BioSoundSCape: Connecting acoustics and remote sensing to study animal-habitat diversity across environmental gradients

2025 Progress Report
June 1, 2024 - February 2, 2025

PI Matthew Clark (Sonoma State University, 1801 E. Cotati Ave, Rohnert Park, CA 94928 USA)

Festus Adegbola (University at Buffalo, USA)
Gary Doran (Jet Propulsion Lab, USA)
António Ferraz (Jet Propulsion Lab, USA)
Gurman Gill (Sonoma State University, USA)
Alan Lee (BirdLife South Africa, South Africa)
David Leland (Point Blue Conservation Science)
John Measey (Univ. Stellenbosch, South Africa)
Leo Salas (Point Blue Conservation Science, USA)
Fabian Schneider (Aarhus University, Denmark)
Colleen Seymour (SANBI, South Africa)
Rose Snyder (Point Blue Conservation Science, USA)
Wendy Schackwitz (Point Blue Conservation Science)
Andrew Turner (CapeNature, South Africa)

80NSSC22K0830

Progress

Acoustic data sharing and archival

In Year 3 we worked to openly share our Cape Region sound archive in the ONRL DAAC as well as develop an associated data description paper. We successfully transferred all acoustic data to the DAAC and collaborated in the development of a data description document and metadata. Co-I Turner and project co-authors submitted the data description paper to Nature Scientific Data entitled *BioSoundSCape: A bioacoustic dataset for the Fynbos Biome*. The paper is currently in review and describes the BioSoundSCape bioacoustic data set of over 900,000 minutes of audio recordings. The BioSoundscape DAAC repository will link to the data paper when published.

Species reference data collection

Since June 2024, a major focus of the BioSoundSCapes project has been on validating the accuracy of acoustics-based species identification deep learning (convolutional neural network, or CNN) models. Through a contract with BirdLife South Africa, contractors Oliver Angus and Campbell Fleming were hired to assess the presence of target amphibian and bird species, respectively, in random minutes of recordings. They performed an exhaustive annotation of the start-second of each instance of a target bird or amphibian in each recording, an effort that took several months to complete. These recordings, which we call “golden validations”, were not part of training the CNN model. Further, the golden validations were collected using a sampling methodology that differs from that used to find labels (i.e., song segments) to train the model. The golden validations were used to compare against the CNN model predictions, in order to determine confidence thresholds that maximize precision. That is, the precision of the model predictions is maximal when we consider as true detections any predictions whose confidence is equal or higher than the threshold. We also worked with citizen scientists Wendy Schackwitz and David Leland, and contractor Fleming to review WildMon CNN output and establish manual confidence thresholds, with the goal to find the threshold at which there were 80% true positive detections. The golden validation and manual CNN thresholds were used by PI Clark and Co-I Salas in our subsequent species detection pipeline.

Soundscape component (ABGI) detection with deep learning

As mentioned in our Year 2 report, PI Clark and collaborating computer science Professor Gurman Gill at Sonoma State worked with students to develop a CNN to detect broad soundscape components: biophony (birds, amphibians, frogs, insects), geophony (wind, rain), anthropophony (human-caused noises), and interference (e.g., branches knocking on the recorder), hereafter ABGI. This work began as class projects in the Spring 2024 semester. Gill and Clark worked with two students over Summer 2024 to refine the CNN model, experimenting with several model architectures, with accuracy tested against a single held-out set. We ultimately found that fine-tuning the BirdNET CNN (Kahl et al., 2021) with our training ABGI data resulted in a learned model with higher performance than other approaches (e.g., training another CNN used in acoustic work, YamNET). Clark applied this model to predict ABGI for all project acoustic data. The precision of this model was 100% with pattern matching data, which were the same domain as used to train the model. However, precision was 50% with a combination of golden validation data from 1-minute annotations combined with pattern matching test data, which is a more realistic representation of accuracy.

Species detections with the WildMon CNN

In Summer 2024, contractor WildMon delivered their first version of the CNN for species detection. This model used our reference data for training and testing based on pattern matching, as described in the Year 2 annual report. The WildMon CNN was based on a fine-tuning of BirdNET (Kahl et al., 2021) and had a reported average precision of 91% across 70 species of birds (52 species), frogs (14 species) and crickets (4 species). However, this accuracy was based on pattern match data that were very similar to the training data, and in our experience are positively biased (Clark et al., 2023). We sought to evaluate accuracy with an independent dataset consisting of 1-min randomly sampled recordings that were exhaustively annotated by our contractors, i.e., the golden validations.

Co-I Salas and PI Clark designed an analytical pipeline to evaluate and correct WildMon CNN model detections for 50 of the original 70 species of birds, frogs, and insects. Using golden validation data, we evaluated the performance of the model and found that it performs as is expected of CNN models, with about 82% precision (lower than the report 91% with pattern matching test data). We further used the golden validation dataset to find species-specific confidence thresholds that maximize the precision of the model, maximize the number of detections, and minimize commission error (i.e., false positive detections). After the use of these thresholds, model precision jumped from 82% to about 97%, but the recall (i.e., the percent of known presences that were detected) declined greatly. Because we know that prediction error is related to the presence of noise in the recordings, we used that information to correct the prediction. We used the ABGI CNN to predict the presence of these soundscape components in our dataset. We then used these predictions and knowledge of true vs false positive predictions from the WildMon CNN species model, and trained an ensemble of 500 random forest models per species of sound-emitting animal to identify the correct detections. This resulted in a "best of both worlds" result, with maximal precision and maximal sensitivity (Fig. 1). We applied the threshold plus post-hoc correction to all the predictions from our sound archive. We are currently working on a manuscript to describe our methodological approach and the accuracy of our models at each step of corrections.

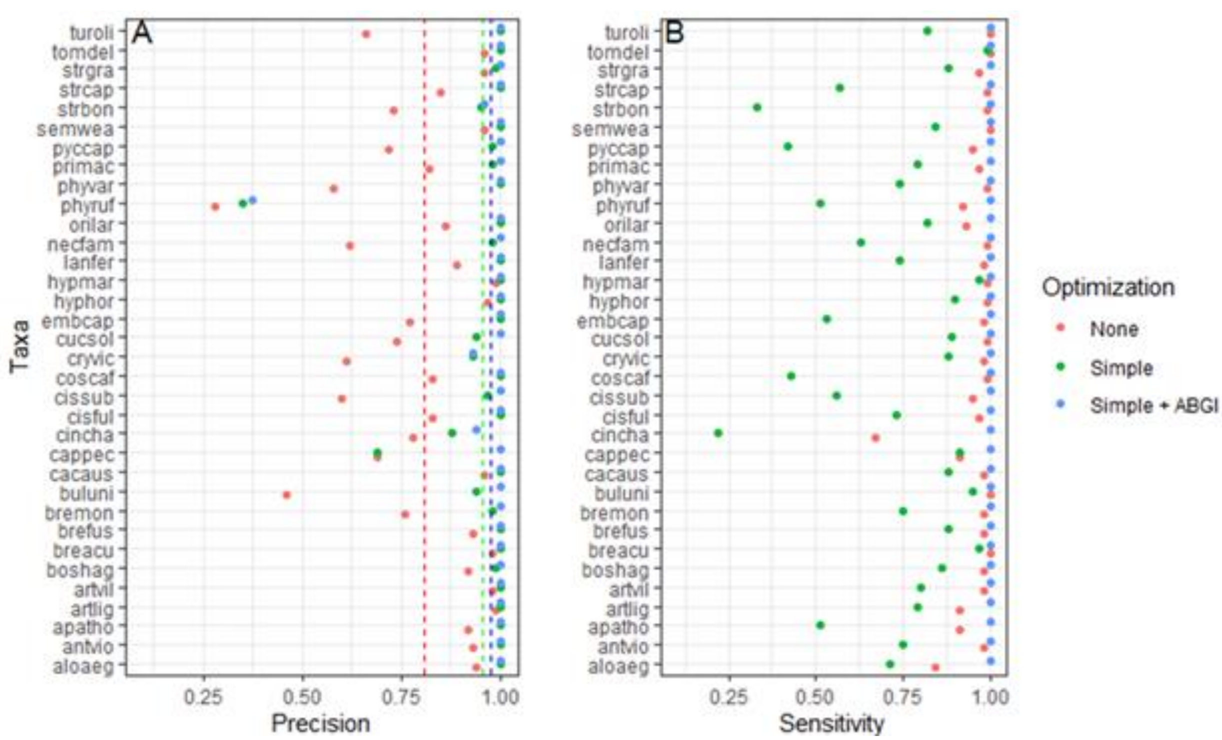

Figure 1. Comparison of three approaches to optimizing the predictive precision of an AI model trained to detect 70 sound-emitting taxa. Results shown include only the taxa with enough validation data to evaluate. The red dots are the detection predictions without any penalization,

with precision ~82%. But because these consider all predictions, model sensitivity is high – i.e., it may be detecting all that is there to be detected, even if making significant commission mistakes in doing so. The result of a simple threshold that maximizes the precision is shown with the green dots. By using a simple threshold we filter out detections that were correct, those with prediction confidence below the threshold, which results in lower sensitivity (see green dots in the sensitivity plot). Finally, the results of applying a post-hoc correction to ensure maximal precision sensitivity are shown with the blue dots. This approach results in precision ~98% and nearly 100% recall with the validation data.

Acoustic data processing

PI Clark has the full archive of BioSoundSCape data on a server at Sonoma State. In advance of Year 3 research goals, Clark processed the 909,309 minutes of sound data to derive a suite of 15 acoustic indices ([Quinn et al., 2023]) and VGGish CNN embeddings ([Sethi et al. 2020]). In Year 3, he also processed sound data with BirdNET embeddings. An ongoing task is to select the best UMAP dimensionality reduction hyperparameters to best separate our sites based on hourly averages of these embeddings (Figs 2 & 3). This will reduce 1024 BirdNET embeddings to several dimensions. These reduced data will then be used to determine functional richness, evenness and divergence ([Carmona et al., 2019]).

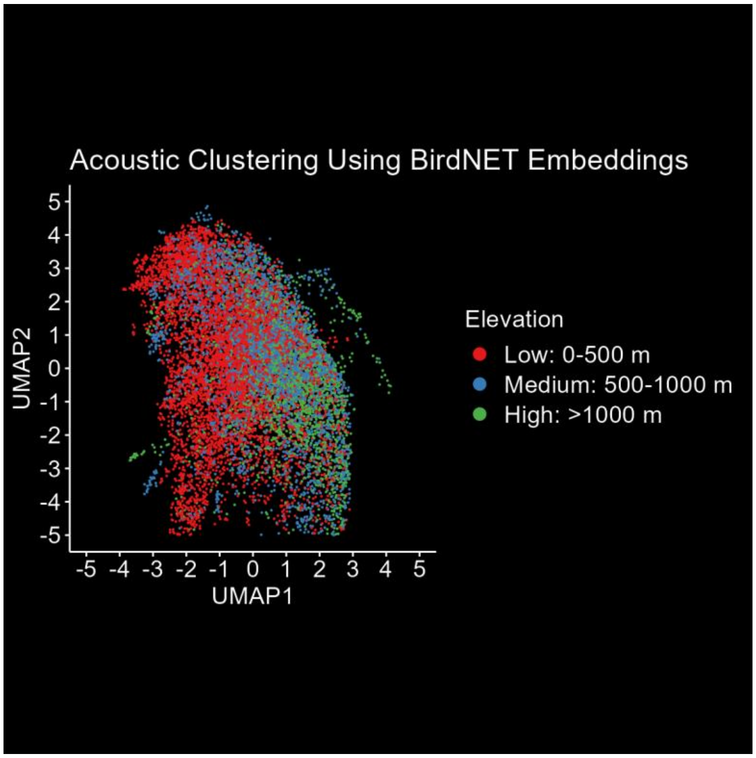

Figure 2. UMAP dimensionality reduction of hourly BirdNET embeddings with color by elevation.

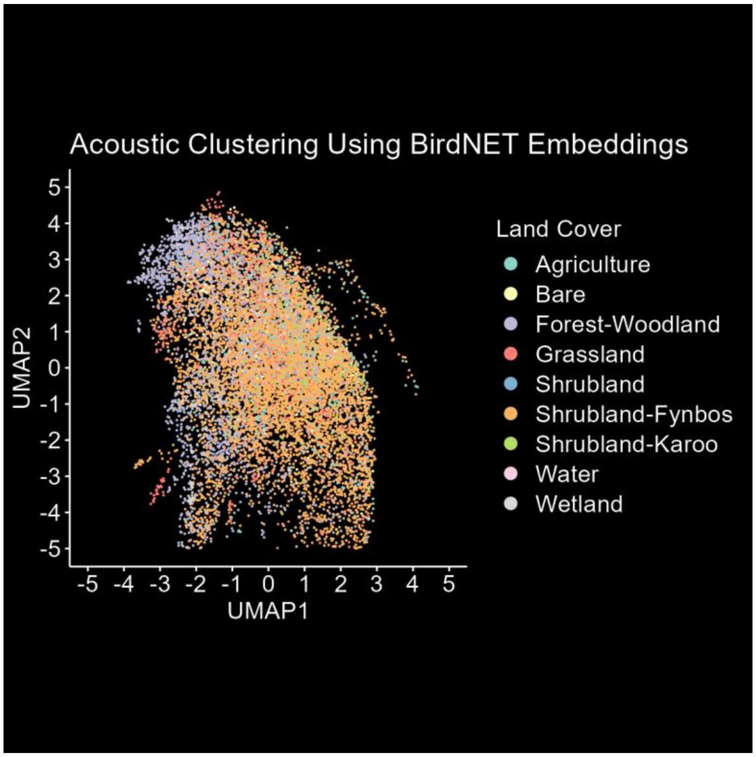

Figure 3. UMAP dimensionality reduction of hourly BirdNET embeddings with color by land cover.

Remote sensing - spectral diversity

In preparation to assess spectral diversity using functional richness, evenness, and divergence metrics, we have explored two approaches for reducing the dimensionality of the full 425-channel AVIRIS NG reflectance data down to 5 or fewer dimensions. The first approach uses a set of ratio-based spectral indices to measure the presence of vegetation, chlorophyll, water, nitrogen, lignin, carotenoids, and other indicators of biological materials (Fig. 4 & 5). The second approach uses principal component analysis to find a set of orthogonal vectors in the space of full-dimensional spectra that explains the most variance across the spectra (Fig. 6). The magnitudes of the top components within each spectrum can be used as another lower-dimensional representation.

We have applied our analyses to sites surrounding acoustic monitoring locations for future comparison between spectral and acoustic diversity using the latest version of the AVIRIS NG data (v2). Our analysis has revealed a need to filter out pixels within a site containing abiotic materials such as rock outcrops or human-made objects, which can produce spurious values in spectral indexes meant for determining vegetation properties. Another finding from our analysis is that some top principal components should be excluded from the lower-dimensional spectral representation, since they represent variations due to illumination or other factors not pertinent to biodiversity. These lessons learned will be incorporated as we proceed to the next step of computing functional diversity metrics from these representations.

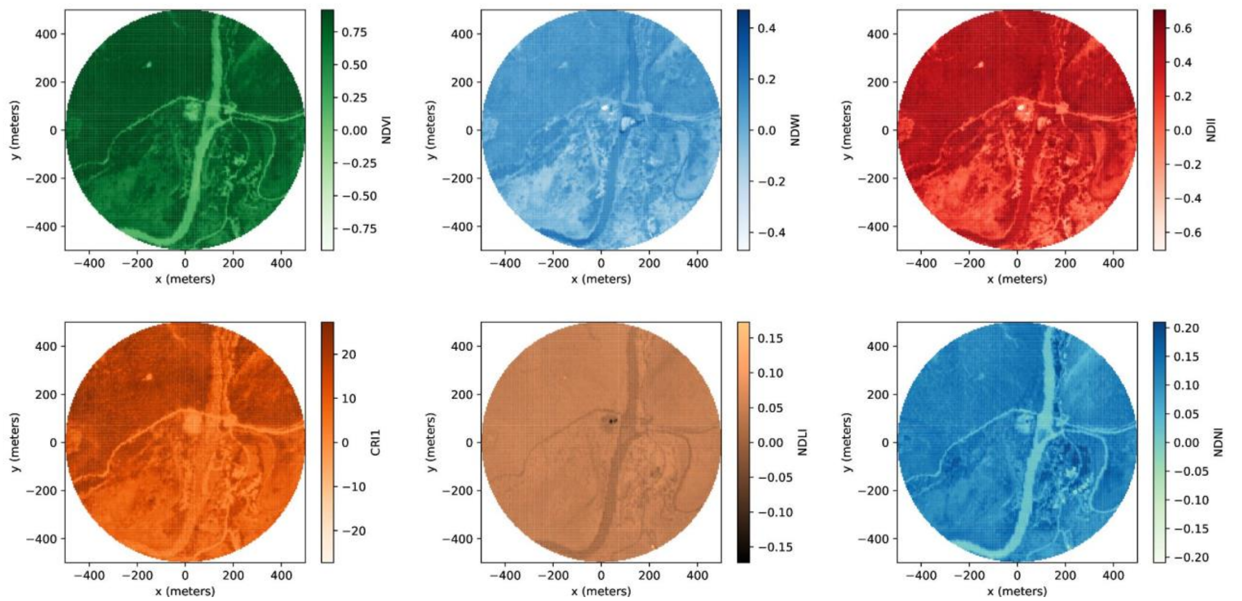

Figure 4. Spectral indices used in the analysis for a 250-m radius around an example site. Indices are Normalized Difference Vegetation Index (NDVI), Water Index (NDWI), Infrared Index (NDII), Carotenoid Reflectance Index 1 (CRI1), Lignin Index (NDLI), and Nitrogen Index (NDNI).

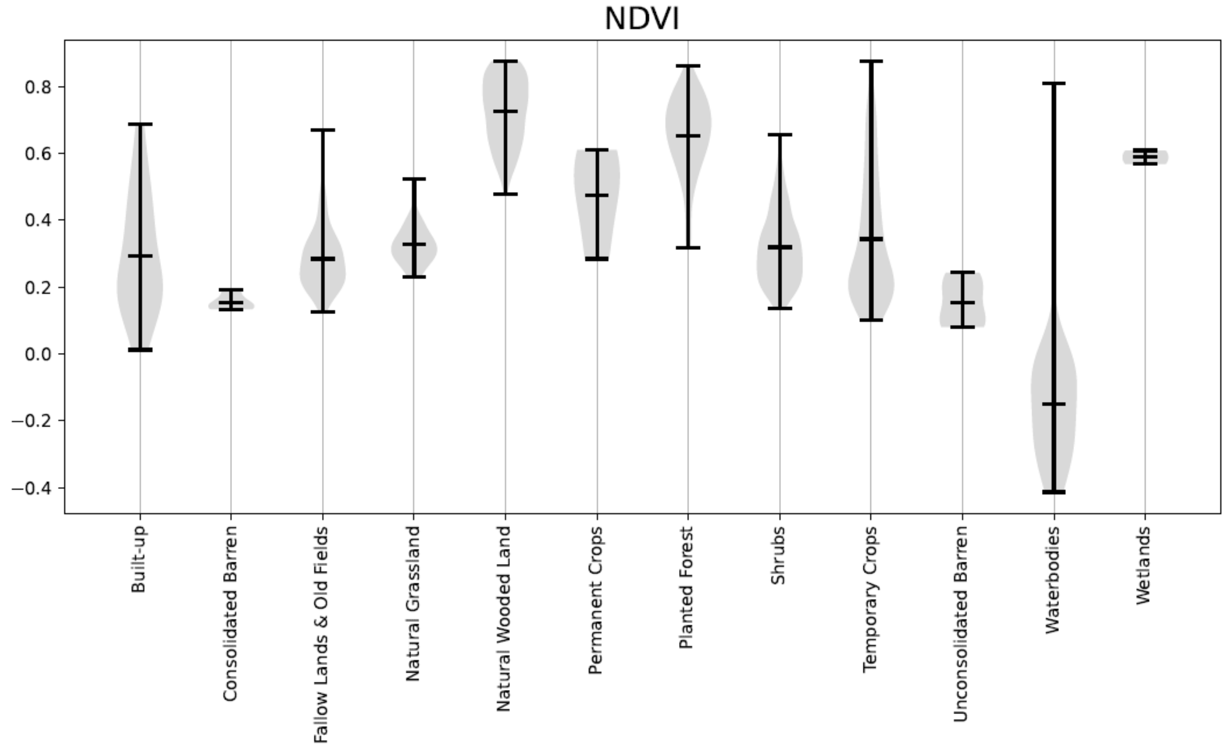

Figure 5. Normalized difference vegetation index (NDVI) statistics for broad land-cover classes. See land cover section for more information on samples.

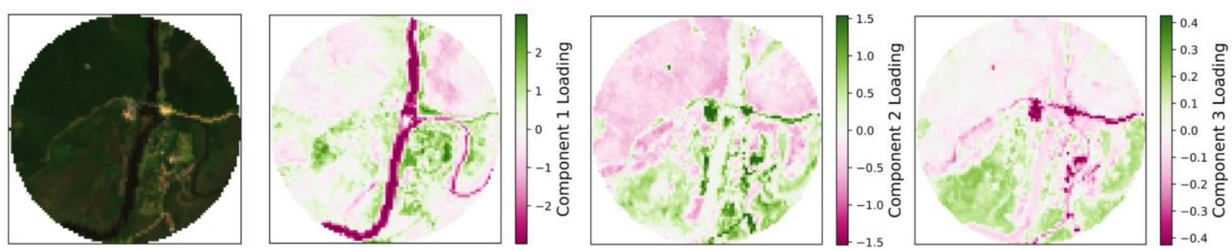

Figure 6. Example PCA component loadings for the top three components.

We also investigated the preliminary products from the Townsend team that is estimating plant traits from AVIRIS NG. These initial estimates are based on models designed using data from outside of South Africa and will be updated with models from the BioSCape vegetation plot chemical assays. Based on available data to date, we found that 364 of our >1000 sites have preliminary traits (Fig. 7). We will continue to explore these initial products along with spectral indices and principal components.

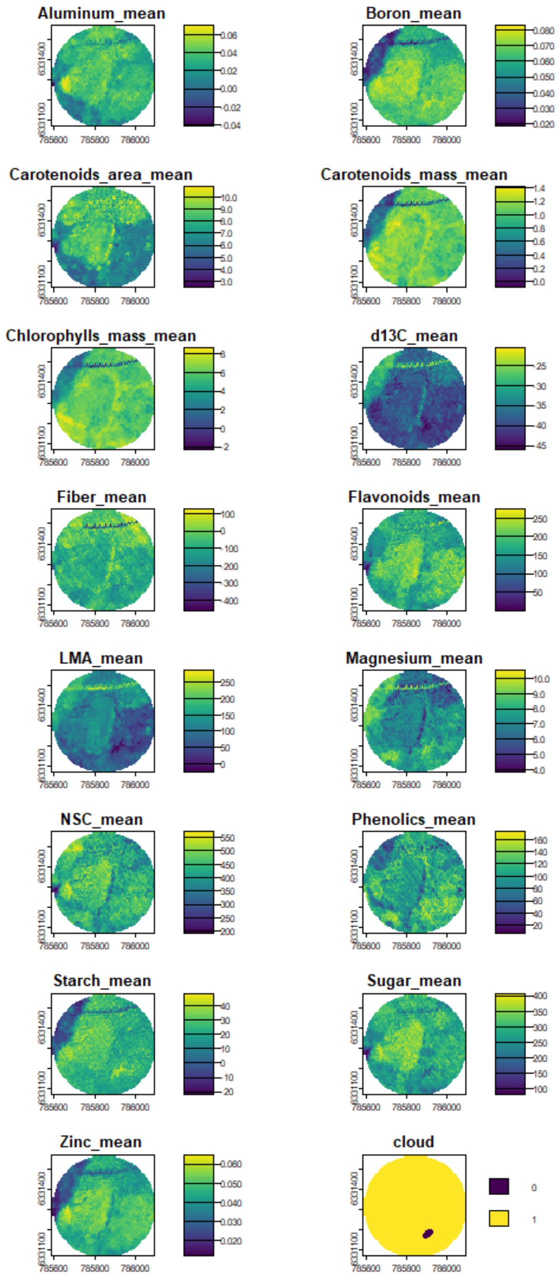

Figure 7. Townsend lab's preliminary plant traits for a 250-m radius area around an example site.

Remote sensing - LVIS structure

In Year 3, Clark organized LVIS L1B lidar metrics (e.g., relative height bins) for our study sites. These data will be used to assess structural richness, divergence and evenness (Carmona et al., 2019) and their relationship to species diversity, similar to the AVIRIS NG spectral approach.

Thanks to pre-planning by the BioSCape and BioSoundSCape team, only 68 of our 1074 plots had no LVIS coverage within a 250-m circular area, and most sites had over 1000 footprints (Fig. 9). This coverage was much more than originally expected and was due to the LVIS flight crew adding several targeted flight lines to acquire data over our sites.

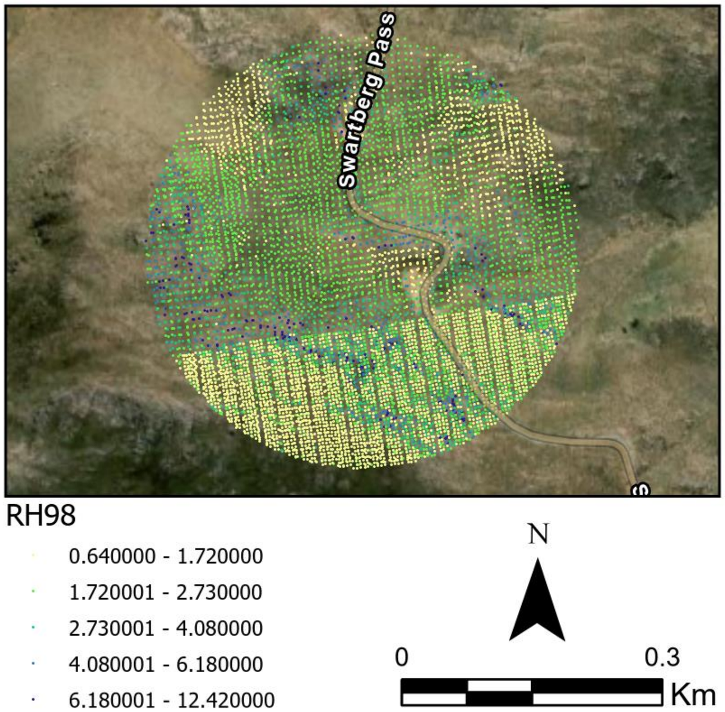

Figure 8. Example LVIS lidar relative height 98 percentile (RH98) for a 250-m radius around an acoustic sample site.

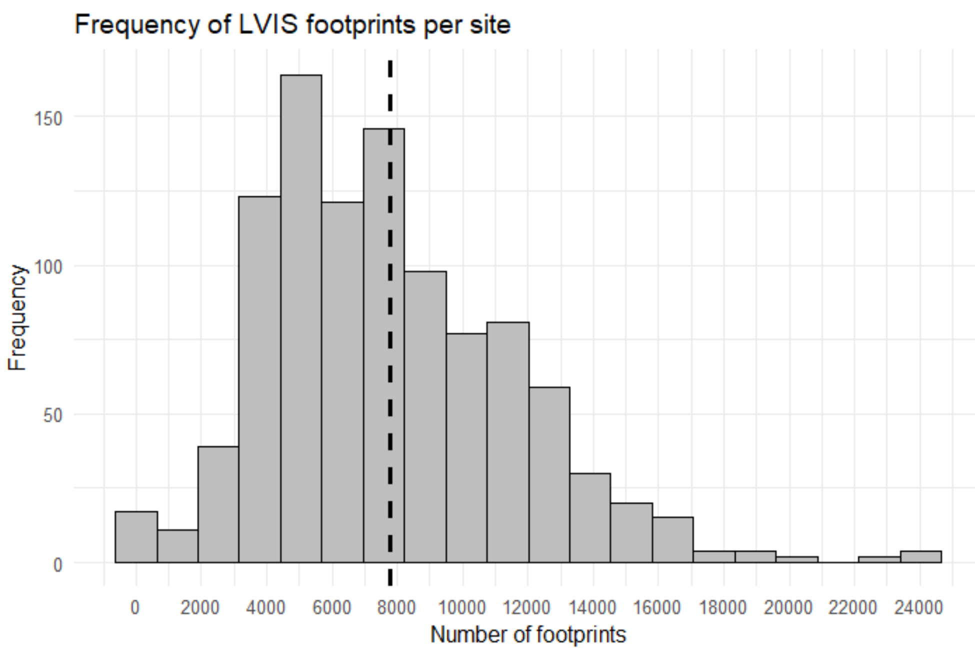

Figure 9. Frequency of the number of LVIS footprints in 250-m circular areas around sites (n=1074 sites). Dotted line is the mean of 7780 footprints.

Remote sensing - land cover

As part of a Computer Science capstone course with Dr. Gurman Gill at Sonoma State, PI Clark engaged with undergraduate students to classify broad land-cover classes in BioSCape AVIRIS NG data. Random 50 x 50-m samples (n=2423) were selected within the 2022 South African National Land Cover raster data product to cover a range of land cover classes across all AVIRIS NG flight boxes. Multiple students evaluated the land-cover class for each sample within Google Earth, with the plurality rule used to assign the final class, and ties re-evaluated or removed. These data were split into training and test sets. One group of students evaluated 2D (spatial) CNN classifiers (see class video), while the other group targeted 1D (pixel-based) CNNs (see class video). Clark is currently working with one exemplary student, Ben Harris, to evaluate all classifiers against a standard test dataset. So far, the best approach is a 1D CNN classifier, with a 78% overall accuracy (Fig. 9). Harris is now experimenting with more traditional 1D machine learning, including random forests and support vector machines. The final classifier with the best accuracy will be applied to pixels within various distances of our sample sites. These land-cover data can help us filter out pixels with vegetation as well as better understand the underlying drivers of spectral diversity within circular areas around sample sites. Further details of this student-led project are found on their GitHub.

Figure 9. 1D CNN land-cover classification of BioSCape AVIRIS NG imagery.

Dissemination and Engagement

Presentations, Conferences, Workshops, and Other Meetings

- PI Clark presented at the 4th African Bioacoustics Community Conference in Cape Town, South Africa in September 2024. His presentation focused on CNN classification of our target bird, frog and insect species that incorporated corrections to improve accuracy using broad soundscape components (biophony, anthropophony, geophony, interference) detected from a separate CNN model.
  Matthew Clark, Darien Labbe, Kathy Yuen, Gurman Gil, Leo Salas, Alan Lee, Daniel Cloete, Oliver Angus, John Measey, Andrew Turner, Colleen Seymour. (2024). Deep learning detection of bird, amphibian and insect species of the Cape Region, South Africa. 4th African Bioacoustics Community Conference, Cape Town, South Africa. September 1-6, 2024.

- Co-I Salas presented a poster at the 2024 Fall Meeting of the American Geophysical Union (AGU). The poster was an update on the CNN acoustics-based species detection and soundscape correction pipeline that we have developed.

  Leonard Salas, Matthew Clark, Andrew Turner, Alan Lee, Campbell Fleming, Colleen Seymour, Daniel Cloete, Darien Labbe, Gurman Gill, David Leland, John Measey, Kathy Yuen, Oliver Angus, Rose Snyder, & Wendy Schackwitz. (2024). What's That Sound? Evaluating and Understanding the Accuracy of AI Models for Sounds and Soundscape Detection, AGU 2024

- Co-I Turner delivered a presentation to the combined African Amphibian Working Group and herpetological Association of Africa highlighting the applied potential of BioSoundSCape through its use of automated amphibian biodiversity monitoring, especially in places where human and financial resources are constrained, which is typically the case in much of Africa.

  Andrew A. Turner, John Measey, Oliver Angus, Colleen Seymour, Alan Lee, Rose Snyder, Festus Adebgola, Leo Salas & Matt Clark. 2024. The BioSoundSCape project: Automating frog detection for ecological monitoring in the Cape Floristic Region. Oral presentation delivered at the combined African Amphibian working Group and Herpetological Association of Africa Symposium, Wilderness, South Africa, November 2024.

- Co-I Lee presented at the Fynbos Forum focused on an analysis of bird density estimates from BioSoundSCape point count data. He received recognition as runner-up for best presentation. This work continues to contribute to refining data analysis methodologies and strengthening biodiversity monitoring capacity within the project.

  Alan T.K. Lee, Matthew Clark, Antonio Ferraz, John Measey, Leo Salas, Fabian Schneider, Colleen Seymour, & Andrew Turner. (2024). Avian Conservation in the Cape Floristic Region: Insights from the BioSCape's BioSoundSCapes Project. 45th Annual Fynbos Forum 6 - 8 August 2024, Protea Hotel Technopark, Stellenbosch.

Publications and Other Products

- Andrew A. Turner, Matthew L. Clark, Leo Salas, Colleen Seymour, Rose L. Snyder, Alan T. K. Lee, Antonio Ferraz, Fabian Schneider , John Measey, Johan Huisamen, Daniël Cloete, Sally D. Hofmeyr, Christina Hagen, David F. Leland, Wendy Schackwitz, Festus Adebgola, Eugéne Hahndiek, Grant S. Joseph, Jacques Van Rooi, Michelle Fuchs, Saskia Thomas, Simphiwe Madlala, Jacky Spiby, & Patty Taljaard. (in review). BioSoundSCape: A bioacoustic dataset for the Fynbos Biome. Nature Scientific Data.

- 904,472 minutes of audio recordings at 1,081 sample sites across the Cape Region shared publicly on the ORNL DAAC.

  Clark, M., L. Salas, R. Snyder, A. Lee, A.A. Turner, C. Seymour, J. Measey, and A. Ferraz. 2024. BioSCape: BioSoundSCape Acoustic Recordings, South Africa, 2023. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/2372

- Bird point counts at acoustic sample sites (n = 884), with a total of 8,639 observations that included 249 species.
- Reference vocalizations (pattern matching) and golden validations (1-min recordings) for 70 species of birds, frogs, and insects and ABGI (anthropophony, geophony, biophony, interference) soundscape components
- Species detection CNN
- Soundscape ABGI CNN
- Complete database with all the metadata on the acoustic surveys.
- Miscellaneous R code files to generate all the results included in this report.
- Sound features from CNN embeddings and acoustic indices.
- GIS stratification factors and locations of acoustic sampling.
- Spectral indices and principal components for sample sites
- Land-cover reference data
- 1D CNN for land-cover classification

Future Plans

Below is our original proposal timeline. Year 1 of the grant was focused on field work planning, while Year 2 was dedicated to acoustic data collection in two field campaigns and subsequent data organization. Also in Year 2, we were awarded an augmentation grant to develop a CNN for species detection, which was not part of the original proposal. This work consumed much of our research time to date. Meanwhile, remote sensing products became available between Year 2 and 3, and our JPL team had a principal investigator leave for a faculty position outside of the US. Remote sensing work began in August, 2024 (Q3 of Year 3).

We are now fully engaged in the remote sensing side of the project but multiple papers, described below, that connect to each other are in various stages of development and will stretch beyond our Year 3 May 31, 2025 end date. We thus request a one-year No Cost Extension to continue working toward our publications.

Table 1. Timeline for proposed research activities. Colors: FCWG, SAWG, RSWG, All

Tasks		Y1	Q2	Q3	Q4	Y2	Q2	Q3	Q4	Y3	Q2	Q3	Q4
1	Planning flights & implementation plan												
2	Designing acoustic sensor network												
3	Acoustic data management												
4	Field campaigns & point counts												
5	Acoustic feature extraction and indices												
6	AVIRIS spectral diversity												
7	LVIS structural diversity												
8	RQ1. Animal-habitat diversity relationships												
9	RQ2. Animal-habitat diversity across disturbance gradients												
10	Data distribution												

Research activities are organized into three working groups (Table 1): Field Campaign Working Group (FCWG), Soundscape Analysis Working Group (SAWG), and Remote Sensing Working Group (RSWG).

Tasks 1-5

We have completed our research activities #1 through #5 as outlined in our original proposal (Table 1).

Tasks #6 & 7

We are behind in our remote sensing processing as AVIRIS NG and LVIS data were released during Year 3. Co-I Ferraz is working with data scientist Gary Doran at JPL to extract AVIRIS NG spectral reflectance data around sample sites and calculate spectral indices and principal components. Doran will evaluate various packages to calculate richness, evenness and divergence measures of spectral habitat diversity. Clark will coordinate with the JPL team to calculate similar measures of diversity for acoustic indices and CNN-based features (e.g, VGGish, BirdNET). Clark will finalize the 1D land-cover CNN and assemble LVIS structural data for plots. Ferraz, Doran, Salas and Clark will work on relating spectral and structural diversity to acoustic diversity and species richness from the WildMon CNN and bird point counts.

Task #8 Habitat-Animal Diversity relationship (RQ1)

We have separated this proposed research question into two papers.

Paper #1. Animal-habitat diversity relationship: Do measures of soundscape diversity correlate with estimates of *in situ* diversity, and does this vary in a predictable manner over time (seasons, times of the day) or space (e.g., elevation, habitat types)? PI Clark will lead this paper, with support from co-I Salas. We will test the linear or power-law relationship between *in situ* data (birds, frogs) and measures of soundscape diversity (evenness, divergence, richness), using several functions to describe the variance in the data. Our underlying hypothesis is that as animal diversity increases, so will soundscape diversity. For *in situ* data, we will use bird point counts from the field campaign and "virtual" frog point counts collected by contractor Angus. We have added new components to this research since the proposal. First, we will also consider as *in situ* data the species detections from our WildMon-developed CNN. Further, our "soundscape" ABGI CNN will be used to assess the impact of anthropophony, geophony and interference on our acoustic diversity. For acoustic features, we will use a subset of acoustic indices and BirdNET embeddings processed with UMAP dimensionality reduction. Since our data include multiple days and we have paired sites in wet and dry seasons, we will also explore different temporal windows (e.g,. dawn vs nighttime, wet vs. dry season) in a factorial design.

Paper #2. Do measures of soundscape diversity correlate with spectral and/or structural diversity? Does the relationship vary in space (e.g., habitat types) and at different spatial scales? Co-I Ferraz will lead this paper, with support from Doran, Salas and Clark. We will test for linear and power-law relationships among remotely-sensed habitat (spectral, structural) and soundscape diversity (best acoustic approach from Paper #1) measures across our GCFR study area, at a variety of spatial scales (e.g., 100, 250, 500, 1000 m). In general, we expect that habitat diversity will correlate with soundscape diversity. We hypothesize that the strongest acoustic-spectral structural relationship may be beyond the AudioMoth sampling range (~50m), as broader scale habitat characteristics might impact faunal diversity.

<u>Task #9 - Habitat-animal diversity along disturbance gradients (RQ2)</u>

As we mainly sampled in remote protected areas, we did not collect many soundscapes with high levels of anthropophony; and thus, we do not expect RQ2 to be a useful pursuit for this project.

<u>Task #10</u>

Paper #3. As discussed above, collaborator Turner submitted a data paper that outlines and summarizes the acoustic dataset. This paper will be revised and published this year.

Paper #4. This is a new paper that stems from our augmentation grant focused on CNN species detection. Salas will lead this paper. We will describe the methods and data sources used to develop the WildMon CNN. Accuracy and confidence thresholds will be presented based on manual thresholds as well as those adjusted after ABGI corrections from the soundscape CNN.

Paper #5. Co-I's Seymour and Lee have been invited to submit a paper for a special issue of the journal Ostrich on AI-based bird surveys. Seymour, Lee, Co-I Salas and others in the team are working on a first draft of a paper addressing the following hypotheses: (a) AI and human observers would not differ in their ability to detect the 50 species of birds we can detect with our trained CNN model - that is, the pattern of detections for each and all species across all sites sampled is highly correlated between humans and the AI model; (b) all 50 species would be similarly detectable by AI and human observers - that is, species poorly detected by humans are also poorly detected by the AI model and vice-versa; (c) AI and human observers do not differ in performance in different habitats (forest and shrubland) - that is, if there is a pattern of detection for any species across habitats in the human surveys, the same pattern is found in the AI predictions; and (d) AI and human observer performance does not differ between the dry and wet seasons (because of weather/timing of bird calls) - that is, if there is a pattern in detections in the human surveys across seasons, this pattern is also present in the AI predictions. We expect to submit the first draft of this paper by April.