

In Defense of Indices: The Case of Bird Surveys

DOUGLAS H. JOHNSON,¹ *United States Geological Survey, Northern Prairie Wildlife Research Center, 204 Hodson Hall, 1980 Folwell Avenue, Saint Paul, MN 55108, USA*

ABSTRACT Indices to population size have come under increasing criticism in recent years, on the grounds that indices might not faithfully represent the entire population. Most criticisms involve surveys of birds, particularly those based on point counts, which is my focus here. A variety of quantitative methods have been developed to reduce the bias of point counts, such as distance sampling, multiple-observer surveys, and time-of-detection methods. I argue that these developments are valuable, in that they enhance understanding of the detection process, but that their practical application may well be limited, likely to intensive studies focusing on a small number of species. These quantitative methods are not generally applicable to extensive, multiple-species surveys. Although criticism of the thoughtless use of indices is welcome, their wholesale rejection is not. (JOURNAL OF WILDLIFE MANAGEMENT 72(4):857–868; 2008)

DOI: 10.2193/2007-294

KEY WORDS bird, census, detectability, distance sampling, index, population size, survey.

Indices have a long history of use in wildlife ecology and management. For example, the age ratio of birds harvested by hunters is taken as an index to recruitment the previous breeding season, the fraction of marked birds that return to a breeding site is viewed as an index to survival during the intervening period, and the success rate of hunters sometimes is treated as an index to the size of harvest. Many measures are used as indices to population size for various species, such as scats, tracks, burrows, visits to scent stations, and catch per unit effort. As early as 1941, Dice (1941) opined that for many practical uses it is sufficient to know the relative abundance of a particular species in different areas or at different times. Caughley (1977:17) described a number of advantages of indices and stated that “the majority of ecological problems can be tackled with the help of indices of density, absolute estimates of density being unnecessary luxuries.” For my purposes, an index is defined to be a variable that correlates strongly with abundance or density of a species in an area (Caughley 1977). Further, I restrict attention to indices of bird populations, especially those based on point counts, which have faced recent scrutiny.

An early trickle of criticisms of such indices (Burnham 1981) has recently morphed into a torrent (Johnson 1995; Thompson et al. 1998; Nichols et al. 2000; Anderson 2001, 2003; Williams et al. 2001; Yoccoz et al. 2001; Farnsworth et al. 2002, 2005; MacKenzie and Kendall 2002; Rosenstock et al. 2002; Ellingson and Lukacs 2003; Norvell et al. 2003; White 2005; Buckland 2006; Kissling and Garton 2006). These and other critics correctly point out that indices are, well, indices; they reflect only the portion of the population that is counted, and that portion may not be constant. Variation in that portion diminishes the value of indices.

Typically the issue is cast in the equation,

$$E(C) = pN,$$

where $E(C)$ is the expected count of some species made during a survey, N is the true number of that species in the

surveyed area at the time of the survey, and p is the proportion of the true number that is recorded (e.g., Nichols 1992, Barker and Sauer 1995, Johnson 1995).

In recent years many authors have emphasized that variation in C reflects not only variation in N but also variation in p , which often is termed detectability. For example, when using an index C of a population N and comparing indices on different occasions (or locations), and the detection probabilities p vary, then the difference in indices is

$$\begin{aligned} C_1 - C_2 &= p_1 N_1 - p_2 N_2 \\ &= p_1(N_1 - N_2) + (p_1 - p_2)N_2, \end{aligned}$$

where the subscript denotes occasion (Johnson 1995). Hence the difference in indices reflects both difference in true population size and difference in detectability. Also, variation in C incorporates variation in both N and p :

$$\text{Var}(C) = p^2 \text{Var}(N) + N^2 \text{Var}(p) + \text{Var}(N) \text{Var}(p),$$

if N and p are independent. If N and p covary, the situation is even further complicated. For an effective index, then, we need p to be independent of N and $\text{Var}(p)$ to be small. Alternatively, the ratio of indices equals the ratio of the counts times the ratio of the detectability values:

$$C_2/C_1 = (N_2/N_1) \times (p_2/p_1).$$

Hence, inequality of the detectability values biases the ratio of the counts as an estimator of the ratio of the population sizes.

Much recent work has focused on estimating detectability, p . It has long been recognized that detectability is lower for animals farther away from the observer than for nearby animals, so considerable effort has been expended on distance sampling to account for that effect. A second remedy is based on framing the detections in a mark-recapture context and employing the tools available for that problem (e.g., Otis et al. 1978). A mark-recapture approach can be taken either with multiple observers (treating each observer as a trapping occasion) or by splitting the time

¹ E-mail: Douglas_H_Johnson@usgs.gov

period into segments (treating each segment as a trapping occasion).

I argue here that, although the criticisms of indices are largely valid, the cures proposed for the disease may be little or no better than the original ailment. The suggested treatments for indices have their own shortcomings. Advocates of treatments typically recognize, but may not fully appreciate, the consequences of those shortcomings. My specific objectives are to 1) review the criticism of indices, specifically bird counts, and 2) evaluate the assumptions, requirements, and applications of proposed alternatives.

AVAILABILITY VERSUS PERCEPTIBILITY

Marsh and Sinclair (1989) distinguished 2 sources of bias associated with nondetection of individuals, which can be considered the components of detection: availability and perceptibility. Availability represents a bird being possible to detect; that is, it is visible or it emits an audible cue such as a song during the count period. Perceptibility incorporates the conditional detection by an observer of that bird, given that it was available to be detected. Availability largely depends on the behavior of the birds, whereas perceptibility is primarily a feature of the observer and the survey conditions. Distinguishing these 2 components is useful conceptually because it emphasizes that the effort focused on perceptibility involves only the latter component, and birds not available for detection still will go unrecorded and unaccounted regardless of any adjustments to account for imperfect perceptibility. The distinction between availability and perceptibility is especially relevant if a population contains individuals that differ with respect to availability.

In reality, however, the distinction between availability and perceptibility is imperfect and depends on many factors, most notably the ability of the observer. If a bird near an observer sings a loud song during the observation period, it clearly is available to be detected. But what about a partial song or a chip note? Some observers will be able to hear that cue and identify the species, but others will not. Likewise, a quiet song given under windy conditions may or may not render the bird available, depending again on the observer. So availability, too, is in the eye (or ear) of the beholder. The distinction between availability and perceptibility is useful conceptually but can break down in practice.

INFLUENCES ON DETECTABILITY

Detectability varies in relation to many factors and it is useful to consider specifically if these factors influence availability, perceptibility, or both (Table 1). Moreover, species vary in regard to both availability and perceptibility.

As one example of imperfect availability, Scott et al. (2005) found for the least Bell's vireo (*Vireo bellii pusillus*) that singing occurred in bouts averaging about 12 minutes, interspersed with quiet periods averaging 6 minutes. Scott et al. (2005) estimated that count periods ≥ 10 minutes were necessary to obtain availability values of 70%. Furthermore,

Diefenbach et al. (2007) found that availability of marked populations of grasshopper sparrows (*Ammodramus sava-narum*) and Henslow's sparrows (*Ammodramus henslowii*) was low during the period when birds are usually surveyed. Availability (i.e., the bird sang and was visible) for Henslow's sparrows fell from 39% (for 5-min counts) and 44% (10-min counts) during late May–mid June to $<10\%$ for either count length in late June. Grasshopper sparrows showed no seasonal decline, but availability was only 10% during 5-minute counts and 19% during 10-minute counts.

McCallum (2005) noted that perceptibility depends on the conspicuousness and abundance of cues. Conspicuousness depends on amplitude of the vocalization, auditory acuity of the observer, attentiveness of the observer, and masking of focal sounds by other sounds. Simons et al. (2007) found under controlled experimental conditions that errors in bird detections were influenced by several factors. One such influencing factor was ambient noise; proportions of birds heard decreased by 28% under breezy conditions, by 41% in the presence of additional background birds, and by 42% with the addition of 10 dB of white noise. On unlimited-radius counts, observers on average undercounted birds, but proportions of actual totals ranged from 81% to 132% among observers. With limited-radius counts (50 m), observers overestimated total population by 17% to 122%. In the same experimental setting, Alldredge et al. (2007b) found that detection probabilities varied not only by distance, singing rate, bird species, and observer, but also by some interactions involving these variables.

McShea and Rappole (1997) found that singing birds within 50 m of observers had an average song rate only half that of singing birds beyond 50 m, which further confounds attempts to estimate density from counts, because availability, as well as perceptibility, varies by distance. Also, Bye et al. (2001) noted significant differences in the numbers of birds recorded with or without the presence of a human observer. The effect of the observer was not even consistent among species; some species were recorded more often and others less often when an observer was present.

APPROACHES TO OVERCOMING IMPERFECT DETECTABILITY

Distance Sampling

Much effort has been devoted to using distance sampling to estimate p because an obvious influence on detection probabilities is the proximity of the bird to an observer. The fundamental assumptions underlying distance sampling are that 1) all birds at zero distance from the observer are detected and 2) detectability declines monotonically with distance (e.g., Buckland et al. 1993). By fitting a detectability curve, it is possible to estimate from the number of birds detected how many birds were actually in the area under observation. Assumption 1 can be relaxed in certain circumstances. Other assumptions needed for distance sampling are that 3) birds are detected at the location they were when the observer began—there is no movement in response to the observer, 4) distance estimates are accurate,

Table 1. Features that affect the availability of birds to be detected or the perceptibility of birds, and example references.

Feature	Example references
Features affecting availability	
Duration of survey period	Granhölm 1983, Johnson 1995, McCallum 2005, Diefenbach et al. 2007
Time of yr, phenology	Best 1981, Wilson and Bart 1985, Diefenbach et al. 2007
Reproductive status	Diehl 1981, Wilson and Bart 1985
Time of day	Robbins 1981
Presence of observer	Bye et al. 2001
Distance from observer	Reynolds et al. 1980, McShea and Rappole 1997
Weather conditions	Anderson and Ohmart 1977, Emlen and DeJong 1981
Density of conspecifics	Diehl 1981, Bart and Schoultz 1984, Verner 1985
Features affecting perceptibility	
Distance from observer	Emlen 1971, Reynolds et al. 1980
Attenuation of signals	McCallum 2005
Habitat features	Richards 1981
Masking of cues by ambient noise	McCallum 2005, Simons et al. 2007
Observer skills	Cyr 1981, Emlen and DeJong 1981, Ramsey and Scott 1981, Sauer et al. 1994
Weather conditions	Anderson and Ohmart 1977
Conspicuousness of bird cues	McCallum 2005
Density of birds	Verner 1985

and 5) birds are distributed independently of the transect line or survey point. Another requirement is that numbers of detections are adequate to estimate a detectability function.

Although distance sampling is ideally suited for certain situations, such as tortoise surveys in which distance from the observer is the primary factor influencing detection, the role of distance sampling for birds has been somewhat controversial (e.g., Hutto and Young 2002, 2003; Ellingson and Lukacs 2003). The requirement for a large number of detections to estimate a detectability function is one concern. Standard recommendations are for 60–100 detections per species (Buckland et al. 2001, Rosenstock et al. 2002), which basically eliminates the use of distance sampling for all but the commonest species, which typically are of lesser conservation interest. Often observations from different species are combined, based on the lack of statistically significant differences between their detection functions. Although this practice may result in adequate sample sizes, its justification requires acceptance of the null hypothesis of no difference between the species. One cannot prove that no difference exists, and attempting to do so is especially problematic with small sample sizes—the reason that pooling is done in the first place. The risk is that species are pooled because they are uncommon, not because their true detectability functions actually are similar. It is better to use equivalence testing, in which the burden of proof is shifted from looking for differences to seeking evidence of similarity, or to use model averaging to appropriately reflect uncertainty about the constancy of detectability values (MacKenzie and Kendall 2002).

Rosenstock et al. (2002) recognized that pooling species must be approached cautiously. Pooling should be based on similarity of features that influence detectability, such as size, behavior, vocalization type and pattern, and micro-habitat use, which is very challenging. Diefenbach et al. (2003) further found that, when sample sizes were adequate (>50), detection functions varied among observers, which suggests that the lack of significant differences between

detection functions was due to small sample sizes, not to the lack of real differences.

If distance is not recorded accurately, estimates based on distance sampling will be markedly affected (as will conclusions from the mark–recapture approaches that use distance as a covariate). The ability of observers to accurately gauge distances to singing birds is highly questionable. Alldredge et al. (2007*b*) asked experienced observers to estimate distances to hidden loudspeakers playing songs of known bird species. Conditions were much more ideal than in real situations; the experiment was conducted in winter, when trees were without leaves that would attenuate and reverberate songs, and observers were allowed ample time to hear a clear signal under quiet ambient noise conditions. Nonetheless, errors in distance estimation were generally large, even after repeated training. Overestimation of distances (typically a doubling) was especially pronounced when the sound source faced away from the observers. Also, errors varied by bird species and individual observer but not by height of sound source in the experiment; the effect of height, however, could be stronger under actual field conditions. Further, errors varied inconsistently with distance, on average negative at some distances and positive at others. The clear message from Alldredge et al. (2007*b*) is that distance estimation is not as straightforward as it might appear.

Another concern with distance sampling involves heterogeneity of detectability. Detectability varies in response to many influences, such as date, phenology, time of day, skills of the observer, weather conditions, habitat features, and behavior of the birds (Table 1). Does that mean, then, that detectability functions needs to be estimated for each combination of these effects? That, of course, would be impossible.

The assumption that all birds at distance zero are detected seems straightforward and easy to satisfy but, in fact, is complicated. Meeting the assumption means that birds need to be detected prior to any movement that they make in

response to the observer. Hence, an observer headed to a location to make a point count must determine the location of birds in the vicinity of that point before arriving there. Also, birds might move in response to an observer approaching the point; if the observer is unaware of initial locations, a biased detection function will result. Some authors have advocated waiting at the point for several minutes before beginning the count (e.g., Ellingson and Lukacs 2003). In addition to the extra time this step would take, it is biological folly to believe birds would return to their original locations while the observer is there. Nelson and Fancy (1999) also noted that such a practice likely still would underestimate bird density near the point. Movement of birds, although generally presumed to be away from an observer, in some species could be toward the observer (e.g., Emlen 1971).

Other concerns about distance sampling have been voiced. Hutto and Young (2002) suggested that detection functions based on repeated surveys could be biased. For example, if one point in a grassland that is surveyed multiple times has a shrub, say, 32 m away from the survey point, there is a good chance that birds will be seen perched on that shrub regularly, resulting in an excess of detections at 32 m. Ellingson and Lukacs (2003) argued that an appropriately randomized design would eliminate that problem, but the reality is that, once a design is selected, randomly or otherwise, it then is fixed, and repeated sampling will indeed result in nonrepresentative results. The problem of nonrepresentative distance estimates can be reduced to some extent, however, by binning the distance data, which, as Hutto and Young (2003) noted, leads to fewer but larger errors.

Another concern with distance sampling is the black box approach taken by many users when analyzing results from distance sampling. Most users rely on Program DISTANCE (Laake et al. 1996, Thomas et al. 2006) for their analysis. In such papers, authors write something like, "data were analyzed using Program DISTANCE. . ." Program DISTANCE indeed is a sophisticated and powerful tool for analyzing distance-sampling data, but such cursory statements about its use hide more than they reveal. Program DISTANCE fits certain key functions, such as half-normal, uniform, and hazard, while also incorporating some parameter adjustments such as cosine or polynomial (Laake et al. 1996, Thomas et al. 2006). Do users truly understand what these are? In his analysis of his own data, Buckland (2006) sometimes chose models different from the one selected by Program DISTANCE based on the lowest Akaike's Information Criterion (AIC) value. Further, decisions made about truncating observations beyond some distance, and grouping observations into intervals, can affect the results, often dramatically. Are users cognizant of the impact of their decisions? Also, AIC usually is used to select model(s) of detectability. Akaike's Information Criterion indicates how well one model fits the data versus another model, but it is silent about the correctness of any model. Most users give no indication that they used the goodness-of-fit procedures available in Program DISTANCE.

Multiple-Observer Methods

The second approach to estimating detectability involves the use of ≥ 2 observers. Multiple-observer methods essentially provide replicate observations so that the powerful machinery of mark-recapture methodology can be used. Three major methods have been advocated: those with 2 dependent observers (Cook and Jacobson 1979, Nichols et al. 2000), 2 independent observers (Caughley 1974, Henny et al. 1977, Magnasson et al. 1978, Pollock et al. 2002), and several independent observers (Allredge et al. 2006).

Nichols et al. (2000) expanded and popularized the 2-dependent-observer approach of Cook and Jacobson (1979). In brief, at each point one individual serves as the primary observer and the other individual as the secondary observer. The primary observer indicates to the secondary which birds he or she detects. The secondary observer notes those birds, as well as any others that the primary observer missed. Individuals alternate roles as primary and secondary observers from one point to the next (Nichols et al. 2000, in press). The 2-dependent-observer method requires short time intervals to assure a closed population.

Treating 2 observers as independent relies on more challenging logistical considerations in the field but affords the opportunity to use the resulting data as if they emanated from a mark-recapture study. Each observer must record birds independently of any cues that the other observer may give when detecting a bird. After the count period, the 2 observers attempt to match the birds that both of them detected. Detections by one observer are treated as the first capture occasion and detections by the other observer as the second. If the observers agree on the birds that both of them saw, then those birds are considered recaptures of marked birds. With these data, Lincoln-Petersen estimators for closed populations can be used (Otis et al. 1978). Two models can be considered: Model M_0 , in which detectability is the same for both observers, and Model M_{obs} , in which detectability varies between observers.

Allredge et al. (2006) included the following assumptions for the independent-observer methods: 1) observations among observers are independent, 2) observers are able to accurately determine recaptures (i.e., there are no matching errors), 3) the detection probability for each species is constant for each observer, and 4) there are no undetected movements of birds into or out of the fixed-radius plot. Studies in experimental situations have found that matching detections is error-prone (Nichols et al., in press).

Allredge et al. (2006) applied the 2-independent-observer method to 4 taxa (3 species plus 2 vireo species combined) recorded along hiking trails in the Great Smoky Mountains National Park and found the best model (lowest AIC value) for the 3 individual species to be M_0 , in which both observers had the same detectability. A model (M_0^d) in which detectability varied with distance, but in the same way for both observers, was the best model for the vireos combined and was competitive for the other species. Estimates of population size of each taxon for both models

Table 2. Abundance estimates for the 2-independent-observer models from Alldredge et al. (2006). Vireos are combined blue-headed vireo (*Vireo solitarius*) and red-eyed vireo (*V. olivaceus*). Birds detected are the total number recorded by the 2 observers. The detectability estimate is based on Model M_0 .

Species	Birds detected	\hat{N} under M_0	\hat{N} under M_0^d	Detectability
Ovenbird (<i>Seiurus aurocapilla</i>)	72	73	73	0.90
Scarlet tanager (<i>Piranga olivacea</i>)	45	48	48	0.77
Tufted titmouse (<i>Baeolophus bicolor</i>)	31	32	32	0.85
Vireos	89	94	94	0.77

were only slightly higher (<7%) than simply the number of birds detected (Table 2).

Having 4 observers simultaneously but independently record data at each point allows an even richer array of models to be fit to the data. Alldredge et al. (2006) considered 4 relevant models appropriate with essentially 4 trapping occasions (Otis et al. 1978): M_0 , in which detectability is the same throughout; M_{obs} , in which detectability varies by observer; M_h , in which there is heterogeneity among the birds in their detectability; and $M_{obs,h}$, in which detectability varies both by observer and by bird.

Alldredge et al. (2006) applied the 4-independent-observer models to the data set described above (Table 3) and found M_0^d (detectability varies only by distance) to be best for ovenbirds and M_h with 2 groups (M_{2h}) best for the other taxa. Model M_{2h} allows heterogeneity among birds by assuming the population of each taxon consists of a group of highly detectable individuals and a second group that are less detectable (Norris and Pollock 1996, Pledger 2000). The other models examined were deemed unacceptable based on AIC values.

Alldredge et al. (2006) identified 2 concerns with multiple-observer methods: correct matching of detections and independence of detections among observers. Kissling and Garton (2006) also recognized the problem of matching detections when >1 observer is involved. Alldredge et al. (2006) suggested that independence would most readily be attained when detections are primarily auditory and there are many detections at each survey point. I suggest that those conditions also would make matching most challenging. Further, it seems likely that the presence of multiple observers would influence bird behavior even more than would one observer.

The limited practicality of 4-observer point counts was recognized when Alldredge et al. (2006) noted 4-observer point counts would be most useful in research situations examining the detection process. Fletcher and Hutto (2006) indicated that multiple-observer approaches account for biases due to observers and visibility but not to distance. In addition, multiple-observer methods, as distance sampling, estimate only the available birds; that is, multiple-observer methods account only for perception problems (McCallum 2005; Diefenbach et al. 2007; Nichols et al., in press).

Time-of-Detection Methods

Multiple-observer methods convert point-count sampling into a mark-recapture problem by treating the observers

analogously to trapping occasions. Another approach splits the time interval of a point count into periods and treats each period as a trapping occasion, thus allowing closed-population mark-recapture methodology to come into play. Removal methods (Farnsworth et al. 2002) as well as more general mark-recapture methods (Alldredge et al. 2007a) have been adapted to point-count surveys.

To use the removal method proposed by Farnsworth et al. (2002), observers divide the count period into ≥ 3 intervals, which may be of unequal length, and keep track of the birds that are first detected in each interval. Once a bird is detected, no further information is recorded about that bird; it is considered removed from the population of birds available to be detected. From such counts Farnsworth et al. (2002) developed 2 estimators of detectability. One estimator allows 2 groups of birds, one with high detectability and one with low detectability; the other estimator assumes that all birds are equally detectable. For the first estimator, it is assumed that all highly detectable birds are recorded during the first time interval (if there are ≥ 4 time intervals, then the detectability of this group can be estimated rather than assumed to be one). From the counts in each of the ≥ 3 time intervals, Program SURVIV (White 1983) is used to estimate the detectability probability for the other group. The other estimator basically assumes that all birds have uniform detectability. Hence it is a subset of the heterogeneous model and can be tested against it with any particular data set. Once detection probabilities are estimated, with either estimator, abundance can be estimated in a straightforward manner.

Farnsworth et al. (2002) were the first to explicitly include availability as an independent parameter in their model and indicated that their model accounted for availability as well as detectability. Seemingly to ensure their model accounted for both, Farnsworth et al. (2002) restricted attention to species with nonnegligible probabilities of singing during the period of the point count. Basically then, Farnsworth et al. (2002) assumed all birds were available for detection. Farnsworth et al. (2002) recognized that their proposed method, which requires long (e.g., 10-min) count periods, might not be suitable for wide-ranging species and could also lead to double-counting of individuals. The fact that birds often sing in irregular bouts, rather than consistently, poses a challenge for the removal method (McCallum 2005).

Alldredge et al. (2007a) generalized the time-of-detection method by incorporating not only the initial detections of all birds but also subsequent detections. The generalized time-

Table 3. Abundance estimates for the 4-independent-observer models from Allredge et al. (2006). Vireos are combined blue-headed vireo (*Vireo solitarius*) and red-eyed vireo (*V. olivaceus*). Birds detected are the total number recorded by the 4 observers.

Species	Birds detected	\hat{N} under M_0	\hat{N} under M_{2h}
Ovenbird (<i>Seiurus aurocapilla</i>)	81	81	83
Scarlet tanager (<i>Piranga olivacea</i>)	55	56	63
Tufted titmouse (<i>Baeolophus bicolor</i>)	39	39	45
Vireos	111	112	134

of-detection method requires that observers carefully track all birds throughout the count period but allows a richer suite of closed-population mark-recapture models (Otis et al. 1978) to be applied to the resulting data. Such models allow for variation associated with 1) the time interval, 2) whether a bird had been detected already, and 3) individual differences in detectability among birds. (Models with all 3 sources of variation cannot be fit without several constraining assumptions, however.) At least 4 intervals are needed to fit models with heterogeneity. In addition, Allredge et al. (2007a) incorporated covariates, such as distance, in the models.

Allredge et al. (2007a) evaluated their method in 2 situations. In one, the selected model led to an estimate only 5% higher than the number of birds actually observed. In the other, estimates for 4 species ranged from 5% to 16% higher than the number observed, which appear to be modest differences considering the additional effort expended.

As did Farnsworth et al. (2002), Allredge et al. (2007a) noted some drawbacks to their method; in particular, the method could lead to biased estimates for wide-ranging species, species that sing infrequently, and in situations with many birds and many species. Contrary to expectations, in a controlled setting the time-of-detection method seemed to estimate only available birds (Nichols et al., in press). Also, observers need to be uniformly attentive during all time intervals; in ≥ 1 operational situation, observers recorded many birds in the early interval but then had a lull, which was followed by an increase in the final interval (B. Mitchell, National Park Service, personal communication).

Methods Involving Repeated Counts

If the same sites (points or transects) are surveyed repeatedly and the population at each site is assumed closed, certain methods can be used to estimate the size of each population. The simplest estimator is the maximum of all the counts at each site. That value, of course, is a minimal estimate of the actual population; its bias will depend on the number of survey occasions, detectability of birds, and their behavior (e.g., some birds may remain inconspicuous if more dominant birds are active). Use of the maximum of repeated counts is common in surveys of lekking species such as grouse (e.g., Walsh et al. 2004).

A somewhat more sophisticated method is the bounded-count estimator (Robson and Whitlock 1964). The

fundamental assumption of this estimator is that the counts are distributed uniformly in the interval $[0, N]$, where N is the true population size at a site. This assumption is equivalent to assuming that the detection rates p are distributed uniformly in the interval $[0, 1]$. A few studies with actual and simulated data suggested that bounded-count estimates were better than maximum counts but often remained inaccurate (Walsh et al. 2004, Johnson et al. 2007).

Royle (2004) recently developed another estimator applicable to situations in which birds are counted multiple times at multiple sites. The estimator is based on a hierarchical model with the true population size N assumed to vary among sites according to a suitable distribution such as Poisson or Negative Binomial, possibly influenced by known explanatory variables. The model for the observed counts at a site on a particular occasion is specified conditional on the local abundance (N), and is assumed to be a realization of a binomial event $Bi(N, p)$, where detectability p might vary in relation to other explanatory variables. This model has performed well in some simulations (Royle 2004) and field applications (Kéry et al. 2005), even when the same explanatory variables influenced population size and detectability (Kéry 2008). The model can be analyzed using conventional methods based on integrated likelihood or Markov chain Monte Carlo; Kéry (2008) provided some R and WinBUGS code for his examples. Royle's (2004) method appears promising and warrants more extensive evaluation.

Combination Methods

It is possible to combine certain methods. A considerable literature on such approaches was noted by Nichols et al. (in press), but only recently have combination methods been applied to bird surveys. For example, Farnsworth et al. (2005) illustrated a method that combined distance sampling with removal sampling. Farnsworth et al. (2005) estimated detectability by dividing the count circle into ≥ 2 concentric rings and assigning the first detection of an individual to a particular ring. One stated assumption of the method is that birds do not move between rings during the count period, which would seem to require fairly short count periods. Another assumption is that all birds of a species have the same probability of vocalizing, which does not change during the count period. Farnsworth et al. (2005) suggested that having > 2 distance categories and time intervals would allow the latter assumption to be relaxed.

The requirement that a count essentially be taken at an instant in time (Buckland 2006) potentially poses a problem for hybrid methods that rely on activity during the count period to estimate availability (McCallum 2005). McCallum (2005) suggested the possibility of taking snapshots (instantaneous counts) during each of the time segments, thus providing estimates of availability without compromising the estimate of detectability.

Kissling and Garton (2006) developed a method that combined double-observer with distance sampling and applied it to breeding birds in southeastern Alaska. They

found densities estimated by that method to be very similar to those from a single-observer variable circular plot method (point counts with distance sampling). The method of Kissling and Garton (2006) gave smaller estimated standard errors, but whether those standard errors accurately reflect uncertainty in the estimates is an open question. Although combination methods are even more labor-intensive and difficult to apply than single methods, the approach does allow, at least conceptually, the separation of the 2 components of detectability (McCallum 2005).

Double-Sampling

Double-sampling basically involves relating 2 types of surveys (one accurate but expensive, another less accurate and less expensive), rather than adjusting a single survey by using ancillary information (e.g., distance, time to detection) collected as part of the extensive survey. There is a rich statistical literature available describing a variety of methods for estimating the relation between results from the 2 types of surveys, which include ratio estimation and regression estimation (e.g., Cochran 1977, Thompson 2002). Double-sampling can be viewed as a way to calibrate an index. Calibration requires knowledge of the true population size, but the reason indices are advocated is because the true population size is difficult to estimate (Lancia et al. 1994).

Double-sampling currently is used in a variety of bird-monitoring situations, notably for adjusting results from extensive aerial surveys with results from more intensive ground surveys (e.g., Jolly 1969, Smith 1995). Eberhardt and Simmons (1987) investigated the method and indicated that double-sampling offered the possibility of strengthening the use of population indices. Bart and Earnst (2002) discussed the methodology and offered some sample size formulas. Unlike other methods I discussed, double-sampling requires an independent and accurate census of (or, as suggested by Pollock et al. [2002], estimates of detectability on) a subset of sample units on which surveys (e.g., point counts) are conducted.

My concern with double-sampling in avian surveys such as point counts is that, because detectability varies in response to so many influences (Table 1), the relation between 2 types of surveys likely varies dramatically as well, so estimating a general relationship is challenging (Dice 1941). For example, if detectability varies according to habitat type, it would seem necessary to develop a calibration factor for each habitat type or to include habitat type in some model for the calibration factor. In addition, double-sampling to obtain accurate estimates on a subset of sample units is labor-intensive (Farnsworth et al. 2005).

ISSUES WITH BIRD SURVEYS

Heterogeneity in Detectability

Heterogeneity in detectability among individual birds is a particularly thorny issue (Link 2003). For approaches that use closed-population mark-recapture methodology, models M_{th} , M_{bh} , M_{th} , and M_{bth} of Otis et al. (1978) involve heterogeneity. There are 2 general ways that heterogeneity has been addressed in such situations. The first, due to

Burnham and Overton (1978), is a jackknife estimator based on the recapture frequencies f_j , where f_j is the number of individuals captured exactly j times ($j = 1, \dots, k$, where k is the no. of trapping occasions or equivalent events). The problem is to estimate f_0 , the number of individuals not captured; from this the total population N is estimated as

$$\hat{N} = \hat{f}_0 + f_1 + f_2 + \dots + f_k.$$

Link (2006) showed that Burnham and Overton's (1978) estimator approximates the result of extrapolating to zero the unique polynomial that fits observed detection frequencies (f_j).

The second way that heterogeneity has been treated is by assuming that the heterogeneous population consists of m subpopulations, each of which is homogeneous (Norris and Pollock 1996, Pledger 2000). Most often the case $m = 2$ has been considered, so it is necessary to estimate the detection probabilities (p_1, p_2) for each subpopulation and the fraction of the total population in each subpopulation ($w_1, w_2 = 1 - w_1$).

Heterogeneity is not easily addressed, however. Relevant to the jackknife approach, Cormack (1979:230) warned of the dangers of estimating f_0 when he suggested that "many different forms of distributions can be found to fit the truncated distribution, while giving rise to vastly different estimates for the zero-frequency class." Link (2003, 2006) clearly demonstrated with realistic examples that different but reasonable models may yield identical observed recapture frequencies but give very different estimates of f_0 and therefore of N . The fact that a model fits the observed data very well tells nothing about how well it estimates f_0 . As Link (2003) observed, the problem has not gone unnoticed, but its extreme consequences appear not to have been fully appreciated. J. D. Nichols (United States Geological Survey, personal communication) has suggested that the method often works in practice, despite its potential for serious error. Analogously for the approach assuming m subpopulations, it seems plausible that conclusions would depend markedly on the value of m assumed, and there is little to guide the selection of m , except that smaller values are more parsimonious.

Length of Count Period

Longer count periods increase the numbers of birds seen, especially birds with low availability or perceptibility. Conversely, longer periods may lead to biased estimates because birds not initially in the count area may enter during the count period (e.g., Johnson 1995). The longer the count period and the greater the speed of birds, the greater is the potential for a biased estimate of density (Scott and Ramsey 1981). Granholm (1983), for example, found that density estimates were 22–56% higher for a 10-minute count than for a 5-minute count. Longer counts may be more suited for sedentary species, especially those that are rare or inconspicuous (Scott and Ramsey 1981). Regardless, longer count periods induce a positive bias, especially for rapidly moving birds.

Allredge et al. (2006) recommended count durations be sufficiently long to ensure that all birds are available (i.e.,

sing at least once during the count) but short enough to minimize the effects of movements. I submit that is an impossible goal in a survey of multiple species.

Line Transect Versus Point Count

Line-transect methods generally are more efficient than point counts, at least in open or semi-open settings (e.g., van Riper 1981). Line-transect methods also tend to be more accurate (Järvinen 1978, Verner 1985). Van Riper (1981) and Buckland (2006) summarized relative advantages of point and line-transect surveys. Errors in estimating distances or violations of other assumptions generate more bias in point counts than in line-transect surveys (Verner 1985, Buckland et al. 2001). Also, it is more likely that birds close to a survey point will be missed than would birds close to the transect line. Although concerns about meeting assumptions are less pronounced for line-transect sampling than for point-count sampling, the latter method is the one most frequently advocated for standardized bird surveys because of other advantages (e.g., Ralph et al. 1993).

Field Evaluations

Several investigators have compared various techniques for surveying birds under field conditions. Rarely was the true population known, although estimates based on marked birds, active nests found, other more-intensive surveys such as territory mapping, or a combination of these are taken to be close to the true value (DeSante 1986, Tarvin et al. 1998, Buckland 2006).

DeSante (1986) compared results from variable circular plot counts to numbers based on spot mapping and nest monitoring for 21 species. He found that, even with 144 count stations in a 48-ha forest plot, errors in estimated density of common species based on point counts with distance sampling ranged from -67% to $+96\%$.

Nelson and Fancy (1999) had the unique opportunity to compare estimates of population size of a species that had been reintroduced and for which the exact number of birds was known. Despite that favorable situation, Nelson and Fancy (1999) had to pool detection distances with those from another site with similar habitat, which resulted in a very large data set comprising 3,146 observations. Further, because of the small number of detections on the study site, Nelson and Fancy (1999) were unable to adjust the detectability functions for effects due to observer, weather conditions, and the like. Nelson and Fancy (1999) conducted 4 surveys and with Program DISTANCE (Laake et al. 1994) obtained estimates that were off by $+24\%$, -34% , $+24\%$, and $+10\%$; had they used only detections from the study site, errors would have been much greater.

Tarvin et al. (1998) found that 30-m fixed-radius point counts were more effective than 100-m fixed-radius point counts or variable circular plots at relating to true number of blue jays (*Cyanocitta cristata*) on small study sites. Jones et al. (2000) investigated the accuracy of 50-m fixed-radius, 100-m fixed-radius, and variable-circular-plot point counts to estimate the actual density of breeding Cerulean warblers (*Dendroica cerulea*) and found that all methods provided

Table 4. Estimates of numbers of 4 species of birds based on 4 methods that involve distance sampling, in comparison to results from territory mapping (from Buckland 2006).

Method	Estimated density			
	Species 1	Species 2	Species 3	Species 4
Point count	103	58	52	129
Snapshot	90	22	60	102
Cue count	71	26	82	121
Line transect	64	26	68	107
Map territories	85	21	84	130

relative measures of density, but the variable-circular-plot technique provided the best absolute assessment of Cerulean warbler density.

Moore et al. (2004) compared removal and double-observer results for 17 bird species and indicated that the 2 methods gave similar results when detectability for observers was high. When detectability was low, however, removal estimates were biased low. Moore et al. (2004) nonetheless suggested the removal method generally should be preferred, in part because it requires only half as many observers, allows more sources of variation in detectability to be modeled, and is less influenced by unavailable birds. The double-observer method, in contrast, was more robust with respect to population closure and may be preferred when less-experienced observers are used.

Buckland (2006) evaluated 5 methods of bird surveys of 4 species in a small woodland-parkland in Scotland. One method was territory mapping, the most labor-intensive and considered the most accurate for an unmarked population (van Riper 1981). The other methods, all of which involved distance sampling, were 1) a 5-minute point count, 2) a snapshot method in which the observer identified and located birds for 3 minutes before the instant of the survey and could take whatever time needed after the survey to verify suspected locations of birds, 3) a cue-count survey in which the number of cues (i.e., songs) were recorded and divided by an estimated rate of cue production, and 4) a line-transect survey with distance estimation. No single method produced reasonably accurate estimates for all species (Table 4).

F. R. Thompson, III, and F. A. La Sorte (United States Forest Service and University of Missouri-Columbia, respectively; unpublished data) compared trend estimates based on raw counts, distance sampling with 3 distance intervals, removal models with 3 time intervals, and a Poisson model. As usual, actual population values were unknown. F. R. Thompson, III, and F. A. La Sorte (unpublished data) noted substantial agreement among the 4 estimators despite strong evidence that detectability was not constant. F. R. Thompson, III, and F. A. La Sorte (unpublished data) observed that results from the 4 methods were most similar when trends were pronounced, which are the situations in which trends are of most interest. Worthy of note is the fact that in most instances estimates from the distance-sampling method were outliers from the others.

DISCUSSION

Some critics have been unduly harsh on indices, for example stating that detectability must remain constant, which is an exaggeration (e.g., Rosenstock et al. 2002). In reality, for indices to serve in a monitoring situation, all that is needed is that the variation in detectability be substantially less than the variation in population size sought to be detected and that it be independent of population size. Anderson (2003:288) argued that, "In an important sense, index values are just 'numbers,' not really 'data'." I maintain that everything except the index values is an assumption or a consequence of an assumption. Caughley and Sinclair (1994:214) further suggested that the assumptions needed to develop population estimates from survey counts required a "leap of faith" by the investigators.

Other authors were unreasonably optimistic about the value of adjustments, for example by indicating that adjustments to raw counts completely correct those counts (e.g., Diefenbach et al. 2003, White 2005). As Engeman (2003:286) noted, "The same influences decried by Anderson (2001) as invalidating indexing methods also could destroy the validity of the assumptions for population estimation procedures." Thompson (2002) said that unadjusted counts will definitely not suffice for species of concern, most of which are uncommon. Yet uncommon species are the ones for which adequate sample sizes for appropriate adjustments will be difficult or impossible.

Some proposed remedies to the deficiencies of indices are themselves problematic. Alldredge et al. (2007c), for example, proposed that attention be restricted to species with songs that are easily identified and localized, which amounts to adapting the objectives of a bird survey to the shortcomings of the tool, rather than employing a tool suited to the objectives.

Sometimes indices truly are not adequate. Comparisons between habitats or between species, for example, are likely to violate the assumption that detectability rates are similar. In such situations, density estimates may well be most appropriate (Nelson and Fancy 1999). Clearly if there are known or expected patterns of variation in detectability, that knowledge should be incorporated in the analysis and interpretation of count data. For example, if an old field proceeds through succession to woodland, detectability of birds at that site would be expected to change, as would the bird community. In such cases, simple indices could well be misleading.

CONCLUSIONS

My review focused on surveys of birds, especially with point counts as indices to population size for monitoring purposes. My conclusions do not necessarily pertain to other indices, which need to be evaluated on their own merits or lack thereof. For example, the Minimum Number Known Alive estimator of population size, a traditional method used in mark-recapture studies of small mammals, has been clearly shown to be inferior to Jolly-Seber estimators (Nichols and Pollock 1983). I have tried to demonstrate that, although

unadjusted counts of birds are mere indices, procedures to adjust such counts to provide estimates of population size may themselves be fraught with problems. Whereas it is easy to show problems with indices, it is much harder to solve those problems realistically (Lancia et al. 1994). Recent work on the detection process has shown that necessary assumptions are difficult to meet in real field settings. Burnham (1981) recognized problems with unadjusted counts and the need to adjust them by their detection probabilities. Burnham (1981:325) argued that "this adjustment only requires appropriate detection distance data." The quarter-century that has passed since Burnham's (1981) argument unfortunately has shown the statement to be idealistic and unrealistic. Many of the procedures require considerably greater personnel and effort, yet may yield only slightly different results (Tables 2, 3, 4).

At present, no method of adjusting bird count data appears to be effective for large-scale, multi-species monitoring surveys. As Buckland (2006:356) noted, "[A] single method is unlikely to work well across all species." Recent efforts to test the efficacy of the methods "provide sobering conclusions about the ability of observers to detect and locate birds in auditory surveys" (Nichols et al., in press). The predominant method of surveying land birds, which relies on rapid, mostly aural, point counts, will seldom meet all the assumptions or requirements of any of the available methods (McCallum 2005). Several techniques, however, may prove useful in individual research studies involving a limited number of species, and in some instances, separating detections of birds into 2 distance categories has been useful (Newson et al. 2005, Cunningham and Johnson 2006).

Some recommendations seem in order. Users of indices, as well as adjusted counts, should be realistic in their expectations. For example, Barker and Sauer (1995) believed that point counts are of best use in early warning monitoring systems, but that population change estimates from the count data should be used as a basis for further, more definitive, study.

Clearly, users of any method should recognize that variability in detectability influences the results. Researchers should attempt to remove as much of that variability as is reasonable by design control (e.g., by restricting counts of birds to certain calendar periods and times of day). If appropriate, it may be useful to reduce the variability further through statistical adjustments as I discussed. But the costs of such adjustments need to be considered, and the adjusted results must be viewed still as approximations to the truth, that is, as indices themselves.

Recent research activities are worthwhile and should continue. Learning more about the detection process will be valuable not only for developing new methods but also for evaluating methods already available. Another potentially productive line of investigation involves assessing the relative roles of variation in population size, variation in availability and perceptibility, relative costs of unadjusted

versus adjusted count data, and sample sizes needed to identify real patterns in bird populations.

An additional topic worthy of study is the value of model-based, rather than sample-based, estimators of detectability (Samuel et al. 1987, White 2005). It may turn out that some simple yet adequate models of detectability could be developed. Model-based estimators would need to have wide generality and yet be based on a few readily measured variables.

Although defending the use of indices, I do not claim their superiority over adjusted estimates. Many of the concerns I raised about adjustments to indices apply equally, if not more so, to the use of unadjusted indices. A key question is whether or not the additional effort needed to adjust the counts is worthwhile; do, say, 20 adjusted estimates tell more about population trends than 50 index values? Also, do the adjustments give the user unwarranted confidence in their accuracy?

Until more sophisticated alternatives to indices have succeeded in real applications in a variety of settings, it is premature to advocate their adoption for extensive surveys (cf., Pollock et al. 2002). Such methods should be extensively field-tested under a variety of situations before recommended for wide adoption. It is appropriate to challenge users of indices to justify their applications. It is not appropriate to condemn the use of indices without offering suitable substitutes, or to disparage as Luddites the users of uncorrected count data.

ACKNOWLEDGMENTS

This paper was presented at the 2006 annual meeting of The Wildlife Society, in a symposium, "Objectives and Metrics for Monitoring Wildlife," sponsored by the Biometrics Working Group. I appreciate the invitation by M. Knutson and J. R. Sauer to participate in that symposium and the invitation by M. L. Morrison to publish the paper. I am grateful to S. T. Buckland, W. A. Link, J. D. Nichols, P. J. Pietz, K. H. Pollock, T. L. Shaffer, and T. R. Simons for comments on earlier drafts of this manuscript. Many reviewers, nonetheless, wanted to distance themselves far from my opinions and hoped to remain undetected.

LITERATURE CITED

- Allredge, M. W., K. H. Pollock, and T. R. Simons. 2006. Estimating detection probabilities from multiple-observer counts. *Auk* 123:1172–1182.
- Allredge, M. W., K. H. Pollock, T. R. Simons, J. A. Collazo, and S. A. Shriner. 2007a. Time-of-detection method for estimating abundance from point-count surveys. *Auk* 124:653–664.
- Allredge, M. W., T. R. Simons, and K. H. Pollock. 2007b. Factors affecting aural detections of songbirds. *Ecological Applications* 17:948–955.
- Allredge, M. W., T. R. Simons, and K. H. Pollock. 2007c. A field evaluation of distance measurement error in auditory avian point count surveys. *Journal of Wildlife Management* 71:2759–2766.
- Anderson, B. W., and R. D. Ohmart. 1977. Climatological and physical characteristics affecting avian population estimates in southwestern riparian communities using transect counts. Pages 193–200 in R. R. Johnson and D. A. Jones, editors. Importance, preservation and management of riparian habitat: a symposium. U.S. Department of Agriculture Forest Service General Technical Report RM-43, Fort Collins, Colorado, USA.
- Anderson, D. R. 2001. The need to get the basics right in wildlife field studies. *Wildlife Society Bulletin* 29:1294–1297.
- Anderson, D. R. 2003. Response to Engeman: index values rarely constitute reliable information. *Wildlife Society Bulletin* 31:288–291.
- Barker, R. J., and J. R. Sauer. 1995. Statistical aspects of point count sampling. Pages 125–130 in C. J. Ralph, J. R. Sauer, and S. Droege, editors. Monitoring bird populations by point counts. U.S. Department of Agriculture Forest Service General Technical Report PSW-GTR-149, Albany, California, USA.
- Bart, J., and S. Earnst. 2002. Double-sampling to estimate density and population trends in birds. *Auk* 119:36–45.
- Bart, J., and J. D. Schoultz. 1984. Reliability of singing bird surveys; changes in observer efficiency with avian density. *Auk* 101:307–318.
- Best, L. B. 1981. Seasonal changes in detection of individual bird species. Pages 252–261 in C. J. Ralph and J. M. Scott, editors. Estimating numbers of terrestrial birds. *Studies in Avian Biology* 6. Cooper Ornithological Society, Camarillo, California, USA.
- Buckland, S. T. 2006. Point-transect surveys for songbirds: robust methodologies. *Auk* 123:345–357.
- 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, United Kingdom.
- Buckland, S. T., K. P. Burnham, D. R. Anderson, and J. L. Laake. 1993. Distance sampling: estimation of abundance of biological populations. Chapman and Hall, London, United Kingdom.
- Burnham, K. P. 1981. Summarizing remarks: environmental influences. Pages 324–325 in C. J. Ralph and J. M. Scott, editors. Estimating numbers of terrestrial birds. *Studies in Avian Biology* 6. Cooper Ornithological Society, Camarillo, California, USA.
- Burnham, K. P., and W. S. Overton. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* 45:343–359.
- Bye, S. L., R. J. Robel, and K. E. Kemp. 2001. Effects of human presence on vocalizations of grassland birds in Kansas. *Prairie Naturalist* 33:249–256.
- Caughley, G. 1974. Bias in aerial survey. *Journal of Wildlife Management* 38:921–933.
- Caughley, G. 1977. Analysis of vertebrate populations. Wiley, London, United Kingdom.
- Caughley, G., and A. Sinclair. 1994. Wildlife ecology and management. Blackwell Science, Cambridge, Massachusetts, USA.
- Cochran, W. G. 1977. Sampling techniques. Third edition. Wiley, New York, New York, USA.
- Cook, R. D., and J. O. Jacobson. 1979. A design for estimating visibility bias in aerial surveys. *Biometrics* 35:735–742.
- Cormack, R. M. 1979. Models for capture–recapture. Pages 217–255 in R. M. Cormack, G. P. Patil, and D. S. Robson, editors. Sampling biological populations. International, Fairland, Maryland, USA.
- Cunningham, M. A., and D. H. Johnson. 2006. Proximate and landscape factors influence grassland bird distributions. *Ecological Applications* 16:1062–1075.
- Cyr, A. 1981. Limitation and variability in hearing ability in censusing birds. Pages 327–333 in C. J. Ralph and J. M. Scott, editors. Estimating numbers of terrestrial birds. *Studies in Avian Biology* 6. Cooper Ornithological Society, Camarillo, California, USA.
- DeSante, D. F. 1986. A field test of the variable circular-plot censusing techniques in a Sierran subalpine forest habitat. *Condor* 88:129–142.
- Dice, L. R. 1941. Methods for estimating populations of mammals. *Journal of Wildlife Management* 5:398–407.
- Diefenbach, D. R., D. W. Brauning, and J. A. Mattice. 2003. Variability in grassland bird counts related to observer differences and species detection rates. *Auk* 120:1168–1179.
- Diefenbach, D. R., M. R. Marshall, J. A. Mattice, and D. W. Brauning. 2007. Incorporating availability for detection in estimates of bird abundance. *Auk* 124:96–106.
- Diehl, B. 1981. Bird populations consist of individuals differing in many respects. Pages 225–229 in C. J. Ralph and J. M. Scott, editors. Estimating numbers of terrestrial birds. *Studies in Avian Biology* 6. Cooper Ornithological Society, Camarillo, California, USA.
- Eberhardt, L. L., and M. A. Simmons. 1987. Calibrating population

- indices by double-sampling. *Journal of Wildlife Management* 51:665–675.
- Ellingson, A. R., and P. M. Lukacs. 2003. Improving methods for regional landbird monitoring: a reply to Hutto and Young. *Wildlife Society Bulletin* 31:896–902.
- Emlen, J. T. 1971. Population densities of birds derived from transect counts. *Auk* 88:323–342.
- Emlen, J. T., and M. J. DeJong. 1981. The application of song detection threshold distance to census operations. Pages 346–352 in C. J. Ralph and J. M. Scott, editors. *Estimating numbers of terrestrial birds*. Studies in Avian Biology 6. Cooper Ornithological Society, Camarillo, California, USA.
- Engeman, R. M. 2003. More on the need to get the basics right: population indices. *Wildlife Society Bulletin* 31:286–287.
- Farnsworth, G. L., J. D. Nichols, J. R. Sauer, S. G. Fancy, K. H. Pollock, S. A. Shriner, and T. R. Simons. 2005. Statistical approaches to the analysis of point count data: a little extra information can go a long way. Pages 736–743 in C. J. Ralph and T. D. Rich, editors. *Bird conservation implementation and integration in the Americas: proceedings of the third international Partners in Flight conference 2002*. U.S. Department of Agriculture Forest Service GTR-PSW-191, Albany, California, USA.
- Farnsworth, G. L., K. H. Pollock, J. D. Nichols, T. R. Simons, J. E. Hines, and J. R. Sauer. 2002. A removal method for estimating detection probabilities from point-count surveys. *Auk* 119:414–425.
- Fletcher, R. J., Jr., and R. L. Hutto. 2006. Estimating detection probabilities of river birds using double surveys. *Auk* 123:695–707.
- Granhölm, S. L. 1983. Bias in density estimated due to movement of birds. *Condor* 85:243–248.
- Henny, C. J., M. A. Byrd, J. A. Jacobs, P. B. McLain, M. R. Todd, and B. F. Halla. 1977. Mid-Atlantic Coast osprey population: present numbers, productivity, pollutant contamination, and status. *Journal of Wildlife Management* 41:254–265.
- Hutto, R. L., and J. S. Young. 2002. Regional landbird monitoring: perspectives from the Northern Rocky Mountains. *Wildlife Society Bulletin* 30:738–750.
- Hutto, R. L., and J. S. Young. 2003. On the design of monitoring programs and the use of population indices: a reply to Ellingson and Lukacs. *Wildlife Society Bulletin* 31:903–910.
- Järvinen, O. 1978. Estimating relative densities of breeding birds by the line-transect method. II. comparison between two methods. *Annales Zoologici Fennici* 15:290–293.
- Johnson, D. H. 1995. Point counts of birds: what are we estimating? Pages 118–123 in C. J. Ralph, J. R. Sauer, and S. Droege, editors. *Monitoring bird populations by point counts*. U.S. Department of Agriculture Forest Service General Technical Report PSW-GTR-149, Albany, California, USA.
- Johnson, D. H., C. E. Braun, and M. A. Schroeder. 2007. The bounded-count method for analysis of lek counts. Pages 25–30 in K. P. Reese and R. T. Bowyer, editors. *Monitoring populations of sage-grouse*. College of Natural Resources Experiment Station Bulletin 88, University of Idaho, Moscow, USA.
- Jolly, G. M. 1969. The treatment of errors in aerial counts of wildlife populations. *East African Agriculture and Forestry Journal* 34:50–55.
- Jones, J., W. J. McLeish, and R. J. Robertson. 2000. Density influences census technique accuracy for Cerulean warblers in eastern Ontario. *Journal of Field Ornithology* 71:46–56.
- Kéry, M. 2008. Estimating abundance from bird counts: binomial mixture models uncover complex covariate relationships. *Auk* 125:in press.
- Kéry, M., J. A. Royle, and H. Schmid. 2005. Modeling avian abundance data from replicated counts using binomial mixture models. *Ecological Applications* 15:1450–1461.
- 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* 123:735–752.
- Laake, J. L., S. T. Buckland, D. R. Anderson, and K. P. Burnham. 1994. DISTANCE user's guide V2.1. Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, USA.
- Laake, J. L., S. T. Buckland, D. R. Anderson, and K. P. Burnham. 1996. DISTANCE 2.2. Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, Colorado, USA. <<http://www.ruwpa.st-and.ac.uk/distance/>>. Accessed 16 Aug 2007.
- Lancia, R. A., J. D. Nichols, and K. P. Pollock. 1994. Estimating the number of animals in wildlife populations. Pages 215–253 in T. A. Bookhout, editor. *Research and management techniques for wildlife and habitats*. The Wildlife Society, Bethesda, Maryland, USA.
- Link, W. A. 2003. Nonidentifiability of population size from capture–recapture data with heterogeneous detection probabilities. *Biometrics* 59:1123–1130.
- Link, W. A. 2006. On identifiability in capture–recapture models—reply. *Biometrics* 62:936–939.
- MacKenzie, D. I., and W. L. Kendall. 2002. How should detection probabilities be incorporated into estimates of relative abundance? *Ecology* 83:2387–2393.
- Magnasson, W. E., G. J. Caughley, and G. C. Sinclair. 1978. A double-survey estimate of population size from incomplete counts. *Journal of Wildlife Management* 42:174–176.
- Marsh, H., and D. F. Sinclair. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. *Journal of Wildlife Management* 53:1017–1024.
- McCallum, A. A. 2005. A conceptual guide to detection probability for point counts and other count-based methods. Pages 754–761 in C. J. Ralph and T. D. Rich, editors. *Bird conservation implementation and integration in the Americas: proceedings of the third international Partners in Flight conference 2002*. U.S. Department of Agriculture Forest Service GTR-PSW-191, Albany, California, USA.
- 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* 68:367–375.
- 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* 121:865–876.
- Nelson, J. T., and S. G. Fancy. 1999. A test of the variable circular-plot methods where exact density of a bird population was known. *Pacific Conservation Biology* 5:139–143.
- Newson, S. E., R. J. W. Woodburn, D. G. Noble, S. R. Baillie, and R. G. Gregory. 2005. Evaluating the Breeding Bird Survey for producing national population size and density estimates. *Bird Study* 52:42–54.
- Nichols, J. D. 1992. Capture–recapture models: using marked animals to study animal populations. *BioScience* 42:94–102.
- 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* 117:393–408.
- Nichols, J. D., and K. H. Pollock. 1983. Estimation methodology in contemporary small mammal capture–recapture studies. *Journal of Mammalogy* 64:253–260.
- Nichols, J. D., L. Thomas, and P. B. Conn. In press. Inferences about landbird abundance from count data: recent advances and future directions. *Environmental and Ecological Statistics*.
- Norris, J. L., and K. H. Pollock. 1996. Nonparametric MLE under two closed-capture models with heterogeneity. *Biometrics* 52:639–649.
- Norvell, R. E., F. P. Howe, and J. R. Parrish. 2003. A seven-year comparison of relative-abundance and distance-sampling methods. *Auk* 120:1013–1028.
- Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. *Wildlife Monographs* 72.
- Pledger, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics* 56:434–442.
- Pollock, K. H., J. D. Nichols, R. T. 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., G. R. Guepel, P. Pyle, T. E. Martin, and D. F. DeSante. 1993. *Handbook of field methods for monitoring landbirds*. U.S. Department of Agriculture Forest Service General Technical Report PSW-GTR-144, Albany, California, USA.
- Ramsey, F. L., and J. M. Scott. 1981. Tests of hearing ability. Pages 341–345 in C. J. Ralph and J. M. Scott, editors. *Estimating numbers of terrestrial birds*. Studies in Avian Biology 6. Cooper Ornithological Society, Camarillo, California, USA.
- Reynolds, R. T., J. M. Scott, and R. A. Nussbaum. 1980. A variable circular-plot method for estimating bird numbers. *Condor* 82:309–313.
- Richards, D. G. 1981. Environmental acoustics and censuses of singing birds. Pages 287–300 in C. J. Ralph and J. M. Scott, editors. *Estimating*

- numbers of terrestrial birds. *Studies in Avian Biology* 6. Cooper Ornithological Society, Camarillo, California, USA.
- Robbins, C. S. 1981. Effect of time of day on bird activity. Pages 275–286 in C. J. Ralph and J. M. Scott, editors. *Estimating numbers of terrestrial birds. Studies in Avian Biology* 6. Cooper Ornithological Society, Camarillo, California, USA.
- Robson, D. S., and J. H. Whitlock. 1964. Estimation of a truncation point. *Biometrika* 51:33–39.
- Rosenstock, S. S., D. R. Anderson, K. M. Giesen, T. Leukering, and M. F. Carter. 2002. Landbird counting techniques: current practice and an alternative. *Auk* 119:46–53.
- Royle, J. A. 2004. *N*-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60:108–115.
- Samuel, M. D., E. O. Garton, M. W. Schlegel, and R. G. Carson. 1987. Visibility bias during aerial surveys of elk in north-central Idaho. *Journal of Wildlife Management* 51:622–630.
- Sauer, J. R., B. G. Peterjohn, and W. A. Link. 1994. Observer differences in the North American Breeding Bird Survey. *Auk* 111:50–62.
- Scott, J. M., and F. L. Ramsey. 1981. Length of count period as a possible source of bias in estimating bird densities. Pages 409–413 in C. J. Ralph and J. M. Scott, editors. *Estimating numbers of terrestrial birds. Studies in Avian Biology* 6. Cooper Ornithological Society, Camarillo, California, USA.
- Scott, T. A., P.-Y. Lee, G. C. Greene, and D. A. McCallum. 2005. Singing rate and detection probability; an example from the least Bell's vireo. Pages 845–853 in C. J. Ralph and T. D. Rich, editors. *Bird conservation implementation and integration in the Americas: proceedings of the third international Partners in Flight conference 2002*. U.S. Department of Agriculture Forest Service GTR-PSW-191, Albany, California, USA.
- Simons, T. R., M. W. Alldredge, K. H. Pollock, and J. M. Wettrouth. 2007. Experimental analysis of the auditory detection process on avian point counts. *Auk* 124:986–999.
- Smith, G. W. 1995. A critical review of the aerial and ground surveys of breeding waterfowl in North America. U.S. Department of the Interior-National Biological Service Biological Science Report 5, Washington, D.C., USA.
- Tarvin, K. A., M. C. Garvin, J. M. Jawor, and K. A. Dayer. 1998. A field evaluation of techniques used to estimate density of blue jays. *Journal of Field Ornithology* 69:209–222.
- Thomas, L., J. L. Laake, S. Strindberg, F. F. C. Marques, S. T. Buckland, D. L. Borchers, D. R. Anderson, K. P. Burnham, S. L. Hedley, J. H. Pollard, J. R. B. Bishop, and T. A. Marques. 2006. DISTANCE 5.0. Research Unit for Wildlife Population Assessment, University of St. Andrews, United Kingdom. <<http://www.ruwpa.st-and.ac.uk/distance/>>. Accessed 16 Aug 2007.
- Thompson, S. K. 2002. *Sampling*. Second edition. Wiley, New York, New York, USA.
- Thompson, W. L. 2002. Towards reliable bird surveys: accounting for individuals present but not detected. *Auk* 119:18–25.
- Thompson, W. L., G. C. White, and C. Gowan. 1998. *Monitoring vertebrate populations*. Academic Press, San Diego, California, USA.
- van Riper, C., III. 1981. Summarizing remarks: comparisons of methods. Pages 217–218 in C. J. Ralph and J. M. Scott, editors. *Estimating numbers of terrestrial birds. Studies in Avian Biology* 6. Cooper Ornithological Society, Camarillo, California, USA.
- Verner, J. 1985. Assessment of counting techniques. *Current Ornithology* 2:247–302.
- Walsh, D. P., G. C. White, T. E. Remington, and D. C. Bowden. 2004. Evaluation of the lek-count index for greater sage-grouse. *Wildlife Society Bulletin* 32:56–68.
- White, G. C. 1983. Numerical estimation of survival rates from band-recovery and biotelemetry data. *Journal of Wildlife Management* 47:716–728.
- White, G. C. 2005. Correcting wildlife counts using detection probabilities. *Wildlife Research* 32:211–216.
- Williams, B. K., J. D. Nichols, and M. J. Conroy. 2001. *Analysis and management of animal populations: modeling, estimation and decision making*. Academic Press, San Diego, California, USA.
- Wilson, D. M., and J. Bart. 1985. Reliability of singing bird surveys: effects of song phenology during the breeding season. *Condor* 87:69–73.
- Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution* 16:446–453.

Associate Editor: Morrison.