NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

**Evaluating observer effects in point count studies:
Using stereo-microphones to investigate the role of species identification
errors on Breeding Bird Survey (BBS) point counts and microphone arrays
to estimate changes in bird behaviour in response to an observer's
presence.**

Margaret Campbell

Thesis submitted to the
School of Graduate Studies and Research, Carleton University
In partial fulfillment of the requirements for the
M.Sc. degree in the
Ottawa-Carleton Institute of Biology
© 2009

Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-60199-0
Our file Notre référence
ISBN: 978-0-494-60199-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

■+■
Canada

ABSTRACT

I evaluated the role of observer effects in point count survey and analysis methods used in North American avian monitoring programs. Skilled observers listening to recordings of Breeding Bird Survey (BBS) routes differed significantly in the number of species detected. On average 11% of the species reported were errors and 23% of the species were missed on the recordings. Field observers reported similar number of species as listeners but fewer than were on the recordings. Site occupancy models failed to correctly account for identification errors and tended to overestimate species richness by 1.0 - 8.7 species and positively bias between 69% - 100% of species specific occupancy estimates. Tests using a microphone array and acoustic localization system with mean localization error of $4.19 \pm 0.52\text{m}$ showed no difference in the mean number of individuals, singing position or onset of singing behaviour of birds in response to an observer conducting a point count.

Acknowledgements

Many people have contributed data or their bird identification skills to this project. Thank you to all the BBS observers who recorded their routes: Bev McBride, Connie Downes, Bernie Ladoucer and Steve Wendt. Many thanks to the 6 volunteers who listened to the BBS recordings and gave me their species lists to evaluate: ED, GH, BL, BM, MM, PM, without them there'd be no second chapter.

Thanks to the Bioacoustics Research Program at Cornell for the use of the ARUs and not freaking out when I brought them back with bear puncture marks. Special thanks to Chris Tessaglia-Hymes for teaching me how to run the ARUs and resolving a multitude of last minute technical glitches and equipment failures; Tom Calupca for the use of BEAST and help with programming the ARUs, extracting and correcting the misaligned data; Ann Warde for her help with data analysis and patient explanations of how to run XBAT and its associated localization functions; and Kathryn Cortopassi for giving me permission to use her localization algorithm.

Thanks to David Okines and Prince Edward Point Bird Observatory for allowing me to live in the VanCott Cottage during my field season, it's an amazing location and wonderful place to do field work. Also thanks to Ryan Zimmerling and Fergus Nichol formerly of Bird Studies Canada for allowing me to tag along with their boreal research crews during my first ill-fated field season.

Thanks to everyone who fed me and kept me moderately sane during this entire project: Mom, Haroldia, Shannon, Anita, Lisa, Akiko, Myra, Isabel, Rob, Amie, Paul, Sarah, Jude, Tabby, Sabrina, Kim-Kim, Jess, Nil, Dave, the angry old man and the rest of Carleton's Fencing Team.

Special thanks to my committee Tom Sherratt and Jeremy Kerr. Finally big thanks to Charles Francis, my supervisor for many, many helpful suggestions and introducing me into the wacky world of bird counting using microphones.

This project was funded by the Canadian Wildlife service of Environment Canada.

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	v
List of Figures	vi
Chapter 1:	
General Introduction	1
Chapter 2:	
Using stereo-microphones to evaluate observer variation in Breeding Bird Survey (BBS) point counts	
Introduction	10
Methods	14
Results	17
Discussion	21
Chapter 3:	
Using a GPS synchronized microphone array and acoustic localization system to investigate observer induced behavioural changes in singing birds during point count surveys	
Introduction	43
Methods	47
Results	53
Discussion	55
Chapter 4:	
General Discussion	77
Literature Cited	87

List of Tables

Chapter 2:

Table 1. Subset of 10 most commonly missed species by observers listening to BBS point count recordings	34
Table 2. Subset of 10 most common detection errors by observers listening to BBS point count recordings	35
Table 3. Model selection table for multiple independent observer models of species richness.	36
Table 4. Model estimates for species specific detection probabilities of 16 target species.	37
Appendix 1. Table of all species detected on 4 BBS routes	38

Chapter 3:

Table 1. Results of localization algorithm tests	74
--	----

List of Figures

Chapter 2:

- Figure 1. Mean number of confirmed species detections and detection errors. 31
- Figure 2. Comparison of mean number of species detected and detection errors on each of 4 BBS routes. 32
- Figure 3. Comparison of mean species richness estimates generated using double and multiple observer models for each of 4 BBS routes. 33

Chapter 3:

- Figure 1. Diagram of layout of adjacent microphone arrays. 66
- Figure 2. Example of localization algorithm. 67
- Figure 3. Algorithm localization error for known whistle trial locations. 68
- Figure 4. Mean number of individuals detected for observer present and absent trials. 69
- Figure 5a, b. QQ-plots of mean signing positions of all individuals for observer absent vs. observer present trials. 70
- Figure 6a,b. Proportion novel first detections of individuals per 30s intervals for observer absent vs. observer present trials 71
- Figure 7a,b. Species specific plots of mean number of individuals detected for observer absent vs. observer present trials. 72
- Figure 8a,b. Species specific plots of mean change in distance from the center of an array for observer present vs. observer absent trials. 73

GENERAL INTRODUCTION

Long term avian population monitoring programs are key elements of many North American management and species at risk conservation plans. Monitoring programs can provide early indications of species or habitats that may require conservation action to maintain biodiversity or ecosystem function. In response to fears of population declines due to exposure to DDT and other pesticides first identified in the 1960s, songbirds have been the primary indicator species for the majority of monitoring programs. Songbirds are useful indicator species since they are a highly diverse group of species, inhabit all major eco-regions in North America and most are migrants, spending part of each year in northern breeding ranges and much of the remainder of the year in more southerly wintering habitats. Changes in migratory songbird populations may provide clues to local changes on breeding or wintering sites or larger continental scale changes in biodiversity or ecosystem stability. The main advantage of using songbirds as indicators is although they are a highly diverse group of species they are relatively easy to identify for observers with some training due to species specific variation in plumage and song.

Point counts are one of the most common methods of data collection in bird population monitoring programs. During a point count an observer records all birds seen or heard from a fixed point over a specific time interval. The basic experimental design of standardized point count protocols are based on 28 recommendations described by Ralph et al. (1995). Point count studies are designed so that the area sampled by each point is maximized but does not

overlap with adjacent points and the time that an observer spends at each point is maximized compared to the travel time between points (Ralph et al 1995). Point counts allow researchers to use a standardized approach to sample a large area and generate indices of population abundance more efficiently than other more labour intensive methods such as territory mapping (Toms et al 2006, Kery and Schmid 2006, Howell et al 2004).

Because of the high sampling area to effort ratio point counts have become the dominant survey method for most avian population monitoring programs. The North American Breeding Bird Survey (BBS) is based on 40+ years of point count data and is recognized as one of the best long term population monitoring data sets in the world. The BBS was established in 1966 and expanded across the USA and into southern Canada in 1967 (Sauer et al 2008). Currently over 4100 road-side survey routes are run annually in North America; the program is internationally co-ordinated by the United States Geological Survey (USGS) Patuxent Wildlife Research Center and the National Wildlife Research Center of the Canadian Wildlife Service, Environment Canada. Each BBS survey route consists of a 39.4 km (24.5 mile) road-side survey along secondary roads with 50 stops spaced every 0.80 km (0.5 miles) (Sauer et al 2008). Observers conduct a standardized 3-minute, 400 m fixed radius (0.25 mile) point count at each of the stops over five hours beginning 30 minutes before local sunrise (Sauer et al 2008).

Count data are most often analyzed as an index of population size (Johnson 2008), where the count (C) is related to the true population size (N) by

the proportion of the true population that is recorded (p) such that $N=C/p$. (Johnson 2008, Nichols et al 2000). One of the challenges in evaluating point count survey data is that many different factors may affect the proportion of the population detected (p). The factors could include natural variation in the songbird communities and observation effects. Habitat effects (Betts et. al 2007, Harris and Haskell 2007, Sallabanks et al 2006), time of year (Maron et al 2005, Selmi and Boulinier 2003, Wilson and Bart 1985), number of individuals (Howell et al 2004, Bart 1985), individual variation in singing rate (McShea and Rappole 1997), and song structure (Nelson 1989, Lemon et al 1983) all affect the detectability of birds in the field. Most studies rely on collecting data under standardized field conditions to minimize the amount of variation in the number of species available for detection during a count. Point counts should be run early in the morning during the breeding season (between May to July depending on latitude) under conditions with minimal wind and no rain; these restrictions result in a relatively short window of opportunity (i.e. 2 – 4 months depending on area surveyed) for data collection (Ralph et al 1995).

Using microphones to record point counts may increase the number and types of songbird communities that can be sampled annually. We know that most of the species detected in the field can also be detected using recordings of a point count (Rempel et al 2005, Hobson et al 2002). One approach to increasing the number of point count surveys conducted annually would be to send relatively unskilled observers into the field with microphones to record point counts, then have the recordings analyzed at a later date by highly skilled

observers. By recording counts the stringent requirements that field observers must be able to identify all species by sight and sound that occur in the area surveyed (Ralph et al 1995) could be relaxed. This method has been used to verify species detected in the field in Peruvian rainforest where species diversity is high and confirmation of species by visual detection can be difficult (Haselmayer and Quinn 2000). Recordings of point counts can also be used as archival material of singing bird populations and as tools for observer training (see McLaren and Cadman 1999).

Recordings may also help to better understand how we design point count surveys and generate better analytical methods. One approach is to use recordings to generate simulated songbird communities. Playback studies have been used to evaluate differences in the number of individuals detected amongst observers during point counts (Bart 1985, Bart and Schoultz 1984). Other simulated communities have been used to test point count methods and assumptions (Alldredge et al 2007d, Simons et al 2006). The major advantage of using simulated songbird populations is that the experimenters control the number of individuals and species present in the simulated community. However, simulations lack the complexity of true song bird populations and full chorus field recordings can also be used to evaluate point count design and analysis methods. A major advantage in analyzing field recordings is that the complexity of individual song type, structure and timing along with multi-species and multi-individual overlap of birds that naturally occur in the field is preserved. There are however, several major disadvantages in using actual recordings; first even after

intensive scanning of recordings there may be some errors in species detections and identifications; second the associated visual and habitat cues typical of field point counts are absent from recordings; third a subset of non-vocal species could be missed on the recordings.

One complication of point count methods is that they rely entirely on observers correctly detecting and identifying individuals and species by sight or sound; there is no way to verify that a species was detected (e.g. scat counts or burrow counts). As a result understanding the role of observer effects is critical to the evaluation of point count data. To date most observer effect studies have focused on how differences in observer skill level (i.e. novice vs. experienced) affect the accuracy of count data both for number of individuals and species detected (McLaren and Cadman 1999, Kendal et al 1996).

There has been relatively little work on evaluating differences amongst skilled observers in the species reported during a count and the frequency of missed and errors in detections made during a count. Nor has there been much research on how detection errors may affect the analysis of point count data. We know that observers will make errors during a count either by failing to detect individuals or species, or by incorrectly identifying species or individuals that are present. For example, there are 286 species of birds that are known to breed in Ontario (Cadman et al 2007); while not all of these species would be encountered during a single count there is a high probability that observers will make detection errors based only on the number of species possible. There are

four detection outcomes possible for each species potentially present in the field (Royle and Link 2006):

1. Present: An individual or species is present and is detected
2. Detection Zero: An individual or species is present but not detected
3. False Positive: An individual or species which is not present is detected (i.e. misidentification error)
4. True Zero: An individual is absent.

Most point count studies focus on species or individuals that are present and correctly detected. The role of missed detections has also received some attention in the ornithological literature. Although it is helpful to know the proportion of missed detections, it may be a philosophical argument: if a bird sings and nobody hears it — does it actually matter in our count statistics? False positive detections pose several interesting challenges in the analysis of point count data and have the potential to play an important role in management and conservation decisions. False positive and misidentification of species can introduce observer generated bias into counts. For example, if an observer consistently misidentifies a species this would inflate counts of the misidentified species while simultaneously negatively biasing counts of the correct species that they failed to detect. Identification bias may be further magnified depending on how common a particular species is in the study area, especially if double (Nichols et al 2000) or multiple observer approaches (Alldredge et al. 2006a) are used to generate diversity estimates of songbird communities or to set management or conservation priorities.

Several modifications to standardized point count protocol have been suggested as methods to improve the data generated during point counts. Distance sampling has been proposed as a modification of standardized point counts by calculating the density of species detected during a count using the distance between the observer and the individual bird detected (Buckland et al 2001, 2004). A key assumption of both standardized point counts and distance methods is that birds do not move or change their singing behaviour in response to an observer (Bart et al. 1995, Buckland et al 2001, 2004). But this assumption has not been adequately evaluated. This is probably due to the difficulty in obtaining an objective sample of bird behaviour when observers are absent. Recent technological advances in microphone arrays and acoustic localization algorithms can be used to estimate the singing positions of birds during when an observer is performing a point count and when observers are truly absent.

An array of synchronously recording microphones can be used to triangulate the approximate location of an individual based on the time delay in the arrival of a single sound to at least 3 different microphones. In order to set up a microphone array suitable for acoustic localization there are several requirements (Mohl et al 2001):

1. The position of the microphones must be known and accurate
2. All recordings must be synchronous
3. The size of the array must be of the same magnitude as the range to be sampled
4. The speed of sound must be known

5. There must be a time marker in the signals at all microphones for alignment.

Simple acoustic localization, without recordings, has been used in the field to estimate the position of booming male Great Bitterns (*Botaurus stellaris*) in reedy marsh habitats (Lefebvre and Poulin, 2003).

Microphone arrays and triangulation were first developed for surveys of aquatic populations (e.g. Mohl et al 2001) but have recently been adapted to terrestrial systems (McGregor 1997, Spiesberger 1999, Kwan et al 2006). Population and migration monitoring of many species of whales rely heavily on acoustic triangulation: e.g. Bowhead whales (*Balaena mysticetus*) (Clark and Ellison 2000) and Humpback whales (*Megaptera novaeangliae*) (Clark and Clapham 2004). With recent reductions in cost and increase in memory available in commercial hard drives it has become easier to sample populations of high frequency complex bird songs in acoustic localization experiments. Linked microphone array experiments are becoming more common in the ornithological literature for evaluating playback systems (Mennil et al 2006) or for experiments in bird behaviour in tropical (Mennill and Verhencamp 2008) and temperate (Foote et al. 2008) ecosystems.

For my M.Sc. project I have tested two assumptions of point count methods with respect to observers. First I have used full chorus stereo-microphone recordings of BBS routes to evaluate variation amongst observers in species identification and the frequency of detection errors in bird surveys and the role these errors play in the analysis of point count data. Second using an

array of GPS synchronized microphones and localization algorithms I tested for changes in song bird behaviour in response to an observer being present during a point count. Finally I review how the results of this research can be used to improve current bird monitoring programs and suggest areas of further research that could improve how the technology I used in this study could be applied in the future.

Using stereo-microphones to evaluate observer variation in BBS point counts

Most major avian survey and monitoring programs rely on skilled birders to identify birds by sight or sound. These include surveys such as the North American Breeding Bird Survey (BBS), provincial and state breeding bird atlases, many regional and local monitoring programs, and various citizen science projects such as Project Nest Watch or Christmas Bird Counts. For quantitative analysis of relative abundance or population trends many of these surveys use point count methods, during which an observer records all birds seen or heard from a fixed point during a specific time interval (Ralph et al. 1995). Although there are many variations in the design of count protocols, for example in survey duration or radius of detection, all approaches are dependent on the ability of the observer to detect and correctly identify birds present at the site.

One of the challenges in the analysis of point count data is dealing with variation among observers in the number of species detected and correctly identified. We expect observers to make a few errors due to omission (failing to detect or identify a species when present) or misidentification (incorrectly identifying a species that is not present) during any count. Bart (1985), using a series of indoor loudspeaker playback experiments of a simulated songbird community, found that even highly skilled observers missed or under-counted approximately 30% of the birds present on recordings, overestimated the number of individuals by 6% due to over-counting, and misidentified an additional 3% of individuals as species that were not present. Using the same playback system

combined with field data, Bart and Schoultz (1984) showed that average observer efficiency in detection decreased as the number of birds increased; as the number of birds singing during a count increases observers tend to make more errors.

Several factors can lead to variation among observers in species identification, including variation in hearing ability, training, past experience with point counts, familiarity with each species and skill at discriminating sounds, Variation in hearing ability can change over time as observers tend to lose their ability to hear high frequency vocalizations as they age (Cyr 1981). Distance effects have also been suggested as a source of observer variation such that “better” observers are able to detect species over a greater range of distances (Alldredge et al. 2007a). Using annotated recordings with known number of species Rempel et al. (2005) found that observers tended to agree on the identification and number of individuals detected of the most common species but identification of rare or uncommon species was more variable.

Three types of approaches have been proposed for dealing with variation among observers. The most common approach is to ignore observer variation in the analysis. This approach assumes that through a combination of standardizing field conditions and providing appropriate training, variation among observers is small relative to spatial or temporal variation in bird abundance (Ralph et. al 1995). Provided that there are large numbers of observers so that the variation can be considered random relative to the scale (either temporal or spatial) of the study then this should lead to little bias (Johnson 2008). However, if a small

number of observers conduct a large number of counts the data may be over-dispersed; this could lead to serious bias in conclusions if observers detecting higher or lower than average numbers are concentrated in particular geographical areas or time periods (Ralph et al 1995).

A second approach is to control for inter-observer variation by analyzing trends within observer and then treating each observer as a covariate in the analysis. This type of analysis is used for BBS point count data (Sauer and Link 2002, Bart et al 2004, Downes and Collins 2008) using either a route regression model (United States Geological Survey – Sauer et al 2008) or model where observer is a blocking term (Canadian Wildlife Service, Downes and Collins 2008). However, there are certain limitations of treating observer as a covariate in trend analysis. First, trend analysis does not adequately account for change in an observer's skill level either due to improvement in bird identification skills or loss of hearing. Kendal et al. (1996) showed that new observers running a BBS route tended to positively bias (i.e. over-estimate) species counts but the bias was removed after the first year of a new observer running the route when they became more adept at performing point counts. Second, there is a loss of statistical power in the analysis if observers change frequently during the survey. In extreme cases some BBS routes are run by different observers each year making it impossible to model observer variation and resulting in these routes being dropped from the analysis (Downes and Collins 2008).

A third approach is to estimate and adjust for variation among observers by estimating detectability (see Sauer et al. 2005, Sauer et al 2004, Thompson

2002, Rosenstock et al 2002, Pollock et al. 2002, Nichols et al. 2000) using approaches such as double observer methods, multiple observer methods or occupancy models. Double observer methods use two observers, either independent (neither knowing what the other reported) or semi-dependent (one observer knowing what the other reported) to estimate the probability that each individual (or species) was detected by the observer (Nichols et al. 2000). Double observer methods have been used in several studies (Forcey et al. 2006, Kissling and Garton 2006, Fletcher and Hutto, 2006). However, Moore et al. (2004) noted that density estimates of less conspicuous species using double observer methods may be significantly smaller than the true density. Multiple observer models are an extension of double observer methods where the detectability estimate is based on the probability that an individual or species is detected by $n > 2$ observers (Alldredge et al 2006 a). Since multiple observer methods have several detection (or capture) occasions rates of detection error and observer heterogeneity can be estimated. In a series of playback experiments Alldredge et al (2006a) showed that, assuming there were no matching errors in detection between observers, multiple observer models that took into account observer variation produced more accurate abundance estimates for their 5 study species. Double and multiple observer models can be thought of as specific types of site occupancy models, where occupancy is a function of detectability across capture occasions within a count rather than as detection occasions on multiple visits to a site (see Kery 2008, Royle and Link

2006, Royle et al 2005, MacKenzie et al 2006, MacKenzie et al. 2004, MacKenzie et al. 2002).

I designed an experiment to quantify observer variation using song data recorded on actual BBS routes. Most studies to date have been based on either playback studies (Bart 1985, Alldredge et al. 2007a) or using data from a limited suite of species (Alldredge et al. 2006a, Diefenbach et al 2004). My objectives were 1) to quantify the amount of variation in detection rates among observers listening to actual point count data including estimating the numbers of species missed and falsely detected; 2) to compare the numbers and diversity of species detected on recordings with those reported by field observers to evaluate the potential for using recordings to expand BBS coverage, particularly in areas where it may be difficult to ensure adequate observer coverage (e.g. Canada's boreal forest); 3) to evaluate the usefulness of multiple observer methods for adjusting for species that were present but not detected and evaluate how these may work in the presence of identification errors.

METHODS

Field recordings

Omni-directional CZM stereo microphones (see Hobson et al. 2002 for details of microphones and Rempel et al. 2005 for configuration details) were used to record BBS routes as they were run in the field. Four routes in Ontario and Quebec were considered in the analysis. Two routes were recorded in eastern Ontario, Glen Tay (BBS #68-113, recorded June 8, 2005, BCR 13:

Lower Great Lakes/St. Lawrence Plains) and Desert Lake (BBS #68-313, recorded June 19, 2005, BCR 13); and two routes in west Quebec, Dumoine (BBS #72-127, recorded July 2, 2004, BCR 12) and Maniwaki (BBS #76-012, recorded June 25, 2004, BCR 12). All routes were recorded directly to iRiver iHP-120 hard drive recorders (iRiver Inc. Seoul, Korea) in WAV format. All field recordings began a few seconds before the start of each point count with a stop announcement and continued for several seconds after the stated end of the count.

I selected a subset of 20 of the 50 stops with minimal road noise from each of the routes we recorded for analysis. Using Cool Edit 2000 (Syntrillium 2000) sound editing software I edited each route into 3-minute intervals, replacing route and stop identifications from the field with new identifications for the experiment. I removed sections of the recordings with excessively loud noise from passing vehicles on 12 stops when it saturated the recordings, but the rest of the recording was retained for the experiment; otherwise recordings were unaltered.

Listening to recordings

I gave the four edited routes to 6 observers with extensive BBS or point count experience. Observers were asked to listen to each route once using high quality stereo headphones and without any other aids (e.g. no use of visualizations, spectrograms or multiple playbacks) and record all species they heard on each stop. I chose this approach, of listening only once to each

recording, to mimic field count conditions since in the field observers cannot rely on extraneous aids (e.g. spectrograms) or the ability to re-listen to portions of a song for individuals that are difficult to identify. However, observers were not asked to try to estimate the number of different individuals of each species present. To reduce variation in listening conditions for the recordings, a set of calibration recordings was developed with excerpts of a few high and low frequency bird calls of various volumes taken from the recordings. Observers were instructed to adjust volumes on their computer/playback devices until they were able to hear all species/frequencies on the recordings.

Data analysis

All statistical analyses were run using R 2.9.0 (R Development Core Team, 2009). I used multi-way ANOVA tests to compare the number of species detected by each observer (treatment). The treatment had 8 levels, 6 observers using rapid scan methods (Observers A – F), 1 field observer (FIELD) and one observer using intensive scanning methods (REC).

I used the RMark (Laake and Rexstad, 2008) package for R to generate and test double and multiple independent observer closed capture models in MARK (White and Burnham, 1999) for estimating species richness. Capture histories for each observer were generated using the uncorrected species detection lists. For all models observer was considered as an independent capture occasion; since there was no contact between observers when generating detection histories. I used the “Huggin’s closed capture” data type in

MARK and generated estimates of population size for an independent double observer model (Double OBS) for each pair of observers, as well as four multiple observer models: basic independent multiple observer models (Basic), multiple observer with heterogeneous detection probabilities (Heterogeneity), multiple observer models incorporating identification errors (Error) and multiple observer models with errors and heterogeneity (Het+Error). The estimated species richness estimates from MARK to the true number of species confirmed to be present on the recordings was compared to determine how well these models estimated species richness in the presence of detection errors.

I also selected the 10 most common species in the data set as well as 6 moderately abundant species (occurring on at least 10% of all stops) and used an occupancy modelling approach (MacKenzie et al. 2006, Royle and Link 2006) with the same closed population estimation models, to estimate the proportion of stops where the species was present. The estimated occupancy was compared with the known proportion of stops for which the species was present on the recordings, to evaluate the effects of observer error on occupancy estimates.

RESULTS

Comparisons among observers

On average, observers reported 9.3 ± 0.1 (mean $\pm$ standard error) species per stop across all routes, but the number reported varied among observers (repeated measures ANOVA, $F_{5,395} = 34.92$, $p < 0.0001$) from 7.8 to 10.6 (Figure

1, 2). This compares with a mean of 12.1 species per stop confirmed to be present on the recordings.

Observers correctly detected 8.0 ± 0.3 of these species per stop, but there was significant variation among observers in the number of correctly identified species ($F_{5, 395} = 33.87$, $p < 0.0001$). On average, observers reported an additional 1.3 ± 0.1 species per stop which I was unable to detect on the recordings; these included false positive detections (which the observers themselves were later unable to relocate) and species misidentifications. There was significant variation in the number of errors amongst observers ($F_{5, 395} = 13.55$, $p < 0.0001$).

Observers were likely to miss both common and rare species on the recordings (Table 1, Appendix 1). Of the 89 species confirmed on the recordings, every species occurring at more than 1 stop was missed by at least one observer on 1 or more stops; 44 species were missed by at least 1 observer on all stops where they occurred. There were 36 species-stop combinations on which I was able to confirm a species on the recordings that was missed by all observers in the study. Missed detections tended to be shared across observers, i.e. species were more likely to be overlooked by multiple observers than just one observer.

Most species were erroneously reported at least once by one or more observers. Error rates ranged up to 85% (for American redstart, *Setophaga ruticilla*) for confirmed species (Table 2, Appendix 1), although observers reported 17 additional species which were not confirmed on the recordings. Most errors were made by only one observer at any given stop, though multiple

observer errors were more common than single observer errors for some species such as Red-eyed Vireo (*Vireo olivaceus*) and Chipping Sparrow (*Spizella passerina*) (Appendix 1).

Comparison of field observers with recordings

Field observers tended to report more species in the field than most of the observers listening to recordings; this varied among routes since each route was run by a different field observer. On all routes the number of species reported by at least one listener was similar to or higher than those reported in the field (Figure 2). For two routes the observer who ran the initial BBS route participated in the microphone study but there was no difference in the average number of species detected in the field or using the recordings by these observers. However, the number of species reported both in the field and on the recordings were still lower than the number confirmed on the recordings.

I was able to confirm from 54% (on route 4, Maniwaiki) to 82% (on route 2, Desert Lake) of the species reported by the field observers on the recordings. Field observers overlooked between 2.2 ± 0.3 (on route 2) and 6.1 ± 0.5 (on route 4) species per stop that were confirmed to be present on the recordings (Figure 2). However, field observers reported between 1.4 ± 0.3 (route 2) and 2.3 ± 0.5 (route 4) additional species per stop that were not confirmed on the recordings. The additional species detected by an observer in the field could have been errors or they could have been detected visually or detectable aurally but not picked up by the microphones because they were too distant or were

masked by other sounds; therefore I considered these detections as unconfirmed rather than errors. However, many of these were species that were also misidentified by one or more listeners on the recordings.

Of the 48 species detected in the field on at least one stop that were not confirmed on the recordings all but 3 (each of which was reported on only 1 to 3 stops) were confirmed to be present on at least one other stop on the recordings; most were more likely to be missed in the field than on the recordings (Appendix 1). There were only a few species, such as Eastern Meadowlark (*Sturnella magna*) and Common Grackle (*Quiscalus quiscula*) that were more likely to be missed on the recordings than in the field (presumably due to visual detections), but even these were frequently detected by sound on recordings at other stops.

Adjusting for observer effects using multiple observer models

On average, mean species richness estimates generated using independent double and multiple observer models were positively biased compared to the true number of species present on the recordings (Figure 3). Estimates from double observer models, averaged across all pairs of observers were positively biased by 1.1 ± 0.2 species (mean $\pm$ standard error). Multiple observer models tended to be even more biased. AIC based model selection in MARK favoured either the heterogeneity or the heterogeneity with error (Het+Error) models (Table 3), but these generally produced the most biased species richness estimates: the heterogeneity model overestimated mean species richness by 8.7 ± 0.4 species, while the Het+Error model overestimated

richness by 6.4 ± 0.5 species for all routes. Multiple observer models assuming neither errors nor heterogeneity overestimated richness by 5.5 ± 0.3 additional species, while models allowing for errors but no heterogeneity overestimated by only 1.0 ± 0.3 species for all routes; however, they were clearly considered inappropriate by AIC model selection (Table 3).

Models of species specific occupancy estimates using double or multiple observer data tended to be positively biased compared to the true number of stops where the species was confirmed to be present (Table 4). For 9 species AIC model selection selected Heterogeneity models as the most appropriate but for all species tested ($n=16$) these models were positively biased and overestimated the true occupancy by 15.0% (12.0 stops). For the remaining 8 species heterogeneity with error (Het+Error) models were selected and occupancy estimates for all species ($n = 16$) were inflated by 10% (8.0 stops). All models produced positively biased occupancy estimates compared to the true value for most species; double observer models for 62.5% of species (10/16), multiple observer and heterogeneity for all species (16/16), error models for 68.8% (11/16) and heterogeneity with error models for 93.8% of species (15/16) tested.

DISCUSSION

Comparisons among observers

There was significant variation in the number of species reported by observers listening to recordings of BBS point counts. Of the species reported

68% of the species detected were confirmed on the recordings but the detection rate varied significantly amongst observers. Using recordings the observers missed 34% of the confirmed species on the recordings, the frequency of missed detections was similar to the estimates of 35% missed detections during a regular point count (Bart and Schoultz, 1984), 30% for loudspeaker playback experiments (Bart, 1985) and between 18 - 37% depending on singing rate for 5 species playback trials (Alldredge et al 2007). It should be noted that comparisons of these results with existing studies are slightly problematic since most studies of observer variation tend to focus on the number of individuals detected during a count rather than the number of species detected.

All of the observers in this study had extensive point count experience yet on most stops observers made both misidentifications and false positive (hearing something that wasn't truly there) errors. These detection errors accounted for 12% of the species detected by any observer. The total error rate for species detection rates is similar to the 9% detection error rate 3% misidentification reported by Bart (1985) for point count surveys. In several cases observers corrected detection errors when asked to double-check their data at a later time. Several of these errors were likely due to observers identifying species after only hearing part of a song or hearing a song overlapping with another sound (either a song or other background or ambient noise).

To date, most studies of observer variation have tended to focus on species specific detection probabilities rather than species richness estimates (e.g. Alldredge et al 2007b, Alldredge et al 2006a, Diefenbach et al. 2004). Often

these studies use simulated songbird communities by playing back recordings of a subset of species. My experimental design was in some ways more complex since I did not know *a priori* what species would be on the recordings. In some ways using recordings of actual point counts is advantageous since it includes a degree of complexity in overlapping songs of multiple species typical of true field point counts. However, there are likely several species which were missed or misidentified during my intensive surveys of the recordings.

I expected variation in both rates and types of species detection errors amongst observers listening to recordings. Factors such as familiarity with certain species, habitats and habitat-species interactions accounted for several detection errors made by the observers. For example, one of the observers was more familiar with boreal habitats than the Great Lakes eco-region test data set, consequently this individual tended to identify more boreal species than were truly present. Some species detection errors may have been eliminated using habitat cues that were not available on the recordings. Pine Warbler (*Dendroica pinus*), Chipping Sparrow, Dark-eyed Junco (*Junco hyemalis*) and Swamp Sparrow (*Melospiza georgiana*) songs tended to be amongst the most common detection errors made by listeners. Habitat cues typical of field point counts may have helped observers identify these species; for instance if there was a stand of pine trees in the vicinity of a stop, observers may have identified a trilling-chip song as a Pine Warbler rather than a Chipping Sparrow. It is interesting to note that McLaren and Cadman (1999) chose to avoid target species with songs that may be confusing for novice observers, such as Chipping Sparrow and Dark-

eyed Junco, in their experiment comparing count data collected by novice and experienced observers. It would appear that without habitat cues even experienced observers often confuse similar species.

Several listeners commented on the difficulty of establishing a mental image of the point count area and maintaining accurate “counts” of individuals and species during the duration of the count. These types of “tracking” errors may have caused listeners to lose track of which species they had detected earlier in the count and then identify the same individual as two different species. The study design required that volunteers to listen to a large number of counts but suggested they take breaks between stops. These breaks were necessary to maintain observer accuracy but may have resulted in similar between session “tracking” errors in species identification.

There was significant variation in species detection rates amongst observers using full chorus recordings. While observers detected a similar proportion of the true number of species available, missed detections and identification errors caused significant variation in the resulting species lists compared to the confirmed species list. Alldredge et al (2007) suggest that differences in detection probability between observers are due to differences in the rate at which detection probabilities decline due to distance (i.e. “better” observers are able to correctly detect species over a greater distance). I did not test the role of distance on detectability, it is possible that species that were more likely to be overlooked in this study were more faint on the recordings and therefore at a greater distance from the microphones. However, these results

indicate that an important part of the variation between observers in detection frequencies is due to identification error rather than detection failure.

Comparison of field observers with recordings

I was able to confirm between 54% and 82% of the species reported by field observers on the recordings. These results are slightly lower than those reported by Hobson et al (2002) of between 83% and 97% similarity between recordings and field observations of the same count. Although some individual observers detected fewer species using recordings than observers in the field, by combining and thoroughly checking the data of all observers against the recordings more species were confirmed on the recordings than detected by any individual listener or the observer in the field. Most species detected were shared amongst field observers and recordings. However there were several cases where species that were missed in the field were detected using recordings and species detected in the field were not confirmed on the recordings.

One of the goals was to compare species richness estimates between field observers and recordings. There were no obvious patterns in the types of species that tended to be detected in the field but not confirmed on the recordings. Several of the species detected only by field observers tended to be species that are typically detected visually as they fly over during a point count such as American goldfinch, eastern meadowlark, barn swallow or tree swallow (Appendix 1). However most of these species were also detected by at least one observer listening to the recordings. Since species richness estimates tended to

be higher using recordings yet still detect the majority of species detected in the field recordings may be a viable option for expanding BBS coverage in under-sampled regions such as Canada's boreal forest, where it may be difficult to get sufficient skilled observers to run field routes.

Recordings have been shown to be an effective tool for resampling and double checking observations in remote locations such as Peruvian rain forest (Hasselmayer and Quinn, 2000) where visual detection of birds is difficult. A similar double checking approach using recordings may be useful for BBS routes in under sampled areas. For instance, less skilled or novice observers could be sent into the field with microphones to record BBS point counts while focusing primarily on visual detections and counts of species. Recordings could then be analyzed or double checked by one or several skilled observers for additional species.

If recordings are to be used to enhance surveys of under sampled areas there should be further research evaluating most cost-effective method of generating accurate species lists. For example, the optimal number of observers needed to listen to recordings to needs to be determined before these types of studies become more common. More observers will likely detect more species but when do the number of errors out-weigh the proportion of novel detections amongst observers? Another consideration is how many times should any individual observer listen to a recording? Most of the detection errors in the data were corrected by observers when they re-listened to the recordings but is this the most efficient method of ensuring correct detections? Or is an approach

similar to the one described here, where the observer assembling the final data set is responsible for double checking all detections against the recording for accuracy, more efficient?

Regardless of which approach is taken to check data for errors observers listening to each recording should be given a description of the habitat where the point count was recorded to aid in identifying species with similar songs but with different habitat associations. Most point count protocols require observers to record specific information about each point count station such as date, start time of count, weather conditions and wind speed. This information along with a brief description of the adjacent habitat could easily be added to a stop announcement at the beginning of each point station recorded. This would give observers using recordings an idea of what is present in adjacent habitat as well as a brief period to adjust to the birds singing in the background without interrupting the flow of data collection or acoustic analysis.

Adjusting for observer effects using multiple observer models

I tested several types of multiple observer models all of which overestimated species ranging from 1.0 – 8.7 additional species per stop compared to the true richness. Abundance estimates for 2 and 4 observer multiple observer models were very similar to the number of individuals detected in the field for 5 target species but models using data from multiple observers (i.e. 4) were better able to account for bias in detection probabilities than models using data from only 2 observers (Alldredge et al 2006). Most studies using

multiple observer methods are used to generate estimates of abundance, i.e. number of individuals of a species detected rather than species richness. Compared to a typical multiple observer model our study design equates the number of species detected with the number of individuals detected and species richness estimates with abundance abundance estimates. However, these changes in the variables modeled make direct comparisons between our data and other studies difficult.

Heterogeneity models both with and without error terms were favoured by AIC as the most parsimonious models for both species richness and occupancy estimates in our study but the estimates tended to be strongly positively biased. These results do not agree with Alldredge et al's (2006) results that multiple observer models that allow for heterogeneity in detectability probability to be the best models for most species in their study. Models that included error terms without heterogeneity tended to be less biased for species richness estimates but were never favoured by AIC. The role of detection errors in species occupancy models was more variable. Models that took into account errors tended to underestimate occupancy for species where only one observer falsely detected the species. However, when several observers made the same species detection error occupancy models with error terms overestimated occupancy. This data indicates that caution should be used when modeling detection probability using occupancy models that do not account for multiple detection errors amongst observers.

Estimates of both species richness and occupancy were the least biased using models that included error terms. At least part of the bias associated with these models is likely a result of the frequency of errors amongst observers. A key assumption of the error models is that errors are rare and that the same errors are unlikely to occur twice or multiple times amongst observers (MacKenzie et al 2006, MacKenzie et al 2004, MacKenzie et al 2002). Alldredge et al (2006) suggest that matching errors (i.e. number of individuals counted) across observers should be minimal if surveys are conducted during the breeding seasons when territories are fixed and birds are not in large flocks, typical of BBS field survey conditions. However, as part of their study Alldredge et al (2006) they chose to lump species of vireo, blue-headed vireo (*Vireo solitarius*) and red-eyed vireo, in their analysis because of likely identification errors. These results indicate that errors are not rare and that the same error will be made multiple times even by skilled observers. In addition, the same detection error was made on at least one stop for 40% (n=36) of the 89 species confirmed on the recordings. While the assumption of relatively few matching errors in the number of individuals detected amongst observers holds, this data shows that at the species level matching detections across observers is more problematic.

Alldredge et al (2006) cited 2 concerns with multiple observer methods, ensuring the independence of observers and correctly matching the number of detections amongst observers for abundance estimation. They suggest that if there are many species per count and detections are primarily auditory then independence should be attained. This study design meant that observers were

truly independent. However, these BBS recordings also fit the characteristics they suggest are necessary to ensure independence: the detections are auditory and there are many species per count. Johnson (2008) suggested that the conditions that need to be met for observer independence would increase the number of matching errors, my results support this suggestion. Of the two concerns suggested by Alldredge et al (2006) independence amongst observers is relatively easy to counter act in the study design (i.e. by using recordings). However, matching errors in species detection are more problematic. The methods to correct for detection errors, at least at the species level, are not adequate and ignoring detection errors result in biased population estimates from multiple observer methods.

It was expected that observers would vary in the number of species detected during a survey and to make a few errors during the count, they are, after all, only human. Detection errors are more common than multiple observer and site occupancy models assume; my data shows that ignoring errors results in biased species richness estimates compared to the true values. Since identification errors are shared amongst observers a “majority rules” approach for determining which species detections are actually errors would not be the best method of correcting detection histories for multiple observer models. Until errors are adequately accounted for (possibly by modelling error rates using a similar approach as heterogeneity estimates) multiple observer methods should be approached with caution for multi-species surveys.

Figure 1. Mean number of confirmed species detections (dark grey) and detection errors (light grey) for 6 observers (A – F) listening to full chorus recordings of 80 BBS routes ($n = 20$ stops per route) compared to the number of species detected using multiple playbacks and spectrograms (REC). There was significant variation among observers in the number of confirmed species detections and the number of detection errors.

Figure 2. Comparison for each of four BBS routes, of the mean number of confirmed species detections (dark grey) and detection errors (light grey) for 6 observers (A – F) listening to full chorus recordings with the mean number of confirmed (dark grey) and unconfirmed (medium grey) species detections from the observers that conducted the BBS routes in the field (FIELD) and the number of species confirmed on the recordings using multiple playbacks and spectrograms (REC). Note that unconfirmed field detections may include species detected visually or that were audible in the field but not captured on the recordings. Data are mean counts for 20 stops per route. Asterisks indicate significant difference (* $p = 0.05$, ** $p = 0.01$, *** $p = 0.001$) in the mean count reported by a listener compared to the number of species reported by the observer in the field.

Figure 3. Species richness estimates (mean $\pm$ standard error) generated using double and multiple observer models run in MARK for 4 BBS routes. Double observer estimates are mean estimates across all double observer pairs ($n = 15$). Most parsimonious models among the multiple observer models (based on AIC) are marked with an asterisk (*) – see Table 3. Solid horizontal line represents the true species richness for each route ($n = 20$ stops per route).

Table 1. Most commonly missed species by observers listening to full chorus recordings of 80 stops from 4 BBS routes indicating number of stops where a species was confirmed, number of stops where the species was missed by only one observer, by more than one observer ($2 \leq n \leq 5$) or by all observers ($n = 6$). Percent missed detections is the total number of missed stops divided by the total number of confirmed detections. Subset of 5 species (of $n = 44$) which were missed by at least one observer on all stops and 5 species of the remaining $n = 45$ confirmed on the recordings (see Appendix 1 for full species table).

Species	Number of stops				% stops with one or more missed
	Confirmed observer	Missed by 1 observer	Missed by 2-5 observers	Missed by all 6 observers	
European Starling	21	1	18	2	100
Common Grackle	15	1	14	0	100
Eastern Kingbird	12	0	10	2	100
Baltimore Oriole	11	2	9	0	100
Great Crested Flycatcher	10	2	7	1	100
Black-and-white Warbler	17	1	15	0	94.12
Magnolia Warbler	13	0	11	1	92.31
Cedar Waxwing	17	3	12	0	88.24
Mourning Warbler	8	1	6	0	87.5
Swainson's Thrush	7	2	4	0	85.71

Table 2. Subset of the 10 species most commonly detected erroneously by observers listening to full chorus recordings of 4 BBS routes indicating number of stops where a species was confirmed on the recordings, number of stops where the species was erroneously reported by exactly one observer, or by more than one observer. Percent error is the total number of stops with a detection error divided by the total number of stops where the species was reported (see Appendix 1 for full species table).

Species	Number of stops				% stops with one or more errors
	Confirmed	Error by 1 observer	Error by 2 – 5 observers		
American Redstart	3	12	5		85
Purple Finch	2	11	0		84.62
Pileated Woodpecker	2	4	4		80
Pine Warbler	1	4	0		80
Blackburnian Warbler	4	6	6		75
Downy Woodpecker	1	3	0		75
Brown Creeper	2	5	0		71.43
Dark-eyed Junco	2	3	2		71.43
Hairy Woodpecker	4	3	6		69.23
Canada Warbler	1	2	0		66.67

Table 3. Model selection table (ΔAIC values) for multiple independent observer models of species richness, with 0 indicating the selected model.

Route	Model			
	Basic	Heterogeneity	Error	Het+Error
1 - Glen Tay	577.2	16.0	312.6	0.0
2 - Desert Lake	610.4	10.4	394.4	0.0
3 - Dumoine	486.2	0.0	366.1	0.0
4 - Maniwaki	623.3	0.0	434.4	1.9

Table 4. Model estimates for species specific detection probability measured as mean number of stops where a species is detected $\pm$ mean standard error for all tops. Table includes total number of stops where a species was detected and total number of confirmed detections. Number of detection errors represents the number of times a species was erroneously detected by only one observer and the number of times more than one observer made the same detection error on the same stop. Model estimates of number of stops where species were detected are followed by the direction of the deviance of the estimate from the confirmed number of stops (+ greater number of stops, - fewer stops). Average detection error is calculated as the mean difference of all species specific estimates from the confirmed number of stops. Model bias is the number of species ($n = 16$) where model estimates were greater than the number of confirmed detections.

Species	Total Number of Stops		Number of Stops with Detection Errors		Model Estimated Number of Stops					
	Reported	Confirmed	Observer = 1	Observer >1	Double OBS	Multiple Observer Models				Het+Error
						Basic	Heterogeneity	Error		
AMRO	60	53	5	2	54.4 (0.48)	+ 60.0 (0.05)	+ 47.2 (1.01)	+ 56.1 (2.10)	+ 59.0 (2.05)	+
AMCR	59	53	4	2	52.9 (0.96)	- 59.0 (0.05)	+ 52.1 (3.19)	+ 52.0 (2.60)	- 58.9 (2.09)	+
SOSP	51	44	5	2	45.1 (0.41)	+ 51.0 (0.08)	+ 38.0 (1.29)	+ 46.8 (2.12)	+ 50.7 (2.19)	+
REVI	51	33	8	10	42.5 (1.23)	+ 51.1 (0.24)	+ 62.0 (1.20)	+ 44.1 (2.58)	+ 48.8 (2.98)	+
COYE	46	29	9	8	37.3 (1.21)	+ 46.0 (0.18)	+ 52.1 (2.23)	+ 38.3 (2.69)	+ 40.3 (2.96)	+
RWBL	45	39	6	0	38.6 (0.66)	- 45.0 (0.08)	+ 28.2 (2.58)	+ 37.5 (2.59)	- 46.4 (2.14)	+
CHSP	40	26	6	8	30.7 (1.96)	+ 40.3 (0.59)	+ 48.5 (3.08)	+ 31.4 (2.90)	+ 40.9 (3.21)	+
BLJA	38	32	6	0	30.0 (0.99)	- 38.2 (0.43)	+ 60.7 (2.98)	+ 28.5 (2.74)	- 30.2 (3.17)	-
YWAR	37	22	13	2	27.8 (0.99)	+ 37.8 (0.99)	+ 43.8 (9.42)	+ 25.2 (3.34)	+ 26.4 (3.80)	+
BCCH	36	25	8	3	25.5 (2.07)	+ 36.2 (0.41)	+ 32.3 (2.23)	+ 24.8 (2.95)	- 35.7 (3.24)	+
WTSP	32	27	4	1	27.5 (0.71)	+ 32.0 (0.06)	+ 42.0 (0.64)	+ 29.4 (1.74)	+ 32.3 (0.64)	+
CSWA	27	19	7	1	19.5 (1.48)	- 27.4 (0.67)	+ 51.3 (2.60)	+ 21.2 (2.59)	+ 30.0 (2.60)	+
AMGO	24	14	10	0	15.5 (0.85)	+ 24.4 (0.69)	+ 25.0 (3.19)	+ 12.7 (2.65)	- 27.8 (3.24)	+
WOTH	21	15	7	0	14.0 (1.22)	- 21.4 (0.64)	+ 30.0 (3.76)	+ 12.6 (2.30)	- 21.9 (1.81)	+
YRWA	20	11	4	5	16.2 (0.74)	+ 20.2 (0.48)	+ 23.5 (3.30)	+ 17.9 (2.33)	+ 23.5 (3.30)	+
WIWR	16	13	3	0	13.2 (0.33)	+ 16.0 (0.07)	+ 18.4 (3.95)	+ 12.0 (1.73)	- 18.6 (3.64)	+
ALFL	12	12	0	0	9.5 (0.63)	- 12.1 (0.32)	+ 16.8 (9.39)	+ 12.1 (0.32)	+ 12.4 (0.76)	+
Average Deviation Across Species					1.9 (0.88)	+ 8.8 (1.25)	+ 12.0 (1.60)	+ 2.0 (1.04)	+ 8.0 (1.21)	+
Percent Species Overestimated					62.5	100	100	68.8	93.8	

Appendix 1. Table of all species detections on a total of 80 stops from 4 Breeding Bird Survey routes. Data includes number of stops where each species was confirmed on the recordings, detected only by the field observer, detected only on the recordings, and both detected in the field and confirmed on the recordings. Also includes number of stops at which one or more of the 6 observers listening to the recordings missed the species when it was confirmed on the recording, or erroneously reported the species when it could not later be confirmed. Percent missed detections is the total number of missed stops divided by the total number of confirmed detections. Percent error is the total number of stops with a detection error divided by the total number of reported detections of the species.

Species	Latin Name	AOU Species Code	Number of stops		% Stops
			Confirmed	Field Only	
American Crow	<i>Corvus brachyrhynchos</i>	AMCR	53	3	9
American Robin	<i>Turdus migratorius</i>	AMRO	53	0	13
Song Sparrow	<i>Melospiza melodia</i>	SOSP	44	1	10
Red-winged Blackbird	<i>Agelaius phoeniceus</i>	RWBL	39	0	9
Red-eyed Vireo	<i>Vireo olivaceus</i>	REVI	33	5	6
Blue Jay	<i>Cyanocitta cristata</i>	BLJA	32	5	11
Common Yellowthroat	<i>Geothlypis trichas</i>	COYE	29	3	5
Chipping Sparrow	<i>Spizella passerina</i>	CHSP	28	7	6
White-throated Sparrow	<i>Zonotrichia albicollis</i>	WTSP	27	4	7
				20	
				22	
				24	
				21	
				27	
				30	
				34	
				40	
				44	
				7	
				10	
				11	
				10	
				0	
				4	
				5	
				2	
				32.1	
				34	
				34.1	
				30.8	
				42.4	
				59.4	
				24.1	
				82.1	
				37	
				33.3	
				1	
				15.6	
				18	
				35.3	
				13.3	
				13.7	
				11.7	
				10.2	
				% One or More Error	
				% One or More Missed	
				Error by 2 – 5 Observers	
				Error by 1 Observer	
				Missed by all 6 Observers	
				Missed by 2-5 Observers	
				Missed by 1 Observer	
				Field and Recording	
				Recording Only	
				Field Only	
				Confirmed	

Black-capped Chickadee	<i>Poecile atricapillus</i>	BCCH	25	7	6	19	2	16	1	9	3	68	32.4
Mourning Dove	<i>Zenaida macroura</i>	MODO	24	1	5	19	8	7	0	8	2	62.5	29.4
Bobolink	<i>Dolichonyx oryzivorus</i>	BOBO	23	3	3	20	6	12	1	3	0	82.6	11.5
Yellow Warbler	<i>Dendroica petechia</i>	YWAR	22	3	7	15	4	12	0	13	2	72.7	40.5
European Starling	<i>Sturnus vulgaris</i>	EUST	21	5	7	14	1	18	2	1	1	100	8.7
Ovenbird	<i>Seiurus aurocapilla</i>	OVEN	21	0	7	14	1	6	2	4	1	42.9	19.2
Chestnut-sided Warbler	<i>Dendroica pensylvanica</i>	CSWA	20	1	10	10	4	12	1	7	1	85	28.6
	<i>Passerculus</i>												
Savannah Sparrow	<i>sandwichensis</i>	SAVS	20	6	3	17	4	8	0	8	1	60	31
Nashville Warbler	<i>Vermivora ruficapilla</i>	NAWA	19	0	7	12	5	8	0	4	2	68.4	24
Red-breasted Nuthatch	<i>Sitta canadensis</i>	RBNH	18	0	9	9	2	8	2	4	0	66.7	18.2
Black-and-white Warbler	<i>Mniotilta varia</i>	BAWW	17	0	10	7	1	15	0	3	0	94.1	15
Cedar Waxwing	<i>Bombicilla cedrorum</i>	CEDW	17	0	9	8	3	12	0	4	0	88.2	19.1
Eastern Meadowlark	<i>Sturnella magna</i>	EAME	17	6	1	16	2	6	1	8	3	52.9	39.3
House Wren	<i>Troglodytes aedon</i>	HOWR	17	0	4	13	4	8	0	4	0	70.6	19.1
Hermit Thrush	<i>Catharus guttatus</i>	HETH	16	0	4	12	0	5	0	4	0	31.3	20
Common Grackle	<i>Quiscalus quiscula</i>	COGR	15	7	4	11	1	14	0	11	2	100	46.4
Wood Thrush	<i>Hylocichla mustelina</i>	WOTH	15	0	2	13	2	9	1	7	0	80	31.8
American Goldfinch	<i>Carduelis tristis</i>	AMGO	14	6	4	10	3	5	0	10	0	57.1	41.7
Magnolia Warbler	<i>Dendroica magnolia</i>	MAWA	13	2	5	8	0	11	1	9	3	92.3	48
Winter Wren	<i>Troglodytes troglodytes</i>	WWR	13	0	4	9	1	3	0	3	0	30.8	18.8
Alder Flycatcher	<i>Empidonax alnorum</i>	ALFL	12	2	7	5	1	9	0	0	0	83.3	0
Eastern Kingbird	<i>Tyrannus tyrannus</i>	EAKI	12	4	4	8	0	10	2	2	0	100	14.3
Baltimore Oriole	<i>Icterus galbula</i>	BAOR	11	2	4	7	2	9	0	6	2	100	42.1
Yellow-rumped Warbler	<i>Dendroica coronata</i>	YRWA	11	2	6	5	4	5	0	4	5	81.8	45
Great Crested Flycatcher	<i>Myiarchus crinitus</i>	GCFL	10	0	5	5	2	7	1	5	0	100	33.3
Veery	<i>Catharus fuscescens</i>	VEER	10	2	2	8	2	6	0	4	1	80	33.3
Warbling Vireo	<i>Vireo gilvus</i>	WAVI	10	4	2	8	1	9	0	4	2	100	37.5
Barn Swallow	<i>Hirundo rustica</i>	BARS	9	4	1	8	0	6	3	2	0	100	18.2
House Sparrow	<i>Passer domesticus</i>	HOSP	9	2	3	6	3	6	0	2	0	100	18.2

Northern Flicker	<i>Colaptes auratus</i>	NOFL	9	4	4	5	0	6	0	6	2	66.7	47.1
Rose-breasted Grosbeak	<i>Pheucticus ludovicianus</i>	RBGR	9	1	3	6	0	8	1	3	1	100	30.8
Mourning Warbler	<i>Oporornis philadelphia</i>	MOWA	8	2	4	4	1	6	0	1	1	87.5	20
Brown Thrasher	<i>Toxostoma rufum</i>	BRTH	7	2	2	5	0	2	0	1	0	28.6	12.5
Brown-headed Cowbird	<i>Molothrus ater</i>	BHCO	7	2	3	4	1	4	2	4	0	100	36.4
Swainson's Thrush	<i>Catharus ustulatus</i>	SWTH	7	0	1	6	2	4	0	4	0	85.7	36.4
Black-throated Blue Warbler	<i>Dendroica caerulescens</i>	BTBW	6	0	1	5	0	2	0	4	0	33.3	40
Golden-crowned Kinglet	<i>Regulus satrapa</i>	GCKI	6	2	2	4	0	5	1	5	1	100	50
Indigo Bunting	<i>Passerina cyanea</i>	INBU	6	0	3	3	1	3	1	5	0	83.3	45.5
Swamp Sparrow	<i>Melospiza georgiana</i>	SWSP	6	2	2	4	0	5	0	7	4	83.3	64.7
Tree Swallow	<i>Tachycineta bicolor</i>	TRES	6	3	4	2	0	4	2	4	0	100	40
Common Loon	<i>Gavia immer</i>	COLO	5	2	0	5	1	1	0	2	0	40	28.6
Eastern Wood Pewee	<i>Contopus virens</i>	EAWP	5	4	1	4	1	4	0	4	1	100	50
Killdeer	<i>Charadrius vociferus</i>	KILL	5	3	1	4	2	1	1	2	0	80	28.6
Blackburnian Warbler	<i>Dendroica fusca</i>	BLBW	4	0	0	4	2	2	0	6	6	100	75
Black-throated Green Warbler	<i>Dendroica virens</i>	BTNW	4	0	2	2	0	2	1	0	0	75	0
Hairy Woodpecker	<i>Picoides villosus</i>	HAWO	4	1	2	2	0	4	0	3	6	100	69.2
White-breasted Nuthatch	<i>Sitta carolinensis</i>	WBNU	4	2	2	2	1	3	0	4	0	100	50
American Redstart	<i>Setophaga ruticilla</i>	AMRE	3	2	2	1	0	2	0	12	5	66.7	85
Blue-headed Vireo	<i>Vireo solitarius</i>	BHVI	3	0	2	1	0	3	0	1	0	100	25
Eastern Phoebe	<i>Sayornis phoebe</i>	EAPH	3	4	1	2	0	1	0	4	1	33.3	62.5
Gray Catbird	<i>Dumetella carolinensis</i>	GRCA	3	2	1	2	0	1	1	4	0	66.7	57.1
Wilson's Snipe	<i>Gallinago delicata</i>	WISN	3	0	0	3	0	2	0	0	0	66.7	0
Yellow-bellied Sapsucker	<i>Sphyrapicus varius</i>	YBSA	3	1	2	1	1	2	0	4	0	100	57.1
Brown Creeper	<i>Certhia americana</i>	BRCR	2	0	0	2	0	1	1	5	0	100	71.4
Dark-eyed Junco	<i>Junco hyemalis</i>	deju	2	1	2	0	0	1	1	3	2	100	71.4
Northern Parula	<i>Parula americana</i>	NOPA	2	0	1	1	0	2	0	0	0	100	0
Northern Waterthrush	<i>Seiurus noveboracensis</i>	NOWA	2	1	0	2	1	0	0	3	0	50	60

Pileated Woodpecker	<i>Dryocopus pileatus</i>	PIWO	2	0	2	0	1	1	0	4	4	100	80
Purple Finch	<i>Carpodacus purpureus</i>	PUIF	2	2	2	0	0	2	0	11	0	100	84.6
Ring-billed Gull	<i>Larus delawarensis</i>	RBGU	2	0	1	1	0	1	1	0	0	100	0
Ruby-throated Hummingbird	<i>Archilochus colubris</i>	RTHU	2	0	2	0	0	2	0	2	0	100	50
Scarlet Tanager	<i>Piranga olivacea</i>	SCTA	2	1	1	1	0	2	0	2	1	100	60
Upland Sandpiper	<i>Bartramia longicauda</i>	UPSA	2	0	2	0	0	2	0	0	0	100	0
Vesper Sparrow	<i>Poocetes gramineus</i>	VESP	2	1	1	1	0	2	0	0	0	100	0
Belted Kingfisher	<i>Ceryle alcyon</i>	BEKI	1	1	0	1	1	0	0	1	0	100	50
Canada Warbler	<i>Wilsonia canadensis</i>	CAWA	1	0	1	0	1	0	0	2	0	100	66.7
Chimney Swift	<i>Chaetura pelagica</i>	CHSW	1	0	1	0	0	1	0	0	0	100	0
Common Nighthawk	<i>Chordeiles minor</i>	CONI	1	0	1	0	0	1	0	0	0	100	0
Common Raven	<i>Corvus corax</i>	CORA	1	1	1	0	0	1	0	2	0	100	66.7
Downy Woodpecker	<i>Picoides pubescens</i>	DOWO	1	2	1	0	0	1	0	3	0	100	75
Eastern Bluebird	<i>Sialia sialis</i>	EABL	1	0	0	1	0	1	0	2	0	100	66.7
Eastern Towhee	<i>Pipilo erythrophthalmus</i>	EATO	1	1	0	1	0	1	0	0	0	100	0
Grasshopper Sparrow	<i>Ammodramus savannarum</i>	GRSP	1	0	0	1	0	0	0	0	0	0	0
Mallard	<i>Anas platyrhynchos</i>	MALL	1	0	0	1	1	0	0	0	0	100	0
Northern Bobwhite	<i>Colinus virginianus</i>	NOBO	1	0	0	1	1	0	0	0	0	100	0
Northern Cardinal	<i>Cardinalis cardinalis</i>	NOCA	1	0	1	0	0	1	0	0	0	100	0
Pine Warbler	<i>Dendroica pinus</i>	PIWA	1	1	0	1	0	1	0	4	0	100	80
Rock Pigeon	<i>Columba livia</i>	ROPI	1	4	0	1	0	0	1	0	0	100	0
Ruby-crowned Kinglet	<i>Regulus calendula</i>	RCKI	1	0	0	1	0	1	0	1	0	100	50
Tennessee Warbler	<i>Vermivora peregrina</i>	TEWA	1	0	1	0	0	1	0	0	0	100	0
Yellow-bellied Flycatcher	<i>Empidonax flaviventris</i>	YBFL	1	0	0	1	0	1	0	1	1	100	66.7
American Bittern	<i>Botaurus lentiginosus</i>	AMBI	0	3	0	0	0	0	0	0	1	N/A	100
Bay-breasted Warbler	<i>Dendroica castanea</i>	BBWA	0	0	0	0	0	0	0	1	0	N/A	100
Black-billed Cuckoo	<i>Coccyzus erythrophthalmus</i>	BBCU	0	1	0	0	0	0	0	1	0	N/A	100

Canada Goose	<i>Branta canadensis</i>	CAGO	0	0	0	0	0	0	0	0	1	0	N/A	100
Cape May Warbler	<i>Dendroica tigrina</i>	CMWA	0	0	0	0	0	0	0	0	1	0	N/A	100
Evening Grosbeak	<i>Coccothraustes vespertinus</i>	EVGR	0	0	0	0	0	0	0	0	1	0	N/A	100
Field Sparrow	<i>Spizella pusilla</i>	FISP	0	0	0	0	0	0	0	0	1	0	N/A	100
Great Blue Heron	<i>Ardea herodias</i>	GBHE	0	0	0	0	0	0	0	0	1	0	N/A	100
Herring Gull	<i>Larus argentatus</i>	HERG	0	0	0	0	0	0	0	0	0	1	N/A	100
House Finch	<i>Carpodacus mexicanus</i>	HOFI	0	2	0	0	0	0	0	0	1	0	N/A	100
Least Flycatcher	<i>Empidonax minimus</i>	LEFL	0	0	0	0	0	0	0	0	2	0	N/A	100
Merlin	<i>Falco columbarius</i>	MERL	0	0	0	0	0	0	0	0	1	1	N/A	100
Olive-sided Flycatcher	<i>Contopus cooperi</i>	OSFL	0	0	0	0	0	0	0	0	1	0	N/A	100
Orange-crowned Warbler	<i>Vermivora celata</i>	OCWA	0	0	0	0	0	0	0	0	1	0	N/A	100
Palm Warbler	<i>Dendroica palmarum</i>	PAWA	0	0	0	0	0	0	0	0	1	0	N/A	100
Pine Siskin	<i>Carduelis pinus</i>	PISI	0	0	0	0	0	0	0	0	1	0	N/A	100
Ruffed Grouse	<i>Bonasa umbellus</i>	RUGR	0	0	0	0	0	0	0	0	1	0	N/A	100

**Using a GPS synchronized microphone array and acoustic
localization system to investigate observer induced behavioural changes in
singing birds during point count surveys**

Point count surveys are one of the most widely used methods of monitoring of songbird populations. Annual point count surveys conducted by thousands of volunteers are the foundation of the North American Breeding Bird Survey (BBS), one of the longest running and largest scale avian monitoring projects in North America (Sauer et al. 2008). Many other major avian monitoring programs including provincial and state run breeding bird atlases (e.g. 2nd Atlas of the Breeding Birds of Ontario 2001-2005, Cadman et al 2007) and most forest bird monitoring programs depend, at least in part, on data collected using point count surveys (e.g. Ontario Forest Bird Monitoring Program, Welsh 1995).

During a point count an observer records the number of individuals of each species detected within a fixed time interval, generally following the recommendations of Ralph et al. (1995) although there are many variations in details of protocols. The duration of point counts varies among surveys, from 3 minutes (e.g., Breeding Bird Survey; Sauer et al. 2008), to 5 minutes (e.g., Ontario Breeding Bird Atlas; Cadman et al, 2007) to 10 minutes (e.g. Ontario Forest Bird Monitoring Program; Welsh 1995) or longer. In some cases the point count is subdivided into multiple time intervals, and the interval when a bird is first detected is recorded (Ralph et al 1995). Some protocols record all birds detected regardless of distance from the center of the point (unlimited radius

counts); while others only record birds within a fixed radius of the point (e.g., 30m, 50m or 100m). Other protocols record the precise distance to each bird detected either by pacing (e.g. Norvell et al. 2003) or by using laser range finders (e.g. Buckland 2006)

Most analyses of point count data treat the number of birds detected per count as an index of population size, where the index is related to true population size with the equation:

$$N = C/p$$

where N is the true population size, C is the number of individuals detected during the point count and p is the probability of detecting an individual from the population (Nichols et al 2002, Barker et al 1993). If one assumes that the detection probability, p, is constant, or at least that it varies proportionally substantially less than the variation of the true population size (Johnson, 2008), then changes in the index can be used to estimate population trends over time (e.g. BBS trend analysis, Sauer et al 2008) or differences in relative abundance across space (e.g. Ontario Breeding Bird Atlas, Cadman et al, 2007). Long term trend data based on count indices are key considerations when identifying species at risk and setting priorities for major management programs (Rosenberg and Blancher 2005, Dunn 2002, Donovan et al. 2002). For instance, Partners in Flight (PIF) uses population trend data, usually from the BBS, in addition to regional, national and global abundance and range data to set conservation priorities for landbird species in Canada (Dunn et al 1999) and the United States (Carter et al. 2000).

However, the assumption that detectability does not change over time has been challenged (e.g. Sauer et al 2004, Thompson 2002, Rosenstock et al. 2002, Nichols et al. 2000) and a variety of approaches have been proposed to estimate detection probabilities. Distance sampling uses the estimated distance of a bird from an observer during a point count to calculate density estimates (Buckland et al 2001, 2004). Multiple observer methods (Alldredge et al 2006a), including double observer methods (Nichols et al 2000), compare detection histories amongst observers conducting simultaneous counts to model detectability. Time of detection methods divide counts into shorter time intervals and use modified capture - recapture methods to determine the probability of an individual being detected during the count using either removal models if only the interval of the first detection is recorded (Farnsworth et al. 2002) or with more general time of detection models if all intervals when an individual is detected are recorded (Alldredge et al 2007b, Alldredge et al 2007c). Mixed methods combine at least two methods such as distance and double observer methods (Kissling and Garton, 2006) to estimate various components of detectability.

Distance sampling has the potential to be used to estimate true population densities provided several key assumptions are met. Two of these assumptions are that observers are able to accurately estimate the distance to a singing bird and that birds do not move or change their singing behaviour in response to an observer (Buckland et al 2001, 2004).

The ability of observers to estimate distance to a bird depends on how a bird is detected, either visually or aurally. Generally, following intensive training

observers are able to accurately gauge distances to birds detected visually either by pacing (Somershoe et al 2006, Norvell et al 2003) or using laser range finders (Buckland 2006). However, the accuracy of distance estimation with respect to birds detected aurally has been questioned. In a series of playback trials using an array of MP3 players Simons et al (2007) showed that observers tend to overestimate population counts by between 17-22% at 50m. Using the same playback system Alldredge et al (2007d) showed that observers made substantial distance estimation errors for all auditory detections with or without training. Observers could not accurately differentiate distances beyond 65m and their estimates of distance to a sound source could be out by a factor of 2 depending on the orientation of the playback speaker (Alldredge et al. 2007d).

The assumption that birds do not move in response to an observer has rarely been tested or quantified. Bart et al. (1995) suggested that any birds that flush within a 50m radius of a point count station when an observer enters or leaves should be counted as present during the count. Others have suggested allowing a short time interval between the arrival of an observer and the start of a point count to allow birds to “adjust” to the observer’s presence (Ellingson and Lukas 2003). One reason for the lack of studies measuring the effect of observers on bird behaviour during a point count (Bye et al 2001) is due to the difficulty of obtaining an objective sample when the observer is absent. To date most studies of observer effect have been based on estimates of detectability period (the time a bird remains near its initial flush point) and approachability (the

minimum distance between an observer and a bird before it flushes; Gutzwiller and Anderson 1999, Gutzwiller et al. 1998).

In this paper I develop a novel approach to test if the singing behaviour of songbirds during a point count is affected by the presence of an observer using an array of synchronously recording microphones that can detect and estimate the positions of birds in the absence of an observer. The objectives were to test the accuracy of an array of synchronized microphones when estimating the position of singing birds and to determine whether the number of individuals available for detection, the position of singing birds or the singing frequency of birds changed in response to the presence of an observer during a point count.

METHODS

Study Site

The study was conducted within the Prince Edward Point (PEPt) National Wildlife Area, a peninsula located on the north-eastern shore of Lake Ontario, Ontario, Canada (43.93° N, 76.88° W). PEPt National Wildlife Area comprises 560ha of mixed habitat including limestone cliffs, rocky beaches, mixed old field, scrub and forest habitat. I set up the experiment in regenerating pasture habitat. I estimated vegetation cover by walking 4 transects and counting the number of trees or shrubs less than 2m tall and over 2m tall from the center of each array once it was deployed to the nearest microphone in each of the cardinal direction. Mean vegetation composition was calculated across all transects on all array locations; mean ground cover consisted of 65.9% grasses or herbs with 31.7%

low shrubs under 2m tall. The site was open canopy with 2.4% cover by trees or shrubs over 2m tall. Shrubs were predominantly juniper (*Juniperus* spp), regenerating cedar (*Thuja* spp) and alder (*Alnus* spp), most trees over 2m tall were mature oak (*Quercus* spp), maple (*Acer* spp), or fruit trees (e.g. apple, *Pyrus* spp) presumably shade trees from regenerating pasture lands, hedgerows or orchards.

Field Methods

Between June 1 – July 4, 2007, I set up microphone arrays consisting of 8 synchronously recording autonomous recording units (ARU) developed by the Bioacoustic Research Program at the Cornell Laboratory of Ornithology, Ithaca, New York. Each ARU consisted of a microphone, amplifier, frequency filter, programmable computer, external power source, hard-drive recorder and Global Positioning System (GPS) receiver. The GPS receiver on each unit continuously records a time code which is used after data extraction to synchronize recordings between all units. The GPS synchronization eliminates the need for wires physically connecting microphones in the field allowing us to move the array much more easily.

Each array consisted of 7 ARUs positioned evenly around a 50m radius circle with the 8th ARU placed in the center (Figure 1). The centers of all arrays were separated by 200m to minimize repeated sampling of the same birds. Each array was laid out using ARC-Pad (ESRI, Redlands CA) and a Thales mobile mapper sub-meter GPS unit (Magellan Inc. Santa Clara CA). Arrays were set up

in the early afternoon and moved after each day's sampling. Each unit was programmed to record daily in the evening (19:00-22:00) and again the following morning (3:30-10:00). A total of 28 locations were sampled inside the National Wildlife Area, but two arrays were dropped from analysis due to the failure of multiple microphones; therefore, the analysis is based on 26 array locations.

Three 10-minute point counts were conducted each morning while the microphones were recording. During each point count, an observer walked directly into the center of the array and began a 10-minute point count approximately 1 minute after arrival. The first point count started approximately at sunrise to sample late dawn chorus and subsequent point count trials were run 1.5 and 3.0 hours later. The observer was absent from the system at all times except during point count trials to minimize the exposure of birds in our study site to the observer.

To test the accuracy of the localization algorithm and verify alignment of the recording files, a series of whistle trials were run in several of the arrays at the end of the morning after the point counts were completed. Whistles were made without a mechanical aid and consisted of a short frequency shift followed by a relatively flat down-slur ranging in frequency from 600 – 4000 Hz. During each whistle trial, we repeated 5-8 whistles from the center and at distances ranging from 43m to 125m from the center of the array. Whistle locations were determined using the Thales sub-metered GPS unit. I calculated localization error as the difference in the estimated location generated using the localization algorithm and the known location obtained from the sub-metered GPS unit.

Data Analysis

Once field trials were completed, I extracted and synchronized the data from the ARUs using BEAST (Tom Calupca, Ithaca, NY. pers. comm), an ARU specific data extraction and synchronization program, and generated 8 channel AIFF files for analysis. All acoustic analysis was done using XBAT, an acoustic analysis software program developed in MATLAB (Math Works Inc., Natick, MA) by the Bioacoustic Research Program at Cornell Laboratory of Ornithology (Harold Figueroa, Ithaca, NY, pers. comm. <http://xbat.org>)

I analyzed songs from mid-morning and late morning trials in the presence and absence of an observer. Early morning trials were omitted from analysis since localizations on early morning data were prone to errors due to the large number of individuals vocalizing at once and the constant overlap of songs which could not be adequately analyzed by the algorithm. I sampled the first 5 minutes of each 10-minute point count when an observer was present (treatment) and a 5-minute period starting 15 minutes before the arrival of the observer (control). During each 5-minute period I attempted to localize every clear non-overlapping vocalization that was detected on at least 3 of the 8 microphone channels. I identified suitable vocalizations through a combined manual and automated process. Songs were initially selected by listening to the recording while visual inspecting the 8-channel sound spectrograms and marked by drawing a box around part or all of the spectrogram on the channel with the strongest signal.

I used the `location_v2p2_measurement_PRE` localization algorithm (Kathryn Cortopassi, Ithaca, NY. pers comm) in XBAT to estimate the singing position of birds in the samples. This algorithm estimates sound source locations in two steps. First the energy distribution of the sound of interest is characterized based on several measures of central tendency; then using a user defined threshold value a series of cross correlations are run to identify matching sounds from other channels of the recording based on a user defined threshold. Then time delays between key acoustic structures (i.e. peak frequencies) are calculated between the sample vocalization and the retained signals from the cross correlation step. These time delays are then used to triangulate the distance from each microphone in the array to the source of a selected vocalization; multiple permutations are run until the algorithm converges to the final estimated position based on a user defined allowable scatter (Kathryn Cortopassi, pers. comm.). The program provides several diagnostic tools to evaluate the estimated locations including diagrams of acoustic characteristics, alignment of time delays between peak frequencies of selected spectrograms, as well as plots of the most probable location compared to the layout of each receiver in the array.

All localization estimates were checked by hand to eliminate spurious results due to algorithm misalignment. Errors were identified when the estimated location did not match the relative intensities of sound on the spectrogram, e.g., the sound localized to a position that did not match the loudest signal on the spectrogram (Figure 2). In most cases, the presence of an error was

corroborated by examining the alignment of the algorithm generated boxes on the each of the spectrograms and checking for a misalignment of key sound features within each box. At times the algorithm failed to detect sounds on the minimum of 3 channels and failed to return position estimates. In these cases, I attempted to re-run detections using lower correlation thresholds. In some cases errors occurred due to song overlap – the algorithm localized on an overlapping song that was not the one we marked. Errors were corrected using a variety of approaches including re-adjusting sampling boxes to eliminate sections of overlap between songs and eliminating channels causing significant bias in estimated positions. If re-adjusting sampling protocol or running the data at a lower threshold failed to correct obvious errors the songs were dropped from further analysis.

All successfully located vocalizations were annotated with a 4-letter species code and a number identifying each individual singing within each time period. Individuals were distinguished based on differences in song type, song timing (two individuals singing in rapid succession), and/or position estimates (after checking for localization errors).

Once all locations were double checked, log files of location data were exported and merged with event identification files the distance from the center of each array was calculated for all localizations. The analysis was restricted to individuals within a 100m radius of the center of the array for several reasons, including the higher probability of successful localization, improved accuracy of

the localization, and minimizing repeated sampling of individual birds from different arrays.

Statistical analyses were carried out using R 2.9.0 (R Development Core Team, 2009). In all analyses, I compared the observer present (treatment) data against the observer absent (control). I analyzed data from mid-morning ($n = 26$ trials) and late morning ($n = 26$ trials) separately. I used paired t-tests to compare the number of all individuals present during each sampling period. I used non-parametric Kolmogorov-Smirnov tests (KS-test) with a boot-strap correction for ties to compare the distribution of singing positions, measured as distance from the center of the point count station, of all birds in the presence and absence of an observer. I also ran KS-tests using a boot-strap correction for ties to compare the distribution of first vocalizations per 30s interval in the presence and absence of an observer.

RESULTS

Array localization tests

Mean localization error for all whistle trials was $4.19 \pm 0.52\text{m}$ (mean $\pm$ standard error); localization error ranged from $1.80 \pm 0.35\text{m}$ inside the array; $1.04 \pm 0.10\text{m}$ at 50m and $10.65 \pm 3.93\text{m}$ for distances greater than 100m from the center of the array (Figure 3). The algorithm failed to localize several whistles and made several obvious localization errors due to misalignment (i.e. $>50\text{m}$ error) that were easily identified using the same algorithm double checking

protocol used for bird songs (Table 1). Failed detections and obvious localization errors were excluded from the accuracy evaluation.

Number and position of individuals

I sampled 13,058 vocalizations. Once failed localizations were excluded, 10,597 localizations remained of approximately 630 individuals for 26 array locations.

I found no effect of observer on the number of individuals of all species detected in either mid-morning (t-test, $t_{25}=-0.574$, $p=0.571$) or late morning (t-test, $t_{25}= -1.093$, $p=0.285$) point counts (Figure 4). However, the mean ($\pm$ standard error) number of individuals available for detection across sample sites was slightly higher for mid-morning trials (observer absent 19.4 ± 1.0 , observer present 18.9 ± 1.0) than late morning trials (observer absent 17.3 ± 1.1 , observer present 16.2 ± 1.0).

I also found no effect of observer on the distribution of mean singing location of individuals of all species in either mid-morning (KS – test, $D_{1,25}=0.044$, $p=0.387$, $p_{boot}=0.707$; Figure 5a) or late-morning trials (KS – test, $D_{1,25}=0.057$, $p=0.243$, $p_{boot}=0.455$; Figure 5b). Individual birds tended to maintain their singing positions throughout the survey regardless of an observer's presence at the point count station.

There was no evidence of a delay in the onset of singing behaviour in response to an observer for mid-morning (KS – test, $D=0.1$, $p=0.905$, $p_{boot}=1$; Figure 6a) or late morning (KS – test, $D_{1,25}=0.4$, $p=0.202$, $p_{boot}=0.324$; Figure 6b)

trials, as measured by the proportion of individuals detected within each 30s time interval after the start of each point count.

Although there was no evidence of an overall observer effect I tested for species specific effects which may have a net effect of canceling out in the overall study design if different species respond in different ways. I analyzed 12 species which occurred on at least 25% ($n \geq 7$) of our sample sites; two species occurred on all 26 sites song sparrow (*Melospiza melodia*) and yellow warbler (*Dendroica petechia*).

There were fewer brown-headed cowbirds (*Molothrus ater*) when an observer was present (t-test, $t_{17}=-2.650$, $p=0.016$) for late morning trials but there was no difference in the mean number of individuals for any other species in response to an observer (Figure 7a, b). Most species did not change their singing positions in response to an observer (Figure 8a, b). However, song sparrows shifted their singing position approximately 4.39 ± 1.97 m away from the center of a point count station in response to an observer (t-test, $t_{25}=0.145$, $p=0.013$, $p_{boot}=0.02$).

Discussion

Array localization

The GPS synchronized array generated fairly accurate estimate of singing positions of birds based on the whistle trials and the analysis of localizations. Localization error estimates for whistle trials were 1.80 ± 0.35 m inside the array are similar to those reported by Mennill et al. (2006) of 2.82 ± 0.26 m in Costa

Rican rainforest using a roughly circular 8-unit array of linked receivers with a mean distance of $75.2 \pm 2.6\text{m}$ between microphones. However, Mennill et al. (2006) do not report localization errors outside the array. Uniquely, this study tested localization error over a range of distances up to 75m beyond the edge of the microphone layout.

The frequency of missed detections and estimates of localization error for whistle trials increased as the distance from the center of the array increased due to the layout of the microphone array and the acoustic structure of the whistles used in the trials. The algorithm missed few whistles and obvious errors were minimal within 50m of the center of the array because of the greater number of microphones near the center. As the distance from the array increased for whistle trials ambient wind noise drowned out some of the test whistles resulting in an increased frequency of missed detections. At times, the whistles were too quiet to be detected on a minimum of 3 microphones in the array and were dropped from analysis by the algorithm. Test whistles were structurally very simple; each whistle began with a short duration increase in frequency followed by a drawn out down-slur. The localization algorithm had few reference points to compare signals between channels when calculating locations of these whistles. I found that the algorithm did a better job of correctly aligning bird songs that generally had greater variation in frequency and a more complex acoustic structure than the test whistles.

There were three types of algorithm generated localization errors when dealing with real songs: between song misalignment, within song misalignment

and missed detections. Between song misalignments occurred when two strong signals from different individuals either overlapped or occurred in rapid succession on two or more channels. Within song alignment occurred when the algorithm mismatched peak characteristics within a signal, for example by mismatching peak 1 with peak 2 between channels. These errors typically occurred with trilled songs, such as chipping sparrow songs (*Spizella passerinea*), where there is relatively little variation in song structure. Both types of misalignment errors were identified when the estimated position generated by the algorithm did not match the strongest signal on the spectrograms but was rather the average position between overlapped signals. Missed detections occurred most frequently with high frequency short duration songs or low volume distant test whistles. In these cases the algorithm failed to detect the sound on the minimum number of channels necessary for localization even though the signal was visible on multiple channels on the spectrograms.

Further ground truthing and array calibration experiments for unlinked, GPS synchronized terrestrial microphone arrays would be an asset to future research. My tests of algorithm localization error were admittedly simplified but adequate for the study system. Future tests must involve multiple playbacks from various locations inside and outside the array using simple tones, frequency modulated artificial playback sounds, and samples of recorded bird songs, to test the accuracy of both the microphones and the localization algorithms. The field site was primarily open canopy old field habitat and there was minimal signal interference due to wind effects: most of the songs analyzed had strong clear

signals with minimal degradation. However, other field sites such as forested sites will likely experience habitat based signal interference such as refraction or reflection of signals due to tree trunks or leaves, additional wind interference due to rustling leaves etc (Spiesberger 1999). Further tests of the accuracy of any localization algorithm should be undertaken in various habitats before remote acoustic monitoring of songbirds using complex vocalizations can become widespread.

Several key factors should be considered when designing experiments using GPS synchronized microphone arrays. First the proportion of canopy cover and satellite availability are key when determining the type of array (linked or GPS synched) used. The study site was regenerating old field habitat at relatively low latitude (N 43.93° W 76.88°) with minimal canopy cover and a good view of the sky; each GPS receiver was able to locate and maintain adequate satellite coverage throughout the experiment. In June 2006, I set up a similar array for a pilot study in boreal forest near Cochrane Ontario, (N 45.15° W 80.50°) where it was significantly more difficult to establish and maintain adequate satellite coverage due, in part, to the increased canopy cover (M. Campbell, unpublished data). The deployment of the ARU array was easier in boreal forest than a linked array but maintaining adequate satellite coverage to ensure alignment of the recordings was at times problematic. Additionally, researchers should consider interactions with non-target wildlife. I had few problems in the old field habitat beyond rodent gnawed GPS cables; however, wildlife in boreal forest habitats caused significantly more problems. Several ARUs were “investigated” by black

bears (*Ursus americanus*) resulting in the destruction of several key parts causing 5 of 8 ARUs to be dropped from the pilot study data set.

The main advantage of a GPS synchronized array over a cable-linked system is the relatively simple task of deploying the array in various habitats. To date most terrestrial acoustic localization studies of birds have used linked receiver arrays where microphones are connected to a central computerized recording station using hundreds of meters of cable (e.g. Mennill and Vehrencamp 2008, Foote et al. 2008, Mennill et al. 2006). I was able to deploy the array alone (i.e. with only one person working) in 2 to 5 hours depending on habitat (M. Campbell, pers. obs.). Some linked arrays may take up to 6 people an entire day to deploy (16-unit array, Jennifer Foote, pers. comm.). I was able to move the array daily in both regenerating old field habitat and boreal forest pilot sites. Deployment of the array in boreal forest took longer (i.e. 4 + hours) using a single observer than in old field habitat (2 – 3 hours) but I was still able to move the array daily over rough terrain in regenerating clear cuts. Power supply vs. mobility is another key consideration when designing any microphone array study. For instance, in the boreal pilot study I used battery packs with a long lifespan, which were heavy and difficult to move, while at the Prince Edward Point, I was able to use lighter rechargeable lead-acid batteries since there was access to a power source for recharging.

Further algorithm development will be critical before microphone array experiments can become widespread in a songbird monitoring context. I recorded a substantial amount of song data for this study of which only a small

fraction was analyzed. All of my analysis relied heavily on manually scanning recordings, identifying species and individuals and double checking localization estimates for major errors. I was unable to use early morning and dawn chorus data in the analysis since the algorithm was unable to generate good position estimates with multiple overlapping songs. One possible method to improve future localization algorithms would be to include a measure of sound intensity (or volume) in the calculations of singing position. The localization algorithm I used did not take into account the intensity of the sample sounds on each channel in the position estimates but treated sounds of different volume equally when estimating singing positions. This resulted in several obvious localization errors where the algorithm correctly aligned the songs (by time) on each channel but did not generate a position estimate closest to the most intense (i.e. loudest) sound picked up by the array. This may be a function of adapting a marine array system (where sound travels more easily) to a terrestrial habitat (where sound tends to degrade more quickly). Algorithm developments that take into account time and additional information on the differences in sound intensity (i.e. volume) differences could improve the accuracy of localization estimates and the amount of time necessary to analyze this type of data in the future.

Number and position of individuals in relation to observers

There was no evidence of any effect of an observer on the number, distribution or behaviour of birds during a point count compared to control observer absent conditions. I detected slightly fewer individuals when an

observer was present during a point count but the difference was non-significant. Similarly, there was no change in the mean singing position of individuals in response to an observer. The mean singing positions of birds in the study site was evenly distributed between 20-100m from the center of the array; there were no birds detected within 10m of the center of our array in any trials. Birds may have been avoiding the ARU set up in the center of each array but the effect was constant for all trials. This is because as the distance from the center of the array increases the area available for birds to occupy increases as a factor of the radius from the center (i.e. by πr^2). There was no evidence of a delay in the onset of singing behaviour in response to an observer. Nor was there a difference in the proportion of novel detections per 30s interval after the start of the count in response to an observer.

There was relatively little evidence of species specific responses to an observer performing a point count. I detected significantly fewer individual brown-headed cowbirds for late morning trials when an observer was present (Figure 7b). But there was no change in the number of individuals available for detection for any other common species. The singing positions of most species were highly variable but did not change in response to an observer. One species, song sparrow, shifted its mean singing position away from the observer during a point count. While the change in position was statistically significant it may not be biologically significant; song sparrows moved 4.39 ± 1.97 m away from an observer which is very close to the mean localization error estimate of 4.19 ± 0.52 m.

Although Gutzwiller et al. (1998) found evidence of observer avoidance in ground associated forest breeding birds in Wyoming I failed to find similar trends in this study site. Regenerating old field habitat should allow birds ample warning to spot an observer, determine if there is a threat and adjust their singing positions. None of the common ground associated species including chipping sparrows (n = 20 arrays) or field sparrows (*Spizella pusilla*, n = 22 arrays) changed their position in response to an observer. Ground associated forest breeding birds may have obstructed sightlines due to undergrowth making an early predator avoidance response an adaptive advantage; a similar response may not have the same advantage in open habitats.

The goal of the study was to track major position movements of singing birds in response to an observer but I found that position estimates for certain types of species tended to be more prone to errors which may have affected my results. Species that tended to sing repeatedly from the same perch (e.g. song sparrows) or sang while foraging over a small area (e.g. yellow warbler) produced very consistent position estimates. However, species that vocalized over a wider area or during flight produced greater variation and uncertainty in position estimates. Specifically, localization estimates for bobolink (*Dolichonyx oryzivorous*) during display flights were highly variable, I sub-sampled within song flights to generate mean position estimates but these estimates are probably oversimplified. Species that tended to change their perches frequently throughout counts like brown-headed cowbirds and American goldfinch (*Carduelis tristis*) tended to produce more variable position estimates and could

have been double counted on the spectrograms. I also found that species with high frequency short duration songs such as brown-headed cowbirds, eastern kingbirds (*Tyrannus tyrannus*) and cedar waxwings (*Bombycilla cedrorum*) tended to be missed more frequently by the algorithm than other species. In order to counteract this variation in position estimation I chose to base the analysis on mean position than raw location data. While this may have masked some movement in response to the observer it is likely not enough to alter the number of individuals detected during the point count.

I sampled time to first successful localization rather than time to first vocalization, which may be a conservative measure of the effect of an observer on singing behaviour during a point count. The time to first localization method may have eliminated several instances where singing individuals would have been detected by a field observer during a regular point count but failed to meet the threshold criteria necessary to generate a localization estimate in the experimental protocol. There were several cases where the first vocalizations of two or more individuals overlapped and were unsuitable for localization estimates but were detectable in the context of a regular point count. Since most birds sang on multiple occasions within the 30s analysis intervals there should be little effect of using first localization rather than first vocalization in the evaluation of behavioral change in response to an observer.

Some authors (Ellingson and Lukas 2003) have suggested delaying the start of a point count for a few minutes to allow the birds to adjust to the presence of an observer this study indicates that this is probably not necessary. Therefore,

I would suggest that when performing a point count observers approach the station, take 30s – 1minute to prepare for the count (e.g. setting up notebooks, taking GPS co-ordinates etc) then begin counting right away. Further, Bart et al's (1995) initial suggestion that any birds that flush when an observer arrives or departs from a count station be counted as present seems to be an adequate adjustment. A "no-waiting" protocol may actually increase the number of count station sampled in future studies by reducing the time an observer has to wait to begin counts.

Although I found no evidence of observer effects on the number of birds present or their behaviour during a point count there are other factors that may cause biases in point counts. Since this study focused on a single relatively open habitat type I cannot address issues of variation detectability due to habitat effects. Habitat effects have been shown to be an important source of bias in count surveys due to overall landscape structure (Betts et al 2007), proximity of roads (Harris and Haskell 2007, Hanowski and Niemi 1995) and vegetation structure (Gonzalo-Turpin et al 2008, Sallabanks et al 2006). Further bias may be introduced into count data due to birds that do not sing during the count. Since this study relied entirely on vocal individuals I could not address this issue however several models have been proposed that account for non-detected (i.e. non-vocal) individuals in counts (Royle and Link 2006, Royle et al 2005).

Microphone arrays and acoustic localization are becoming powerful tools in behavioural studies of songbirds (e.g. Mennill and Vehrencamp 2008, Foote et al. 2008) and could play a significant role in population monitoring, especially if

combined with distance methods for density estimation studies. The results of my experiment show that the general assumption of no change in position of birds in response to an observer is valid. However, the inability of any observer to accurately gauge distances for auditory detections (Alldredge et al. 2007) calls into question the usefulness of distance sampling methods in habitats where most species are detected aurally. My data indicates that distance estimates derived from microphone array localizations are a significant improvement on distance estimates made by observers in the field (see Alldredge et al 2006). Since, accurate estimation is critical for distance based density estimation (Buckland et al 2001, 2004) the combination of distance sampling methods and acoustic localization in multi-species systems could be a powerful tool for large scale landbird monitoring and conservation programs in the future.

Figure 1. Diagram of layout of adjacent microphone arrays. Each array has a central microphone unit (ARU) surrounded by 7 ARUs distributed along a circle of 50m radius. The distance between the center ARUs of adjacent arrays was a minimum of 200m.

Figure 2. Example of alignment error in localization algorithm. The green box on channel 4 is a selection of a common yellowthroat song. The blue box on channel 5 is correctly aligned, but those on channels 2, 3 and 6 are incorrectly aligned; on channels 3 and 6, the algorithm proposed an alternative alignment (red lines) that was, in fact, correct, but it failed to find the correct alignment on channel 2. The alignment error was corrected by re-adjusting the green sampling box closer to the start of the song segment. Note the strongly overlapping partial song sparrow song (conspicuous in the sonograms from channels 2 and 3) and cricket (conspicuous on channels 5 and 6).

Figure 3. Distribution of algorithm localization error for whistle trials from known locations in Prince Edward Point National Wildlife Area, June 2007. Localization error is calculated as the distance in meters between algorithm generated position and known whistle location measured using a sub-metered GPS unit. Obvious failures of the localization algorithm were excluded.

Figure 4. Mean number of individuals detected with 95% confidence interval bars for observer absent (control) trials and observer present (treatment) trials across 26 array locations in Prince Edward Point National Wildlife Area, June – July 2007.

Figure 5a, b. QQplot of mean singing positions of all individuals for observer present (treatment) vs. observer absent (control) trials. Singing position is measured as distance from center of array (m). Null effect line (slope = 1) plots no-difference in mean distance from center in response to an observer. Pooled data from 26 array locations analyzed separately for mid-morning and late morning trials in Prince Edward Point National Wildlife Area, June – July 2007.

Figure 6a, b. Proportion of novel first detections of individual birds per 30 second interval after the start of paired mid-morning observer absent (dark grey) and observer present (light grey) trials. Pooled data from 26 array locations analyzed separately for mid-morning and late morning trials in Prince Edward Point National Wildlife Area, June – July 2007.

Figure 7a, b. Mean number of individuals detected with 95% confidence for species occurring on $\geq 25\%$ ($n \geq 7$) of all array locations for observer absent (dark grey) trials and observer present (light grey) trials. Numbers below species labels indicate number of arrays where species were detected. Pooled data from 26 array locations analyzed separately for mid-morning and late morning trials in Prince Edward Point National Wildlife Area, June – July 2007; 4 letter species codes correspond to species names in Table 2.

Figure 8a, b. Plot of mean change in distance from center of array (m) for species occurring on $\geq 25\%$ ($n \geq 7$) of array locations. Solid bars show mean difference and whiskers show maximum difference; positive values represent observer absent (control) trials, negative values represent observer present (treatment) trials. Numbers below species labels show number of arrays where a species was detected. Pooled data from 26 array locations analyzed separately for mid-morning and late morning trials in Prince Edward Point National Wildlife Area, June – July 2007; 4 letter species codes correspond to species names in Table 2.

Table 1. Results of localization algorithm tests of whistle trials from known locations in several array layouts from Prince Edward Point National Wildlife Area, June 2007. Chart shows total number of available whistles (N), number of whistles algorithm failed to detect and localize (N miss), number of obvious localization errors (N wrong) and number of the original whistles that were available for localization analysis (N kept) and included in Figure 1.

Trial	Location	N	N miss	N wrong	N kept
1	Center	27	0	1	26
2	50m edge	35	0	5	30
3	<100m outside	30	0	12	18
4	>100m outside	35	7	9	19
5	>125m outside	22	9	6	7

General Discussion

The goal of my M.Sc. project was to use recent technological advances in sound recording and analysis to investigate various aspects of bird survey methods using point counts. In this chapter I summarize the results of my research and present several recommendations for using recordings in avian monitoring programs. These recommendations include methods to use stereo-microphones for recording boreal BBS routes, suggestions for dealing with detection errors made by observers during point counts, and further applications for microphone arrays and acoustic localization studies.

Summary of major findings

First I evaluated variation in the number of species detected, number of missed species and the number of detection errors amongst skilled observers listening to recordings of BBS routes. There were significant differences in the overall number of species detected, the number that could be confirmed on the recordings and the number of detection errors (both false positive detections and misidentifications) amongst observers. I found that field observers detected slightly fewer species than were confirmed on the recordings but not all the species that were detected in the field could be confirmed on the recordings. I then ran a series of multiple observer and site occupancy models to test how detection errors may affect estimates of species richness and site occupancy. Uncorrected detection errors resulted in overestimates of species richness of between 1.0 and 8.7 additional species depending on the model compared to the

true number of species detected. Species specific occupancy models were no better. Of the 16 species tested between 69% and 100% of the species had occupancy rates overestimated. Current occupancy based models do not adequately deal with detection errors at the species level and produce biased estimates. Therefore until methods that better account for identification errors are designed occupancy models should be used with caution especially in multi-species systems. Simple index methods, such as those used in BBS analysis, are a more conservative method of analyzing count data which are less biased by identification errors.

The second part of my project focused on evaluating the effect of an observer's presence on the detectability and behaviour of birds during a point count. I used an array of GPS synchronized microphones and localization algorithm to compare estimates of the number of individuals available, the mean singing position of individual birds and the timing of onset of singing behaviour when an observer was present and absent from the count station. Part of the experiment involved testing the accuracy of the position estimates generated by the localization algorithm; mean localization error was $4.19 \pm 0.52\text{m}$. This is a significant improvement on distance estimates made by observers in the field, auditory position estimates can range from 59.7m to 202.2m away from their true location (Alldredge et al 2007d). There was no difference in the number of individuals detected, their mean singing position or the onset of singing behaviour in response to an observer. My findings of no effect of observer on the singing behaviour, mean singing position or number of individuals available for

detection during a count validate a key assumption of point count methodology that birds do not change their singing position in response to an observer (Ralph et al 1995, Buckland et al 2001, 2004).

Using stereo-microphones to record boreal BBS routes

One of the goals of the M.Sc. projects was to evaluate the potential usefulness of using high quality stereo microphone to expand the number of BBS routes that could be run in parts of Canada's boreal forests where it is normally difficult to maintain consistent observer coverage. One approach would be to send untrained observers to record the routes without attempting to conduct simultaneous point counts, then have a skilled observer analyze the recordings at a later time. This type of approach fails to deal with non-vocal species that may be overlooked during the point count. However, the vast majority of species that were detected in the field but not confirmed on the recordings occurred on at least one other stop in the experimental test data. If the detections are analyzed using a typical trend analysis protocol, where population estimates are based on data across all stops on a route, the vast majority of the species should be accounted for in the final population trend estimates.

A second approach would be to send moderately skilled observers into the field with stereo microphones to record BBS routes while simultaneously performing a regular point count. These types of counts could use a modified protocol; observers in the field would be instructed to focus on visual detections, counting individuals and identifying species that may be confused on the

recordings (i.e. differentiating chipping sparrows from pine warblers). This protocol is unlike a traditional point count where observers attempt to record all the species detected on the route. The modified point count detections lists and recordings could then be given to second skilled observer after the field season to analyze, verify the field observer's detections and add any species that were missed in the field.

One major challenge when analyzing recorded BBS routes is dealing with ambient background noise related to habitat, wind effects or road noise. Simons et al (2007) showed that the proportion of individuals detected by observers decreased as ambient noise increased. Since the BBS is a road-side survey the most likely source of ambient noise will be passing vehicles. Road noise was not a major issue in the study design since I selected recordings from a subset of 20 stops per route with relatively little background noise or passing vehicles. On all of the routes there were several stops where the noise of passing vehicles completely saturated the recordings. On these stops all that was audible for several seconds as a vehicle passed the microphone was engine noise. Although attempts were made to edit out passing vehicles on some stops this caused small gaps in the recordings where the loudest part of the vehicle interference was removed and the noise of the approaching vehicle could never be completely removed. Other sources of ambient background noise include wind effects which are amplified on the microphones compared to field conditions. One set of routes from Alberta that were edited for testing but never used were almost impossible to listen to because of the near constant wind interference.

Several types of filters have been suggested to reduce the amount of ambient noise on recordings. The most promising filters work by setting the mean background noise equal to zero and extracting sounds that exceed the background threshold (see Baker and Logue 2007). However, since passing vehicles completely saturate the recording filters would not be able to remove vehicle noise at its peak and they could remove parts of the singing bird chorus when the engine noises exceed the background threshold but the bird song does not. However, editing sound files to remove or filter passing vehicles or other ambient noise can be quite labour intensive and time consuming.

One approach for dealing with road noise could be to have the observer in the field record the number of passing vehicles during the 3-minute count. A maximum number of vehicles threshold could be set so that any stop where the number of passing vehicles that exceeds the threshold number would be dropped from analysis. Alternatively the recording period could be extended beyond the standard 3-minute length to cover the periods when vehicles interrupted the recordings. Both of these approaches would solve the problem of trying to analyze data from poor recordings, but would result in some routes that may not adhere to BBS protocol which requires each point count to be 3-minutes long and each route to consist of 50 stops. There is a trade off between strictly following the experimental protocol and having a greater number of routes from otherwise under sampled regions available for the BBS. One area of future research could focus on determining what proportion of recorded BBS routes would be adequate to replicate a more traditional approach to analyzing BBS

point counts. This may include creating adjusted methods for analysing BBS data collected using recordings that corrects for any stops that need to be dropped or extended due to vehicle noise.

Further research on microphone quality is needed to determine if microphone quality affects the range of detection for any point count study. For example if higher quality microphones pick up sounds over a greater radius than a lower quality microphone the effective population sampled by each microphone would not be equal and direct comparisons of routes recording using different microphones would be invalid. While I was able to detect the test whistles for the array calibration trials approximately 150m away from the closest microphone I do not have any estimates of the effective range of the stereo-microphones used for recording the BBS routes. Data on the effective range of detection of stereo-microphones may clarify some of the detection differences between species detected by observers in the field and those listening to recordings.

Microphone quality may also affect what species are detected on recordings. Songs may be clearer on recordings from higher quality microphones compared to lower quality microphones resulting in a difference in the total number of species available for detection based only on the microphone used. Ideally, observer variation in species detection rates should be tested using data recorded from the same population using several microphones of variable recording quality. Initially I had included data from a single BBS route recorded using a high and low quality microphone to compare microphone effects on species detected by observers listening to recordings. However the microphone

differences added a confounding factor in the analysis of modified point count models and were dropped from analysis. Experiments testing microphone effects are essential if recording BBS routes in under sampled areas becomes a standard monitoring protocol.

Dealing with detection errors made by observers

The analysis of recordings is also problematic because of missed detections and detection errors made by observers when listening to recordings. I have shown that the effects of errors are augmented when using multiple observers in a capture-recapture framework for data analysis. Occupancy models do not adequately correct for overlooked birds because errors are common amongst observers. The approach I used of generating master lists by combining the species lists of several observers, then verifying each species on the recordings is extremely time consuming and impractical for operational purposes. One approach may be to ask several observers to listen to the recordings once, amalgamate the lists then ask the observers to double check their own detections against the group list. While this will likely not solve the problem of conflicting species identifications at least it would remove several of the obvious detection mistakes, since when asked to double check the recordings most observers in the study failed to confirm several of the species they had initially detected. Further research could include investigating an alternative approach to multiple observer analysis using intensive independent double observer analysis of recordings. This may reduce the number of errors

but could potentially increase the frequency of conflicting species identifications amongst observers.

Another approach may be to correct for detection errors using a double sampling protocol. Double sampling compares the ratio of birds detected using rapid sampling methods to the total number of birds detected using intensive surveys on a sub-sample of the rapid survey data to estimate true density (Bart and Earnst 2002). This approach could be used to correct for errors in multi-observer recording surveys, observers could be asked to listen to all recordings once without double checking, then asked to re-listen to a subset of the stops intensively to look for detection errors etc. The ratio of the difference between uncorrected and intensively scanned recordings could then be used to estimate the number of detection errors and number of missed species made by any observer when listening to the recordings. This could eliminate the labour intensive process where all recordings need to be intensively scanned to create master species lists. There may also be an added benefit that observers who correct their own data using double sampling methods would improve their detection skills by learning what errors they tend to make as well as providing a range of observer specific detection probability estimates for correct, missed and error type detections.

A third approach to dealing with observer error may be to include probability estimates for detection error rates at the observer level rather than at the species level. Most proponents of modified point count methods argue that estimated species specific (Diefenbach et al 2007) or grouped species (Alldredge

et al 2007e) detection probabilities need to be included in the analysis of point count statistics. Rather than estimate detection probabilities for species which are thought to be heterogeneous and highly dependent on vegetation structure (Gonzalo-Turpin et al 2008) detection probability based on observer variation could be included. Detection probability could be thought of as made up of the probability that the suite of species reported by any one observer will include a fraction of correct detections (p_{correct}), missed detections (p_{miss}) and some detection errors (p_{error}). These probabilities could then be modelled based on regional or habitat specific observer variance experiments. The resulting probability estimates could then be included in the count equation:

$$N = Cp + \epsilon$$

Where N is the true population estimate, C is the number of individuals counted, p is the probability of detection (Nichols et al 2002) and ϵ is some residual error. If we then split the probability of detection into sub-categories based on factors such as observer variation, habitat etc and re-write the equation:

$$N = C * (p_{\text{obs}} + p_{\text{habitat}}) + \epsilon$$

Then define the observer based detection probability term as follows:

$$p_{\text{obs}} = p_{\text{correct}} + p_{\text{miss}} + p_{\text{error}}$$

And the final equation is of the form:

$$N = C * (p_{\text{correct}} + p_{\text{miss}} + p_{\text{error}} + p_{\text{habitat}}) + \epsilon$$

This type of approach would need to consider various habitat effects, vegetation differences amongst sites and would likely only be useful within relatively similar habitats or eco-region types and songbird communities. However, at this time I

am unsure if this approach would be practical or even mathematically appropriate, it may be an improvement on the occupancy based framework or it may be a naive approach/theory.

Applications for microphone arrays and acoustic localization

The relative portability and ease of deployment of GPS linked arrays and the accuracy of the localization algorithms make these types of systems ideal for population monitoring in remote areas. Provided an adequate power supply is available, the between unit synchronization remains consistent, and no bears take an interest in the array then long term monitoring studies in remote areas could be undertaken at a relatively low cost compared to the amount of data collected. If distance sampling theory is combined with algorithm generated position estimates, acoustic localization systems could be used to remotely monitor population densities throughout the breeding season. One approach could be to use the array to estimate the density and layout of breeding territories in areas where intensive nesting and territory mapping studies are impractical, then use the population density when developing conservation strategies for boreal songbird populations. Acoustic arrays have been suggested as a method to study two species of endangered songbirds (golden-cheeked warbler, *Dendroica chrysoparia* and black-capped vireo *Vireo atricapilla*) in remote areas of Texas (Fristrup 2004).

Further studies testing the accuracy of positions estimates generated using localization algorithms and GPS synchronized arrays should be undertaken.

Additional testing is especially critical for arrays that are deployed in areas where habitat features may alter the accuracy of location estimates (i.e. due to sounds reflecting off of leaves). My localization tests were relatively simple, they focused on a random subset of whistle trial locations using acoustically simple whistles. The Bioacoustics Research Program at the Cornell Laboratory of Ornithology began a systematic 13 unit ARU calibration experiment testing localization error up to 100m from the center of an array using playback of pure tones and samples of bird song in 2007 but abandoned the trial due to alignment issues in the data (C. Tessaglia-Hymes, pers. comm.). Calibration experiments should be undertaken since they are critical to any future acoustic localization studies using both GPS synchronized and linked arrays. The role of wind speed and direction and habitat effects including sound reflecting off leaves will likely increase the localization error for any array and should be evaluated in these calibration trials.

Finally, advances in algorithm research in voice recognition and signal identification may be key to using recordings in future monitoring programs. Both studies generated an enormous amount of acoustic data for analysis. I collected over 200 hours (1.2 weeks) of bird song recordings from Prince Edward Point in a month but only analysed 9.3 hours (560 minutes). This is a consequence of having to listen to all recordings, identify and sample all singing individuals and double check all position estimates by hand. This was an insanely labour intensive process. Recently a protocol for the use of sound detection algorithm for XBAT has been described and successfully tested for a series of 3 species of woodpecker calls (Swiston and Mennill, 2009). Initially, I had attempted to use

the same system for my songbird data but after testing the detection algorithm I found that it made more errors than correct detections. These errors were due to the overlap of songs of several species singing at the same time, variation between songs of the same individual and the acoustic complexity of the songs sampled. The frequency of identification errors generated using the detection algorithm meant that it was faster to perform all acoustic analysis by hand. I did not keep close track of how many hours of labour this required, however it was probably close to 1 hour per 5 minute count to identify and annotate all individuals singing then run the localization, double check and correct the localization errors. If recordings of bird songs are to be used as a tool for monitoring, there must either be further advances in algorithm development for species identification or any project will require several skilled individuals to perform the acoustic analysis.

Overall Conclusions

The focus of my M.Sc. research was using novel technology to evaluate a number of assumptions of current field methodology in avian population monitoring programs. This research has lead to several new directions where technology could be applied to improve monitoring programs. Novel applications of technology could include using recordings to expand the sampling area of traditional index based monitoring programs into areas that are often difficult to sample or maintain adequate coverage between years (i.e. boreal BBS routes). Population density monitoring could be greatly improved using acoustic

localization rather than traditional distance estimation methods. Acoustic arrays could also be used to map the breeding territories and monitor local populations in sensitive areas where repeated human intrusion is difficult or dangerous (i.e. endangered species monitoring in remote or restricted areas).

To facilitate the implementation of these types of technologies in monitoring programs future research should focus on the following priorities. First, evaluations of the effects of microphone quality on the number and types of species available for detection using recordings are critical. Microphone effects will determine the design of any future monitoring programs as well as how to compare data collected using microphones to traditional point counts. Second advances in algorithm design with respect to song recognition in multi-species systems need to be made to make analysis of recording data more practical. Third analytical models that take into account the frequency of detection errors made by observers would improve estimates of species specific detection probabilities and overall population abundance or species richness estimates. A final consideration should be what to do with the acoustic data once it has been collected and analyzed. One option would be to set up an archive of the recordings and the matching lists of confirmed species. This archival data could be used as both as training material for observer to access as well as a source of data for further observer effect studies.

Literature Cited

- Allredge, M.W., T.R. Simons and K.H. Pollock. 2007a. Factors affecting aural detections of songbirds. *Ecological Applications* 17(3): 948-955.
- Allredge, M.W., K.H. Pollock, T.R. Simons, J.A. Collazo, and S.A. Shriner. 2007b. Time-of-detection method for estimating abundance from point-count surveys. *Auk* 124 (2): 653-664.
- Allredge, M.W., T.R. Simons, K.H. Pollock and K. Pacifici. 2007c. A field evaluation of the time-of-detection method to estimate population size and duration for aural avian point counts. *Avian Conservation and Ecology - Écologie et conservation des oiseaux* 2 (2): 13.
- Allredge, M.W., T.R. Simons, and K.H. Pollock. 2007d. A field evaluation of distance measurement error in auditory avian point count surveys. *Journal of Wildlife Management* 71 (8): 2759-2766.
- Allredge, M.W., K.H. Pollock, T.R. Simons and S.A. Shriner. 2007e. Multiple-species analysis of point count data: a more parsimonious modelling framework. *Journal of Applied Ecology* 44, 281-290.
- Allredge, M.W., K.H. Pollock and T.R. Simons. 2006. Estimating detection probabilities from multiple-observer point counts. *Auk* 123 (4): 1172-1182.
- Baker, M.C. and D.M. Logue. 2007. A comparison of three noise reduction procedures applied to bird vocal signals. *Journal of Field Ornithology*. 78 (3): 240-253.
- Barker, R.J., J.R. Sauer., and W.A. Link. 1993. Optimal allocation of point-count sampling effort. *Auk*, 110 (4): 752-758.
- Bart, J., K.P. Burnham, E.H. Dunn, C.M. Francis, and C.J. Ralph. 2004. Goals and strategies for estimating trends in landbird abundance. *Journal of Wildlife Management* 68: 611-626.
- Bart, J., B. Collins, and R.I.G. Morrison. 2003. Estimating population trends with a linear model. *Condor* 105: 367-372.
- Bart, J., and S. Earnst. 2002. Double sampling to estimate density and population trends in birds. *Auk*. 119 (1): 36-45.
- Bart, J. 1985. Causes of recording errors in singing bird surveys. *Wilson Bulletin*. 97(2): 161-172.

- Bart, J. and J.D. Schoultz. 1984. Reliability of singing bird surveys: changes in observer efficiency with avian density. *Auk* 101: 307-318.
- Betts, M.G., D. Mitchell, A.W. Diamond, and J. Bety. 2007. Uneven rates of landscape change as a source of bias in roadside wildlife surveys. *Journal of Wildlife Management* 71 (7): 2266-2273.
- Buckland, S.T. 2006. Point-transect surveys for songbirds: robust methodologies. *Auk* 123 (2): 345-357.
- Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas. 2001. Introduction to distance sampling. Oxford University Press, Oxford.
- Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas, EDS. 2004. Advanced distance sampling. Oxford University Press, Oxford.
- Bye, S.L., R. J. Robel, K.E. Kemp. 2001. Effect of human presence on the vocalizations of grassland birds in Kansas. *Prairie Naturalist*. 33 (4): 249-256.
- Cadman, M.D., D.A. Sutherland, G.G. Beck, D. Lepage and A.R. Couturier (eds.) 2007. Atlas of the Breeding Birds of Ontario, 2001-2005. Bird Studies Canada, Environment Canada, Ontario Field Ornithologists, Ontario Ministry of Natural Resources and Ontario Nature, Toronto, xxii + 706 pp.
- Carter, M.F., W.C. Hunter, D.N. Pashley and K.V. Rosenberg. 2000. Setting conservation priorities for landbirds in the United States: The Partners in Flight Approach. *Auk* 117(2): 541-548.
- Clark, C.W., and P.J. Clapham. 2004. Acoustic monitoring on a humpback whale (*Megaptera novaeangliae*) feeding ground shows continual singing into late spring. *Proceedings of the Royal Society of London B*. 271, 1051-1057.
- Clark, C.W. and W.T. Ellison. 2000. Calibration and comparison of the acoustic location methods used during the spring migration of the bowhead whale, *Balaena mysticetus*, off Pt. Barrow, Alaska, 1984-1993. *Journal of the Acoustical Society of America*. 107 (6): 3509-
- Cortopassi K.A., 2006. Automated and Robust Measurement of Signal Features. Bioacoustic Research Program, Cornell Laboratory of Ornithology. Ithaca NY. <http://www.birds.cornell.edu/brp/research/algorithm>

- Cortopassi, K.A., 2004. Energy Distribution Measurement. Bioacoustic Research Program, Cornell Laboratory of Ornithology. Ithaca NY.
<http://www.birds.cornell.edu/brp/research/algorithm>
- Cyr, A. 1981. Limitation and variability in hearing ability in censusing birds. Pp. 327-333 in *Estimating numbers of terrestrial birds* (C.J. Ralph and J.M. Scott, eds.). *Studies in Avian Biology*. 6.
- Diefenbach, D.R., M.R. Marshall, J.A. Mattice and D.W. Brauning. 2007. Incorporating availability for detection in estimates of bird abundance. *Auk* 124 (1): 96-106.
- Donovan, T.M., C.J. Beardmore, D.N. Bonter, J.D. Brawn, R.J. Cooper, J.A. Fitzgerald, R. Ford, S.A. Gauthreaux, T.L. George, W.C. Hunter, T. E. Martin, J. Price, K.V. Rosenberg, P.D. Vickery, and T.B. Wigley. 2002. Priority research needs for the conservation of Neotropical migrant landbirds. *Journal of Field Ornithology* 73 (4): 329-450.
- Downes, C.M., and B.T. Collins. 2008. Canadian Bird Trends Web site Version 2.2. Canadian Wildlife Service, Environment Canada, Gatineau, Quebec, K1A 0H3
- Dunn, E.H. 2002. Using decline in bird populations to identify needs for conservation action. *Conservation Biology*, 16 (6), 1632 -1637.
- Dunn, E.H., D.J.T. Huseell, and D.A. Welsh 1999. Priority-setting tool applied to Canada's landbirds based on concern and responsibility for species. *Conservation Biology* 13 (6): 1404-1415.
- Farnsworth, G.L., K.H. Pollock, J.D. Nichols, T.R. Simons, J.E. Hines and J.R. Sauer. A removal model for estimating detection probabilities from point-count surveys. *Auk* 119 (2): 414-425.
- Fletcher, R.J. and R.L. Hutto. 2006. Estimating detection probabilities of river birds using double surveys. *Auk*. 123 (3): 695-707.
- Foot, J.R., L.P. Fitzsimmons, D.J. Mennill, and L.M. Ratcliffe. 2008. Tied to the nest: male black-capped chickadees decrease dawn chorus movement behaviour when their mate is fertile. *Animal Behaviour*. 76 (4): 1227-1233.
- Forcey, G.M., J.T. Anderson, F.K. Ammer, and R.C. Whitmore. 2006. Comparison of two double-observer point-count approaches for estimating breeding bird abundance. *Journal of Wildlife Management* 70 (6): 1674-1681.

- Fristrup, K.M. 2004. Acoustic monitoring of threatened and endangered species in inaccessible areas. Strategic Environmental Research and Development Program (SERDP) number CS – 1185. <http://www.birds.cornell.edu/brp/pdf-documents/InaccessbileSERDP.pdf>
- Gonzalo-Turpin, H., C. Sirami, L. Brotons, L. Gonzalo and J.L. Martin. 2008. Teasing out biological effects and sampling artifacts when using occupancy rate in monitoring programs. *Journal of Field Ornithology* 79 (2): 159-169.
- Harris, J.B.C, and D.G. Haskell. 2007. Land cover sampling biases associated with roadside bird surveys. *Avian conservation and ecology - Écologie et conservation des oiseaux*. 2 (2) 12
- Haselmayer, J. and J.S. Quinn. 2000. A comparison of point counts and sound recording as bird survey methods in Amazonian southeast Peru. *Condor* 102: 887-893.
- Hobson, K.A., R.S. Rempel, H. Greenwood, B. Turnbull and S.L. Van Wilgenburg. 2002. *Wildlife Society Bulletin* 30 (3): 709-720.
- Howell, C.A., P.A. Porneluzi, R.L. Clawson, and J. Faaborg. 2004. Breeding density affects point-count accuracy in Missouri forest birds. *Journal of Field Ornithology*. 75 (2): 123-133.
- Johnson, D.H. 2008. In Defence of indices: the case of bird surveys. *Journal of Wildlife Management* 72 (4): 857-868.
- Kendal, W.L., B.G. Peterjohn, and J.R. Sauer. 1996. First-time observer effects in the North American breeding bird survey. *Auk* 113(4): 823-829.
- Kéry, M. 2008. Estimating abundance from bird counts: binomial mixture models uncover complex covariate relationships. *Auk* 125 (2): 336-345.
- Kéry, M. and H. Schmid. 2006. Estimating species richness: calibrating a large avian monitoring programme. *Journal of Applied Ecology*. 43: 101-110.
- Kissling, M.L. and E.O. Garton. 2006. Estimating detection probability and density from point-count surveys: a combination of distance and double-observer sampling. *Auk* 123 (3): 735-752.
- Kwan, C. K.C. Ho, G. Mei, Y. Li, Z. Ren, R. Xu, Y. Zhang, D. Lao, M. Stevenson, V. Stanford, and C. Rochet. 2006. An automated acoustic system to monitor and classify birds. *EURASIP Journal on Applied Signal Processing*. 1-19.
- Lefebvre, G. and B. Poulin. 2003. Accuracy of bittern location by acoustic triangulation. *Journal of Field Ornithology*. 74 (3): 305-311.

- Lemon, R.E., J. Struger, and M.J. Lechowicz. 1983. Song features as species discriminates in American warblers (Parulidae). *Condor* 85: 308-322.
- MacKenzie, D.I., J.D. Nichols, J.A. Royle, K.H. Pollock, L.L. Bailey and J.E. Hines. 2006. *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Elsevier. USA.
- MacKenzie, L.L. Bailey, and J.D. Nichols. 2004. Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology* 73: 546-555.
- MacKenzie, D.I., J.D. Nichols, G.B. Lachman, S. Droege, J.A. Royle, and C.A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83 (8): 2248-2255.
- Maron, M., A. Lill, D.M. Watson and R. MacNally. 2005. Temporal variation in bird assemblages: how representative is a one-year snapshot? *Austral Ecology* 30: 383-394.
- McGregor, P.K., T. Dabelsteen, C.W. Clark, J.L. Bowser, J.P. Tavares, and J. Holland. 1997. Accuracy of a passive acoustic location system: empirical studies in terrestrial habitats. *Ethology Ecology & Evolution*. 9 (3): 269-286.
- McLaren, M.A., and M.D. Cadman. 1999. Can novice volunteers provide credible data for bird surveys requiring song identification? *Journal of Field Ornithology* 70 (4): 481-490.
- McShea, W.J. and J.H. Rappole. 1997. Variable song rates in three species of passerines and implications for estimating bird populations. *Journal of Field Ornithology*. 68 (3): 367-375.
- Mennill, D.J and S.L. Vehrencamp. 2008. Context-dependent functions of avian duets revealed by microphone-array recordings and multispeaker playback. *Current Biology* 18, 1314-1319.
- Mennill, D.J., J.M. Burt, K.M. Fristrup, and S.L. Vehrencamp. 2006. Accuracy of an acoustic location system for monitoring the position of duetting songbirds in tropical forest. *Journal of the Acoustical Society of America*. 119 (5): 2832-2839
- Mhøl, B, M. Wahlberg, and A. Heerfordt. 2001. A large-aperture array of nonlinked receivers for acoustic positioning of biological sound sources. *Journal of the Acoustical Society of America*. 109 (1): 434-437.

- Moore, J.E., D.M. Scheiman and R.K. Swihart. 2004. Field comparison of removal and modified double-observer modeling for estimating detectability and abundance of birds. *Auk* 121 (3): 865-876.
- Nelson, D.A. 1989. The importance of invariant and distinctive features in species recognition of bird song. *Condor* 91: 120-130
- Nichols, J.D., J.E. Hines, J.R. Sauer, F.W. Fallon, J.E. Fallon, and P.J. Heglund. 2000. A double-observer approach for estimating detection probability and abundance from point counts. *Auk* 117 (2): 393-408.
- Norvell, R.E., F.P. Howe, and J.R. Parrish. 2003. A seven-year comparison of relative-abundance and distance-sampling methods. *Auk* 120 (4): 1013-1028.
- Pollock, K.H., J.D. Nichols, T.R. Simons, G.L. Farnsworth, L.L. Bailey, and J.R. Sauer. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* 13: 105-119.
- Ralph, C.J., S. Droege, and J.R. Sauer. 1995. Managing and monitoring birds using point counts: standards and applications. In: *Monitoring bird populations by point counts* (C.J. Ralph, J.R. Sauer, and S. Droege, eds.), pp. 161-169. U.S. Department of Agriculture Forest Service General Technical Report (PSW-GTR-149), Pacific Southwest Research Station, Albany, CA
- Rempel, R.S., K.A. Hobson, G. Holborn, S.L. Van Wilgenburg, and J. Elliott. 2005. Bioacoustic monitoring of forest songbirds: interpreter variability and effects of configuration and digital processing methods in the laboratory. *Journal of Field Ornithology*. 76 (1): 1-11.
- Rosenberg, K.V. and P.J. Blancher 2005. Setting numerical population objectives for priority landbird species. Pages 57-67 in *Bird Conservation Implementation and Integration in the Americas: Proceedings of the Third International Partners in Flight Conference* (C.J. Ralph and T.D. Rich, Eds.). U.S. Department of Agriculture Forest Service General Technical Report PSW-GTR-191
- Rosenstock, S.S., D.R. Anderson, K.M. Giesen, T. Leukering, and M.F. Carter. 2002. Landbird counting techniques: current practices and an alternative. *Auk* 119 (1): 46-53.
- Royle, J.R. and W.A. Link. 2006. Generalized site occupancy models allowing for false positive and false negative errors. *Ecology* 87 (4): 835-841.

- Royle, J.R., J.D. Nichols, and M. Kéry. 2005. Modelling occurrence and abundance of species when detection is imperfect. *OIKOS* 110: 353-359.
- Sallabanks, R., J.B. Hauffer, and C.A. Mehl. 2006. Influence of forest vegetation structure on avian community composition in west-central Idaho. *Wildlife Society Bulletin* 34(4): 1079-1093.
- Sauer, J.R., J.E. Hines and J. Fallon. 2008. The North American Breeding Bird Survey, Results and Analysis 1966 -2007. Version 5.15.2008. USGS Patuxent Wildlife Research Center, Laurel, MD
- Sauer, J.R., W.A. Link, J.D. Nichols, J.A. Royle. 2005. Using the North American Breeding Bird Survey as a tool for conservation: a critique of Bart et al. (2004). *Journal of Wildlife Management* 69(4): 1321-1326.
- Sauer, J.R., W.A. Link, and J.A. Royle. 2004. Estimating population trends with a linear model: technical comments. *The Condor* 106:435-440
- Sauer, J.R. and W. A. Link. 2002. Hierarchical modeling of population stability and species group attributes from survey data. *Ecology* 86 (6): 1743-1751
- Sauer, J.R., B.G. Peterjohn, S. Schwartz, and J.E. Hines 1995. The Grassland Bird Home Page. Version 95.0. Patuxent Wildlife Research Center, Laurel, MD.
- Selmi, S. and T. Boulinier. 2003. Does time of season influence bird species number determined from point-count data? A capture-recapture approach. *Journal of Field Ornithology*. 74(4): 349-356
- Simons, T.R., M.W. Alldredge, K.H. Pollock and J.M. Wettroth. 2007. Experimental analysis of the auditory detection process on avian point counts. *Auk* 124 (3): 986-999.
- Somershoe, S.G., D.J. Twedt, B. Reid. 2006. Combining breeding bird survey and distance sampling to estimate density of migrant and breeding birds. *Condor* 108: 691-699.
- Spiesberger, J.L. 1999. Locating animals from their sounds and tomography of the atmosphere: experimental demonstration. *Journal of the Acoustical Society of America*. 106 (2): 837-846
- Swinston, K.A. and D.J. Mennill. 2009. Comparisons of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed and putative Ivory-billed woodpeckers. *Journal of Field Ornithology*. 80 (1): 42-50.

- Thompson, W.L. 2002. Towards reliable bird surveys: accounting for individuals present but not detected. *Auk* 119 (1): 18-25.
- Toms, J.D., F.K.A. Schmiegelow, S.J. Hannon, and M.A. Villard. 2006. Are point counts of boreal songbirds reliable proxies for more intensive abundance estimators? *Auk*, 123 (2): 438-454.
- Welsh, D.A. 1995, An Overview of the Forest Bird Monitoring Program in Ontario, Canada. In: *Monitoring bird populations by point counts* (C.J. Ralph, J.R. Sauer, and S. Droege, eds.), pp. 93-97. U.S. Department of Agriculture Forest Service General Technical Report (PSW-GTR-149), Pacific Southwest Research Station, Albany, CA
- White and Burnham 1999, MARK, Department of Fishery and Wildlife Biology, Colorado State University, Fort Collins, Colorado, USA
<http://www.cnr.colostate.edu/~gwhite/mark/mark.htm>
- Wilson, D.M. and J. Bart. 1985. Reliability of singing bird surveys: effects of song phenology during the breeding season