

Knight, E. C., K. C. Hannah, and J. DeMoor. 2022. In the still of the night: revisiting Eastern Whip-poor-will surveys with passive acoustic monitoring. *Avian Conservation and Ecology* 17(1):21. <https://doi.org/10.5751/ACE-02080-170121>
Copyright © 2022 by the author(s). Published here under license by the Resilience Alliance.

Research Paper

In the still of the night: revisiting Eastern Whip-poor-will surveys with passive acoustic monitoring

Elly C. Knight¹, Kevin C. Hannah² and Jonathan DeMoor

¹Department of Biological Sciences, University of Alberta, ²Canadian Wildlife Service, Environment and Climate Change Canada

ABSTRACT. Recent advances in passive acoustic monitoring warrant the review of survey protocols because passive acoustic monitoring can increase sampling effort with minimal additional cost. In particular, protocols for nocturnal species should be re-evaluated because automated processing with signal recognition is expected to perform well for these species and surveys conducted by human observers are often limited by safety concerns. We revisited the best survey practices for the Eastern Whip-poor-will (*Antrostomus vociferus*), a nocturnal species of conservation concern. Whip-poor-will surveys are typically limited to nighttime, but also to times of high lunar illumination because their calling rate is associated with moonlight levels. We used automated recognition to extract Whip-poor-will detections from a dataset of autonomous recording unit (ARU) recordings from sites with known Eastern Whip-poor-will occupancy in Ontario, Canada. Temperature and time relative to sunset had particularly strong quadratic effects on detectability, with detectability maximized at 13 °C and 4 hours after sunset. Moon altitude and day of year had positive effects on detectability, while wind speed had negative effects on detectability. We found constraining surveys by optimal values of those detectability covariates was worthwhile only up until 10 recordings, at which point the cumulative probability of detecting an Eastern Whip-poor-will at each site was equal between constrained and unconstrained nocturnal recordings. The number of recordings required to reach an asymptote for detectability was between 81 and 97, depending on recording length. We provide objective-specific recommendations for Eastern Whip-poor-will surveys and suggest unconstrained passive acoustic monitoring as the preferred survey method for many objectives. Given the rise of passive acoustic monitoring, survey practices for many species should be revisited because the increases in sampling effort provided by ARUs can improve cumulative detection probability and potentially outweigh the advantages of limiting surveys to times and dates of optimal detectability.

Dans le calme de la nuit : réévaluer les inventaires d'Engoulevent bois-pourri au moyen de suivis acoustiques passifs

RÉSUMÉ. Les progrès récents dans le domaine des suivis acoustiques passifs justifient la révision des protocoles d'inventaire car le suivi acoustique passif permet d'augmenter l'effort d'échantillonnage à un coût supplémentaire minime. Les protocoles pour les espèces nocturnes devraient être réévalués en particulier, parce que le traitement automatisé de la reconnaissance de signaux devrait être performant pour ces espèces et les relevés menés par des observateurs humains sont souvent limités par la sécurité de ces derniers. Nous avons réexaminé les meilleures pratiques d'inventaire pour l'Engoulevent bois-pourri (*Antrostomus vociferus*), une espèce nocturne dont la conservation est préoccupante. Les inventaires d'Engoulevents bois-pourri sont généralement limités à la nuit et aussi aux périodes de forte illumination lunaire, car la fréquence de leurs appels est associée aux niveaux de lumière lunaire. Nous avons utilisé la reconnaissance automatique pour extraire les détections d'Engoulevent bois-pourri d'enregistrements provenant d'enregistreurs automatisés (EA) installés à des sites dont l'occupation par cette espèce est connue en Ontario, au Canada. La température et le temps par rapport au coucher du soleil ont eu des effets quadratiques particulièrement forts sur la détectabilité, la détectabilité maximale se situant à 13 °C et à 4 heures après le coucher du soleil. L'altitude de la lune et le jour de l'année ont eu des effets positifs sur la détectabilité, tandis que la vitesse du vent a eu des effets négatifs sur celle-ci. Nous avons constaté que le fait de restreindre les relevés aux valeurs optimales de ces covariables de détectabilité ne valait la peine que jusqu'à 10 enregistrements, après quoi la probabilité cumulée de détecter un engoulevent à chaque site était égale entre les enregistrements nocturnes restreints ou non restreints. Le nombre d'enregistrements requis pour atteindre une asymptote de détectabilité se situait entre 81 et 97, selon la longueur de l'enregistrement. Nous offrons des recommandations spécifiques aux objectifs pour les inventaires d'Engoulevent bois-pourri et proposons que le suivi acoustique passif sans restriction soit la méthode d'inventaire privilégiée pour de nombreux objectifs. Étant donné l'essor des suivis acoustiques passifs, les pratiques d'inventaire pour de nombreuses espèces devraient être réévaluées, car l'augmentation de l'effort d'échantillonnage fournie par les EA peut améliorer la probabilité de détection cumulative et l'emporter potentiellement sur les avantages de limiter les inventaires aux heures et aux dates de détectabilité optimale.

Key Words: *autonomous recording unit (ARU); citizen science; detection probability; Eastern Whip-poor-will; environmental assessment; monitoring; nightjar; nocturnal; passive acoustic monitoring; survey*

INTRODUCTION

Successful management of avian species includes regular evaluation of survey best practices in an adaptive management framework (Yoccoz et al. 2001, Nichols and Williams 2006, Lindenmayer and Likens 2009). Part of this process of revisiting survey protocols should include quantifying and understanding probability of detection (hereafter “detectability”; Kéry and Schmidt 2008). Understanding the conditions where detectability is maximized can result in improved precision of statistical estimates such as population trend (Diefenbach et al. 2007, Sauer et al. 2017), occupancy, or density (MacKenzie et al. 2002, Rosenberg et al. 2017). Imperfect detectability on surveys can also have further consequences for species, such as exclusion from biodiversity reporting (North American Bird Conservation Initiative Canada 2019), environmental impact assessment, range delineation, or other assessments.

Nightjars (Family Caprimulgidae) are nocturnal birds that require specialized protocols to maximize detectability. Obviously, constraining observations to darkness is important for detecting nightjars and has been shown to be critical for detecting population declines relative to dawn surveys (Knight et al. 2021). Surveying during optimal lunar conditions is another well-established survey suggestion for nightjars. Vocal activity of Common Poorwills (*Phalaenoptilus nuttallii*; Brauner 1953, Woods and Brigham 2008), Chuck-will’s-widows (*Antrostomus carolinensis*; Harper 1938, Cooper 1981), Eastern Whip-poor-wills (*Antrostomus vociferus*; Cooper 1981, Mills 1986, Wilson and Watts 2006), Red-necked Nightjars (*Caprimulgus ruficollis*; Reino et al. 2015), Common Pauraque (*Nyctidromus albicollis*), and Little Nightjar (*Setopagis parvula*; Pérez-Granados et al. 2022) have all been positively associated with moonlight levels, including lunar phase or percent lunar illumination and lunar altitude. Recommendations have been made for surveys to be conducted around the full moon (Reino et al. 2015), and specifically when there is at least 50% lunar illumination (Wilson and Watts 2006). Citizen science monitoring protocols for nightjars, including the recently formalized Canadian Nightjar Survey, now recommend or require surveying within one week of the full moon (Knight et al. 2019). Lunar phase or illumination has also been included as a detectability covariate in occupancy studies (Farrell et al. 2017, Vala et al. 2020).

Avian survey methods are changing because of recent advances in passive acoustic monitoring, and so it follows that best survey practices should be revisited. Autonomous recordings units (ARUs) facilitate surveying sites many more times than human observers can at little to no extra effort or cost (Shonfield and Bayne 2017, Gibb et al. 2018). ARUs are useful for surveying nightjars because frequent nocturnal surveys in remote areas can be both logistically challenging and dangerous. Automated computer recognition of acoustic signals allows efficient processing of large amounts of acoustic recordings. Unlike the diurnal period, the nocturnal soundscape is much quieter, reducing the likelihood of sound masking, i.e., overlap of sounds in time and frequency band. Reduced sound masking during the nocturnal period makes it much more straightforward for a computer to separate nightjar calls from the rest of the soundscape (Zwart et al. 2014, Knight et al. 2017, Pérez-Granados and Schuchmann 2020). Once detected, their frequent, simple vocalizations are relatively easy for computer algorithms to

classify. Automated recognition facilitates the collection of large, highly detailed survey datasets with a time stamp for every vocalization. Other parameters such as relative sound level can also be derived concurrently and can be used to estimate the distance of detection for every vocalization (Yip et al. 2020). Although automated recognition does require visual or aural validation to remove false positive detections from the results, it can be approximately five times more efficient than aural processing alone (Knight et al. 2017), and there are post-processing approaches available that can further increase this efficiency (Balantic and Donovan 2020, Knight et al. 2020).

Given these improvements in survey efficiency, constraining nightjar surveys by lunar conditions or other covariates (hereafter “constrained survey”) may not always be the most effective survey recommendation. If increasing survey effort is of relatively minimal additional cost—as is the case with passive acoustic monitoring—conducting surveys that are not restricted to specific times or conditions (hereafter “unconstrained survey”) may have a higher cumulative probability of detection because they allow for the collection of more information. In fact, an early assessment of ARU and recognizer technology for the Eastern Whip-poor-will concluded that passive acoustic monitoring and automated recognition could enable monitoring on “more dates than possible using established field protocols” (Clark and Frstrup 2009:8). Maximizing survey effort and, thus, potentially cumulative probability of detection is particularly important for objectives that require confirmation of species presence or absence such as environmental assessment or range delineation.

Our goal was to revisit best survey practices for the Eastern Whip-poor-will for both ARU and human surveys. The Eastern Whip-poor-will is a nightjar species that has a Near Threatened global conservation status (Cink et al. 2020), is listed as Threatened under Canada’s Species at Risk Act, and is on the U.S. Fish and Wildlife Service’s (USFWS) list of Birds of Conservation Concern. Developing standardized survey protocols has been identified as a high-priority knowledge gap for the Eastern-Whip-poor-will (Environment and Climate Change Canada 2018). We were interested in informing ARU protocols because passive acoustic monitoring has been identified as a potential survey method for this species given its extensive range and nocturnal habits (Environment and Climate Change Canada 2018). We were also interested in informing human surveys, including the recently formalized citizen science Canadian Nightjar Survey. We used automated recognition to extract Whip-poor-will detections from a high temporal resolution dataset of ARU recordings from sites in southern Ontario with known Eastern Whip-poor-will occupancy. First, we used an occupancy analysis framework to determine which temporal and weather covariates best predicted Whip-poor-will detectability. We then examined the effects of sampling effort (in this case, recording length and number of recordings) on detectability and occupancy estimates and cumulative probability of detection. Next, we examined whether constraining surveys by detectability covariates, including percent lunar illumination, or using all available recordings resulted in higher cumulative probability of detection. We compared constrained and unconstrained surveys with mixed effects logistic regression across a range of recording lengths and number of recordings. We then used a nonlinear least squares growth model to determine the sample size at which cumulative probability of

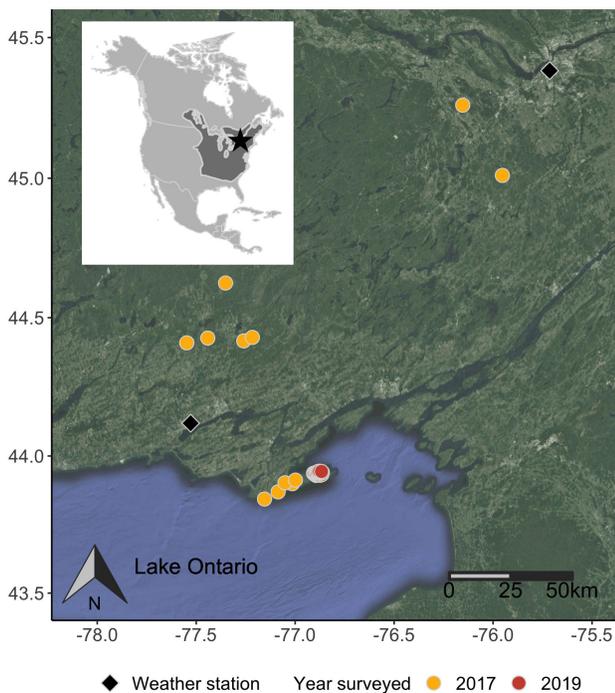
detection asymptoted to inform environmental assessment. Based on our results, we provide recommendations for Eastern Whip-poor-will survey protocols in an objective-based framework by differentiating between objectives that require confirmation of presence-absence and objectives that can accommodate imperfect detection.

METHODS

Study area and sites

We studied detectability of Eastern Whip-poor-wills near the center of their breeding range in eastern Ontario, Canada (Fig. 1). Our study sites included two areas: (i) the transition zone between the Canadian Shield and Mixedwood Plains ecozones (Crins et al. 2009), and (ii) the southern portion of Prince Edward County. We selected these areas as part of a (i) grassland bird breeding phenology study in 2017, and (ii) a systematic survey of breeding birds on the Prince Edward Point National Wildlife Area (PEP hereafter) in 2019. Although the surveys were focused on diurnal passerine birds, we also included crepuscular and nocturnal sampling in those areas given the high potential suitability of the habitat for Eastern Whip-poor-will. The vegetation community in both study areas consisted mostly of open shrubland and thicket, dominated by red cedar (*Juniperus virginiana*), Red-osier dogwood (*Cornus stolonifera*), and prickly ash (*Zanthoxylum americanum*), surrounded by mature hardwood forest.

Fig. 1. Study area and sites in Ontario, Canada with known Eastern Whip-poor-will (*Antrostomus vociferus*) occupancy where autonomous recording units were deployed to collect acoustic recordings. Dark grey shading on the inset map shows the breeding range of the Eastern Whip-poor-will.



At those two study areas, we selected 32 study sites with known Eastern Whip-poor-will occupancy. At the grassland area, initial suitability of 18 sites was evaluated using Google Earth imagery, where sites containing open grassland with < 50% forest canopy closure within a 100 m radius of each selected site were considered further. Six of the initial 18 grassland sites were removed after ground-truthing because of unsuitable habitat (n = 4) and excessive vehicular noise (n = 2), resulting in 12 locations for ARU deployment. At PEP, we divided the area into a grid of hexagons with 250 m between centroids and randomly selected hexagons with at least 750 m between centroids to maintain spatial independence of survey locations, resulting in 20 locations for ARU deployment.

Audio recording collection

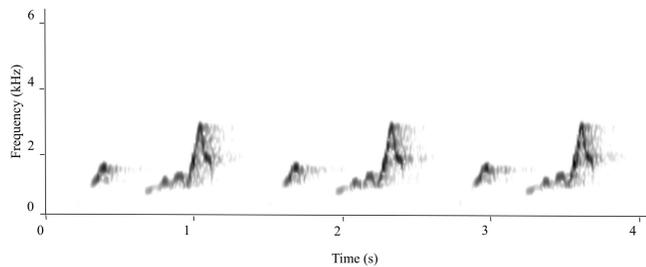
All recordings were made using ARUs (Model SM2+, Wildlife Acoustics, Maynard, MA, USA). We used a sampling rate of 16 kHz, or double the maximum frequency of a typical Whip-poor-will song (Cink et al. 2020), and a bit depth of 16 bits. We used the same model of ARU between years to ensure a comparable signal-to-noise ratio (e.g., Darras et al. 2020), and conducted routine microphone testing and replacement between years to ensure microphone sensitivity and potential detection radius was not compromised (Turgeon et al. 2017). We affixed ARUs to vertical tree trunks or fence posts at a height of approximately 1.5 m and removed branches, leaves or other debris near microphones that might impede clear recordings. In 2017, we collected recordings every third day at 30, 90, and 150 minutes after sunset from 16 April to 30 July, resulting in 108 10-minute recordings per site. In 2019, we collected recordings every second day at every hour on the hour from sunset to sunrise from 17 May to 16 July, resulting in 279 five-minute recordings per site. Premature battery failure of some ARUs resulted in fewer recordings at some sites, particularly in 2017 (Appendix 1).

Recognizer construction

We used Song Scope software (Wildlife Acoustics, Maynard, MA, USA) to construct a recognizer for the Eastern Whip-poor-will. Song Scope is a signal detection recognizer that extracts Mel Frequency Cepstral Coefficients from each detected signal and computes the overall score using Hidden Markov Models. Although Song Scope was recently discontinued by its manufacturer, we chose it because it remains freely available and was shown to perform well for another nightjar species, the Common Nighthawk (*Chordeiles minor*; Knight et al. 2017). We selected 80 clips of full Eastern Whip-poor-will songs (Fig. 2) from a dataset of high-quality recordings collected across southern Ontario from 2015 to 2020 using autonomous recording units (Model SM2+, Wildlife Acoustics, Maynard, MA, USA) and a shotgun microphone system (Nagra SD digital audio recorder, Sennheiser ME66 shotgun microphone). We also explored building recognizers with just the second note of the song phrase (Fig. 2) and a longer clip of multiple song phrases, but found the single full phrase performed best in preliminary evaluation. We used only loud clips recorded at close-range to maximize the probability of detection at 0 m and facilitate large-scale use of the recognizer (Knight and Bayne 2019, Knight et al. 2020). We removed any clips that were not fully detected by the signal detection process in Song Scope (setting available in

Appendix 2), leaving 60 clips for recognizer training. The recognizer file we constructed for Song Scope software is available in Appendix 3.

Fig. 2. Spectrogram of the song of the Eastern Whip-poor-will (*Anrostomus vociferus*). Spectrogram shows three separate song phrases, each with two notes (“Whip”, “poor-will”).



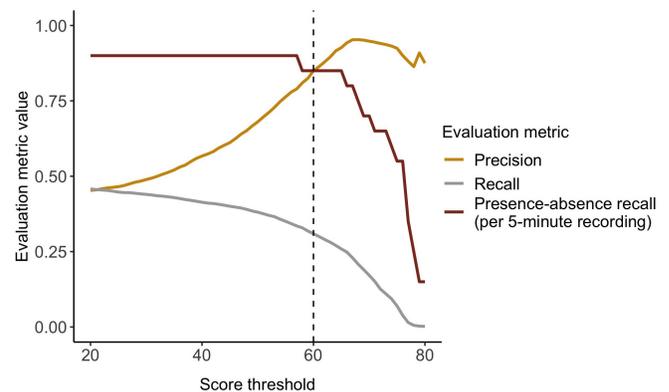
Recognizer evaluation

We followed Knight et al. (2017) to evaluate the precision and recall of our recognizer and select a score threshold for processing. We randomly selected a test dataset of five-minute nighttime recordings where we knew Eastern Whip-poor-wills were present, visually scanned them for calls, and categorized them as present or absent. We continued this process until we had 20 presence and 20 absence recordings. We visually and/or aurally processed the presence recordings and counted the number of Eastern Whip-poor-will calls in each recording. We then used our trained Song Scope recognizer to scan those 40 selected recordings with a score threshold of 20 and a quality threshold of 20. We visually validated the recognizer results to separate true and false positives and found the recognizer detected Eastern Whip-poor-wills in 18 of the 20 (90%) presence recordings. Of the 3525 Eastern Whip-poor-will calls we detected in the test dataset by sight/sound, 1626 (46%) were detected by the recognizer. We then compared the number of detections in the recognizer results to the number of detections in the benchmark data across all possible score thresholds to determine precision (proportion of true positives vs false positives) and recall (proportion of true positives vs false negatives; Knight et al. 2017; Fig. 3). We also determined the recall of Eastern Whip-poor-will presence per recording across all possible score thresholds. Our results indicated the call-level recall of the recognizer was low, however, so we also examined the relationship between recall and sound energy, which is a proxy for detection distance (Hedley et al. 2020, Yip et al. 2020), to ensure this low recall was simply due to a lower effective detection radius than that of a human observer (Appendix 4).

Acoustic data processing

We scanned our entire set of recordings with our Song Scope recognizer with a score threshold of 60 and a quality threshold of 0. We selected a score threshold of 60 to balance false positives and false negatives at the recording level. Our recognizer evaluation suggested a score threshold of 60 resulted in detected of 30.9% of individual Eastern Whip-poor-will calls but 85% of five-minute recordings, as compared with a human listener (Fig. 3). We visually validated the recognizer results to separate true and false positives.

Fig. 3. Evaluation of Eastern Whip-poor-will (*Anrostomus vociferus*) call detection for a recognizer built in Song Scope software across multiple score thresholds. Precision is the proportion of recognizer hits that are true detections. Recall is the proportion of target species vocalizations detected by the recognizer. Presence-absence recall is the proportion of five-minute recordings containing Eastern Whip-poor-wills in which the recognizer detected the target species. Results are from a test dataset of 40 nocturnal recordings, 20 of which contained Eastern Whip-poor-will vocalizations. Dashed line represents the score threshold selected for acoustic data processing.



Detectability covariates

We collected temporal, solar, lunar, and weather detectability covariates for each recording at each study site. We used the “suncalc” package (Thieurmel and Elmarhraoui 2019) to calculate time since sunset and moon fraction for every recording. We used the “weathercan” package (LaZerte and Albers 2018) to retrieve weather variables from nearby weather stations. Of the available weather stations in the study area, we selected the nearest one that had the most complete hourly dataset available (Appendix 5). For each recording, we retrieved hourly wind speed and temperature data and daily total precipitation data. We also quantified potential weather effects on perceptibility via sound masking or degradation. We used the “hardRain” package (Metcalf et al. 2020) to quantify the signal-to-noise ratio (StN) and power spectrum density (PSD) between 4.4 and 5.6 kHz for each one-minute interval of recording. We chose that signal band because it has been shown to be effective at classifying heavy rainfall (Metcalf et al. 2020). We then checked all potential covariates for variance inflation and correlation and removed any variables with VIF > 5 or correlation > 0.7.

Statistical analysis

First, we used an occupancy modeling framework in the “unmarked” package (Fiske and Chandler 2011) in R version 4.0.3 (R Core Team 2020) to determine which covariates predicted Eastern Whip-poor-will detectability. We randomly sampled 50 recordings from each of the 32 study sites to even out sampling across the study sites (Appendix 1). We fit two sets of occupancy models: one with all potential solar, lunar, and temporal covariates, and one with weather covariates. We included time relative to sunset and temperature as second order polynomials based on predicted relationships for these two covariates. We also

Table 1. Occupancy model selection results for detectability (P) of Eastern Whip-poor-wills (*Antrostomus vociferus*) from acoustic recordings. Results are the mean and standard deviation (SD) of 100 bootstrapped models for each covariate set (1) temporal, lunar, and solar (day = day of year, altitude = lunar altitude, sunset = time relative to sunset); (2) weather (PSD = power spectrum density, StN = signal to noise ratio, temperature = temperature in degrees Celsius, wind = wind speed, rain = total daily precipitation). Bold indicates the model with the highest mean model weight that was thus selected for subsequent analyses. Occupancy ~1 for all models.

Model terms	K	Mean weight (SD)	Mean Δ AICc(SD)	% times top model
Temporal, lunar, and solar covariates				
$P \approx \text{day} + \text{altitude} + \text{sunset}^2$	6	0.27 (0.21)	2.54 (5.03)	34
$P \approx \text{altitude} + \text{sunset}^2$	5	0.16 (0.18)	11.00 (16.29)	25
$P \approx \text{fraction} + \text{altitude} + \text{sunset}^2$	6	0.15 (0.12)	2.68 (4.93)	15
$P \approx \text{day} + \text{fraction} + \text{altitude} + \text{sunset}^2$		0.09 (0.12)	5.62 (2.27)	4
Weather covariates				
$P \approx \text{PSD} + \text{StN} + \text{temperature}^2 + \text{wind}$	7	0.09 (0.12)	4.04 (3.47)	19
$P \approx \text{StN} + \text{temperature}^2 + \text{wind}$	6	0.06 (0.10)	5.14 (3.69)	16
$P \approx \text{PSD} + \text{StN} + \text{temperature}^2 + \text{wind} + \text{rain}$	8	0.05 (0.09)	13.30 (14.30)	13
$P \approx \text{PSD} + \text{wind}$	4	0.05 (0.08)	4.70 (3.19)	20

included an interaction between moon altitude and fraction. For each set, we fit one global model and all potential combinations of the covariates in that model. We bootstrapped this visit selection and model fitting process 100 times. From each set, we selected the model with the highest model weight across the 100 bootstraps. We then combined all the covariates from the two best fitting models into a final model. We again randomly sampled 50 5-minute visits and fit them to this final model. We bootstrapped visit sampling and model fitting 100 times. We summarized the 100 bootstraps to obtain mean and standard error estimates for the coefficients in our final model.

We then examined the effects of sampling effort on Eastern Whip-poor-will detectability and occupancy estimates. To examine the effects of sampling effort, we constructed occupancy models for a range of recording lengths (1 to 5 minutes) and number of recordings (1 to 30, 40, 50, 60, 70, 80, 90, 100). For each combination of these two sampling effort parameters, we randomly sampled the appropriate number of recordings from each study site and the appropriate recording length from each of those recordings, starting at the beginning of the recording. We fit an occupancy model to the validated recognizer data from those recordings, including the detectability covariates from the final model of our previous analysis. We bootstrapped this sampling and model fitting process 100 times and calculated the mean occupancy and detectability estimates and 95% confidence intervals (CI) across those bootstraps.

Next, we examined the effects of constraining surveys by detectability covariates. In other words, how is sampling effort affected by using only recordings from times and days when the probability of detection is high, for example when surveys are conducted by a human observer? We restricted recordings to those collected between sunset and 7 hours after sunset, when the moon altitude was greater than 0.6 radians, when the temperature was between 7 and 20 degrees Celsius, and when the wind speed was less than 19 km/h. These thresholds were selected following recommendations from the Canadian Nightjar Survey and the results from the previous analysis (Table 1, Fig. 4). We repeated the previous analysis, with the exception that we only selected up to 10 recordings per study site because that was the minimum available after constraining our dataset by our detectability covariates.

We compared the cumulative probability of detection of constrained and unconstrained surveys using logistic regression. We used a binomial response variable of whether an Eastern Whip-poor-will was detected at each site for each bootstrap and included site as a random effect. We included length of recording, number of recordings, and covariate approach (constrained/unconstrained) as covariates and all two-way interactions. We only used bootstraps with up to 10 recordings to allow direct comparison of the two approaches.

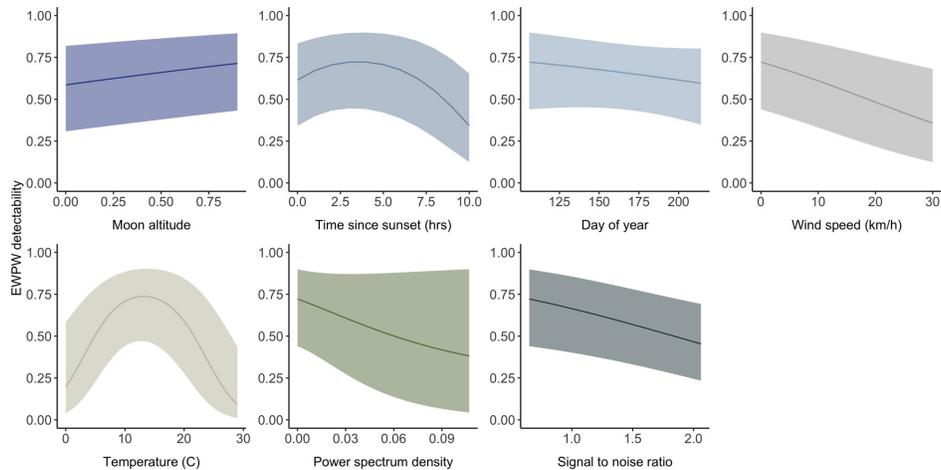
We also determined the proportion of the 32 sites at which a Whip-poor-will was detected per bootstrap and used an asymptotic nonlinear least squares growth model (Von Bertalanffy 1957) to model the proportion of sites with detections across number of recordings for each of the recording lengths (1 to 5 minutes). We determined the number of recordings required to reach the predicted asymptote of proportion of sites with detections as the x-intercept of 99% of the asymptote. We repeated this analysis for recording length across number of visits (every 10 visits from 10 to 100).

RESULTS

The recognizer reported 115,954 potential Eastern Whip-poor-will detections, of which 52,704 were validated as true positives and 63,250 were removed as false positives (precision = 0.45). Of those false positives, 56,663 (89.6%) detections occurred during the dawn chorus just before sunrise. The number of Eastern Whip-poor-will detections per minute of recording ranged from 0 to 336 (mean = 4.76, SD = 25.51). The mean rate of Eastern Whip-poor-will detections per site varied from 0.02/min to 30.54/min (mean = 5.03/min, SD = 8.08/min).

The model with highest mean model weight across bootstraps for solar, lunar, and temporal covariates included everything except moon fraction; it included day of year, time relative to sunset, and moon altitude (Table 1). The model with the highest mean model weight across bootstraps for weather covariates included everything except total daily rain; it included temperature, wind speed, StN, and PSD (Table 1). Moon altitude had a positive effect on Eastern Whip-poor-will detectability, while day of year, wind speed, StN, and PSD all had negative effects, with wind speed having the strongest effect (Fig. 4). Temperature and time relative

Fig. 4. Mean and 95% confidence interval of effects of temporal, lunar, solar, weather, and recording covariates on the detectability of Eastern Whip-poor-wills (*Antrostomus vociferus*), as predicted by occupancy models. Predictions for each covariate are made while holding all other covariates at values that maximize detectability (i.e., maximum for moon altitude, mean for time since sunset and temperature, and minimum for day of year, wind speed, power spectrum density, and signal-to-noise ratio).

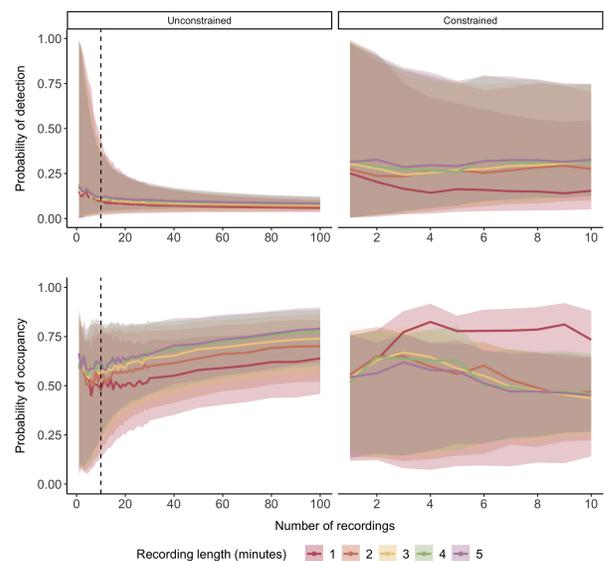


to sunset had quadratic effects on detectability, with the highest probability of detection at 13 °C and 4 hours after sunset, respectively.

The mean probability of detection varied between 0.059 for an unconstrained survey of 100 1-minute recordings and 0.325 for a constrained survey of 10 5-minute recordings (Fig. 5). The mean probability of detection per recording across all combinations of recording length and number of recordings was 0.103 when all recordings were included and 0.268 when recordings constrained by detectability covariates were included. Recording length affected mean probability of detection, with shorter recordings having a lower probability of detection, particularly for 1-minute recordings when surveys were constrained by detectability covariates (Fig. 5). Mean probability of detection was slightly higher for unconstrained surveys with few recordings; however, the main effect of number of recordings was to reduce the uncertainty of the detectability estimate.

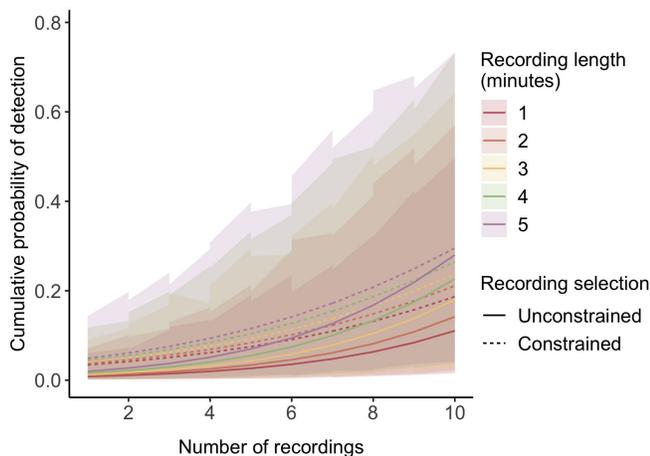
The mean probability of occupancy estimates varied between 0.445 for an unconstrained survey of 9 2-minute recordings and 0.821 for a constrained survey of five 1-minute recordings (Fig. 5). The mean probability of occupancy per recording across all combinations of recording length and number of recordings was 0.608 when all recordings were included and 0.583 when only recordings were constrained by detectability covariates. In general, the occupancy estimates were higher for longer recordings and increasing numbers of recordings; however, the one-minute constrained recordings had the highest occupancy estimates, likely because the model corrected for the particularly low detectability of this survey effort combination. For both constrained and unconstrained surveys, mean probability of occupancy peaked between 1 and 5 surveys, and then stabilized; however, all differences were well within the mean 95% confidence intervals of the bootstrapped model estimates.

Fig. 5. Effects of sampling effort (recording length and number of recordings) for autonomous recording unit surveys of Eastern Whip-poor-wills (*Antrostomus vociferus*). Probability of detection and probability of occupancy estimates are mean predictions and 95% confidence intervals from occupancy models of 100 bootstrapped datasets. Two model sets were examined: one in which all nocturnal recordings in the dataset were included (“unconstrained”), and one for which recordings were constrained by covariates that had been shown to affect detectability (“constrained”). The dotted line on the unconstrained model set indicates the maximum number of recordings for the constrained model set and is included to facilitate comparison between the two plots with differing x-axis scales.



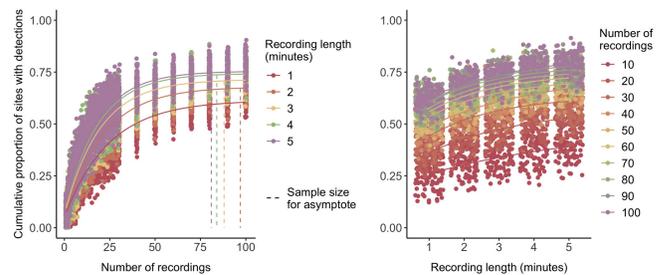
The cumulative probability of detection ranged between 0.7% for one unconstrained 1-minute recording to 30.7% for 10 constrained 5-minute recordings (Fig. 6). The cumulative probability of detection for constrained surveys was more than twice as high as that of unconstrained surveys when only one recording was sampled (4.6% vs 1.9% for 5-minute recordings); however, as sample size increased, this difference decreased until they were nearly equal for 10 recordings (28.7% vs 27.9% for 5-minute recordings).

Fig. 6. Effects of sampling effort (recording length and number of recordings) for autonomous recording unit surveys of Eastern Whip-poor-wills (*Antrostomus vociferus*). Cumulative probability of detection was defined as whether an Eastern Whip-poor-will was detected within a set of randomly selected recordings. Mean and 95% confidence intervals of cumulative probability of detection were calculated from 100 bootstraps for each of the recording length and number of recording combinations. Two model sets were examined: one in which all nocturnal recordings in the dataset were included (“unconstrained”), and one for which recordings were constrained by covariates that had been shown to affect detectability (“constrained”). Shaded intervals represent the 95% confidence interval and are shown only for the unconstrained model set for visualization (both model sets had nearly identical confidence intervals).



For number of recordings, the proportion of sites at which an Eastern Whip-poor-will was detected reached an asymptote between 0.679 and 0.754, depending on recording length (Figure 7). No asymptote was reached for one-minute recordings. The minimum sample size required to reach 99% of that asymptote was lowest for longer recordings; minimum sample size was 81, 84, 88, and 97 recordings for five-minute through two-minute recordings. For recording length, the predicted values for up to five minutes recording length did not reach the asymptote predicted by the nonlinear least squares growth model, but the maximum proportion of sites at which an Eastern Whip-poor-will was detected was between 0.480 and 0.837 for five-minute recordings. The proportion of sites at which an Eastern Whip-poor-will was detected did not reach one because there were several sites at which an Eastern Whip-poor-will was detected in only a few recordings (1 recording: 3 sites, 2 recordings: 1 site, 3 recordings: 3 sites).

Fig. 7. Effects of sampling effort (recording length and number of recordings) for autonomous recording unit surveys of Eastern Whip-poor-wills (*Antrostomus vociferus*). Cumulative probability of detection was defined as whether an Eastern Whip-poor-will was detected in any of the randomly selected recordings in each of 100 bootstraps for each of the recording length and number of recording combinations. Lines are the mean model predictions from an asymmetric nonlinear least squares growth model. The sample size required to reach the predicted asymptote was calculated as the x-intercept of 99% of the asymptote.



DISCUSSION

We used time series ARU survey data from sites with known Eastern Whip-poor-will occupancy to revisit survey protocols for this species. Temperature and time relative to sunset had particularly strong effects on detectability, with detectability maximized at intermediate values of both. Moon altitude had positive effects on detectability, while day of year, wind speed, and two acoustic measurements of potential sound masking or attenuation from weather (StN: signal-to-noise ratio; PSD: power spectrum density) negatively affected detectability. We found constraining surveys by optimal values of those detectability covariates was worthwhile only up until 10 recordings, at which point the cumulative probability of detecting an Eastern Whip-poor-will at each site was equal between constrained and unconstrained nocturnal recordings. The number of recordings required to reach an asymptote for proportion of sites with detections was between 53 and 65 recordings, depending on recording length; however, we did not find an asymptote for recording length.

The design and evaluation of survey methods should be performed in an objective-specific framework (O'Connor et al. 2000, Yoccoz et al. 2001, Nichols and Williams 2006), and so we recommend survey protocols should be selected based on the project objectives and logistical constraints (Table 2). Choosing between constrained versus unconstrained and human observer versus ARU surveys should therefore be done on a case-by-case basis because there are trade-offs for each choice. In general, human surveys have higher detectability because of a larger effective detection radius (Yip et al. 2017a, Darras et al. 2018), but we showed here that passive acoustic monitoring has higher overall detectability because it is more efficient for collecting time series data (among many other benefits; Rempel et al. 2013, Shonfield and Bayne 2017, Darras et al. 2019). Passive acoustic monitoring does, however, require specialized equipment and

Table 2. Recommendations for autonomous recording unit (ARU) and human surveys of Eastern Whip-poor-wills (*Antrostomus vociferus*) along a gradient of sampling objectives from imperfect detection to confirmation of presence or absence.

Sampling objective	Potential applications	ARU surveys (per season)	Human surveys	Notes
Imperfect detection	Occupancy modeling, population monitoring	Up to 80 unconstrained nighttime recordings of at least 5 minutes length each	Constrain surveys: - Temperature: 7–20 °C - Wind: < 20 km/h - Time: nautical twilight - Moon: > 50% illumination - Caveat: may be unsafe in remote areas	- More recordings or visits will reduce uncertainty of statistical estimates. - Collection of detectability covariate data is required to account for imperfect detection (Rempel et al. 2013).
Confirm presence or absence	Environmental assessment	At least 80 unconstrained nighttime recordings of at least 5 minutes length each	- Number of surveys required may be unrealistic for human surveyors - Call playback surveys may improve probability of detection	- Caveat: perfect cumulative probability of detection unlikely - Recordings > 5 minutes may yield higher cumulative probability of detection than reported here

training, which can be unrealistic for survey programs that cover large study areas.

On one end of the objectives spectrum, surveys requiring confirmation of species presence or absence should use protocols that maximize cumulative probability of detection (e.g., Pérez-Granados et al. 2018; Table 2). Undue harm to Eastern Whip-poor-wills is of particular concern to environmental managers because this species is listed as Threatened under Canada’s Species at Risk Act, and nest disturbance is prohibited under Canada’s Migratory Bird Conservation Act. Our results suggested at least 80 recordings of at least 5 minutes length each are required to reach an asymptote in cumulative probability of detection (Table 2). We note, however, that our asymptote was not at a cumulative probability of detection of 1.00, but rather at 0.75 because there were several sites at which an Eastern Whip-poor-will was only detected a handful of times. These detections could be of prospecting individuals; however, the date when most the detections occurred (June) is beyond the territory settlement phase of the breeding period in this region. Instead, we suggest these are individuals at the extreme periphery of their territory, which is typically 2–10 ha (Fitch 1958), or that these males are non-territorial “floaters” that were not breeding (Hunt 2016). There is a possibility that samples longer than the 5-minute recordings we used could increase the proportion of sites at which an Eastern Whip-poor-will was detected; we did not find an asymptote of cumulative probability of detection for recording length. Guidance is mixed on whether fewer long recordings (Rempel et al. 2013, Sugai et al. 2020) or more short recordings (Cook and Hartley 2018) result in higher detection probability. Our results suggest that recordings should be at minimum two minutes long for occupancy modeling because our occupancy estimate for 1-minute recordings was higher than our other estimates, suggesting unstable parameterization.

On the other end of the spectrum, surveys for which the intended objective can accommodate imperfect detection can have much more flexible protocols (Table 2). Although maximizing cumulative probability of detection is desirable because model estimates have higher precision when detectability is maximized (MacKenzie et al. 2002, Diefenbach et al. 2007), there are ways to account for imperfect detection: hierarchical models in

occupancy modeling (MacKenzie et al. 2002), detectability covariates in population trend estimation, e.g., observer ability (Link and Sauer 2016), and offsets or correction factors in density estimation (Buckland et al. 1995, Sóllymos and Lele 2016, Sóllymos et al. 2018), to name a few.

If using passive acoustic monitoring, we suggest Eastern Whip-poor-will surveys should move beyond the previous reliance on moon phase and use an unconstrained protocol for most applications. We showed that cumulative probability of detection is maximized by conducting many unconstrained visits. In other words, using all survey data and accounting for imperfect detection with the appropriate covariates results in higher overall detectability than only surveying when conditions are optimal. If using this unconstrained approach, we recommend quantifying poor recording quality with measurements of signal-to-noise ratio and power spectrum density (Metcalfe et al. 2020), as we showed both had significant effects on Eastern Whip-poor-will detectability.

Using passive acoustic monitoring and an unconstrained approach to survey Eastern Whip-poor-wills and other nocturnal species should be feasible for most applications if equipment is available. ARUs are a valuable tool for surveying nocturnal species like nightjars (Frommolt and Tauchert 2014, Shonfield et al. 2018, Duchac et al. 2020) because nocturnal human surveys for nightjars are often restricted to roadsides for safety considerations (Takats et al. 2001, Knight et al. 2019), which can result in a biased understanding of habitat relationships, occupancy, and population size (Pankratz et al. 2017, Yip et al. 2017b). The time required to process ARU recordings can be an obstacle for passive acoustic monitoring; however, we showed here automated recognition was an effective method for our study. Although our recognizer precision was not particularly high (0.45; compared to > 70% in Knight et al. 2017, Pérez-Granados and Schuchmann 2020, for other nightjar species), we did not find validation an onerous process and automated processing was much more efficient than manual review for processing our dataset of over 650 hours of audio recordings. Furthermore, the majority (89.6%) of our false positives occurred during dawn chorus, which we sampled to ensure we captured the full range of Eastern Whip-poor-will availability for detection; practitioners using ARUs to

monitor Eastern Whip-poor-wills could greatly increase the precision of the recognizer by omitting dawn chorus sampling. Our recognizer recall rate was low for individual calls (0.31; compared to 0.74, 0.85 in Pérez-Granados and Schuchmann 2020, for other nightjar species); however, the recall rate at the five-minute recording level was quite high (0.85), likely due to the high call rate of the Eastern Whip-poor-will (Mulaik et al. 1953). The recording-level recall rate was the evaluation metric of interest for our study because we used detection/non-detection at the recording level as our response variable, and so we are confident in our results. Studies that wish to use call rate or time-to-first detection as a response variable for applications like density estimation should explore more cutting-edge algorithms like convolutional neural networks that might yield higher recall values (Stowell et al. 2019).

On the other hand, surveys by human observers may be advantageous if there is a large survey area to cover and/or repeat visits are not feasible (Klingbeil and Willig 2015). For example, range-wide population monitoring programs like the North American Breeding Bird Survey and the Canadian Nightjar Survey rely on single visit surveys by citizen scientists to monitor avian population trends (Downes et al. 2016, Hudson et al. 2017, Knight et al. 2019). Nightjar detectability during human observer surveys may also be improved by using call playbacks (e.g., Zuberogoitia et al. 2020). If surveys are conducted by human observers, the observation process should be constrained by detectability covariates because more than 10 repeat visits are unlikely for human observer protocols (Table 2). We showed temperature and wind have the strongest effects on Eastern Whip-poor-will detectability, with detections maximized at intermediate values. Other authors have found effects of weather (wind, rain) on Eastern Whip-poor-will detectability (Farrell et al. 2017, Vala et al. 2020). Cool weather is likely particularly important for nightjar activity levels because many species, including Eastern Whip-poor-wills, are able to undergo partial torpor (Lane et al. 2004). Red-necked Nightjar (*Caprimulgus ruficollis*) detectability has also been shown to be particularly sensitive to temperature (Camacho 2013). We therefore recommend that the Canadian Nightjar Survey be constrained to between 13 and 20 °C. We also showed time relative to sunset had an effect on Eastern Whip-poor-will detectability, with detectability maximized at 4 hours after sunset; however, the confidence intervals on the effects of time relative to sunset were quite wide. We therefore recommend surveys be conducted anytime when the sun is less than 6 degrees below the horizon, such as during nautical or astronomical twilight, or night. Finally, we showed that moon altitude had a positive effect on Eastern Whip-poor-will detectability, which is contrary to other studies that have found an effect of moon illumination (Cooper 1981, Mills 1986, Wilson and Watts 2006). The origin of using moon illumination instead of moon altitude to constrain nightjar surveys dates back to Mills (1986) who actually found a stronger effect of altitude but opted to recommend moon illumination because all moons with high illumination also have high altitude. We also recommend surveys continue to be constrained by moon illumination because it is a much easier constraint to interpret and implement, which is important for the success of monitoring programs based around citizen science (Parsons et al. 2011, McKinley et al. 2017).

Revisiting survey protocols for acoustic species is timely because of the rapid proliferation of passive acoustic monitoring and our

results about the context-dependency of survey recommendations suggest revisitation may also be warranted. Regular evaluation of survey protocols relative to desired outcomes, particularly for long-term monitoring programs, is also important because it encourages managers to revisit the objectives and design of the program as part of an adaptive management framework (Yoccoz et al. 2001, Nichols and Williams 2006, Lindenmayer and Likens 2009). We showed that constraining the observation process by detectability covariates may no longer be an optimal survey method because of the large volume of information that can be collected if constraints are not applied, as suggested previously by others. We suggest that revisiting survey protocols for other nocturnal species may be warranted because automated recognition is likely to work well for those species, and there are several examples already published (Zwart et al. 2014, Knight et al. 2017, Shonfield et al. 2018). The method we present here of using ARU data from sites of known occupancy to understand detectability trade-offs and refine survey protocols can be applied to other taxa. We suggest this approach could enhance wildlife management for other species because improved survey protocols can provide more efficient use of resources and more precise statistical estimates that inform management decisions.

Responses to this article can be read online at:
<https://www.ace-eco.org/issues/responses.php/2080>

Acknowledgments:

Thanks to Subject Editor W. Thogmartin and two anonymous reviewers for their thoughtful comments that improved our manuscript. We thank M. Cadman and N. Spencer for assisting with fieldwork and data management. We thank T. Trustham for permission and logistical support in accessing field locations managed by Quinte Conservation. We thank the Hastings Prince Edward Land Trust for permission to access the Miller Family Nature Reserve. Finally, we thank D. Sadler and M. Shapiera with Environment and Climate Change Canada for permission and logistical support in accessing Prince Edward Point National Wildlife Area.

LITERATURE CITED

- Balantic, C. M., and T. M. Donovan. 2020. Statistical learning mitigation of false positives from template-detected data in automated acoustic wildlife monitoring. *Bioacoustics* 29:296-321. <https://doi.org/10.1080/09524622.2019.1605309>
- Brauner, J. 1953. Observations on the behavior of a captive poor-will. *Condor* 55:68-74. <https://doi.org/10.2307/1365027>
- Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake. 1995. Distance sampling: estimating abundance of biological populations. Chapman & Hall, London, UK.
- Camacho, C. 2013. Behavioural thermoregulation in man-made habitats: surface choice and mortality risk in Red-necked Nightjars. *Bird Study* 60:124-130. <https://doi.org/10.1080/00063-657.2012.753400>

- Cink, C. L., P. Pyle, and M. A. Patten. 2020. Eastern-whip-poor-will (*Antrostomus vociferus*), version 1.0. In P. G. Rodewald, editor. Birds of the world. Cornell Lab of Ornithology, Ithaca, New York, USA. <https://doi.org/10.2173/bow.whip-p1.01>
- Clark, C. W., and K. M. Fristrup. 2009. Advanced technologies for acoustic monitoring of bird populations. Report prepared for Department of Defense Strategic Environmental Research and Development Program. Cornell University, Ithaca, New York, USA.
- Cook, A., and S. Hartley. 2018. Efficient sampling of avian acoustic recordings: intermittent subsamples improve estimates of single species prevalence and total species richness. Avian Conservation and Ecology 13(1):21. <https://doi.org/10.5751/ACE-01221-130121>
- Cooper, R. J. 1981. Relative abundance of Georgia Caprimulgids based on call-counts. Wilson Bulletin 93:363-371.
- Crins, W. J., P. A. Gray, P. W. C. Uhlig, and M. C. Wester. 2009. The ecosystems of Ontario, Part I: eozones and ecoregions. Ontario Ministry of Natural Resources, Inventory, Monitoring and Assessment, Peterborough, Ontario, Canada.
- Darras, K., P. Batáry, B. Furnas, A. Celis-Murillo, S. L. V. Wilgenburg, Y. A. Mulyani, and T. Tschardt. 2018. Comparing the sampling performance of sound recorders versus point counts in bird surveys: a meta-analysis. Journal of Applied Ecology 55:2575-2586. <https://doi.org/10.1111/1365-2664.13229>
- Darras, K., P. Batáry, B. J. Furnas, I. Grass, Y. A. Mulyani, and T. Tschardt. 2019. Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide. Ecological Applications 29:e01954. <https://doi.org/10.1002/eap.1954>
- Darras, K., F. Deppe, Y. Fabian, A. Kartono, A. Angulo, B. Kolbrek, Y. Mulyani, and D. Prawiradilaga. 2020. High microphone signal-to-noise ratio enhances acoustic sampling of wildlife. PeerJ 8:e9955. <https://doi.org/10.7717/peerj.9955>
- Diefenbach, D. R., M. R. Marshall, J. A. Mattice, and D. W. Brauning. 2007. Incorporating availability for detection in estimates of bird abundance. Auk 124:96-106. <https://doi.org/10.1093/auk/124.1.96>
- Downes, C. M., M.-A. R. Hudson, A. C. Smith, and C. M. Francis. 2016. The Breeding Bird Survey at 50: scientists and birders working together for bird conservation. Avian Conservation and Ecology 11(1):8. <https://doi.org/10.5751/ACE-00855-110108>
- Duchac, L. S., D. B. Lesmeister, K. M. Dugger, Z. J. Ruff, and R. J. Davis. 2020. Passive acoustic monitoring effectively detects Northern Spotted Owls and Barred Owls over a range of forest conditions. Condor 122:duaa017. <https://doi.org/10.1093/condor/duaa017>
- Environment and Climate Change Canada. 2018. Recovery strategy for the Eastern Whip-poor-will (*Antrostomus vociferus*) in Canada. Environment and Climate Change Canada, Ottawa, Ontario, Canada.
- Farrell, C. E., S. Wilson, and G. Mitchell. 2017. Assessing the relative use of clearcuts, burned stands, and wetlands as breeding habitat for two declining aerial insectivores in the boreal forest. Forest Ecology and Management 386:62-70. <https://doi.org/10.1016/j.foreco.2016.11.026>
- Fiske, I., and R. Chandler. 2011. unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. Journal of Statistical Software 43(10):1-23. <https://doi.org/10.18637/jss.v043.i10>
- Fitch, H. S. 1958. Home ranges, territories, and seasonal movements of vertebrates of the Natural History Reservation. University of Kansas Publications Museum of Natural History 11:63-326.
- Frommolt, K.-H., and K.-H. Tauchert. 2014. Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. Ecological Informatics 21:4-12. <https://doi.org/10.1016/j.ecoinf.2013.12.009>
- Gibb, R., E. Browning, P. Glover-Kapfer, and K. E. Jones. 2018. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods in Ecology and Evolution 3:992-17.
- Harper, F. 1938. The Chuck-will's-widow in the Okefinokee region. Oriole 3:9-14.
- Hedley, R. W., S. J. Wilson, D. A. Yip, K. Li, and E. M. Bayne. 2020. Distance truncation via sound level for bioacoustic surveys in patchy habitat. Bioacoustics 30:303-323. <https://doi.org/10.1080/09524622.2020.1730240>
- Hudson, M.-A. R., C. M. Francis, K. J. Campbell, C. M. Downes, A. C. Smith, and K. L. Pardieck. 2017. The role of the North American Breeding Bird Survey in conservation. Condor 119:526-545. <https://doi.org/10.1650/CONDOR-17-62.1>
- Hunt, P. D. 2016. Extra-territorial movements by Eastern Whip-poor-wills. North American Bird Bander 41(3):97-102.
- Kéry, M., and B. R. Schmidt. 2008. Imperfect detection and its consequences for monitoring for conservation. Community Ecology 9:207-216. <https://doi.org/10.1556/ComEc.9.2008.2.10>
- Klingbeil, B. T., and M. R. Willig. 2015. Bird biodiversity assessments in temperate forest: the value of point count versus acoustic monitoring protocols. PeerJ 3:e973 <http://dx.doi.org/10.7717/peerj.973>
- Knight, E. C., and E. M. Bayne. 2019. Classification threshold and training data affect the quality and utility of focal species data processed with automated audio-recognition software. Bioacoustics 28:539-554. <https://doi.org/10.1080/09524622.2018.1503971>
- Knight, E. C., K. C. Hannah, R. M. Brigham, J. McCracken, G. Falardeau, M.-F. Julien, J.-S. Guénette, A. Manthorne, A. Sidler, G. J. Foley, J. McKnight, K. S. Laurent, P. Sinclair, and R. Pankratz. 2019. Canadian Nightjar survey protocol. WildResearch, Vancouver, British Columbia, Canada.
- Knight, E. C., K. C. Hannah, G. Foley, C. Scott, R. M. Brigham, and E. Bayne. 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. Avian Conservation and Ecology 12(2):14. <https://doi.org/10.5751/ACE-01114-120214>

- Knight, E. C., A. C. Smith, R. M. Brigham, and E. M. Bayne. 2021. Combination of targeted monitoring and Breeding Bird Survey data improves population trend estimation and species distribution modelling for the Common Nighthawk. *Ornithological Applications* 123:duab005. <https://doi.org/10.1093/ornithapp/duab005>
- Knight, E. C., P. Sólymos, C. Scott, and E. M. Bayne. 2020. Validation prediction: a flexible protocol to increase efficiency of automated acoustic processing for wildlife research. *Ecological Applications* 30:e02140. <https://doi.org/10.1002/eap.2140>
- Lane, J. E., R. M. Brigham, and D. L. Swanson. 2004. Daily torpor in free-ranging Whip-poor-wills (*Caprimulgus vociferus*). *Physiological and Biochemical Zoology* 77:297-304. <https://doi.org/10.1086/380210>
- LaZerte, S. E., and S. Albers. 2018. weathercan: Download and format weather data from Environment and Climate Change Canada. *Journal of Open Source Software* 3(22):571. <https://doi.org/10.21105/joss.00571>
- Lindenmayer, D. B., and G. E. Likens. 2009. Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in Ecology & Evolution* 24:482-486. <https://doi.org/10.1016/j.tree.2009.03.005>
- Link, W. A., and J. R. Sauer. 2016. Modeling participation duration, with application to the North American Breeding Bird Survey. *Communications in Statistics-Theory and Methods* 45:6311-6320. <https://doi.org/10.1080/03610926.2014.957854>
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248-2255. [https://doi.org/10.1890/0012-9658\(2002\)083\[2248:ESORWD\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2)
- McKinley, D. C., A. J. Miller-Rushing, H. L. Ballard, R. Bonney, H. Brown, S. C. Cook-Patton, D. M. Evans, R. A. French, J. K. Parrish, T. B. Phillips, S. F. Ryan, L. A. Shanley, J. L. Shirk, K. F. Stepenuck, J. F. Weltzin, A. Wiggins, O. D. Boyle, R. D. Briggs, S. F. C. III, D. A. Hewitt, P. W. Preuss, and M. A. Soukup. 2017. Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation* 208:15-28. <https://doi.org/10.1016/j.biocon.2016.05.015>
- Metcalf, O. C., A. C. Lees, J. Barlow, S. J. Marsden, and C. Devenish. 2020. hardRain: An R package for quick, automated rainfall detection in ecoacoustic datasets using a threshold-based approach. *Ecological Indicators* 109:105793. <https://doi.org/10.1016/j.ecolind.2019.105793>
- Mills, A. M. 1986. The influence of moonlight on the behavior of goatsuckers (Caprimulgidae). *Auk* 103(2):370-378. <https://doi.org/10.1093/auk/103.2.370>
- Mulaik, S., W. Lee, and J. E. Stillwell. 1953. Analysis of the call of the Whip-Poor-Will. *Auk* 70:367-368. <https://doi.org/10.2307/4081331>
- Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. *Trends in Ecology & Evolution* 21:668-673. <https://doi.org/10.1016/j.tree.2006.08.007>
- North American Bird Conservation Initiative Canada. 2019. The state of Canada's birds. Environment and Climate Change Canada, Ottawa, Ontario, Canada.
- O'Connor, R. J., E. Dunn, D. H. Johnson, S. L. Jones, D. Petit, K. Pollock, C. R. Smith, J. L. Trapp, and E. Welling. 2000. A programmatic review of the North American Breeding Bird Survey. USGS Patuxent Wildlife Research Center, Patuxent, Maryland, USA.
- Pankratz, R. F., S. Hache, P. Sólymos, and E. M. Bayne. 2017. Potential benefits of augmenting road-based breeding bird surveys with autonomous recordings. *Avian Conservation and Ecology* 12(2):18. <https://doi.org/10.5751/ACE-01087-120218>
- Parsons, J., R. Lukyanenko, and Y. Wiersma. 2011. Easier citizen science is better. *Nature* 471:37. <https://doi.org/10.1038/471037a>
- Pérez-Granados, C., G. Bota, D. Giralt, and J. Traba. 2018. A cost-effective protocol for monitoring birds using autonomous recording units: a case study with a night-time singing passerine. *Bird Study* 65(3):338-345. <https://doi.org/10.1080/00063657.2018.1511682>
- Pérez-Granados, C., and K.-L. Schuchmann. 2020. Illuminating the nightlife of two Neotropical nightjars: vocal behavior over a year and monitoring recommendations. *Ethology Ecology & Evolution* 32:466-480. <https://doi.org/10.1080/03949370.2020.1753117>
- Pérez-Granados, C., K.-L. Schuchmann, and M. I. Marques. 2022. Addicted to the moon: vocal output and diel pattern of vocal activity in two Neotropical nightjars is related to moon phase. *Ethology Ecology & Evolution* 34:66-81. <https://doi.org/10.1080/03949370.2021.1886182>
- R Core Team. 2020. R: A language and environment for statistical computing. Version 4.0.3. R Foundation for Statistical Computing, Vienna, Austria. [online] URL: <https://www.R-project.org/>
- Reino, L., M. Porto, J. Santana, and T. S. Osiejuk. 2015. Influence of moonlight on nightjars' vocal activity: a guideline for nightjar surveys in Europe. *Biologia* 70:968-973. <https://doi.org/10.1515/biolog-2015-0099>
- Rempel, R. S., C. M. Francis, J. N. Robinson, and M. Campbell. 2013. Comparison of audio recording system performance for detecting and monitoring songbirds. *Journal of Field Ornithology* 84:86-97. <https://doi.org/10.1111/jof.12008>
- Rosenberg, K. V., P. J. Blancher, J. C. Stanton, and A. O. Panjabi. 2017. Use of North American Breeding Bird Survey data in avian conservation assessments. *Condor* 119:594-606. <https://doi.org/10.1650/CONDOR-17-57.1>
- Sauer, J. R., D. K. Niven, K. L. Pardieck, D. J. Ziolkowski Jr., and W. A. Link. 2017. Expanding the North American Breeding Bird Survey analysis to include additional species and regions. *Journal of Fish and Wildlife Management* 8:154-172. <https://doi.org/10.3996/102015-JFWM-109>
- Shonfield, J., and E. M. Bayne. 2017. Autonomous recording units in avian ecological research: current use and future applications. *Avian Conservation and Ecology* 12(1):14. <https://doi.org/10.5751/ACE-00974-120114>

- Shonfield, J., S. Heemskerk, and E. M. Bayne. 2018. Utility of automated species recognition for acoustic monitoring of owls. *Journal of Raptor Research* 52:42-55. <https://doi.org/10.3356/JRR-17-52.1>
- Sólymos, P., and S. R. Lele. 2016. Revisiting resource selection probability functions and single-visit methods: clarification and extensions. *Methods in Ecology and Evolution* 7:196-205. <https://doi.org/10.1111/2041-210X.12432>
- Sólymos, P., S. M. Matsuoka, S. G. Cumming, D. Stralberg, P. Fontaine, F. K. A. Schmiegelow, S. J. Song, and E. M. Bayne. 2018. Evaluating time-removal models for estimating availability of boreal birds during point count surveys: sample size requirements and model complexity. *Condor* 120:765-786. <https://doi.org/10.1650/CONDOR-18-32.1>
- Stowell, D., T. Petrusková, M. Šálek, and P. Linhart. 2019. Automatic acoustic identification of individuals in multiple species: improving identification across recording conditions. *Journal of the Royal Society Interface* 16:20180940. <https://doi.org/10.1098/rsif.2018.0940>
- Sugai, L. S. M., C. Desjonquères, T. S. F. Silva, and D. Llusia. 2020. A roadmap for survey designs in terrestrial acoustic monitoring. *Remote Sensing in Ecology and Conservation* 6:220-235. <https://doi.org/10.1002/rse2.131>
- Takats, D. L., C. M. Francis, G. L. Holroyd, J. R. Duncan, K. M. Mazur, R. J. Cannings, W. Harris, and D. Holt. 2001. Guidelines for nocturnal owl monitoring in North America. *Beaverhill Bird Observatory and Bird Studies Canada, Edmonton, Alberta, Canada.*
- Thieurmél, B., and A. Elmarhraoui. 2019. *suncalc*: Compute sun position, sunlight phases, moon position and lunar phase. R package version 0.5.0. [online] URL: <https://CRAN.R-project.org/package=suncalc>
- Turgeon, P. J., S. L. Van Wilgenburg, and K. L. Drake. 2017. Microphone variability and degradation: implications for monitoring programs employing autonomous recording units. *Avian Conservation and Ecology* 12(1):9. <https://doi.org/10.5751/ACE-00958-120109>
- Vala, M. A., G. W. Mitchell, K. C. Hannah, J. Put, and S. Wilson. 2020. The effects of landscape composition and configuration on Eastern Whip-poor-will (*Caprimulgus vociferous*) and Common Nighthawk (*Chordeiles minor*) occupancy in an agroecosystem. *Avian Conservation and Ecology* 15(1):24. <https://doi.org/10.5751/ACE-01613-150124>
- Von Bertalanffy, L. 1957. Quantitative laws in metabolism and growth. *Quarterly Review of Biology* 32(3):217-231. <https://doi.org/10.1086/401873>
- Wilson, M. D., and B. D. Watts. 2006. Effect of moonlight on detection of Whip-poor-wills: implications for long-term monitoring strategies. *Journal of Field Ornithology* 77(2):207-211. <https://doi.org/10.1111/j.1557-9263.2006.00042.x>
- Woods, C. P., and R. M. Brigham. 2008. Common Poorwill activity and calling behavior in relation to moonlight and predation. *Wilson Journal of Ornithology* 120:505-512. <https://doi.org/10.1676/06-067.1>
- Yip, D. A., E. M. Bayne, P. Sólymos, J. Campbell, and D. Proppe. 2017b. Sound attenuation in forest and roadside environments: implications for avian point-count surveys. *Condor* 119:73-84. <https://doi.org/10.1650/CONDOR-16-93.1>
- Yip, D. A., E. C. Knight, E. Haave-Audet, S. J. Wilson, C. Charchuk, C. D. Scott, P. Sólymos, and E. M. Bayne. 2020. Sound level measurements from audio recordings provide objective distance estimates for distance sampling wildlife populations. *Remote Sensing in Ecology and Conservation* 6(3):301-315. <https://doi.org/10.1002/rse2.118>
- Yip, D. A., L. Leston, E. M. Bayne, P. Sólymos, and A. Grover. 2017a. Experimentally derived detection distances from audio recordings and human observers enable integrated analysis of point count data. *Avian Conservation and Ecology* 12(1):11. <https://doi.org/10.5751/ACE-00997-120111>
- Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution* 16(8):446-453. [https://doi.org/10.1016/S0169-5347\(01\)02205-4](https://doi.org/10.1016/S0169-5347(01)02205-4)
- Zuberogitia, I., J. E. Martínez, J. A. González-Oreja, C. González de Buitrago, G. Belamendia, J. Zabala, M. Laso, N. Pagaldai, and M. V. J. Franco. 2020. Maximizing detection probability for effective large-scale nocturnal bird monitoring. *Diversity and Distributions* 26:1034-1050. <https://doi.org/10.1111/ddi.13075>
- Zwart, M. C., A. Baker, P. J. K. McGowan, and M. J. Whittingham. 2014. The use of automated bioacoustic recorders to replace human wildlife surveys: an example using nightjars. *PLoS ONE* 9:e102770. <https://doi.org/10.1371/journal.pone.0102770>



Appendix 2. Recognizer construction details.

Table A2.1. Parameter settings for recognizers built in Song Scope software.

Parameter	Setting
Sample rate (Hz)	16000
FFT size	256
FFT overlap	$\frac{3}{4}$
Frequency minimum	16
Frequency range	40
Amplitude gain (dB)	0
Background filter (s)	1
Max syllable (ms)	300
Max syllable gap (ms)	300
Max song (ms)	1016
Dynamic range (dB)	20
Algorithm	2.0
Maximum complexity	32
Maximum resolution	10

Appendix 3. Examination of relationship between recognizer recall and mean relative sound level (RSL).Examination of relationship between recognizer recall and mean relative sound level (RSL).

Appendix 3 can be downloaded from the website

<https://www.ace-eco.org/vol17/iss1/art21/>

Appendix 4. Examination of relationship between recognizer recall and mean relative sound level (RSL).

We examined the relationship between recall and sound energy, which is a proxy for detection distance (Yip et al. 2020, Hedley et al. 2020), to ensure low recall of our recognizer was simply due to a lower effective detection radius than a human observer. We measured the sound energy, or loudness of each Eastern Whip-poor-will detection in our test dataset with two methods. First, we assigned a loudness ranking (Table A3.1) to each calling bout during manual counting of Eastern Whip-poor-will calls in our test dataset. We defined a calling bout as a duration of consecutive “whip-poor-will” calls with less than one second in between successive calls. Second, we measured the relative sound level (RSL) in each detected call with Song Scope software while running the recognizer.

Table A3.1. Rankings used to describe the loudness of Eastern Whip-poor-will calling bouts during manual counting of calls in audio recordings.

Loudness ranking	Description	Total # of calling bouts	Total duration of calling bouts (minutes)	Total number of calls
1	Barely perceptible	16	4.75	240
2	Barely perceptible to quiet	13	8.10	426
3	Quiet	29	8.23	455
4	Quiet to medium	10	13.17	777
5	Medium	3	1.67	113
6	Medium to loud	2	4.80	270
7	Loud	23	16.58	191
8	Very loud	4	3.25	191
	Not assessed	1	0.63	38

We then visually evaluated the relationship between recall and each sound energy measurement across the range of possible score thresholds. One bout of 38 calls was missed during loudness ranking (Table A3.1) and was eliminated from analysis of loudness ranking. We also tested the relationship between score threshold and each of the sound energy measurements for true positive detections using a linear mixed effects model with file name as a random effect. We compared each model to a null model using Akaike’s information criterion (AIC).

There was a clear relationship between recall and both sound energy metrics, suggesting that recall is related to detection distance (Figure A3.1). Furthermore, the model including sound energy had an AIC weight of 1.0 for both metrics (loudness ranking: $\Delta AIC = 283.1$, RSL: $\Delta AIC = 57.8$).

We conclude that the recall of our recognizer is simply due to a lower detection radius than a human observer, which is concordant with previous studies (Knight and Bayne 2018). We

suggest the recording-level recall of the recognizer is much higher, however, due to the high number of calls that Eastern Whip-poor-wills produce (Mulaik et al. 1953).

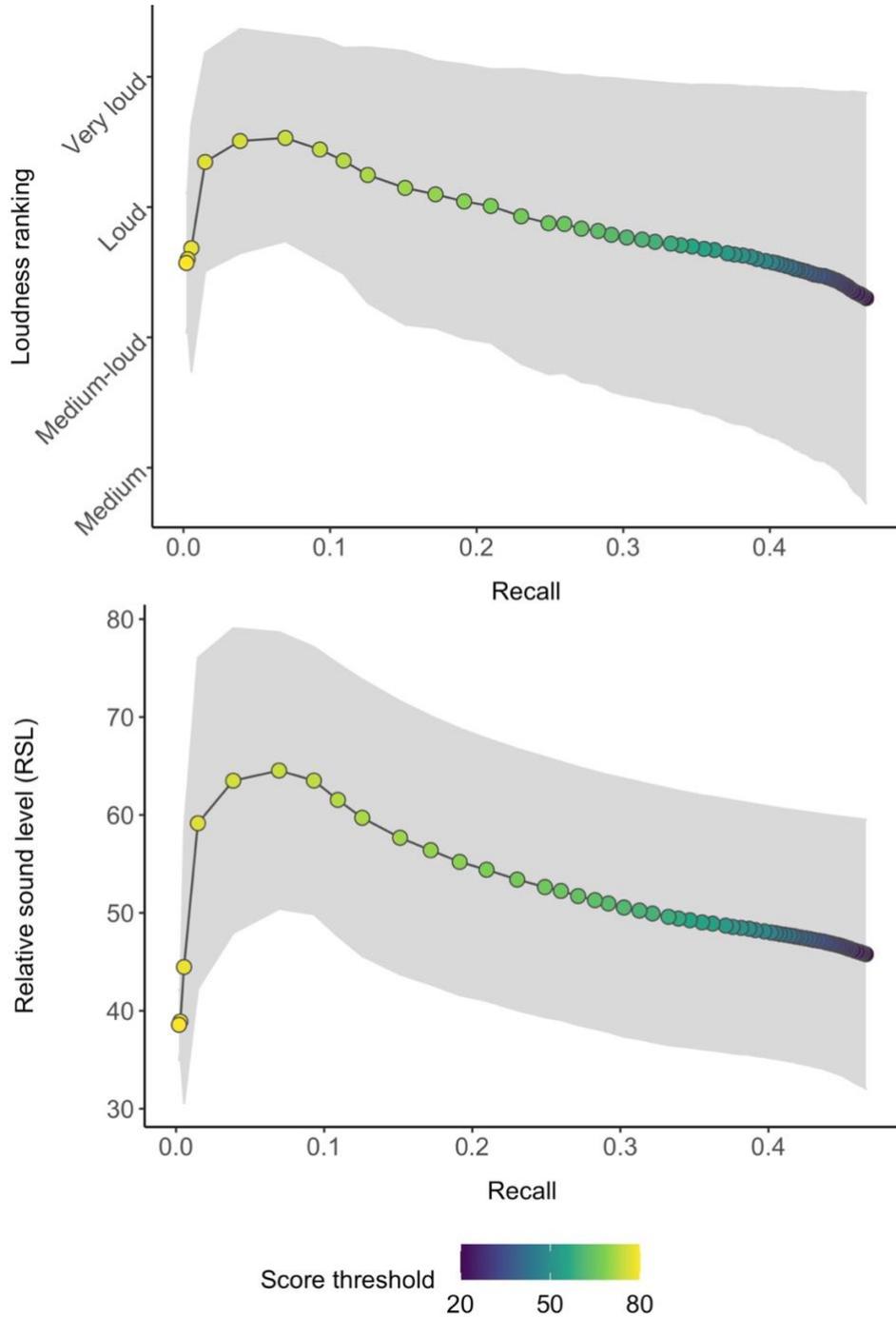


Figure A3.1. Relationship between recognizer recall and two metrics of sound energy (loudness ranking, relative sound level) across a range of score thresholds for an Eastern Whip-poor-will recognizer built in Song Scope software. Points are the mean value for each score threshold and grey shading is the standard deviation.

Literature Cited

- Hedley, R. W., S. J. Wilson, D. A. Yip, K. Li, and E. M. Bayne. 2020. Distance truncation via sound level for bioacoustic surveys in patchy habitat. *Bioacoustics* 30:1–21. doi: 10.1080/09524622.2020.1730240.
- Knight, E. C., and E. M. Bayne. 2018. Classification threshold and training data affect the quality and utility of focal species data processed with automated audio-recognition software. *Bioacoustics* 28:539–554. doi: 10.1080/09524622.2018.1503971.
- Mulaik, S., W. Lee, and J. E. Stillwell. 1953. Analysis of the Call of the Whip-Poor-Will. *Auk* 70:367–368.
- Yip, D. A., E. C. Knight, E. H. Audet, S. J. Wilson, C. Charchuk, C. D. Scott, P. Solymos, and E. M. Bayne. 2019. Sound level measurements from audio recordings provide objective distance estimates for distance sampling wildlife populations. *Remote Sensing in Ecology and Conservation* 71:2759–15. doi: 10.1002/rse2.118.