EXECUTIVE SUMMARY

Exploring the Use of
Concurrent Sound Recordings
To Improve the Reliability of the BBS

Final Report to
Patuxent Wildlife & Research Center
Breeding Bird Survey
12100 Beech Forest Road
Laurel, Maryland 20708-4038

Under Purchase Order No. 02ERSA0327

by

D. Archibald McCallum, Ph.D.
Applied Bioacoustics
P.O.Box 51063
Eugene, OR 97405
(541) 221-2112
mccalluma@qwest.net

Keywords: point count, bioacoustics, detection probability, availability, detectability, distance sampling, identification error, counting error, singing rate, bird census.

TABLE OF CONTENTS

ABSTRACT

This study produced heuristic estimates of accuracy, availability, and detectability on eight runs of an 18-stop simulated BBS route. The results suggest that availability, a reflection of observer-independent, natural variation in bird activity, is the major source of bias in BBS counts, rather than observer-specific effects. Recordings made by observers with low-end (< $100) equipment may be useful for estimating availability, if estimates can be based on loud, nearby birds. Better quality equipment costing < $1000 may produce recordings that allow post-processors to detect all species detectable by observers, facilitating the estimation of availability and the proportion of species overlooked by observers.

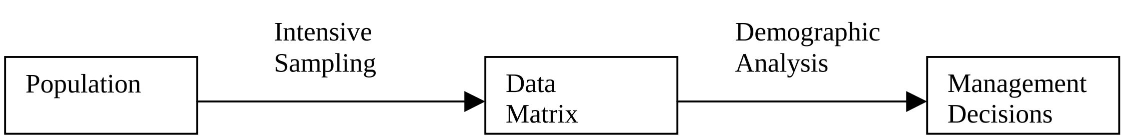
INTRODUCTION

THE NEED FOR SAMPLING

Decisions whether to and how to manage wildlife populations must be based on information. Data must be gathered on the population, or surrogate population, and then organized into a data matrix for analysis and subsequent decision-making. In the diagrams that follow, the boxes represent states of the system, the arrows and associated labels indicate transfer functions. Importantly, the data matrix, although a mere artifact of the population at a moment or brief interval in time, can be stored and retrieved for later re-analysis. Accuracy of this artifact can have a critical effect on the success of management decisions (see below).

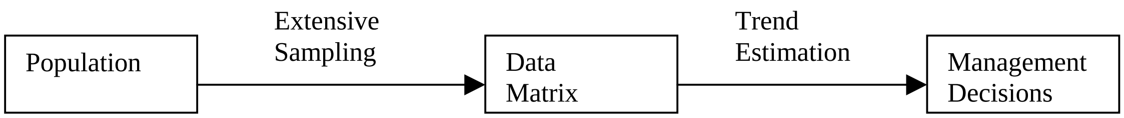
Intensive Sampling

Intensive management requires demographic data. Obtaining such data requires intensive sampling, which is time-consuming and economically expensive. A *de facto* two-tiered approach to management of nongame bird species in North America is a consequence of this fact. Intensive sampling, involving nest-finding, color-banding of individuals, and perhaps radio-tracking of dispersing individuals, can produce a reasonably representative data matrix, although at a cost that is ordinarily only undertaken when trying to save a species or population from extinction. Although the presence of the investigators may indeed change the dynamics of the population itself, the data themselves are largely unbiased.



Extensive Sampling

Other species are monitored for decreasing population trends that may place them in the imperiled category, which would then lead to intensive study and management. Estimates of trend are based mostly on brief surveys conducted over wide geographic areas. Because of the briefness of these surveys, investigators are unlikely to bias population dynamics, but the potential for the resulting data matrix to be biased is high.



Correcting Bias in Extensive Samples

N, the number of individuals present in an area, is estimated in such surveys from a sample count, *C*, of that area. The expected value of the count is given by $E(C) = Np$, where *p* is the "detection probability" (Nichols et al. 2000, Farnsworth et al. 2002) or "index ratio" (Bart and Earnst 2002). Use of *C* (raw count data) to estimate the change in *N* over time (i.e., trend), requires the "proportionality assumption" (Thompson 2002) that a trend in *p* does not exist (J. Bart, pers. com). Populations in different areas and populations counted with different methods cannot be compared quantitatively with indices (i.e., *C* values)(Bart and Earnst 2002). In order to make inferences from count data without "discomfort with the knowledge that such inferences depend upon untested assumptions" (Nichols et al. 2000: 394), it is necessary to estimate *N*, preferably with methods that are grounded in statistical theory (Thompson 2002).

The standard form of such an estimate is given by

$$\hat{N} = \frac{C}{\hat{p}} \tag{1}$$

(Nichols et al. 2002, equation 4). These parameters apply to the birds of a sex, species, area, or indeed any group that has a common value of *p* (Nichols et al. 2000). *P* is the probability of detecting a typical individual. It can be thought of as the average detection probability of all the individuals that reside in the area being surveyed, although is it never estimated in this way. Instead, it is estimated from population parameters

Components of Detection Probability

Although *p* is potentially a function of a number of factors, it is useful for aural surveys to subdivide it into two main components (Farnsworth et al. 2002):

$$p = p_s \, p_{d|s} \tag{2}$$

where p_s is the probability that an average bird sings (or produces some other detectable cue), and $p_{d|s}$ is the probability it is detected, given that it sings. Recognizing the distinction between these two component probabilities, "availability" and "detectability," respectively, is essential for designing sampling schemes with the potential to provide unbiased estimates of N.

Components of "Detectability"

Unlike intensive survey methods, (e.g., territory-mapping, nest-finding protocols), “rapid survey” methods (e.g., point counts, line transects) define a single detection of an individual as sufficient to count that individual. Further detections of that individual do not change *C*. Indeed, count periods are intentionally made brief to minimize the possibility of double-counting of an individual (e.g., Buckland et al. 2001). So, detectability ($p_{d|s}$) is actually the probability that a bird will be detected *at least once* during its active periods. Therefore

$$p_{d|s} = 1\text{-}(1\text{-}\ p_{1d})^{s} \tag{3}$$

where p_{1d} is the probability of detecting an average cue; and *s* is the number of cues, i.e., songs or other detectable acts, it actually produces during the count period.

P_{1d} is a measure of conspicuousness, i.e., it captures reductions in detectability due to the following four factors:

- **Amplitude** of the vocalizations of the average individual. Amplitude diminishes as the square of the distance between the source and the detector, so detectability is strongly influenced by distance and correlated factors, such as sound blocking structures.
- **Auditory acuity** of the observer. Acuity is frequency-dependent in all humans, and is greatest at 1-2 kHz, below the frequencies of most bird sounds. One reason that *p* differs among species is the varying degree to which the birds’ sounds fall outside this 1-2 kHz band. Moreover, individual humans vary in both general acuity and frequency response (Emlen and DeJong 1981). Obviously, *p* is observer-specific for these reasons, even if not for others.
- **Attentiveness** of the observer. Because point counts are typically conducted in real time, i.e., the observer cannot rewind and hear or see any cues a second time, it is standard practice for an observer to attempt to focus on a single singer, identify it, and then move on to another. This means that the listening time of the observer is divided among all the singers, some of which will be missed if they cease vocalizing before the observer has a chance to attend to them.
- **Masking** of focal sounds by other sounds, including ambient noise, speech of the observer and any assistants present, vocalizations of other species, and vocalizations of nonfocal individuals of the focal species.

High amplitude mitigates the other three causes. If a sound is loud, it is more likely to be noticed and more likely to mask other sounds then to be masked. Amplitude is directly related to distance, while the other three factors come into play because of the low amplitude of sounds from distant sources.

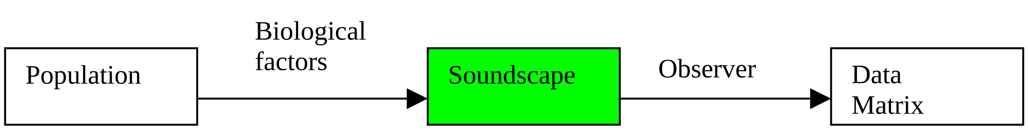
Parameter *s* is the number of cues produced during a count period, independent of their intensity. High singing rates (*s/m*) mitigate all four causes of non-detection, by giving the observer multiple opportunities to make the single detection that is needed to count an individual.

Equation (3) quantifies the intuitive relationship between singing rate and the likelihood of detecting an individual. The good news from this equation is that even inconspicuous cues (low p_{1d}) can result in detection when they are numerous (high *s*). Equation 3 also shows why the dawn chorus may not be the optimal time to conduct a survey. Singing rates (*s/m)* tend to be highest at this time, and owing to correlation of *s* among individuals, masking by other individuals may reduce p_{1d}, offsetting the advantage of high *s*.

The main source of trend data in the United States and southern Canada is the Breeding Bird Survey (BBS), which is based on 50 3-min point counts repeated once annually. The BBS, other point-count protocols, and indeed most survey methods applied to territorial land birds, rely heavily on songs and other vocalizations to detect the birds present in the population of interest. Approximately 90% of detections in such surveys are based on sound (Buckland et al. 2001).

THE SOUNDSCAPE

A variety of biological factors (e.g., time of day, pairing status, number of conspecifics) influences which individual birds sing and at what time. The resulting soundscape has objective reality, but because it is a transform of the population itself, any data matrix estimated from it becomes an artifact of the soundscape rather than the population. In the worst case, the soundscape may be a misrepresentation of the demographic vitality of the population. This unwelcome possibility could be reality if, for example, males with mates are the least likely males to sing. In polygynous species and those socially monogamous species in which mated males pursue extra-pair fertility, attractive males may continue singing throughout the season, and using the soundscape as a proxy for the population may be less problematic. Regardless, these biological factors are independent of any effort to sample the soundscape, i.e., the resulting concerns would pertain even if the soundscape were known without error. If there are lawful relationships between population characteristics and the soundscape, e.g., a seasonal change in the relationship between number of females present and number of males singing, then it may be possible to devise correction factors to improve the picture of the population that can be gained from the soundscape.



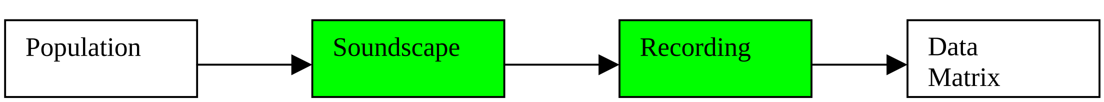
Even when the information available in the soundscape is positive, i.e., the amount of singing is correlated with the number of birds present, sampling the soundscape introduces additional, but tractable, problems, as follows: (1) Because many species of birds have repertoires of song-types and because some species sound similar to many observers who are competent to distinguish them visually, the potential for identification error is nontrivial. (2) When the sample is very brief, as in a 3-min stop on the BBS, additional statistical factors related to sampling error may lead to bias in the data matrix. (3) Point-count theory requires only one detection of an individual for it to be “counted” on the survey. Many birds sing at once, and the brief period available may not afford the observer sufficient time to hear all of them, lowering detectability. Moreover, when the sample is taken from a single point, as is usually the case, the fall-off in detectability owing to attenuation of sound with distance must also be accounted for. These three factors will be referred to as “accuracy,” “availability,” and “detectability,” respectively.

Factors 2 and 3, subsumed under the rubric of "detection probability," are currently an area of active research. The goal is to estimate correction factors that can be applied to the raw data matrix to make it a more accurate reflection of the population. The need for such correction factors is demonstrated by a U. S. National Park Service study in southeastern Utah. Long-play digital recordings detected 45 species at one location where point counts detected 11 (Daw and Ambrose 2003).

Included in the needed correction factors is one for "availability," the probability that a bird that is present will sing during the count period (Farnsworth et al. 2002, in press; McCallum in press). In models introduced to date, corrections for unavailability treat all males (i.e., potential singers) in the population as equivalent, and hence do not get at the difference in detectability of mated and unmated males. A well-developed methodology for correcting for distance effects on detectability exists (Buckland et al. 2001), and is now receiving wide attention in the bird-monitoring community.

For brief, sound-based surveys, it is now technically and economically feasible to make electronic recordings of the soundscape for post-processing. The amount of post-processing can vary from one-time audition (in an attempt to delay the time and relocate the place at which the observer experiences the soundscape, while minimizing any change in the amount of information available to the observer / interpreter) to detailed bioacoustic analysis.

Although an audio-recording is a more objective record of the soundscape than a field notebook or a point-count data sheet, it is undeniably, like them, an artifact, interposed between the soundscape and data matrix, and care must be taken to ensure that it is an accurate representation of the soundscape at the time of recording.



Making a recording of the soundscape requires a (1) a microphone system, (2) a recording platform, and (3) storage media. All microphones are electronic devices that transduce variation in pressure waves to variation in voltage. Recording platforms amplify the time-varying voltage that is supplied by the microphone system and either encode it on magnetic tape (analog recorders), or convert it to binary-number format (digital recorders) and store the number stream on one of several possible types of media (magnetic tape, mini-discs, memory chip).

PROS AND CONS OF RECORDING POINT-COUNTS

Recording the soundscape has numerous advantages over simply listening to it and taking notes in real time, as is currently done with the BBS and other point-count protocols. The most important of these is that an objective representation of the soundscape can be stored and retrieved later for re-analysis. This is the standard practice in science. Data that cannot be independently verified violate a basic principle of science. As the consequences of management decisions based on survey data become more and more costly for powerful economic interests, the potential for survey results to be challenged in court increases. Recording the soundscape rather than simply taking notes on it produces a record that can be examined independently by contesting and neutral parties, thereby potentially averting harm to imperiled wildlife populations from legal challenges to the data matrices used to justify their protection.

Recording the soundscape also makes reanalysis possible, even decades later, for different research objectives from those intended at the time the recordings were stored. The usefulness of old bird-egg collections for demonstrating the negative effects of pesticide poisoning, and of old collections of study skins for molecular genetic analyses are but two examples of valuable but unintended uses of archived samples of biological material.

Recording also mitigates the logistic problem of an insufficient number of field-workers able to identify all bird species in the study area, by ear, in real time (Rempel et al. 2005, http://flash.lakeheadu.ca/~rrempel/CVX/). Even the simplest post-processing (listening to the tape) can reduce identification errors through use of an expert on the sounds of the birds in the study area in lieu of inexperienced field workers. Using more than one such expert permits estimation of the level of confidence in the data matrix. Alternatively, a less experienced observer can listen to a tape repeatedly until convinced of his/her identifications, or use visualization techniques (e.g., spectrograms) to increase the likelihood of correct identifications.

When post-processing is not limited to a single audition of the tape, it increases detectability by allowing for independent examination of contemporaneous sounds from different individuals. Post-processing is especially advantageous when the sounds are visualized with spectrograms, which separate the sounds’ frequencies.

Because the amplitude of a sound is attenuated as the square of the distance from the source to the microphone, distance from the bird to the microphone is encoded in the recordings. Deriving the distance to singers from tapes would make it possible to use the distance method (Buckland et al. 2001) to estimate detectability, a major step toward decreasing bias in the population estimates derived from point counts (Rosestock et al. 2002). Decoding this information requires a distance-amplitude function, a curve that gives the true distance as a function of the amplitude on a calibrated tape. Such functions are likely to differ for species, call-type, direction the bird is facing (relative to the microphone), and vegetation type (because vegetation absorbs or reflects acoustic energy).

The potential liabilities of recording all fall under the headings of “availability” and “detectability.” Clearly, visual information is not available in a recording. Moreover, several potential auditory short-comings of recordings must be minimized. In typical point-counts, the number of individuals is counted. When individuals are detected with aural cues, counters typically rely on differences in amplitude and bearing of the sounds to distinguish them. Amplitude differences are likely to be recorded faithfully, but it is an open question whether low amplitude sounds are as detectable on a tape as they are in the field. This depends primarily upon the sensitivity of the microphone system used (see below).

Potentially a more serious problem for recordings is the loss of directional information in recordings. Single-channel (“monaural”) recordings are technically the easiest to acquire, but do not encode directional cues. Two-channel (“stereo”) recordings are technically feasible with consumer equipment, and have been used with satisfaction by the Centre for Northern Forest Ecosystem Research (Rempel et al. 2005, http://flash.lakeheadu.ca/~rrempel/CVX/). The more channels, the more directional information, but recording more then two channels requires expensive professional equipment and cumbersome multiple-microphone set-ups (AVAIL: Fig. 3).

In summary, post-processing of recordings is likely to increase the accuracy of identifications made in the field because multiple passes through the data are possible, while reducing the count of species and individuals through the sacrifice of visual information. The numbers of individuals and species detected can be biased positively or negatively, depending primarily upon the sensitivity of the recording system. The number of individuals detected is likely to be biased negatively through the loss of directional information in single-channel recordings, but corrections for lowered detectability owing to distance can be obtained from tapes, at least in principle.

OBJECTIVES OF THIS STUDY

A number of empirical questions arise from the foregoing discussion. The following were addressed in this study:

QUESTIONS

1. Do typical BBS data need correction factors for accuracy, availability, and detectability?

2. Can omni-directional recordings made by the observer while running the BBS be used to estimate any of these correction factors?

3. If the answer to any of the above is yes, how good (expensive) does the recording equipment have to be to produce acceptable recordings?

REPORTS

Four reports were submitted for this study. Rather than referring to them below as "McCallum, unpublished a, b, c, and d." They will be cited as follows:

SIM McCallum, D. A. 2003. The relationship between cue abundance and cue availability, and its impact on detectability during point count surveys: A Monte Carlo simulation study

EQUIP McCallum, D. A. 2004. The pros and cons of audio-taping point counts: Equipment considerations.

POST McCallum, D. A. 2005. Using observer-recorded tapes to enhance a BBS survey: Augmentation but not substitution.

AVAIL McCallum, D. A. 2005. The relative contributions of bird availability and observer detection rates to repeatability of BBS results.

Two published papers by the author, although not funded by USGS, are essential to this discussion. They will be referred to as follows:

DET McCallum, D. A. In Press [2005]. A conceptual guide to detection probability for point counts and other count-based survey methods. *In* Ralph, C. J., and T. D. Rich (Eds.). Bird Conservation Implementation and Integration in the Americas. USDA Forest Service Gen. Tech. Rep. PSW-GTR-191.

BEVI Scott, T. A., P.-Y. Li, G. C. Greene, and D. A. McCallum. In Press [2005]. Singing rate and detection probability: An example from the Least Bell's Vireo (*Vireo belli pusillus*). *In* Ralph, C. J., and T. D. Rich (Eds.). Bird Conservation Implementation and Integration in the Americas. USDA Forest Service Gen. Tech. Rep. PSW-GTR-191.

Other published works will be cited conventionally.

RESULTS AND DISCUSSION

I. DO TYPICAL BBS DATA NEED CORRECTION FACTORS FOR ACCURACY, AVAILABILITY, AND DETECTABILITY?

Differences in ability to identify birds by sound, differences in availability of birds for detection, and differences in attentiveness to the birds that are singing are all likely to exist, but the significance of such differences has not been documented. While the fall-off in sound amplitude with distance to a singing bird is a physical fact, it requires a correction factor only if observers differ significantly in auditory acuity. Do they? The following specific questions were addressed in the report.:

A. What percentage of birds detected by realtime observers is misidentified, and how much do observers vary in accuracy?

This question could not be addressed adequately in the study for methodological reasons. We explored four options for synonymizing the sounds on which field observations were based with sounds on the recordings (AVAIL). It was decided that speaking the identification onto the observer's tape or speaking it into a separate dictation system were too obtrusive and would jeopardize the applicability of the results to actual BBS data. The compromise alternative was to map all detections in hopes that individuals could be synonymized by location on the three-channel recordings.

Unfortunately, the map option proved unsatisfactory for this purpose (AVAIL), although it did provide useful information for assessing the detection distance of recordings (POST). A single data set cannot be optimal for all questions, and other results indicate that preserving the realism of the simulated BBS runs was a good decision. Mis-identification is probably a far less important source of bias than detection rate and availability.

On the other hand, speaking an occasional species name into a microphone is no more intrusive than speaking names of all detections to a data-recorder, which some observers do, and which is the essence of the double observer method (Nichols et al. 2000). The experience of this study indicates that assessing identification error will require a one-for-one synonymy between the referent (cue) and the assessment, as in normal academic testing. (Imagine giving a fill-in-the-blanks test in a class, and letting the students write all the answers at the bottom of the page; this is an example of lack of synonymy between referent and response.)

This may be more readily done as a test administered on a computer than in a field test. Preparing such a computerized test is straightforward and within the current capabilities of Applied Bioacoustics. The challenge in this approach is the great variability of song-types found among so many species of birds. Ideally, the test should be limited to the song-types found on the observer's BBS route. The only way to obtain an unbiased representation of the song-types encountered on the route is to record them on the route. Then an observer-specific test must be prepared with these route-specific sounds.

An attractive alternative is to have the observer make a simple recording, as done by the observers in this study (POST:methods), and speak the identification of haphazardly chosen sounds into the microphone. After the survey, the observer could "grade" the test, noting the species requiring some study. This approach would treat the problem as a training issue. Alternatively, the test could be "graded" by another party, leading to a correction matrix that could be applied to that year's data.

Before either the indoor or outdoor alternatives are implemented, an objective assessment of misidentification error rate is needed. An attempt to do this in the field is likely to show that misidentification error is insignificant compared to availability and detectability errors. A simple survey would show whether observers try to minimize misidentification by ignoring cues they cannot identify, i.e., by shunting the problem from accuracy to detectability.

B. Are all birds resident at a stop available for detection during a three-minute BBS stop, and if not, what proportion of the population does emit a detectable aural cue during such a time period?

Empirical studies of availability framed in a monitoring context are few and far between, probably because the concept of availability is so new, dating in print from 2002 (Farnsworth et al. 2002). Before that date, theorists conflated availability and detectability. For example, the double observer method (Nichols et al. 2000) was claimed to estimate detection probability, P, but in fact it has no means of accounting for silent or invisible birds (DET). One explicit attempt to relate availability to detection probability was the study by Scott et al. (BEVI) of the Least Bell's Vireo. Following a single individual, presumably but not necessarily the same bird, for 4 hr per morning over seven days distributed throughout the breeding season, they found daily per-minute availability to range from 0.36 to 0.96.

This study compared detection of over 40 species on eight visits to 18 stops (AVAIL). The eight visits were conducted in 11 days, so phenological changes in pairing status and nest-stage, while no doubt present, are far from maximal. This rare level of replication made it possible to estimate availability directly, rather than indirectly. All but eight species had availability rates < 0.5. The maximal availability rate was 0.75 for the Prairie Warbler. These rates may have been low partially because the study was conducted in late June and early July, at least one month after the BBS should be run in the study area. Many individual birds, however, were singing at high rates when they did sing, and it seems likely that the periods of silence documented in this study are representative of the bout structure of most of these species.

No doubt variation in availability has been observed by every researcher who has conducted a behavioral study of a marked population of a single species, and perhaps it has been quantified in many cases. For example, for a study of territorial behavior I quantified the amount of singing by different individual Mountain Chickadees at different hours and stages of the nesting cycle (McCallum unpubl.), but these data have not found their way into the monitoring literature. A massive search of the behavioral literature for such data would be worthy of funding.

Behavioral ecologists are currently producing more data on availability in single species because of the popularity of research on the function of birdsong. One study that explicitly addressed detection issues is Walker (2000, also talk at 2002 PIF meeting). An emerging pattern is the use of one type of vocal advertisement (e.g., “Type 1 song in parulids) to attract mates and another type (e.g., Type 2 song in parulids) to repel rivals, although perhaps also to promote extra-pair copulation (pers. obs.). The general pattern is that the former kind of cue declines markedly in availability with pairing. Those males continuing to sing a mate-attraction song well into the breeding season (e.g., the Bell’s Vireo studied in BEVI), are likely to be either low-quality males, e.g., yearlings, and/or to have a low-quality territory. Given that the BBS is conducted well into the breeding season for most species, the possibility that the survey counts mostly losers is arresting.

C. What percentage of birds available for detection on the BBS is not detected (or ignored), and how much do detection rates vary?

Overall detection rates for the four observers were 0.760, 0.836, 0.900, and 0.905, all well above most availability rates (AVAIL). In fact, species-specific detection rates were double the availability rates for most species. Only one species had higher availability than detectability, the Prairie Warbler (AVAIL).

D. How much do observers vary in auditory acuity?

In a field experiment, detection of 20 test sounds by seven observers varied from 60% to 95% at 50 m, but the range was reduced to 25%-50% at 200 m (EQUIP). Observers varied in acuity in laboratory tests (Emlen and DeJong 1981) as well. Some of the birds missed by observers on the simulated BBS routes (AVAIL) were indeed represented by faint sounds, but others were quite conspicuous. I did not attempt to break the miss rate down into an acuity component and an attention component, but it seems that neither was 0.

2. CAN OMNI-DIRECTIONAL RECORDINGS MADE BY THE OBSERVER WHILE RUNNING THE BBS BE USED TO ESTIMATE ANY OF THESE CORRECTION FACTORS?

A. Which of these correction factors can be estimated directly from BBS data?

Before addressing estimation of correction factors from recordings, one must first ask if such correction factors can be estimated directly from BBS data of any kind, realtime or post-processed.

Accuracy

The correctness of the identification of an individual bird to species is verifiable only if the bird itself, or some artifact of it, is available for independent examination by an arbiter, preferably any arbiter. A specific arbiter may therefore accompany the observer and produce a realtime assessment of the accuracy of the observer's identifications. But, the standard in science in general and ornithology in particular is that that arbiter's conclusions should therefore be open to scrutiny, by anyone. In other words, the only way to objectively assess the accuracy of a field survey is to bring back artifacts of all the individuals counted, labeled by species, age, sex, etc., which can be examined and compared to reference material at any time thereafter by any arbiter.

Availability

By definition, availability cannot be estimated directly from a single survey. Availability is the proportion of all individuals that are targets (e.g., territorial males) of the count that produce a cue during the count. Individuals that do not produce a cue are by definition not counted, and yet their number must be known for the calculation of availability to be performed. This is why some form of the double sample method (Bart and Earnst 2002) must be conducted for availability to be assessed directly (DET).

To put it another way, N must be known for availability to be estimated directly, a requirement that is problematic when the object of the survey is to estimate N. The essence of the double sample method is that N is estimated independently of "brief surveys" with intensive surveys (Bart and Earnst 2002). In the current study, N was estimated as the maximum of the eight counts of the species at the stop. It should be recalled that the sole use of this estimate was the production of a heuristic estimate of availability. Availability was not in turn used to estimate N.

This underlines one of the benefits of distinguishing availability from detectability. Availability is the component of detection probability that cannot be estimated directly from single visits to a location. Detectability, the other component, can be (see below). It may be possible to estimate availability indirectly from single-survey data if some factor that is highly correlated with availability can be measured during single surveys. This is the strategy of the removal method (Farnsworth et al. 2002), which uses the repeatability of cues among temporal subdivisions of the single survey to estimate availability (DET). Although cue rate and availability may be correlated empirically, they are not synonymous, nor do they have a necessary functional relationship. Estimates of availability from cue rates are therefore indirect and correlative.

Detectability

The major components of detectability are distance effects and observer effects. (See detailed decomposition in Introduction.) The former are estimable from single-survey data with the distance method (Buckland et al. 2001); the latter with the double observer method (Nichols et al. 2000).

1. Distance Effects

Buckland et al. (2001), Rosenstock et al. (2002), McCallum (in press), and Farnsworth et al. (in press) discuss the assumptions of the distance method. In addition to those well-understood mathematical requirements, the sheer difficulty of estimating distance to a singing bird deserves mention. The observer must either see it and estimate distance visually or make the estimate on the basis of sound amplitude. As the vast majority of birds is not seen, observers clearly must rely on sound to estimate distance. This means accounting for the amplitude of the sound produced by a singing bird of species X, the variation in this amplitude with the orientation of the bird, and the attenuation of this amplitude by vegetation and wind.

It would be pardonable to express some doubt about the ability of observers to perform this feat accurately during a three-minute BBS stop. So, to paraphrase terminology introduced by Richard Lewontin (pers. comm., Fritz Taylor, 1983), the distance method is theoretically sufficient, but its empirical sufficiency for aural detections can be questioned. It should be remembered that the distance method was developed for shipboard visual detection of cetaceans. The book describing this method (Buckland et al. 2001) shows an observer using a range finder on the deck of a ship. Visual detection is based on incident light that is reflected off the target. Aural detection is based, almost exclusively on intentional signals produced by the target. Much of the difficulty of estimating distance from sounds is the result of intentional variation in broadcast direction and in signal amplitude.

Add to this the 6-dB lowering of amplitude of a sound that is passed through beech forest, compared to the same sound passing over an open field (EQUIP). In the open field the same sound is twice as loud as in a forest, requiring the observer who is estimating distance from sound amplitude to adjust for the amount of forest in the soundscape.

2. Observer Effects

The second-observer method is theoretically sufficient to estimation of misses owing to observer inattention. Although designed for visual surveys, it has been adapted to sound-based surveys (Nichols et al. 2000). Because it requires that the observer call out the species of every bird to the scribe / second observer, it may lower detection rate. Its utility for sound-based counts therefore requires further study.

B. Do concurrent omni-directional recordings produce a detectable record of all sounds audible to on-scene observers? Can recordings be used to count individuals?

The short answer to this question is "yes" for detectability, but it is conditional on the equipment used, which is discussed in detail below. Counting more than two or three individuals of most species from recordings is not practicable in most cases, and may be impossible in some.

C. Can concurrent omni-directional recordings be used to improve identification accuracy on point counts?

Such recordings can be used readily to assess accuracy as long as the observer dictates enough information onto the recording to assure correct linkage of the song with the singer, e.g., "That last trill was a Chipping Sparrow." Resulting data can be used to calculate correction factors that adjust the counts for each species in the survey. I was unable, however, to demonstrate this potential in the present study because the linkage between song and singer was not available (AVAIL).

D. Can concurrent omni-directional recordings be used to estimate availability?

Concurrent recordings can be used to count singing rates as well as presence/absence (i.e., availability) of a cue in different temporal segments of the time period. Such estimates are likely to be more accurate than field estimates because recordings allow revisitation.

Singing rates are not likely to be as useful for this purpose as sequential presence/absence of cues. In a Monte Carlo simulation of the relationship of cue abundance to availability (SIM), I found that "recapture rate," the probability that a bird singing in a given minute had been singing in the previous minute, when averaged over at least 25 samples, explained 90% of the variation in true availability. In contrast, one-minute estimates of singing rate could explain no more than 44% of the variation in true availability, regardless of the number of such estimates that were averaged. This was because continuation probability has a profound influence on availability, but proved uncorrelated with short-term data.

Even if only two 1-min count periods were used with the Farnsworth et al. (2002) model, a sample of 25 counts would yield a very serviceable estimate of availability (SIM). Using more than two periods, of more than 1-min duration, as proposed (Farnsworth et al. 2002), should improve the estimate, although this assumption should be tested. If this can be done in the field, it can certainly be done from tapes, assuming that individuals can be recognized on the tapes. Differences in amplitude and song-type help distinguish individuals on tapes, as they do in the field. A modification of the removal method that allowed observers or interpreters to confine estimates of availability to selected, more-conspicuous individuals, would make the method more amenable to post-processing and easier to implement in the field. Using conspicuous birds to estimate availability would reduce the potential for low detectability to bias estimates of availability.

Although the results of the simulation are heartening for the removal model, it is sobering to note that real-life data from the Least Bell's Vireo did not support the feasibility of estimating availability with the removal model (BEVI). In that study, contrary to this one, a minimum of 10 random 1-min samples of singing rate explained 80% of the variation in availability. That is good, of course, but, short-term estimates of "recapture rate," calculated exactly as was done in this study, were useless for estimating availability (BEVI). Why the difference? The Bell's Vireo averaged 3 songs per minute, equivalent to a probability of singing of 0.1 in the model (SIM), i.e., well within the simulated parameter space. The number of runs per hour (a measure of realized continuation probability) produced by the vireo were also in the parameter space of the simulation model. This suggests that further study of real birds is needed to tie down the relationship between availability and the short-term distribution of cues. This could be done with the recordings made for this study.

E. Can concurrent omni-directional recordings be used to estimate detectability of singing birds?

Detectability is the observer-specific component of detection probability. Although it has numerous components (DET, and Introduction above), the simplest decomposition is into a component that can be estimated with the distance method (Buckland et al. 2001) and a component that can be estimated with the double observer method (Nichols et al. 2000).

This study (AVAIL) showed that even highly-skilled and experienced BBS observers miss conspicuously singing birds. Regardless of the causes of these misses, e.g., attention to other birds, forgetting to record data, high-frequency hearing loss, this source of bias in BBS counts can be mitigated with the estimate of P produced by the double observer method. Although this method was once suggested for estimating P in its entirety (Nichols et al. 2000), it is now clear that it estimates only the observer-specific component of detectability (Farnsworth et al. 2002, DET). In this study, post-processing of observer-recorded single-channel tapes by a person lacking previous experience with spectrograms did reveal some of the birds missed by the observers (POST). Post-processing of multi-channel professional recordings by an experienced bioacoustician revealed even more misses.

Remembering that availability seems to be a much more serious source of bias in BBS data than detectability (AVAIL), some bias could probably be reduced by having observers record their counts, or a subset of them, and then listen to the tapes afterward for species they missed. Alternatively, a professional bioacoustician could post-process the tapes or a sample of them to produce estimates of miss-rates. Results of the present study suggest, though, that reducing bias due to low availability is a much more serious problem.

F. Can concurrent omni-directional recordings be used to estimate distance to singing birds?

The distance method is being promoted actively in the North American point-counting community. This is good as long as it is acknowledged that the distance method does not produce an estimate of the availability component of detection probability. In fact, this was first pointed out, albeit rather unobtrusively, by the developers of the method (Buckland et al. 2001:189).

Further, the difficulty of estimating distance to singing birds has perhaps been underestimated (see above). Estimating distance from a recording is equally challenging. This study documented considerable variation in the performance of several recording setups (EQUIP), which is analogous to inter-observer differences in acuity. It documented, as expected, average differences in sound transmission between forests and fields (EQUIP). For recordings to be used to estimate distance to singing birds, species-specific distance-amplitude functions must be compiled, and recording equipment must be calibrated and then used successfully by observers. This use of observer-recorded tapes seems much less plausible than using them to estimate miss-rate and accuracy. It can be done, but it is highly technical and the gain may not be worth the expense. Again, estimating distance in the field is also difficult.

3. IF THE ANSWER TO ANY OF THE ABOVE IS “YES,” HOW GOOD (EXPENSIVE) DOES THE RECORDING EQUIPMENT HAVE TO BE TO PRODUCE ACCEPTABLE RECORDINGS?

I tested two inexpensive General Electric cassette recorders (< $50) against two professional Sony recorders (> $1000), all analogue. Simultaneously I tested two Optimus microphones (< $50) against two Sennheiser professional microphones (>$400).

All four recorders recorded at higher than rated speed, but the greatest deviation in frequency response was 4%, equivalent to shifting a C note to a C# (EQUIP). This level of bias should have no effect on bird identification. Professional microphones and professional recorders were both more sensitive (produced higher-amplitude recordings of the same sounds) than the low-end alternatives. The effects were additive, but variation within recorder brand was surprisingly high (EQUIP). If upgrading from the cheapest setup, one could realize approximately the same gain in sensitivity by upgrading to either the Sennheiser microphone ($400) or the Sony recorder ($1200, but see below for advances in recorder technology).

Data from other parts of this study suggest that these relative differences in sensitivity do have practical consequences for detection of birds. Recordings made with the Sennheiser microphones during human audition tests yielded detection rates that were favorably comparable to those of realtime human observers (EQUIP). Working in the blind with the tape made with the low-end recorder, I detected a median number of sounds at 50 and 150 m (although not at 100 and 200 m). Using the tape made on the professional analogue recorder, I improved my detection rate. When I made a directed search for the test sounds, I found more test sounds than six of seven observers. The latter result is important because it shows that the system is as sensitive as the auditory systems of most humans. In contrast, the recordings made in this context with the low-end microphone were too faint to score.

Observer-recorded tapes were made with the same low-end recorders. On most days, two observer tapes were made, one with the low-end microphone, one with the professional microphone. As the interpreter who post-processed these tapes was allowed to choose the better of the two tapes, it is not possible to compare the performance of the two microphones. It does appear, however, that these systems were less sensitive to low amplitude sounds than the human observers, because the interpreter detected a lower number of birds than the observers at all distances, while exceeding the observers' counts of individuals within 50 m, for many species.

Overall, then, equipment totaling less than $100 in cost captures the sounds of many birds, but not enough of them to substitute for a realtime observer. Tapes made with comparable equipment may be used to check accuracy of identification and to estimate availability of nearby birds with the removal method (Farnsworth et al. 2002), but should not be used to estimate detectability or the number of birds present.

Professional equipment costing more than $1500 may capture all the bird sounds that many BBS observers would hear. Counting all of the individuals would remain very challenging to impossible for species with more than two or three individuals per stop, because of the loss of directional information. But, the benefits of post-processing are significant (AVAIL), and presence-absence data equal or superior in completeness to those of realtime observers could be extracted efficiently from omni-directional recordings made with good equipment. Such tapes can also be used in lieu of a second observer (Nichols et al. 2000) to estimate the miss rate of the realtime observer for species, but not for individuals.

Recorder technology has improved since the equipment tests were performed in 2002. Technical improvements in mini-disc recorders, some of which are less than $300 in cost, makes them superior to the $1200 analogue recorders I used:

> When recordings made on these [MD=mini-disc] machines are compared to similar recordings created by high-quality analog cassette recorders the results typically favor the MD version. The total lack of tape hiss, wider frequency bandwidth of the MD system, and lower overall distortion levels all contribute to favoring the MD. Low-level high-frequency information that the MD ATRAC system might

> have masked would have been buried in tape hiss on the analog cassette version so frequency masking is not an issue when these two formats are compared.
>
> – Macaulay Library, Cornell Laboratory of Ornithology
> http://www.birds.cornell.edu/MacaulayLibrary/contribute/equipMd.html

It is therefore likely that recordings made with equipment costing less than $700 at May 2005 prices (http://www.stithrecording.com) can be used to estimate availability with the removal method, estimate miss-rate of conspicuous birds with the second-observer method, and check the correctness of questionable identifications. Of these, availability is by far the most important. Amplitude-based distance estimates should be viewed with reservation, whether obtained by ear or from recordings.

Acknowledgments

I would like to thank Keith Pardieck and Bruce Peterjohn for their able administration of this contract, and Mary Gustafson, Sam Droege, Keith Pardieck and Bruce Peterjohn for participating as observers in this study. I would like especially to thank Glenn Johnson for the diligence, professionalism, and good nature with which he performed the post-processing.

LITERATURE CITED

Bart, J., and S. Earnst. 2002. Double sampling to estimate density and population trends in birds. Auk 119:36-45.

Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas. 2001. Introduction to distance sampling: Estimating abundance of biological populations. Oxford Univ. Press, Oxford, UK.

Daw, S., and S. Ambrose. 2003. Using digital recordings to supplement avian surveys, Canyonlands National Park, 2002. USDI, National Park Service, Southeast Utah Group, unpublished final report.

Emlen, J. T., and M. J. DeJong. 1981. The application of song detection threshold distance to census operations. Studies in Avian Biology No. 6:346-352

Farnsworth, G.L., K.H. Pollock, J.D. Nichols, T.R. Simons, J.E. Hines, and J.R. Sauer. 2002. A removal model for estimating detection probabilities from point-count surveys. Auk 119:414-425.

Farnsworth, G.L., J.D. Nichols, J.R. Sauer, S. G. Fancy, K.H. Pollock, S.A.Shriner, and T.R. Simons. In Press. Statistical approaches to the analysis of point count data: A little extra information can go a long way. *In* Ralph, C.J., and T.D. Rich (Eds.). Bird

Conservation Implementation and Integration in the Americas. USDA Forest Service Gen. Tech. Rep. PNW-GTR-191.

McCallum, D. A. In Press. A conceptual guide to detection probability for point counts and other count-based survey methods. *In* Ralph, C.J., and T.D. Rich (Eds.). Bird Conservation Implementation and Integration in the Americas. USDA Forest Service Gen. Tech. Rep. PNW-GTR-191.

Nichols, J. T., 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 117:393-408.

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 76:1–11.

Rosenstock, S. S., D. R. Anderson, K. M. Giesen, T. Leukering, and M. F. Carter. 2002. Land bird counting techniques: Current practices and an alternative. Auk 119:46-53.

Scott, T. A., P.-Y. Li, G. C. Greene, and D. A. McCallum. In Press. Singing rate and detection probability: An example from the Least Bell's Vireo (*Vireo belli pusillus*). *In* Ralph, C.J., and T.D. Rich (Eds.). Bird Conservation Implementation and Integration in the Americas. USDA Forest Service Gen. Tech. Rep. PNW-GTR-191.

Thompson, W. T. 2002. Toward reliable bird surveys: Accounting for individuals present but not detected. Auk 119:18-25.

Walker, B. L. 2000. The structure, use, and function of song categories in the Brewer's Sparrow. M.S. thesis. University of Montana, Missoula, Montana. 114 pages.