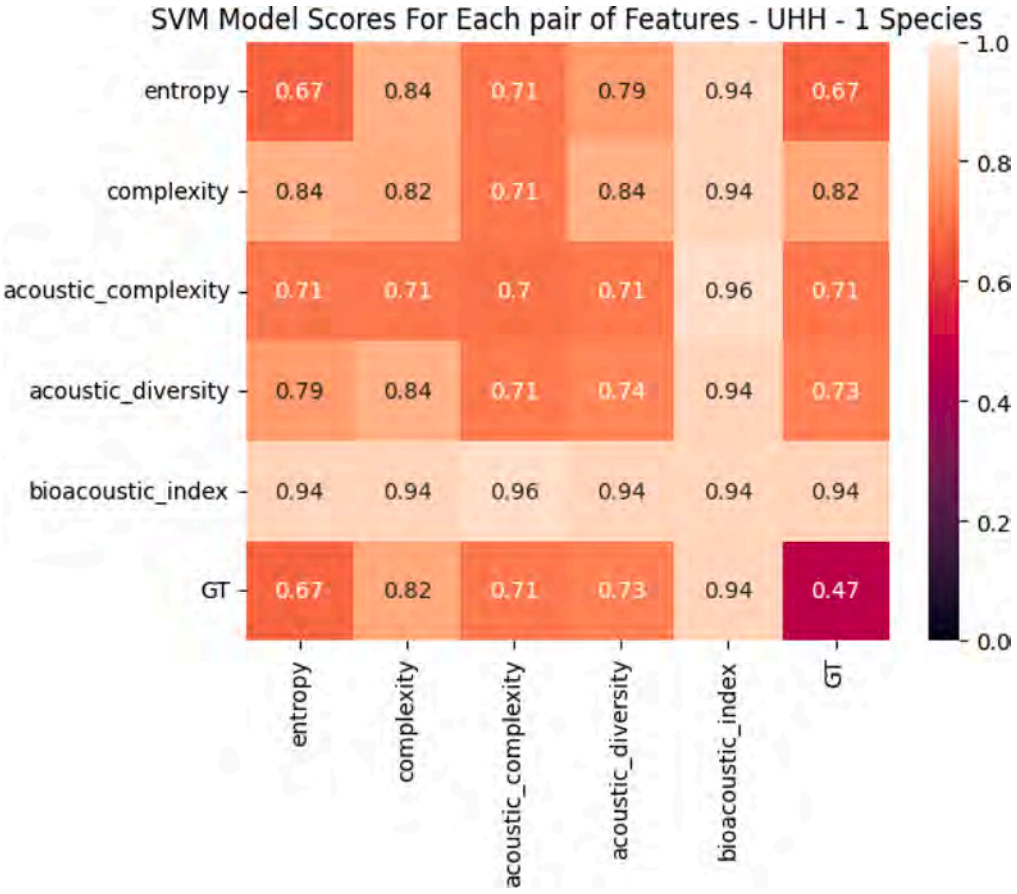

Fig. A.18. SVM accuracy scores for UHH for only 1 bird species. All p-values were 0.

Table B.24
Correlations between the indices and "NumSpecies" to verify hypothesis regarding how it may influence results based on soundscape recordings. Correlations appear mostly positive, however no region nor feature has a particularly strong nor significant correlation with species density according to Cohen (1988).

	Entropy	Complexity	ACI	ADI	BI	NumSpecies
PER corr	−0.126807	0.282577	0.126337	−0.082122	0.284148	1.000000
p-value	0.023288	0.000000	0.023808	0.142710	0.000000	0.000000
UHH corr	0.167971	0.366352	0.096017	0.156766	0.131999	1.000000
p-value	0.002656	0.000000	0.087368	0.005081	0.018526	0.000000
SNE corr	0.127313	0.030041	0.134398	0.077426	0.142152	1.000000
p-value	0.022738	0.592371	0.016142	0.167064	0.010900	0.000000
POW corr	0.238041	0.229768	0.112904	−0.102900	0.274918	1.000000
p-value	0.000017	0.000033	0.043565	0.066003	0.000001	0.000000
NES corr	0.015857	−0.021487	0.219789	−0.028446	0.142084	1.000000
p-value	0.777507	0.701783	0.000073	0.612175	0.010939	0.000000
HSN corr	−0.008080	0.062260	0.273359	0.057668	0.044837	1.000000
p-value	0.885517	0.266802	0.000001	0.303758	0.424098	0.000000
SSW corr	−0.072630	0.037779	−0.067832	−0.012565	0.070691	1.000000
p-value	0.195016	0.500688	0.226254	0.822841	0.207238	0.000000

Table C.25
Coefficients of Huber regression over all soundscapes. Results are similar to OLS.

Index	Const	Entropy	Complexity	ACI	ADI	BI	NumSpecies
PER coef	0.8371	−0.0051	0.0021	0.0012	0.0009	−0.0004	0.0928
p-value	0.000	0.000	0.136	0.286	0.460	0.752	0.000
UHH coef	1.0022	0.0042	−0.0223	−0.0013	0.0080	0.0174	0.2798
p-value	0.000	0.591	0.005	0.791	0.195	0.000	0.000
SNE coef	0.8270	−0.0141	−0.0028	−0.0005	0.0103	−0.0026	0.1225
p-value	0.000	0.000	0.457	0.781	0.000	0.126	0.000
POW coef	0.9717	−0.0117	−0.0305	−0.0073	0.0085	−0.0185	0.1862
p-value	0.000	0.000	0.000	0.069	0.003	0.000	0.000
NES coef	0.7158	2.714e−16	2.175e−16	−6.531e−16	−1.088e−15	1.349e−16	0.0321
p-value	0.000	0.174	0.253	0.000	0.000	0.220	0.000
HSN coef	0.8872	−0.0045	−0.0017	−0.0101	0.0016	0.0011	0.2409
p-value	0.000	0.000	0.036	0.000	0.044	0.066	0.000
SSW coef	0.7243	−6.939e−18	−5.638e−17	8.327e−17	−2.255e−17	−1.388e−17	0.0529
p-value	0.000	0.897	0.224	0.072	0.618	0.723	0.000

Table C.26
Coefficients of Huber regression over only samples filtered for birds out of the 2000 sample per soundscapes. Compared to Table C.25, each coefficient other than NumSpecies starts to become more influential in the regression.

Index	Const	Entropy	Complexity	ACI	ADI	BI	NumSpecies
PER coef	0.8506	−0.0054	0.0025	−0.0011	−0.0008	−0.0016	0.0847
p-value	0.000	0.001	0.113	0.402	0.562	0.238	0.000
UHH coef	1.1334	0.0187	−0.0444	−0.0108	0.0140	0.0396	0.2578
p-value	0.000	0.142	0.000	0.192	0.166	0.000	0.000
SNE coef	0.8919	−0.0237	−0.0123	0.0034	0.0159	−0.0051	0.1060
p-value	0.000	0.000	0.077	0.342	0.000	0.139	0.000
POW coef	0.9757	−0.0089	−0.0315	−0.0127	0.0033	−0.0218	0.1830
p-value	0.000	0.003	0.000	0.001	0.249	0.000	0.000
NES coef	0.7536	0.0164	−0.0109	−0.0204	0.0059	−0.0131	0.0330
p-value	0.000	0.002	0.026	0.000	0.021	0.000	0.000
HSN coef	1.2422	−0.1710	0.0056	−0.0532	0.0284	0.0266	0.1908
p-value	0.000	0.000	0.797	0.000	0.119	0.065	0.000
SSW coef	0.7872	−0.0053	−0.0086	−0.0028	0.0041	−0.0007	0.0488
p-value	0.000	0.319	0.081	0.433	0.267	0.824	0.000

Table C.27
Coefficients of Huber regression over samples of only 1 bird selected for the second round of the experiment.

index	Const	Entropy	Complexity	ACI	ADI	BI
SSW coef	0.7562	−0.0020	−0.0005	0.0005	0.0035	−0.0047
p-value	0.000	0.481	0.848	0.803	0.069	0.005
HSN coef	1.1418	−0.1217	−0.0420	−0.0004	0.0168	0.0058
p-value	0.000	0.000	0.000	0.960	0.075	0.443
PER coef	0.7492	−0.0084	−0.0016	0.0047	0.0070	7.905e−06
p-value	0.000	0.000	0.142	0.000	0.000	0.994
UHH coef	0.9469	0.0099	−0.0463	0.0089	0.0127	0.0333
p-value	0.000	0.301	0.000	0.167	0.123	0.000
SNE coef	0.8248	−0.0024	−0.0198	−0.0135	0.0215	−0.0036
p-value	0.000	0.646	0.000	0.000	0.000	0.163
POW coef	0.7210	−0.0079	−0.0311	−0.0211	0.0189	−0.0214
p-value	0.000	0.001	0.000	0.000	0.000	0.000
NES coef	0.7425	0.0033	2.097e−05	−0.0135	0.0065	−0.0094
p-value	0.000	0.365	0.995	0.000	0.000	0.000

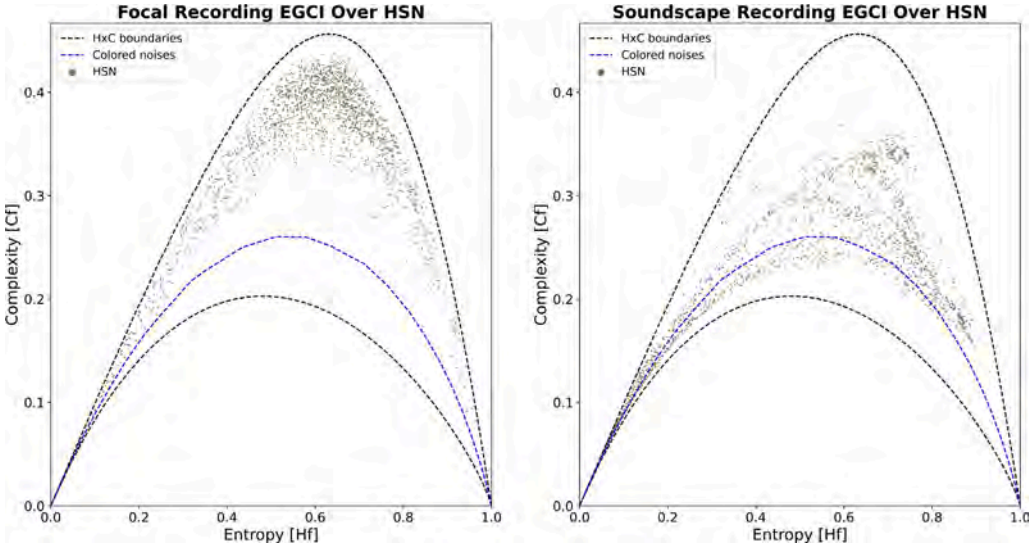

Fig. D.19. HSN focal by soundscape EGCI visualization.

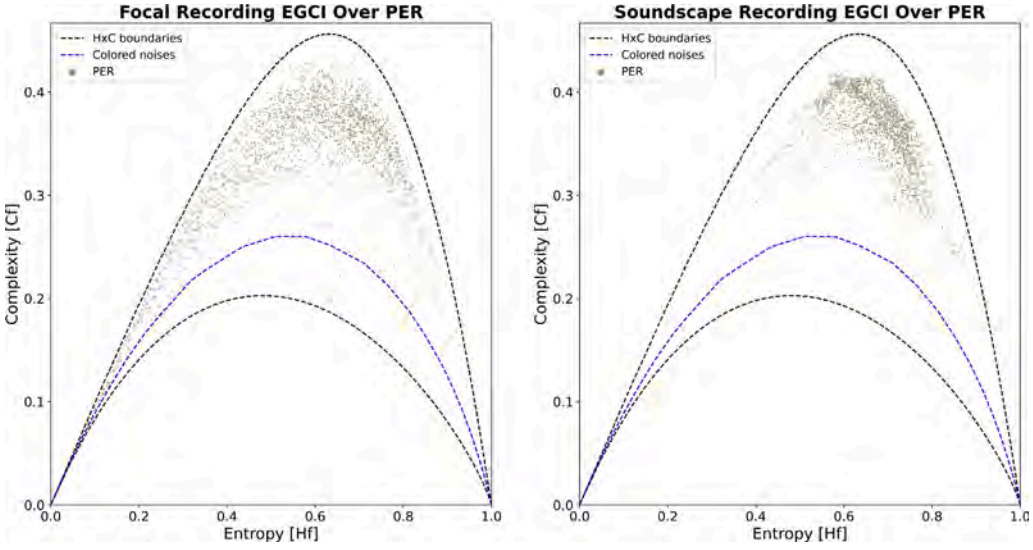

Fig. D.20. PER focal by soundscape EGCI visualization.

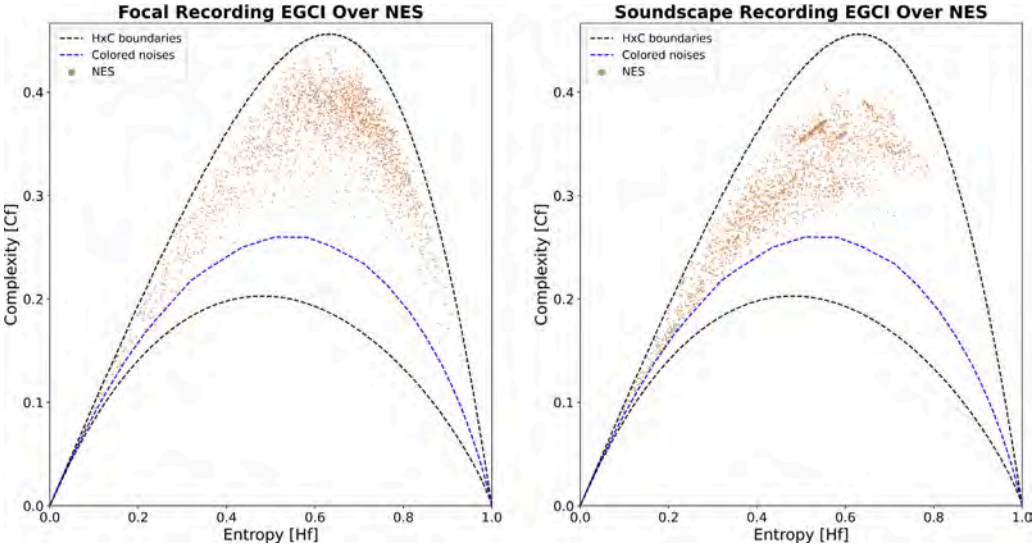

Fig. D.21. NES focal by soundscape EGCI visualization.

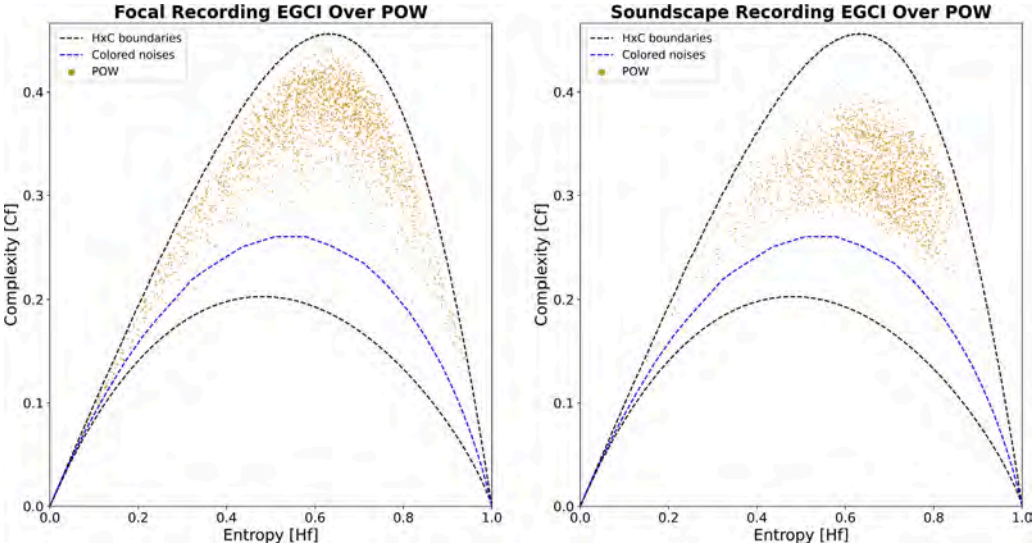

Fig. D.22. POW focal by soundscape EGCI visualization.

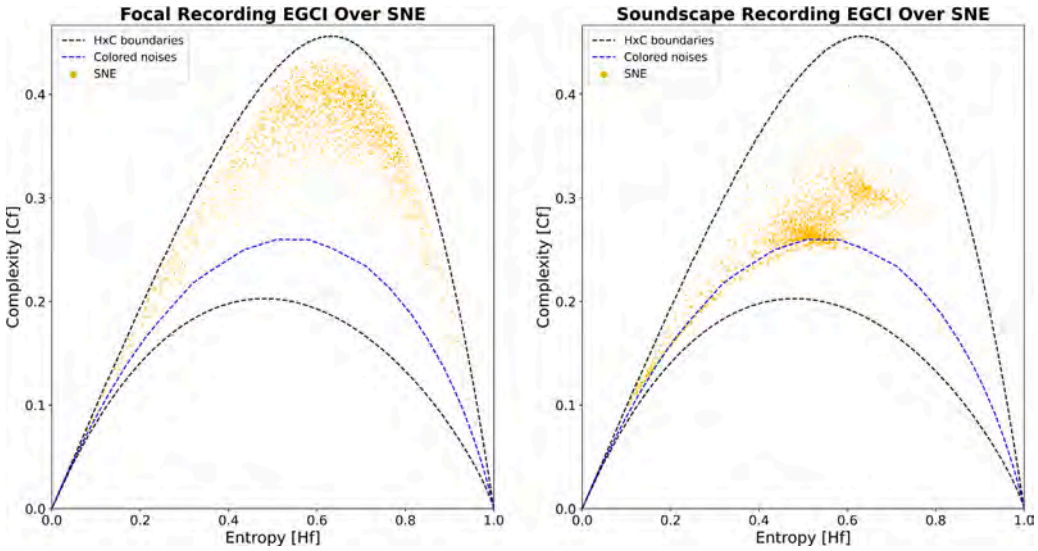

Fig. D.23. SNE focal by soundscape EGCI visualization.

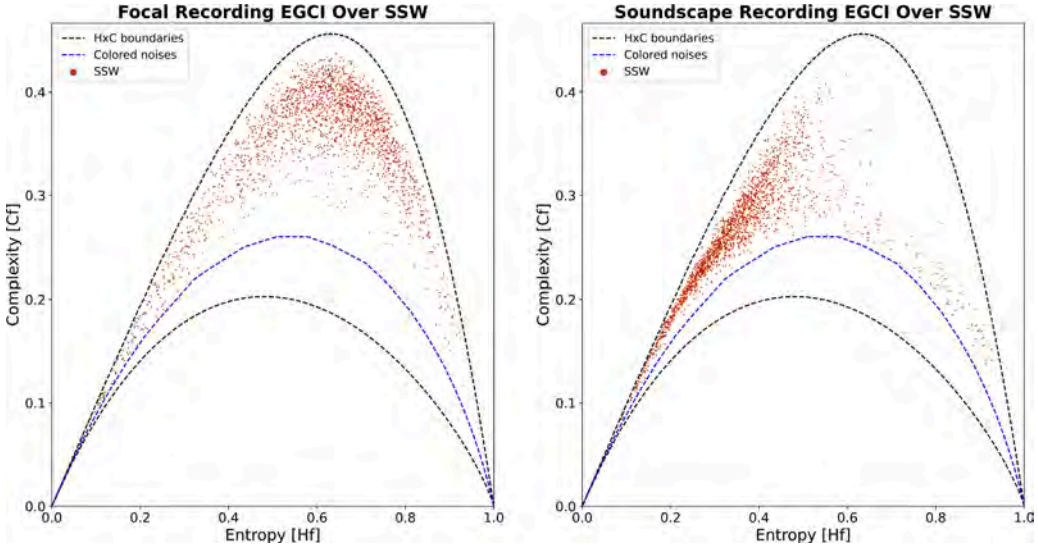

Fig. D.24. SSW focal by soundscape EGCI visualization.

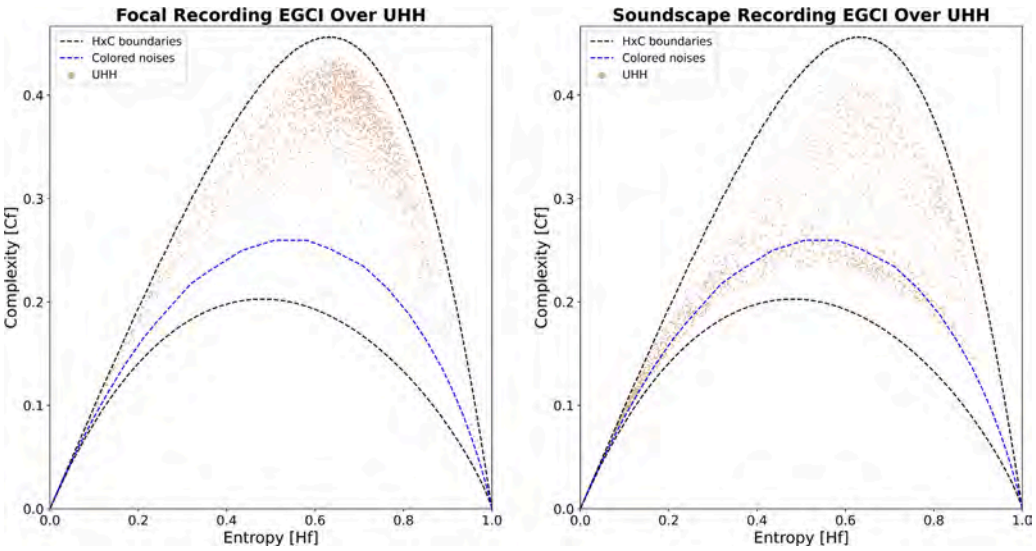

Fig. D.25. UHH focal by soundscape EGCI visualization.

References

Bhagoj, A.N., Cullina, D., Sehwag, V., Mittal, P., 2021. Lower bounds on cross-entropy loss in the presence of test-time adversaries - lower-bounds-icml21. URL: https://www.princeton.edu/~pmittal/publications/lower-bounds-icml21.

Bidarouni, A.L., Abeßer, J., 2024. Towards domain shift in location-mismatch scenarios for bird activity detection. (ISSN: 2076-1465) pp. 1267–1271. http://dx.doi.org/10.23919/EUSIPCO63174.2024.10715313, URL: https://ieeexplore.ieee.org/abstract/document/10715313.

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 (8), 2137–2144. http://dx.doi.org/10.1890/07-0004.1, URL: https://onlinelibrary.wiley.com/doi/abs/10.1890/07-0004.1.

Boudiaf, M., Denton, T., Merrienboer, B.V., Dumoulin, V., Triantafillou, E., 2023. In Search for a Generalizable Method for Source Free Domain Adaptation. PMLR, pp. 2914–2931, URL: https://proceedings.mlr.press/v202/boudiaf23a.html.

Bradfer-Lawrence, T., Desjonqueres, C., Eldridge, A., Johnston, A., Metcalf, O., 2023. Using acoustic indices in ecology: Guidance on study design, analyses and interpretation. Methods Ecol. Evol. 14 (9), 2192–2204. http://dx.doi.org/10.1111/2041-210X.14194, URL: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14194, arXiv:https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.14194.

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. 10 (10), 1796–1807. http://dx.doi.org/10.1111/2041-210X.13254, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13254.

Budka, M., Sokołowska, E., Muszyńska, A., Staniewicz, A., 2023. Acoustic indices estimate breeding bird species richness with daily and seasonally variable effectiveness in lowland temperate Białowieża forest. Ecol. Indic. 148, 110027. http://dx.doi.org/10.1016/j.ecolind.2023.110027, URL: https://www.sciencedirect.com/science/article/pii/S1470160X23001693.

Bustamante, N., Garitano-Zavala, A., 2024. Natural patterns in the dawn and dusk choruses of a neotropical songbird in relation to an urban sound environment. Animals 14 (4), 646. http://dx.doi.org/10.3390/ani14040646, URL: https://www.mdpi.com/2076-2615/14/4/646.

Casson, R.J., Farmer, L.D., 2014. Understanding and checking the assumptions of linear regression: a primer for medical researchers. Clin. Exp. Ophthalmol. 42 (6), 590–596. http://dx.doi.org/10.1111/ceo.12358, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ceo.12358.

Chasmai, M., Shepard, A., Maji, S., Van Horn, G., 2024. The inaturalist sounds dataset. In: Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., Zhang, C. (Eds.), Advances in Neural Information Processing Systems. Vol. 37, Curran Associates, Inc., pp. 132524–132544, URL: https://proceedings.neurips.cc/paper_files/paper/2024/file/ef3713b8d72266e803f9346088fed92d-Paper-Datasets_and_Benchmarks_Track.pdf.

Clark, M.L., Salas, L., Baligar, S., Quinn, C.A., Snyder, R.L., Leland, D., Schackwitz, W., Goetz, S.J., Newsam, S., 2023. The effect of soundscape composition on bird vocalization classification in a citizen science biodiversity monitoring project. Ecol. Inform. 75, 102065. http://dx.doi.org/10.1016/j.ecoinf.2023.102065, URL: https://www.sciencedirect.com/science/article/pii/S1574954123000948.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, second ed. Routledge, New York, http://dx.doi.org/10.4324/9780203771587.

Colonna, J.G., Carvalho, J.R.H., Rosso, O.A., 2020. Estimating ecoacoustic activity in the Amazon rainforest through information theory quantifiers. PLOS ONE 15 (7), e0229425. http://dx.doi.org/10.1371/journal.pone.0229425, URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0229425.

Denton, T., Wisdom, S., Hershey, J.R., 2021. Improving bird classification with unsupervised sound separation. http://dx.doi.org/10.48550/arXiv.2110.03209, URL: http://arxiv.org/abs/2110.03209. arXiv:2110.03209 [eess].

Dohi, K., Imoto, K., Harada, N., Niizumi, D., Koizumi, Y., Nishida, T., Purohit, H., Endo, T., Yamamoto, M., Kawaguchi, Y., 2022. Description and discussion on DCASE 2022 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques. URL: https://dcase.community/documents/workshop2022/proceedings/DCASE2022Workshop_Dohi_63.pdf.

Dufourq, E., Batist, C., Foquet, R., Durbach, I., 2022. Passive acoustic monitoring of animal populations with transfer learning. Ecol. Inform. 70, 101688. http://dx.doi.org/10.1016/j.ecoinf.2022.101688, URL: https://www.sciencedirect.com/science/article/pii/S1574954122001388.

Farina, A., Krause, B., Mullet, T.C., 2024. An exploration of ecoacoustics and its applications in conservation ecology. BioSystems 245, 105296. http://dx.doi.org/10.1016/j.biosystems.2024.105296, URL: https://www.sciencedirect.com/science/article/pii/S0303264724001813.

Farina, A., Righini, R., Fuller, S., Li, P., Pavan, G., 2021. Acoustic complexity indices reveal the acoustic communities of the old-growth mediterranean forest of Sasso Fratino integral natural reserve (central Italy). Ecol. Indic. 120, 106927. http://dx.doi.org/10.1016/j.ecolind.2020.106927, URL: https://www.sciencedirect.com/science/article/pii/S1470160X20308669.

Gasc, A., Sueur, J., Jiguet, F., Devictor, V., Grandcolas, P., Burrow, C., Depraetere, M., Pavoine, S., 2013. Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? Ecol. Indic. 25, 279–287. http://dx.doi.org/10.1016/j.ecolind.2012.10.009, URL: https://www.sciencedirect.com/science/article/pii/S1470160X12003603.

Ghani, B., Denton, T., Kahl, S., Klinck, H., 2023. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. Sci. Rep. 13 (1), 22876. http://dx.doi.org/10.1038/s41598-023-49989-z, URL: https://www.nature.com/articles/s41598-023-49989-z.

Ghani, B., Kalkman, V.J., Planqué, B., Vellinga, W.-P., Gill, L., Stowell, D., 2025. Impact of transfer learning methods and dataset characteristics on generalization in birdsong classification. Sci. Rep. 15 (1), 16273. http://dx.doi.org/10.1038/s41598-025-00996-2, URL: https://www.nature.com/articles/s41598-025-00996-2.

Gil, D., Llusia, D., 2020. The Bird Dawn Chorus Revisited. Springer International Publishing, pp. 45–90. http://dx.doi.org/10.1007/978-3-030-39200-0_3.

Goëau, H., Glotin, H., Vellinga, W.-P., Planqué, R., Joly, A., 2016. Lifeclef bird identification task 2016: The arrival of deep learning.

Goëau, H., Glotin, H., Vellinga, W.-P., Planqué, R., Joly, A., 2017. Lifeclef bird identification task 2017. In: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum. Dublin, Ireland, URL: https://hal.science/hal-01629175.

Goëau, H., Kahl, S., Glotin, H., Planqué, R., Vellinga, W.-P., Joly, A., 2018. Overview of birdclef 2018: monospecies vs. soundscape bird identification. In: CEUR Workshops Proceedings. Avignon, France, URL: https://hal.science/hal-02189229.

Haupert, S., Ulloa, J.S., scikit maad, Gil, J.F.L., Novoa, S.A., Suarez, G.A.P., Aumond, P., 2025. Scikit-maad/scikit-maad: New stable release :v1.5.1. http://dx.doi.org/10.5281/zenodo.15192421.

Heinrich, R., Rauch, L., Sick, B., Scholz, C., 2025. Adversarial training improves generalization under distribution shifts in bioacoustics. URL: https://arxiv.org/abs/2507.13727, arXiv:2507.13727.

Jordal, I., 2025. Iver56/audiomentations. URL: https://github.com/iver56/audiomentations. original-date: 2019-02-12T16:36:24Z.

Kahl, S., Denton, T., Klinck, H., Glotin, H., Goeau, H., 2021a. Overview of birdclef 2021: Bird call identification in soundscape recordings. In: Guglielmo, F., Nicola, F., Alexis, J., Maria, M., Florina, P. (Eds.), Proceedings of the Working Notes of CLEF 2021. CEUR-WS, Bucharest, Romania, pp. 1437–1450, URL: https://hal.science/hal-05182984.

Kahl, S., Wood, C.M., Eibl, M., Klinck, H., 2021b. Birdnet: A deep learning solution for avian diversity monitoring. Ecol. Inform. 61, 101236. http://dx.doi.org/10.1016/j.ecoinf.2021.101236, URL: https://www.sciencedirect.com/science/article/pii/S1574954121000273.

Kramer, H.A., Kelly, K.G., Whitmore, S.A., Berigan, W.J., Reid, D.S., Wood, C.M., Klinck, H., Kahl, S., Manley, P.N., Sawyer, S.C., Peery, M.Z., 2024. Using bioacoustics to enhance the efficiency of spotted owl surveys and facilitate forest restoration. J. Wildl. Manag. 88 (2), e22533. http://dx.doi.org/10.1002/jwmg.22533, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jwmg.22533.

Liang, J., Nolasco, I., Ghani, B., Phan, H., Benetos, E., Stowell, D., 2024. Mind the domain gap: A systematic analysis on bioacoustic sound event detection. (ISSN: 2076-1465) pp. 1257–1261. http://dx.doi.org/10.23919/EUSIPCO63174.2024.10714948, URL: https://ieeexplore.ieee.org/abstract/document/10714948.

López-Ruiz, R., Mancini, H.L., Calbet, X., 1995. A statistical measure of complexity. Phys. Lett. A 209 (5), 321–326. http://dx.doi.org/10.1016/0375-9601(95)00867-5, URL: https://www.sciencedirect.com/science/article/pii/0375960195008675.

Mair, L., Ruete, A., 2016. Explaining spatial variation in the recording effort of citizen science data across multiple taxa. PLOS ONE 11 (1), e0147796. http://dx.doi.org/10.1371/journal.pone.0147796, URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0147796.

Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 18 (1), 50–60. http://dx.doi.org/10.1214/aoms/1177730491, URL: https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-18/issue-1/On-a-Test-of-Whether-one-of-Two-Random-Variables/10.1214/aoms/1177730491.full.

Marvin, P., 2017. XC358862 Buff-bellied Pipit (Anthus rubescens). URL: https://xeno-canto.org/358862.

Massey, F.J., 1951. The Kolmogorov-Smirnov test for goodness of fit. J. Amer. Statist. Assoc. 46 (253), 68–78. http://dx.doi.org/10.2307/2280095, URL: https://www.jstor.org/stable/2280095.

van Merriënboer, B., Hamer, J., Dumoulin, V., Triantafillou, E., Denton, T., 2024a. Birds, bats and beyond: evaluating generalization in bioacoustics models. Front. Bird Sci. 3.

van Merriënboer, B., Hamer, J., Dumoulin, V., Triantafillou, E., Denton, T., 2024b. Birds, bats and beyond: evaluating generalization in bioacoustics models. Front. Bird Sci. Volume 3 - 2024, http://dx.doi.org/10.3389/fbirs.2024.1369756, URL: https://www.frontiersin.org/journals/bird-science/articles/10.3389/fbirs.2024.1369756.

Michaud, F., Sueur, J., Sèbe, F., Le Cesne, M., Haupert, S., 2025. Acoustic detection of a nocturnal bird with deep learning: the challenge of low signal-to-noise ratio. Ecol. Indic. 181, 114475. http://dx.doi.org/10.1016/j.ecolind.2025.114475, URL: https://www.sciencedirect.com/science/article/pii/S1470160X25014074.

Mutanu, L., Gohil, J., Gupta, K., Wagio, P., Kotonya, G., 2022. A review of automated bioacoustics and general acoustics classification research. Sensors 22 (21), 8361. http://dx.doi.org/10.3390/s22218361, URL: https://www.mdpi.com/1424-8220/22/21/8361.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011a. 5.2. Permutation feature importance. URL: https://scikit-learn.org/stable/modules/generated/sklearn.inspection.permutation_importance.html.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011b. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Penar, W., Magiera, A., Klocek, C., 2020. Applications of bioacoustics in animal ecology. Ecol. Complex. 43, 100847. http://dx.doi.org/10.1016/j.ecocom.2020.100847, URL: https://www.sciencedirect.com/science/article/pii/S1476945X19301606.

Pérez-Granados, C., 2023. A first assessment of birdnet performance at varying distances: A playback experiment. Ardeola 70 (2), 257–269. http://dx.doi.org/10.13157/arla.70.2.2023.sc1.

Perktold, J., Seabold, S., Sheppard, K., ChadFulton, Shedden, K., jbrockmendel, j grana6, Quackenbush, P., Arel-Bundock, V., McKinney, W., Langmore, I., Baker, B., Gommers, R., yogabonito, s scherrer, Zhurko, Y., Brett, M., Giampieri, E., yl565, Millman, J., Hobson, P., Vincent, Roy, P., Augspurger, T., tvanzyl, alexbrc, Hartley, T., Perez, F., Tamiya, Y., Halchenko, Y., 2024. Statsmodels/statsmodels: Release 0.14.2. http://dx.doi.org/10.5281/zenodo.10984387, URL: https://zenodo.org/records/10984387.

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 (3), 868–873. http://dx.doi.org/10.1016/j.ecolind.2010.11.005, URL: https://www.sciencedirect.com/science/article/pii/S1470160X10002037.

Pijanowski, B.C., Farina, A., Gage, S.H., Dumyahn, S.L., Krause, B.L., 2011. What is soundscape ecology? Anintroduction and overview of an emerging new science. Landsc. Ecol. 26 (9), 1213–1232. http://dx.doi.org/10.1007/s10980-011-9600-8.

Quinonero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N.D., 2022. Dataset Shift in Machine Learning. MIT Press, Google-Books-ID:MBZuEAAAQBAJ.

Rauch, L., Schwinger, R., Wirth, M., Heinrich, R., Huseljic, D., Herde, M., Lange, J., Kahl, S., Sick, B., Tomforde, S., Scholz, C., 2025a. BirdSet: A large-scale dataset for audio classification in avian bioacoustics. http://dx.doi.org/10.48550/arXiv.2403.10380, URL: http://arxiv.org/abs/2403.10380. arXiv:2403.10380 [cs].

Rauch, L., Schwinger, R., Wirth, M., Heinrich, R., Huseljic, D., Herde, M., Lange, J., Kahl, S., Sick, B., Tomforde, S., Scholz, C., 2025b. DBD-research-group/BirdSet · datasets at hugging face. URL: https://huggingface.co/datasets/DBD-research-group/BirdSet.

Rosso, O.A., Larrondo, H., Martin, M.T., Plastino, A., Fuentes, M.A., 2007. Distinguishing noise from chaos. Phys. Rev. Lett. 99 (15), 154102.

Shiner, J.S., Davison, M., Landsberg, P.T., 1999. Simple measure for complexity. Phys. Rev. E 59 (2), 1459.

Somervuo, P., Lauha, P., Lokki, T., 2023. Effects of landscape and distance in automatic audio based bird species identification. J. Acoust. Soc. Am. 154 (1), 245–254. http://dx.doi.org/10.1121/10.0020153.

Stowell, D., 2018. Computational Bioacoustic Scene Analysis. Springer International Publishing, Cham, pp. 303–333. http://dx.doi.org/10.1007/978-3-319-63450-0_11.

Teixeira, D., Maron, M., van Rensburg, B.J., 2019. Bioacoustic monitoring of animal vocal behavior for conservation. Conserv. Sci. Pract. 1 (8), e72. http://dx.doi.org/10.1111/csp2.72, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/csp2.72.

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

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. Scipy 1.0: Fundamental algorithms for scientific computing in python. Nature Methods 17, 261–272. http://dx.doi.org/10.1038/s41592-019-0686-2.

Wood, C.M., Barceinas Cruz, A., Kahl, S., 2023. Pairing a user-friendly machine-learning animal sound detector with passive acoustic surveys for occupancy modeling of an endangered primate. Am. J. Primatol. 85 (8), e23507. http://dx.doi.org/10.1002/ajp.23507, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ajp.23507.

Wood, C.M., Socolar, J., Kahl, S., Peery, M.Z., Chaon, P., Kelly, K., Koch, R.A., Sawyer, S.C., Klinck, H., 2024. A scalable and transferable approach to combining emerging conservation technologies to identify biodiversity change after large disturbances. J. Appl. Ecol. 61 (4), 797–808. http://dx.doi.org/10.1111/1365-2664.14579, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1365-2664.14579.

Y, G.D., Nair, N.G., Satpathy, P., Christopher, J., 2019. Covariate shift: A review and analysis on classifiers. In: 2019 Global Conference for Advancement in Technology. GCAT, pp. 1–6. http://dx.doi.org/10.1109/GCAT47503.2019.8978471.

Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D., 2018. Mixup: Beyond empirical risk minimization. URL: https://arxiv.org/abs/1710.09412, arXiv:1710.09412.