

Predicting Breeding Status of a Forest Songbird from Singing Rate

by

Emily Upham-Mills

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Ecology

Department of Biological Sciences

University of Alberta

© Emily Upham-Mills, 2018

Abstract

For male breeding songbirds, song rate varies throughout the breeding season and tends to be correlated with breeding-cycle stages. Although these patterns have been well documented, to our knowledge, this relationship has not been used to predict a bird's breeding status through acoustic monitoring. The first objective of this study was to determine if variation in song rate can be used to predict the breeding status of the Olive-sided Flycatcher (*Contopus cooperi*; OSFL), a Species at Risk in Canada. In 2016, song rates from 27 male OSFLs in Alberta and the Northwest Territories were collected from human observers (n = 454 5-min counts), and breeding status (i.e. single, paired, and feeding young) was monitored throughout the breeding season. I evaluated the predictive ability of three modeling approaches (i.e. regression, hierarchical, and machine learning) using model sensitivity and specificity. The hierarchical model was the best at predicting all three breeding statuses, with 69%, 50% and 87% sensitivities and 80%, 82% and 78% specificities for predicting single, paired, and feeding young, respectively. This resulted in a mean sensitivity of 69%, compared with 54% and 50% from the regression and machine learning models, respectively. A second objective was to use the hierarchical modelling framework to predict breeding status from song rates collected by Autonomous Recording Units (ARUs) processed using automatic recognition software. For 24 of these OSFLs, I collected 4,302 5-min song counts and used daily song rate to compare the relationship of rates and breeding status as determined by ARUs versus human-observers. We then tested four hierarchical models accounting for imperfect detection. Song rates derived from ARU data followed a similar pattern to that of human-observer song rates, where single males had the higher rates, paired males had lower rates, and those feeding young had lowest rates, but the absolute values for rates were much lower with ARUs. All ARU data predictive models

performed poorly at predicting single (sensitivity range 0 – 7%) and well at predicting paired (sensitivity range 77 – 84%). The ARU models had mixed success at predicting feeding young (sensitivity range of 25 – 68%) but adjusting for imperfect detection did not improve model sensitivity to predict any breeding statuses. Low predictive ability was likely due to the low detectability of ARUs (e.g. bird movement out of detection range of ARU) and the automatic recognition software we used. Considering the high predictive ability of models using human-observer data and that the challenges currently associated with our acoustic processing methods can be addressed, I recommend that the breeding status of forest birds should be monitored using acoustic data. I provided a hierarchical modelling framework than can be applied to other species and improved to account for bird movement or number of conspecifics. This novel approach could provide a cost-effective tool to infer much needed demographic information over large spatial extents, and inform species status assessments, recovery strategies, and management plans for many species of conservation interest.

Preface

This thesis is an original work by Emily Upham-Mills. The research projects, of which this thesis was a part, were partly conducted on the traditional lands of the Dehcho and Tłı̨chǫ First Nations in the Northwest Territories, from whom I received support to conduct this work. I received research ethics approval from the University of Alberta Research Ethics Board (No. AUP00001523), and approval from the Government of Northwest Territories (No. WL500435, 29-April-2016).

Chapters 2 and 3 of this thesis are intended for publication soon and were co-authored by multiple people. Co-authors are listed at the start and contributions are summarized at the end of each chapter.

Dedication

For Mom and Dad.

Acknowledgements

Funding for this research was provided by the CREATE-Environmental Innovation graduate scholarship by the Natural Sciences and Engineering Council of Canada, Queen Elizabeth II Graduate Scholarship, Alberta Graduate Student Scholarship, the Department of Biological Sciences at the University of Alberta and the Federal Student Work Experience Program, and grants from the Alberta Conservation Association, Northern Scientific Training Program, University of Alberta Northern Research Award, Society of Canadian Ornithologists and Bird Studies Canada. Thank you to the Canadian Wildlife Service Northern Region landbird program (Environment & Climate Change Canada) for extensive in-kind support and funding for field technicians in the Northwest Territories and for the Bioacoustics Unit (University of Alberta) for in-kind and field technician support in Alberta.

Thank you to the person who suggested I take on a Masters studying boreal birds in the Northwest Territories and join the huge academic family that is the Bayne lab, Sam Haché. You reviewed every word of every abstract, chapter, poster and PowerPoint I generated throughout this thesis, spent many Friday afternoon phone calls or bumpy winter road drives talking science with me, and mentored with 110%. Thank you to Erin Bayne for accepting me into the lab and pushing me to become the modeller I never thought I could be. Thank you to my committee member Cindy Paszkowski for being a positive supporter from the get-go, and to my external examiner Mark Lewis for stepping in on short notice. Thank you to my co-authors/statistics gurus for your endless patience in explaining scary terms and concepts to me and revising our papers-to-be: Jody Reimer, Subhash Lele and Andy Crosby.

Thank you to my hard-working field assistants Alex Stone, Lucas Short, Emily Trendos, and Orla Osborne for long hours with a frustrating species. You collected so much data for this

project and the field work was so much more fun with the crew of you! Thanks for the additional field (and mental) support from Elly Knight, Rhiannon Pankratz, Michelle Knaggs and Hedwig Lankau. And Elly, thank you for always being there for me as a devoted friend and academic mentor, for inviting me to nighthawk happy hour, and for helping me keep life in perspective. Thank you for the Olive-sided Flycatcher ideas and enthusiasm from Julie Hagelin and Tara Stehelin. Thank you to the dedicated volunteers who processed hundreds of hours of recordings: Steph Parth, Natasha Annich, Logan Bradley, Peter Hoyer, Keara Nelson, Kathleen Enns and Jackie Twil.

Thank you to my fellow basement-dwellers Connor, Jocelyn, Nick, Jill, Laura and Carrie-Ann for the goofy times and science talks, and the rest of the Bayne lab for all the memories over the past three years. Thank you to my roommates, Tonya, Thea, and Laura for being so supportive and positive, and to Caitlin for inspiring talks and being my last-minute editing queen. Thank you, Jeremiah, for reminding me how cool birds are, for believing so hard in the purpose of my thesis, and for helping me to eat good food, take breaks to play and stay focused in the final days of the thesis.

Thank you, Mom, Dad, Greg, and Grace, for being the most supportive family a grad student could ask for.

Table of Contents

List of Tables	x
List of Figures	xi
Chapter 1: General introduction.....	1
Migratory songbird monitoring for status assessment and focal species	1
Demographic studies for forest songbirds.....	2
Breeding status and singing behaviour.....	3
Acoustic technology.....	3
Study species.....	4
Research objectives.....	6
Chapter 2: Can singing rate be used to predict male breeding status of forest songbirds? A comparison of three calibration models.....	8
Introduction.....	9
Methods.....	12
Study area	12
Sampling design	14
Methods of calibration.....	16
Model Evaluation	19
Results.....	20
Discussion	23
Author's contributions.....	27
Chapter 3: Using acoustic recording units to monitor breeding status of a forest songbird.....	29
Introduction.....	30
Methods.....	33
Study species	33
Field methods	34
Acoustic data collection.....	34
Acoustic data processing.....	35
Statistical analysis.....	36
Results.....	40
Author's contributions.....	50
Chapter 4: General conclusion.....	51

Summary	51
Model improvement using different song metrics	52
Model improvement using different breeding status classes	52
Recommendations for model application to other songbird species.....	53
1) Consider the research question.....	53
2) Consider behaviour of the focal species.....	54
3) Calibration.....	54
4) Validation.....	54
Conclusions	55
Literature Cited.....	57
Appendices.....	67
Appendix 1: Model selection results for each of the three modelling approaches from Chapter 2 (multinomial logistic regression, hierarchical model and classification tree), including individual component models of the hierarchical model	67
Appendix 2: Hierarchical model bug (Bayesian inference Using Gibbs Sampling) file for use with ‘rjags’ code (R package) used in Chapter 2, which identifies the model specification, data specification and prior distributions in the model to implement the Markov Chain Monte Carlo (MCMC) sampler	70
Appendix 3: Decision process for selecting a score threshold for use with an Olive-sided Flycatcher Song Scope recognizer to optimize recognizer output validation time while maintaining high true positive rate.....	71
Appendix 4: Hierarchical model bug (Bayesian inference Using Gibbs Sampling) files for use with ‘rjags’ code (R package) used in the four hierarchical models of Chapter 3, which identifies the model specification, data specification and prior distributions in the model to implement the Markov Chain Monte Carlo (MCMC) sampler.....	73
Appendix 5: A pilot study of quantifying Olive-sided Flycatcher songs in ARU recordings manually versus with automatic recognition to determine the most efficient processing method for breeding status – song rate studies	77

List of Tables

Table 1. 1. Glossary of terms used to describe predictive abilities of models. Descriptions are based on the R documentation for package ‘caret’, function confusionMatrix (Kuhn et al., 2017).	7
Table 2. 1. Breeding status classification descriptions and observational cues.....	15
Table 2. 2. Performance evaluation metrics of the three model types (MR = multinomial logistic regression, HM = hierarchical model, CT = classification tree).....	22
Table 3. 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 for ARU and human-observer data. SR = song rate, BS = breeding status, 1 ind = random effect of individual.	41
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 human-observer song rate hierarchical model (Chapter 2).	45

List of Figures

Figure 1. 1. A spectrogram of two consecutive Olive-sided Flycatcher songs recorded from an ARU.	5
Figure 2. 1. Locations (red dots) where Olive-sided Flycatcher territories were monitored in 2016 in the Northwest Territories (n=18) and northern Alberta (n=9), Canada.	14
Figure 2. 2. Prevalence of predicted breeding statuses for the three model types (MR = multinomial logistic regression, HM = hierarchical model, CT = classification tree) and the true breeding statuses.	23
Figure 3. 1. Predicted Olive-sided Flycatcher song rates for three breeding status classes from: A) a generalized linear (GLM) regression using human-observer* data, B) a generalized linear mixed effects (GLMM) model using human-observer data, C) a GLM using ARU data, and D) a GLMM using ARU data. GLMMs included random effect of individual (n = 28 individuals, and 24 individuals, for human-observer and ARU data, respectively). * Human-observer models included a time co-variate (minutes past sunrise), which was held at the median value for these predictions (113 minutes).	42
Figure 3. 2. Predicted song rates using ARU data for three breeding statuses over a range of sound levels (scaled and mean-centered).	43
Figure 3. 3. Prevalence of observed versus predicted breeding statuses for four ARU data models and the human-observer data model from Chapter 2.	44

Chapter 1: General introduction

Migratory songbird monitoring for status assessment and focal species

Large-scale declines in North American migratory songbird populations (Bohning-Gaese et al., 1993; Michel et al., 2016) have highlighted the importance of understanding the ecological processes driving population change throughout a species full life annual cycle (Donovan & Flather, 2002). Millions of these migratory songbirds are breeding each year in the boreal forest of North America (Wells et al., 2014), but very little is known on their population dynamics in this region which makes it difficult to determine the status of these species (Machtans et al., 2014). Forest songbird monitoring programs are typically limited to standardized species counts (i.e. North American Breeding Bird Survey; Sauer et al., 2017), with inference on habitat quality limited to population density estimates (Haché et al., 2014). Density may be an indicator of habitat quality, as long as individuals select habitat according to an ideal free distribution (Fretwell & Lucas Jr, 1970). However density often does not always indicate habitat quality (Van Horne, 1983).

An example of a boreal breeding songbird experiencing an important decline in population size is the Olive-sided Flycatcher (*Contopus cooperi*; OSFL), which is designated as Threatened under Canada's *Species at Risk Act* (S.C. 2002, c.29). However, the causes for decline in this species are not well-understood. This important knowledge gap has been identified in the Schedule of Studies of the Recovery Strategy of this species as one component preventing the identification of critical habitat to inform conservation actions to allow the species to recover (Environment and Climate Change Canada, 2016a). For example, Robertson & Hutto (2007) found that OSFLs breeding in the northern Rocky Mountains of Montana occurred at higher densities in selectively harvested forests, but nest success was half that of pairs in burned

forests of the same region. Thus, results from density models may not always depict spatial variation in habitat quality. Finding a cost-effective way to monitor variation in breeding success across a species range is required to understanding population trends and inform conservation efforts.

Demographic studies for forest songbirds

To quantify population viability throughout a species' range, researchers need to measure demographic processes (Johnson, 2007). This involves estimating reproduction (e.g. breeding status and successful fledging) and survival among populations associated with different biophysical attributes at different spatial scales (i.e. habitat types) to infer habitat quality (Johnson, 2007). However, this information is time-consuming to collect for songbirds (Martin & Geupel, 1993). Locating nests and identifying breeding status is especially challenging for forest songbirds with large territories and cryptic, hard-to-access nests like the OSFL. A current standard method to monitor bird nests is to use the Breeding Biology Research & Monitoring Database (BBIRD; Martin et al., 1997), which involves a large time commitment to find nests and many repeated visits to monitor nesting status. Another standardized method called MAPS (Monitoring Avian Productivity and Survivorship) combines mist-netting, banding and intensive point counts to estimate survival and reproductive rates of local breeding birds (DeSante et al., 1993). This requires standard effort every year, with hundreds of hours required throughout the breeding season. Because these traditional demographic monitoring techniques are so time consuming, they are rarely used at a larger scale, such as informing conservation of species across its entire range.

Breeding status and singing behaviour

Male songbirds sing mainly to attract females and defend their breeding territories (Armstrong, 1973; Thorpe, 1961). Researchers commonly use these audible cues to conduct surveys (Bibby et al., 2000) to estimate population densities, distributions, and trends (Robbins et al., 1989). Many factors affect the accuracy of acoustic songbird surveys, including song rate, distance, species and observer (Aldredge et al., 2007; Sólymos et al., 2013). Singing rate from an individual varies across the breeding season, which affects the probability of detecting these individuals (D. M. Wilson & Bart, 1985; Wright, 1997). This pattern has spurred much research into how and why song rate varies throughout the breeding cycle of these species (Slagsvold, 1977; Wilson & Bart, 1985).

Variation in song rate throughout the breeding season is strongly linked to breeding status (Wilson & Bart, 1985). Upon arrival to the breeding grounds in the springtime, male songbirds sing more frequently and consistently with the purpose of attracting females. After pairing, song rate tends to decline significantly (Gibbs & Wenny, 1993), with further decreases upon incubation (Lampe & Espmark, 1987; Wilson & Bart, 1985; Wright, 1997). Although these patterns create challenges for abundance estimates from acoustic surveys (Gibbs & Wenny, 1993), they also provide opportunities to study breeding behaviour which can be used to provide useful demographic information. In this thesis, I explore the potential to predict breeding status from song rate using the OSFL as a focal species.

Acoustic technology

Acoustic data collection by human observers in the field is limited by the time it takes to access sites on multiple visits and is negatively impacted by inclement weather. Fortunately, in

recent years, technological advances have facilitated higher efficiency and more fine scale detail in acoustically monitoring breeding birds. Autonomous Recording Units (ARUs) are audio recording devices which can be deployed to a stationary location to collect a large volume of acoustic data based on a pre-programmed recording schedule. These units are increasing in popularity in the field of avian ecology. Some examples of their use include conducting acoustic surveys (Shonfield & Bayne, 2017), monitoring breeding phenology of bird populations (e.g. Colbert et al., 2015; Digby et al., 2013; Jahn et al., 2017) and monitoring nest status in owls (Kozłowski, 2005). ARUs provide a large detection range and can record over a large time period, making them well suited for studying rare and low density species, or species with large territories (e.g. Campos-Cerqueira & Aide, 2016). The high volumes of acoustic recording data that are collected by ARUs can now be processed using species-specific recognition software, which scans audio recordings quickly to detect sounds and identify target species using pattern-matching algorithms. The advantages of ARUs provide an opportunity to collect detailed information on singing behaviour, such as song or call rate (Digby et al., 2013), leading to potential use in inferring breeding status. A primary challenge of using ARUs is that of imperfect detection, where the probability of detecting a bird decreases with distance from the recording unit (Yip et al., 2017). Thus, detectability is an important consideration for acoustic-based research, especially using ARUs which are stationary and will not capture all songs produced by birds as they move in and out of detection range.

Study species

The OSFL is a neotropical migratory songbird that typically breeds in lowland coniferous forests of the Canadian boreal and montane coniferous or coastal regions of the Pacific coast of

Canada and the United States (Altman & Sallabanks, 2012). OSFLs breed in habitat along forest edges and open areas such as wetlands, harvested forests and recent burns, that provide tall trees which provide perches for singing and open spaces for feeding on aerial insects (Altman & Sallabanks, 2012). In boreal regions, they typically defend territories in conifer stands, recently burned forest, and shrubby patches (Haché et al., 2014). OSFLs are monogamous birds that defend large territories (up to 40-45 ha) compared to other songbird species (Altman & Sallabanks, 2012), with territories from conspecifics being located at least >100 m apart (Robertson et al., 2009). They usually build an open cup nest in branches near the top of tall live or dead conifer trees, lay 3-4 eggs, and only lay one clutch per year (Altman & Sallabanks, 2012). OSFLs sing a single song type, described as a loud, clear whistle or the onomatopoeia “quick, three-beers!” (Fig. 1.1).



Figure 1. 1. A spectrogram of two consecutive Olive-sided Flycatcher songs recorded from an ARU.

In 2010, the OSFL was designated as a Threatened species in Schedule 1 of the *Species at Risk Act* (S.C. 2002, c. 29) of Canada. This designation was due to a widespread population decline of about 75% since around 1970 (Environment and Climate Change Canada, 2016a). Reasons for the decline remains largely unknown, but theories include a decrease in aerial insect prey abundance, fire suppression, deforestation and land conversion in the wintering grounds, and habitat loss in the breeding grounds due to industry (i.e. forestry and oil and gas sectors) and

urban development (Environment and Climate Change Canada, 2016a). More information is needed on demographic processes of this species to determine if conditions occurring on the breeding grounds could limit population growth (see Schedule of Studies in the Recovery Strategy for the Olive-sided Flycatcher [Environment and Climate Change Canada, 2016a]).

A 2-year study in Alaska indicated significant variation in song rate as a function of breeding status in the OSFL (Wright, 1997). As predicted, males sang at high rates before pairing, low rates after pairing, moderate rates when incubating, and very low rates when feeding young, and males that never paired sang at higher rates throughout the season (Wright, 1997). All males also sang at higher rates and most consistently around sunrise (Wright, 1997). Female OSFLs can also sing, but infrequently, and their songs are weaker and soft, lazy or slurred (Wright, 1997).

Research objectives

In this thesis, the overall objective was to determine if song rate can be used to predict breeding status in the OSFL in its northwestern boreal breeding range. In Chapter 2, I determined if song rate quantified by human observers conducting song counts in the field can be used to predict three breeding status classes for males: single, paired and feeding young. I evaluated the predictive abilities of three statistical methods commonly used for classification: 1) a multinomial logistic regression, 2) a hierarchical model, and 3) a classification tree (see Table 1.1 for glossary of terms used to describe predictive abilities throughout the thesis). In Chapter 3, I used the best modelling framework identified from Chapter 2 and applied this analytical framework to song rate data collected from ARUs to predict the breeding status of OSFLs. Specifically, I compared predictive ability of four models differing in adjustments for imperfect

detection. To my knowledge, this is the first study to use song rate to predict breeding status of a forest songbird and to highlight the potential of ARU technology to collect broad-scale demographic data on a species of conservation interest. In Chapter 4, I summarized the success of the hierarchical model using human-observer song counts, recommended an approach for applying the model to other species, and described improvements to study design and acoustic processing methods to improve model performance using ARUs to predict breeding status. I conclude by describing how my findings contribute to conservation management of vocal species in expanding the potential for large-scale demographic data collection using bioacoustics.

Table 1. 1. Glossary of terms used to describe predictive abilities of models. Descriptions are based on the R documentation for package ‘caret’, function confusionMatrix (Kuhn et al., 2017).

	Reference	
Predicted	Event	No Event
Event	A	B
No Event	C	D

Term	Synonym	Formula
Sensitivity	True positive rate; recall	$A/(A + C)$
Specificity	False positive rate	$D/(B + D)$
Prevalence (Observed)	Proportion of observed events	$(A + C)/(A + B + C + D)$
Prevalence (Predicted)	Proportion of predicted events	$(A + B)/(A + B + C + D)$

Chapter 2: Can singing rate be used to predict male breeding status of forest songbirds? A comparison of three calibration models

Emily J. Upham-Mills^{1*}

Jody Reimer^{1,2}

Samuel Haché³

Subhash R. Lele²

Erin M. Bayne¹

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

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

² Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada

* corresponding author: uphammil@ualberta.ca

Introduction

Ecologists often desire information on the state of an organism or environment that can be challenging to measure directly. Because of this, many state variables require indirect measurements (Stephens et al., 2015). For example, leaf area index, an important metric of forest function, is often measured indirectly through recording light absorption patterns because of the high cost of directly measuring leaf dimensions (Olivas et al., 2013). Other examples include using indicator species to track changes in the state of the environment (Lindenmayer & Likens, 2011), or satellite tracking data as an indirect measure of wildlife feeding behaviour (Robinson et al., 2007). Although use of indirect metrics is common in ecology (Stephens et al., 2015), the methods of statistical analysis used to create the relationship between the state variable of interest and the indirect proxy are often oversimplified. For simplicity, it may be tempting to ignore causal dependencies when analysing these relationships. However, this can result in incorrect conclusions or low predictive accuracy due to the error distributions implied by standard statistical models (e.g., regression models). Statistical calibration, which can be described as the reverse process of regression, aims to estimate an independent variable (the cause) from a dependent variable (the effect) (Osborne, 1991). Despite acknowledgements of the importance of calibration in some fields (for example, water quality health [Hall & Smol, 1992; Ter Braak & Barendregt, 1986] and paleoecology [ter Braak, 1995]), relatively few ecological studies have used such approaches for creating effective indirect measurement techniques (Biondi & Waikul, 2004).

In avian ecology, calibration models describing the relationship between a male songbird's breeding status and his behaviour (cause and effect, respectively) may provide a novel way to monitor male breeding status indirectly. This information is required to inform sound

conservation planning (Anders & Marshall, 2005), but direct measurements of pairing success, nest success, and fledging rates for songbirds is expensive and logistically challenging (Martin et al., 1997). Thus, empirical data on breeding status is only available for a few species and over relatively small spatial extents (e.g. Christoferson & Morrison, 2001; Dussourd & Ritchison, 2003; Haché, Villard, & Bayne, 2013; Holmes, Sherry, Marra, & Petit, 1992). Indirect measures of breeding status, such as observing non-agonistic behaviour towards conspecifics to confirm pairing status and observing adult birds carrying food to confirm presence of young, have been suggested as an approach to decrease time and effort to estimate metrics such as fledging success (Hunt et al., 2017; Vickery et al., 1992). However, such methods are still time consuming for many species occurring at low abundance and with large breeding territories and have not yet been rigorously calibrated.

We propose a simpler indirect measure of a songbird's breeding status: inferring breeding status from singing behaviour. Songs in passerines are primarily used by a male to attract a female and to defend a territory against conspecifics (Armstrong, 1973; Thorpe, 1961). For many species, males tend to sing at high rates when they are unpaired, with declines in singing rate as their breeding status changes (i.e. unpaired to paired, mated to nest building, egg laying to incubating, incubating to feeding nestlings, etc.; Dussourd & Ritchison, 2003; Gibbs & Wenny, 1993; Liu et al., 2007). We refer to this pattern as the breeding status - song rate relationship, or the "BSSR" relationship. While several studies have described the BSSR relationship, to our knowledge, none have attempted to use song rate to predict breeding status (although Staicer et al. [2006] suggested the possibility). We explore three different calibration models to assess the use of song rate to predict breeding status (single, paired, and feeding young) of the Olive-sided Flycatcher (*Contopus cooperi*; OSFL). Specifically, our objective was to test predictive accuracy

of these three BSSR calibration models. We used OSFL song rate data from a study conducted in the Northwest Territories, Canada. This species is designated as Threatened under the *Species at Risk Act* (SARA; 1994), and has experienced an overall population decline of 70% between 1970 and 2015 (Environment and Climate Change Canada, 2017). Therefore, finding a cost-effective way to monitor breeding success is a priority in recovery planning (Environment and Climate Change Canada, 2017).

Previous work has suggested that song rates and detection probabilities for the OSFL are influenced by breeding status, time of day, and day of year (Wright, 1997). Time of day is an important predictor for singing activity in songbirds, as most males sing the most around sunrise and song production declines throughout the day (Stacier et al., 1996). Date is also an important variable for predicting breeding status, because migratory birds will be single upon first arriving at the breeding ground, then will be more likely to be paired or have active nests as the days advance. Furthermore, latitude may affect breeding timing because more northern breeding sites will have later arrival times. We therefore test the importance of each of these predictors in the BSSR models.

First, we perform a multinomial logistic regression of breeding status against song rate and temporal covariates. Arguably the simplest model we consider, this model conflates the causal relationship assumed between breeding status and song rate, and the causal mechanisms behind the temporal predictors. Secondly, we use a hierarchical model, defined as a sequence of probability models arranged to describe conditionally dependent random variables (Kéry & Royle, 2016a). This modelling approach is useful for complex ecological modelling because of its ability to account for multiple sources of uncertainty (Cressie et al., 2009). In the case of BSSR calibration, a hierarchical model can be used to deconstruct the cause and effect

relationships into one component that accounts for temporal variation in breeding status probabilities throughout the breeding season and a second component that models how breeding status and time of day affect song rate. Our third approach was to use a classification and regression tree (CART) model (Breiman et al., 1984), to predict breeding status from song rate, time of day, date and latitude. CART is a machine learning approach which has been recommended as a powerful method for modelling complex ecological data because of its ability to deal with nonlinear relationships and high order interactions (De'Ath & Fabricius, 2000). CART models are comprised of a series of binary splits, based on predictor variable values, to partition data into smaller groups and increase the proportion of any one class (i.e. categorical value) in each group (Kuhn & Johnson, 2016).

Our objective was to determine the best modelling framework to predict accurately a songbird's breeding status. We measured relative success of the three models by comparing prediction sensitivity and specificity for the three breeding status classes using k-fold cross validation. We conclude by discussing the strengths and weaknesses of the top performing model and how this model may be further developed for use with autonomously recorded acoustic data.

Methods

Study area

The study took place in northern Alberta and the Northwest Territories (Fig. 2.1) between May 30 – July 22, 2016. The sampling locations in Alberta were ~80 km north of Fort McMurray, in the Mid Boreal Mixedwood Ecoregion (Strong & Leggat, 1992) in stands comprised of upland jack pine (*Pinus banksiana*) forest, bog and fen wetlands (dominated by black spruce, *Picea mariana* and tamarack, *Larix laricina*). Dominant understory shrubs

included rose (*Rosa acicularis*), alder (*Alnus spp.*) and aspen (*Populus tremuloides*) in upland sites and blueberry (*Vaccinium myrtilloides*), Labrador tea (*Ledum groenlandicum*), dwarf/bog birch (*Betula spp.*), and willow (*Salix spp.*) in lowland sites. The Northwest Territories study area ranged from ~ 30 km south of Fort Providence to Behchokò, with site access off highway 3 (Fig. 2.1). It was within the Great Slave Lowland Mid-Boreal Ecoregion (Ecosystem Classification Group, 2009), which is dominated by wetlands (bogs and fens) and scattered patches of upland mixed-wood and jack pine forests. Some sampling locations were in stands that had burned in 2014 and 2015. Shrub species composition was similar to Alberta.

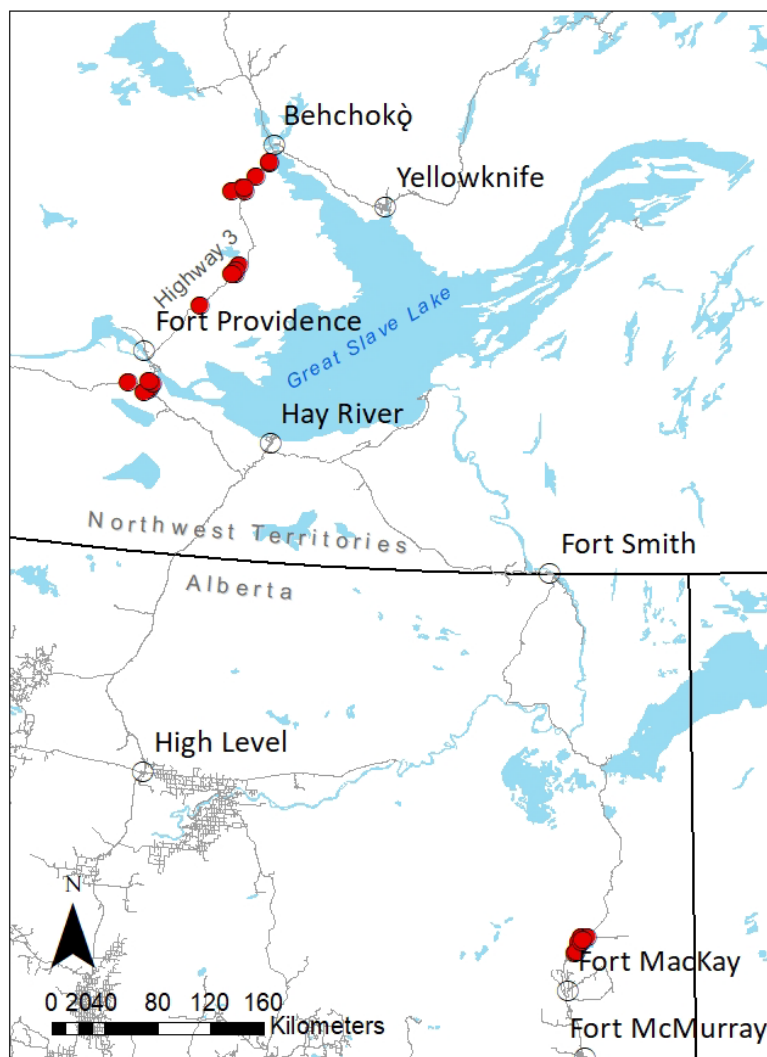


Figure 2. 1. Locations (red dots) where Olive-sided Flycatcher territories were monitored in 2016 in the Northwest Territories (n=18) and northern Alberta (n=9), Canada.

Sampling design

Potential sampling locations were selected based on known locations of territorial male Olive-sided Flycatchers (OSFLs) obtained during the previous two breeding seasons (Pankratz et al., 2017; Pardieck et al., 2015, E. Bayne unpublished data, Knaggs et al. in prep.). The spatial extent of our study area was selected to represent southern (latitude 57°) and northern locations (latitude 62°) to account for variation in daily activity levels. Between the last week of May and the first week of June 2016, potential sampling locations were monitored using call playback surveys to confirm arrival and settlement of OSFLs. Call playback surveys consisted of 5 minutes of listening, followed by 30 seconds of playback, 2 minutes of listening, 30 seconds of playback and a final 5 minutes of listening. If call playback surveys were conducted at a location on three separate days before June 8, 2016, and no OSFL was detected, or only once prior to the last visit, a location was deemed to not overlap a territory. Alternatively, if an OSFL was detected twice or more, the potential territory was considered occupied and included in our sampling locations. While monitoring potential territories, we detected additional males in other nearby locations (n =5) and these locations were added to the sampling design. This resulted in 19 sampling locations in the Northwest Territories and 9 in Alberta.

For each territory, we conducted repeated visits approximately once per week, resulting in between 1 to 10 visits per bird (6 ± 2.43 ; mean \pm SD). During each visit, we assessed breeding status (i.e. single, paired [including being paired with no known nest, nest building or incubating] or feeding young; Table 2.1). When breeding status could not be confirmed in the field, the

status was back calculated based on average breeding timing for the species (15 days for incubation, 19 days for nestling period; Wright, 1997) using breeding status information from previous and subsequent visits. Dates when status could not be back calculated with confidence (i.e. there were not enough dates with field-confirmed breeding status) were excluded from the analysis. A visit lasted 1 hour and was conducted between sunrise and the first 6 hours after sunrise. During each visit, we also measured song rate, i.e. the mean singing rate based off four 5-minute song counts. At time 0, 15, 30 and 45 minutes from the start of the visit, the number of “quick, three beers” songs sung by the male were counted in a 5-minute period. Song counts were only conducted when observers were close enough to see the male and were cancelled if he flew away during the count period. Song count data were recorded until June 30 and July 8, in Alberta and the Northwest Territories, respectively. In both study areas, territories were revisited once or twice between July 8 - 22 to confirm nest contents if not yet confirmed. Nests were located for 12 of the 28 males monitored in this study. For these males, nest contents were confirmed using a telescopic PVC pole (maximum height 17 m) with a video camera, which provided a live feed to a handheld monitor on the ground (The Peeper Cam, <http://www.ibwo.org/camera.php>, David Luneau, Arkansas).

Table 2. 1. Breeding status classification descriptions and observational cues.

Classification	Description	Behavioural evidence
Single	Unpaired male defending a breeding territory	No interactions with a female detected on that day and during previous visits
Paired	Immediately after a female settled on a male’s territory, no known nest	First observation of a female present on a male OSFL’s territory, but nest not located
	Nest building – one or both adults seen building a nest	Either or both adults seen carrying nesting material
	Incubating - 15-day period pre-hatching when female spends most of her time sitting on eggs in the nest (Altman & Sallabanks, 2012)	Female directly observed sitting on nest, male observed feeding a female on the nest or when nestlings were

		observed during a later visit (backdated incubation period based on hatch date)
Feeding young	One or more eggs have hatched, and nestlings are present	A parent observed holding/carrying insect and/or feeding an insect to nestlings, or direct observations of nestlings

Methods of calibration

For the three model types, i.e. multinomial logistic regression, hierarchical model, and classification tree, model selection was conducted on a set of candidate models allowing for identification of meaningful predictor variables. Song rate, time, ordinal date, and latitude were considered as candidate predictors for each model type. Song rate was the average 5-minute song count collected within one sampling-hour (n=4) during a visit. Time was calculated by subtracting time of sunrise from the mean time of the song counts. Both date and time were mean-centered and scaled by their standard deviation. Latitude was a binary categorical variable, representing either the northern (Northwest Territories) or southern sites (Alberta).

All models were built using the R statistical programming language (R Core Team, 2013, R Foundation for Statistical Computing, Vienna, Austria).

Multinomial logistic regression

Multinomial logistic regression models were built using the `multinom` function in R package ‘`nnet`’ (Ripley & Venables, 2016; Venables & Ripley, 2002). We compared six *a priori* models (Appendix 1, Table 1), ranging from the simplest models with a single predictor, to the full model which included song rate, date, time as an interaction with song rate, and latitude as an interaction with date, as predictors of breeding status. We used the lowest Akaike’s Information

Criterion (AIC) value (with a difference >2) to select the best supported model (Burnham & Anderson, 2002).

Hierarchical model

The hierarchical model structure follows from Bayes' theorem (Kéry & Royle, 2016b). This approach allows relating the probability of an individual having a given breeding status, conditional on its song rate, to the probability of that individual having a given song rate, conditional on its breeding status (Eq. 1).

Equation 1.

$$P(bs|sr, date, latitude, time) \propto \underbrace{P(bs|date, latitude)}_{\text{Component A}} \underbrace{P(sr|bs, time)}_{\text{Component B}}$$

Where *bs* is breeding status, *sr* is song rate and \propto means “proportional to”, with the constraint that the sum of the left-hand side probabilities for each breeding status must sum to 1.

Component A of Eq. 1 is the marginal probability of observing each breeding status on a given day at a given latitude, while component B is the conditional probability relating song rate to breeding status and time of day. Equation 1 describes all co-variate relationships which we considered, but final models did not necessarily include all co-variables. The breeding status variable in our model is known and observable, unlike state-space models (Patterson et al., 2008), hence we were able to conduct model selection on each component separately. This avoided the difficulty in model selection for state-space models (i.e. where AIC is not properly defined and Deviance Information Criterion [DIC], typically used with Bayesian model selection, is problematic; Hooten & Hobbs, 2015).

We first conducted model selection for component A, using the lowest AIC value within 2 to select the top model. For this component, we compared three *a priori* multinomial logistic

regression models (MLR) relating the marginal probability of each breeding status to date and latitude (Appendix 1, Table 2).

For component B, we used a generalized linear model (GLM) for song rate, with time and breeding status as predictor variables. We considered 12 *a priori* models, using either breeding status as a single covariate or both breeding status and time. For both options, we tested using a Poisson, zero-inflated Poisson, a negative binomial, and a zero-inflated negative binomial song rate distribution (Appendix 1, Table 3). For component B, model sensitivity for the three predicted classes was used to select the top model, instead of AIC, with the purpose of maximizing model ability to predict individual breeding statuses from song rate.

Normal approximations of the parameter estimates and their standard error values from the top-ranked MLR (component A) and GLM (component B) were used as priors for the hyperparameters of model components. A common practice is to use so-called non-informative priors, but they have issues (Lele, 2014; Northrup & Gerber, 2018). We chose to use informative priors as is suggested by Hamilton (1986) and Harris (1989) in the context of prediction.

Cumulatively, selected models for components A and B comprised the top hierarchical model used to predict breeding status class probability densities. These calibration distributions (i.e. posterior distributions) were generated using Markov Chain Monte Carlo methods from the package ‘rjags’ in R (Plummer, Stukalov, et al., 2016). We generated 5 Markov chains, discarding the first 1,000 values as the burn-in, followed by 10,000 iterations. We used the Gelman-Rubin diagnostic to test for convergence of the chains to a posterior distribution (Brooks & Gelman, 1998; Spiegelhalter et al., 1995).

Classification tree

The classification tree model (CT; Brieman et al., 1984) was built using R package ‘rpart’ (Therneau et al., 2018), using the Gini index as the impurity index (Wu et al., 2008). A set of classification trees were built to include a range of sizes, from unpruned (i.e. the tree with the highest number of branches, created using the default complexity parameter value of 0.01) to fully pruned (i.e. the tree with the smallest number of branches). We conducted model selection by choosing the classification tree which predicted with the highest mean sensitivity after K-fold cross validation (i.e. “leave-one-group-out”, process described in more detail below). This model selection process is the best when the research objective behind the generation of classification trees is prediction (De’Ath & Fabricius, 2000).

Model Evaluation

Breeding status predictions were made using “leave-one-group-out” K-fold cross validation, using one individual OSFL as the “group-out”. Specifically, all observations from one individual OSFL were removed from the dataset, leaving a training set with observations from 27 males which was used to obtain predictive distributions. The model selection process was not repeated for each validation fold. Breeding status predictions were then made on observations from the “group-out” OSFL and the process was repeated for each OSFL, resulting in a 28-fold cross validation for each model type. The output from each model was a probability mass function for each sampling time, describing the probability of the individual having each of the three breeding statuses. We then used the breeding status with the highest probability as the predicted status.

We used the following performance statistics to compare the predictive accuracy from the top model in each model type: over-all model accuracy (i.e. proportion of correctly classified song rates), sensitivity (i.e. true positive rate for each breeding status class), mean sensitivity (mean taken across the three breeding status classes), specificity (i.e. false positive rate for each breeding status class), and mean specificity (mean taken across all three breeding status classes). See Chapter 1, Table 1.1 for functions describing each prediction evaluation term. We used specificity and sensitivity as predictive measures because both are prevalence-independent test characteristics, meaning that their values do not depend on the prevalence of a value in the dataset. We tested whether predicted breeding status prevalence was significantly different from prevalence of observed breeding statuses by performing two tests of marginal homogeneity: a Bhapkar test for overall results (Bhapkar, 1966) and McNemar tests for each of the three predicted classes (McNemar, 1947).

Results

We collected on average 20 (± 7.70 ; $n = 545$) 5-minute song counts per territorial male, from which we calculated 160 mean 5-minute song rates. The breeding statuses associated with these song rates were comprised of 26 (16.3%) instances where males were classified as single, 111 (69.4%) as paired, and 23 (14.4%) when feeding young.

Multinomial logistic regression

The top multinomial model included song rate and date as independent variables (Appendix 1, Table 1). Song rate had a significant negative effect on the probability of being paired versus single (log odds ratio = 0.94, $p < 0.001$). OSFLs are more likely to be paired or

feeding young than single later in the breeding season. Date had a significant positive effect on the probability of being paired (log odds ratio = 1.07, $p = 0.045$) or feeding young (log odds ratio = 1.74, $p < 0.001$) versus single.

Hierarchical model

The top ranked full model included time, date, and song rate (see Appendix 2 for BUGS-language script file for this model). For component A, breeding status was best modelled by date (Table 2 in Appendix 1), where the probabilities of: 1) being single is highest in early dates, 2) being paired is highest in the middle of the date range, and 3) feeding young is highest in later dates. The component B model which resulted in the highest mean sensitivity included both breeding status and time as predictors of song rate, using a Poisson distribution of mean rounded song rate (see Table 3 in Appendix 1 for contrasting AIC values).

Classification tree

The top CT model from the model selection (i.e. the one that best predicted all three breeding statuses) had four splits and included all predictor variables (ranked importance: date, song rate, time and latitude; Appendix 1, Fig. 1).

Model performance comparison

Based on the highest mean sensitivity, the top model type was the hierarchical model (69%), followed by the multinomial logistic regression and classification tree with mean sensitivity of 54% and 50%, respectively (Table 2.2). Overall accuracy values for the three models did not produce the same ranking, in that the multinomial model had the highest accuracy

(74%), followed by the classification tree (71%), and the hierarchical model (59%). Overall accuracy is highly biased by unequal class distribution in the dataset (i.e. prevalence), with models predicting higher prevalence of paired birds than observed.

All three models overpredicted some breeding statuses, indicated by specificity values other than 1 (Table 2.2). The multinomial logistic regression and classification tree both overpredicted paired at a high rate (specificity values of 0.37 and 0.29, respectively) compared to the hierarchical model (0.82 specificity). The marginal frequencies (i.e. predicted breeding status prevalence) of the multinomial logistic model and the classification tree were similar (< 10% predictions of single and feeding young and > 80% for paired), while those for the hierarchical model predicted a lower prevalence of paired individuals (Figure 2.2), with the true prevalence lying between these two extremes. Marginal homogeneity between the predicted and the observed breeding statuses was similar for all three models, with prevalence of single and paired differing significantly from true prevalence (Table 4 in Appendix 1). Prevalence of feeding young did not differ significantly between predicted and true breeding statuses for the classification tree and multinomial logistic regression, but the predicted value differed significantly from the true breeding status for the hierarchical model.

Table 2. 2. Performance evaluation metrics of the three model types (MR = multinomial logistic regression, HM = hierarchical model, CT = classification tree).

Model	Accuracy (95% CI)	Sensitivity single	Sensitivity paired	Sensitivity feeding young	Mean Sensitivity	Specificity single	Specificity paired	Specificity feeding young	Mean Specificity
MR	0.74 (0.67 – 0.81)	0.19	0.92	0.52	0.54	0.96	0.37	0.97	0.77
HM	0.59 (0.51 – 0.66)	0.69	0.50	0.87	0.69	0.80	0.82	0.78	0.80
CT	0.71 (0.63 – 0.78)	0.08	0.89	0.52	0.5	0.93	0.29	0.98	0.73

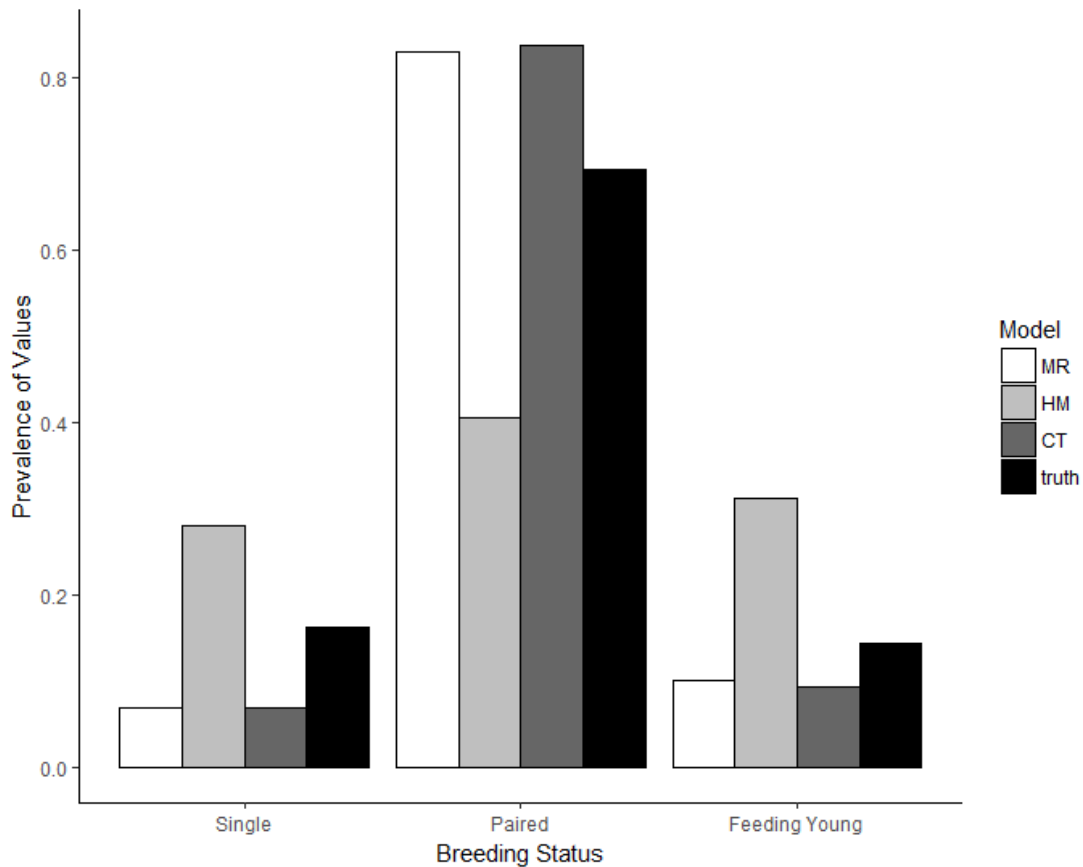


Figure 2. 2. Prevalence of predicted breeding statuses for the three model types (MR = multinomial logistic regression, HM = hierarchical model, CT = classification tree) and the true breeding statuses.

Discussion

The top modelling approach in our comparison was the hierarchical model (hereafter referred to as the BSSR model), which predicted all three breeding statuses correctly at a higher rate (i.e. sensitivity) than the regression and CART models and was the best at not overpredicting any given breeding status (i.e. specificity). The challenge with measuring breeding status indirectly is the statistical calibration of the underlying behavioural mechanism when breeding

status causes changes in singing rate. The hierarchical structure of the BSSR model allowed us to address this challenge while also accounting for daily variation in singing rate through the hierarchical series of conditional probability statements. This study provides an example of how to create a relatively accurate predictive model through statistical calibration for an indirect measurement of a biological state. Our work answers the call for more fundamental studies to better understand and represent the underlying mechanisms in indirect measurements in ecology (Lindenmayer & Likens, 2011; Stephens et al., 2015).

The pattern we observed in how singing rate changes with breeding status in the OSFL is similar to patterns observed by Wright (1997) in an OSFL population in Alaska. In both study areas, in different years, unpaired males sang at the highest rates, males who had paired and were engaged in initial breeding activities (i.e. nest building and incubating) sang less and males feeding young rarely sang and when they did, they sang few songs. OSFLs in both studies also sang most around sunrise and much less as time since sunrise increased. This suggests the song rate component of the hierarchical BSSR model can be used in different study areas for OSFL research and that time of day is an important song rate predictor. The other component of the hierarchical BSSR model however, models the probability of each breeding status given the date, or general species breeding timing (i.e. phenology) for that latitude and year. Although latitude was tested as a predictor in the breeding status component model, it was not significant in model selection. This result suggests limited regional variation in breeding phenology, contrary to our expectation that phenology would shift in our more northern study area. Environmental conditions during migration and at the breeding grounds can change breeding timing for a species among years, especially with the warming effects of climate change (Visser et al., 2004). Mean dates of OSFL pairing and of feeding young in 2016 in northern Alberta and the Northwest

Territories were comparable to those reported from other OSFL populations and phenology tended to not vary beyond a week between 1995 and 1996 from one study (Wright, 1997). Although there may not be extreme variation in phenology between breeding regions or among years in OSFL, it may be important to verify breeding phenology for the region and year of interest for future application of the model.

The BSSR model we produced is a simple version which can be used as a baseline on which to add parameters to improve predictive ability. Additional covariates which could be added to the model to account for variation in song rate include density of conspecifics (Lampe & Espmark, 1987), temperature (Gottlander, 1987), and singing rate or predicted (or known) breeding status of a given individual in the previous visit. Model improvement may also be achieved by using other song metrics. For example, instead of song rate, length of song bout, time of first song, afternoon singing rates, song count conditional on at least one song, or a combination of other song metrics might improve accuracy. The BSSR model we constructed is constrained to use between OSFL spring arrival and late July when the first round of breeding OSFLs are feeding young, but if the research objective was to predict breeding status after pairing (i.e. incubating eggs, feeding young, and fledging), the use of call rates instead of song rates may be a more precise indicator of nest status because calls represent activity at the nest (J. Hagelin and J. Wright, unpublished data). Though this would only be possible when nest locations were known, as calls are only detected by observers when in close proximity of a nest.

The sensitivity values (i.e. true positive rates) for the BSSR model were 69%, 50% and 87% to predict single, paired and feeding young respectively. To our knowledge, no other studies have used calibration methods to predict breeding status from song rate, so we are unable to compare predictive ability with those from other models. However, we can compare the BSSR

model predictive ability to that of other breeding bird reproductive indices. Vickery et al. (1992) designed a method to measure reproductive success, representing five statuses ranging from unpaired to fledged young, based on breeding-behaviours. This index provided a reasonable measure of reproductive success for grassland songbirds compared to more intensive nest monitoring at the same study area (27% predicted fledged vs 42% truly fledged; Vickery et al., 1992). When adapted to integrate nest monitoring with breeding behaviours for three forest breeding birds, the index provided correct breeding status predictions for 61 – 79% of the visits (Christoferson & Morrison, 2001). Our hierarchical model had a similar predictive success, without the need for extensive nest searching and behavioural observations, although our model is constrained to predict three breeding classes.

Monitoring song rate over a larger portion of the breeding season would improve certainty in predictions for individual birds. However, collection of song rate data by human observers on a fine temporal scale (i.e. daily versus once per week) would take a large amount of time, and ultimately be infeasible. A promising alternative method for collecting a larger amount of song rate data is using autonomous recording units (ARUs), which are increasing in popularity for bird research (e.g. Shonfield & Bayne, 2017, Pankratz et al. 2017). There are three important advantages of using this technology to predict breeding status from song rate: 1) daily acoustic surveys of a target location can be conducted for the entire breeding season; 2) large quantities of acoustic data can be processed using automatic recognition software; and 3) acoustic data can be permanently stored, which can be reanalyzed later. The advantage connected to future reanalysis reflects the fact that automatic recognition software, used to detect species of interest efficiently, is still improving and future processing may improve detection rates on recordings. This technology is being applied over large spatial extents and acoustic data are becoming readily

available for many regions. For example, the Alberta Biodiversity Monitoring Institute (www.abmi.ca) has been monitoring breeding birds in Alberta since 2003 and breeding season-long recordings are available from across the province (Alberta Biodiversity Monitoring Institute, 2012). Thus, large-scale demographic analyses based on temporal variation in song rates could be conducted for our focal species if the hierarchical model can be adapted for ARU-based song rates. ARU data has some challenges however, primarily associated with imperfect detection probabilities related to bird movement away from the detection limit of the ARU. This would have to be accounted for in the modelling approach, and the hierarchical model provides the framework to include such uncertainty and is an area of active investigation.

This study was the first attempt to predict a male songbird's breeding status using his singing rate, and our results provide a new method to monitor breeding status in a migratory songbird. We highlighted the importance of considering the calibration problem in ecological prediction modeling and demonstrated the advantage of using hierarchical modeling over conventional predictive model types (i.e. multinomial logistic regression and classification tree) to improve the sensitivity of predicting target classes. Future studies should aim at testing a similar approach to predict breeding status from song rates for other songbird species. We demonstrated that monitoring birdsong to infer songbird breeding status shows promise and warrants further investigation, especially if the model can be further developed for application with non-invasive ARU monitoring.

Author's contributions

EU, EB, and SH conceived the ideas and designed the field sampling methodology; EU collected the data; SL and JR conceived the ideas for the statistical analysis methodology; EU and JR

analyzed the data; EU led the writing of the manuscript. All authors contributed critically to the drafts.

Chapter 3: Using acoustic recording units to monitor breeding status of a forest songbird

Emily J. Upham-Mills^{1*}

Jody Reimer^{1,2}

Samuel Haché³

Andrew D. Crosby¹

Erin M. Bayne¹

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

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

² Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB
T6G 2G1, Canada

* corresponding author: uphammil@ualberta.ca

Introduction

The primary functions of singing in male songbirds during the breeding season are to defend a territory and to attract a mate (Catchpole & Slater, 2008; Collins, 2004). Singing rates for males vary throughout the breeding season as these functions change from mainly intersexual attraction to intrasexual competition (Lampe & Espmark, 1987; Otter & Ratcliffe, 1993). Specifically males tend to sing at high rates when unpaired and lower rates immediately after pairing, with gradual decreases in song rate as birds progress from nest building through incubating and feeding young (Gibbs & Wenny, 1993; Hayes et al., 1986; Lein, 1978; M. B. Robbins et al., 2009; Stacier et al., 1996; Wright, 1997). This variation in singing rate results in temporal variation in detectability of birds by human observers conducting acoustic surveys (Johnson, 2008; Nichols et al., 2009). While researchers have quantified singing rate of forest songbird males throughout the breeding season based on acoustic surveys (Sólymos et al., 2013), such data have rarely been used as a predictor of breeding phenology and success (but see Staicer et al., 2006). The ability to monitor breeding status efficiently in individual birds using acoustic data would fill important gaps in our understanding of population dynamics of most forest songbirds (e.g. Environment and Climate Change Canada 2016; Environment Canada 2015; Haché, Villard, and Bayne 2013) and thus inform species status assessments (e.g. Environment and Climate Change Canada 2018).

In Chapter 2, we presented the advantages of using a hierarchical modelling framework to predict breeding status of the Olive-sided Flycatcher (*Contopus cooperi*; OSFL), a neotropical migratory songbird, as a function of date, song rate, and time of day by using field observations. This model correctly predicted breeding status classes (single, paired and feeding young) with an average sensitivity of 70%. This result suggests that the breeding status – song rate relationship

(hereafter “BSSR”) can be used to monitor the breeding status of this Species at Risk in Canada (Environment and Climate Change Canada, 2016a). However, this type of modelling (hereafter “human-observer model”) requires song rate data from field observers who must track focal individuals multiple times throughout the breeding season, which can be time consuming and costly. Thus, these types of 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 number of years (but see Scott Sillett and Holmes 2002). The presence of human observers may also alter singing behaviour, resulting in biased data, although the degree of this bias, to our knowledge, has never been quantified (Bye et al., 2001; Gutzwiller et al., 1994, but see Campbell & Francis, 2012).

The advent of autonomous recording unit (ARU) technology to supplement human observers in collecting acoustic data (Pankratz et al., 2017; Shonfield & Bayne, 2017; Yip et al., 2017) may provide a cost-effective way to quantify the BSSR relationship over large spatial scales. 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 do not alter the behaviour of birds being monitored (Darras et al., 2018; Müller et al., 2006). New technology also allows for bulk processing of the larger amounts of recordings generated for sound identification and species classification (Shonfield & Bayne, 2017). The performance and design of these species recognition models are constantly improving (Knight et al., 2017). Additionally, recent research has show that measurements of relative sound level in the recordings can be used to estimate distances of birds from recorders and thus standardize ARU surveys; a lack of standardization has limited inference in ARU studies (Darras et al., 2018; Yip et al. in review). Long-term monitoring programs have started using ARUs to document changes in breeding phenology of

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 produce abundant data on singing behaviours passively, there are uncertainties about how bird movement affects detectability (but see Matsuoka et al. 2012), especially, when song rates are derived from recordings collected from permanent sampling locations.

This goal of this study was to evaluate the ability of the BSSR relationship model to predict breeding status in the OSFL, using song rate data collected from ARUs. The first step towards using ARU data in this predictive model was to calibrate the existing modelling framework based on acoustic data collection by human observers (Chapter 2) to ARU data by determining if underlying singing patterns produced using ARU data are comparable to patterns produced with human observer data. This calibration step is important in building a model for indirect measurement of a biological state (Stephens et al., 2015). Therefore, our first objective was to test for differences in effect size, and changes in song rates among breeding statuses (single, paired, and feeding young) of OSFLs from ARU-based versus human-based song rates. The second objective was to predict breeding status using ARU data and a modified hierarchical human-observer model (Chapter 2) based on the top ARU data song rate model determined by objective 1. Because OSFLs occupy territories which are larger than the detection range of an ARU (Altman & Sallabanks, 2012), we expected detectability issues. Thus, 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) only song rates with known presence (hereafter the “zero-truncated model”); 3) a two-step model where the probability of detection is first modeled followed by the song rate model (hereafter the “zero-

inflated model”), and 4) relative sound level as a predictor (i.e. proxy for the relative distance of a male to the ARU; hereafter the “distance model”). We tested for an effect of relative distance of a male because lower song rates should be recorded for birds farther away from an ARU than birds singing closer to a unit. If the detection from ARUs at fixed locations resembles the detection by human observers, then the “no adjustment model” should perform similarly to the human-observer model (Chapter 2). Alternatively, if imperfect detection of the birds in ARU recordings alters song rates, models accounting for imperfect detection should have improved predictive performance. We assessed model performance by comparing sensitivity and specificity values for predicting each of the three breeding statuses.

Methods

Study species

OSFL are an appropriate model species to test the use of the BSSR relationship for prediction for three major reasons: 1) The OSFL 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 should affect detection positively, possibly counteracting the challenge of movement (i.e. the bird is loud enough on average that it can be heard from most spots on the territory regardless of where it is singing on the territory); 2) territories are typically large and non-overlapping (Altman & Sallabanks, 2012), unlike many songbirds with smaller territory sizes and thus, avoiding challenges associated with distinguishing between individuals; and 3) The OSFL song is a simple crisp, clear song, described by the mnemonic “quick, three-beers”, with little variation among individuals

(Robertson et al., 2009), making them easy to detect and distinguish from other species of songbirds.

Field methods

ARU recordings and breeding status data were collected from breeding OSFL males in northern Alberta (n = 8) and the Northwest Territories (n = 16), Canada from May 30 - July 22, 2016. Each week, the territorial males were tracked for 1-hour observation periods to confirm breeding statuses (single, paired, and feeding young) and document song rates (every 15 minutes; n = 4 per visit). For a more detailed description of the study area, breeding territory selection, and field methods, refer to Chapter 2 methods.

Acoustic data collection

Acoustic recordings were collected using Song Meters (SM4s and SM2s; Wildlife Acoustics Inc.) deployed in 24 breeding territories when males were considered to have settled (i.e. a male was found singing in the same area on two separate occasions between May 30 and June 8, 2016). One ARU was deployed near a primary singing perch at each territory, where males were more often observed singing (Wright, 1997). Each ARU was preprogrammed to record daily, for one continuous recording period between 15 minutes before sunrise until 30 minutes after sunrise (i.e. 45 minutes per day per ARU), to maximize OSFL song detection (Wright, 1997). Three ARUs were deployed in the Northwest Territories prior to May 5, 2016, which was > 3 weeks before spring arrival of the first migrants (E. Upham-Mills, unpublished data). Locations for these ARUs were based on detections from individuals during the previous breeding season (Knaggs, 2018). Average deployment date for the other 21 ARUs, placed at

primary singing perches, was June 10 (± 5 days; SD), 2016. All ARUs were retrieved in August or September 2016, resulting in an average number of 85 (± 53) days deployed/unit.

Acoustic data processing

We processed the ARU recordings using automated recognition software (hereafter "recognizer"; McLeod, 2015) built with the software Song Scope (Wildlife Acoustics, Inc.). A recognizer scans audio recording files using a moving window with a pattern-matching algorithm to produce a list of candidate target sounds (i.e. detections). The recognizer was used at a quality threshold of 50 and score threshold of 70, both of which are values to quantify "fit" of the detection to the pattern the algorithm is attempting to match (see Appendix 3 for score threshold decision process). Each detection was assigned as a true or false positive after visual and audio inspections of the spectrograms.

For this study, our sampling unit was the mean song count from nine consecutive 5-minute song counts collected between 15 minutes before sunrise and 30 minutes after sunrise for a given day (hereafter "song rate"). For each true positive song detection, the relative sound level value (i.e. "level" in the Song Scope output) was extracted and the mean level value was calculated for each song rate (i.e. the mean of all level values for each song included in a 5-minute song count was calculated, then the mean level of the four song counts was used for the song rate). Data were removed from the analysis when the breeding status was not confirmed or recordings were incomplete (e.g. full SD cards, empty batteries, etc.).

Statistical analysis

Our first objective was to compare singing rates as determined by ARUs versus human-observers to calibrate the hierarchical model built in Chapter 2 to ARU data. We wanted to differentiate between variation in singing rate due to individual effect of OSFLs (i.e. certain birds sing more than others) and effect of individual ARUs (i.e. a combination of the local sound environment and the individual bird's movement and singing patterns) to inform the song rate component of the hierarchical model, thus we evaluated the inclusion of random effect in models for both datasets. We modelled both the ARU and the human-observer song rates using one generalized linear regression (GLM) and one generalized linear mixed model (GLMM) each. All models included the response variable of song rates using a Poisson distribution, as predicted by breeding status. The human-observer GLM and GLMM both included a covariate for time relative to sunrise, because human observer song rates were not collected at a standardized time of day, as was the case for ARUs. Lastly, for the ARU model, we also tested for the effect of sound level to determine if it would improve model fit. We used Akaike's Information Criterion (AIC) to rank model fit for each dataset (ARU and human-observer) and considered a decrease in $AIC > 2$ as evidence for improved fit. We modelled GLMs and GLMMs using the R packages 'stats' and 'glmmADMB', respectively (Skaug et al., 2018). We explored coefficient values and significance (α level of $P < 0.05$) of each predictor and plots of fitted values from these models to compare their relative fit to the data and appropriateness for use in the predictive hierarchical models (objective 2).

For the second objective, we determined which of the four approaches to correcting ARU data (i.e. the no-adjustment model, the zero-truncated model, the zero-inflation model, and the distance model) produced the best predictive performance within the hierarchical framework. 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 framework adapted from the following human-observer model (Chapter 2):

Equation 1.

$$P(\textit{breeding status}|\textit{song rate}, \textit{date}, \textit{time}) \propto \underbrace{P(\textit{breeding status}|\textit{date})}_{\text{Component A}} \underbrace{P(\textit{song rate}|\textit{breeding status}, \textit{time})}_{\text{Component B}}$$

where the left-hand side probabilities for each breeding status conditional on song rate are proportional to the product of two components on the right-hand side. Component A is the marginal probability of observing each breeding status on a given day, and component B is the probability of song rate conditional on breeding status and time of day (i.e. the model tested in objective 1).

The no adjustment model used the hierarchical model with ARU data, but made no adjustment for imperfect detection of bird songs. Each observed song rate i followed a Poisson distribution with a mean of λ_i :

Equation 2.

$$\textit{song rate}_i \sim \textit{Poisson}(\lambda_i)$$

Mean song rate λ_i was log-normally distributed with a mean conditional on breeding status for a given bird ($\textit{breeding status}_j$; single, paired, or feeding young) a fixed effect, individual bird as a random effect ($\gamma_{\textit{bird},i}$):

Equation 3.

$$\log(\lambda_i) \sim \gamma_{\textit{bird},i} + \beta_0 * \textit{breeding status}_i,$$

where $\gamma_{\textit{bird},i}$ was normally distributed with a mean of 0 and an empirical Bayes estimate of the variance. $\textit{breeding status}_j$ followed a multinomial distribution conditional on ordinal date (\textit{date}_i):

Equation 4.

$$breeding\ status_i \sim multinomial(p_{single}, p_{paired}, p_{feeding\ young} | date_i).$$

The probability of each breeding status (p_{single} , p_{paired} , and $p_{feeding\ young}$) depended on the date and the three probabilities were constrained to sum to 1. The no adjustment model was built using JAGS language (packages ‘R2jags’ [Su & Yajima, 2015] and ‘UIjags’ [Kellner, 2017] see Appendix 4 for details).

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 $song\ count_i$.

A disproportionate number of zeros 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. 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 5.

$$p_{detection_i} \sim Bernoulli(\psi)$$

Equation 6.

$$song\ count_i | p_{detection_i} \sim Poisson(\lambda_i) * p_{detection_i}$$

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

Lastly, the distance model built on the zero-inflation model by adding a relative sound level ($level_i$) as a predictor to account for lower song rates resulting from increasing distances of a bird from the ARU. The full structure of the distance model consisted of Eqn. 4 – 6 and:

Equation 7.

$$\log(\lambda_i) \sim \gamma_{bird,i} + \beta_0 * breeding\ status_i + \beta_1 * level_i$$

using the same parameterization as Eqn. 3, with the addition of the $level_i$ covariate, which followed a Cauchy distribution using a location of 0, precision of 0.4, and degrees of freedom of 1.

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 five parallel chains. We ensured convergence using the Gelman-Rubin diagnostic (GR), considering a threshold of $GR > 1.1$ to indicate convergence (Gelman & Rubin, 1992; Plummer, Best, 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 song rates from one bird. The subset of song rates removed from the dataset (i.e. the test data) were 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 28 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.

Results

For each territory, we collected an average of 896 minutes of recordings (± 502 ; mean \pm SD) producing a sample size of 478 5-minute song rates (19.9 ± 11.2 song rates/territory).

For objective 1, we found that the three breeding status classes were significant predictors of song rate in both the ARU and human-observer GLMs (i.e. with no random effect of individual; Table 3.1). However, the datasets differed in two major ways: 1) song rates were lower in the ARU dataset (Fig. 3.1a,c), and 2) the order of the highest to lowest song rate as a response to the three breeding statuses was different (Fig. 3.1a,c). In the human-observer data, the song rates were predicted to be the highest when birds were single, then when paired, and lowest when feeding young (Fig. 3.1a). The order differed with ARU data; here paired birds had the highest predicted song rate, followed by single and feeding young (Fig. 1c). Adding the random effect of bird to the human-observer data had a minimal effect on effect size, standard error, or significance of breeding status effect on song rate (Fig. 3.1b) which was not true for the ARU-data (Fig. 3.1d; Table 3.1). With the ARU dataset, random effect changed the order of single, paired and feeding young song rate predictions to align with the human-observer dataset (Fig 3.1a,d), but the effect of single and paired status on song rate were no longer significant (Table 3.1). The distance proxy (i.e. mean sound level) had a significant positive effect on documented song rate (Fig. 3.2, Table 3.1).

AIC values for both the ARU and human observer datasets decreased after random effect of individual was included in the model (Table 3.1). Adding sound level as a covariate to the ARU model further decreased the AIC value, indicating a better model fit (Table 3.1).

Table 3. 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 for ARU and human-observer data. SR = song rate, BS = breeding status, 1|ind = random effect of individual.

Model	Data	Predictor variables					AIC value
		Time (Minutes from Sunrise)	Breeding Status: Single	Breeding Status: Paired	Breeding Status: Feeding Young	Level (Relative Sound Level)	
SR ~ BS + time	Human	-0.48 ± 0.03*	2.84 ± 0.05*	1.91 ± 0.04*	1.09 ± 0.11*	-	2277.8
SR ~ BS + time + 1 ind		-0.55 ± 0.03*	2.93 ± 0.23*	1.39 ± 0.20*	0.53 ± 0.24*	-	1725.1
SR ~ BS	ARU	-	-0.32 ± 0.13*	0.44 ± 0.05*	-1.88 ± 0.25*	-	1985.3
SR ~ BS + 1 ind		-	0.06 ± 0.53	-0.32 ± 0.33	-3.05 ± 0.42	-	1246.1
SR ~ BS + level + 1 ind		-	0.00 ± 0.52	-0.34 ± 0.32	-2.96 ± 0.40*	0.23 ± 0.07*	1237.9

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

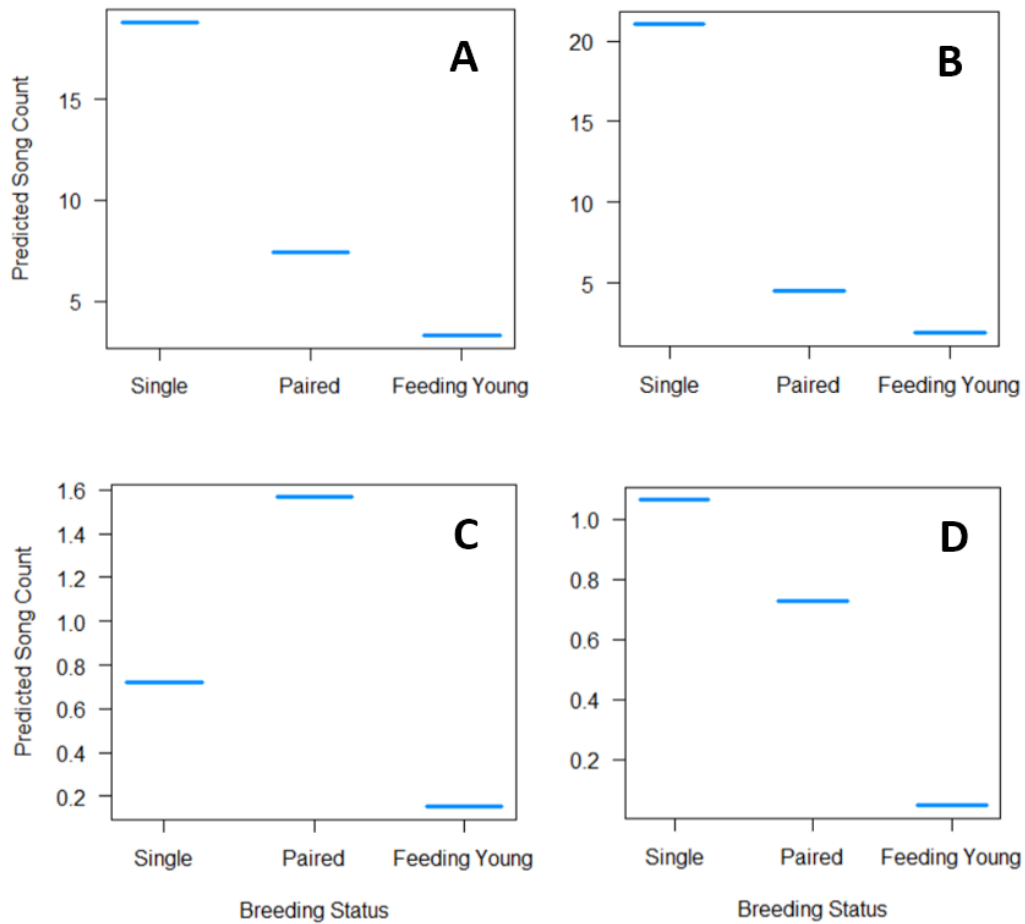


Figure 3. 1. Predicted Olive-sided Flycatcher song rates for three breeding status classes from: A) a generalized linear (GLM) regression using human-observer* data, B) a generalized linear mixed effects (GLMM) model using human-observer data, C) a GLM using ARU data, and D) a GLMM using ARU data. GLMMs included random effect of individual (n = 28 individuals, and 24 individuals, for human-observer and ARU data, respectively). * Human-observer models included a time co-variate (minutes past sunrise), which was held at the median value for these predictions (113 minutes).

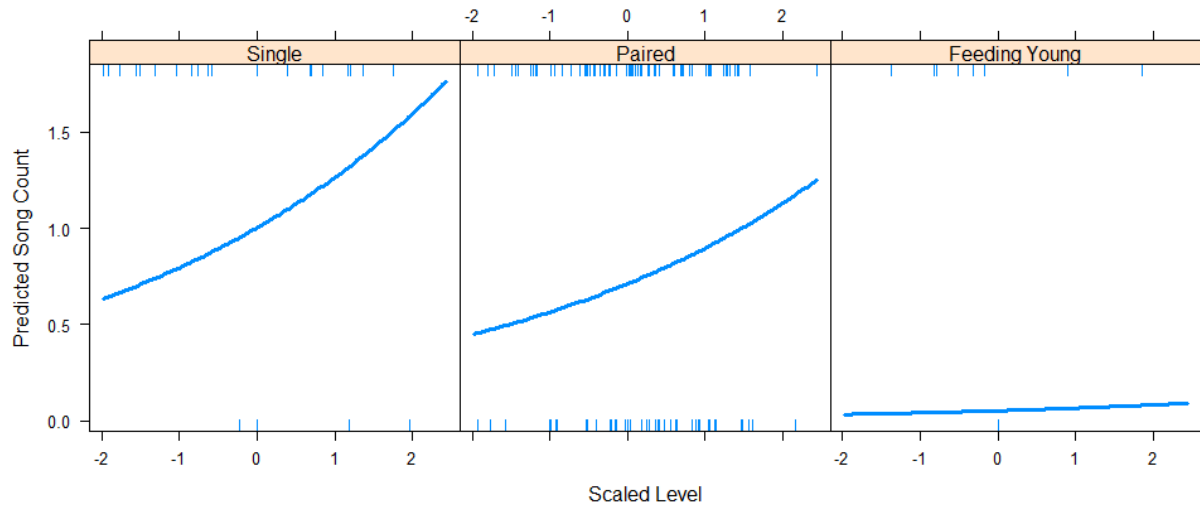


Figure 3. 2. Predicted song rates using ARU data for three breeding statuses over a range of sound levels (scaled and mean-centered).

The four hierarchical models based on ARU data rarely predicted the breeding status single correctly (sensitivity <0.07) compared to models based on human-observer 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 human-observer 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 $0.34 (\pm 0.39)$, $0.49 (\pm 0.41)$, and $0.44 (\pm 0.41)$ 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 human-observer model, which underpredicted paired by more than 30%. All four ARU models underpredicted single by $\geq 10\%$, which was opposite to the human-observer model. The zero-truncated model was the only model to overpredict feeding young, as did the human-observer model.

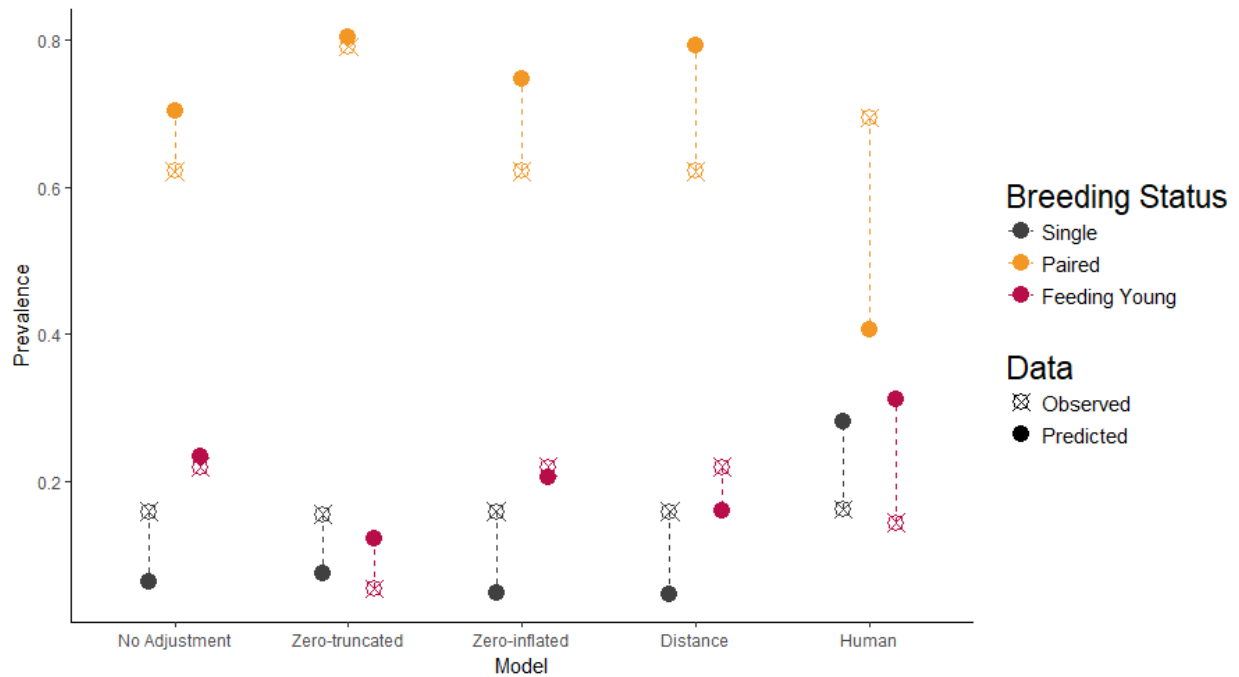


Figure 3. 3. Prevalence of observed versus predicted breeding statuses for four ARU data models and the human-observer data model from Chapter 2.

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 human-observer 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
		<i>Specificity</i>	<i>0.94</i>	<i>0.46</i>	<i>0.89</i>
Zero-truncated	Removed zeros from dataset	Sensitivity	0.00	0.77	0.25
		<i>Specificity</i>	<i>0.91</i>	<i>0.06</i>	<i>0.89</i>
Zero-inflated	Modelled song detection/non-detection first, then song rate	Sensitivity	0.04	0.83	0.61
		<i>Specificity</i>	<i>0.95</i>	<i>0.39</i>	<i>0.91</i>
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
		<i>Specificity</i>	<i>0.95</i>	<i>0.28</i>	<i>0.92</i>
Human-based song rate model (Chapter 2)	None (no need because observers always confirmed bird presence)	Sensitivity	0.69	0.50	0.87
		<i>Specificity</i>	<i>0.80</i>	<i>0.82</i>	<i>0.78</i>

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 human-observer 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, 2016a).

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 collected the data; JR, AC and EU conceived the ideas for the statistical analysis methodology; EU and AC analyzed the data; EU led the writing of the manuscript. All authors contributed critically to the drafts.

Chapter 4: General conclusion

Summary

The purpose of this study was to determine the best modelling approach to predict breeding status from singing rate in the OSFL and test its use with ARUs. I successfully addressed the first challenge of statistically calibrating the cause and effect relationship between breeding status and singing rate, respectively, to derive an indirect measure of breeding status. I found that a hierarchical modelling framework, based on song rates determined by human observers, produced highly sensitive breeding status predictions. This was in contrast with the multinomial logistic regression and classification tree models, which did not account for the true relationship between cause and effect and produced low sensitivity values for the breeding status classes of conservation interest, single and feeding young. When I applied the ARU-derived data to the hierarchical model, sensitivity values for predicting the three breeding statuses (single, paired and feeding young) were low. Song rates collected from ARUs likely did not represent the actual singing behaviours due to low detectability. Thus, the ARU-based data did not reflect the true relationship between singing rate and breeding status. The detectability issues I identified were low accuracy of the recognizer to detect OSFL songs in the recordings and movement of the individual bird in and out of the detection range of an ARU. A large portion of these detectability issues observed in our study can be accounted for through improvements in the acoustic data processing (i.e. better species recognition with different software or human listening) and modifications to my study design (i.e. increasing number of ARUs in an area). Results from my thesis provides an important contribution to the field of bioacoustics by providing a proof of concept for a new approach to collect valuable and efficient demographic monitoring of forest songbirds to better inform conservation actions.

Model improvement using different song metrics

I used song rate as a metric of singing behaviour to infer breeding status in the Olive-sided Flycatcher, but this metric is not the only option to quantify singing behaviour in birds. I propose that the modelling framework used in this study can likely be applied to other songbird species, but the model will first have to be calibrated for those species. For many songbird species, singing behaviour is significantly different between single and paired individuals (Dussourd & Ritchison, 2003; Gibbs & Wenny, 1993; Radesäter et al., 1987; Stacier et al., 1996). For example, Yellow-breasted Chats (*Icteria virens*) spend much less time singing after pairing (Dussourd & Ritchison, 2003). Additionally, incremental declines in percent of time spent singing between arrival on breeding territory and post-fledging indicate that this species may be a candidate for predicting breeding status using singing behaviour. Song rate within singing bouts does not vary with breeding status for Yellow-breasted Chats (Dussourd & Ritchison, 2003). Thus, a predictive model for this species could use percent of time spent singing instead of the metric I used for OSFL, mean number of songs in a 5-minute period (i.e. singing rate). With 170 pairs of the southern mountain population of Yellow-breasted Chat (subspecies *auricollis*) remaining in Canada (Environment and Climate Change Canada, 2016c) and a classification of Endangered by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC), our methods could be applied to inform conservation of this threatened taxon.

Model improvement using different breeding status classes

My study was designed to predict only three breeding status classes (single, feeding young and paired) because I wanted to test model accuracy for a well-established behavioural relationship between breeding status and song rate in the Olive-sided Flycatcher (Wright, 1997).

Singing behaviour was not as well quantified when nests either fail or successfully fledge young (Altman & Sallabanks, 2012; Wright, 1997). For some species, there is evidence suggesting that singing behaviour may be differential at these later states because males tend to re-initiate high singing rates when his mate dies or abandons (Amrhein et al., 2002; Stacier et al., 1996). Thus, predicting reproductive success in addition to breeding status may be achievable by calibrating a song behaviour model to predict fledge or fail status later in the breeding season. It is possible that predictions of breeding status may be more accurate using multiple sets of binary responses, e.g. paired versus unpaired early in the breeding season and fledged versus failed later in the season (conditional on pairing status). Thus, future studies could test modelling singing behaviour of multiple binary states in different phases of the breeding cycle instead of a multinomial response for predictive purposes. A hierarchical modelling framework provides the flexibility to account for these conditional relationships.

Recommendations for model application to other songbird species

1) *Consider the research question.* The study design influences both the scale of inferences that can be generated and the predictive ability of the model. If the objective is to make predictions about individual birds, then minimizing detectability issues and maximizing temporal coverage (within breeding season) and spatial coverage (within territory) will be important to ensure the best predictive ability. Alternatively, if the goal is to make predictions at a population scale, maximizing coverage over large spatial extent may be more important than intensive monitoring (i.e. high frequency visits by human observers or many ARUs recording daily) at a few locations.

2) *Consider behaviour of the focal species.* Territory size, peak of acoustic activity levels (within day and breeding season), and densities of conspecific are all life history traits important to understand to inform appropriate sampling design (i.e. human vs. ARU-based data, recording schedule, and need for individual recognition through visual or vocal identification). For example, if the species has a large territory, spatial coverage within a territory or accounting for bird movement in the model may be required to ensure predictive ability. This could be accounted for by increasing the number of ARUs deployed on a territory or monitoring bird movement in the area with ARUs by deploying radio transmitters or GPS units. For species with small territories, individual recognition within recordings may be the challenge if conspecific densities are high enough to result overlap of detectable individuals. In this case, acoustic processing would need to be able to distinguish individuals.

3) *Calibration.* Quantifying the behavioural mechanisms underlying an observed biological state is a key step in ensuring correct and accurate modeling for prediction of this state (Stephens et al., 2015). In the case of BSSR models, this means determining the song metric with the strongest response to breeding status for the species of interest. Calibration of a model for a given time and location may also be important if the data used to build a model was from a different time or place even if it is from the same focal study species.

4) *Validation.* Testing the predictive ability of the model is also an important step in predictive modelling. This can be accomplished in a many ways, including holdout methods, cross-validation, and bootstrapping (Kohavi, 1995). An important part of validation is interpreting predictive ability, which can be subjective and depends on the research question in mind. In our study, I decided to optimize prediction of the breeding statuses classes of conservation interest,

single and feeding young. However, the need to optimize some classes over others may not be required if the data are comprised of classes with equal prevalence.

Researchers should consider these steps when designing a model to acoustically monitor breeding status, but I also recommend taking advantage of the few existing studies monitoring demographics in vocal animals by placing ARUs at study areas for future use in model-building and monitoring.

Conclusions

Acoustic monitoring to infer breeding status in a songbird is a novel approach, despite the relatively large amount of research into functions of vocalization in birds (Boncoraglio & Saino, 2007; Collins, 2004; Nowicki & Searcy, 2004) and growing popularity of bioacoustics in avian conservation research (Fristrup & Mennitt, 2012; Shonfield & Bayne, 2017). The wealth of literature on studies in singing behaviour functions associated with breeding status indicates that any of these empirical observations can likely be used to indirectly measure breeding status, and possibly reproductive success, in an abundance of vocal species. Advancements of ARU and acoustic processing technology, and models for predicting breeding status such as those presented in this thesis, will promote efficient collection of large-scale data on population dynamics. Existing ARU recordings can be mined for information on breeding status for focal species. Land managers could use such information to infer habitat quality, impact of anthropogenic disturbance, and effects of climate change on vocal species. Bioacoustics methods could address significant knowledge gaps in the recovery of many Species at Risk songbirds in Canada (Environment and Climate Change Canada, 2016b, 2016c, 2016a) by providing an efficient method to collect information on breeding success and the factors affecting it. The analytical framework and management implications presented in this thesis are not limited to

avian systems because it could be applied to any vocal species, such as cetaceans (Mellinger et al., 2007), primates (Todt et al., 1988) and other mammals (Lynch et al., 2013).

Literature Cited

- Alberta Biodiversity Monitoring Institute. (2012). *Monitoring activity report: lower Athabasca planning region 2010 season*.
- Allredge, M. W., Simons, T. R., & Pollock, K. H. (2007). Factors affecting aural detections of songbirds. *Ecological Applications*, *17*(3), 948–955. <https://doi.org/10.1890/06-0685>
- Altman, B., & Sallabanks, R. (2012). Olive-sided Flycatcher (*Contopus cooperi*), version 2.0. Retrieved from <https://doi-org.login.ezproxy.library.ualberta.ca/10.2173/bna.502>
- 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>
- Anders, A. D., & Marshall, M. R. (2005). Increasing the accuracy of productivity and survival estimates in assembling landbird population status. *Conservation Biology*, *19*(1), 66–74.
- 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>
- Armstrong, E. A. (1973). *A Study of Bird Song* (2nd ed.). Dover Publications.
- 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.
- Bhapkar, V. P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. *American Statistical Association Journal*, *61*:313, 228–235.
- Bibby, C. J., Burgess, N. D., Hill, D. A., & Mustoe, S. (2000). *Bird census techniques*. London: Elsevier.
- Biondi, F., & Waikul, K. (2004). DENDROCLIM2002: A C ++ program for statistical calibration of climate signals in tree-ring chronologies. *Computers & Geosciences*, *30*, 303–311. <https://doi.org/10.1016/j.cageo.2003.11.004>
- 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>
- Bohning-Gaese, K., Taper, M. L., & Brown, J. H. (1993). Are declines in North American insectivorous songbirds due to causes on the breeding range? *Conservation Biology*, *7*(1), 76–86. Retrieved from <http://www.jstor.org/stable/2386644>
- Boncoraglio, G., & Saino, N. (2007). Habitat structure and the evolution of bird song: A meta-analysis of the evidence for the acoustic adaptation hypothesis. *Functional Ecology*, *21*(1), 134–142. <https://doi.org/10.1111/j.1365-2435.2006.01207.x>
- Brieman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Boca Raton, FL: CRC press.

- Brooks, S. P., & Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434–455. <https://doi.org/10.1080/10618600.1998.10474787>
- Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed)*. *Ecological Modelling* (Vol. 172). <https://doi.org/10.1016/j.ecolmodel.2003.11.004>
- 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>
- Campos-Cerqueira, M., & Aide, T. M. (2016). Improving distribution data of threatened species by combining acoustic monitoring and occupancy modelling. *Methods in Ecology and Evolution*, 7(11), 1340–1348. <https://doi.org/10.1111/2041-210X.12599>
- Catchpole, C. K., & Slater, P. J. B. (2008). *Bird song: Biological 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>
- Christoferson, L. L., & Morrison, M. L. (2001). Integrating methods to determine breeding and nesting status of 3 western songbirds. *Wildlife Society Bulletin*, 29(2), 688–696. Retrieved from <http://www.jstor.org/stable/3784196>
- 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., & Tschardtke, 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>
- De'Ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, 81(11), 3178–3192. [https://doi.org/10.1890/0012-9658\(2000\)081\[3178:CARTAP\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2)

- DeSante, D. F., Williams, O. E., & Burton, K. M. (1993). *The Monitoring Avian Productivity and Survivorship (MAPS) Program: Overview and Progress*. In: Finch, Deborah M.; Stangel, Peter W. (eds.). *Status and management of neotropical migratory birds: September 21-25, 1992, Estes Park, Colorado*. Gen. Tech. Rep. R.
- 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>
- Donovan, T., & Flather, C. (2002). Relationships among North American songbird trends, habitat fragmentation, and landscape occupancy, 12(2), 364–374.
- 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.
- Ecosystem Classification Group. (2009). *Taiga plains*. Yellowknife, NT, Canada.
- 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. (2016a). *Recovery strategy for Olive-sided flycatcher (Contopus cooperi) in Canada*. *Species at Risk Act Recovery Strategy Series*. Ottawa.
- Environment and Climate Change Canada. (2016b). *Recovery strategy for the Canada warbler (Cardellina canadensis) in Canada*. *Species at Risk Act Recovery Strategy Series*. Ottawa.
- Environment and Climate Change Canada. (2016c). *Recovery Strategy for the Yellow-breasted Chat auricollis subspecies (Icteria virens auricollis) (Southern Mountain population) in Canada*. *Species at Risk Act Recovery Strategy Series*. Ottawa.
- Environment and Climate Change Canada. (2017). North American Breeding Bird Survey - Canadian Trends Website, Data-version 2015. Environment and Climate Change Canada, Gatineau, Quebec, K1A 0H3.
- 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 subsongbird, are individually distinctive but do not vary geographically. *Bioacoustics*, 22(2), 137–151. <https://doi.org/10.1080/09524622.2012.740174>
- Fretwell, S. D., & Lucas Jr, H. L. (1970). On territorial behaviour and other factors influencing habitat distribution in birds. I. Theoretical development. *Acta Biotheoretica*, 14, 16–36.
- Fristrup, K. M., & Mennitt, D. (2012). Bioacoustical monitoring in terrestrial environments.

- Acoustics Today*, (July), 16–24.
- 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.
- Gottlander, K. (1987). Variation in the song rate of the male pied flycatcher *Ficedula hypoleuca*: causes and consequences. *Animal Behaviour*, 35(4), 1037–1043.
[https://doi.org/10.1016/S0003-3472\(87\)80160-4](https://doi.org/10.1016/S0003-3472(87)80160-4)
- 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., Solymos, P., Bayne, E. M., Fontaine, T., Cumming, S. G., Schmiegelow, F., & Stralberg, D. (2014). Analyses to support critical habitat identification for Canada Warbler, Olive-sided Flycatcher, and Common Nighthawk, Final Report 2. Boreal Avian Modelling Project, 157.
- 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.
- Hall, R. I., & Smol, J. P. (1992). A weighted — averaging regression and calibration model for inferring total phosphorus concentration from diatoms in British Columbia (Canada) lakes, 417–434.
- Hamilton, J. D. (1986). A standard error for the estimated state vector of a state-space model. *Journal of Econometrics*, 33, 387–397.
- Harris, I. (1989). Predictive fit for natural exponential families. *Biometrika*, 76(4), 675–684.
- 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>
- Holmes, R. T., Sherry, T. W., Marra, P. P., & Petit, K. E. (1992). Multiple brooding and productivity of a Neotropical migrant, the Black-throated blue warbler (*Dendroica caerulescens*), in an unfragmented temperate forest. *The Auk*, 109(2), 321–333. Retrieved from <https://www.jstor.org/stable/4088201>
- Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, 85(1), 3–28.
- Hunt, A. R., Bayne, E. M., & Haché, S. (2017). Forestry and conspecifics influence Canada Warbler (*Cardellina canadensis*) habitat use and reproductive activity in boreal Alberta, Canada. *The Condor*, 119(4), 832–847. <https://doi.org/10.1650/CONDOR-17-35.1>

- Jahn, O., Ganchev, T. D., Marques, M. I., & Schuchmann, K.-L. (2017). Insights into the behavioral ecology of a tropical bird. *PLoS ONE*, *12*(1), 1–29. <https://doi.org/10.1371/journal.pone.0169041>
- 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>
- Johnson, M. D. (2007). Measuring habitat quality: A review. *The Condor*, *109*, 489–504.
- 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.
- Kéry, M., & Royle, J. A. (2016a). Distribution, Abundance, and Species Richness in Ecology. *Applied Hierarchical Modeling in Ecology*, 3–18. <https://doi.org/10.1016/B978-0-12-801378-6.00001-1>
- Kéry, M., & Royle, J. A. (2016b). *What Are Hierarchical Models and How Do We Analyze Them?* *Applied Hierarchical Modeling in Ecology*. <https://doi.org/10.1016/B978-0-12-801378-6.00002-3>
- 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. D. (2018). *Effects of burn severity and time since fire on songbird communities in the northern boreal forest*. University of Alberta. <https://doi.org/10.15713/ins.mmj.3>
- 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.
- Kozłowski, C. P. (2005). *Embryonic vocalizations and parent-offspring communication in the Eastern screech owl*. Cornell University.
- Kuhn, M., & Johnson, K. (2016). *Applied Predictive Modeling* (5th ed.). New York: Springer Nature.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... Hunt, T. (2017). 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>
- Lele, S. R. (2014). *Is non-informative Bayesian analysis appropriate for wildlife management: survival of San Joaquin Kit Fox and declines in amphibian populations?* Edmonton, Alberta. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1502/1502.00483.pdf>
- Lindenmayer, D. B., & Likens, G. E. (2011). Direct measurement versus surrogate indicator species for evaluating environmental change and biodiversity loss. *Ecosystems*, 14(1), 47–59. <https://doi.org/10.1007/s10021-010-9394-6>
- Liu, W., Kroodsma, D. E., & Yasukawa, K. (2007). Dawn and daytime singing behavior of Chipping sparrows (*Spizella passerina*). *The Auk*, 124(1), 44–52.
- Lynch, E., Angeloni, L., Fristrup, K., Joyce, D., & Wittemyer, G. (2013). The use of on-animal acoustical recording devices for studying animal behavior. *Ecology and Evolution*, 3(7), 2030–2037. <https://doi.org/10.1002/ece3.608>
- 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.
- Martin, T. E., & Geupel, G. R. (1993). Nest-monitoring plots: methods for locating nests and monitoring success. *Journal of Field Ornithology*, 64(4), 507–519. Retrieved from <http://www.jstor.org/stable/4513862> <http://about.jstor.org/terms>
- Martin, T. E., Paine, C., Conway, C. J., Hochachka, W. M., Allen, P., & Jenkins, W. (1997). BBIRD (Breeding Biology Research & Monitoring Database) Field protocol. *Montana Cooperative Wildlife Research Unit, University of Montana, Missoula, MT*, 59812, 59812(406), 1–64.
- Matsuoka, S. M., Bayne, E. M., Sóllymos, 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/>
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2), 153–157.
- Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P., & Matsumoto, H. (2007). An overview of fixed passive acoustic observation methods for cetaceans. *Oceanography*,

20(4), 36–45.

- Michel, N. L., Smith, A. C., Clark, R. G., Morrissey, C. A., & Hobson, K. A. (2016). Differences in spatial synchrony and interspecific concordance inform guild-level population trends for aerial insectivorous birds. *Ecography*, 39(8), 774–786. <https://doi.org/10.1111/ecog.01798>
- 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>
- Northrup, J. M., & Gerber, B. D. (2018). A comment on priors for Bayesian occupancy models. *PLoS ONE*, 13(2), 1–14. <https://doi.org/https://doi.org/10.1371/journal.pone.0192819>
- Nowicki, S., & Searcy, W. A. (2004). Song function and the evolution of female preferences: Why birds sing, why brains matter. *Annals of the New York Academy of Sciences*, 1016, 704–723. <https://doi.org/10.1196/annals.1298.012>
- Olivas, P. C., Oberbauer, S. F., Clark, D. B., Clark, D. A., Ryan, M. G., O'Brien, J. J., & Ordoñez, H. (2013). Comparison of direct and indirect methods for assessing leaf area index across a tropical rain forest landscape. *Agricultural and Forest Meteorology*, 177, 110–116. <https://doi.org/10.1016/j.agrformet.2013.04.010>
- Osborne, C. (1991). Statistical Calibration: A review. *International Statistical Review*, 59(3), 309–336. <https://doi.org/10.1021/pr8011107>
- 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., Sóllymos, P., & Bayne, E. M. (2017). Potential benefits of augmenting road-based breeding bird surveys with autonomous recordings. *Avian Conservation and Ecology*, 12(2), art18. <https://doi.org/10.5751/ACE-01087-120218>
- Pardieck, K. L., Ziolkowski Jr., D. J., Lutmerding, M., & Hudson., M.-A. R. (2015). *North American Breeding Bird Survey Dataset 1966 - 2015*. Retrieved from <https://doi.org/10.5066/F76972V8>
- Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., & Matthiopoulos, J. (2008). State – space models of individual animal movement. *Trends in Ecology & Evolution*, 87-. <https://doi.org/10.1016/j.tree.2007.10.009>
- 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&doi=10.1648/0273-8570\(2002\)073\[0387:CDBPAS\]2.0.CO;2](http://www.bioone.org/perlserv/?request=get-document&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.
- Plummer, M., Stukalov, A., & Denwood, M. (2016). rjags: Bayesian graphical models using MCMC. Retrieved from <https://cran.r-project.org/web/packages/rjags/rjags.pdf>
- Radesäter, T., Jakobsson, S., Andbjer, N., Bylin, A., & Nyström, K. (1987). Song rate and pair formation in the willow warbler, *Phylloscopus trochilus*. *Animal Behaviour*, 35(6), 1645–1651. [https://doi.org/10.1016/S0003-3472\(87\)80057-X](https://doi.org/10.1016/S0003-3472(87)80057-X)
- Ripley, B., & Venables, W. (2016). nnet: Feed-forward neural networks and multinomial log-linear models. Retrieved January 15, 2018, from <http://www.stats.ox.ac.uk/pub/MASS4/>
- Robbins, C. S., Sauer, J. R., Greenberg, R. S., & Droege, S. (1989). Population declines in North American birds that migrate to the neotropics. *Proceedings of the National Academy of Sciences*, 86(19), 7658–7662. <https://doi.org/10.1073/pnas.86.19.7658>
- 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.
- Robertson, B. A., & Hutto, R. L. (2007). Is selectively harvested forest an ecological trap for Olive-sided flycatchers? *The Condor*, 109(1), 109–121. [https://doi.org/10.1650/0010-5422\(2007\)109\[109:ISHFAE\]2.0.CO;2](https://doi.org/10.1650/0010-5422(2007)109[109:ISHFAE]2.0.CO;2)
- Robinson, P. W., Tremblay, Y., Crocker, D. E., Kappes, M. A., Kuhn, C. E., Shaffer, S. A., ... Costa, D. P. (2007). A comparison of indirect measures of feeding behaviour based on ARGOS tracking data. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 54(3–4), 356–368. <https://doi.org/10.1016/j.dsr2.2006.11.020>
- Sauer, J. R., Pardieck, K. L., Ziolkowski, D. J., Smith, A. C., Hudson, M. A. R., Rodriguez, V., ... Link, W. A. (2017). The first 50 years of the North American Breeding Bird Survey. *The Condor*, 119(3), 576–593. <https://doi.org/10.1650/CONDOR-17-83.1>
- 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), art14. <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.”
- Slagsvold, T. (1977). Bird song activity in relation to breeding cycle, spring weather, and environmental phenology. *Ornis Scandinavica*, 8(2), 197–222. Retrieved from <http://www.jstor.org/stable/3676105>
- 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>
- Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. (1995). *BUGS: Bayesian inference using Gibbs sampling. Version 0.5*. Cambridge, UK: MRC Biostatistics Unit, Institute of Public Health.
- Stacier, C. A., Spector, D. A., & Horn, A. G. (1996). *The dawn chorus and other diel patterns in acoustic signaling*. In: Kroodsma DE, Miller EH (eds) *Ecology and evolution of acoustic communication in birds*. Ithica: 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>
- Stephens, P. A., Pettorelli, N., Barlow, J., Whittingham, M. J., & Cadotte, M. W. (2015). Management by proxy? The use of indices in applied ecology. *Journal of Applied Ecology*, 52(1), 1–6. <https://doi.org/10.1111/1365-2664.12383>
- 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>
- Strong, W. L., & Leggat, K. R. (1992). *Ecoregions of Alberta. Alberta Forestry, Lands and Wildlife*. Edmonton, Alberta.
- Su, Y., & Yajima, M. (2015). Package ‘`R2jags`’: Using R to run “JAGS.”
- ter Braak, C. J. F. (1995). Non-linear methods for multivariate statistical calibration and their use in palaeoecology: a comparison of inverse (k-nearest neighbours, partial least squares and weighted averaging partial least squares) and classical approaches. *Chemometrics and Intelligent Laboratory Systems*, 28(1), 165–180. [https://doi.org/10.1016/0169-7439\(95\)80048-E](https://doi.org/10.1016/0169-7439(95)80048-E)
- Ter Braak, C. J. F., & Barendregt, L. G. (1986). Weighted averaging of species indicator values: Its efficiency in environmental calibration. *Mathematical Biosciences*, 78(1), 57–72. [https://doi.org/10.1016/0025-5564\(86\)90031-3](https://doi.org/10.1016/0025-5564(86)90031-3)
- Therneau, T., Atkinson, B., & Ripley, B. (2018). `rpart`: Recursive partitioning and regression trees. Retrieved January 15, 2018, from <https://cran.r-project.org/web/packages/rpart/rpart.pdf>
- Thorpe, W. H. (1961). *Bird-song*. Cambridge: Cambridge University Press.

- Todt, D., Goedeke, P., & Symmes, D. (1988). *Primate vocal communication* (1st ed.). Berlin Heidelberg: Springer-Verlag. <https://doi.org/10.1007/978-3-642-73769-5> e-ISBN-13:
- Van Horne, B. (1983). Density as a misleading indicator of habitat quality. *The Journal of Wildlife Management*, 47(4), 893–901.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. (J. Chambers, W. Eddy, W. Hardle, S. Sheather, & L. Tierney, Eds.) (4th ed.). New York, NY: Springer Science + Business Media.
- Ver Hoef, J. M. (1996). Parametric empirical Bayes methods for ecological applications. *Ecological Applications*, 6(4), 1047–1055.
- Vickery, P. D., Hunter Jr., M. L., & Wells, J. V. (1992). Use of a new reproductive index to evaluate relationship between habitat quality and breeding success. *The Auk*, 109(4), 697–705.
- Visser, M. E., Both, C., & Lambrechts, M. M. (2004). Global climate change leads to mistimed avian reproduction. *Advances in Ecological Research*, 35(04), 89–110. [https://doi.org/10.1016/S0065-2504\(04\)35005-1](https://doi.org/10.1016/S0065-2504(04)35005-1)
- Wells, J., Childs, D., Reid, F., Smith, K., Darveau, M., & Courtois, V. (2014). *Boreal birds need half: Maintaining North America's bird nursery and why it matters*.
- Wildlife Acoustics, I. (2011). Song Scope: Bioacoustics software version 4.0 documentation.
- Wilson, D. M., & Bart, J. (1985). Reliability of singing bird surveys: effects of song phenology during the breeding season. *The Condor*, 87(1), 69–73.
- 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.
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... Dan, J. H. (2008). *Top 10 algorithms in data mining*. <https://doi.org/10.1007/s10115-007-0114-2>
- 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>

Appendices

Appendix 1: Model selection results for each of the three modelling approaches from Chapter 2 (multinomial logistic regression, hierarchical model and classification tree), including individual component models of the hierarchical model

Table 1. Model selection for the effects of the predictor variables song rate, date, time and latitude on breeding status (single, paired, feeding young) using six *a priori* multinomial logistic regressions and their residual degrees of freedom, residual deviance and AIC values. The top model, indicated in bold, was identified as the lowest AIC value by a difference of > 2 AIC.

Predictor Variables	Resid. Df	Resid. Dev	AIC
Breeding Status ~ Song Rate	4	241.8392	249.8392
Breeding Status ~ Date	4	185.2757	193.2757
Breeding Status ~ Song Rate + Date	6	170.4263	182.4263
Breeding Status ~ Date + Song Rate*Time	10	164.6213	184.6213
Breeding Status ~ Song Rate + Date*Latitude	10	168.1566	188.1566
Breeding Status ~ Song Rate*Time + Date*Latitude	14	161.9928	189.9928

Table 2. Model selection by AIC values for the first component of the hierarchical model; a multinomial logistic regression model to predict breeding status. The top model, indicated in bold, was identified as the lowest AIC value by a difference of > 2 AIC.

Model	AIC
Breeding Status ~ 1	268.9
Breeding Status ~ date	193.3
Breeding Status ~ date + latitude	200.3

Table 3. Model selection and AIC values for second component of the hierarchical model. Two underlying models were tested using six different model families, including Poisson, Zero-inflated Poisson with one zero-inflation co-efficient, Zero-inflated Poisson allowing zero-inflation to vary with breeding status, Negative Binomial, Zero-inflated Negative Binomial with

one zero-inflation co-efficient, and Zero-inflated Negative Binomial allowing zero-inflation to vary with breeding status.

Model	Family	AIC
Song Rate ~ Breeding Status	Poisson	2539.2
	Zero-inf. Poisson 1	1868.1
	Zero-inf. Poisson bs	1857.2
	Neg. Binom.	948.7
	Zero-inf. Neg. Binom. 1	950.6
	Zero-inf. Neg. Binom. bs	946.9
Song Rate ~ Breeding Status + Time	Poisson	2277.8
	Zero-inf. Poisson 1	1768.8
	Zero-inf. Poisson bs	1758.0
	Neg. Binom.	936.4
	Zero-inf. neg. Binom. 1	938.4
	Zero-inf. Neg. Binom. bs	935.8

Table 4. Marginal homogeneity test results, comparing proportions of predicted breeding status classes outputted by each model to the true breeding status proportions. *significance at $p < 0.05$.

Model	Bhapkhar χ^2	Bhapkar P -value	McNemar χ^2 (single)	McNemar P -value (single)	McNemar χ^2 (paired)	McNemar P -value (paired)	McNemar χ^2 (feeding young)	McNemar P -value (feeding young)
MR	13.15	0.00 *	7.26	0.01 *	11.03	0.00 *	2.4	0.12
HM	43.66	0.00 *	9.26	0.00 *	31.64	0.00 *	20.48	0.00 *
CT	12.26	0.00 *	5.94	0.01 *	10.3	0.00 *	3.5	0.06

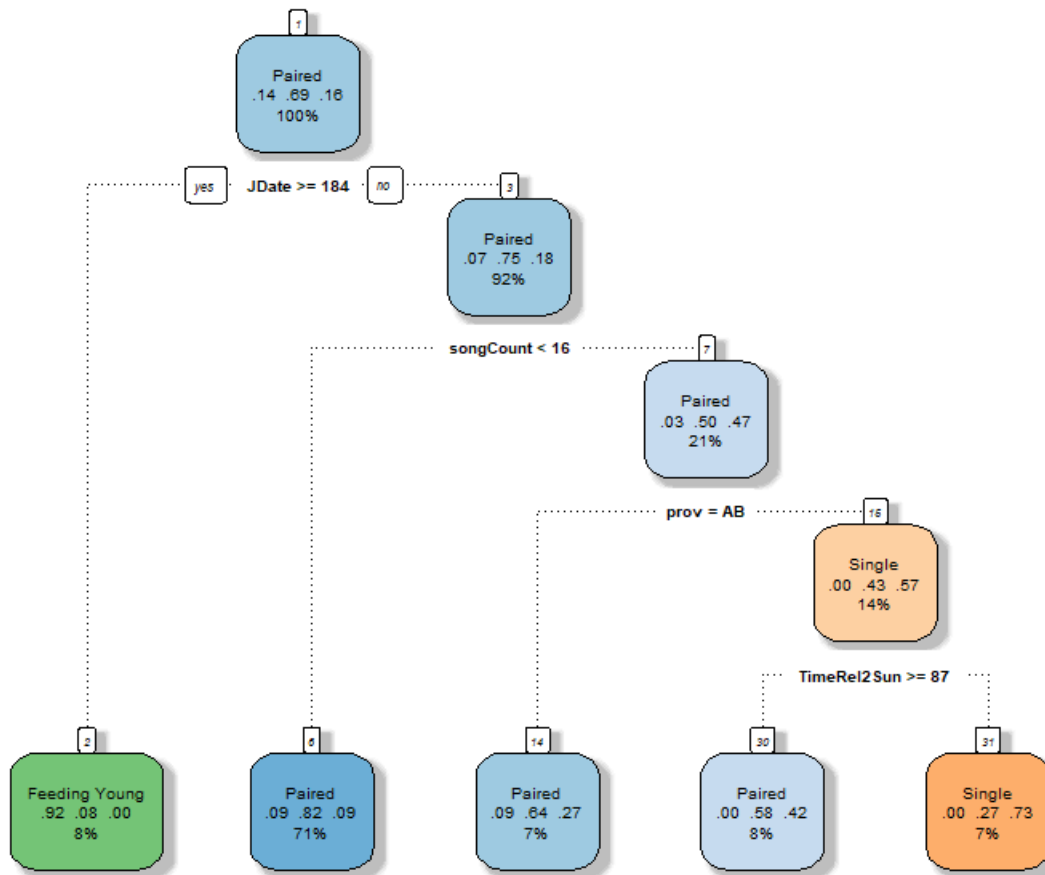


Figure 1. Top classification tree model. The class (single, paired, feeding young) in each box is class with the highest probability for that node (and would be the final prediction, for the terminal nodes). At each split, the predictor variable and split value is indicated, with a left split agreeing and right split disagreeing. The three probabilities in each box represent the probability of feeding young, paired and single in that split and the percentage value represents the proportion of all observations in that node.

Appendix 2: Hierarchical model bug (Bayesian inference Using Gibbs Sampling) file for use with

'rjags' code (R package) used in Chapter 2, which identifies the model specification, data

specification and prior distributions in the model to implement the Markov Chain Monte Carlo

(MCMC) sampler

JAGS language for the top hierarchical model in Chapter 2 used to predict breeding status using ARU data song rates.

note: the order of breeding states is: FY, P, S

```
model{
  ### Likelihood
  for (i in 1:N){
    songcount[i] ~ dpois(lambda[i])
    log(lambda[i]) <- beta0[breedingstatus[i]] +
    beta1*abs(TimeRel2Sun[i])

    breedingstatus[i] ~ dcat(p[i,1:3])
    for (j in 1:3){
      p[i,j] <- delta[i,j]/sum(delta[i,])
      log(delta[i,j]) = alpha[i,j]
    }
    alpha[i,1] <- eta[1] + eta[4]*JDate[i]
    alpha[i,2] <- eta[2]
    alpha[i,3] <- eta[3] + eta[5]*JDate[i]
  }

  # Priors
  # etas are the log odds of each state (with paired as the reference
  here)
  eta[1] ~ dnorm(eta_FY, 1/(eta_FY_SE^2))
  eta[2] ~ dnorm(0, 1000)
  eta[3] ~ dnorm(eta_S, 1/(eta_S_SE^2))
  eta[4] ~ dnorm(eta_FY_date, 1/(eta_FY_date_SE^2))
  eta[5] ~ dnorm(eta_S_date, 1/(eta_S_date_SE^2))

  # exp(beta0) is the expected song count (for each state)
  beta0[1] ~ dnorm(beta0_FY, 1/(beta0_FY_SE^2))
  beta0[2] ~ dnorm(beta0_P, 1/(beta0_P_SE^2))
  beta0[3] ~ dnorm(beta0_S, 1/(beta0_S_SE^2))

  beta1 ~ dnorm(beta_1, 1/(beta_1_SE^2))
}

# FY, P, S
```

Appendix 3: Decision process for selecting a score threshold for use with an Olive-sided Flycatcher Song Scope recognizer to optimize recognizer output validation time while maintaining high true positive rate

I used a subset of recordings processed using the OSFL recognizer (McLeod, 2015) to determine the most efficient way to validate Olive-sided Flycatcher (OSFL) recognizer outputs. I combined outputs from 11 ARUs, representing recordings with high (ie. 2086/3299) and low (ie. 27/4649) proportions of true positive detections (Table 1).

Table 1. Counts of true and false positives used to optimize OSFL recognizer validation.

ARU name	Count of false positives	Count of true positives	Total count of detections
OSFL_JN_08_1	4649	27	4677
OSFL_JN_08_2	1297	214	1511
OSFL_JN_08_3	3299	2086	5385
OSFL_JN_09_1	3982	147	4129
OSFL_JN_09_2	4305	385	4690
OSFL_JN_09_3	6376	935	7311
OSFL_JN_09_6	2612	73	2685
OSFL_JN_10_1	3864	17	3881
OSFL_JN_11_2	6594	624	7218
OSFL_JN_11_3	2593	3622	6215
OSFL_JN_08_2	639	23	662
Total	40210	8153	48364

I investigated the descriptive statistics of acoustic measurements produced in the recognizer output along (level, score and quality; Wildlife Acoustics, 2011) with each detection to determine which measures are associated with true positive classification (Table 2) of OSFL.

Table 2. Descriptive statistics of level, quality and score from OSFL recognizer outputs (n = 48,364 detections), including average and standard deviation.

True (y) or false (n) positive	Average of Level	SD of Level	Average of Quality	SD of Quality	Average of Score	SD of Score
N	37.80	10.22	61.81	9.75	68.66	5.32
Y	55.25	10.51	67.43	11.68	75.46	4.86

I found that both high level and high score are associated with true positive classification.

I then removed all detections with a score of less than 70 and observed descriptive statistics of level and quality (Table 3).

Table 3. Mean, standard deviation and range of level and mean quality for detections with score ≥ 70 .

True (y) or false (n) positive	Average of Level	SD of Level	Min of Level	Max of Level	Average of Quality
n	50.34	12.94	21	74	66.92
y	55.99	9.89	22	88	67.51

I found that level is higher when score ≥ 70 , but by an average of 5, and large SD, therefore not helpful for further discrimination between true and false positives.

Lastly, I calculated time saved if only detections with score ≥ 70 were validated instead of the entire dataset. If technicians had only validated this subset, they would have had to process 23,397/48,367 detections (i.e. 48%), decreasing validation time to half that of the full dataset while retaining a large portion of true positives.

References

- McLeod, L. (2015). Olive-sided Flycatcher Song Scope Recognizer “OSFL_8_laprtraining(50_60).ssr.” Retrieved from <http://bioacoustic.abmi.ca/resources/reports/>
- Wildlife Acoustics, Inc. (2011). Song Scope: Bioacoustics software version 4.0 documentation.

Appendix 4: Hierarchical model bug (Bayesian inference Using Gibbs Sampling) files for use with 'rjags' code (R package) used in the four hierarchical models of Chapter 3, which identifies the model specification, data specification and prior distributions in the model to implement the Markov Chain Monte Carlo (MCMC) sampler

JAGS language for the four hierarchical models used to predict breeding status using ARU data song rates.

```
#####
# Model 1: the "no adjustment" model
# note: for all three models, the order of breeding states is: FY, P,
# S
model{
  ### Likelihood
  for (i in 1:N){
    songcount[i] ~ dpois(lambda[i])
    log(lambda[i]) <- gamma[birdID[i]] + beta0[breedingstatus[i]]

    breedingstatus[i] ~ dcat(p[i,1:3])
    for (j in 1:3){
      p[i,j] <- delta[i,j]/sum(delta[i,])
      log(delta[i,j]) = alpha[i,j]
    }
    alpha[i,1] <- eta[1] + eta[4]*JDate[i]
    alpha[i,2] <- eta[2]
    alpha[i,3] <- eta[3] + eta[5]*JDate[i]
  }

  # Priors
  # etas are the log odds of each state (with paired as the reference
  here)
  eta[1] ~ dnorm(eta_FY,1/(eta_FY_SE^2))
  eta[2] ~ dnorm(0,1000)
  eta[3] ~ dnorm(eta_S,1/(eta_S_SE^2))
  eta[4] ~ dnorm(eta_FY_date,1/(eta_FY_date_SE^2))
  eta[5] ~ dnorm(eta_S_date,1/(eta_S_date_SE^2))

  # exp(beta0) is the expected song count (for each state)
  beta0[1] ~ dnorm(beta0_FY, 1/(beta0_FY_SE^2))
  beta0[2] ~ dnorm(beta0_P, 1/(beta0_P_SE^2))
  beta0[3] ~ dnorm(beta0_S, 1/(beta0_S_SE^2))

  # Prior distribution on random effect of individual
  for(k in 1:nbird){
    gamma[k] ~ dnorm(0, 1/alpha.var)
  }
}
#####
```

```

# Model 2: the "zero-truncated" model

model{
  ### Likelihood
  for (i in 1:N){
    songcount[i] ~ dpois(lambda[i])T(1,)
    log(lambda[i]) <- gamma[birdID[i]] + beta0[breedingstatus[i]]

    breedingstatus[i] ~ dcat(p[i,1:3])
    for (j in 1:3){
      p[i,j] <- delta[i,j]/sum(delta[i,])
      log(delta[i,j]) = alpha[i,j]
    }
    alpha[i,1] <- eta[1] + eta[4]*JDate[i]
    alpha[i,2] <- eta[2]
    alpha[i,3] <- eta[3] + eta[5]*JDate[i]
  }

  # Priors
  # etas are the log odds of each state (with paired as the reference
  here)
  eta[1] ~ dnorm(eta_FY, 1/(eta_FY_SE^2))
  eta[2] ~ dnorm(0, 1000)
  eta[3] ~ dnorm(eta_S, 1/(eta_S_SE^2))
  eta[4] ~ dnorm(eta_FY_date, 1/(eta_FY_date_SE^2))
  eta[5] ~ dnorm(eta_S_date, 1/(eta_S_date_SE^2))

  # exp(beta0) is the expected song count (for each state)
  beta0[1] ~ dnorm(beta0_FY, 1/(beta0_FY_SE^2))
  beta0[2] ~ dnorm(beta0_P, 1/(beta0_P_SE^2))
  beta0[3] ~ dnorm(beta0_S, 1/(beta0_S_SE^2))

  # Prior distribution on random effect of individual
  for(k in 1:nbird){
    gamma[k] ~ dnorm(0, 0.0001)
  }
}

#####
# Model 3: the "zero-inflated" model

model{
  ### Likelihood
  for (i in 1:N){
    z[i] ~ dbern(psi)
    songcount[i] ~ dpois(lambda[i]*z[i])
    log(lambda[i]) <- gamma[birdID[i]] + beta0[breedingstatus[i]]

    breedingstatus[i] ~ dcat(p[i,1:3])
    for (j in 1:3){
      p[i,j] <- delta[i,j]/sum(delta[i,])
      log(delta[i,j]) = alpha[i,j]
    }
    alpha[i,1] <- eta[1] + eta[4]*JDate[i]
    alpha[i,2] <- eta[2]
    alpha[i,3] <- eta[3] + eta[5]*JDate[i]
  }
}

```

```

# Priors
# etas are the log odds of each state (with paired as the reference
here)
eta[1] ~ dnorm(eta_FY,1/(eta_FY_SE^2))
eta[2] ~ dnorm(0,1000)
eta[3] ~ dnorm(eta_S,1/(eta_S_SE^2))
eta[4] ~ dnorm(eta_FY_date,1/(eta_FY_date_SE^2))
eta[5] ~ dnorm(eta_S_date,1/(eta_S_date_SE^2))

# exp(beta0) is the expected song count (for each state)
beta0[1] ~ dnorm(beta0_FY, 1/(beta0_FY_SE^2))
beta0[2] ~ dnorm(beta0_P, 1/(beta0_P_SE^2))
beta0[3] ~ dnorm(beta0_S, 1/(beta0_S_SE^2))

# psi is the probability that the bird sings (or is detected by the
ARU and the automation algorithm)
psi ~ dunif(0, 1)

# Prior distribution on random effect of individual
for(k in 1:nbird){
  gamma[k] ~ dnorm(0, 1/alpha.var)
}
}

#####
# Model 4: the "distance" model

model{
  ### Likelihood
  for (i in 1:N){
    z[i] ~ dbern(psi)
    songcount[i] ~ dpois(lambda[i]*z[i])
    log(lambda[i]) <- gamma[birdID[i]] + beta0[breedingstatus[i]] +
beta1.phi*level[i]

    breedingstatus[i] ~ dcat(p[i,1:3])
    for (j in 1:3){
      p[i,j] <- delta[i,j]/sum(delta[i,])
      log(delta[i,j]) = alpha[i,j]
    }
    alpha[i,1] <- eta[1] + eta[4]*JDate[i]
    alpha[i,2] <- eta[2]
    alpha[i,3] <- eta[3] + eta[5]*JDate[i]
  }
}

# Priors
# etas are the log odds of each state (with paired as the reference
here)
eta[1] ~ dnorm(eta_FY,1/(eta_FY_SE^2))
eta[2] ~ dnorm(0,1000)
eta[3] ~ dnorm(eta_S,1/(eta_S_SE^2))
eta[4] ~ dnorm(eta_FY_date,1/(eta_FY_date_SE^2))
eta[5] ~ dnorm(eta_S_date,1/(eta_S_date_SE^2))

# exp(beta0) is the expected song count (for each state)
beta0[1] ~ dnorm(beta0_FY, 1/(beta0_FY_SE^2))
beta0[2] ~ dnorm(beta0_P, 1/(beta0_P_SE^2))

```

```
beta0[3] ~ dnorm(beta0_S, 1/(beta0_S_SE^2))

# psi is the probability that the bird sings (or is detected by the
ARU and the automation algorithm)
psi ~ dunif(0, 1)

# Prior distributions on detection
beta1.phi ~ dt(0, 0.4, 1)

# Prior distribution on random effect of individual
for(k in 1:nbird){
  gamma[k] ~ dnorm(0, 1/alpha.var)
}
}
```

Appendix 5: A pilot study of quantifying Olive-sided Flycatcher songs in ARU recordings manually versus with automatic recognition to determine the most efficient processing method for breeding status – song rate studies

Objectives:

- 1) Determine which song metrics can be extracted manually using a spectrogram annotation software (Syrinx, henceforth “manual” method) vs. automatic recognition software (Song Scope, henceforth “automatic recognition” method).
- 2) Determine the optimum processing method (manual vs. automatic recognition) to extract Olive-sided Flycatcher singing behaviour to model the relationship between singing behaviour and breeding status.

Methods:

Acoustic processing

Song Scope - The recognizer (McLeod, 2015) was run at score-quality threshold 50-60, then validated to score ≥ 70 . True positives were totalled for each 5-min period at the start of every hour.

Syrinx - Spectrograms for each 5-minute sample period were visually scanned for any OSFL vocalization, including songs (“quick, three-beers”), calls (“pip-pip-pip”) and other miscellaneous OSFL sounds (ie. chattering and purring sounds and beak snapping). Visual identification was also confirmed by listening to clips, when required. If multiple individuals were detected (based off overlapping songs/calls or obvious distance differences), they were recorded separately.

Recordings selection

I selected two OSFL territories that were monitored in 2016, had already been processed using Song Scope, and had relatively high song counts (Table 1). The ARUs selected had been well-placed on the OSFL territories (Fig.1); they were at the centre of the where the OSFLs were detected during multiple 1-hour focal studies throughout the breeding season and were placed near the nest. I also wanted to test the effect of multiple individuals on recognizer detections and AB-08 was known to have an immediately adjacent territory. For each male, I listened to recordings collected on 4 days throughout the breeding season, approximately 1 week apart, which overlapped the “paired”, “incubating” stages, as well as the “feeding young” stage for JN-10 (Table 2). For each day, I processed nine 5-minute periods at the start of every hour between 3:00 am and 11:05 am.

Table 1. Total counts for all ARUs from 2016 which were run and validated using Song Scope. Rows in white represent ARUs where an OSFL was detected at least once, whereas grey rows are ARUs with no true positives produced from Song Scope processing. The two ARUs selected for this study are highlighted in yellow.

ARU	Sum 5-min song count	Mean 5-min song count	StdDev 5-min song count	Max 5-min song count
OSFL-AB-02-1	934	7.41	16.57	108
OSFL-NB-10-2	773	4.52	15.74	131
OSFL-AB-08-1	573	1.82	6.42	50
OSFL-JN-11-3	327	2.02	6.19	43
OSFL-NB-08-1	207	1.10	2.65	22
OSFL-SB-14-1	193	1.07	3.25	22
OSFL-AB-06-2	168	0.50	3.40	42
OSFL-AB-09-1	147	1.09	2.43	12
OSFL-SB-01-2	147	2.04	2.86	12

OSFL-NB-01-2	122	0.71	2.38	14
OSFL-AB-07-1	93	0.49	1.10	6
OSFL-JN-10-2	79	0.63	1.94	18
OSFL-SB-04-1	66	0.29	1.52	16
OSFL-JN-09-3	62	0.30	1.13	9
OSFL-SB-03-1	47	1.31	3.88	16
OSFL-JN-08-3	38	2.11	5.80	18
OSFL-AB-01-1	10	0.07	0.45	4
OSFL-NB-20-2	9	0.14	0.56	3
OSFL-NB-09-1	8	0.07	0.45	3
OSFL-AB-04-1	4	0.02	0.14	1
OSFL-AB-10-1	1	0.01	0.12	1
OSFL-AB-01-2	0	0.00	0.00	0
OSFL-AB-02-2	0	0.00	0.00	0
OSFL-AB-04-2	0	0.00	0.00	0
OSFL-AB-06-3	0	0.00	0.00	0
OSFL-AB-08-2	0	0.00	0.00	0
OSFL-AB-09-2	0	0.00	0.00	0
OSFL-AB-11-1	0	0.00	0.00	0
OSFL-AB-11-2	0	0.00	0.00	0
OSFL-JN-08-1	0	0.00	0.00	0
OSFL-JN-08-2	0	0.00	0.00	0
OSFL-JN-09-1	0	0.00	0.00	0
OSFL-JN-09-2	0	0.00	0.00	0
OSFL-JN-09-4	0	0.00	0.00	0
OSFL-JN-09-5	0	0.00	0.00	0
OSFL-JN-09-6	0	0.00	0.00	0
OSFL-JN-11-2	0	0.00	0.00	0
OSFL-JN-11-4	0	0.00	0.00	0
OSFL-JN-14-1	0	0.00	0.00	0
OSFL-JN-14-2	0	0.00	0.00	0
OSFL-JN-14-3	0	0.00	0.00	0
OSFL-JN-14-4	0	0.00	0.00	0
OSFL-JN-18-1	0	0.00	0.00	0
OSFL-JN-18-2	0	0.00	0.00	0
OSFL-NB-08-2	0	0.00	0.00	0
OSFL-NB-09-2	0	0.00	0.00	0
OSFL-NB-10-1	0	0.00	0.00	0
OSFL-NB-20-1	0	0.00	0.00	0
OSFL-NB-20-3	0	0.00	0.00	0
OSFL-NB-20-4	0	0.00	0.00	0
OSFL-SB-01-3	0	0.00	0.00	0
OSFL-SB-03-2	0	0.00	0.00	0

OSFL-SB-04-2	0	0.00	0.00	0
OSFL-SB-11-2	0	0.00	0.00	0
OSFL-SB-11-3	0	0.00	0.00	0
OSFL-SB-14-2	0	0.00	0.00	0

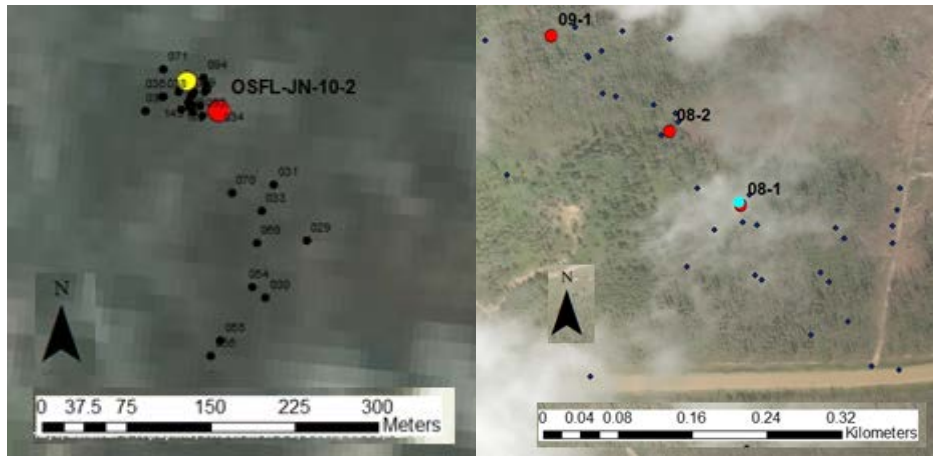


Figure 1. Placement of the ARUs (red dots) in relation to nest locations (yellow dot on the left, blue dot on the right) and other used locations (GPS points collected every 5 minutes during hour-long follows). JN-10 is an OSFL territory in the NWT, with no known neighbouring OSFL males nearby. AB-08 (right) is a territory near McClelland fen, AB and was immediately adjacent to OSFL territory AB-09. For this analysis, ARU 08-1 was used (closest to nest).

Table 2. Sample dates for each OSFL territory and associated breeding status of the male OSFL occupying that territory. For each date, nine 5-minute periods were sampled (every hour, between 3 am – 11 am).

OSFL territory	Date	Breeding Status	Count of 5-min recordings
AB-08	June 10, 2016	Paired	9
AB-08	June 17, 2016	Paired	9
AB-08	June 23, 2016	Paired	9
AB-08	June 30, 2016	Incubating	9
JN-10	June 11, 2016	Paired	9

JN-10	June 19, 2016	Incubating	9
JN-10	June 26, 2016	Incubating	9
JN-10	July 2, 2016	Feeding Young	9
TOTAL			72

Objective 1: Determine which song metrics can be extracted manually using manual vs. automatic recognition methods.

1a. What is the false negative rate (i.e. missed OSFL songs) using the Song Scope recognizer?

Table 3. Number of songs detected using automatic recognition vs. manual methods for the same recordings and the percentage of songs missed by automatic recognition compared with manual processing.

	Song Scope (automatic recognition)	Syrinx (manual/visual)	% of songs missed
JN-10 (only 1 male)	44	256	83%
AB-08 (up to 3 males)	98	1170	92%

1b. How much more song/call information can I extract?

The number of OSFL songs detected by human listening, using Syrinx, was much higher than the number of songs detected using Song Scope automatic recognition (Fig.2). Song Scope is unable to distinguish between multiple individuals, but when they were identified using Syrinx, I see that more individuals result in more singing (Fig. 3).

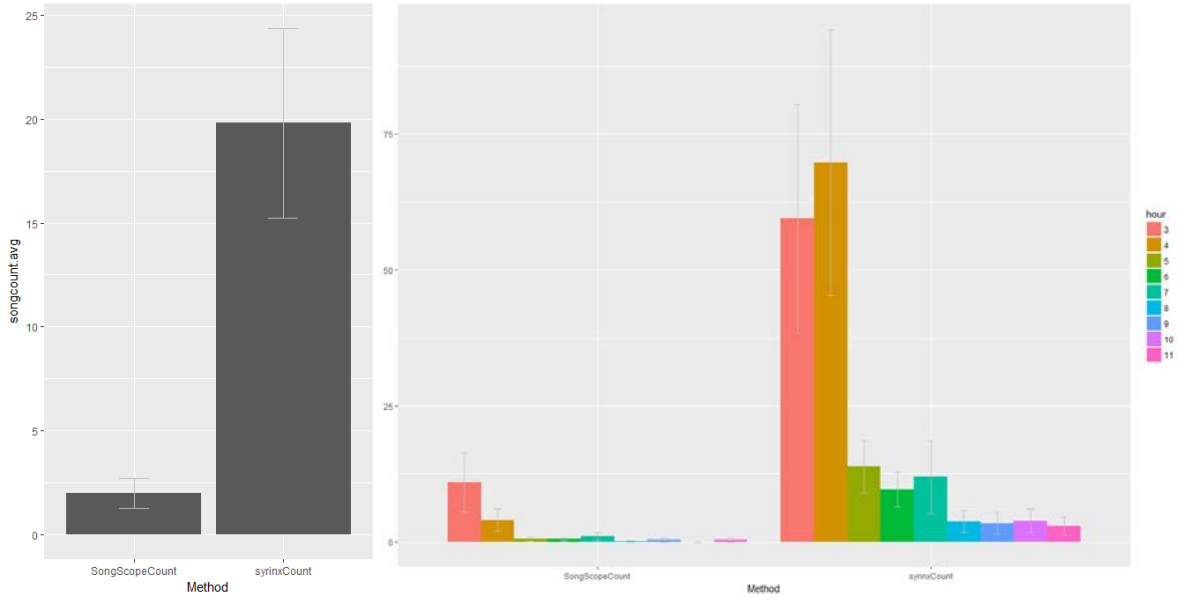


Figure 2. Mean 5-minute song count \pm SE when the same 72 recordings were processed using Song Scope (automatic recognition) vs Syrnix (manual detection). The left plot is all counts combined, and the right plot is mean song counts as a function of sample hour.

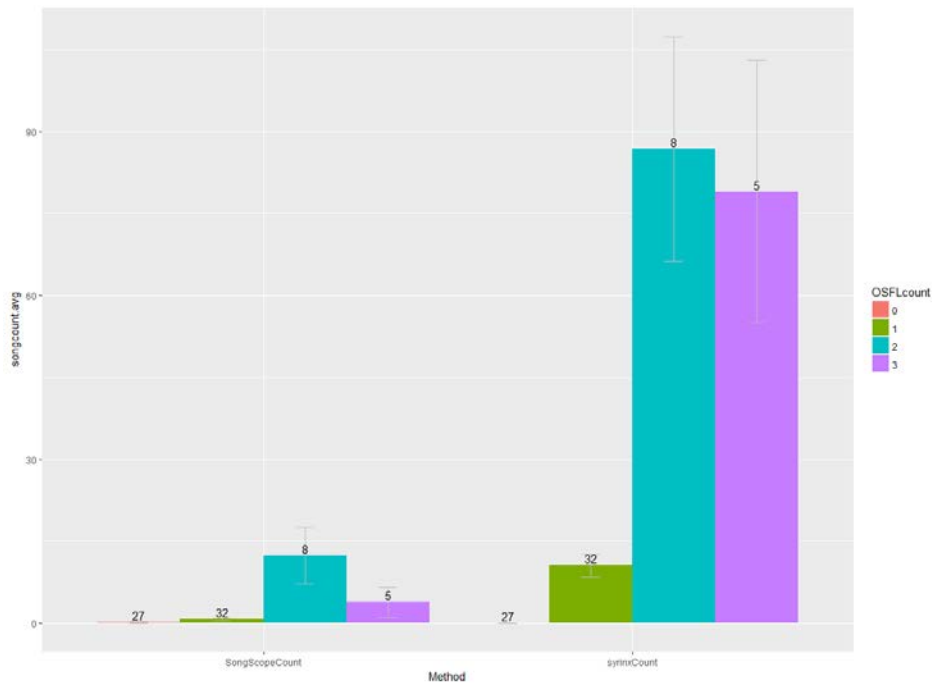


Figure 3. Mean 5-minute song count \pm SE when the same 72 recordings were processed using Song Scope (automatic recognition) vs Syrnix (manual detection), as a function of the number of

individual OSFLs detected singing on the recording. The number above each bar is the sample size.

Objective 2: *Determine the optimum processing method (manual vs. automatic recognition) to extract Olive-sided Flycatcher singing behaviour to model the relationship between singing behaviour and breeding status.*

2a. *How long does manual processing of recordings take?*

Average processing time using my current song and pip detection methods in Syrinx is ~ 1.5 processing minutes/1 minute of audio. I have calculated the number of days (comprised of 6 hours of processing/day) that it would take to complete a few different scenarios with different sampling options (varying daily sample rate and number of samples per day; table 2). The maximum standard recording size to sample is limited to 5 minutes because that was the maximum length of time the ARU was programmed to record. If sampling is only once per day, I would select 3 am, to maximize the likelihood of capturing a singing bout (based off my results so far and other studies on OSFL singing behaviour).

Discussion:

- Detecting Olive-sided Flycatcher (OSFL) songs using an automated recognition software (Song Scope, by Wildlife Acoustics, recognizer built by McLeod [2015]) results in missing >80% of songs that could be detected in the Autonomous Recording Unit (ARU) recordings, based on manual processing of audio recordings (Syrinx)
- Multiple OSFLs on a recording cannot be identified using Song Scope, but can with Syrinx, and singing conspecific neighbours likely impact singing behaviour

- The processing speed for identifying vocalizations using Syrinx is ~1.5 processing minutes/1 audio minute (example sample regime: to sample 19 individuals, for 67 days, for 5-minutes a day, would require 13 6-hour work days to process)