# Using acoustic recording units to remotely monitor breeding status of an elusive yet vocal forest songbird

Emily J. Upham-Mills1\*

Andrew D. Crosby1

Jody R. Reimer2

Samuel Haché3

Tara Stehelin4

Erin M. Bayne1

1 Department of Biological Sciences, University of Alberta, Edmonton, AB, T6G 2E9, Canada

2 Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA

3 Environment & Climate Change Canada, Yellowknife, NT X1A 2P7, Canada

4 School of Science,Yukon University, Whitehorse, YT, Y1A 5K4, Canada

\* corresponding author: [uphammil@ualberta.ca](mailto:uphammil@ualberta.ca)

## Abstract

To be completed once analysis and conclusions finalized

## Introduction

The primary functions of singing in male songbirds during the breeding season are to defend a breeding territory and to attract and maintain mates (Catchpole & Slater, 2008; Collins, 2004). Singing rates of males vary throughout the breeding season, possibly reflecting a shift in function from primarily intersexual functions, such as attraction of males or other females for copulation, to intrasexual functions such as territory maintenance, mate guarding, and male-male competition (Lampe & Espmark, 1987; Otter & Ratcliffe, 1993). Specifically, males tend to sing at high rates when unpaired and lower rates following pairing (Gibbs & Wenny, 1993; Robbins et al., 2009; Stacier et al., 1996), with gradual decreases in song rate as birds progress from nest building through incubating and feeding young, although this pattern can depend on time of day (Hayes et al., 1986; ; Upham-Mills et al., 2020; Wright, 1997). This variation in singing rate results in seasonal and diurnal temporal variation in detectability of birds by human observers conducting acoustic surveys (Johnson, 2008; Nichols et al., 2009). While researchers have quantified singing rates of forest songbird males throughout the breeding season based on acoustic surveys (Sólymos et al., 2013), the potential of using such data to predict breeding phenology and success has only begun to be explored (Staicer et al., 2006; Upham-Mills et al., 2020). The ability to monitor breeding status in individual birds using acoustic data would fill important data gaps in our understanding of population dynamics of most forest songbirds by greatly increasing sample sizes and thus inform species status assessments in a more quantitative approach (e.g. Environment and Climate Change Canada 2016; Environment and Climate Change Canada 2018). This could be especially valuable for species that are difficult to study by traditional methods because they are rare, occupy territories in challenging terrain, or have cryptic reproductive behaviours.

A hierarchical modelling framework has been used to predict breeding status classes (single, paired, and feeding young) of male Olive-sided Flycatchers (*Contopus cooperi*; OSFL) as a function of date, time of day, and song rate (Upham-Mills et al., 2020). Results from this study using human-based field observations (hereafter “field-based model”) suggested that the breeding status–song rate relationship (hereafter “BSSR relationship”) can be used to monitor demographics for this Schedule A *Threatened* (since 2007) Species at Risk in Canada (Environment and Climate Change Canada, 2016). However, field-based models require song rate data from focal individuals tracked multiple times throughout the breeding season, which can be time consuming and costly. Thus, these datasets of demographic studies are generally limited to small spatial extents (e.g. Amrhein, Korner, and Naguib 2002; Dussourd and Ritchison 2003; Haché, Villard, and Bayne 2013) and low numbers of years (but see Sillett and Holmes 2002). The presence of human observers may also alter the behaviour of breeding males, although this bias remains poorly quantified (Bye et al., 2001; Gutzwiller et al., 1994, Campbell & Francis, 2012).

The advancement of autonomous recording unit (ARU) technology may provide a cost-effective way to quantify the BSSR relationship over large spatial and temporal scales by allowing the monitoring of breeding males with limited effort from human observers (e.g. Pankratz et al., 2017; Shonfield & Bayne, 2017; Yip et al., 2017). ARUs can be preprogrammed to record for long time periods, on specific days of the year and times of day, and simultaneously in many locations. They also likely do not alter the behaviour of birds being monitored once they are set up (Darras et al., 2018). Long-term monitoring programs have started using ARUs to document changes in distribution, population trends, and demographic parameters such as peak nest initiation in bird populations based on singing or calling behaviour (Colbert et al., 2015; Digby, Towsey, Bell, & Teal, 2013), but, to our knowledge, these studies have not monitored the breeding status of individual breeding birds. Although ARUs have the potential to passively produce abundant data on singing behaviours, there are uncertainties about how bird movement affects detectability (Matsuoka et al. 2012; Yip et al., 2020), especially when song rates are derived from recordings collected from permanent sampling locations. This could be a potential concern for species with very large breeding territories, such as the OSFL (Stehelin, 2020).

The goal of this study was to evaluate the ability of the BSSR relationship model to predict breeding status of male OSFLs using song rate data collected from ARUs deployed at multiple sites across the western boreal portion of Canada. We counted songs in ARU recordings by visually scanning spectrograms and modified a hierarchical field-based BSSR model (Upham-Mills et al., 2020) to fit ARU-based data. We anticipated decreased performance of the ARU-based model relative to the field-based model due to imperfect song detection caused by bird movements beyond the detection radius of ARUs deployed at fixed locations. However, we also suspected some or all of the negative effects of imperfect detection might be counterbalanced by much larger volumes of data available to build the model, possibly resulting in an equally or better performing model. We generated four predictive hierarchical models with different adjustments for imperfect detection using: 1) no adjustment for imperfect detection (hereafter the “no-adjustment model”); 2) a random effect to adjust for site-specific acoustic conditions (i.e., ARU placement at the site relative to average bird distance, weather, interspecific acoustic activity) (hereafter the “detection-bias model”), 3) only song rates with known presence (hereafter the “zero-truncated model”); and 4) a two-step model where the probability of detection is first modelled followed by the song rate model (hereafter the “zero-inflated model”). If the detection from ARUs at fixed locations resembles the detection by human observers, then the no-adjustment model should perform similarly to the field-based model (Upham-Mills et al., 2020). Alternatively, if imperfect detection of the birds in ARU recordings significantly alters song counts, models accounting for imperfect detection should have improved predictive performance. In both cases, the larger volume available from ARUs versus field observations should improve model performance. We assessed model performance by comparing sensitivity and specificity values for predicting each breeding status.

## Methods

### Study species

We used OSFL as a focal species to test the use of the BSSR relationship with ARU data for three main reasons. First, the OSFL song is highly recognizable. It is a loud three-note whistle comprising one short soft note at a lower pitch and two higher-pitched accented and drawn out notes, often described by the mnemonic “quick, three-beers” (Altman, 1997; Taverner, 1928) (Figure 1). This song also has little variation among individuals (Robertson et al., 2009), making them easy to detect and distinguish from other species of songbirds. Secondly, this song carries farther in the forest than most songbirds (effective detection radius = 121.3 ± 16.9 m; maximum detection distance = 400 m; Matsuoka et al. 2012, or 624 m; Wright 1997), which increases detection probabilities. Lastly, territories are typically large and non-overlapping (Altman & Sallabanks, 2012), unlike many songbirds with smaller territory sizes, precluding challenges associated with distinguishing between individuals.

### Study area and sampling design

ARU recordings and breeding status data were collected from territorial OSFL males in the western boreal forest of Canada. Specifically, data were collected in northern Alberta (57°29’ N, 111°27’W; n = 16) in 2016 and 2017, the Northwest Territories (61°50’ N, 116°40’W; n = 16) in 2016, and Yukon (60°43’ N, 135°4’W; n = 9) in 2017. Sites were selected based on previous studies confirming presence of multiple OSFLs in the areas (Knaggs et al., 2020; Pardieck et al., 2016, Stehelin, 2020; E. Bayne, *unpublished data*,). Breeding territories in Alberta and the Northwest Territories were situated in bog (dominated by black spruce, *Picea mariana*)and fen wetlands (dominated by black spruceand tamarack, *Larix laricina*) and upland stands dominated by jack pine (*Pinus banksiana*) or a mix of jack pine and trembling aspen (*Populus tremuloides*). Many of these sites had burned in 2014 and 2015. In Yukon, breeding territories were found in white spruce (*Picea glauca*) dominated forests, with some black spruce and intermittent lodgepole pine (*Pinus contorta*) and subalpine fir (*Abies lasiocarpa*) at higher elevations. Some Yukon sites also contained varying amounts of trembling aspen and balsam poplar (*Populus balsamifera*).

Breeding status and ARU recording data were collected at 41 OSFL territories between May 11 and August 10 in 2016 and 2017. Five of the territories were monitored in both years, for a total of 46 unique territory-years, hereafter called territories. Similarly, we refer to unique bird-year combinations as birds. Male OSFLs were observed approximately once per week, for at least 20 minutes per observation, to confirm breeding statuses (i.e., single, paired, and feeding young). Breeding status was confirmed from behavioural evidence, such as observing only one OSFL present on a given day and all past observation days (i.e., single), both members of a pair building a nest (i.e., paired) or an adult carrying food (i.e., feeding young; see Upham-Mills et al. 2020 for more information). Data from days when breeding status could not be determined were removed from the analysis. Acoustic recordings were collected using Song Meters (SM2+, SM3, and SM4; Wildlife Acoustics Inc., Maynard, MA, U.S.A.) deployed in territories after males were considered to have settled (i.e., a male was found singing in the same area on at least two separate occasions between May 10 and June 8). One ARU was deployed per territory near a primary singing perch. OSFLs typically have 2-3 primary singing perches on a territory, where males were more often observed singing (Wright, 1997). ARUs were attached at a height of approximately 1.5 m on trees with a diameter less than the width of the ARU. Each ARU was preprogrammed to record daily, for between 10 and 60 minutes per hour between 45 minutes before sunrise until 15 minutes after sunrise, to maximize OSFL song detection probability (Wright, 1997). ARUs recorded in stereo at a sampling rate of 44.1 kHz and a 16-bit resolution. Average deployment date for all ARUs across both years was June 7 (+ 12 days; SD) and units recorded for an average of 68 days (+ 22 days; SD).

### Acoustic data processing

We used a two-minute song count sample interval to maximize the extent of data being processed, while maintaining an adequate probability of detection and song rate accuracy. Wright (1997) found marginal improvements in detectability between a 5-minute and a 3-minute listening interval, and we found little difference in mean song rate when counting songs in 2 minutes of an ARU recording versus 4 or 5 minutes (E. Upham-Mills, *unpublished data*).

A random subsample of 4,500 2-minute recordings was selected for processing, stratified by breeding status: 1,500 recordings when an OSFL was confirmed as single, 1,500 when paired, and 1,500 when feeding young. We also stratified by bird, to maximize the subsamples from each individual OSFL. We processed the ARU recordings by visually scanning spectrograms, which is an efficient and accurate approach for processing recordings (Digby et al., 2013; Joshi et al., 2017), particularly when assessing temporal patterns in call frequency (Swiston & Mennill, 2009). We used R software (R Core Team, 2020) and the package ‘seewave’ (Sueur et al., 2008) to create spectrograms representing 2 minute recording segments. For ease and efficiency of processing, a spectrogram grid approach was used (J. Kennedy, *in prep*; Figure 1). Each spectrogram in the grid had a frequency range of 1.5 - 5.5 kHz and used a short time Fourier transform window length of 1000. Spectrograms were combined into a JPEG file with dimensions of 13 x 8 inches and a resolution of 500, and were visually assessed on a 27 inch computer monitor.

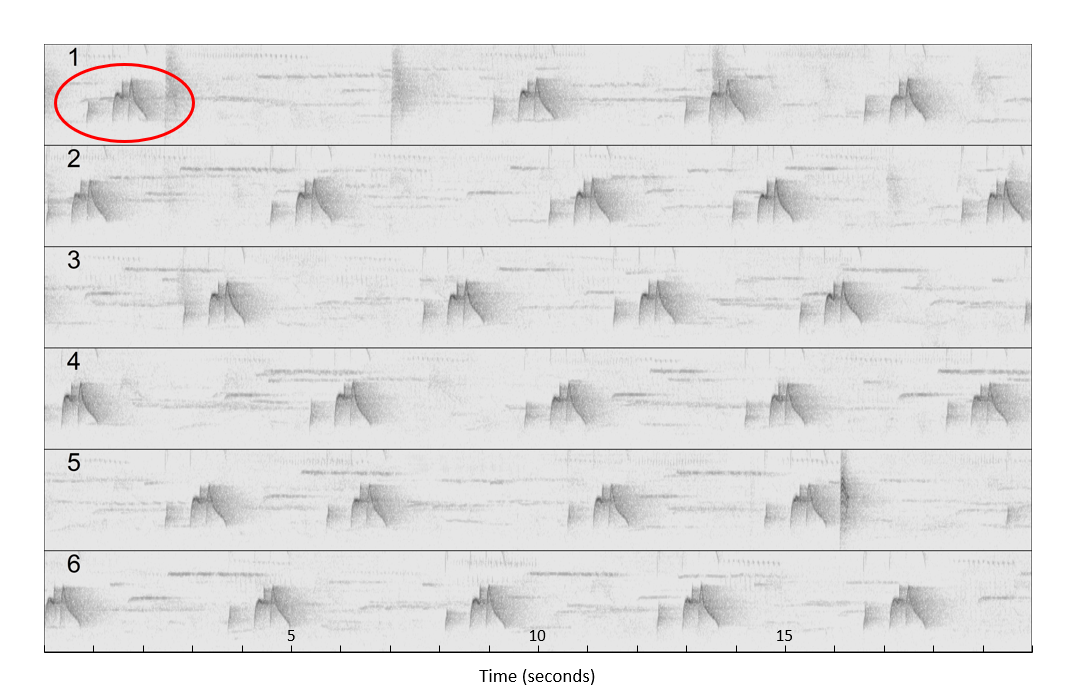


Figure 1. Example of a 2-minute recording represented as a single spectrogram grid of six consecutive 20-second panels. A single OSFL song is circled in red.

One experienced observer (EU) scanned all recordings and counted every visible OSFL song. If a sound in the spectrogram resembled that of an OSFL song but required confirmation, the observer listened to the section of recording containing the candidate song to identify the sound. When a second OSFL was detected in the spectrogram, only the OSFL nearest to the ARU (i.e., the one with the loudest song), was counted.

Following spectrogram processing, we calculated full-season detection rates for each bird, i.e., the total sample days with at least one song detected divided by the total number of sample days. We then determined a threshold full-season detection rate to remove birds with low detection rates under the assumption that they were not actively defending an area overlapping the detection radius from the ARU.

### Statistical analysis

Prior to creating predictions using the hierarchical models, we first conducted data exploration to determine how song rate was affected by breeding status, time past sunrise, and random effect of the site using ARU-based song counts. We compared a generalized linear model, a generalized linear mixed model, a zero-inflated model, and a zero-truncated model where the response variable was song rate using a Poisson distribution. We used Akaike’s Information Criterion (AIC) to rank model fit for each model and considered a decrease in AIC >2 as evidence for improved fit. We also explored coefficient values and significance (α level of *P* < 0.05) of each predictor to compare their relative fit to the data.

For all hierarchical models, we used a parametric empirical Bayesian approach and priors were derived from maximum likelihood models fit using the ARU data (Cressie et al., 2009; Morris, 2012; Ver Hoef, 1996). Models were built using a hierarchical framework with two component models adapted from the following field-based model (Upham-Mills et al., 2020):

**Equation 1.**



where the left-hand side probabilities for each breeding status (*BS*) conditional on song rate (*SR*), day (*D*) and time (*T*), are proportional to the product of two components on the right-hand side. Component A (i.e., the ecological process component) is the marginal probability of observing each breeding status on a given day, and component B (i.e., the observation component), is the probability of song rate conditional on breeding status and time of day.

The ecological process component model assumed that breeding status was a function of ordinal day, where breeding status could be one of: single, paired, or feeding young. We considered each bird-day combination to be an independent sample, and so the component A model was:

**Equation 2.**

*BSi* ~ *multinomial*(**p***i,*1:3),

where *BSi* is the breeding status for bird-day *i*, and **p***i,*1:3 is the vector of probabilities for each breeding status. Each probability, 𝛼, was log(𝛼) = 𝜇 + 𝜇1*D*, where the 𝜇’s are the vector of coefficients and *D* is ordinal day.

The four models, differing in how they deal with imperfect detection (i.e., no adjustment, detection-bias, zero-inflated, and zero-truncated), used the same component A but used variations of the observation component of the hierarchical model. We will first describe the simplest version of this component, in the no-adjustment model, then describe how the other three models differed from it. The no-adjustment observation model connected singing rate to breeding status, so that song count within a given period of time on bird-day *i, Ni*, was:

**Equation 3.**

*Ni* ~ *Poisson* (λ*i*), and

**Equation 4.**

log (λ*i*) = 𝛽0*BSi* + 𝛽1*T*,

where λ*i* is expected song count, 𝛽0*BSi* is the coefficient for breeding status *i*, and *T* is time in hours relative to sunrise. Because multiple song count subsamples were often taken on the same day, we created a separate observation model for each subsample, so that the song count for each bird-day *i* at subsample *j*, *Nij*, had a specific λ*ij* as a function of breeding status and time relative to sunrise.

The detection-bias model incorporated a random effect () for the site in the observation component to incorporate variability due to site-specific characteristics, such as individual bird behaviour, ARU placement on the territory, and local condition:

**Equation 5.**

log (λ*i*) = 𝛶bird,i + 𝛽0*BSi* + 𝛽1*T*,

A disproportionate number of zeros in the song counts can arise from two processes: 1) birds are within an ARU detection range and silent or 2) birds are out of detection range of the ARU and processing method. The zero-inflation model adjusted for this imperfect detection by first modelling the probability of detection as a Bernoulli distribution, then modelling detected song rate conditional on probability of detection:

**Equation 6.**

**Equation 7.**

followed a uniform distribution from 0 to 1, followed by detected song rates modelled as a Poisson distribution with mean λi.

The zero-truncated model adjusted for imperfect detection by constraining song rate analysis to sampling sessions when a bird was confirmed to be present (i.e., at least one song). We therefore removed cases where song rate was zero and used a zero-truncated Poisson distribution for .

The Monte Carlo Markov Chain (MCMC) specifications for all models used a burn-in period of 1,000 iterations, followed by 10,000 iterations, on three parallel chains. All models were built using JAGS language (package ‘jagsUI’; Kellner, 2019). We monitored for convergence using the Gelman-Rubin diagnostic (GR), considering a threshold of GR > 1.1 to indicate convergence (Gelman & Rubin, 1992; Plummer et al., 2016).

We compared the relative predictive ability of the four predictive hierarchical models using a modified K-fold cross-validation (Arlot & Celisse, 2010), where for each fold we removed all data associated with one bird. The subset of data removed from the dataset (i.e. the test data) was then used for predictions from the model built using the remaining dataset (i.e. the training data). This process was repeated until each bird had been removed, for a total of predictions from 41 datasets for each model. We used K-fold validation because it is a common method for accuracy estimation of classifiers (Kohavi, 1995). We assessed prediction accuracy (or predictive ability) by comparing true (i.e. sensitivity) and false (i.e. specificity) positive rates for each breeding status from each model. Values were displayed in a contingency table (i.e. confusion matrix) built using the package ‘caret’ in R (Kuhn et al. 2017). We emphasized the importance of the ability of each model to predict single and feeding young correctly as this information can reflect important differences in habitat quality among breeding territories.

We sourced parameter estimates and standard errors for the empirical priors for each model and conducted preliminary data exploration using the r packages ‘stats’, ‘lme4’, ‘glmmTMB’, and ‘VGAM’ (Bates et al., 2015; Brooks et al., 2017; R Core Team, 2020; Yee et al., 2015).

## Results

Approximately 512 hours of recordings were collected, equating to 15,640 potential 2-minute recording samples. Following spectrogram processing of 4,500 samples, we determined a full-season detection rate threshold of 30% was a natural break in the data to separate birds detected on a regular basis throughout the sample period from those that were detected intermittently. This process resulted in the removal of 5 birds and updated the data set to 924 single samples, 1458 paired samples, and 1445 feeding young samples. The final sample comprised 3827 recordings from 41 birds, with an average of 93 (+ 35; SD) 2-minute recordings per bird.

Regardless of the model used to explore the effects of breeding status and time relative to sunrise on song count, the three breeding status classes and time were always significant predictors of song rate (Table 1). According to AIC values, the zero-inflated model fit the data best when using data including the zeros. The order of the highest to lowest song rate as a response to the three breeding statuses changed when a site random effect was added. In the fixed effects model, the song rates were predicted to be the highest when birds were paired, then when single, and lowest when feeding young (Table 1). The beta-coefficients for breeding status generally increased with every iteration of the model, particularly after dealing with zeros through zero-inflation or truncation.

Table 1. Beta-coefficients, standard errors and p-values from generalized linear regression models describing the effect of breeding status and other covariates on song rate from ARU recordings. All models used a Poisson or distribution for song rate.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Predictor variables** | | | |  |
| **Associated Hierarchical Model** (and type of model) | Hours from Sunrise | **Breeding Status:**  Single | **Breeding Status:** Paired | **Breeding Status:** Feeding Young | AIC value |
| **No-adjustment**  (generalized linear model) | -1.78\* | 1.54 + 0.01\* | 1.92 + 0.01\* | 0.90 + 0.01\* | 62492 |
| **Detection-bias**  (generalized linear mixed effects model) | -1.80\* | 1.96 + 0.16\* | 1.51 + 0.14\* | 0.49 + 0.14\* | 46599 |
| **Zero-inflated**  (zero-inflated mixed effects model) | -1.28\* | 2.46 + 0.08\* | 2.33 + 0.08\* | 1.92 + 0.07\* | 23651 |
| **Zero-truncated♰**  (zero-truncated model) | -1.31\* | 2.47 + 0.01\* | 2.50 + 0.01\* | 2.12 + 0.02\* | - |

\* indicates significance at α of *P* < 0.05.

**♰** Zero-truncated mixed effects model R packages were only available using an Bayesian approach, so no mixed effects were used in the zero-truncated model. No AIC was used because it was not comparable to other models due to using a different (i.e., truncated) data set.

The four hierarchical models based on ARU data rarely predicted the breeding status single correctly (sensitivity <0.07) compared to models based on field-based song rates (sensitivity = 0.69; Table 3.2). ARU models had good predictive accuracy for paired (sensitivity ranging from 0.77 – 0.84) compared with the field-based model (0.50). However, specificity values for paired (<0.46) were lower than those for the other breeding statuses (specificity > 0.89; Table 3.2). The no adjustment and zero-inflated models had high sensitivities for feeding young (0.68 and 0.61, respectively), while the zero-truncated model performed poorly, with the lowest sensitivity (0.25) of all models for predicting feeding young.

Overall, the best predictive model was the no adjustment model, which had the highest mean sensitivity (0.52 + 0.39) for all three breeding statuses versus values of 0.34 (+ 0.39), 0.49 (+ 0.41), and 0.44 (+ 0.41) for the zero-truncated, zero-inflated and distance models, respectively. This model also had the highest mean specificity (0.76 + 0.26), followed by the zero-inflated (0.75 + 0.31), distance (0.72 + 0.38) and the zero-truncated (0.62 + 0.49) models.

All ARU models followed similar patterns of over and under-predicting the occurrence of specific breeding statuses (Fig. 3.3). The no adjustment, zero-inflated and distance models all overpredicted paired, whereas the zero-truncated model predicted similar prevalence of paired statuses compared to observed breeding status prevalence. This is different from the field-based model, which underpredicted paired by more than 30%. All four ARU models underpredicted single by ≥ 10%, which was opposite to the field-based model. The zero-truncated model was the only model to overpredict feeding young, as did the field-based model.

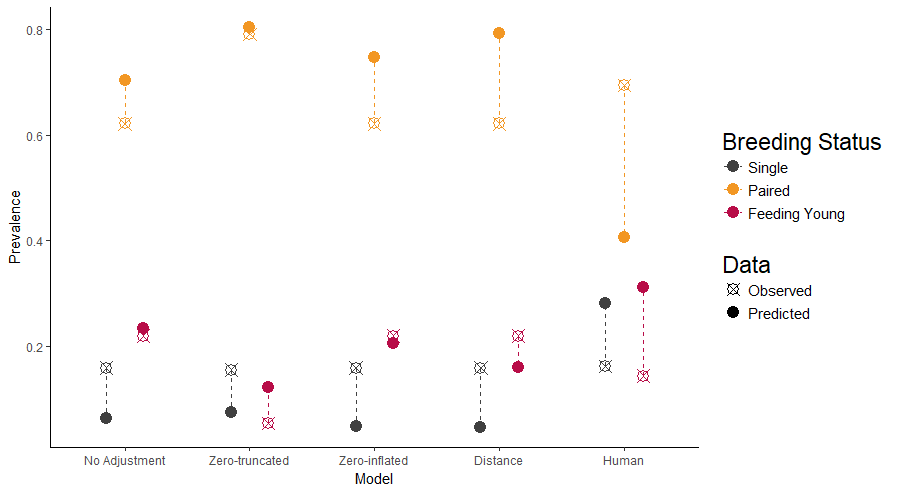


Figure 3. 3. Prevalence of observed versus predicted breeding statuses for four ARU data models.

Table 3. 2. Sensitivity and specificity values for predicting breeding statuses using ARU-based song rate data in four hierarchical models as compared to a field-based song rate hierarchical model (Chapter 2).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Adjustment for imperfect detection | Prediction Evaluation Value | Single | Paired | Feeding  Young |
| No adjustment | None | Sensitivity | 0.07 | 0.80 | 0.68 |
| *Specificity* | *0.94* | *0.46* | *0.89* |
| Zero-truncated | Removed zeros from dataset | Sensitivity | 0.00 | 0.77 | 0.25 |
| *Specificity* | *0.91* | *0.06* | *0.89* |
| Zero-inflated | Modelled song detection/non-detection first, then song rate | Sensitivity | 0.04 | 0.83 | 0.61 |
| *Specificity* | *0.95* | *0.39* | *0.91* |
| Distance | Model song detection/non-detection first, then song rate as a function of mean level (i.e. relative sound level) | Sensitivity | 0.03 | 0.84 | 0.46 |
| *Specificity* | *0.95* | *0.28* | *0.92* |
| Human-based song rate model  (Chapter 2) | None (no need because observers always confirmed bird presence) | Sensitivity | 0.69 | 0.50 | 0.87 |
| *Specificity* | *0.80* | *0.82* | *0.78* |

**Discussion**

In this study, song counts derived from ARU recordings processed with automatic recognition were much lower than those recorded by human observers in the field. We expected that a BSSR model using song count data from ARUs would be constrained by limited detectability, i.e. values of zero could reflect a silent bird or a bird singing beyond the perceptual range of an ARU deployed at a permanent location. However, even after adjusting ARU data to account for imperfect detection through zero-truncation, zero-inflation, and by adding sound level as a proxy for distance, predictive performance of ARU-based models did not improve. These results suggest that the current BSSR model with ARU data has important limitations and improvements in both technology and sampling design are warranted. Still we do propose that the ARU is a promising tool to monitor population dynamics of forest songbird over large spatial extent in a cost-effective way to inform conservation actions.

Detection of sound on an acoustic recording can be inhibited by multiple extrinsic factors related to the location of the ARU: 1) the structure of the environment affecting sound attenuation (Yip et al., 2017); 2) amount of precipitation and wind intensity and direction; and 3) species richness and activity levels of the bird community within the range of an ARU. An additional factor affecting song rates detected from an ARU is density of conspecifics. This can artificially inflate song rate for the target male when neighbouring male songs are captured on the ARU. However, this challenge can be overcome if individual recognition is possible for the species of interest, which is increasingly available with recent technological advancements (Ehnes & Foote, 2015; Foote et al., 2013; Kirschel et al., 2011; Wilson, 2018). Increased density can also change male singing behaviour by countersinging behaviour producing higher song rates in the target individual, independently of breeding status (Penteriani et al., 2002). We assumed multiple conspecifics in an area would not be an issue for OSFL due to large, non-overlapping territories, but some of our ARUs did contain songs from multiple individuals. Future studies using ARU data in BSSR models should account for these different sources of variations which may improve predictive ability.

Another important source of detection error is low detectability of the recognizer. The OSFL recognizer used in this study (McLeod, 2015) performs well when the objective is to identify presence of an OSFL at a given ARU. Individuals were detected by the recognizer for each ARU where a human interpreter confirmed the presence of an individual (Bayne and McLeod, unpublished data). However, the recognizer misses >80% of the songs present in the recordings (i.e. false negative rate), when compared to the performance of human interpreters (Appendix 5). If predicted song rates for single are ~20-fold lower for ARU data compared to field-based data, as with the false-negative rate we observed, then the difference in song rates could be much closer (i.e. 4-fold lower) if recordings were processed by human listeners instead of a computer algorithm. Thus, adjusting the recognizer to achieve higher detectability may greatly improve sensitivity to predict breeding status. We only tested one automated recognition software currently available. Many more exist (e.g. Raven Pro [Cornell Laboratory of Ornithology, Ithaca, New York, USA], R package “monitoR” [Hafner and Katz 2018], and Kaleidoscope Pro [Wildlife Acoustics, Maynard, Massachusetts, USA]) and statistical models are quickly evolving to minimize false negative rates (Knight et al. 2017). For example, Chambert et al. (2018) present a hierarchical model to use an automatic recognition algorithm to first detect the target species, then incorporate a subset of *post-hoc* validated data to optimize processing and improve recognizer accuracy. Lower false negative rates can also be achieved by lowering the score threshold of a recognizer to increase sensitivity to candidate detections (Knight et al., 2017), but this comes at the efficiency expense of having a considerably larger number of detections to validate. We recommend future users of the BSSR model should consider either the use of human interpreters (e.g. Joshi, Mulder, and Rowe 2017) or evaluate the available species recognizers (see Knight et al. [2017] for recommendations) to ensure the lowest false negative rate in song detection.

The difference in predicted song count between single and paired males based on ARU data was negligible compared to the differences detected by human observers. We suspect this is the reason that all hierarchical models based on ARU data had a low rate of correctly predicting a status of single. Single males are expected to sing significantly more than paired males, therefore differences in song counts between ARU and human observer data should be explained by differences in detectability generated by bird movements. Our results suggest a larger propensity for single birds to move over larger areas than paired birds. There is evidence for differential movement associated with changes in breeding status in songbirds (e.g. Bayne and Hobson 2001; Liu, Kroodsma, and Yasukawa 2007). We posit that single males searching for a mate may not only sing at higher rates but may spend more time covering their entire territories and beyond. Potential differences in movement patterns associated with breeding status should be quantified acoustically and these variables could be integrated into the hierarchical models. For example, the mean relative sound level (i.e. what we used as a proxy for average distance from the ARU) could be added to the multinomial component of the hierarchical model (i.e. instead of the song count regression component), where breeding status is predicted by ordinal date, to use as a proxy for movement around the territory. However, the importance of variation in detectability related to movement across breeding statuses is likely species-specific and/or a function of territory size. For example, the BSSR approach using a single ARU may be more suited to smaller passerines that tend to defend territories of < 1 ha (e.g. Red-eyed Vireo, *Vireo olivaceus* [Marshall & Cooper 2004]; Cerulean Warbler, *Setophaga cerulea* [Robbins et al., 2009]; Golden-winged Warbler, *Vermivora chrysoptera* [Streby et al., 2012]). Fortunately, technology is available to calibrate the effect of breeding status on bird movement, and movement on detectability. Options include either increasing acoustic coverage using an array of ARUs (Blumstein et al., 2011), allowing acoustic localization of birds (Wilson, 2018), tracking bird movements using telemetry or GPS technology, or using a combination of microphone and GPS (e.g. lightweight GPS-ARUs have been used to simultaneously track movement and vocalizations in Common Nighthawks, *Chordeiles minor*, Knight et al., unpublished data).

We have identified challenges that need to be addressed to account for ARU-related detection error. Thus, ARU-based models might not be currently appropriate to predict and monitor the breeding status of an individual based on song counts. However, the BSSR model is well suited to integrate larger acoustic datasets that cover a large number of breeding territories to predict proportion of territories that have achieved more advanced breeding statuses (e.g. feeding young) and, perhaps, infer differences in habitat quality. For example, hundreds of ARUs are being deployed annually across Alberta (Alberta Biodiversity Monitoring Institute 2012, www.abmi.com) since 2003, and in the Northwest Territories since 2013 (Haché & Pankratz, unpublished data). Information on breeding phenology and where large proportions of OSFL territories produce young would provide important demographic information given the status and population trends of this species is largely unknown in northern boreal regions owing to the sparse coverage of the North American Breeding Bird Survey (Machtans et al., 2014). Thus, already available multi-year ARU data covering large spatial extents could be processed to provide information on breeding success for this Species at Risk in Canada to help the status reassessment process (Environment and Climate Change Canada, 2016).

This is, to our knowledge, the first study demonstrating how song rate from ARUs data processed with automatic recognition software can predict breeding status of a forest songbird. The hierarchical modelling approach we used provides a flexible framework to include additional parameters to take full advantage of the behavioural and detectability information available in acoustic recordings. ARUs are being used across North America to monitor bird occurrence or abundance (Shonfield & Bayne, 2017), but this new tool can potentially be applied to monitor population dynamics of forest birds for a broad range of species and ecosystems and could revolutionize how avian ecologists monitor populations. Bioacoustics technology is improving quickly, which provides great opportunity for researchers to document, in a cost-effective way, much needed information about population dynamics for status assessments and inform conservation initiatives to address the growing global biodiversity crisis (Singh, 2002).

## Author’s contributions

EU, EB, and SH conceived the ideas and designed the field sampling methodology; EU and TS collected the data; JR, AC and EU conceived the ideas for the statistical analysis methodology; EU processed the data; EU and AC analyzed the data; EU led the writing of the manuscript. All authors contributed critically to the drafts.

# Literature Cited

Alberta Biodiversity Monitoring Institute. (2011). *Monitoring activity report: lower Athabasca planning region 2010 season. Biodiversity Monitoring Report.* Edmonton, AB, Canada.

Altmas, B. (1997). Olive-sided Flycatcher in western North America Status Review. U.S. Fish and Wildlife Service, Portland, Oregon, USA.

Altman, B., & Sallabanks, R. (2012). *Olive-sided Flycatcher (*Contopus cooperi*), version 2.0*. (P. G. Rodewald, Ed.), *The Birds of North America*. Ithaca, New York, USA: Cornell Lab of Ornithology.

Amrhein, V., Korner, P., & Naguib, M. (2002). Nocturnal and diurnal singing activity in the nightingale: Correlations with mating status and breeding cycle. *Animal Behaviour*, *64*(6), 939–944. https://doi.org/10.1006/anbe.2002.1974

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistical Surveys*, *4*, 40–79. https://doi.org/10.1214/09-SS054

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Bayne, E. M., & Hobson, K. A. (2001). Movement patterns of adult male Ovenbirds during the post-fledging period in fragmented and forested boreal landscapes. *The Condor*, *103*(2), 343–351. Retrieved from http://www.bioone.org/doi/full/10.1650/0010-5422%282001%29103%5B0343%3AMPOAMO %5D2.0.CO%3B2

Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., … Ali, A. M. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*, *48*, 758–767. https://doi.org/10.1111/j.1365-2664.2011.01993.x

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Maechler, M., & Bolker, B.M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. The R Journal, 9(2), 378-400.

Bye, S. L., Robel, R. J., & Kemp, K. E. (2001). Effects of human presence on vocalizations of grassland birds in Kansas. *Prairie Naturalist*, *33*, 249– 256.

Campbell, M., & Francis, C. M. (2012). Using microphone arrays to examine effects of observers on birds during point count surveys. *Journal of Field Ornithology*, *83*(4), 391–402. https://doi.org/10.1111/j.1557-9263.2012.00389.x

Catchpole, C. K., & Slater, P. J. B. (2008). *Bird song: Biologial themes and variations* (Second). Cambridge, UK: Cambridge University Press. https://doi.org/10.1071/MU952208b

Chambert, T., Waddle, J. H., Miller, D. A. W., Walls, S. C., & Nichols, J. D. (2018). A new framework for analysing automated acoustic species detection data: Occupancy estimation and optimization of recordings post-processing. *Methods in Ecology and Evolution*, *9*(3), 560–570. https://doi.org/10.1111/2041-210X.12910

Colbert, D. S., Ruttinger, J. A., Streich, M., Chamberlain, M., Conner, L. M., & Warren, R. J. (2015). Application of autonomous recording units to monitor gobbling activity by wild Turkey. *Wildlife Society Bulletin*, *39*(4), 757–763. https://doi.org/10.1002/wsb.577

Collins, S. (2004). *Vocal fighting and flirting: The functions of birdsong*. *Nature’s Music: The Science of Birdsong*. Elsevier Inc. https://doi.org/10.1016/B978-012473070-0/50005-0

Cressie, N., Calder, C. A., Clark, J. S., Ver Hoef, J. M., & Wikle, C. K. (2009). Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling, *19*(3), 553–570.

Darras, K., Furnas, B. J., Fitriawan, I., Mulyani, Y., & Tscharntke, T. (2018). Estimating bird detection distances in sound recordings for standardising detection ranges and distance sampling. *Methods in Ecology and Evolution*, 1–11. https://doi.org/10.1111/2041-210X.13031

Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, *4*(7), 675–683. https://doi.org/10.1111/2041-210X.12060

Dussourd, N., & Ritchison, G. (2003). Singing behavior of male yellow-breasted chats: repertoires, rates, reproductive success, and a comparison with other wood-warblers. *The Wilson Bulletin*, *115*(1), 52–57.

Ehnes, M., & Foote, J. R. (2015). Comparison of autonomous and manual recording methods for discrimination of individually distinctive Ovenbird songs. *Bioacoustics*, *24*(2), 111–121. https://doi.org/10.1080/09524622.2014.994228

Environment and Climate Change Canada. (2015). *Recovery strategy for the Common Nighthawk (Chordeiles minor) [Proposed]*. *Species at Risk Act Recovery Strategy Series*. Ottawa.

Environment and Climate Change Canada. (2016). *Recovery strategy for Olive-sided flycatcher (Contopus cooperi) in Canada*. *Species at Risk Act Recovery Strategy Series*. Ottawa.

Environment and Climate Change Canada. (2018). Status of birds in Canada 2014. Retrieved September 6, 2018, from https://wildlife-species.canada.ca/bird-status/index-eng.aspx?sY=2014&sL=e

Foote, J. R., Palazzi, E., & Mennill, D. J. (2013). Songs of the Eastern Phoebe, a suboscine songbird, are individually distinctive but do not vary geographically. *Bioacoustics*, *22*(2), 137–151. https://doi.org/10.1080/09524622.2012.740174

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–511.

Gibbs, J. P., & Wenny, D. G. (1993). Song output as a population estimator: effect of male pairing status. *Journal of Field Ornithology*, *64*(3), 316–322.

Gutzwiller, K. J., Wiedenmann, R. T., Clements, K. L., & Anderson, S. H. (1994). Effects of human intrusion on song occurrence and singing consistency in subalpine birds. *The Auk*, *111*(1), 28–37. Retrieved from http://www.jstor.org/stable/4088502

Haché, S., Villard, M., & Bayne, E. M. (2013). Experimental evidence for an ideal free distribution in a breeding population of a territorial songbird. *Ecology*, *94*(4), 861–869. Retrieved from http://www.jstor.org/stable/23436299

Hafner, S. D., & Katz, J. (2018). A short introduction to acoustic template matching with monitoR.

Hayes, J. P., Probst, J. R., & Rakstad, D. (1986). Effect of mating status and time of day on Kirtland’s warbler song rates. *The Condor*, *88*(3), 386–388. https://doi.org/10.2307/1368892

Johnson, D. H. (2008). In Defense of Indices: The Case of Bird Surveys. *Journal of Wildlife Management*, *72*(4), 857–868. https://doi.org/10.2193/2007-294

Joshi, K. A., Mulder, R. A., & Rowe, K. M. C. (2017). Comparing manual and automated species recognition in the detection of four common south-east Australian forest birds from digital field recordings. *Emu - Austral Ornithology*, *117*(3), 233–246. https://doi.org/10.1080/01584197.2017.1298970

Kellner, K. (2017). Package ‘ jagsUI ’: A wrapper around “rjags” to streamline “JAGS” analyses.

Kirschel, A. N. G., Cody, M. L., Harlow, Z. T., Promponas, V. J., Vallejo, E. E., & Taylor, C. E. (2011). Territorial dynamics of Mexican Ant-thrushes Formicarius moniliger revealed by individual recognition of their songs. *Ibis*, *153*(2), 255–268. https://doi.org/10.1111/j.1474-919X.2011.01102.x

Knaggs, M., Haché, S., Nielsen, S. E., Pankratz, R. F., & Bayne, E. (2020). Avian response to wildfire severity in a northern Boreal region. *Forests*, *11*(12), 1–23. https://doi.org/10.3390/f11121330

Knight, E. C., Hannah, K. C., Foley, G. J., Scott, C. D., Brigham, R. M., & Bayne, E. (2017). Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation and Ecology*, *12*(2).

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*, (March 2001), 1137–1143.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., … Hunt, T. (2017). Package ‘caret’: classification and regression training. Retrieved January 15, 2018, from https://cran.r-project.org/web/packages/caret/caret.pdf

Lampe, H. M., & Espmark, Y. O. (1987). Singing activity and song pattern of the Redwing *Turdus iliacus* during the breeding season. *Ornis Scandinavica*, *18*(3), 179–185. https://doi.org/10.2307/3676764

Lein, M. R. (1978). Song variation in a population of chestnut-sided warblers (Dendroica pensylvanica): its nature and suggested significance. *Canadian Journal of Zoology*, *56*(6), 1266–1283. https://doi.org/10.1139/z78-182

Liu, W., Kroodsma, D. E., & Yasukawa, K. (2007). Dawn and daytime singing behavior of Chipping Sparrows (*Spizella passerina*). *The Auk*, *124*(1), 44–52.

Machtans, C. S., Kardynal, K. J., & Smith, P. A. (2014). How well do regional or national Breeding Bird Survey data predict songbird population trends at an intact boreal site? *Avian Conservation and Ecology*, *9*(1), art5. https://doi.org/10.5751/ACE-00649-090105

Marshall, M. R., & Cooper, R. J. (2004). Territory size of a migratory songbird in response to caterpillar density and foliage structure. *Ecology*, *85*(2), 432–445.

Matsuoka, S. M., Bayne, E. M., Sólymos, P., Fontaine, P. C., Cumming, S. G., Schmiegelow, F. K. a., & Song, S. J. (2012). Using binomial distance-sampling models to estimate the effective detection radius of point-count surveys across boreal Canada. *The Auk*, *129*(2), 268–282. https://doi.org/10.1525/auk.2012.11190

McLeod, L. (2015). Olive-sided Flycatcher Song Scope Recognizer “OSFL\_8\_laprtraining(50\_60).ssr.” Retrieved from http://bioacoustic.abmi.ca/resources/reports/

Morris, C. N. (2012). Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association*, *78*(381), 47–55. Retrieved from https://doi.org/10.1080/01621459.1983.10477920

Müller, C., Jenni-Eiermann, S., Blondel, J., Perret, P., Caro, S. P., Lambrechts, M., & Jenni, L. (2006). Effect of human presence and handling on circulating corticosterone levels in breeding blue tits (Parus caeruleus). *General and Comparative Endocrinology*, *148*(2), 163–171. https://doi.org/10.1016/j.ygcen.2006.02.012

Nichols, J. D., Thomas, L., & Conn, P. B. (2009). *Inferences about landbird abundance from count data: recent advances and future directions. Modeling demographic processes in marked populations*. (D. L. Thomson et al., Ed.). New York: Springer Science+Business Media. https://doi.org/10.1007/978-0-387-78151-8

Otter, K., & Ratcliffe, L. (1993). Changes in singing behavior of male black-capped chickadees (Parus atricapillus) following mate removal. *Behavioral Ecology and Sociobiology*, *33*(6), 409–414. https://doi.org/10.1007/BF00170256

Pankratz, R. F., Hache, S., Sólymos, P., & Bayne, E. M. (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

Pardieck, K. L., Ziolkowski, D. J. J., Hudson, M.-A. R., & Campbell, K. (2016). North American Breeding Bird Survey Dataset 1966 - 2015, version 2015.0. https://doi.org/10.5066/F7C53HZN

Penteriani, V., Penteriani, V., Gallardo, M., Gallardo, M., Cazassus, H., & Cazassus, H. (2002). Conspecific density biases passive auditory surveys. *Journal of Field Ornithology*, *73*(4), 387–391. Retrieved from http://www.bioone.org/perlserv/?request=get-document&amp;doi=10.1648/0273-8570(2002)073[0387:CDBPAS]2.0.CO;2

Plummer, M., Best, N., Cowles, K., Vines, K., Sarkar, D., Bates, D., … Magnusson, A. (2016). Package “coda”: output analysis and diagnostics for MCMC.

R Core Team. (2020). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

Robbins, M. B., S Ny Ri, R. D., Pape, M., Benz, B. W., Papes, M., & Robbins, M. B. (2009). Song rates, mating status, and territory size of Cerulean warblers in Missouri Ozark riparian forest. *The Wilson Journal of Ornithology*, *121*(2), 283–289. https://doi.org/10.1676/08-100.1

Robertson, B. A., Fontaine, J. J., & Loomis, E. (2009). Seasonal patterns of song structure variation in a suboscine passerine. *The Wilson Journal of Ornithology*, *121*(4), 815–818.

Scott Sillett, T., & Holmes, R. T. (2002). Variation in survivorship of a migratory songbird throughout its annual cycle. *Journal of Animal Ecology*, *71*(2), 296–308. https://doi.org/10.1046/j.1365-2656.2002.00599.x

Shonfield, J., & Bayne, E. M. (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

Singh, J. S. (2002). The biodiversity crisis: A multifaceted review. *Current Science*, *82*(6), 638–647.

Skaug, H., Fournier, D., Nielsen, A., Magnusson, A., & Bolker, B. (2018). Package ‘ glmmADMB ’: Generalized linear mixed models using “AD Model Builder.”

Sólymos, P., Matsuoka, S. M., Bayne, E. M., Lele, S. R., Fontaine, P., Cumming, S. G., … Song, S. J. (2013). Calibrating indices of avian density from non-standardized survey data: Making the most of a messy situation. *Methods in Ecology and Evolution*, *4*(11), 1047–1058. https://doi.org/10.1111/2041-210X.12106

Stacier, C. A., Spector, D. A., & Horn, A. G. (1996). The dawn chorus and other diel patterns in acoustic signaling. In D. E. Kroodsma & E. H. Miller (Eds.), *Ecology and evolution of acoustic communication in birds* (pp. 426–453). Ithica, New York, USA: Cornell University Press.

Staicer, C. A., Ingalls, V., & Sherry, T. W. (2006). Singing behavior varies with breeding status of American Redstarts (*Setophaga ruticilla*). *The Wilson Journal of Ornithology*, *118*(4), 439–451. https://doi.org/10.1676/05-056.1

Stehelin, T. 2020. PhD thesis...

Streby, H. M., Loegering, J. P., & Andersen, D. E. (2012). Spot-mapping underestimates song-territory size and use of mature forest by breeding Golden-winged warblers in Minnesota , USA. *Wildlife Society Bulletin*, *36*(1), 40–46. https://doi.org/10.1002/wsb.118

Su, Y., & Yajima, M. (2015). Package ‘ R2jags ’: Using R to run “JAGS.”

Sueur, J., Aubin, T., & Simonis, C. (2008). Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics*, *18*, 213–226. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/09524622.2008.9753600

Swiston, K. A., & Mennill, D. J. (2009). Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. *Journal of Field Ornithology*, *80*(1), 42–50. https://doi.org/10.1111/j.1557-9263.2009.00204.x

Taverner, P.A. 1928. Birds of western Canada. Canada Department of Mines and National Museum of Canada. Bull. No. 41, Biol. Series No. 10. Ottawa. 379 pp.

Upham-Mills, E. J., Reimer, J. R., Haché, S., Lele, S. R., & Bayne, E. M. (2020). Can singing rate be used to predict male breeding status of forest songbirds? A comparison of three calibration models. *Ecosphere*, *11*(1). https://doi.org/10.1002/ecs2.3005

Ver Hoef, J. M. (1996). Parametric empirical Bayes methods for ecological applications. *Ecological Applications*, *6*(4), 1047–1055.

Wilson, S. J. (2018). Use of an acoustic location system to understand how presence of conspecifics and canopy cover influence Ovenbird (Seiurus aurocapilla) space use near reclaimed wellsites in the boreal forest of Alberta. *Avian Conservation and Ecology*, *13*(2), 4. Retrieved from https://doi.org/10.5751/ACE-01248-130204

Wright, J. M. (1997). Preliminary study of Olive-sided Flycatchers, July 1994 - April 1997. *Alaska Department of Fish and Game. Final Research Report. Endangered Species Conservation Fund Federal Aid Studies SE-3-3, 4 and 5*, (April), 34.

Yee, T. W., Stoklosa, J., & Huggins, R. M. (2015). The VGAM Package for Capture-Recapture Data Using the Conditional Likelihood. Journal of Statistical Software, 65(5), 1-33. URL https://www.jstatsoft.org/v65/i05/.

Yip, D. A., Bayne, E. M., Sólymos, P., Campbell, J., & Proppe, D. (2017). Sound attenuation in forest and roadside environments: Implications for avian point-count surveys. *The Condor*, *119*(1), 73–84. https://doi.org/10.1650/CONDOR-16-93.1

Yip, D. A., Leston, L., Bayne, E. M., Sólymos, P., & Grover, A. (2017). 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

Yip et al., 2020 Sound level measurements from audio recordings provide objective distance estimates for distance sampling wildlife populations