

## A REMOVAL MODEL FOR ESTIMATING DETECTION PROBABILITIES FROM POINT-COUNT SURVEYS

GEORGE L. FARNSWORTH,<sup>1,4</sup> KENNETH H. POLLOCK,<sup>2</sup> JAMES D. NICHOLS,<sup>3</sup>

THEODORE R. SIMONS,<sup>1</sup> JAMES E. HINES,<sup>3</sup> AND JOHN R. SAUER<sup>3</sup>

<sup>1</sup>Cooperative Fish and Wildlife Research Unit, Department of Zoology, North Carolina State University, Raleigh, North Carolina 27695, USA;

<sup>2</sup>Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, USA; and

<sup>3</sup>U.S. Geological Survey, Biological Resources Division, Patuxent Wildlife Research Center, Laurel, Maryland 20708, USA

**ABSTRACT.**—Use of point-count surveys is a popular method for collecting data on abundance and distribution of birds. However, analyses of such data often ignore potential differences in detection probability. We adapted a removal model to directly estimate detection probability during point-count surveys. The model assumes that singing frequency is a major factor influencing probability of detection when birds are surveyed using point counts. This may be appropriate for surveys in which most detections are by sound. The model requires counts to be divided into several time intervals. Point counts are often conducted for 10 min, where the number of birds recorded is divided into those first observed in the first 3 min, the subsequent 2 min, and the last 5 min. We developed a maximum-likelihood estimator for the detectability of birds recorded during counts divided into those intervals. This technique can easily be adapted to point counts divided into intervals of any length. We applied this method to unlimited-radius counts conducted in Great Smoky Mountains National Park. We used model selection criteria to identify whether detection probabilities varied among species, throughout the morning, throughout the season, and among different observers. We found differences in detection probability among species. Species that sing frequently such as Winter Wren (*Troglodytes troglodytes*) and Acadian Flycatcher (*Empidonax virescens*) had high detection probabilities (~90%) and species that call infrequently such as Pileated Woodpecker (*Dryocopus pileatus*) had low detection probability (36%). We also found detection probabilities varied with the time of day for some species (e.g. thrushes) and between observers for other species. We used the same approach to estimate detection probability and density for a subset of the observations with limited-radius point counts. Received 23 February 2000, accepted 9 October 2001.

**RESUMEN.**—El muestreo mediante conteos por punto es un método popular para coleccionar datos sobre distribución y abundancia de aves. Sin embargo, los análisis de estos datos generalmente ignoran diferencias potenciales en la probabilidad de detección. Aquí adaptamos un modelo de remoción para estimar directamente la probabilidad de detección de aves en conteos por punto. El modelo supone que la frecuencia con que las aves cantan es un factor principal que influye la probabilidad de detección. Esto puede ser apropiado en muestreos en donde la mayoría de las detecciones son por sonido. El modelo requiere que los conteos sean divididos en varios intervalos de tiempo. Los conteos por punto duran por lo general 10 min, donde el número de aves registradas es dividido en aquellas observadas durante los primeros 3 min, los 2 min subsiguientes y los últimos 5 min. Desarrollamos un estimador de máxima probabilidad en relación a la detectabilidad de las aves registradas durante conteos divididos en dichos intervalos. Esta técnica puede ser fácilmente adaptada a conteos por punto divididos en intervalos de cualquier duración. Aplicamos este método a conteos de radio ilimitado realizados en el Parque Nacional Great Smoky Mountains. Empleamos criterios de selección del modelo para identificar si las probabilidades de detección variaban entre especies, a lo largo de la mañana, a lo largo de las estaciones y entre diferentes observadores. Encontramos diferencias entre las especies en la probabilidad de detección. Las especies que cantan con frecuencia, como *Troglodytes troglodytes* y *Empidonax virescens*, tuvieron

---

<sup>4</sup> Present address: Department of Natural Sciences, University of Houston-Downtown, One Main Street, Houston, Texas 77002-1001, USA. E-mail: farnsworthg@uhd.edu

una alta probabilidad de detección (~90%), mientras que las especies que realizan pocas llamadas, como *Dryocopus pileatus*, tuvieron una baja probabilidad de detección (36%). Encontramos también que la probabilidad de detección varió en relación a la hora del día para algunas especies (e.g. Tóridos) y entre observadores para otras. Empleamos el mismo procedimiento para conteos por punto de radio limitado para estimar la probabilidad de detección y la densidad en un subconjunto de las observaciones.

POINT-COUNT SURVEYS are routinely used to gather information about breeding birds. This technique involves using a standardized methodology to record all birds heard or seen during a fixed amount of time at many widely spaced count locations. This method is widely used because it is an efficient way to collect count data over a large area. However, there is considerable controversy about what inferences may be made on the basis of such data (e.g. see Burnham 1981, Johnson 1995).

Typically most investigators consider point-count surveys to represent an index of bird abundance that can be used to make comparisons between datasets (Lancia et al. 1994, Ralph et al. 1995). That requires an assumption that the probability of detection is the same for each data set being compared. For example, if two habitat types were sampled, the ratio of the counts would reflect the ratio of abundance only if the detection probabilities were the same in both habitats. Similarly, if the same location were sampled in two different years and there was a change in the number of birds counted, this change could only be interpreted as a change in population size if the detection probabilities were the same in both years. In that way, counts are often used as estimates of relative abundance.

Criticism of such analyses that use counts as an index centers on the assumption of equal detectability between datasets (Burnham 1981, Wilson and Bart 1985, Johnson 1995, Barker and Sauer 1995). Such criticism may be valid because many factors have been shown to affect detectability. For example, numbers of birds detected on point-count surveys can be affected by time of season (e.g. Best 1981, Ralph 1981, Skirvin 1981) and time of day (e.g. Robbins 1981, Skirvin 1981, Bart and Herrick 1984) presumably because of variations in singing frequency. Wilson and Bart (1985) found that the singing frequency of House Wrens (*Troglodytes aedon*) changed throughout the nesting cycle. McShea and Rappole (1997) found singing frequency for Ovenbird, Wood Thrush, and

Northern Cardinal (*Cardinalis cardinalis*) varied with distance to observer and with habitat type (fragmented vs. contiguous forest). Physical attributes of habitat such as foliage density can also affect an observer's ability to hear and identify bird song (Richards 1981). Differences in detectability are also related to skill and experience of observers (Sauer et al. 1994). Variations in detectability due to those and other factors have cast doubt on the use of counts as indexes.

One method for overcoming the assumption of equal detectability is distance sampling. It relies on the notion that detectability declines with distance from the observer (Reynolds et al. 1980, Buckland et al. 1993). The variable circular plot technique uses detection distance to estimate the detection probability and bird density from point-count surveys. However, estimating distances to all birds seen or heard can be difficult and imprecise. A method that does not require distance measures was recently developed by Nichols et al. (2000). That method estimates detection probabilities using two observers collecting data simultaneously on point-count surveys.

In point count surveys where birds are primarily detected by song, probability that a bird will be recorded during a count can be thought of as the product of two probabilities: (1) the probability the bird sings during the count, and (2) the probability the bird is detected given that it sings. Distance sampling models the decline in the second component of the probability (that a bird is detected given that it sings) with increasing distance from the observer (Reynolds et al. 1980, Buckland et al. 1993). Similarly, the double-observer approach (Nichols et al. 2000) models probability that a bird is recorded given that it could be detected by at least one observer. To be recorded by one observer, a bird must sing. Neither of those approaches deals with the first component of detection probability. We propose that if counts are separated into time intervals, we can estimate the product of both components of detec-

tion probability using the approach of a removal experiment (Moran 1951, Seber 1982). A removal experiment typically traps and removes animals from a population in discrete time periods (trap sessions). As animals are removed from the population, fewer will be available for capture in the subsequent trap sessions. The decline in numbers caught through time can then be used to estimate the initial population size.

The simplest application of this approach to point counts can be illustrated with just two time intervals of equal duration. Suppose an observer records all birds seen or heard in the interval  $(0, t)$  and continues the point count, recording any additional birds detected in the interval  $(t, 2t)$ . At the end of the point count, we define  $x_1$  as number of birds counted in the first time interval and  $x_2$  as number of new birds (not detected in period 1) detected in the second time interval. The expected value of the random variable  $x_1$  is  $E(x_1) = Np_1$  where  $N$  is the total number of birds within the detection radius of the observer and  $p_1$  is the detectability for an individual bird in the first time period. The expected value of the second random variable  $x_2$  is  $E(x_2) = N(1 - p_1)p_2$ , where  $p_2$  is the detectability in the second time period. The  $(1 - p_1)$  term is needed because in order for a bird to be first counted in the second time interval, it must have been missed in the first time interval. Let us assume that detectability for the two intervals is the same (i.e.  $p_1 = p_2$ ) because the duration of each interval is the same. Solving the above equations produces the following moment estimator for  $N$  (Zippin 1958):

$$\hat{N} = \frac{x_1^2}{x_1 - x_2} \quad (1)$$

Note that the estimator can fail if  $x_1 \leq x_2$  which is possible when  $p$  is small. This is a good approximation to the maximum-likelihood estimator discussed in Otis et al. (1978), which can be computed numerically using program CAPTURE (White et al. 1982).

We present this two-sample removal estimator to illustrate the approach with the simplest possible situation. In practice, we recommend using more than two intervals, which permits us to relax the assumption of equal detectability ( $p_1 = p_2$ ). Program CAPTURE can produce maximum-likelihood estimates for  $N$ , as well as the estimated variance of  $\hat{N}$ , using

model  $M_b$  (described in Otis et al. 1978 and White et al. 1982), as long as each of the time intervals is the same length. Model  $M_b$  estimates the capture probability of unmarked animals in a closed population capture-recapture experiment. Here we present a more general model that allows for the count intervals to have variable length (i.e. the detection probabilities in the different intervals need not be the same). This model is therefore a generalization of the model  $M_b$ . We illustrate the technique with several examples derived from field data.

## METHODS

We developed models capable of estimating detectability when a point count is divided into three or more intervals of variable length. A common method for recording data at point counts is to separate number of birds counted into those first observed within the first 3 min, those first observed within the next 2 min, and those first observed within the final 5 min. This procedure was recommended by Ralph et al. (1995) and was originally designed to allow results from 10 min counts to be comparable with those from studies employing 3 and 5 min counts. We define  $x_1$  as number of birds counted in the first interval,  $x_2$  as number of birds counted in the second interval,  $x_3$  as number of birds counted in the third interval, and  $x$  as total number of birds counted in the full 10 min ( $x = \sum_{i=1}^3 x_i$ ).

*Estimating detectability.*—We developed two estimators for detection probability, one that allows for heterogeneity (variation in the detectability within the population of birds sampled) and one that does not. We describe the most general model ( $M_c$ ) that incorporates heterogeneity first because the reduced model ( $M$ ) is a simplified version of this model.

First we divide the population of birds ( $N$ ) within the detection radius of an observer into two groups. Group 1 is composed of the birds that are easily detected and group 2 includes those more difficult to detect. The probability that a randomly selected bird is a member of group 2 (hence the expected proportion of the population in group 2) is defined as  $c$ . We assume that all members of group 1 will be detected within the first time interval. We also define probability of failing to detect a member of group 2 within one minute as  $q$ . The expected value for number of birds detected within the first time interval of three minutes is therefore:

$$\begin{aligned} E(x_1) &= N(1 - c) + Nc(1 - q^3) \\ &= N(1 - cq^3) \end{aligned} \quad (2)$$

All of the members of group 1 plus some from group 2 will be detected in that interval. The probability that a bird in group 2 will be missed during the first

3 min is  $q^3$ , and the complement of this ( $1 - q^3$ ) is thus the probability of being detected at least once in the first interval. Similarly, the expected value of the number of birds initially detected within the next time interval (2 min) is:

$$E(x_2) = Ncq^3(1 - q^2) \quad (3)$$

The birds first recorded during the middle interval must be missed in the first 3 min and not missed in the subsequent 2 min. Thus, these are all members of group 2. Finally, expected number of birds counted in the last 5 min of the count is:

$$E(x_3) = Ncq^5(1 - q^5) \quad (4)$$

For a bird to be first counted in the final interval, it must be a member of group 2, and it must be missed during the first 5 min and not missed during the last 5 min. The expected total number of birds counted by the end of the full 10 min is:

$$\begin{aligned} E(x.) &= N[c(1 - q^{10}) + (1 - c)] \\ &= N(1 - cq^{10}) \end{aligned} \quad (5)$$

The cumulative probability of detecting a bird during the full 10 min count is thus  $p = 1 - cq^{10}$ . This can be described as a full multinomial distribution with probability density function:

$$\begin{aligned} f(x_1, x_2, x_3 | N) \\ &= \frac{N!}{x_1!x_2!x_3!(N - x.)!} [1 - cq^3]^{x_1} [cq^3(1 - q^2)]^{x_2} \\ &\quad \times [cq^5(1 - q^5)]^{x_3} [cq^{10}]^{N - x.} \end{aligned} \quad (6)$$

However,  $N$  cannot be directly observed, so we condition on the total number of birds counted ( $x.$ ). The conditional probability that bird  $y$  was detected within the first interval given that it was detected in the entire 10 min is:

$$\pi_1 = P(y \in x_1 | y \in x.) = \frac{1 - cq^3}{1 - cq^{10}} \quad (7)$$

The conditional probability of first detecting bird  $y$  within the second interval given that it was detected in the entire 10 min is:

$$\pi_2 = P(y \in x_2 | y \in x.) = \frac{cq^3(1 - q^2)}{1 - cq^{10}} \quad (8)$$

Finally the conditional probability of first detecting bird  $y$  in the third interval given that it was detected in the entire 10 min is:

$$\pi_3 = P(y \in x_3 | y \in x.) = \frac{cq^5(1 - q^5)}{1 - cq^{10}} \quad (9)$$

Therefore the conditional multinomial has the probability density function:

$$f(x_1, x_2, x_3 | x.) = \frac{x.!}{x_1!x_2!x_3!} (\pi_1)^{x_1} (\pi_2)^{x_2} (\pi_3)^{x_3} \quad (10)$$

Now we can find the estimates of  $c$  and  $q$  that maximize the following likelihood function:

$$\begin{aligned} L(c, q | x_1, x_2, x_3) \\ &\propto \left[ \frac{1 - cq^3}{1 - cq^{10}} \right]^{x_1} \left[ \frac{cq^3(1 - q^2)}{1 - cq^{10}} \right]^{x_2} \left[ \frac{cq^5(1 - q^5)}{1 - cq^{10}} \right]^{x_3} \end{aligned} \quad (11)$$

We used program SURVIV (White 1983) to find the values of  $c$  and  $q$  that maximized the above likelihood function. SURVIV also computed associated estimates of the variances and covariance of  $c$  and  $q$ . We then reparameterized the SURVIV model to estimate the total detectability for the full 10 min ( $\hat{p}$ ) and its associated standard error. This parameter,  $\hat{p}$ , incorporates both group membership and the group detection probabilities and specifies the probability that an individual bird randomly selected from  $N$  is detected during the 10 min sampling period ( $p = 1 - cq^{10}$ ).

This model represents a modified special case of more general mixture models in which detectability of members of group 1 is estimated (not assumed to be one, as is done here). Norris and Pollock (1996) and Pledger (2000) fit full two-point mixture models to capture-recapture and removal data for closed populations. Because counts were divided into three intervals in our example, we were unable to fit those full two-point mixture models. At least four intervals are necessary to use the full two-point mixture models.

The model described here represents the most general (fully parameterized) model possible under this sampling design. This model can then be tested against more specific (reduced parameter) models. One such model constrains  $c$  to be equal to 1 and thus represents a model that does not attempt to incorporate heterogeneity (i.e. all birds are members of group 2). In addition, when different datasets are compared, the most general model will treat the estimates of  $c$  and  $q$  differently for each dataset. A more specific model will constrain the estimates of both  $c$  and  $q$  to be equal for both datasets ( $c_1 = c_2$ ;  $q_1 = q_2$ ). We used Akaike's Information Criterion (AIC; Burnham and Anderson 1998) for model selection and chose the model that most parsimoniously fit the data (i.e. model with minimum AIC).

*Estimating density.*—Once detection probability is estimated with either model, abundance can then be estimated as:

$$\hat{N} = \frac{x.}{\hat{p}} \quad (12)$$

When limited-radius point counts are used, this estimate of abundance can be used to estimate density simply as:

$$\hat{D} = \frac{\hat{N}}{A} \quad (13)$$

where  $A$  is the total area sampled (sum of the areas within each limited-radius count). Assuming that the counts ( $x$ ) are from a binomial distribution and independent of  $\hat{p}$ , an estimate of the variance of density at the sampled points is (after Nichols et al. 2000):

$$\widehat{\text{VAR}}(\hat{D}) = \frac{(x.)^2 \widehat{\text{VAR}}(\hat{p})}{A^2 \hat{p}^4} + \frac{(x.)(1 - \hat{p})}{A^2 \hat{p}^2} \quad (14)$$

If the estimated density were to be extrapolated to a defined area from which counts were sampled, a more formal estimate of the variance of density should include the variance associated with differences in counts over the sampled locations within the larger space (e.g. Thompson 1992). For the purposes of illustration in this paper, we will only consider the sampling variance represented by Eq. 14.

Model assumptions are as follows:

1. There is no change in the population of birds within the detection radius during the point count (i.e. the population is closed).
2. There is no double-counting of individuals.
3. All members of group 1 are detected in the first interval.
4. All members of group 2 that have not yet been detected have a constant per minute probability of being detected.
5. If counts with limited-radius are used, observers accurately assign birds to within or beyond the radius used.

**Field data.**—We applied this approach to data collected in Great Smoky Mountains National Park from 1993 to 1995. We conducted counts at 258 locations up to three times within each year for a total of 1,221 point counts (some locations were not surveyed in every year). Surveys were located in closed-canopy deciduous hardwood forests. In those areas with high canopy (20–30 m) and dense vegetation, most detections were recorded by ear. Our population of interest was therefore vocalizing birds (i.e. birds with non-negligible probability of vocalizing during the period of the point count). For the four songbird species to be discussed at length, Ovenbird, Black-throated Green Warbler, Red-eyed Vireo, and Black-throated Blue Warbler (see Appendix for scientific names), we recorded singing males, ignoring nonsong vocalizations. For each count, the total number of birds counted was divided into those detected within the first 3 min, the subsequent 2 min, and the final 5 min as described above. In addition, birds were recorded as within 50 m from the observer or beyond 50 m.

**Unlimited-radius counts.**—For the 15 most frequently detected species, we fit the data to four models. Model  $M^c$ , the most general model, estimated sepa-

rate parameters for each species and included heterogeneity (using the term  $c$  described above). Model  $M_a$  allowed heterogeneity among individuals but did not fit separate parameters for the different species. Model  $M^s$  fit the data separately for the different species, but had the constraint  $c = 1$ . Thus this model did not incorporate heterogeneity. And model  $M$ , with only a single parameter, did not account for heterogeneity ( $c = 1$ ), nor did it fit different estimates of  $q$  for each species.

For each of the four most frequently recorded species (Ovenbird, Black-throated Green Warbler, Red-eyed Vireo, and Black-throated Blue Warbler) we examined how temporal changes in bird activity influenced detectability. We compared detectability between those points conducted early in the morning (sunrise to 0745 EST; 610 points) with those conducted late in the morning (0746 to 1000; 611 points). If birds sing more frequently in early morning, we would expect them to have higher detectabilities at that time. In addition to the four species above, we also examined detection probability using combined data for Wood Thrush and Veery. Thrushes sing more frequently early in the morning than later in the day. Therefore they should have a higher detection probability earlier in the morning. We tested each of four models with data from the four most common species and the combined data for the thrushes. Model  $M^c$  incorporated heterogeneity and estimated different parameters for early morning and late morning. Model  $M_a$  incorporated heterogeneity but did not distinguish between early and late morning counts. Model  $M^s$  estimated different parameters for early and late morning, but did not incorporate heterogeneity. And model  $M$  did not incorporate heterogeneity or distinguish between early and late morning.

We also compared detectability between counts conducted at different times during the breeding season. Different bird species may have different peaks in singing frequency due to differences in nesting behavior, which will change their detectability. Species that nest early and attempt only one brood may have reduced singing frequencies by late spring compared to species that raise multiple broods. We separated the counts conducted on or before 20 May (early spring; 563 points), and those conducted after 20 May (late spring; 658 points). We fit four models to examine seasonal changes in detectability. Model  $M^c$  incorporated heterogeneity and estimated different parameters for early and late spring. Model  $M_a$  incorporated heterogeneity but did not distinguish between early and late spring. Model  $M^s$  estimated different parameters for the early and late spring, but did not incorporate heterogeneity. And model  $M$  did not incorporate heterogeneity or distinguish between early and late spring.

Observer variability related to differences in skill or hearing acuity is another potential factor affecting



detection probability. We compared detectability using the three observers for which we had the largest number of counts (observer 1 conducted 255 counts, observer 2 conducted 200, and observer 3 conducted 178). Again we fit four models for data from the four most common species. Model  $M_c^o$  incorporated heterogeneity and estimated different parameters for the different observers. Model  $M_c$  incorporated heterogeneity but did not distinguish between observers. Model  $M^o$  estimated different parameters for each observer, but did not incorporate heterogeneity. And model  $M$  did not incorporate heterogeneity or distinguish between observers.

The analyses described above only address issues of detection probability. For those analyses, each point count was considered an independent trial with respect to the estimate of detection probability, and unlimited-radius counts were used. Multiple counts were conducted at the same locations so we were not able to estimate abundance or density.

**Limited-radius counts.**—To estimate abundance and density, we restricted the analysis to two visits to each location in 1994 (155 locations). Each visit was treated as a separate sampling of abundance for the 155 counts. Only birds detected within 50 m of the observer were included because an estimate of density requires a measurement of the area sampled. Counts were separated by approximately two weeks between the first and second visit to each point location. It therefore seemed reasonable to expect the true abundance (and density) to be the same for those two visits. The same four species were used in this analysis. For each species, model selection was performed to choose between a model that incorporated heterogeneity ( $M_c$ ) and a model that did not ( $M$ ) as described above. Using the estimated detection probability ( $\hat{p}$ ) from the most parsimonious model, we estimated abundance and density for each visit. We compared estimated density between visits for each species by estimating difference between densities for the two periods.

## RESULTS

For the 15 most frequently detected species, the most parsimonious model was  $M_c^o$  (AIC values:  $M_c^o = 249.5$ ,  $M_c = 370.3$ ,  $M^o = 570.7$ , and  $M = 728.8$ ). Thus, there was strong evidence of differences in estimated detectability for different species (Fig. 1). Also, heterogeneity appeared to be an important component of the detectability requiring an estimate for parameter  $c$  for each species.

We did not find evidence of a change in detection probability at different times of the morning for three species. Model  $M_c$  received the most support for the data for Ovenbird,

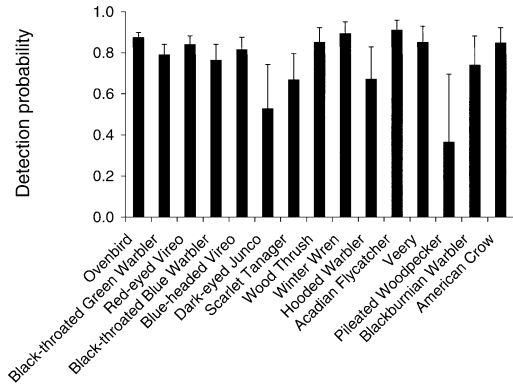


FIG. 1. Estimated detection probability during unlimited-radius counts for the 15 most frequently encountered species, ordered from most common (left) to least common (right). Error bars represent one estimated standard error. See Appendix for scientific names and count data.

Black-throated Green Warbler, and Red-eyed Vireo. However, model  $M_c^o$  received the most support for Black-throated Blue Warbler and the combined data for thrushes (Wood Thrush and Veery). Estimated detection probability appeared to decrease later in the morning for those species (Table 1). We did not find evidence of an influence of time of season on detectability for three of the four species tested. Model  $M_c$  received the most support for the data for Ovenbird, Black-throated Green Warbler, and Black-throated Blue Warbler. Model  $M_c^o$  was best supported for Red-eyed Vireo; however the estimated detection probability for early spring was only slightly lower than that for late spring (Table 2).

We found evidence of different detection probabilities for different observers for two species (model  $M_c^o$  received the most support). Observer 1 had higher estimated detectability than observers 2 and 3 for Ovenbird and Black-throated Green Warbler. Detectability for Red-eyed Vireo and Black-throated Blue Warbler did not appear to vary among observers (model  $M$  received the most support; Table 3). When testing Black-throated Blue Warbler, we were unable to estimate the parameters for each observer under model  $M_c^o$  due to the small number of detections during the middle time interval by observer 2. Model  $M_c$  was best supported when observers 1 and 3 were compared, suggesting no difference in estimated detection probability for these two observers.

TABLE 1. Detection probabilities during unlimited-radius counts conducted in early (at or before 0745) and late (after 0745) morning. AIC for model that received the most support is underlined. Selection of model  $M_c$  indicates different estimated detection probabilities for early and late morning. See text for description of models.

Time of count	Counts <sup>a</sup>			AIC of models				$\hat{p}$ (SE)	
	$x^1$	$x^2$	$x^3$	$M_c^t$	$M_c$	$M^t$	$M$	Model $M_c^t$	Model $M_c$
<b>Ovenbird</b>									
Early morning	1,003	170	233	35.5	<u>32.9</u>	115.6	114.4	0.90 (0.03)	0.87 (0.03)
Late morning	836	134	216					0.83 (0.08)	—
<b>Black-throated Green Warbler</b>									
Early morning	591	123	403	34.7	<u>31.4</u>	89.1	87.5	0.80 (0.07)	0.79 (0.05)
Late morning	631	118	202					0.78 (0.08)	—
<b>Red-eyed Vireo</b>									
Early morning	502	101	149	33.6	<u>30.5</u>	63.5	62.5	0.86 (0.05)	0.84 (0.04)
Late morning	426	90	143					0.81 (0.07)	—
<b>Black-throated Blue Warbler</b>									
Early morning	287	74	100	<u>32.5</u>	35.2	69.3	67.6	0.87 (0.05)	0.76 (0.08)
Late morning	405	65	140					0.51 (0.29)	—
<b>Thrush species (Wood Thrush and Veery)</b>									
Early morning	376	66	89	<u>31.3</u>	33.6	50.7	55.7	0.90 (0.04)	0.85 (0.05)
Late morning	199	41	74					0.72 (0.17)	—

<sup>a</sup>  $x_i$  is number of birds first detected in the  $i$ th interval.

*Estimating density.*—Using the subset of independent limited-radius counts, the best supported model for Ovenbird and Black-throated Green Warbler included heterogeneity (model  $M$ ) for both visits. The selected model for Black-throated Blue Warbler did not include heterogeneity (model  $M$ ) for either visit. And the best supported model for Red-eyed Vireo was  $M$  for the first visit and  $M_c$  for the second visit. Esti-

mates of detection probability ranged from 0.81 (0.19 SE) for Red-eyed Vireo to 0.97 (0.02 SE) for Black-throated Blue Warbler (Table 4). And the estimates of density ranged from 0.49 (0.02 SE) singing birds per hectare for Black-throated Blue Warbler to 1.93 (0.16 SE) singing birds per hectare for Ovenbird. Estimates of the difference in density between visits ( $\Delta\hat{D}$ ) ranged from 0.04 birds per hectare for Black-throated Blue War-

TABLE 2. Detection probabilities during unlimited-radius counts conducted in early (on or before 20 May) and late (after 20 May) spring. AIC for model that received the most support is underlined. Selection of model  $M_c^b$  indicates different estimated detection probabilities for early and late spring. See text for description of models.

Time of count	Counts <sup>a</sup>			AIC of models				$\hat{p}$ (SE)	
	$x_1$	$x_2$	$x_3$	$M_c^b$	$M_c$	$M^b$	$M$	Model $M_c^b$	Model $M_c$
<b>Ovenbird</b>									
Early spring	927	163	223	35.5	<u>32.8</u>	116.3	114.3	0.90 (0.03)	0.87 (0.03)
Late spring	912	141	226					0.84 (0.05)	—
<b>Black-throated Green Warbler</b>									
Early spring	629	131	209	34.7	<u>31.4</u>	89.4	87.5	0.81 (0.06)	0.79 (0.05)
Late spring	593	110	194					0.76 (0.09)	—
<b>Red-eyed Vireo</b>									
Early spring	443	103	159	<u>33.6</u>	35.0	63.7	67.0	0.82 (0.06)	0.84 (0.04)
Late spring	485	88	133					0.85 (0.06)	—
<b>Black-throated Blue Warbler</b>									
Early spring	289	56	109	32.5	<u>29.8</u>	63.2	62.1	0.65 (0.20)	0.76 (0.08)
Late spring	403	83	131					0.82 (0.07)	—

<sup>a</sup>  $x_i$  is number of birds first detected in the  $i$ th interval.

TABLE 3. Detection probabilities during unlimited-radius counts conducted by different observers. AIC for model that received the most support is underlined. Selection of model  $M_c$  indicates different estimated detection probabilities for different observers. See text for description of models.

Observer	Counts <sup>a</sup>			AIC of models				$\hat{p}$ (SE)	
	$x_1$	$x_2$	$x_3$	$M_c$	$M$	$M^o$		Model $M_c$	Model $M_c$
Ovenbird									
Observer 1	451	57	83	<u>47.2</u>	53.6	92.4	102.8	0.90 (0.05)	0.87 (0.04)
Observer 2	319	49	78					0.84 (0.09)	—
Observer 3	261	56	83					0.85 (0.07)	—
Black-throated Green Warbler									
Observer 1	295	46	62	<u>45.4</u>	51.6	60.1	71.0	0.91 (0.05)	0.86 (0.05)
Observer 2	187	44	65					0.84 (0.08)	—
Observer 3	146	37	59					0.80 (0.12)	—
Red-eyed Vireo									
Observer 1	218	27	54	43.8	<u>42.8</u>	67.4	67.7	0.68 (0.29)	0.80 (0.08)
Observer 2	151	32	49					0.83 (0.11)	—
Observer 3	137	31	47					0.83 (0.11)	—
Black-throated Blue Warbler <sup>b</sup>									
Observer 1	153	21	36	28.0	<u>25.7</u>	42.0	41.8	0.81 (0.17)	0.79 (0.13)
Observer 3	123	22	39					0.76 (0.21)	—

<sup>a</sup>  $x_i$  is number of birds first detected in the  $i$ th interval.  
<sup>b</sup> Models could not estimate separate parameters for observer 2.

bler to 0.39 birds per hectare for Red-eyed Vireo. The associated 95% confidence intervals for  $\Delta\hat{D}$  included zero for all species.

DISCUSSION

Application of a removal model to point count surveys divided into time intervals offers a promising new approach for estimating

detectability. Detectability estimates allow for comparisons among datasets without having to resort to using counts as an index of abundance. One strength of that procedure is that it can be applied to existing data as we have done here. In addition, it may be incorporated into future studies with no additional cost and without much additional training.

TABLE 4. Estimation of density for two visits to the same count locations in 1994. AIC for model that received the most support is underlined. The low-AIC model was used to estimate  $p$  and  $D$ . Confidence interval (95%) for estimated difference in density,  $\Delta\hat{D}$ , included zero for all species.

Visit	Counts <sup>a</sup>			AIC		$\hat{p}$ (SE)	$\hat{D}$ (SE)	$\Delta\hat{D}$ (95% CI)
	$x_1$	$x_2$	$x_3$	$M_c$	$M$		Birds ha <sup>-1</sup>	Birds ha <sup>-1</sup>
Ovenbird								
First visit	141	29	39	<u>14</u>	16	0.89 (0.07)	1.93 (0.16)	0.12 (−0.27, 0.51)
Second visit	152	23	29	<u>14</u>	17	0.93 (0.06)	1.81 (0.11)	—
Black-throated Green Warbler								
First visit	96	18	26	<u>14</u>	15	0.87 (0.11)	1.32 (0.17)	0.26 (−0.14, 0.66)
Second visit	118	26	32	<u>14</u>	14	0.91 (0.06)	1.58 (0.11)	—
Red-eyed Vireo								
First visit	64	19	25	<u>13</u>	12	0.92 (0.03)	0.96 (0.04)	0.39 (−0.23, 1.01)
Second visit	94	15	25	<u>13</u>	<u>17</u>	0.81 (0.19)	1.35 (0.31)	—
Black-throated Blue Warbler								
First visit	43	11	9	<u>12</u>	<u>10</u>	0.97 (0.02)	0.53 (0.01)	0.04 (−0.01, 0.10)
Second visit	35	9	12	12	<u>10</u>	0.93 (0.04)	0.49 (0.02)	—

<sup>a</sup>  $x_i$  is number of birds first detected in the  $i$ th interval.



Many of the differences we observed in detectability were probably due to differences in singing frequency. The species with the highest estimated detectabilities were Winter Wren ( $\hat{p} = 0.89 \pm 0.06 \text{ SE}$ ) and Acadian Flycatcher ( $\hat{p} = 0.91 \pm 0.05 \text{ SE}$ ). Winter Wrens and Acadian Flycatchers sing frequently making them easy to detect on point count surveys. Conversely, Pileated Woodpeckers, which give loud but infrequent vocalizations, had the lowest detectability estimate ( $\hat{p} = 0.36 \pm 0.33 \text{ SE}$ ) among the 15 most frequently detected species. We also found a decrease in detection probability for thrushes from early morning to late morning counts (Table 1), probably due to a tendency for thrushes to sing most frequently early in the morning and only sporadically later in the day.

We found differences in the detection probabilities between observers for certain species. That may reflect differences in hearing abilities among observers. Different observers may have different sensitivities to songs of particular species, allowing them to detect a greater proportion of those species than other observers. Observers may also show favoritism to some species, recording those species preferentially when uncertain about identification (Bart 1985). Detectability of a species declines with increasing density of that species, and that may be especially true when observers are recording multiple species at the same time (Scott and Ramsey 1981, Bart and Shoultz 1984). The observers in our field study recorded all individuals of all species detected. The average number of detections per point count was high (11.8 birds) and observers' abilities to discriminate individuals may have been compromised on some counts. We found that observers often focused their attention on one species at a time to help them discriminate individuals of that species. If observers dedicate different amounts of effort (time focused on one species) for different species, that could lead to observer differences in detectability.

In our examples, we have conducted separate modeling efforts for individual species. We have done so because that is the traditional approach to estimating parameters of animal populations. Species differences are often suspected to be so large as to preclude simultaneous modeling of multiple species. However, as noted by Nichols et al. (2000), it may be reasonable to estimate detectability by groups of

species expected *a priori* to exhibit similar detection probabilities. Similarly, we estimated the influence on detectability separately for each of the factors discussed (time of day, time of season, and observer). Our approach could be applied to a more general model capable of examining multiple factors simultaneously and detecting interactions among them (for example, an interaction between time of day and time of season). However, such an analysis would require a larger data set than we used here. Parsimonious modeling of detection probability might include multiple species with the same detection probabilities, observer differences with parallel effects for different species, and possibly interactions between species, observers, and other factors. Such modeling of multiple species and interaction among factors can be readily implemented using the general modeling framework that we have presented and should be an area of active research.

The estimates for density were developed from a small data set (155 limited-radius counts). The two visits to each count location should represent two samples from the same population. Therefore we did not expect the estimates of density to differ between visits. Indeed, although estimates of density differed among species, they did not differ within species for the two visits. For the purposes of this article, we only dealt with sampling variance at the actual locations. This was adequate for our tests because we were testing estimated density for two visits to the same count locations. We did not have independent estimates of density (e.g. spot-map data or nest locations) with which we could compare these density estimates. Future work should attempt to compare estimates generated from this removal sampling procedure to known density and to results of other estimation procedures (e.g. see Tarvin et al. 1998).

**Model assumptions.**—Assumption 1: there is no change in the population of birds within the detection radius during the point count. This assumption of closure may not be met for some species during a 10 min count. This should be less of a problem for small breeding songbirds, such as Wood Warblers, with their relatively small territories and high singing rates. However for larger ranging species like Pileated Woodpecker and American Crow, this assumption is more likely to be violated. The model

will work equally well for point counts of shorter duration (e.g. 5 min) that are more likely to meet the assumption of closure, provided the count is divided into three or more intervals.

Assumption 2: there is no double-counting of individuals. The somewhat long duration of the point counts used in this analysis (10 min) may lead to violation of this assumption. However, observers were trained to be conservative in this regard. Similarly, there may have been identification errors. Such phantom detections would result in recording species that are not present and inflating the number of individuals recorded for some species that are present (see Bart and Schoultz 1984, Bart 1985). Whereas violations of assumptions 1 and 2 may present problems for this analysis, they are not unique to this method; they are also necessary for virtually any analysis of point-count data, even naïve analyses that do not adjust for detectability.

Assumptions 3 and 4: all members of group 1 are detected in the first interval; all members of group 2 have a constant per minute probability of being detected. These two assumptions are likely to be violated to some degree. This modeling device should not be interpreted literally. With the available count data divided into three time intervals, this was the best way we could address the issue of heterogeneity of detection probability. Our procedure included model selection using AIC to choose between models with one group ( $c = 1$ ) and models with two groups. The critical parameter when estimating abundance in the face of heterogeneous detection probabilities among individuals is the coefficient of variation of the distribution of detection probabilities (Carothers 1973). Carothers (1973) was the first to note that this variation could be modeled adequately using a two-point distribution, and Pledger (2000) successfully exploited this approach as well. Specifically, Pledger (2000) demonstrated that using a model with two groups (each with a homogenous detection probability) was adequate to provide an unbiased estimate of population size even when the population was composed of many such groups. Our model constrained group 1 to have a detection probability of one because our counts were only divided into three intervals.

Another way in which these assumptions may be violated is if the detection probability

varies through time during the count. For example, even if there were two uniform groups, members of group 2 could have higher detectability in the fourth minute than in the eighth minute, although it may not be easy to develop a plausible biological story for such variation, especially with likely variation in starting times of different point counts. Even when assumptions such as these are likely to be violated, use of this model-based approach is likely to be far more robust than index methods that assume counts to be a constant fraction of the sampled populations (e.g. see Nichols and Pollock 1983).

Assumption 5: if counts with limited-radius are used, observers accurately assign birds to within or beyond the radius used. In order to estimate density, some measure of area sampled is necessary. In this study, observers were trained to estimate distance to birds and assign each detection to within or beyond 50 m from the observer. Even with training and experience, it is often difficult to estimate distances to birds on the basis of hearing songs. This assumption is also required for distance sampling and virtually any method of density estimation.

*Recommendations and future work.*—We constructed this particular model with three time intervals because this count procedure was recommended by Ralph et al. (1995). We hope this will facilitate its use in analyzing existing data. However, our model is flexible enough to accommodate data collected in other ways. We recommend that future surveys be designed to include four or more time intervals of equal duration. That would allow the use of full two-point mixture models and would simplify the mathematics (see Pledger 2000). For example, a 10 min point count divided into five intervals of 2 min each would allow use of the more general model and goodness-of-fit tests for all three models ( $M$ ,  $M_c$ , and the full two-point mixture model). However, to avoid violation of assumptions 1 and 2, short counts may be preferable in some instances. Perhaps a 5 min count divided into 1 min intervals would be appropriate.

In addition, combining the removal approach with other current methods may provide improved estimates of detectability and density. For example, a procedure that combines removal sampling with distance sampling would al-

low estimation of two separate components of detectability: probability a bird sings and probability song is detected as a function of distance from observer. Similarly, combining the removal approach with double-observer sampling could also address two components of detectability: probability a bird sings and probability a song is detected by at least one observer. Perhaps all three techniques could be combined into one study with known density to evaluate the merits of each method and the various combinations thereof.

#### ACKNOWLEDGMENTS

The U.S. Geological Survey, National Park Service, and Environmental Protection Agency provided funding for this project. Jon Bart, Jeff Brawn, and Jay Rotella provided useful suggestions on an earlier draft of this manuscript. We also thank the many members of the field crews that collected the survey data: David Aborn, Jeff Clark, John Garesche, Brad Goodner, Susan Moegenburg, Kerry Rabenold, Susan Shriner, and Barbara Stedman. Examples of the SURVIV code and equations used to estimate  $\hat{p}$  and  $\hat{D}$  are available online at [www.mbr-pwrc.usgs.gov/software/CountRemoval.html](http://www.mbr-pwrc.usgs.gov/software/CountRemoval.html).

#### LITERATURE CITED

- 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.
- BART, J. 1985. Causes of recording errors in singing bird surveys. *Wilson Bulletin* 97:161–172.
- BART, J., AND J. HERRICK. 1984. Diurnal timing of bird surveys. *Auk* 101:384–387.
- BART, J., AND J. D. SCHOULTZ. 1984. Reliability of singing bird surveys: Change in observer efficiency with avian density. *Auk* 101:307–318.
- BEST, L. B. 1981. Seasonal changes in detection of individual bird species. *Studies in Avian Biology* 6:252–261.
- BUCKLAND, S. T., K. P. BURNHAM, D. R. ANDERSON, AND J. L. LAKE. 1993. *Density Estimation Using Distance Sampling*. Chapman and Hall, London.
- BURNHAM, K. P. 1981. Summarizing remarks: Environmental influences. *Studies in Avian Biology* 6:324–325.
- BURNHAM, K. P., AND D. R. ANDERSON. 1998. *Model Selection and Inference: A Practical Information Theoretic Approach*. Academic Press, New York.
- CAROTHERS, A. D. 1973. Effects of unequal catchability on Jolly-Seber estimates. *Biometrics* 29:79–100.
- 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.
- LANCIA, R. A., J. D. NICHOLS, AND K. H. POLLOCK. 1994. Estimating the number of animals in wildlife populations. Pages 215–253 in *Research and Management Techniques for Wildlife and Habitats* (T. Bookhout, Ed.). Wildlife Society, Bethesda, Maryland.
- MC SHEA, 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.
- MORAN, P. A. P. 1951. A mathematical theory of animal trapping. *Biometrika* 38:307–311.
- 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* 117:293–408.
- NICHOLS, J. D., AND K. H. POLLOCK. 1983. Estimation methodology in contemporary small mammal capture–recapture studies. *Journal of Mammalogy* 64:253–260.
- NORRIS, J. L., AND K. H. POLLOCK. 1996. Nonparametric maximum likelihood estimators for population size 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.
- RALPH, C. J. 1981. An investigation of the effect of seasonal activity levels on avian censusing. *Studies in Avian Biology* 6:265–270.
- RALPH, C. J., S. DROEGE, AND J. R. SAUER. 1995. Managing and monitoring birds using point counts: Standards and applications. Pages 161–168 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.
- 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. *Studies in Avian Biology* 6:297–300.
- ROBBINS, C. S. 1981. Effect of time of day on bird activity. *Studies in Avian Biology* 6:275–286.

- 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. Effect of abundant species on the ability of observers to make accurate counts of birds. *Auk* 98:610–613.
- SEBER, G. A. F. 1982. Estimation of Animal Abundance and Related Parameters. Macmillan, New York.
- SKIRVIN, A. A. 1981. Effect of time of day and time of season on the number of observations and density estimates of breeding birds. *Studies in Avian Biology* 6:271–274.
- 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.
- THOMPSON, S. K. 1992. Sampling. John Wiley and Sons Inc., New York.
- 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., D. R. ANDERSON, K. P. BURNHAM, AND D. L. OTIS. 1982. Capture-recapture and removal methods for sampling closed populations. Los Alamos National Laboratory Publication LA-8787-NERP, Los Alamos, New Mexico.
- WILSON, D. M., AND J. BART. 1985. Reliability of singing bird surveys: Effects of song phenology during the breeding season. *Condor* 87:69–73.
- ZIPPIN, C. 1958. The removal method of population estimation. *Journal of Wildlife Management* 22: 82–90.

Associate Editor: J. Brawn

# APPENDIX. Scientific names and total counts for species in Figure 1.

Common name	Scientific name	Counts <sup>a</sup>		
		$x_1$	$x_2$	$x_3$
Ovenbird	<i>Seiurus aurocapillus</i>	1,839	304	449
Black-throated Green Warbler	<i>Dendroica virens</i>	1,222	241	403
Red-eyed Vireo	<i>Vireo olivaceus</i>	928	191	292
Black-throated Blue Warbler	<i>Dendroica caerulescens</i>	692	139	240
Blue-headed Vireo	<i>Vireo solitarius</i>	645	129	207
Dark-eyed Junco	<i>Junco hyemalis</i>	425	90	188
Scarlet Tanager	<i>Piranga olivacea</i>	392	98	182
Wood Thrush	<i>Hylocichla mustelina</i>	340	59	91
Winter Wren	<i>Troglodytes troglodytes</i>	351	49	71
Hooded Warbler	<i>Wilsonia citrina</i>	233	62	114
Acadian Flycatcher	<i>Empidonax virescens</i>	257	46	60
Veery	<i>Catharus fuscescens</i>	235	48	72
Pileated Woodpecker	<i>Dryocopus pileatus</i>	144	53	116
Blackburnian Warbler	<i>Dendroica fusca</i>	169	45	77
American Crow	<i>Corvus brachyrhynchos</i>	159	51	71
Other species		1,162	332	723

<sup>a</sup>  $x_i$  is number of birds detected in the  $i$ th interval.