ESTIMATING DETECTION PROBABILITIES FROM MULTIPLE-OBSERVER POINT COUNTS

Mathew W. Alldredge,[1,4] Kenneth H. Pollock,[2] and Theodore R. Simons[3]

[1]*Departments of Zoology and Biomathematics, Campus Box 7617, North Carolina State University, Raleigh, North Carolina 27695, USA;*
[2]*Departments of Zoology, Biomathematics, and Statistics, Campus Box 7617, North Carolina State University, Raleigh, North Carolina 27695, USA; and*
[3]*U.S. Geological Survey, North Carolina Cooperative Fish and Wildlife Research Unit, Department of Zoology, Campus Box 7617, North Carolina State University, Raleigh, North Carolina 27695, USA*

Abstract.—Point counts are commonly used to obtain indices of bird population abundance. We present an independent-observer point-count method, a generalization of the dependent-observer approach, based on closed-population capture–recapture methods. The approach can incorporate individual covariates, such as detection distance, to account for individual differences in detection probabilities associated with measurable sources of variation. We demonstrate a negative bias in two-observer estimates by comparing abundance estimates from two- and four-observer point counts. Models incorporating data from four independent observers were capable of accounting for this bias. Modeling individual bird differences in detection probabilities produced abundance estimates 15–21% higher than models that did not account for individual differences, in four out of five data sets analyzed. Although independent-observer methods are expensive and impractical for large-scale applications, we believe they can provide important insights into the sources and degree of perception bias (i.e., probability of detecting an individual, given that it is available for detection) in avian point-count estimates. Therefore, they may be useful in a two-stage sampling framework to calibrate larger surveys based on single-observer estimates. *Received 20 July 2004, accepted 5 January 2006.*

Key words: bird surveys, capture–recapture, detection probability, multiple observer, point counts.

Estimación de Probabilidades de Detección a Partir de Conteos en Puntos Hechos por Varios Observadores

Resumen.—Los conteos en puntos son empleados comúnmente para obtener índices de la abundancia de las poblaciones de aves. En este estudio presentamos un método de conteo en puntos para observadores independientes el cual es una generalización del enfoque de observadores dependientes basada en métodos de captura-recaptura para poblaciones cerradas. El enfoque puede incorporar covariables (ej. la distancia de detección) para tener en cuenta las diferencias en las probabilidades de detección entre individuos asociadas con fuentes de variación medibles. Comparando conteos hechos por dos y por cuatro observadores, demostramos que existe un sesgo negativo en los parámetros estimados con base en datos de dos observadores. Los modelos que incorporaban datos de cuatro observadores independientes tuvieron la habilidad de corregir este sesgo. En cuatro de los cinco conjuntos de datos analizados, al modelar las diferencias en las probabilidades de detección de las aves se obtuvieron parametros estimados de 15 a 21% mayores que los obtenidos mediante modelos que no tuvieron

[4]E-mail: mwalldre@unity.ncsu.edu

en cuenta las diferencias individuales. Aunque los métodos que emplean observadores independientes son costosos y poco prácticos para aplicaciones a gran escala, creemos que éstos pueden proveer información importante para los parámetros estimados con base en conteos en puntos en cuanto a las fuentes que causan el sesgo de percepción (i.e., la probabilidad de detectar un individuo, dado que está disponible para ser detectado) y en cuanto a la magnitud de éste. Por lo tanto, estos métodos podrían ser de utilidad si se incorporan en diseños con muestreos en dos etapas para calibrar censos más amplios basados en parámetros estimados a partir de censos hechos por un solo observador.

Point counts are used extensively as indices of bird diversity and abundance, and to assess habitat relationships and population response to environmental change or management (Ralph et al. 1995b, Thompson 2002). They are used across a spectrum of scales, from long-term continental-scale surveys such as the Breeding Bird Survey (Robbins et al. 1986, Sauer et al. 1997) to short-term site-specific studies (Ralph et al. 1995b).

There are fundamentally two approaches to abundance estimation using count data. The first approach generates an abundance index using a standardized approach to control for known sources of bias (e.g., weather, observer skill, time of day and season; Conroy 1996, Sauer et al. 1997, Williams et al. 2002). Comparisons of index counts across space require the assumption that the probability of detection is similar across sites, and comparisons of counts across time assume there is no systematic trend in detection rates over time. Assumptions of constant detection probability have long been questioned (Burnham 1981, Wilson and Bart 1985, Barker and Sauer 1995, Johnson 1995). The second approach uses statistical methods that estimate detection probabilities for various covariates, such as species, observers, and habitats (Nichols et al. 2000, Farnsworth et al. 2002, Rosenstock et al. 2002, Thompson 2002). Estimated detection probabilities are used to reduce the bias of unadjusted counts. For a complete discussion of the approaches to point-count surveys and the need to account for detection probabilities, see Thompson (2002).

The probability of detection has two components (Marsh and Sinclair 1989): the probability of being available for detection ($\hat{p}_a$; i.e., if detections are auditory, the probability that a bird sings during the count interval) and the probability of detecting a bird, given that it is available ($\hat{p}_d$). As with many of the other methods for estimating abundance, the multiple-observer method only estimates the probability of detection given

that an individual is available. For auditory point counts, this means that birds must sing during the count to be detected. Availability can be estimated, but most abundance estimates assume that all birds present at the point have some chance of detection during the count. Conducting surveys early in the breeding season, when singing rates are high, is one way to reduce bias associated with availability.

Here, we focus on multiple-observer methods of estimating detection probabilities from point-count surveys where detections are primarily auditory. Nichols et al. (2000) suggested that, if it were possible to maintain independence among observers, a completely independent-observer approach would provide more modeling flexibility and benefits over their dependent-observer approach. Because independent-observer approaches apply closed-population capture–recapture models, they are more efficient (smaller variance) than the dependent-observer approaches (Alldredge 2004), which rely on removal models (Seber 1982).

The standard multiple-observer point-count methods do not allow for differences in detection probabilities among birds, which is a very limiting assumption. For example, it is likely that detection probabilities decrease with increasing distance from the observer (Buckland et al. 2001). To meet the constant-detection-probability assumption with the multiple-observer method in this situation would require using a limited-radius point count. Of course, there are many sources of variation other than distance that would cause detection probabilities to differ among individual birds.

Among the key developments that we present are models that incorporate individual differences (heterogeneity) in detection probabilities. Individual heterogeneity indicates that each individual in the population has a possible unique capture (detection) probability.

Accounting for individual heterogeneity is important, because ignoring it will cause a negative bias in population estimates resulting from a positive bias in detection probability estimates (White et al. 1982, Johnson et al. 1986, Williams et al. 2002).

We first used covariate models (Johnson et al. 1986) to account for observable or measurable individual differences in detection probability (observable heterogeneity). These models are important because the covariates associated with differences in detection probabilities often have biological explanations. The significance of covariates can be tested using model selection procedures (Burnham and Anderson 2002). Detection distance (Buckland et al. 2001), habitat differences (Nichols et al. 2000), and singing rate (Wilson and Bart 1985, McShea and Rappole 1997) are examples of measurable covariates that may cause variations in detection probabilities.

Unfortunately, there are factors that cannot be observed or measured during point counts that may also affect detection probabilities. The second set of models that we use account for unobservable or unmeasurable individual differences in detection probability (unobservable heterogeneity). These factors include variation associated with age, social status, innate levels of activity, physical condition, and genetic factors. Because it is not possible to control for unobservable sources of variation that affect count data (Burnham 1981, Johnson 1995), models that incorporate these unobservable sources of heterogeneity are potentially very useful.

Our objectives here are (1) to describe an independent-observer method for two observers, providing an example that illustrates the use of covariates for estimating detection probabilities, and (2) to present a more general model using four independent observers that allows for observable and unobservable individual heterogeneity, providing an example using two-point finite-mixture models of heterogeneity. We separated the independent-observer models for two observers from the models using four independent observers to illustrate the effect of heterogeneity on detection probabilities and population estimates.

Methods

Field methods and data.—Example analyses are based on 70 point counts conducted by four observers along low-use hiking trails in Great Smoky Mountains National Park during June 1999. All observers had been conducting point counts on the study site for the previous six weeks, during which their identification and distance-estimation skills were periodically validated. Before each count, observers estimated a 50-m-radius circle using a laser rangefinder; they began the count immediately thereafter.

Observers conducted variable circular-plot 3-min point counts (Reynolds et al. 1980) between dawn and 1015 hours, only in good weather (no rain or excessive wind), which is consistent with the recommendations for point-count methodology detailed by Ralph et al. (1995a). During each 3-min count, observers recorded bird detections on an unlimited-radius plot, mapping the locations of all detections. Points were spaced a minimum of 250 m apart, and the location and movement of all individual birds detected were mapped to avoid double counting.

Observers were instructed to conduct their counts independently of the other observers, not looking at the other observers or keying in on their movements. Observers were spaced out around the point so that they could not see each other's data sheets. Most (>95%) of observations were auditory, so cues from other observers were minimal. Birds were also abundant in this area (>10 observations per point), so observers had little opportunity to cue off of each other. Because of these factors, we are confident that the independence assumption was met for this survey.

Following each count, observers compared their data sheets to determine the total number of birds detected and match birds seen in common. Detection histories for each bird are based on these matches. There are three observable detection histories for a two-independent-observer point count (detected only by observer 1, detected only by observer 2, or detected by both observers). Similarly, there are 15 observable detection histories for four independent observers.

We used the first two observers from the full four-observer data set to construct two-observer data sets. Two-observer analyses were also done for the other two observers to confirm consistency, but these data are not presented. For illustrative purposes, we present analyses, using both two-observer and four-observer methods, for five species: Blue-headed Vireo (*Vireo solitarius*), Red-eyed Vireo (*V. olivaceus*), Tufted Titmouse (*Baeolophus bicolor*), Ovenbird

(*Seiurus aurocapilla*), and Scarlet Tanager (*Piranga olivacea*). The two vireo species are analyzed as a group because their songs are similar and we believe that identification errors between these two species are likely. All data were truncated by discarding 10% of the largest detection distances for each species. This gives a reference distance that can be used for density estimation and ensures a nonzero probability of detection for all birds across all four observers.

Two independent observer models.—Data for two independent observers are analyzed using Lincoln-Petersen closed-population capture–recapture models (Otis et al. 1978, Seber 1982). Relevant capture–recapture models for these data are model M_0 (equal detection probability between observers) and model M_{obs} (unequal detection probability between observers; M_t of Otis et al. [1978]). See Otis et al. (1978) and White et al. (1982) for a description of these capture–recapture models.

The assumptions of the independent-observer models are (1) that observations among observers are independent, (2) that counts within fixed-radius plots are accurate, (3) that assignments of detection histories x_{11}, x_{10}, and x_{01} are accurate (i.e., there are no matching errors), (4) that the detection probability for each species is constant for each observer, and (5) that there is no undetected movement into or out of the fixed-radius plot.

Models incorporating detection distance and other covariates require conditioning the probability of detecting a bird on both availability and the individual covariates. Probability of detection by observer j, given availability for an individual i, can be represented as a function of a covariate:

$$\text{Logit}(p_{ij}) = a_j + \beta_j \delta(r_i) \qquad (1)$$

where α_j is the intercept, β_j is the slope, and $\delta(r_i)$ is a function of the covariate (such as detection distance). Using detection distance as a covariate allows for four additional models: model M_0^d (equal intercept and slope terms between observers), model M_0^{*d} (equal intercept between observers but different slope), model M_{obs}^d (unequal intercept between observers but similar slope), and model M_{obs}^{*d} (unequal intercept and slope between observers).

When covariates are included in the model, it is necessary to use the generalized Horvitz-Thompson (Horvitz and Thompson 1952) estimator of population size (Huggins 1989, 1991; Alho 1990):

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{\hat{p}_i} \qquad (2)$$

where n is the number of birds detected and p_i is the detection probability of an individual bird. The program MARK (White and Burnham 1999) provides the Horvitz-Thompson estimate of population size as a "derived parameter," when using the "Huggins closed captures" data type.

Two-observer data sets were analyzed using MARK with the Huggins closed captures data type. The *a priori* set of candidate models included the six models previously discussed for two-observer data. The most parsimonious models were selected using second-order Akaike's Information Criterion (AIC$_c$), an information-theoretic approach with an adjustment for small sample size (Burnham and Anderson 2002).

Four or more independent observer models.—When four or more sampling periods (in our case, observers) are used in closed capture–recapture experiments, there are conceptually eight models available for analysis (Otis et al. 1978, Pollock et al. 1990, Williams et al. 2002). Only four of these are reasonable models for independent-observer point-count data: M_0 (equal capture probability), M_{obs} (observer variation in capture probability), M_h (individual capture heterogeneity), $M_{obs,h}$ (observer variation and individual capture heterogeneity).

Behavioral response models (Otis et al. 1978) are probably not relevant to analysis of independent-observer point-count data. A behavioral response is a response by an individual to capture (in our case, detection) that makes it either more or less likely to be captured after first capture. Because observations on point counts are done simultaneously and independently, we assume that detections by one observer do not affect detections of other observers.

The general assumptions for the four-independent-observer models are the same, except that the assumption of constant detection probability among individuals is no longer necessary because individual differences among birds can be modeled with data from four or more observers. Using three observers may provide more precise estimates than two observers,

but it does not provide the data necessary to fit the heterogeneity models.

Different models reflect different assumptions about the sources of variability in the data. Model M_0 has the most restrictive assumptions, requiring that the probability of detection is the same for all individuals in the population and that there are no differences among observers in their ability to detect individuals. Model M_{obs} is less restrictive, in that it allows for differences among observers but still requires an assumption that all individuals in the population have equal detection probabilities for a given observer.

Model M_h allows for heterogeneity in detection probabilities of individual birds but no observer differences. Model $M_{obs,h}$ provides for both observer differences and heterogeneity in the detection probabilities of individual birds. Three estimators are available to estimate abundance in the presence of individual heterogeneity. The jackknife estimator for model M_h (Burnham and Overton 1978, 1979) and Chao's estimator for model M_h and model $M_{obs,h}$ (Chao and Lee 1992, Chao et al. 1992) can be run using the program CAPTURE (see Acknowledgments), but not with MARK. The alternative likelihood-based approach involves finite-mixture models of heterogeneity (Norris and Pollock 1996, Pledger 2000), which can be parameterized in MARK and can include individual covariates. We use the notation M_{xh} for the finite-mixture models, where x is the number of mixtures in the model.

Incorporating observable individual heterogeneity into the independent-observer models using covariates with four or more observers is identical to the procedure described above for the two-observer situation using the generalized Horvitz-Thompson estimator.

We modeled the four-observer data sets using the same six models used for the two-observer data sets and also incorporated the heterogeneity

models. We ran the jackknife and Chao estimators of the capture–recapture heterogeneity model (M_h) using CAPTURE, and we ran the finite-mixture heterogeneity models in MARK. In all cases, the three methods gave similar estimates. The estimates from the two-point mixture models (M_{2h} and $M_{obs,2h}$) are reported because model selection is based on AIC criteria from the entire suite of models, including those with detection-distance covariates.

RESULTS

Two-independent-observer examples.—The total number of birds detected ranged from 31 (Tufted Titmouse) to 89 (vireos) for the two-observer data sets. The AIC_c scores selected either model M_0 or model M_0^d for all data sets (Table 1). When model M_0 was selected, model M_0^d was always a reasonable alternative model, judging from differences in AIC_c ($\Delta AIC_c < 2$). By contrast, when model M_0^d was selected, model M_0 was not necessarily a reasonable alternative model (e.g., with Ovenbirds; Table 1).

Model M_0 detection probabilities for individual observers was high, ranging from 0.77 (Scarlet Tanager and vireos) to 0.90 (Ovenbird) (Table 2). Population estimates were similar between models M_0 and M_0^d. Standard errors were small because of high detection probabilities.

Four-independent-observer examples.—Raw counts from the four-observer data (Table 3) were higher than those from the two-observer abundance estimates (Table 2), which suggests individual heterogeneity in detection probabilities for all species. Model M_h (jackknife estimator) was selected as the most parsimonious model for all four data sets using CAPTURE. A two-point mixture model for model M_h was selected as the most parsimonious model for all data sets, except the Ovenbird data set, using MARK (Table 4).

TABLE 1. ΔAIC_c values for two-independent-observer data sets using all six candidate models (see text). Smaller ΔAIC_c values indicate more parsimonious models, with 0 indicating the selected model. AIC_c weights are in parentheses.

Species	Models					
	M_0	M_{obs}	M_0^d	M_0^{*d}	M_{obs}^d	M_{obs}^{*d}
Ovenbird	5.93 (0.03)	7.91 (0.01)	0 (0.65)	3.26 (0.13)	3.25 (0.13)	5.35 (0.05)
Tanager	0 (0.40)	2.03 (0.15)	1.06 (0.24)	2.76 (0.10)	3.15 (0.08)	4.80 (0.04)
Titmouse	0 (0.33)	2.14 (0.11)	0.10 (0.31)	2.31 (0.10)	2.31 (0.10)	4.60 (0.03)
Vireo	0 (0.33)	1.77 (0.14)	0.63 (0.24)	1.63 (0.15)	2.42 (0.01)	3.60 (0.05)

TABLE 2. Abundance estimates (*N*) for the two-independent-observer examples. Birds detected are the totals between the two observers. Model M_0 was selected as the most parsimonious for all data sets except the Ovenbird. Standard errors are in parentheses.

Species	Birds detected	Model M_0 detectability	N Model M_0	N Model M_0^d
Ovenbird	72	0.90 (0.027)	73 (0.94)	73 (1.43)
Tanager	45	0.77 (0.055)	48 (2.09)	48 (2.28)
Titmouse	31	0.85 (0.052)	32 (0.98)	32 (1.22)
Vireo	89	0.77 (0.039)	94 (2.86)	94 (3.03)

TABLE 3. Abundance estimates (*N*) for the four-independent-observer examples. Birds detected are the totals among the four observers. Detection probabilities are given by *p* for model M_0 and p^{group1} and p^{group2} for model M_{2h}. The proportion of the population in group 1 is given by pr(group1). Standard errors are in parentheses.

Data set	Birds detected	Model M_0		Model M_{2h}			
		p	N	pr(group1)	p^{group1}	p^{group2}	N
Ovenbird [a]	81	0.82 (0.022)	81 (0.305)	0.41 (0.059)	0.51 (0.060)	1.0 (≈0.0)	83 (1.704)
Tanager	55	0.66 (0.033)	56 (0.894)	0.54 (0.086)	0.30 (0.108)	0.92 (0.047)	63 (6.32)
Titmouse	39	0.70 (0.038)	39 (0.592)	0.53 (0.090)	0.28 (0.111)	0.97 (0.032)	45 (5.56)
Vireo	111	0.67 (0.023)	112 (1.243)	0.55 (0.055)	0.25 (0.067)	0.93 (0.026)	134 (12.00)

[a] Selected model was model M_0^d and the estimated population size was 82 (SE = 0.889).

TABLE 4. Model selection for the four-independent-observer examples giving the ΔAIC_c values for all eight candidate models. Model M_{2h} and model $M_{obs,2h}$ are based on two-point mixture models of heterogeneity. The smaller ΔAIC_c values indicate a more parsimonious model, with 0 indicating the selected model. AIC_c weights are in parentheses.

Data set	Models							
	M_0	M_{obs}	M_0^d	M_0^{*d}	M_{obs}^d	M_{obs}^{*d}	M_{2h}	$M_{obs,2h}$
Ovenbird	48.5 (0.0)	53.5 (0.0)	**0.0 (0.95)**	8.4 (0.01)	7.4 (0.02)	12.5 (0.0)	10.3 (0.01)	37.7 (0.0)
Tanager	23.4 (0.0)	29.4 (0.0)	24.1 (0.0)	29.0 (0.0)	30.1 (0.0)	35.2 (0.0)	**0.0 (1.0)**	22.8 (0.0)
Titmouse	28.9 (0.0)	34.1 (0.0)	26.0 (0.0)	30.3 (0.0)	31.2 (0.0)	35.8 (0.0)	**0.0 (0.96)**	6.2 (0.04)
Vireo	71.2 (0.0)	77.2 (0.0)	60.2 (0.0)	65.6 (0.0)	66.3 (0.0)	71.7 (0.0)	**0.0 (1.0)**	36.4 (0.0)

Estimates of abundance were similar (differences < 2) among the jackknife, Chao, and two-point mixture estimators of the heterogeneity model.

The finite-mixture parameter estimates for all species were close to 0.5, indicating that approximately half of the population was in each detection group (Table 3). One group of the finite mixture was always highly detectable (probability of detection > 0.90), whereas the other group was generally much harder to detect (probability of detection ≤ 0.30). The Ovenbird was the exception to this pattern (probability of detection = 0.51). Population estimates (Table 3) from

model M_{2h} (two-point mixture) ranged from 15% (Tufted Titmouse) to 21% (vireos) higher than the estimates from the selected four-observer model without heterogeneity. The selected model for the Ovenbird, which used distance to model observable heterogeneity, gave the same population estimate as the two-point mixture model.

DISCUSSION

An analysis of independent-observer point-count data suggests potential biases associated with individual differences in detection

probabilities among birds. Covariate and heterogeneity models provide a means to account for these individual differences in detection probabilities. The four-observer point-count method is the most sensitive to sources of variability in detection probability.

In addition, the independent-observer approach is not constrained by the assumption that detection probability is one at detection distance zero (Buckland et al. 2001), even when detection distance is used as a covariate. This restrictive assumption of distance sampling may not be reasonable for all surveys, especially when birds occur high in a forest canopy.

The approach can also accommodate other factors affecting the detection process, such as singing rate (Wilson and Bart 1985, McShea and Rappole 1997). McShea and Rappole (1997) found that singing rates were lower for birds closer to an observer. If singing rate affects the probability of detection, there is an additional source of variation in the data that will cause heterogeneity in detection probabilities that cannot be accounted for by distance information.

Assumptions.—Multiple-observer approaches assume that the survey population is closed (i.e., no movement into or out of the survey area during the count) and that birds are not double counted. These assumptions are common to all point counts, and they are sensitive to count duration. Double counting arises when the same bird is counted at more than one location and is counted as two or more individuals. Reducing count duration limits the probability of birds moving into, out of, or within the survey area during the count. Unfortunately, reducing count duration also reduces the total number of detections on the count. On our study sites, detections are generally auditory; thus, if birds have a low singing rate, short-duration counts can significantly limit the proportion of birds that are available to count. Optimal count durations should be long enough to ensure that all birds are available for detection (i.e., sing at least once during the count), and short enough to minimize the effects of movement.

Independent-observer methods require that observations are independent among observers. The primary–secondary observer method of Nichols et al. (2000) requires only that the observations by the primary observer are independent of detections made by the secondary observer. Many of the issues involved with independence between observers are similar between the two methods. Independence is violated if an observer obtains cues from other observers. This could occur if one observer is writing down an observation (Nichols et al. 2000), estimating a distance, or moving in a manner that would draw the attention of other observers. Nichols et al. (2000) suggested that violations of the independence assumption are most likely when there are few birds at a point or when most observations are visual. In other words, it is harder to obtain cues from other observers when detections are auditory. Our data are based almost exclusively (>95%) on auditory detections.

We believe that the independence assumption, especially with four observers, is the most critical assumption of the multiple-observer method. Failure to meet this assumption may induce heterogeneity in the data and bias abundance estimates. Therefore, we recommend that multiple-observer methods are most practical when detections are primarily auditory and there are a large number of observations at each point.

Nichols et al. (2000) viewed differences in observers' ability to detect birds at different distances as the most serious source of error in the primary–secondary observer method, and they recommended using fixed-radius plots. We used distance estimates as covariates in our models to account for this problem. Using fixed-radius plots to obtain constant detection probabilities is not necessary when more precise distance information is available.

The process of matching observers' observations is another source of error in multiple-observer methods. If surveys are conducted during the breeding season, when territories are fixed and birds are not in large flocks, matching errors should be minimal. The presence of matching errors should be carefully evaluated for all surveys using this method. If many birds of one species are present at a point, it may be difficult to determine which birds are seen in common. Matching errors is another source of variation that may lead to apparent heterogeneity in the data. We suggest a conservative set of rules for matching, because failure to match birds seen in common will lead to overestimates of abundance.

Example analyses.—Model selection for the two-observer data sets suggests some important characteristics of these data. First, we note that models that incorporated differences in

detection probability between observers were not selected. This implies that detection probability based on auditory cues was similar between two highly trained observers. The largest difference in estimated detection probabilities between observers for any data set was 8%, for the vireo group. Models incorporating differences between observers were reasonable alternative models for several species. This suggests that models of observer differences should be included in analyses of point-count data. Multiple-observer models may prove useful for species that are hard to detect or when observers' abilities vary widely. Incorporating detection distance as a covariate also provided important improvements to multiple-observer models.

Comparing two-observer analyses with those from the other two observers shows almost identical detection probabilities and population estimates. However, two-observer abundance estimates were low, and detection probability estimates were high compared with the four-observer estimates, which is consistent with the expected biases associated with individual animal heterogeneity. Although estimates based solely on a two-observer approach may appear to be reasonable (as in our examples), caution is advised because no assessment of potential bias can be made with these data.

Heterogeneity models were generally selected as the "best" models. Potential causes of heterogeneity in detection probabilities for point-count surveys are violation of the independence assumption, combining types of detection (e.g., auditory and visual), estimating detection probability for a group of species, and individual variation in singing rates. Violation of the independence assumption caused by observers' cuing off each other would bias observed capture histories. This violation would result in more capture histories where the majority of observers detected an individual bird and fewer histories where only one observer detected an individual bird than expected under models of equal detection probability. This would result in a positive bias in detection probabilities and an underestimation of abundance. Comparisons of observed to expected capture histories indicate that violations of the independence assumption are not the main factor associated with heterogeneity in these data.

Combining multiple types of detections is a potential source of individual heterogeneity

that should also be considered carefully when studies are designed. For example, the detection probabilities of songs may be very different from the detection probabilities of calls or visual observations.

Variability in singing rates of individual birds is another potential source of heterogeneity in detection probabilities (Wilson and Bart 1985, McShea and Rappole 1997). A bird that sings frequently is more likely to be detected than one that sings only once during a survey period. Differences in singing rates may be associated with differences in pairing status or nest stage. For example, unpaired males may sing more frequently than paired males, and paired males that have not nested are likely to sing at higher rates than a paired male that is incubating or caring for a brood (Wasserman 1977, Krebs et al. 1981, Lein 1981, Wilson and Bart 1985). This presents a situation where the use of finite-mixture models is biologically reasonable because individual heterogeneity can be associated with biological groups on the basis of an individual's reproductive stage.

Unaccountable sources of heterogeneity can arise from factors such as whether a bird is facing toward or away from the observer, or from more complex interactions, such as the effects of vegetation on sound attenuation. Differences in song characteristics, vegetative cover or density, topographic features, and background noise all affect the intensity and attenuation of auditory cues (Richards 1981). Likewise, the presence of conspecific territorial males, competitors, predators, and a bird's mating and nesting status could all affect the singing behavior of individual birds. Stratification and careful survey design can provide partial control of some of these sources of heterogeneity, but they will always be present in avian survey data.

Although heterogeneity models can account for unobservable heterogeneity, they should be interpreted with caution. Link (2003) showed that alternative models of heterogeneity may predict similar observations from populations of different size, making it impossible to distinguish among alternative heterogeneity models. We are encouraged that all three heterogeneity estimators gave similar results for our data, but the estimates may still be biased. The best approach to account for heterogeneity is to determine what the major sources of heterogeneity are and either control for them to eliminate the effect

or measure them during the survey so they can be used as covariates in the analysis.

Although we recognize that the assumption of independence associated with this approach is very stringent and that the method is not practical for many field situations, the analysis provides valuable insights into the detection process and sources of variation that may bias point-count data. When the number of birds at each point is sufficient (in our experience, at least eight birds), observers are too busy recording and tracking birds to cue off other observers, and the independence assumption should be met. Two-observer point counts are practical for many field situations and could be used for monitoring purposes. Four-observer point counts are not practical and are probably limited in use to research situations where a better understanding of the detection process is important.

Recommendations

Independent-observer approaches are useful tools among the suite of statistical methods available for analyzing point-count data. Benefits of the approach include the number of candidate models available for analysis; ease of implementation with existing software; ability to run a suite of models, including covariate models, within MARK; and the application of information-theoretic approaches to model selection. In particular, the incorporation of detection distance, a relatively well-studied methodology, as a covariate within an independent-observer framework may provide more precise abundance estimates.

We recommend conducting pilot studies involving multiple observers collecting multiple types of data before any implementation of a large-scale survey. Such a study will allow an assessment of the various sources of individual heterogeneity and selection of the most efficient methodology. Pilot-study results can provide guidance, based on factors determined to be affecting the detection process, for allocation of distance sampling using one, two, or more than three observers.

Further investigation of distance covariate models to develop models with different detection functions, including those that are not strictly monotonically decreasing, is warranted. These models could account for observer effects on detection probabilities of birds near the point. Detection distance could also be used as a covariate in the finite-mixture models of individual heterogeneity.

Finally, models combining independent-observer and time-of-detection (Farnsworth et al. 2002) approaches would be a useful extension of this work. Combined models would allow separation of the probability of availability from the probability of detection given availability. For species with high singing rates, this may not provide much useful information, because availability is often near one. For less vocal species, this additional component of the detection process may have significant effects on abundance estimates.

Literature Cited

Alho, J. M. 1990. Logistic regression in capture–recapture models. Biometrics 46:623–635.

Alldredge, M. W. 2004. Avian point count surveys: Estimating components of the detection process. Ph.D. dissertation, North Carolina State University, Raleigh.

Barker, R. J., and J. R. Sauer. 1995. Statistical aspects of point count sampling. Pages 125–130 in Monitoring Bird Populations by Point Counts (C. J. Ralph, J. R. Sauer, and S. Droege, Eds.). U.S. Department of Agriculture, Forest Service General Technical Report PSW-GTR-149.

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 University Press, New York.

Burnham, K. P. 1981. Summarizing remarks: Environmental influences. Pages 324–325 in Estimating Numbers of Terrestrial Birds (C. J. Ralph and J. M. Scott, Eds.). Studies in Avian Biology, no. 6.

BURNHAM, K. P., AND D. R. ANDERSON. 2002. Model Selection and Multimodel Inference: A Practical Information-theoretic Approach, 2nd ed. Springer-Verlag, New York.

BURNHAM, K. P., AND W. S. OVERTON. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. Biometrika 65:625–633.

BURNHAM, K. P., AND W. S. OVERTON. 1979. Robust estimation of population size when capture probabilities vary among animals. Ecology 60:927–936.

CHAO, A., AND S.-M. LEE. 1992. Estimating the number of classes via sample coverage. Journal of the American Statistical Association 87:210–217.

CHAO, A., S.-M. LEE, AND S. L. JENG. 1992. Estimating population size for capture–recapture data when capture probabilities vary by time and individual animal. Biometrics 48:201–216.

CONROY, M. J. 1996. Abundance indices. Pages 179–192 *in* Measuring and Monitoring Biological Diversity: Standard Methods for Mammals (D. E. Wilson, F. R. Cole, J. D. Nichols, R. Rudran, and M. Foster, Eds.). Smithsonian Institution Press, Washington, D.C.

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.

HORVITZ, D. G., AND D. J. THOMPSON. 1952. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47:663–685.

HUGGINS, R. M. 1989. On the statistical analysis of capture experiments. Biometrika 76:133–140.

HUGGINS, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. Biometrics 47:725–732.

JOHNSON, D. H. 1995. Point counts of birds: What are we estimating? Pages 117–123 *in* Monitoring Bird Populations by Point Counts (C. J. Ralph, J. R. Sauer, and S. Droege, Eds.). U.S. Department of Agriculture, Forest Service General Technical Report PSW-GTR-149.

JOHNSON, D. H., K. P. BURNHAM, AND J. D. NICHOLS. 1986. The role of heterogeneity in animal population dynamics. Proceedings of the International Biometrics Conference 13:1–15.

KREBS, J. R., M. AVERY, AND R. J. COWIE. 1981. Effect of removal of mate on the singing behavior of Great Tits. Animal Behaviour 29:635–637.

LEIN, M. R. 1981. Display behavior of Ovenbirds (*Seiurus aurocapillus*). II. Song variation and singing behavior. Wilson Bulletin 93:21–41.

LINK, W. A. 2003. Nonidentifiability of population size from capture–recapture data with heterogeneous detection probabilities. Biometrics 59:1123–1130.

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.

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.

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.

NORRIS, J. L., III, AND K. H. POLLOCK. 1996. Nonparametric MLE under two closed capture–recapture models with heterogeneity. Biometrics 52:639–649.

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, no. 62.

PLEDGER, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. Biometrics 56:434–442.

POLLOCK, K. H., J. D. NICHOLS, C. BROWNIE, AND J. E. HINES. 1990. Statistical inference for capture–recapture experiments. Wildlife Monographs, no. 107.

RALPH, C. J., S. DROEGE, AND J. R. SAUER. 1995a. Managing and monitoring birds using point counts: Standards and applications. Pages 161–168 *in* Monitoring Bird Populations by Point Counts (J. C. Ralph, J. R. Sauer, and S. Droege, Eds.). U.S. Department of Agriculture, Forest Service General Technical Report PSW-GTR-149.

RALPH, C. J., J. R. SAUER, S. DROEGE, EDS. 1995b. Monitoring Bird Populations by Point Counts. U.S. Department of Agriculture,

Forest Service General Technical Report PSW-GTR-149.

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 297–300 *in* Estimating Numbers of Terrestrial Birds (C. J. Ralph and J. M. Scott, Eds.). Studies in Avian Biology, no. 6.

ROBBINS, C. S., D. BYSTRAK, AND P. H. GEISSLER. 1986. The Breeding Bird Survey: Its First Fifteen Years, 1965–1979. U.S. Fish and Wildlife Service Resource Publications, no. 157.

ROSENSTOCK, S. S., D. R. ANDERSON, K. M. GIESEN, T. LEUKERING, AND M. F. CARTER. 2002. Landbird counting techniques: Current practices and an alternative. Auk 119:46–53.

SAUER, J. R., J. E. HINES, G. GOUGH, I. THOMAS, AND B. G. PETERJOHN. 1997. The North American Breeding Bird Survey, Results and Analysis, version 96.4. U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland.

SEBER, G. A. F. 1982. The Estimation of Animal Abundance and Related Parameters, 2nd ed. Edward Arnold, London.

THOMPSON, W. L. 2002. Towards reliable bird surveys: Accounting for individuals present but not detected. Auk 119:18–25.

WASSERMAN, F. E. 1977. Mate attraction function of song in the White-throated Sparrow. Condor 79:125–127.

WHITE, G. C., D. R. ANDERSON, K. P. BURNHAM, AND D. L. OTIS. 1982. Capture–recapture removal methods for sampling closed populations. Los Alamos National Laboratory Publication LA-8787-NERP, Los Alamos, New Mexico.

WHITE, G. C., AND K. P. BURNHAM. 1999. Program MARK: Survival estimation from populations of marked animals. Bird Study 46 (Supplement):S120–S139.

WILLIAMS, B. K., J. D. NICHOLS, AND M. J. CONROY. 2002. Analysis and Management of Animal Populations. Academic Press, San Diego, California.

WILSON, D. M., AND J. BART. 1985. Reliability of singing bird surveys: Effects of song phenology during the breeding season. Condor 87: 69–73.