

UCLA

UCLA Electronic Theses and Dissertations

Title

Using Autonomous Recording Units and Image Processing to Investigate Patterns in Avian Singing Activity and Nesting Phenology

Permalink

<https://escholarship.org/uc/item/92p9z0qp>

Author

Malamut, Evelyn Julia

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Using Autonomous Recording Units and Image Processing
to Investigate Patterns in Avian Singing Activity
and Nesting Phenology

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Bioinformatics

by

Evelyn Julia Malamut

2022

© Copyright by
Evelyn Julia Malamut
2022

ABSTRACT OF THE THESIS

Using Autonomous Recording Units and Image Processing
to Investigate Patterns in Avian Singing Activity
and Nesting Phenology

by

Evelyn Julia Malamut

Master of Science in Bioinformatics

University of California, Los Angeles, 2022

Professor Van M Savage, Chair

In order to investigate how birds are adapting to a changing environment, it is imperative that we have a better understanding of avian nesting phenology. The research outlined in this paper is intended to be a first step toward developing procedures and data processing pipelines that can be adapted to monitor nesting phenology at a continental scale. Here, we investigate the effectiveness of bioacoustic monitoring approaches to study the relationship between singing activity and nesting phenology. Using a mix of xeno-canto and field data recordings, we developed and applied species-specific, image-based song classification models for three bird species: Hooded Warblers, Grasshopper Sparrows, and Gray Vireos. We compared our proposed classifiers to BirdNET, a multi-species classifier developed by the Cornell Lab of Ornithology. Our proposed classifiers received an average AUC score for the ROC

curve of 0.89, 0.89, and 0.81, and an average AUC score for the precision-recall curve of 0.85, 0.79, and 0.85, respectively. Our proposed Hooded Warbler and Grasshopper Sparrow classifier performed better or the same as the BirdNET classifiers on at least 4/5 accuracy metrics, whereas the BirdNET classifier outperformed our Gray Vireo classifier on all accuracy metrics. Using our proposed Hooded Warbler and Grasshopper Sparrow classifiers, along with the BirdNET Gray Vireo classifier, we explored the relationship between singing activity and nesting phenology using meta-data collected at each study site. We observed little to no significant difference in song occurrence and singing rate within the standard nesting cycle for all three species. For Gray Vireos, there was a significant decrease in song occurrence after failure, fledgling, and at a nest without eggs compared to the song occurrence and singing rate during the nesting cycle. This distinct difference was not observed for Hooded Warblers and Grasshopper Sparrows. Overall, the methods proposed in this paper can be broadly adapted and applied to any vocal avian species. Given that this research is the first step in a potentially much larger project, there are many opportunities to fine-tune the approaches and extend the analysis.

The thesis of Evelyn Julia Malamut is approved.

Nandita Garud

Morgan W Tingley

Van M Savage, Committee Chair

University of California, Los Angeles

2022

TABLE OF CONTENTS

1	Introduction	1
2	Methods	6
2.1	Data Collection	6
2.2	Species Phenology and Vocalization Description	8
2.2.1	Hooded Warbler	8
2.2.2	Grasshopper Sparrow	10
2.2.3	Gray Vireo	13
2.3	Identification of Vocalizing Species	15
2.3.1	Training Data	16
2.3.2	Annotation Process	17
2.3.3	Model Training Process and Predictions	19
2.4	Application to Phenology Parameters	22
3	Results	26
3.1	Classifier Performance	26
3.2	Results of Singing Activity and Phenology Review	33
4	Discussion	43
4.1	Classifier Considerations	43
4.1.1	Potential Modifications and Future Considerations	45

4.2	Phenology Considerations	46
4.2.1	Hooded Warbler	47
4.2.2	Grasshopper Sparrow	48
4.2.3	Gray Vireo	49
4.2.4	Future Research and Use Consideration	50

LIST OF FIGURES

1	Autonomous Recording Unit at Pennsylvania Study Site.	7
2	Hooded Warbler Singing Patterns from Field Data.	10
3	Grasshopper Sparrow Singing Patterns from Field Data.	12
4	Gray Vireo Singing Patterns from Field Data.	14
5	Annotation of Grasshopper Sparrow Song in RavenLite.	18
6	Spectrogram Before and After Augmentation.	21
7	Data Processing Pipeline.	24
8	Recall vs Threshold.	27
9	Precision vs Threshold.	28
10	Precision Recall Curve.	30
11	F1 Scores Curve.	31
12	Receiver Operating Curve.	32
13	Pairwise Comparisons of Song Occurrence Between Phenophases.	35
14	Pairwise Comparisons of the Average Singing Rate Between Phenophases.	36
15	Song Occurrence Across the Nesting Cycle.	39
16	Average Singing Rate Across the Nesting Cycle.	40
17	Song Occurrence Across the Season.	41
18	Average Singing Rate Across the Season.	42

LIST OF TABLES

1	Study Site Nest Fates and Clutch Size.	15
2	Duration of Phenological Stage Parameters for Each Species.	23
3	Optimal Threshold Values and Classifier Accuracy Metrics.	33

GLOSSARY

Autonomous Recording Unit (ARU) - Self-contained audio recording device that researchers deploy to monitor the environmental acoustics of an area (ex. Audiomoth)

Data Processing Pipeline - a set of data processing elements connected in series, where the output of one element is the input of the next one. The elements of a pipeline are often executed in parallel or in time-sliced fashion

F1 score - a balanced measure of overall classifier performance combining precision and recall. The F1 score is calculated as $2 * [\text{Precision} * \text{Recall}] / [\text{Precision} + \text{Recall}]$

Nesting Phenology - the study of periodic events or stages and their timing and duration in the nesting cycle

Spectrogram - a visual depiction of audio signal frequency as it varies with time

Phenoday - a day of the nesting cycle timeline. Date of first egg laid is day zero

Phenophase - a specific event in the nesting cycle (ex. incubation)

Precision - the proportion of apparent detections that correspond to true focal species songs

Recall - the proportion of true target species song detections out of the total amount of songs from the focal species

Song Occurrence - the fraction of recordings that contain at least one song out of the total number of recordings for a certain grouping

Threshold - a confidence value above which a clip would be considered to have a bird of interest song present

ACKNOWLEDGMENTS

I would like to thank my professor and mentor, Morgan Tingley, for his invaluable guidance and encouragement over the course of this research. I would additionally like to thank all members of the Tingley Lab for their support, wisdom, and infectious enthusiasm for birds. Finally, I would like to thank UCLA undergraduate Kylie Lim for her work annotating hours of recordings for this study.

We at UCLA acknowledge the Gabrielino/Tongva peoples as the traditional land caretakers of Tovaangar (the Los Angeles basin and So. Channel Islands). As a land grant institution, we pay our respects to the Honuukvetam (Ancestors), ‘Ahihirom (Elders) and ‘Eyoohiinkem (our relatives/relations) past, present and emerging.

1

Introduction

Nesting phenology describes the study of periodic events or stages that occur predictably in the seasonal nesting cycle of oviparous species. Understanding the nesting phenology of avian species allows us to investigate how birds are adapting to climate change along with other important avian behaviors. It also provides us with useful insights on potential threats to species and phenological mismatch. Phenological mismatch occurs when species change the timing or duration of phases in their life cycle, which may then lead to disruptions in the life cycles of species they interact with. Many species have adapted their nesting phenology with the changing climate, generally by starting the nesting cycle earlier in the season [1–3]. In order to study these potential changes and create conservation strategies, it is imperative that we have a better understanding of continent-wide multi-species nesting phenology. However, there is currently no such program to undertake these studies for birds in the United States. The research outlined in this paper is based on a pilot study that is intended to be a first step toward developing procedures and data processing pipelines that can be adapted to monitor nesting phenology at a continental scale.

Specifically, in this paper, we were interested in the nesting phenology of migratory birds throughout the breeding season. The nesting cycle of most bird species adhere to a typical pattern of nest building, followed by egg laying, incubation, caring

for hatchlings, and caring for fledglings. Hatchling care differs from fledgling care in that hatchlings are not yet able to leave the nest, whereas fledglings have left the nest and are fed in various locations throughout the area. We use the term “phenophase” to indicate a specific period in the nesting cycle. At any stage in this process, there is a chance of failure due to predation, brood parasitism, or other causes of egg mortality. When these types of failures occur, depending on resource availability and seasonal timing, birds may start this nesting process again at a new nest. Previous work supports that vocalization activity changes with breeding cycle and is a likely signal of change in phenophase [4–8]. Building upon this well-documented relationship, we investigate, in this paper, the effectiveness of bioacoustic monitoring approaches to study the relationship between nesting phenology and vocalization activity.

Passive acoustic monitoring is a data collection technique, in which recording devices take samples of a soundscape without a human present. It has become an effective and widely used means to study ecological bioacoustics. Passive acoustic monitoring utilizes autonomous recording units (ARU) to collect audio recordings. ARUs are self-contained audio recording devices that researchers deploy to monitor the environmental acoustics of an area. They can be programmed to record at times of interest throughout the day, month, or year and generate a permanent, consistent record of a soundscape that researchers can continually reference and analyze. Researchers use this capability to take repeated samples of an area with less human effort and disturbance of the environment. ARUs help reduce inter-researcher variation in collecting data and allow us to study ecological bioacoustics in novel ways [9]. ARUs can provide a more accurate representation of a soundscape, sounds from an environment without a particular focus, than can be achieved with continual

human disturbance. They are also ideal for use in study sites that researchers cannot monitor well or where staff have limited bird identification expertise or time. ARUs can greatly aid research on species distributions, phenology, effects of anthropogenic noise, and social networks [10]. Researchers have used ARU recordings to detect minute variations in vocalization patterns that humans may not be able to perceive and identify birds in landscapes that are rare or endangered [10, 11].

Researchers are generally interested in isolating a particular signal from the data, such as vocalization by a species of interest. Currently, identification of species from ARU data is often done by researchers who listen to the recordings in order to identify a focal species [12]. However, this approach is time-consuming and prone to inter-researcher bias [13]. Recently, some researchers have developed machine learning approaches to identify species from audio recordings [14–18]. Many of these machine learning approaches employ image processing techniques on the collected data. Image processing involves performing operations on an image to extract useful information from it using some type of computer program or algorithm. In this context, this procedure generally involves building a machine learning classification model which takes spectrograms with annotated vocalizations as input and generates output that informs us when our focal species is vocalizing within the time period of the recording. This type of audio processing is commonly used to study animal and specifically bird acoustics [14, 19, 20].

The BirdCLEF challenge evaluates current bird vocalization identification techniques by providing a large-scale annotated vocalization training set and inviting participants to develop identification methods. The most recent (2021) dataset consisted of xeno-canto recordings and soundscape recordings. In 2015 and 2016, the dataset consisted of only xeno-canto recordings. In 2017, the dataset added sound-

scapes, and in 2019, the dataset was completely comprised of soundscape recordings. Xeno-canto is a collaborative online database where people can upload audio clips of known birds and provide information on the recording such as length, date, time, location, type of vocalization, and other notes (xeno-canto.org) [21]. For the 2015 competition, Joly et al. employed a nearest neighbors approach and Lasseck et al. utilized decision trees for classification [16, 17]. In 2016, Sprengel et al. relied on the similarity between human speech and bird vocalization to identify bird vocalizations, using speech processing techniques [15]. Their method resulted in an average precision score of 0.69. Lasseck, the winner of the BirdCLEF challenge in 2018 and 2019, developed a deep convolutional neural network in 2019 that achieved a mean average precision of 0.36 and a retrieval mean average precision of 0.75 [18].

Outside of the BirdCLEF challenge, Kahl et al. developed a deep learning neural network called BirdNET that can identify 3,000 of the world’s common avian species [14]. BirdNET, developed by the Cornell Lab of Ornithology, is now the most common bird ID software used in the scientific community that can process data on a large scale. While these large-scale identifiers are impressive and useful for public education, it remains uncertain whether broad-scale, multi-taxa classifiers (like BirdNET) are more effective for purpose-driven science focused on just a few focal species. Consequently, for the purposes of the present investigation, we have chosen to build our own data processing pipeline with the goal of optimizing performance on detecting focal species. Developing our own methods gives us more flexibility to manipulate specific parameters and procedures relative to the species of interest.

Here, we built individual species classifiers, evaluated those classifiers relative to BirdNET, and conducted preliminary investigations of the relationship between singing activity and breeding phenology for three migratory species, the Hooded War-

bler (*Setophaga citrina*), Grasshopper Sparrow (*Ammodramus savannarum*), and Gray Vireo (*Vireo vicinior*). We then developed a data processing pipeline, which involves the use of species-specific, image-based acoustic convolutional neural network classifiers trained on a mix of publicly accessible published reference recordings and new field data to identify each species from ARU data. The classifiers were built using open-source software (opensoundscape.org) that allows users to generate custom data processing pipelines to detect the presence of species in field recordings [22]. Using classified singing activity from field recordings, we asked the following questions: can we detect changes in singing activity in response to changes in nest state (e.g., when a nest fledges, or when a nest fails)? Can we detect a difference between recordings taken at a nest versus recordings from random locations? Do we see patterns emerge from singing activity in relation to phenophase and can we determine differences in singing activity between different phenological stages? Are there factors other than phenophase change that affect singing rate throughout the nesting cycle or season? Do we see overall trends in the relationship between singing and phenology among bird species? This research is ongoing and as we collect more data, we hope to better understand these relationships.

Methods

2.1 Data Collection

The data used in this research were collected as part of a pilot program to monitor bird phenology and vocalization in the summer of 2021. The data came from four study sites across the United States that vary widely in ecosystem and associated bird communities. We collected data on Gray Vireos in the pinyon-juniper woodlands of New Mexico, Grasshopper Sparrows in the prairies of Kansas, Hooded Warblers in the mixed deciduous forests of Pennsylvania, and Seaside Sparrows (*Ammodramus maritimus*) in coastal salt marsh of Connecticut. These four study species were chosen ad hoc as all four were already being studied as part of intensive nest monitoring programs by collaborators who were willing to work with us on this study. Even though species choice was not random or planned, the species do span a wide phylogenetic breadth of the oscine radiation, and have differing ecologies and breed in widely different habitats (which may differ in their soundscape qualities). Due to time constraints, analyses of Seaside Sparrow data are not included in this paper.

Throughout the season, we collected phenology and vocalization data on the birds at the monitored nest sites. We summarized the nest data into seven variables describing monitoring (nest location, nest status, deployment and removal of ARU,

recording number, and SD memory card number), and four variables describing phenology and productivity (date of first egg, clutch size, number of host bird eggs, number of cowbird eggs, fledge/fail status, and date of fledge or fail). The vocalization data were collected using Audiomoths, a brand of autonomous recording unit. Each Audiomoth was programmed to record for 15 minutes at the start of every hour for five hours of the day. The first recordings began at 5:30 AM local time and the last recordings began at 11:30 AM local time. The Audiomoths were placed in plastic bags to prevent water damage and animal tampering. They were then attached to trees or other stationary objects throughout the study site (Figure. 1).



Figure 1: Autonomous Recording Unit at Pennsylvania Study Site. The ARU is placed in a plastic bag sealed with duct tape to avoid water damage and animal tampering and attached to a tree.

In April 2021, each study site was provided with 25 AudioMoths. In May–June, ten ARUs were placed at quasi-random locations around the study area. The other 15 ARUs were placed within five to ten meters of a nest, often within 1 day of nest discovery. Most of the nest monitoring ARUs were placed during nest construction, egg laying, or early incubation. At some sites, the nest monitoring ARUs were redeployed to new nests after fledging or failure of a nest. AudioMoths were removed from sites once breeding ceased (by September).

2.2 Species Phenology and Vocalization Description

Each of the species descriptions and parameters used in the phenology analysis were drawn from relevant literature and reviewed by researchers familiar with the populations of each species at their specific study site.

2.2.1 Hooded Warbler

Hooded Warblers are a long-distance migratory species that typically nest in mixed hardwood forests with dense understory. In Pennsylvania, females settle onto pre-established male territories from May 7th to June 4th. Nest-building typically lasts five to six days. After the nest is lined, the female will lay eggs on consecutive days until the full brood is established [23]. Incubation then starts on the day that the last egg is laid and spans 12 days (Chapman 1907, Bent 1953). After this, the eggs generally hatch over a period of one to two days. After eight to ten days (the caring for hatchlings stage), the young typically fledge [23].

A study found that 44% of Hooded Warblers with successful first nests have

second broods, with the date of the first egg ranging from June 2nd to July 19th [24]. For a second brood, the female remains with the same male and establishes a new nest on the same territory, typically ten to fourteen days after fledging of the first brood [23]. At any stage, nest failure may occur due to brood parasitism, predation, or other factors. In our study, 60% (25/42) of nests associated with ARUs failed. Of those nests, 10% (4/42) were affected by brood parasitism, with a 75% (3/4) fail rate after parasitism (Table. 1). However, most nests failed from predation. In addition, some lined nests never received eggs or the nests were abandoned for some other reason. After a nest failure, Hooded Warblers will abandon the nest and attempt to re-nest with the same mate in the same territory [23].

Hooded Warbler males sing starting at arrival on the breeding grounds and throughout the breeding season. Hooded Warblers sing to attract mates and for territorial defense [23]. Singing by females is sparsely documented, and believed to be very rare [25]. Hooded Warblers have four to nine distinct singing patterns, which can be grouped into “repeat” songs and “mixed” songs. The repeat songs consist of about 4-5 notes repeated at a rate of 5-6/min (Figure. 2a). The mixed songs are sung at irregular intervals at a rate of 10-12/min (Figure. 2b)[26].

Before mating, Hooded Warblers spend much of their time (50-60%) singing [23]. In terms of territorial defense, males will increase their singing rate in response to a neighbor’s song [26]. However, after mating, the time spent singing drops to 20% during nest building and incubation. At this time, males are also less likely to respond to a neighbor’s song [23, 26]. The amount of vocalization rises to 30% during egg-laying, but then decreases during incubation, feeding hatchlings, and feeding fledglings. The stage of the nesting cycle and the hour of the day influences the rate of singing more than the day of the season or age of the individual [26].

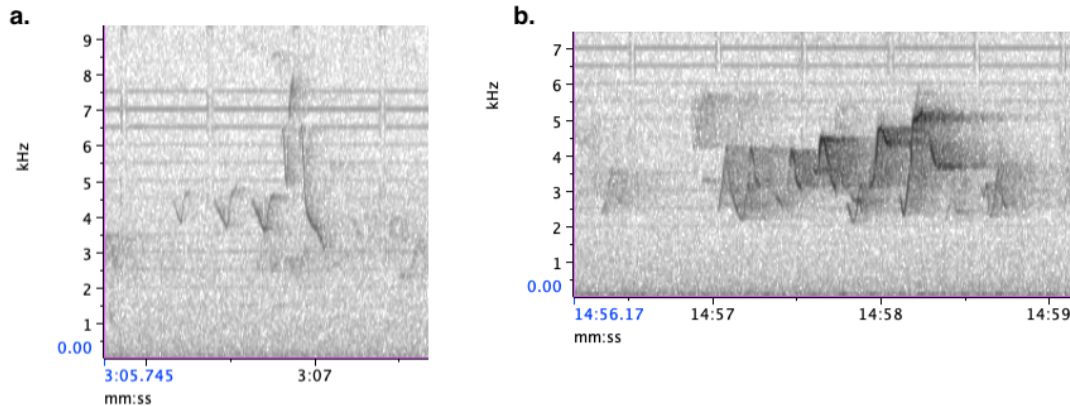


Figure 2: Hooded Warbler Singing Patterns from Field Data. (a) depicts a repeat mode song type (b) represents one example of a mixed mode song type.

Males sing all day, but the highest rates of singing occur in the early morning (dawn) and decrease throughout the day [23]. Females generally build their nests in locations on a male's territory that are farthest away from boundaries with other neighbors. Males tend to sing on perches and often high in the canopy throughout the territory. Unmated males spend 80-100% of their time singing, depending on if a neighbor is singing [26].

2.2.2 Grasshopper Sparrow

Grasshopper Sparrows often inhabit grasslands with mixed vegetation and bare ground, specifically tallgrass and shortgrass prairie in the East and Midwest United States [27]. They generally settle on breeding grounds in mid-April. Males arrive three to five days before females, and pair formation commences soon after females arrive [28]. Nest building lasts two to three days and eggs are laid on consecutive days

after nest completion [29]. In a study conducted in Iowa, incubation typically lasted for eleven days and the hatchling phase lasted nine days before fledging [30]. However, at our study site, the Grasshopper Sparrows fledge around five to six days after incubation [31]. Grasshopper Sparrows commonly produce two broods. Typically, the first brood occurs in late May and the second occurs in early July [28].

In our study, brood parasitism by Brown-headed Cowbirds (*Molothrus ater*) affected 53% (9/17) of the nests in the ARU-associated nests, resulting in a failure rate of 78% (7/9) for those nests. In total, 82% (9/17) of nests associated with ARUs failed (Table. 1). Mammal predation and trampling by livestock were other common causes of failure. After nest failure, Grasshopper Sparrows will frequently attempt to re-nest multiple times throughout the breeding season [28].

Grasshopper Sparrow males sing throughout the mating and nesting season. They have three distinct songs: a primary song, a sustained song, and a trill. Females trill and call, but do not sing the primary or sustained song [32]. For the purposes of this study, we have focused only on the primary and sustained song. The primary song is mostly territorial, while the sustained song is more related to attracting a mate and maintaining a pair bond [32]. The primary song is a 1-2 second buzzing preceded by two to three staccato notes (Figure. 3a). The sustained song is a series of buzzy notes that vary in pitch, which lasts for 5 to 15 seconds (Figure. 3b) [32]. Often, the sustained song is preceded by the primary song, which might be used for the dual purpose of territorial defense and mate attraction [29]. Males commonly sing from fixed perches or on the ground, usually at the edge of the territory [29].

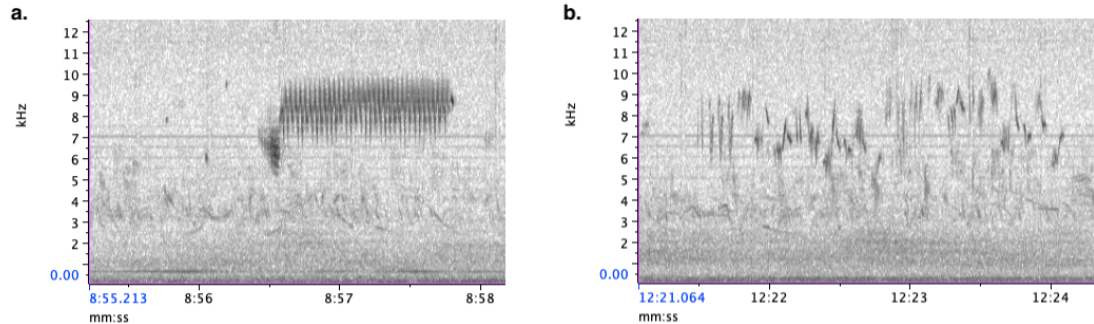


Figure 3: Grasshopper Sparrow Singing Patterns from Field Data. (a) depicts the primary song type (b) represents the sustained song type.

When males first arrive on the breeding grounds, they almost exclusively sing the primary song in the morning, and do not sing later in the day. As more males arrive, they sing the primary song more frequently throughout the day [32]. Within approximately two weeks of arriving, the males start singing the sustained song (usually in the evening) mixed with the primary song [32]. When territories are established, the males start removing the primary song introduction from the sustained song [32]. In general, the number of sustained songs increases when birds pair and the number of primary songs remains mostly unchanged [29]. After mating, time spent singing decreases, but does not stop entirely. During egg-laying and incubation, the male sings both songs frequently throughout the day [32]. This increase in singing during those phases could suggest that the male is seeking extra-pair copulation [29]. During the hatchling care phase, when feeding the young, Grasshopper Sparrows tend to sing less [32].

2.2.3 Gray Vireo

Gray Vireos are a short-distance migratory species that inhabit pinyon pine-Utah juniper stands, one-seed juniper savannas, mixed juniper-oak (*Quercus* spp.) woodlands, and desert riparian communities in New Mexico [33]. They have been identified as a species of conservation concern since 1978 and are one of the least-studied songbirds in North America [34].

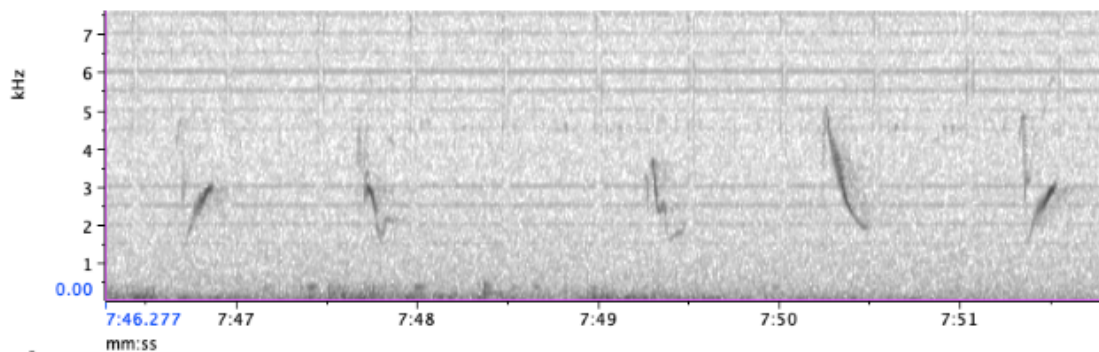
Gray Vireos typically arrive on the breeding grounds in late April through early May and form pairs within one to two days after females arrive [35, 36]. After pair formation, nest building begins, which generally lasts five to six days [37]. In a study collected from 32 nesting records in New Mexico, the median date of nest construction for Gray Vireos was May 19th [33]. Eggs are laid on consecutive days after the nest is complete [38]. The incubation period lasts around 11-14 days and the nestling period is typically 12 days, but ranges from 9-15 days, after which the young fledge [33, 35, 37, 39]. After fledging, the young chicks are still dependent on adult care [40]. Gray Vireos may make up to six nesting attempts in the season, however it is unclear if second brooding occurs at our study site [33–35].

In our study, brood parasitism by Brown-headed Cowbirds affected 24% (5/21) of the ARU-associated nests, causing failure for each parasitized nest. In total, 71% (15/21) of nests resulted in failure (Table. 1). Most of these failures can be attributed to nests never being active (ie. the nest was built, but no eggs were laid). Others failed due to predation or brood parasitism. After a nest failure, Gray Vireos will abandon the current nest, and re-nest in a new location [36].

Gray Vireos sing when establishing and maintaining their territory, while attracting a mate, during pair formation, and when informing a female of the male's

location [37]. Female Gray Vireos also sing, generally when females and males are both at the nest. Gray Vireos sing in distinct syllables, which is the smallest unit of sound produced by a bird [41]. The syllables consist of sounds that modulate up and down [42]. Gray Vireos have a primary song (Figure. 4a), as well as a complex song (Figure. 3b). The song length of the Gray Vireo is 0.18-0.35 seconds, and they sing at a rate of 24-41 songs per minute [43].

a.



b.

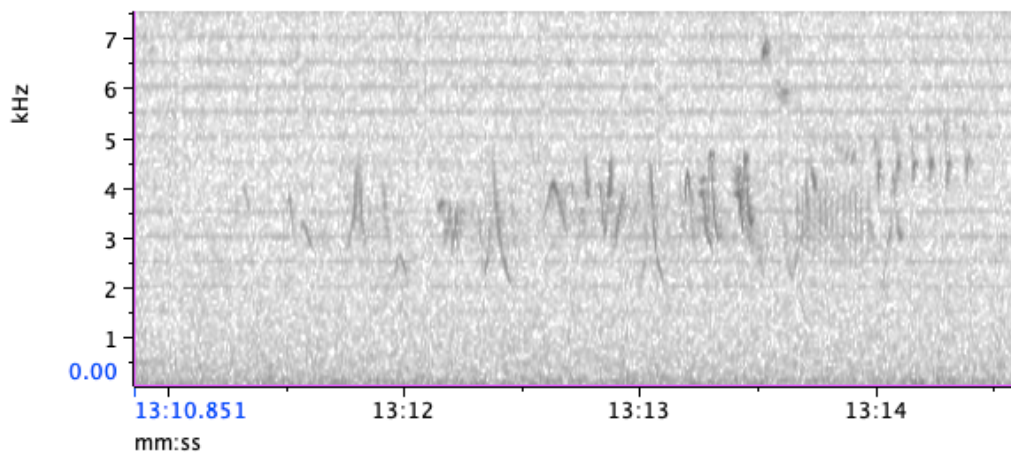


Figure 4: Gray Vireo Singing Patterns from Field Data. (a) depicts the primary song type (b) represents the complex song type.

Males sing throughout the day at all stages of the nesting cycle [34]. Males sing to females during nest construction, and the female will sing in response to the male at the nest. During incubation, both male and female Gray Vireos will sing from the nest [37]. Males tend to sing early in the morning, and generally stop by mid-day. They may resume singing in the late afternoon, but at a slower pace. This daily pattern seems to be consistent throughout the breeding season [37]. Unmated males will sing throughout the day, and at a faster rate than mated males [37].

Table 1: Study Site Nest Fates and Clutch Size. For each species, researchers in the area collected daily or weekly meta-data on the nests associated with ARUs. This table outlines the aggregate fates and clutch sizes of the ARU-associated nests for each species.

	Hooded Warbler	Grasshopper Sparrow	Gray Vireo
ARU-associated Nests	42	17	21
Avg. Clutch Size	2.65 ± 1.03	2.92 ± 0.79	2.54 ± 1.05
Brood Parasitized Nests	4	9	5
Failed Nests	25	14	15

2.3 Identification of Vocalizing Species

In order to study song activity, we needed to first identify when our bird of interest was vocalizing in a given audio recording. Using Opensoundscape, we developed a data processing pipeline to identify each species from ARU data. We utilized this pipeline to build classifiers for our three focal species.

2.3.1 Training Data

To make our classifiers easier to train, more robust, and useful from year to year or in different study sites, we chose to train them mostly on xeno-canto recordings. Field recordings are long, noisy, and may only contain sparse bird vocalizations, which makes annotation quite difficult and time consuming. Using xeno-canto recordings (for which the bird of interest has already been identified as present in the recordings) greatly accelerates the process of annotating and allows for researchers that do not have considerable experience with the vocalizations of the birds of interest to annotate more effectively. For this analysis, we annotated recordings for both calls and songs, but only trained the classifiers using song vocalization annotations.

We trained our model with data that included a mix of recordings in which the bird of interest was singing (positives) and recordings in which the bird of interest was not singing (negatives). For the positives, we used xeno-canto recordings for which the bird of interest had already been identified as present. We annotated the first 60 seconds of each xeno-canto recording that included our bird of interest to reduce the time spent annotating and get a broader diversity of songs. For the negatives, we used a mix of field data recordings from randomly placed ARUs and xeno-canto recordings. To approximate the vocalizations from birds that might be heard on our audio recordings, we chose three to four recordings of four to five birds that were common in the study site of the bird of interest. For all of the negative recordings, we excluded sections that contained any songs from our bird of interest. We generated a training dataset that we used as input for our classifier by annotating audio clips visualized as spectrograms from the xeno-canto recordings and our field data.

2.3.2 Annotation Process

For each spectrogram, we annotated the portions of the recording in which the bird of interest was vocalizing using RavenLite, a free software for viewing and manipulating spectrograms (ravensoundsoftware.com) [44]. Generally, we categorized bird vocalizations as a call or song. However, some birds have common variations in their songs or calls that make these vocalizations distinct from the other calls or songs. We annotated vocalizations separately, depending on the commonality and distinctiveness of a vocalization pattern. For each bird, we annotated the song as their alpha code followed by either “song” or “call.” For Grasshopper Sparrows, we included a “complex_song” category as well. RavenLite has a point and drag interface that we used to box and label the vocalizations. RavenLite automatically converts these boxes into table entries with information on the start and end times of the vocalizations (Figure. 5). We annotated for both song and call, but only included the songs in our analysis.

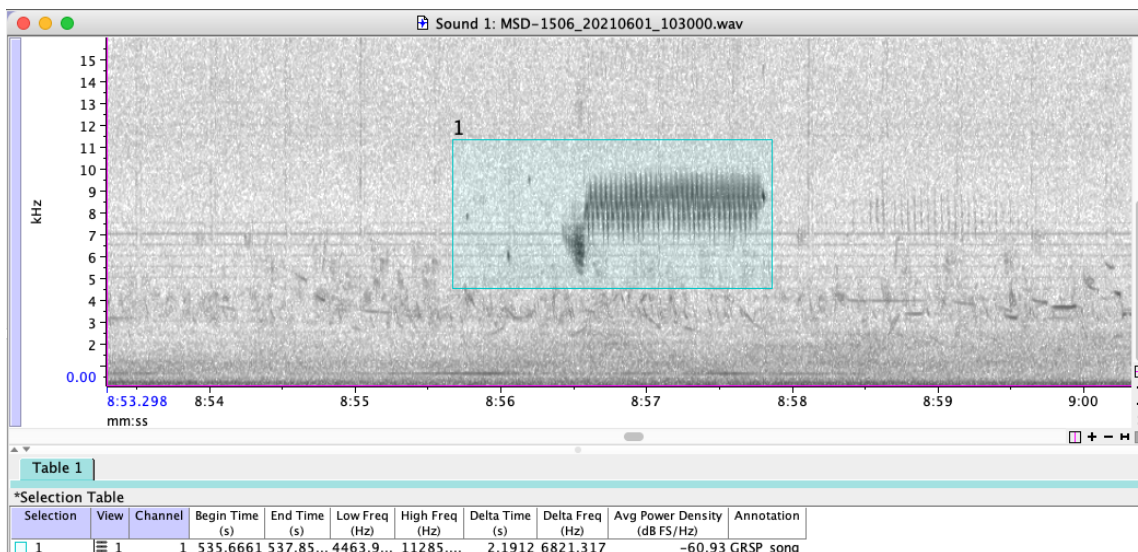


Figure 5: Annotation of Grasshopper Sparrow Song in RavenLite. The y-axis represents frequency in hertz (pitch) and the x-axis represents time. The blue boxed region surrounds the Grasshopper Sparrow song (three introductory notes followed by a series of buzzy notes), which has a duration of ~ 2.2 seconds beginning at second 535.7 (minute 8:55.7) and ending at second 537.9 (minute 8:57.9). The blue box corresponds to the Table entry under Selection 1.

To analyze the data, we used Python 3 with a virtual environment from the Opensoundscape repository (opensoundscape.org) in a Python 3 kernel [22]. We split the recordings into 3 second clips, which we labeled as positive or negative, depending on the functions and parameters below. We counted a positive 3-second clip as one song by the bird of interest. We chose a clip length of 3 seconds so that each clip was large enough to contain a song, but small enough to not contain multiple songs. To generate sets of clips from audio recordings and the corresponding “one-hot encoded labels,” we adapted the code from the Opensoundscape documentation

to work with our data.

We paired the recordings with their corresponding annotation file from Raven-Lite. We split the audio into 3-second clips using the *audio.split_and_save* function. Then, we generated labels using the *one_hot_labels_like* function with the parameters: *min_label_overlap* and *min_label_fraction* set to 3 and 0.5, respectively. For this analysis, we only used the song labels as the classes for each species, including the complex song label for Grasshopper Sparrows. We aggregated the one-hot encoded labels for each class into presence/absence labels that we used for model training.

2.3.3 Model Training Process and Predictions

We created our model training dataset by combining the 3-second clips labeled as present from the positive training data and the 3-second clips labeled as absent from the negative training data. Our negatives consisted of the xeno-canto bird negatives and the field negatives (~ 15 min), with an 80:20 split. We then resampled the positives and negatives to have an equal number of each (300 positives and 300 negatives for a total of 600 clips for each species).

We augmented our model with field data from randomly placed ARUs that excluded songs from our bird of interest (~ 60 min). We refer to this data as the “overlay data,” because the overlay data and training data form a composite image that is X% overlay image and Y% training image. Overlay augmentation transforms the pixel values of the images using a weighted average (Figure. 6). We use this augmentation to make the training data representative of field data, which may have songs that are quieter or oriented differently than the training data. This step is especially important when training with xeno-canto data, since xeno-canto recordings

come from a variety of microphones and audio equipment and are sourced from a variety of locations where the soundscape may not be the same as the soundscape in the field data of interest. Additionally, they are often “targeted” by the recording equipment, which reduces background noise and makes the target species’ songs louder.

For each species, we first split the (~ 600) annotated data samples into training and validation, with an 80:20 split. We preprocessed the training data with the Opensoundscape CnnPreprocessor class, which loads audio files, creates spectrograms, applies specified augmentations, and returns a pytorch tensor. We set the *overlay_weight* parameter to 0.7; thus, the resulting training samples had a 70% contribution from the overlay samples and a 30% contribution from the original training samples. We used these augmented tensors to train our model.

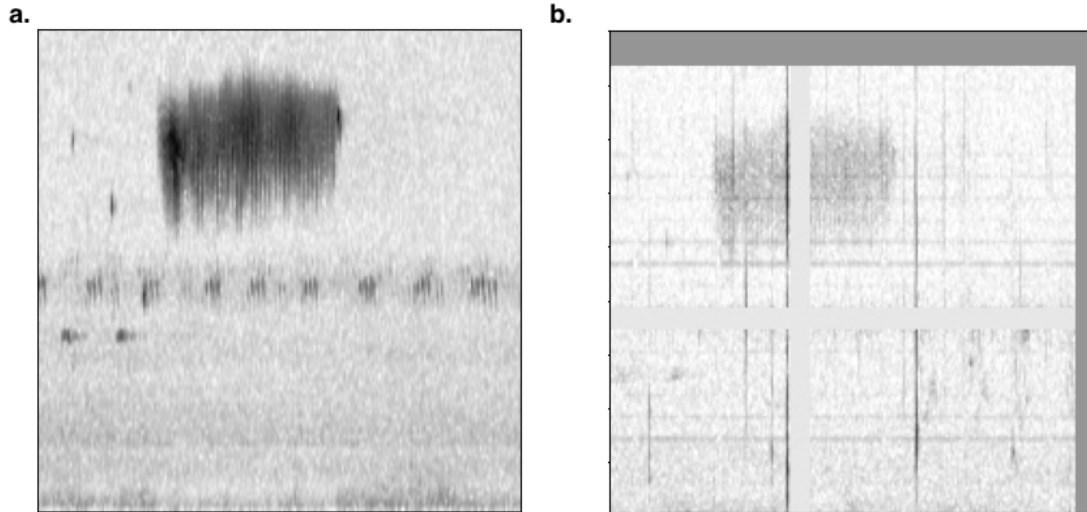


Figure 6: Spectrogram Before and After Augmentation. (a) 3-second clip of a Grasshopper Sparrow song tensor (b) 3 second clip of a Grasshopper Sparrow song tensor after augmentation with 70% contribution from the overlay sample.

For each species-specific model, we used a single-target Pytorch ResNet-18 model with a softmax activation layer. We used pre-trained weights (image-net.org) and the default learning rate, learning schedule, and regularization weight decay parameters. We trained the model with 100 epochs and a batch size of 64. The model assigned each 3-second clip a classification confidence score, with scores closer to 1 meaning greater confidence that the clip contains a song from the bird of interest. The best model was chosen based on the F1 score calculated at each epoch from the training validation data. The training took approximately 20 minutes per species.

The final step of this process was to use our best trained model to identify if the bird of interest was vocalizing in recordings from our field data. For Hooded

Warblers, we predicted on $\sim 3,276$ hours of field recordings. For Gray Vireos, we predicted on $\sim 2,556$ hours of field recordings. For Grasshopper Sparrows, we predicted on $\sim 1,600$ hours of field recordings. For each species, we split these recordings into 3-second clips to predict the presence or absence of the bird of interest in these clips. Prediction took approximately 4 hours per species.

2.4 Application to Phenology Parameters

For each species, we derived estimates of the time spent in each phenological nesting stage from the literature and the researchers familiar with the species at each site (Table. 2). We combined these estimates with the meta-data collected at each study site to generate a nesting phenology timeline for each of the ARU-associated nests at the site. The date of first egg, the fledge/fail date, and the clutch size were identified from the meta-data, while the first day of nest building, the start of incubation, and the hatch date were approximated.

We estimated the nest construction start date by subtracting the estimated nest construction time from the date of the first egg. We estimated the incubation date by adding the number of egg laying days (eggs laid per day * clutch size) to the date of the first egg. We estimated the start of the hatchling care phase by adding the incubation time to the calculated incubation date. We determined the location, phenophase, phenoday, and day of the season for each field data recording by parsing the time, day, and SD memory card number from the filename and aligning those values to the nesting phenology timeline for each nest.

Table 2: Duration of Phenological Stage Parameters for Each Species.

These nesting phenology phase durations were estimated from the literature or from researchers familiar with the focal species at their study site. All values are in days, unless otherwise noted.

	Hooded Warbler	Grasshopper Sparrow	Gray Vireo
Nest Building	5	3	5
Hatching	1	1	1
Laying	1 egg per day	1 egg per day	1 egg per day
Incubation	12	11	14
Hatchling Care	9	6	12

To explore the relationship between the singing activity and nesting phenology, we first needed to assign a threshold confidence value, above which a clip would be considered to have a song from the bird of interest present. We annotated field data and randomly selected 500 clips, with different ratios of positive and negative clips. We generated ten datasets from 1 hour of recordings, with two of each ratio type. The ratios were 400:100, 300:200, 250:250, 200:300, and 100:400. We chose these different ratios to explore how the classifiers performed with different levels of positive and negative clips and approximate potential splits in the full datasets. An optimal threshold was calculated based on the highest average F1 score across the reviewed subset of the field data with different splits of positive and negative clips. The average precision and average recall associated with the optimal threshold are shown in (Table. 3).

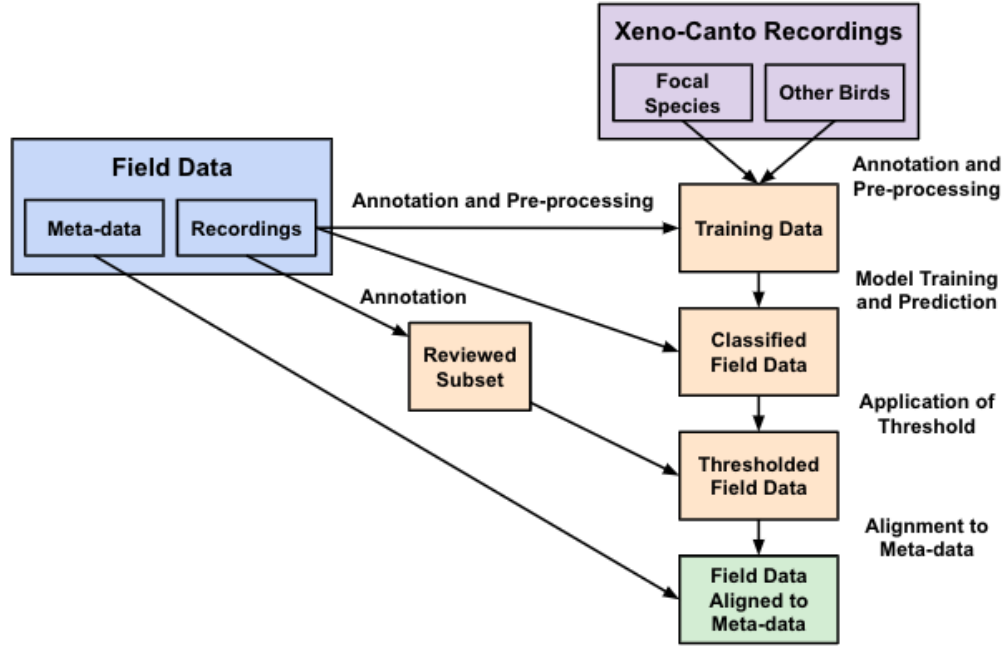


Figure 7: Data Processing Pipeline. Meta-data on phenology and ARU recordings are collected in the field (Section. 2.1). A few field data recordings, along with xeno-canto recordings of focal species and a few common birds in the area are annotated, pre-processed, and combined to constitute the training data (Section. 2.3.2). The classification model is trained on the training data, and the best model is selected to predict focal species identification on the field data recordings. An optimal threshold is calculated using a reviewed subset of the field data with different splits of positive and negative clips. The classified data is then thresholded and aggregated from 3-second clips into full recordings (Section. 2.3.3). For each recording, the singing rate and song occurrence is calculated. Next, each recording is aligned to a nest location and phenophase or to a randomly placed ARU. Then, phenological analysis begins (Section. 2.4). The inputs to the data processing pipeline are in blue and purple, the intermediate results are in orange, and the output is in green.

After selecting a threshold, we employed a hierarchical process to investigate the relationship between singing activity and nesting phenology. First, we binarized our singing rate data to consider the relationship between any singing activity and phenological stage by looking at the fraction of recordings with at least one song. We refer to this fraction of recordings as the song occurrence. This measure provides us with some insight into the territoriality of the birds and whether they are singing near the nest. Then, we removed the zero frequencies to examine the relationship between singing rate, given that the birds are singing, and phenological stage. We compared the singing activity between phenophases by running analysis of variance (ANOVA) on the binarized data and singing rates grouped by phenophase and performed Tukey HSD on the output of the ANOVA. We explored the effect of daily time differences on these relationships. Lastly, we compared the across-nesting cycle fluctuations in singing activity for data collected at nests to the across-season fluctuations in singing activity for data collected from randomly placed ARUs.

Results

3.1 Classifier Performance

For the Hooded Warbler and Grasshopper Sparrow, the recall from the proposed classifiers was consistently high across thresholds (0.0-0.99). The recall for the BirdNET classifier dropped significantly as the threshold values increased. The recall did not seem to vary with positive and negative splits. In contrast, for Gray Vireos, the recall of the proposed classifier was higher than the BirdNET classifier, but not significantly (Figure. 8).

For the Hooded Warbler and Grasshopper Sparrow, the BirdNET classifier had a much higher precision at splits with more negative clips, but as the number of positive clips increased, the difference between the BirdNet classifier and proposed classifier became minimal. Both the BirdNET and proposed classifiers for Hooded Warblers and Grasshopper Sparrows had better precision as the number of positive clips increased. For Gray Vireos, the BirdNET classifier had a significantly higher precision at all threshold levels than the proposed classifier. The Hooded Warbler and Grasshopper Sparrows performed similarly across thresholds for all splits, whereas the proposed Gray Vireo classifier increased precision dramatically at higher thresholds for all splits (Figure. 9).

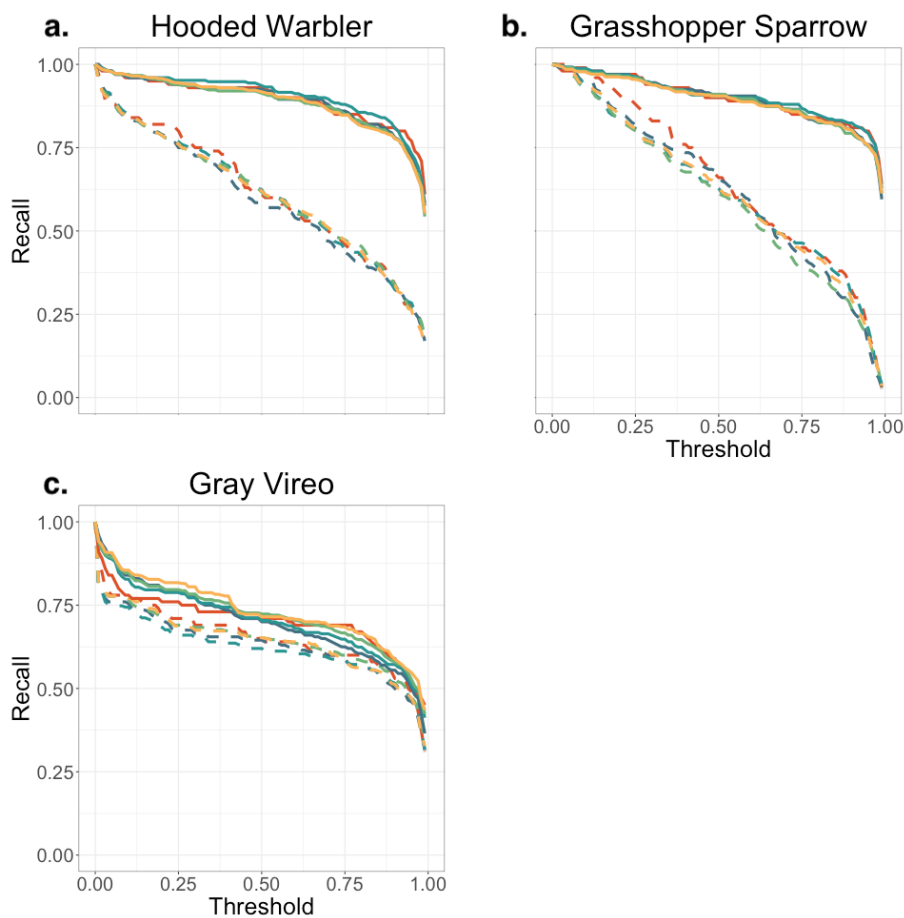


Figure 8: Recall vs Threshold. Recall or true positive rate is calculated as $[\text{True Positives}]/[\text{True Positives} + \text{False Negatives}]$, considering only clips with confidence \geq threshold. Recall represents the proportion of songs from the bird of interest in the dataset that were detected by the classifiers out of the total amount of songs from the bird of interest in the dataset. Each plot corresponds to a species. The solid line represents the proposed classifier and the dotted line represents the BirdNET classifier. The color of the line corresponds to the ratio of negative and positive clips for that dataset. Red=400:100, dark blue=300:200, light blue=250:250, green=200:300, yellow=400:100.

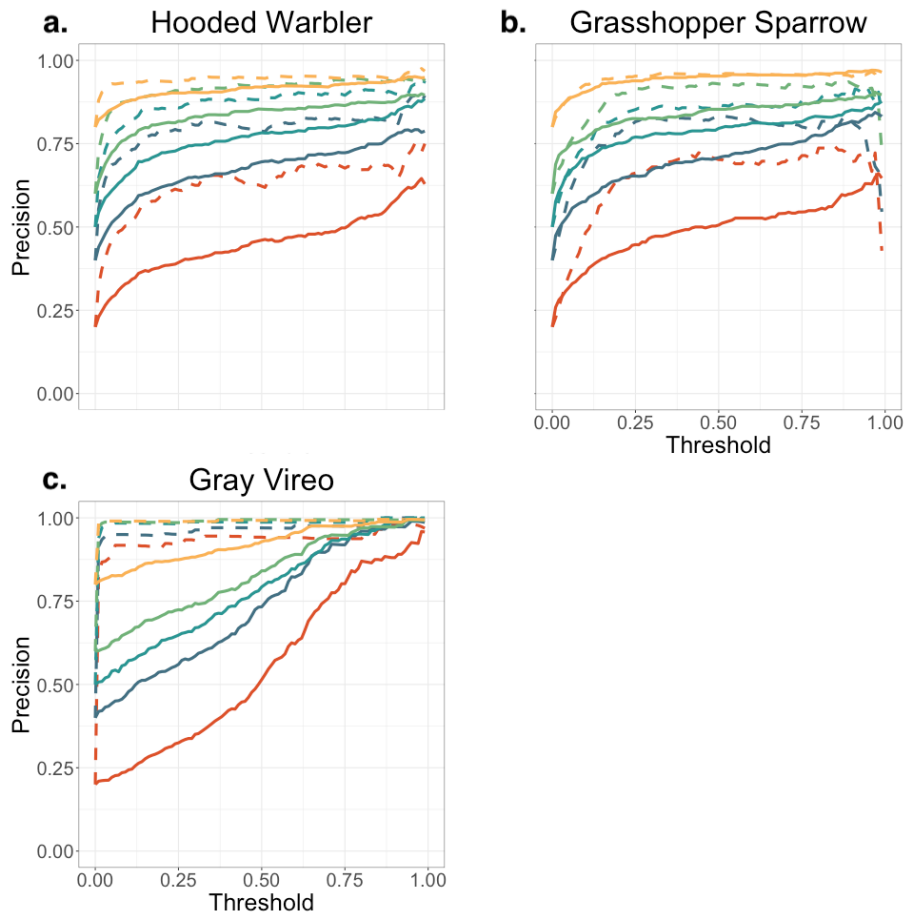


Figure 9: Precision vs Threshold. Precision or True Positive Rate is calculated as $[\text{True Positives}]/[\text{True Positives} + \text{False Positives}]$, considering only clips with $\text{Confidence} \geq \text{Threshold}$. Precision represents the proportion of apparent detections that correspond to true target species calls. Each plot corresponds to a species. The solid line represents the proposed classifier and the dotted line represents the BirdNET classifier. The color of the line corresponds to the ratio of negative and positive clips for that dataset. Red=400:100, dark blue=300:200, light blue=250:250, green=200:300, yellow=400:100.

The area under the precision-recall curve (AUC) serves as a general measure of model performance; a classifier with perfect precision and perfect recall would have an $AUC = 1$. The precision recall curves and AUC values were calculated using the PRROC package in R, which interpolates values across the full range of threshold values (Figure. 10). The average precision-recall AUC values by proposed classifier ranged from 0.64 to 0.96 for Hooded Warbler, from 0.59 to 0.93 for Grasshopper Sparrow, and from 0.71 to 0.94 for Gray Vireo (Table. 3).

The F1 scores were consistently high for the proposed Hooded Warbler and Grasshopper Sparrow classifier at all thresholds, while the F1 scores for the BirdNET classifiers decreased at higher thresholds, due to the poor recall. At lower thresholds, the proposed Gray Vireo classifier had much lower F1 scores than the BirdNET classifier, but at high thresholds they had similar values (Figure. 11).

The receiver operating characteristic (ROC) curve demonstrated that the BirdNET classifiers and proposed classifiers had generally good performance on distinguishing true positives (songs from the species of interest) from false positives (noise). The ROC curve balances the true positive rate against the false positive rate. The true positive rate is calculated as $[\text{True Positives}]/[\text{True Positives} + \text{False Negatives}]$. The false positive rate is calculated as $[\text{False Positives}]/[\text{False Positives} + \text{True Negatives}]$. The area under the (ROC) curve corresponds to the probability that the classifier will assign a higher score to a randomly chosen true positive than to a randomly chosen true negative. The proposed Hooded Warbler and Grasshopper Sparrow classifiers mean ROC AUC values greater than 0.88, whereas the mean ROC AUC for Gray Vireos was 0.81 (Figure. 12). The average ROC AUC, precision-recall AUC, average maximum F1 scores, optimal threshold values, and average precision and recall scores associated with the optimal threshold are listed in Table 3.

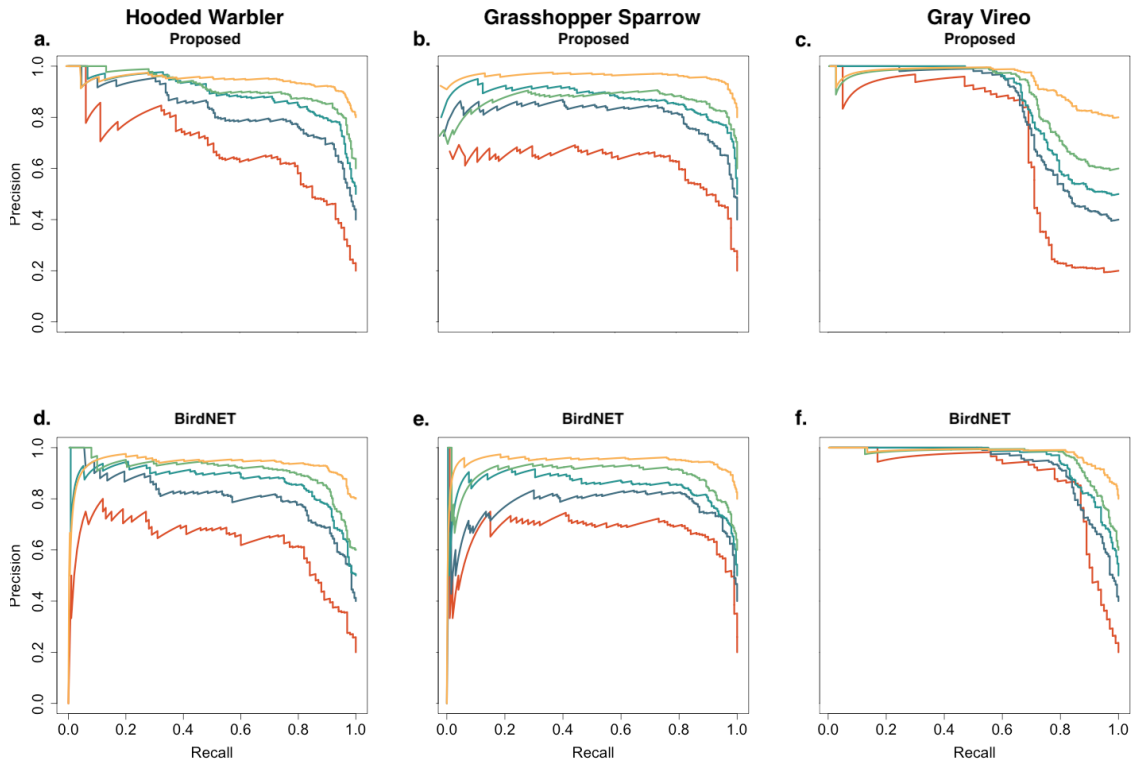


Figure 10: Precision Recall Curve. The precision-recall curve illustrates the tradeoff between sensitivity (recall) and specificity (precision). For each species, the top plot corresponds to the precision recall curve for the proposed classifiers and the bottom plot corresponds to the precision recall curve for the BirdNET classifiers. The color of the line corresponds to the ratio of negative and positive clips for that dataset. Red=400:100, dark blue=300:200, light blue=250:250, green=200:300, yellow=400:100. The precision-recall curves were generated using the PRROC package in R, which interpolates values across the full range of threshold values.

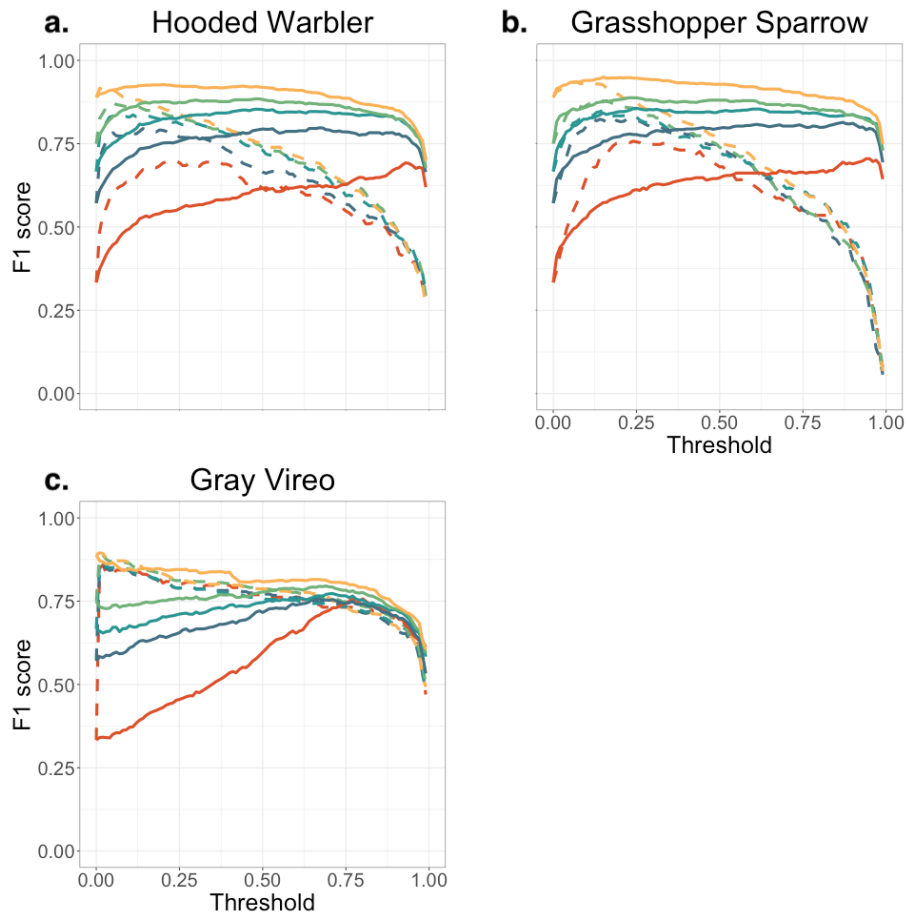


Figure 11: F1 Scores Curve. F1 score is interpreted as a balanced measure of overall classifier performance combining precision and recall. The F1 score is calculated as $2 * [\text{Precision} * \text{Recall}] / [\text{Precision} + \text{Recall}]$, considering only clips with $\text{Confidence} \geq \text{Threshold}$. The highest point on each curve represents the threshold at which the model gives the best overall performance for each species if we consider precision and recall to be equally important. Each plot corresponds to a species. The solid line represents the proposed classifier and the dotted line represents the BirdNET classifier. The color of the line corresponds to the ratio of negative and positive clips for that dataset. Red=400:100, dark blue=300:200, light blue=250:250, green=200:300, yellow=400:100.

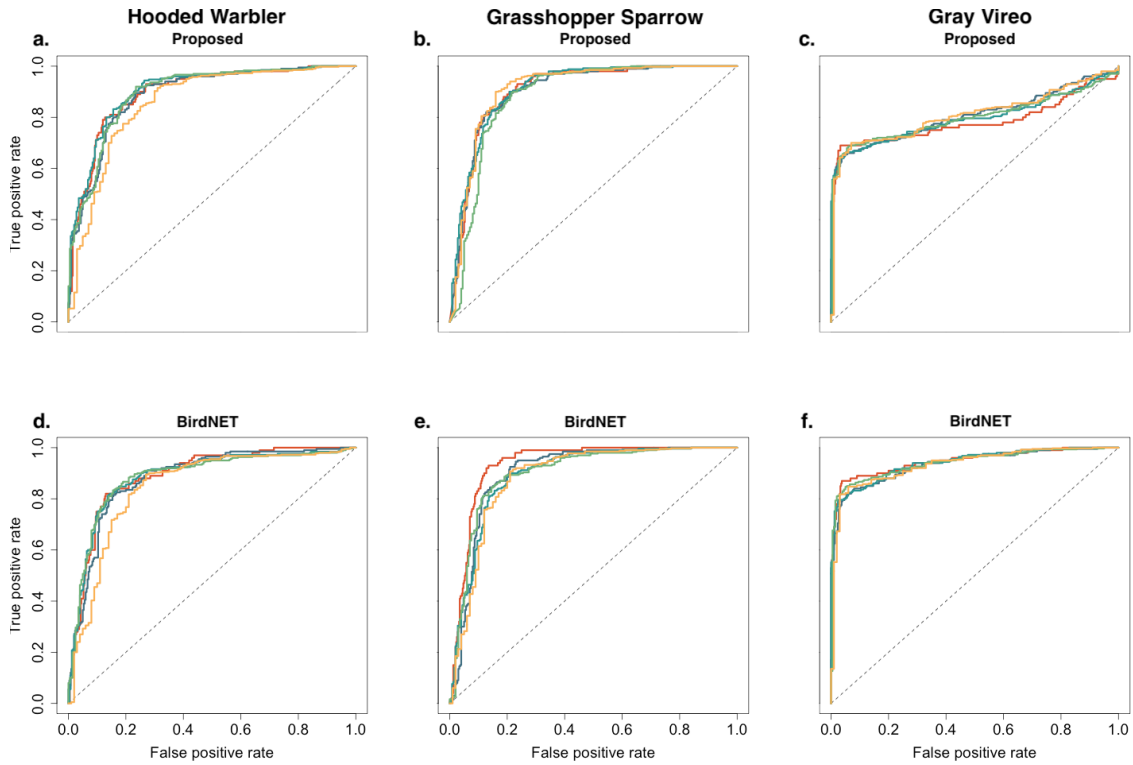


Figure 12: Receiver Operating Curve. The ROC curve plots the true positive rate against the false positive rate. The diagonal line represents a classifier with no prediction value. The top plot corresponds to the proposed classifier for a species and the bottom plot corresponds to BirdNET classifier for each species. The color of the line corresponds to the ratio of negative and positive clips for that dataset. Red=400:100, dark blue=300:200, light blue=250:250, green=200:300, yellow=400:100. The ROC curves were generated using the PRROC package in R, which interpolates values across the full range of threshold values.

Table 3: Optimal Threshold Values and Classifier Accuracy Metrics. All values were calculated with the reviewed subset of the field data for each species with all positive:negative splits. The average AUC for the ROC and Precision Recall (PR) curve were generated using the PRROC package in R, which interpolates values across the full range of threshold values. The average maximum F1 score was calculated across all splits. The optimal threshold value was selected based on the highest associated average F1 score across splits. The average precision and average recall are associated with the optimal threshold, and were calculated across splits.

	Hooded Warbler		Grasshopper Sparrow		Gray Vireo	
	Proposed	BirdNET	Proposed	BirdNET	Proposed	BirdNET
Max F1 Score	0.83 ± 0.09	0.83 ± 0.08	0.84 ± 0.08	0.83 ± 0.08	0.79 ± 0.05	0.87 ± 0.01
ROC AUC	0.89 ± 0.02	0.88 ± 0.02	0.89 ± 0.02	0.89 ± 0.02	0.81 ± 0.01	0.94 ± 0.01
PR AUC	0.85 ± 0.11	0.83 ± 0.12	0.79 ± 0.11	0.80 ± 0.11	0.85 ± 0.07	0.94 ± 0.04
Threshold	0.48	0.14	0.60	0.15	0.71	0.02
Precision	0.74 ± 0.17	0.81 ± 0.14	0.78 ± 0.14	0.77 ± 0.14	0.90 ± 0.08	0.96 ± 0.04
Recall	0.93 ± 0.01	0.82 ± 0.02	0.90 ± 0.01	0.90 ± 0.03	0.67 ± 0.02	0.81 ± 0.02

3.2 Results of Singing Activity and Phenology Review

To investigate singing activity and phenology, we chose to use the proposed Hooded Warbler and Grasshopper Sparrow classifiers, because they performed better than or equal to the BirdNET classifier on 4/5 accuracy metrics. Additionally, for both the Hooded Warbler and Grasshopper Sparrow recall curves, the BirdNET classifier recall for all positive:negative splits rapidly declined as threshold values increased,

which caused us to question the effectiveness of the BirdNET classifier's recall abilities, even at lower thresholds. We chose to use the BirdNET classifier for Gray Vireos, since it outperformed our classifier on every accuracy metric.

For all three species, there was a significant decrease in song occurrence after fledging compared to most stages during the nesting cycle. The singing rate also decreased after fledging for Hooded Warblers and Gray Vireos. However, the rate did not decrease for Grasshopper Sparrows. For all three species there was no significant variation in song occurrence or singing rate between phases in the nesting cycle before fledging, except for a significant increase in singing rate during the nest construction phase for the Gray Vireos and a significant decrease in singing rate for Hooded Warblers while caring for hatchlings (Figure. 13, 14).

For Gray Vireos, after nest failure, song occurrence and singing rate was significantly lower than song occurrence and singing rate during most phases of the nesting cycle before fledging. For Grasshopper Sparrows, song occurrence after nest failure was significantly lower, but singing activity was not. For Hooded Warblers, we did not observe a significant decrease in song occurrence or singing rate after nest failure or at randomly placed ARUs. For Grasshopper Sparrows, the song occurrence at randomly placed ARUs was significantly lower than song occurrence during most stages of the nesting cycle before fledging, but the singing rate was not. For Gray Vireos, the song occurrence and singing rate at randomly placed ARUs was significantly lower than the song occurrence during most stages of the nesting cycle before fledging. Overall, the singing rate for Grasshopper Sparrows was much higher than the singing rate of Hooded Warblers and Gray Vireos (Figure. 13, 14).

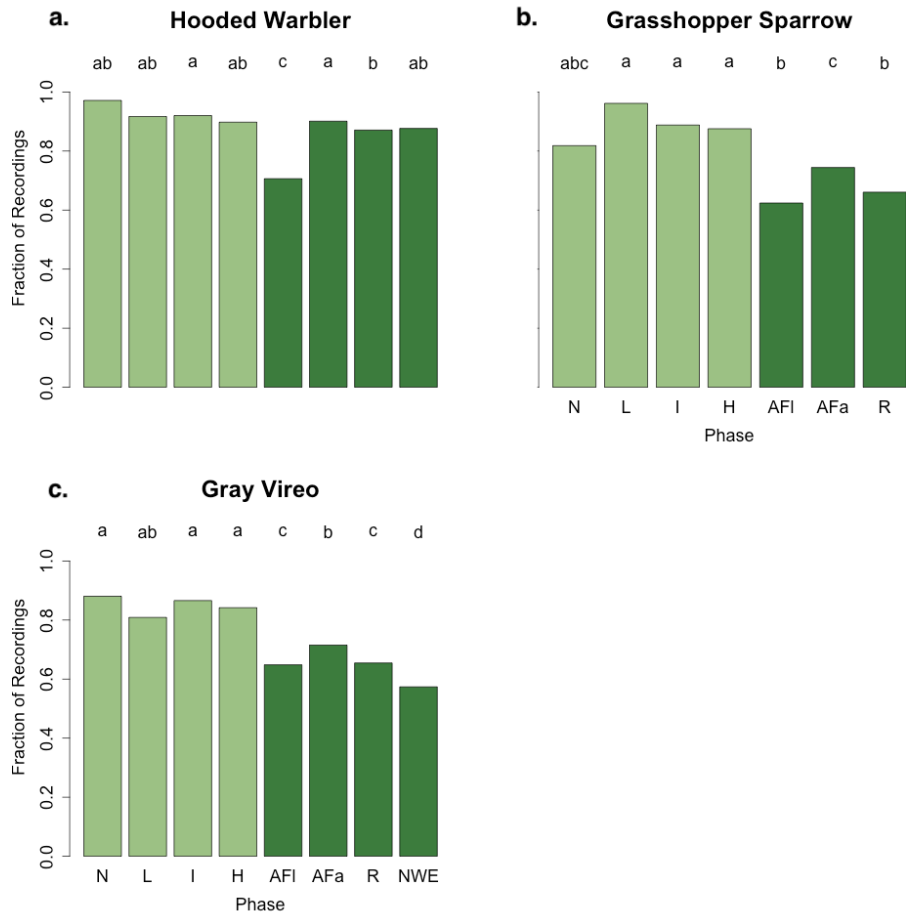


Figure 13: Pairwise Comparisons of the Song Occurrence Between Phenophases. The height of each bar represents the proportion of recordings that contain songs occurring in a certain phenophase (x-axis). The label at the top of each bar is a pair-wise classification of significance. If two phases are associated with the same label letter, there is not a significant difference in song occurrence between those phases. The stages are: N = Nest Building, L = Egg Laying, I = Incubation, H = Caring for Hatchlings, AFI = After Fledge, AFa = After Fail, R = Randomly placed ARU recordings, NWE = Nest without Eggs (Never active build). The significance values and labels were generated using ANOVA and Tukey HSD.

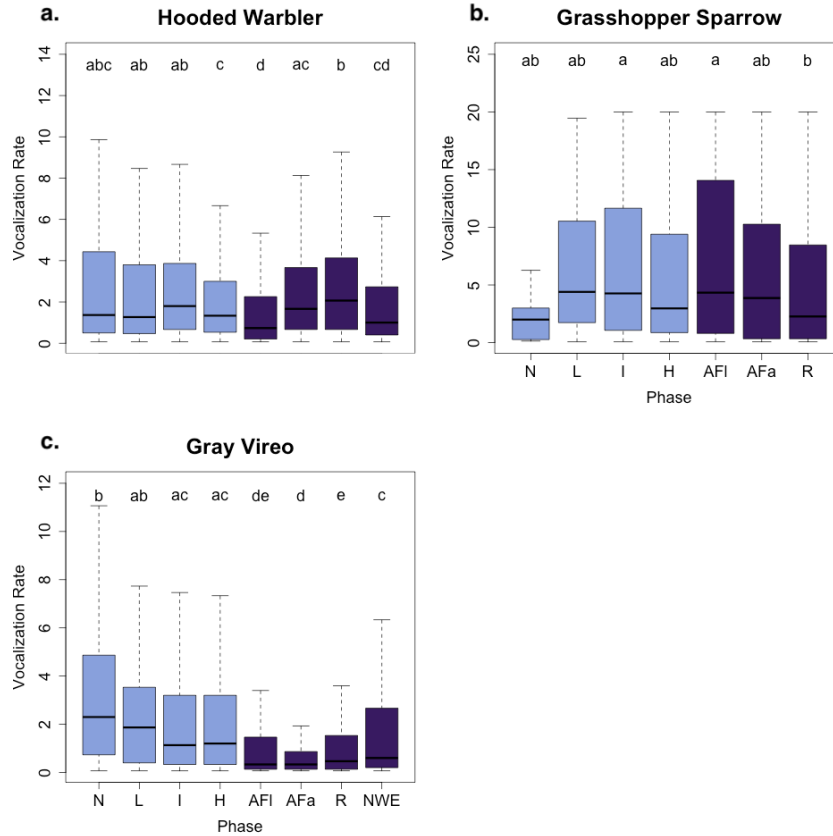


Figure 14: Pairwise Comparisons of the Average Singing Rate Between Phenophases. Each boxplot represents the distribution of singing rates (per minute) from recordings taken during a certain phase (outliers were removed for clarity). The label at the top of each boxplot is a pair-wise classification of significance. If two phases are associated with the same label letter, there is not a significant difference in singing rate between those phases. The stages are: N = Nest Building, L = Egg Laying, I = Incubation, H = Caring for Hatchlings, AFI = After Fledge, AFa = After Fail, R = Randomly placed ARU recordings, NWE = Nest without Eggs (Never active build). The significance values and labels were generated using ANOVA and Tukey HSD.

In terms of daily variation in singing, we observed that Hooded Warblers had a significantly higher song occurrence between 6:30 - 9:30 AM and a lower song occurrence at 5:30 AM, and after 10:30 AM. Although the birds had a lower song occurrence at 5:30 AM, the singing rate was higher at 5:30 AM and decreased throughout the day. We observed that Grasshopper Sparrows had a higher song occurrence at 5:30 AM and 10:30 AM, but from 6:30 AM-9:30 AM the song occurrence was lower and not significantly distinct within those hours. However at 11:30 AM, there was significantly lower song occurrence. The singing rate was much lower at 6:30 and 7:30 AM compared to 5:30 and 8:30 - 11:30 AM. We observed that Gray Vireos sang the most from the hours of 6:30 AM to 9:30 AM. The song occurrence was lower at 5:30 AM and after 10:30 AM. Their singing rate was lowest at 5:30 AM, and mostly consistent throughout the day when they were singing.

With regard to daily variation in singing across the nesting cycle, we observed a high song occurrence for every species throughout the nesting cycle. After fledging, the song occurrence started to decrease for Hooded Warblers. For Grasshopper Sparrows and Gray Vireos, there was not enough data from nests to conclusively state results. During the nesting cycle, we observed a singing rate around three songs per minute for Hooded Warblers. For Grasshopper Sparrows, we observed a singing rate of about four to seven songs per minute. Lastly, for Gray Vireos, we observed a singing rate around one to three songs per minute (Figure. 15, 16).

At randomly placed ARUs, recording the soundscape of the study site throughout the season, there was a general decline in song occurrence throughout the season for Hooded Warblers and Gray Vireos (especially at the end of the season). We did not observe this for Grasshopper Sparrows, which had consistently high song occurrence at randomly placed ARUs throughout the season and then a dramatic termination

of song occurrence in late July. However, in terms of singing rate, both Hooded Warblers and Gray Vireos had mostly consistent rates at randomly placed ARUs, which decreased at the end of the season. The Grasshopper Sparrow singing rate at the randomly placed ARUs showed an increase around late June through mid-July, and decreased in late July (Figure 17, 18).

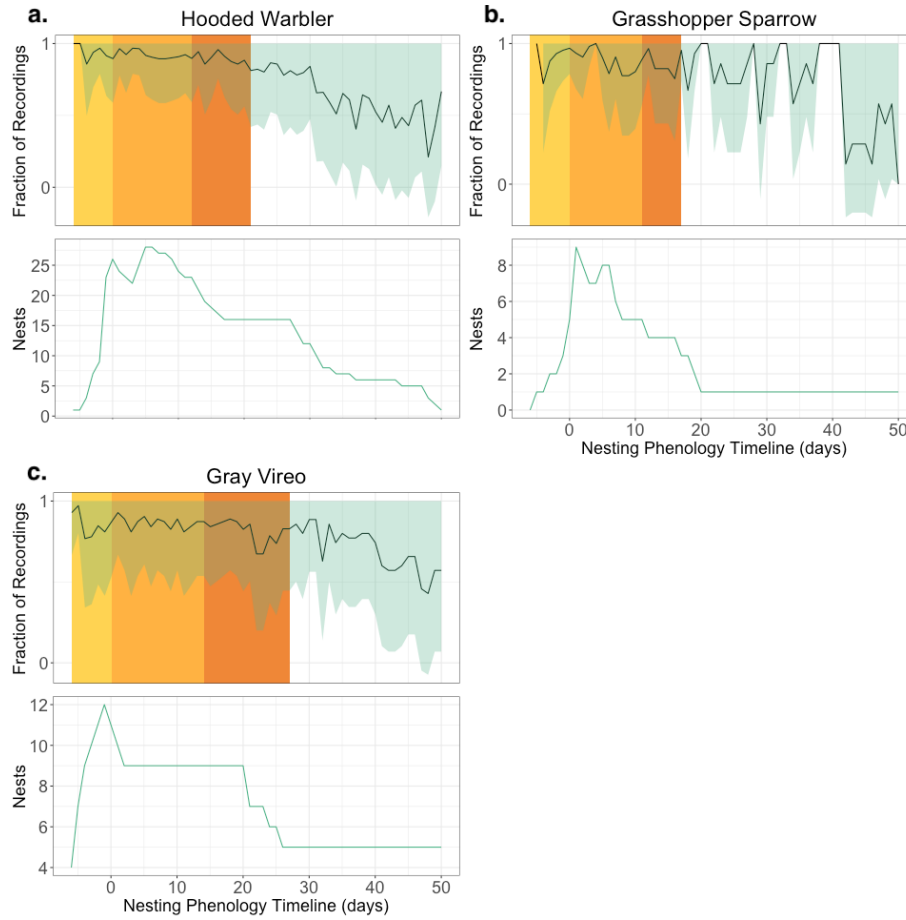


Figure 15: Song Occurrence Across the Nesting Cycle. The top graph of a, b, c represents the average fraction of recordings with songs (song occurrence) plotted against the nesting phenology day (phenoDay) from ARU-associated nests. Day zero is associated with the date when the first egg is laid. The yellow rectangle corresponds to the nest construction and egg laying phase duration, the orange rectangle corresponds to the incubation phase duration, and the darker orange rectangle corresponds to the caring for hatchling phase duration (Table. 2). The bottom graph of a, b, c represents the number of nests for which there are recordings for that day of the nesting cycle. Fewer nests means that the average song occurrence is a less reliable site-wide estimate.

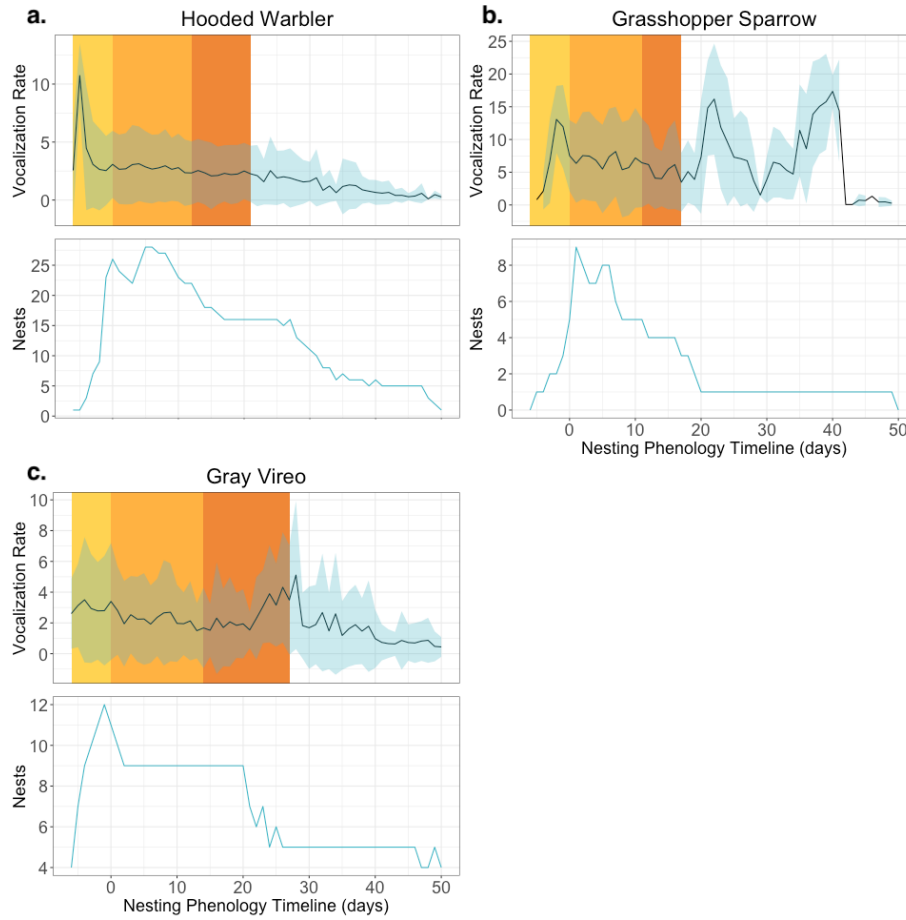


Figure 16: Average Singing Rate Across the Nesting Cycle. The top graph of a, b, c represents the average singing rate plotted against the nesting phenology day (phenoDay) from ARU-associated nests. Day zero is associated with the date when the first egg is laid. The yellow rectangle corresponds to the nest construction and egg laying phase duration, the orange rectangle corresponds to the incubation phase duration, and the darker orange rectangle corresponds to the caring for hatchling care phase duration (Table. 2). The bottom graph of a, b, c represents the number of nests for which there are recordings for that day of the nesting cycle with nonzero singing rates removed. Fewer nests means that the singing rate is a less reliable site-wide estimate.

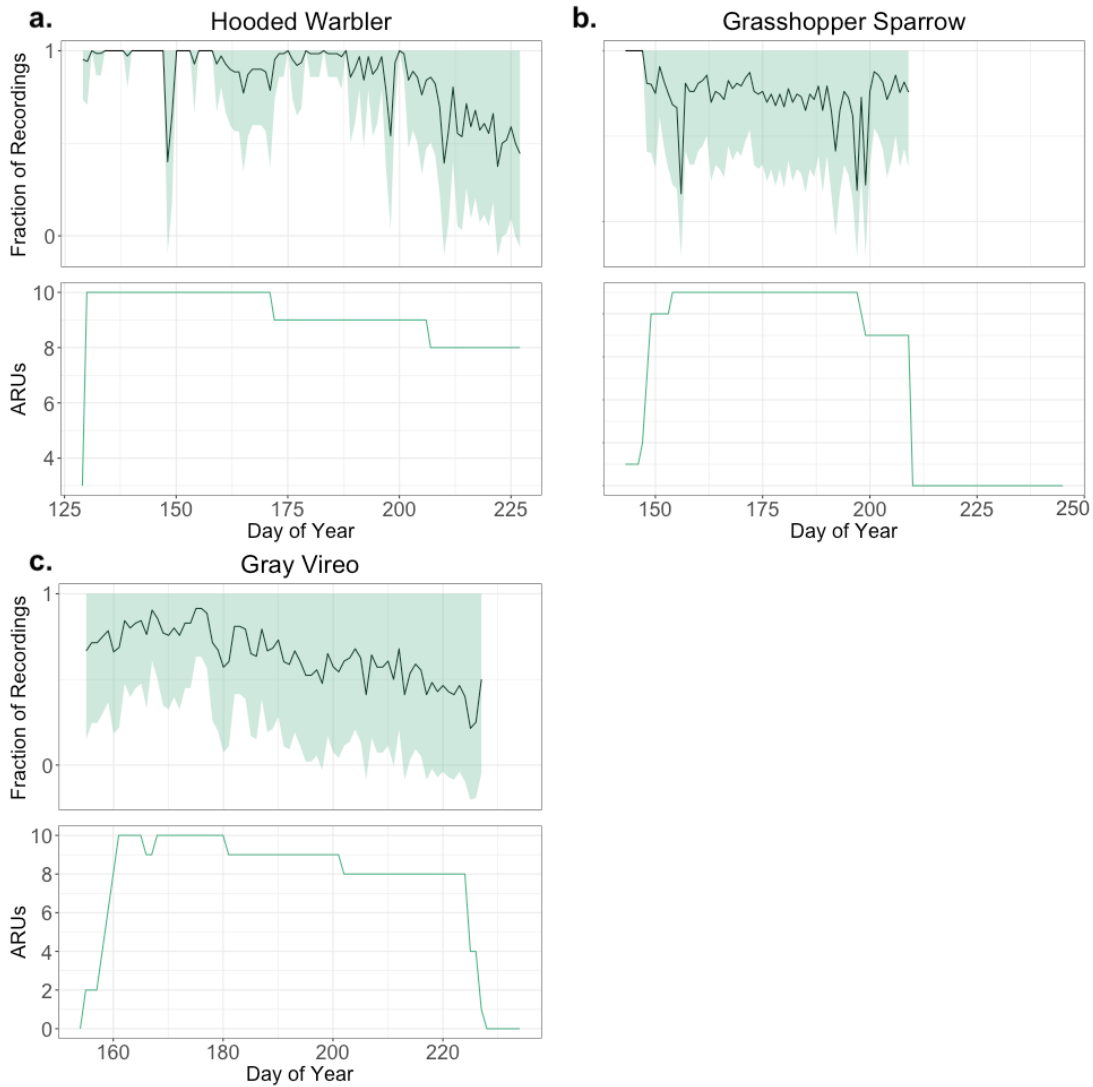


Figure 17: Song Occurrence Across the Season. The top graph of a, b, c represents the average fraction of recordings with songs (song occurrence) plotted against the day of the season from randomly placed ARUs. May 1st = 121, June 1st = 152, July 1st = 182, and August 1st = 213. The bottom graph of a, b, c represents the number of ARUs with recordings containing songs for that day of the season. Fewer ARUs mean that the average song occurrence is a less reliable site-wide estimate.

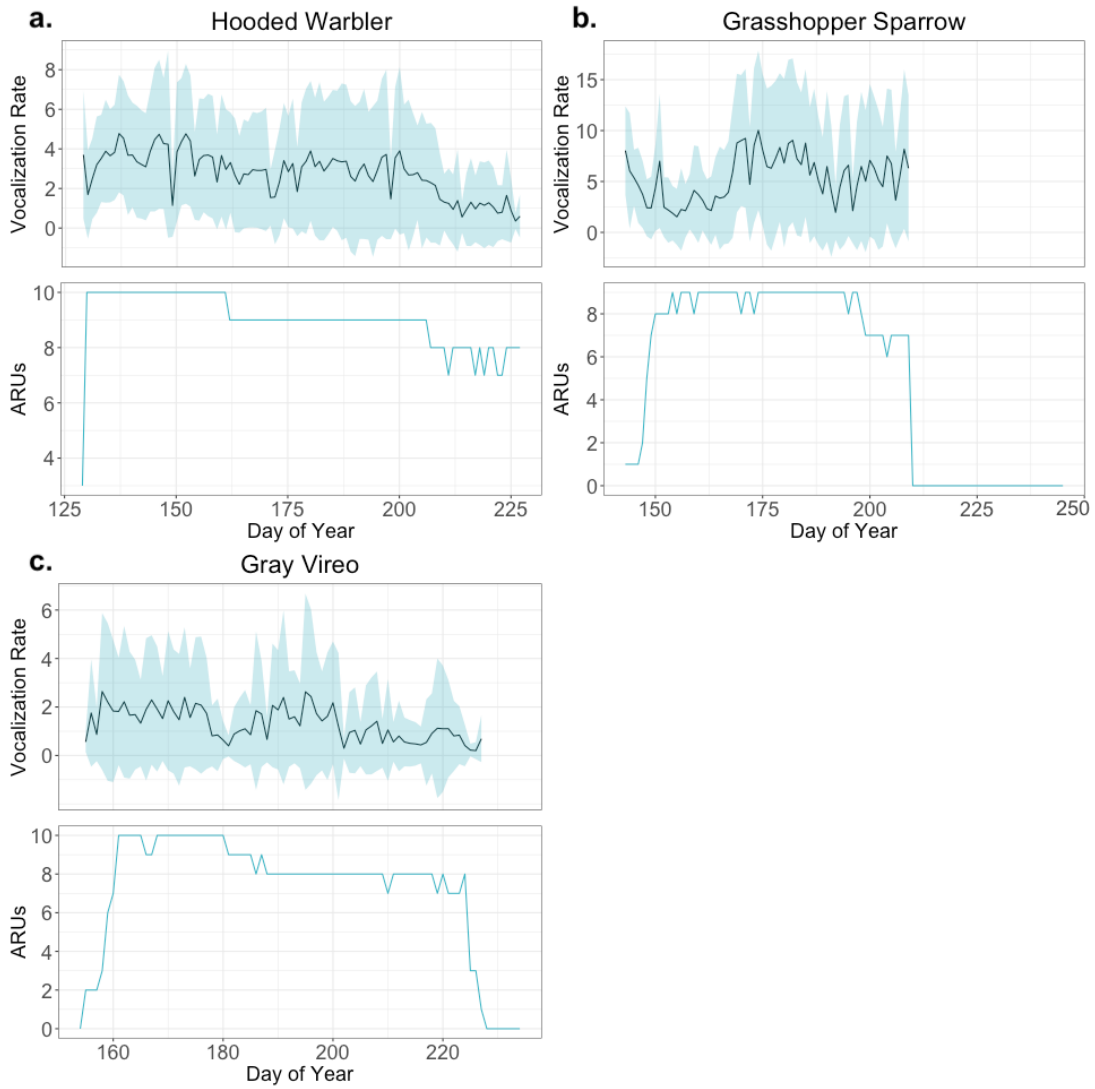


Figure 18: Average Singing Rate Across the Season. The top graph of a, b, c represents the average singing rate plotted against the day of the season from randomly placed ARUs. May 1st = 121, June 1st = 152, July 1st = 182, and August 1st = 213. The bottom graph of a, b, c represents the number of ARUs with recordings for that day of the season with nonzero singing rates removed. Fewer ARUs mean that the average singing rate is a less reliable site-wide estimate.

Discussion

4.1 Classifier Considerations

While the BirdNET Hooded Warbler classifier had the same average F1 score and a higher average precision, our proposed Hooded Warbler classifier had a far greater recall and greater ROC AUC and precision-recall AUC. The proposed Grasshopper Sparrow classifier had a higher average F1 score and average precision than the BirdNET classifier, but the same ROC AUC, the same average recall, and a lower precision-recall AUC. Overall, there was no large significant differences between the accuracy metrics of the Grasshopper Sparrow classifier and BirdNET classifier. For Hooded Warblers, the only large difference between the proposed and BirdNET classifier was average precision and recall (Table. 3).

For the Hooded Warbler and Grasshopper Sparrow, the BirdNet classifier recall quickly declined, falling far below the proposed classifiers almost immediately (Figure. 8). This result may be related to the BirdNET classifiers having a harder time identifying vocalizations from the bird of interest in noisy soundscapes. Unfortunately, this type of detection is imperative to answer our research questions. The BirdNET classifiers may be trained on a much larger dataset which is why their precision was so high. This high precision and low recall is reflected in the very low

optimal threshold values for the BirdNET classifiers (Table. 3).

For Gray Vireos, we noticed in the validation recordings and in the general field data that the recordings had very little background noise and the Gray Vireo vocalizations were very clear. There did not seem to be an abundance of sounds from other birds or sources in the area, unlike in the Grasshopper Sparrow recordings in which there were a lot of wind, insects, and other birds, and the Hooded Warbler recordings in which there were a lot of other birds. We think the BirdNET classifier performed better than our classifier on the Gray Vireo data because the lack of background noise may have increased the recall of BirdNET classifier substantially. In general, for the Gray Vireos, both classifiers did not have high recall. We think this is due to the variation in types and lengths of song syllables. Therefore, in regions where there is not much background noise and/or the birds sing in or near the nest, using a BirdNET classifier might be sufficient.

Interestingly, we would have suspected that the BirdNET classifier would have a high false positive rate, because they should also be identifying calls from the focal species as presences. These presences would be classified as false positives, because we only annotated for songs in our reviewed subset of the field data. However, this was not demonstrated in the ROC curves (Figure. 12b, d, f). We hypothesized that the BirdNET classifier does not perform very well at detecting calls of species. If we had trained our classifier on calls, they may have substantially outperformed the BirdNET classifier.

We noticed that all classifiers had difficulty registering quieter sounds. We expect that this difference in confidence was due to the training data since, as was mentioned earlier (Section. 2.3.3), xeno-canto recordings are often targeted at the species of

interest, reducing background noise and making the target species songs louder. The reason this poses an interesting problem is that for a species such as the Gray Vireo that sings from the nest, the classifier will accurately identify those songs and may not register songs from other Gray Vireos in the area. If we are only interested in songs from the birds using the nest, this difference in confidence may be beneficial, since presumably the classifier would not identify Gray Vireo calls from far away. However, for a species such as the Grasshopper Sparrow or Hooded Warbler that sing from perches, the classifier might not be able to register songs from the birds associated with a certain nest because they are singing far away from the nest.

4.1.1 Potential Modifications and Future Considerations

Our classification procedure could potentially be improved by considering and experimenting with a few possible adaptation strategies. Firstly, our singing rate was based on the number of 3-second clips that contain a song within a recording. However, some songs are shorter than 3 seconds and some are longer. This discrepancy in song length and clip length could result in situations where songs are double counted or missed. This limitation could potentially be addressed by varying the clip size depending on the species and its song length.

With regard to the training data, the accuracy of the model would likely be improved if the person annotating the xeno-canto training data had greater experience with the vocalizations of the bird of interest. One could also explore manipulating the overlay parameter weighting, the positive and negative split, or including the negative xeno-canto birds in the overlay dataset to see if recall and confidence values improve. Given that the BirdNET approach has excellent precision, but struggles

with recall, we could possibly experiment with using our proposed classifiers and BirdNET in tandem to accurately detect and identify focal species.

In terms of the validation data, we also did not explicitly annotate the reviewed data with a loud or soft label, but this may be helpful in the future to determine how well the classifier is registering softer sounds. With this data, one could also attempt to measure distance from an ARU using the intensity or pixel value of the spectrogram as a proxy. If we are only interested in vocalizations within a certain radius of an ARU, we could devise a threshold value to remove vocalizations that are too quiet.

4.2 Phenology Considerations

While we uncovered some interesting preliminary results, this research is ongoing and the analysis and conclusions are subject to change as we learn more. It is also worth noting that this data was collected from one year and from as few as 17 nests (Table. 1).

Our hierarchical approach for investigating singing activity and nesting phenology provided us with insight into different behavioral cues for each species. Song occurrence informs us of how often birds are singing throughout the day, which relates generally to territoriality and proximity to the nest while singing. An increase in song occurrence could signify a bird warning others to stay away or raising an alarm. It could also mean that the bird is spending more time near the nest. Singing rate measures the number of songs per minute given that a bird is already vocalizing. Singing rate could inform us about aggression signaling or the level of danger that a threat poses [45]. Song occurrence and singing rate could also be measures of mate

attraction or male quality. However, at our nest sites, it is unlikely that the ARUs are registering these signals since mating has already commenced [46].

4.2.1 Hooded Warbler

For Hooded Warblers, we think that the noted significant decrease in singing rate and song occurrence after fledge happens because a successful fledging will most likely occur at some point later in the season. At this time, other Hooded Warbler males may have already settled into their own nests and territories. The probability that a female will attempt a second brood or reneest rapidly declines in July [23, 47]. It seems likely that after a successful fledge, late in the season, there is less of a need to defend the territory from other males, thus the reduction in song.

We observed that there was not a significant difference in song occurrence and singing rate at a nest after failure, at a nest without eggs, and from ARUs at randomly placed locations compared to the other phenophases. We hypothesized that this may have occurred for a few different reasons. Our first theory was that because failures can occur anytime throughout the season, after a failure, there are still other Hooded Warblers in the area that the male at the nest is defending his territory against. Our second theory was that failure caused by cowbirds could potentially result in delayed abandonment of the nest. Interestingly, the singing rate at randomly placed ARUs was higher than during the hatchling phase, after fledge, after failure, and at a nest without eggs. This result could be due to the randomly placed ARUs registering more unmated males that do not have a set territory [23].

Another factor to consider is that females frequently nest near territory boundaries and Hooded Warblers do not always sing near the nest. An ARU placed at a

nest might be registering songs of multiple males regardless of the nest stage. This possibility may also explain why we did not observe any significant difference in song occurrence or singing rate throughout the standard nesting cycle (nest building, laying, incubating, and caring for hatchlings), since changes in singing activity between these phases are documented in the literature [26].

4.2.2 Grasshopper Sparrow

The higher song occurrence during the nesting cycle compared to after fledging or after failure are what we would expect to see given that after fledging or a nest failure, Grasshopper Sparrows generally abandon the nest. With regard to the randomly placed ARUs, we observed increased song occurrence in or near the nest rather than in random locations throughout the site. Grasshopper Sparrows generally perch 10-20 meters from the nest, so randomly placed ARUs that are not within a male's territory might not register songs.

We observed that the mean singing rate during egg laying and incubation was higher than during nest construction and hatchling care, which was consistent with the literature [32]. However, the difference was not significant. In general, the singing rate was not significantly different for any nest state or at randomly placed ARUs; thus, Grasshopper Sparrows seemed to change when they sang, but did not seem to markedly change the rate at which they sang.

We observed that the singing rate was low in the beginning of June (152-165), increased in late June through mid July (165-190), and then decreased again from late July through August (190+). This pattern could possibly be explained by the timing of second broods, which tend to occur in early July and may correspond to

an increase in singing rate due to mate attraction, mate protection, or increased aggression [28]. The decline is consistent with Grasshopper Sparrow behavior at the study site, since they have been observed to sing less frequently toward the end of the breeding season (mid- to late-July and after).

4.2.3 Gray Vireo

The decrease in singing rates and song occurrence during the nesting cycle compared to after fledging, after a nest failure, and at randomly placed ARUs are what we would expect to see given that Gray Vireos sing at the nest throughout the nesting cycle, and abandon the nest after failure and fledging. Within the nesting cycle, we did not observe a significant difference in song occurrence. However, the singing rate during incubation and caring for hatchlings was significantly lower than during nest construction. During nest construction, females respond to the male song, which might explain this difference. The increase in singing rate might be the ARU registering the female and male call and response.

The decrease in song occurrence and singing rate throughout the nesting cycle and season are consistent with the literature [37]. After phenoday 20 in the nesting cycle, there is not enough data to state a conclusion from the results. At randomly placed ARUs, the amount of singing seems to stay constant from mid- to late June and then starts to decrease throughout August, which is consistent with the literature [37].

While at first glance, the Gray Vireo singing rate throughout the nesting season did not seem consistent with the literature estimate of 24-41 songs per minute. However, we noted in the field data that Gray Vireos did not often sing for the entire 15

minutes [43]. They tended to sing in two-to-3-minute bouts and then stop, which might explain why the singing rate is deflated in our results.

4.2.4 Future Research and Use Consideration

Given that this research is the first step in a potentially much larger project, there are many opportunities to extend the analysis and fine-tune the approaches. One potential extension could be to investigate the relationship between different types of vocalizations and phenology. This extension of the research could be undertaken by including annotations for calls, trills, or other types of vocalizations and training separate classifiers on each type of vocalization or training one multi-target model. This idea could be further extended to look at differences in vocalization and vocalization rates between different sexes or ages.

With regard to the phenology measures, a further adaptation of this approach could involve the integration of more information from the observations taken by researchers at the study site or more realistic sampling distributions of phenological phase durations. This adaptation would make the analysis more biologically relevant and include more realistic nest-specific variation.

In terms of exploring the relationship between vocalization rate and phenological day of day of the year, we could attempt to fit a curve to the data using a generalized additive model (GAM) or another non-linear approach. As we acquire more data over the years, we will be able to construct more robust and accurate models to interpret these relationships. At this stage, there were not enough data to generate a reliable model. With an accurate model, we could potentially predict the phenological timeline or even phenological stage of a bird from a series of time-stamped

recordings. We might also be able to determine when a nest has fledged or failed from changes in vocalization or vocalization rate.

Most importantly, for the continuation of the project, the method outlined in this paper can be broadly applied to any vocal avian species. With some minor modifications, it could also be adapted to study the phenology of any species that has modular singing rates throughout its breeding cycle. The classifier can be trained on any annotated audio recordings and the phenology analysis can be adapted to include any number of stages for any time duration.

Bibliography

1. Socolar, J. B., Epanchin, P. N., Beissinger, S. R. & Tingley, M. W. Phenological shifts conserve thermal niches in North American birds and reshape expectations for climate-driven range shifts. *Proceedings of the National Academy of Sciences* **114**. Publisher: Proceedings of the National Academy of Sciences, 12976–12981. <https://www.pnas.org/doi/10.1073/pnas.1705897114> (2022) (Dec. 2017).
2. Youngflesh, C. *et al.* Circumpolar analysis of the Adélie Penguin reveals the importance of environmental variability in phenological mismatch. en. *Ecology* **98**. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ecy.1749>, 940–951. ISSN: 1939-9170. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ecy.1749> (2022) (2017).
3. Cotton, P. A. Avian migration phenology and global climate change. *Proceedings of the National Academy of Sciences* **100**. Publisher: Proceedings of the National Academy of Sciences, 12219–12222. <https://www.pnas.org/doi/abs/10.1073/pnas.1930548100> (2022) (Oct. 2003).
4. Slagsvold, T. Bird Song Activity in Relation to Breeding Cycle, Spring Weather, and Environmental Phenology. *Ornis Scandinavica (Scandinavian Journal of Ornithology)* **8**. Publisher: [Nordic Society Oikos, Wiley], 197–222. ISSN: 0030-5693. <https://www.jstor.org/stable/3676105> (2022) (1977).
5. Wilson, D. M. & Bart, J. Reliability of Singing Bird Surveys: Effects of Song Phenology during the Breeding Season. *The Condor* **87**, 69–73. ISSN: 1938-5129. <https://doi.org/10.2307/1367132> (2022) (Feb. 1985).

6. Greig-Smith, P. W. Seasonal Patterns of Song Production by Male Stonechats *Saxicola torquata*. *Ornis Scandinavica (Scandinavian Journal of Ornithology)* **13**. Publisher: [Nordic Society Oikos, Wiley], 225–231. ISSN: 0030-5693. <https://www.jstor.org/stable/3676303> (2022) (1982).
7. Gil, D., Graves, J. A. & Slater, P. J. B. Seasonal patterns of singing in the willow warbler: evidence against the fertility announcement hypothesis. en. *Animal Behaviour* **58**, 995–1000. ISSN: 0003-3472. <https://www.sciencedirect.com/science/article/pii/S0003347299912113> (2022) (Nov. 1999).
8. Hegelbach, J. & Spaar, R. Annual variation in singing activity of the Song Thrush (*Turdus philomelos*), with comments on high postbreeding song output. *Journal of Ornithology* **141**, 425–434 (Oct. 2000).
9. Sugai, L. S. M., Silva, T. S. F., Ribeiro Jr, J. W. & Llusia, D. Terrestrial Passive Acoustic Monitoring: Review and Perspectives. *BioScience* **69**, 15–25. ISSN: 0006-3568. <https://doi.org/10.1093/biosci/biy147> (2022) (Jan. 2019).
10. Blumstein, D. T. *et al.* Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. en. *Journal of Applied Ecology* **48**. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2664.2011.01993.x>, 758–767. ISSN: 1365-2664. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2664.2011.01993.x> (2022) (2011).
11. Wood, C. M., Gutiérrez, R. J. & Peery, M. Z. Acoustic monitoring reveals a diverse forest owl community, illustrating its potential for basic and applied ecology. *Ecology* **100**. Publisher: [Wiley, Ecological Society of America], 1–3. ISSN: 0012-9658. <https://www.jstor.org/stable/26785451> (2022) (2019).

12. Furnas, B. J. Rapid and varied responses of songbirds to climate change in California coniferous forests. en. *Biological Conservation* **241**, 108347. ISSN: 0006-3207. <https://www.sciencedirect.com/science/article/pii/S000632071931211X> (2022) (Jan. 2020).
13. Symes, L. B. *et al.* Analytical approaches for evaluating passive acoustic monitoring data: A case study of avian vocalizations. en. *Ecology and Evolution* **12**. ISSN: 2045-7758. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.8797> (2022) (2022).
14. Kahl, S., Wood, C. M., Eibl, M. & Klinck, H. BirdNET: A deep learning solution for avian diversity monitoring. en. *Ecological Informatics* **61**, 101236. ISSN: 1574-9541. <https://www.sciencedirect.com/science/article/pii/S1574954121000273> (2022) (Mar. 2021).
15. Sprengel, E., Jaggi, M., Kilcher, Y. & Hofmann, T. Audio based bird species identification using deep learning techniques. *Eidgenossische Technische Hochschule (ETH) Zurich*, 547–559 (2016).
16. Joly, A., Leveau, V., Champ, J. & Buisson, O. Shared nearest neighbors match kernel for bird songs identification -LifeCLEF 2015 challenge (Aug. 2015).
17. Lasseck, M. Improved Automatic Bird Identification through Decision Tree based Feature Selection and Bagging (Sept. 2015).
18. Lasseck, M. *Bird Species Identification in Soundscapes* in *CLEF* (2019).
19. Ruff, Z. J., Lesmeister, D. B., Duchac, L. S., Padmaraju, B. K. & Sullivan, C. M. Automated identification of avian vocalizations with deep convolutional neural networks. en. *Remote Sensing in Ecology and Conservation* **6**. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/rse2.125>, 79–92. ISSN: 2056-3485.

- <https://onlinelibrary.wiley.com/doi/abs/10.1002/rse2.125> (2022) (2020).
20. Stastny, J., Munk, M. & Juranek, L. Automatic bird species recognition based on birds vocalization. *EURASIP Journal on Audio, Speech, and Music Processing* **2018** (Dec. 2018).
 21. Vellinga, W.-P. *Xeno-canto - Bird sounds from around the world* en. Type: dataset. 2022. <https://www.gbif.org/dataset/b1047888-ae52-4179-9dd5-5448ea342a24> (2022).
 22. Rhinehart, T., Lapp, S. & Kitzes, J. Identifying and building on the current state of bioacoustics software. *The Journal of the Acoustical Society of America* **151**. Publisher: Acoustical Society of America, A27–A27. ISSN: 0001-4966. <https://asa.scitation.org/doi/abs/10.1121/10.0010544> (2022) (Apr. 2022).
 23. Chiver, I., Evans Ogden, L. J. & Stutchbury, B. J. Hooded Warbler (*Setophaga citrina*), version 1.0. en. *Birds of the World*. Publisher: Cornell Lab of Ornithology, Ithaca, NY, USA. <https://birdsoftheworld.org/bow/species/hoowar/cur/introduction> (2022) (2020).
 24. Evans Ogden, L. J. & Stutchbury, B. J. M. Constraints on Double Brooding in a Neotropical Migrant, the Hooded Warbler. *The Condor* **98**, 736–744. ISSN: 1938-5129. <https://doi.org/10.2307/1369855> (2022) (Nov. 1996).
 25. Ogden, L. J. E., Neudorf, D. L. H., Pitcher, T. E. & Stutchbury, B. J. M. FEMALE SONG IN THE HOODED WARBLER. *Northeastern Naturalist* **10**. Publisher: Eagle Hill Institute, 457–464. ISSN: 1092-6194, 1938-5307. <https://bioone.org/journals/northeastern-naturalist/volume-10/issue->

- 4/1092-6194_2003_010_0457_FSITHW_2.0.CO_2/FEMALE-SONG-IN-THE-HOODED-WARBLER/10.1656/1092-6194(2003)010[0457:FSITHW]2.0.CO;2.full (2022) (Dec. 2003).
26. Wiley, R. H., Godard, R. & Thompson, A. D. Use of Two Singing Modes by Hooded Warblers as Adaptations for Signalling. *Behaviour* **129**. Publisher: Brill, 243–278. ISSN: 0005-7959. <https://www.jstor.org/stable/4535197> (2022) (1994).
27. Wiens, J. A. Pattern and Process in Grassland Bird Communities. *Ecological Monographs* **43**. Publisher: Ecological Society of America, 237–270. ISSN: 0012-9615. <https://www.jstor.org/stable/1942196> (2022) (1973).
28. Vickery, P. D. Grasshopper Sparrow (*Ammodramus savannarum*), version 2.0. en. *Birds of the World*. Publisher: Cornell Lab of Ornithology, Ithaca, NY, USA. <https://birdsoftheworld.org/bow/species/graspa/cur/introduction> (2022) (1996).
29. Vickery, P. D. Grasshopper Sparrow (*Ammodramus savannarum*), version 1.0. en. *Birds of the World*. Publisher: Cornell Lab of Ornithology, Ithaca, NY, USA. <https://birdsoftheworld.org/bow/species/graspa/cur/introduction> (2022) (2020).
30. Hovick, T. J. *et al.* Effects of fire and grazing on grasshopper sparrow nest survival. en. *The Journal of Wildlife Management* **76**. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jwmg.243> (2022) (2012).
31. Winnicki, S. Growing up prairie: ecological drivers of grassland songbird nestling development (Unpublished master’s thesis). en. *Denison University*, 232 (2019).

32. Smith, R. L. THE SONGS OF THE GRASSHOPPER SPARROW. en. *THE WILSON BULLETIN* **71**, 12 (1959).
33. Delong, J. & Iii, S. Status Report and Biological Review of the Gray Vireo in New Mexico. *New Mexico Department of Game and Fish* (Dec. 2006).
34. Hargrove, L. & Unitt, P. Gray Vireo (*Vireo vicinior*) status assessment and nest monitoring to investigate causes of decline in California. *Wildlife Branch, California Department of Fish and Wildlife, Wildlife Branch Wildlife Branch, Nongame Wildlife Program Report 2014-01*. www.dfg.ca.gov/wildlife/nongame/publications/ (2014).
35. Fischer, S. E. Post-fledging and Migration Ecology of Gray Vireos (*Vireo vicinior*) and Using ArtScience to Explore Gender and Identity (Unpublished master's thesis). *University of Toledo* (2020).
36. Stevens, H. C. & Fischer, S. E. Novel nest construction behavior in Gray Vireos (*Vireo vicinior*). *The Wilson Journal of Ornithology* **130**. Publisher: Wilson Ornithological Society, 1020–1023. ISSN: 1559-4491. <https://www.jstor.org/stable/26984222> (2022) (2018).
37. Barlow, J. C., Leckie, S. N. & Baril, C. T. Gray Vireo (*Vireo vicinior*), version 1.0. en. *Birds of the World*. Publisher: Cornell Lab of Ornithology, Ithaca, NY, USA. <https://birdsoftheworld.org/bow/species/gryvir/cur/introduction> (2022) (2020).
38. Corman, T. E. & Wise-Gervais, C. *The Arizona Breeding Bird Atlas* en. Google-Books-ID: 4gaHj7s87g4C. ISBN: 978-0-8263-3379-7 (UNM Press, 2005).
39. Pierce, L. J. S. Gray Vireo (*Vireo vicinior*) Recovery Plan. *Conservation Services Division New Mexico Department of Game and Fish* (May 2007).

40. Fischer, S., Granillo, K. & Streby, H. Post-fledging survival, movements, and habitat associations of Gray Vireos in New Mexico. en. *Avian Conservation and Ecology* **17**. Publisher: The Resilience Alliance. ISSN: 1712-6568. <https://www.ace-eco.org/vol17/iss1/art13/> (2022) (Mar. 2022).
41. Werden, A. H. The implications of geographic and interseasonal variation in the song of the Gray Vireo (*Vireo vicinior*) (Unpublished master's thesis). *University of Toronto* (1991).
42. Hoffmann, R. & Brooks, A. *Birds of the Pacific states; containing brief biographies and descriptions of about four hundred species, with especial reference to their appearance in the field*, English. OCLC: 403956 (Houghton Mifflin Co., Boston and New York, 1927).
43. Borrer, D. J. Yellow-Green Vireo in Arizona, with Notes on Vireo Songs. en. *The Condor* **74**, 80. ISSN: 00105422. <https://academic.oup.com/condor/article/74/1/80-86/5202696> (2022) (1972).
44. For Conservation Bioacoustics at the Cornell Lab of Ornithology, K. L. Y. C. Raven Lite: Interactive Sound Analysis Software (Version 2.0.3) [Computer software]. *Ithaca, NY: The Cornell Lab of Ornithology*. <https://ravensoundsoftware.com/>.
45. Linhart, P., Jaška, P., Petrusková, T., Petrusek, A. & Fuchs, R. Being angry, singing fast? Signalling of aggressive motivation by syllable rate in a songbird with slow song. en. *Behavioural Processes* **100**, 139–145. ISSN: 0376-6357. <https://www.sciencedirect.com/science/article/pii/S0376635713001927> (2022) (Nov. 2013).

46. Ballentine, B., Hyman, J. & Nowicki, S. Vocal performance influences female response to male bird song: an experimental test. *Behavioral Ecology* **15**, 163–168. ISSN: 1045-2249. <https://doi.org/10.1093/beheco/arg090> (2022) (Jan. 2004).
47. Mumme, R. L. The trade-off between molt and parental care in Hooded Warblers: Simultaneous rectrix molt and uniparental desertion of late-season young. *The Auk* **135**, 427–438. ISSN: 1938-4254. <https://doi.org/10.1642/AUK-17-240.1> (2022) (July 2018).