

Statistical Aspects of Point Count Sampling¹

Richard J. Barker and John R. Sauer²

Abstract: The dominant feature of point counts is that they do not census birds, but instead provide incomplete counts of individuals present within a survey plot. Considering a simple model for point count sampling, we demonstrate that use of these incomplete counts can bias estimators and testing procedures, leading to inappropriate conclusions. A large portion of the variability in point counts is caused by the incomplete counting, and this within-count variation can be confounded with ecologically meaningful variation. We recommend caution in the analysis of estimates obtained from point counts. Using our model, we also consider optimal allocation of sampling effort. The critical step in the optimization process is in determining the goals of the study and methods that will be used to meet these goals. By explicitly defining the constraints on sampling and by estimating the relationship between precision and bias of estimators and time spent counting, we can predict the optimal time at a point for each of several monitoring goals. In general, time spent at a point will differ depending on the goals of the study.

Most observational studies of birds have a goal of characterizing the numbers or distribution of birds through time or space using observations of the number of birds counted at randomly selected points, or clusters of points. Most biologists analyze these data using standard sampling estimation procedures (Cochran 1977) and assume that the counts can be used in place of exact measurements of bird abundance at the points. Unfortunately, most point counts miss over 50 percent of the individual birds at any point (Sauer and others, 1994a). Not only are the data incomplete counts, or indices of bird abundance, but usually we cannot estimate the proportion of birds counted, or detection probabilities, associated with each sample of counts. Burnham (1981) stated that without estimating detection probabilities, the use of counts as indices of abundance is scientifically unsound and unreliable. He also suggested that estimation of detection probabilities could be achieved relatively easily using detection distance data; unfortunately this view is probably overly optimistic for the majority of bird species that large-scale monitoring programs are designed to cover. Ironically, in these programs samples are collected in many habitats by many observers, and detection probabilities are unlikely to be comparable among points.

Therefore, it is critical that we address the question of how unmodeled detection probabilities affect inference based on point counts, and how sampling at points should be conducted to minimize possible inefficiencies and biases associated with incomplete counts. In this paper we develop a

model for point count sampling and consider how incomplete counting affects some of the common estimators of abundance and population trend. We then describe optimal allocation of sampling effort in point counts under our model.

Model

In this section, we develop a model for how animals are sampled at points. All statistical analyses require an underlying model, and we use the model to assess the effects of incomplete counts on commonly used estimation procedures. The model reflects a view of how an underlying population of individuals at a point is counted. Development of the model requires some discussion of the mathematical formalism, which should be of interest to quantitative ecologists.

Let c_i denote a count at the i th sampling location (point) from a population of N_i animals. Among points assume that animals are independent and identically distributed with mean μ and variance σ^2 . To model the incomplete counting process we assume the c_i are binomial random variables with parameters N_i and detection probability p , thus

$$E[c | N, p] = Np, \quad (1)$$

$$E[c] = \mu p, \quad (2)$$

$$Var(c | N, p) = Np(1-p), \quad (3)$$

and

$$Var(c) = \mu p(1-p) + p^2 \sigma^2. \quad (4)$$

The E notation represents mathematical expectation.

Under this model, double counting is not allowed, and p is constrained between 0 and 1. Alternatively, one could model the counts conditional on N as Poisson random variables with parameter pN which would allow p to exceed 1, as may occur in some cases (Bart and Schoultz 1984). The Poisson model may also be advantageous in more complex models, such as models for the unconditional distribution of bird counts in space or time, as certain mixtures of the Poisson distribution can lead to other well known distributions such as the negative-binomial or the Neyman type A . As our intention is to discuss the sampling process, we consider only the binomial model, which is conceptually easy to understand in the point count context.

There are two basic ways to conceptualize what we refer to as incomplete counting. In the first case, as represented by unlimited distance point counts, all animals noted by the observer at a point are recorded. The region sampled at the i th point can then be described as that area lying within a circle of some unknown radius about the observer, the boundary of which marks the point at which detection probability can effectively be considered as zero. The number of animals associated with this point is then the collection of birds that are "located" within this circle, and c_i is the count obtained from this population. The second case, represented by limited

¹ An abbreviated version of this paper was presented at the Workshop on Monitoring Bird Population Trends by Point Counts, November 6-7, 1991, Beltsville, Maryland.

² Statistician, Florida Cooperative Fish and Wildlife Research Unit, Department of Wildlife and Range Science, University of Florida, Gainesville FL 32611, present address: Department of Statistics, Massey University, Palmerston North, New Zealand; and Research Wildlife Biologist, Patuxent Wildlife Research Center, USDI National Biological Service, Laurel, MD 20708

distance sampling, is described in a similar manner, except at the time of the count only those animals present within some fixed radius of the observer are recorded. The detection probability, as described above, is now the product of the probability that any one of the N_i birds is located within the fixed distance of the observer at the time of the count, and the probability that it is detected by the observer given that it is present within the fixed radius. This latter probability is usually assumed to be 1.

Estimators

In this section, we use the model to evaluate the possible effects of incomplete counting on estimators of various population parameters. Data from point counts are used to estimate: (1) mean relative abundance and its variance; (2) differences between relative abundances for populations in different habitats or regions; (3) population trends; and (4) species richness. Point count data are often reasonable surrogates for total population sizes for all of these population parameters, but we demonstrate that the potential exists for analyses based upon point counts to not accurately reflect the real population.

Estimation of Total Counts

The mean and variance of the counts are given by expressions (2) and (4) above. Because p appears in both formulae, as detection probabilities decrease, both the average count and the variance of the counts decrease. Thus counts from populations with inherently low detection probabilities tend to also have low absolute variability. An implication of this, which we explore in a later section, is that using variance as a measure of count quality is not a good idea. Considering the coefficient of variation of the counts (CV_c), however, from (2) and (4) it can be shown that

$$CV_c^2 = \frac{1-p}{\mu p} + CV_N^2, \quad (5)$$

which increases without bound as p tends to zero. Thus, not counting all the animals that are present increases the relative variability of the counts. If one is interested in modeling demographic processes, the information of interest is contained within the variance component described by the term CV_N^2 in expression (5). Clearly, however, if detection probabilities are not estimated, this information is confounded with the additional term associated with p . It is important to be aware of the presence of this additional component when interpreting index data. For example, variance associated with p can dominate the total variance, particularly for counts with low detection probabilities. Differences in variances over time or regions based on point counts therefore may reflect differences in the variability of counts, not real population differences in variance.

Ratio Estimator of Population Change

If it is assumed that detection probabilities remain constant through time, it is possible to obtain reasonable

estimates of population change between 2 years using the ratio estimator

$$\hat{\beta} \approx \frac{\sum_{i=1}^n c_{i,t+1}}{\sum_{i=1}^n c_{i,t}}, \quad (6)$$

where t indexes year. Barker, Link, and Sauer (personal communication) used the model described in the first section to derive analytical expressions for bias and variance of the estimator (6). They showed that if $p_{t+1} = p_t$, then

$$Bias(\hat{\beta}) \approx \frac{\beta}{n\mu} \frac{\{1-p\}}{p}, \quad (7)$$

and

$$Var(\hat{\beta}) \approx \frac{\beta(1+\beta)}{\mu} \frac{\{1-p\}}{p} + \sigma_\beta^2 \left\{ 1 + \frac{\sigma^2}{\mu^2} \right\}, \quad (8)$$

where β denotes the ratio μ_{t+1}/μ_t , σ_β^2 denotes the variance of N_{t+1}/N_t among points, and σ^2 denotes the variance of the number of animals among points in the first year. It is evident from expression (7) that without complete counts the ratio estimator (6) is biased, with the extent of bias determined by the detection probability (p) and the number of points sampled (n). In the case that p_{t+1} is exactly equal to p_t , the bias tends to zero for large samples of points. However, if this condition is violated, even if the p_t and p_{t+1} are random variables with the same expected values, the expression (7) contains another term, and the estimator is no longer unbiased for large sample sizes. Thus it is critically important to consider the assumption of constant detection probabilities over time when using incomplete counts to estimate population changes. Numerous studies have demonstrated that observers and environmental conditions affect the number of birds counted (Ralph and Scott 1981); thus it is likely that in practice an assumption of constant detection probabilities through time is unreasonable. Therefore, bias in estimation of population change from point counts will always exist, but can be minimized by large sample sizes if p does not change over time. Use of statistical methods such as the base-year method that estimate trends as products of these ratio estimators should be avoided (Barker, Link, and Sauer, personal communication).

Comparison of Average Counts by Habitat or Regions

If it can be assumed that detection probabilities are identical between study areas that have been sampled using point counts, then it is possible to use the count data to test for differences in bird abundance between those areas. Under our model, power of a two-sided z -test for a difference in

means between two sets of counts with identical detection probabilities and equal sampling effort is given by:

$$Power = 1 - \Phi \left[\frac{\{z_{\alpha/2}\sigma_0 - pk\mu_i\}}{\sigma_1} \right] + \Phi \left[\frac{\{-z_{\alpha/2}\sigma_0 - pk\mu_i\}}{\sigma_1} \right], \quad (9)$$

where $\Phi(z)$ denotes the standard normal cumulative density function evaluated at z , σ_0 denotes the standard deviation of the difference in means under the null hypothesis ($H_0: \mu_i = \mu_j$), σ_1 denotes the standard deviation of the difference in means under the alternative hypothesis ($H_1: \mu_i \neq \mu_j$), and where $\mu_j = (1-k)\mu_i$. The standard deviation of the difference in mean counts under the null and alternative hypotheses can easily be computed using expression (4). To demonstrate the effect of detection probabilities on power we generated power curves for a population with $\mu_1 = 20$ birds, $\sigma_1^2 = \sigma_2^2 = 20$, and for $p = 1, 0.75, 0.5,$ and 0.25 (fig. 1). The incomplete nature of the counting process leads to a reduction in power of the hypothesis tests, as shown by the decline in steepness of the curves as the p becomes smaller. More seri

equal detection probabilities between areas would have. Such a failure leads to a shift in the curve away from the origin leading to both increased power and increased Type I error rates (fig. 2). Even small changes in detection probabilities can invalidate the statistical tests. Thus, a change in detection probabilities between treatments can never be separated from a real population difference, and the null hypothesis gets rejected at a greater rate even when no difference in population means exists.

Allocation of Sampling Effort

Several components of the design of point count studies remain controversial. One primary source of disagreement is the optimal length of time spent sampling at points. In our opinion, consistency in design is desirable for monitoring programs, and one consequence of this workshop is a set of standards for the design of point count studies. However, selection of point count duration should be viewed as a statistical problem, for which one uses available information and a model to predict a sampling procedure with optimal properties. In this section, we discuss a method of optimizing point count duration and review how different goals for a survey can lead to different point count durations.

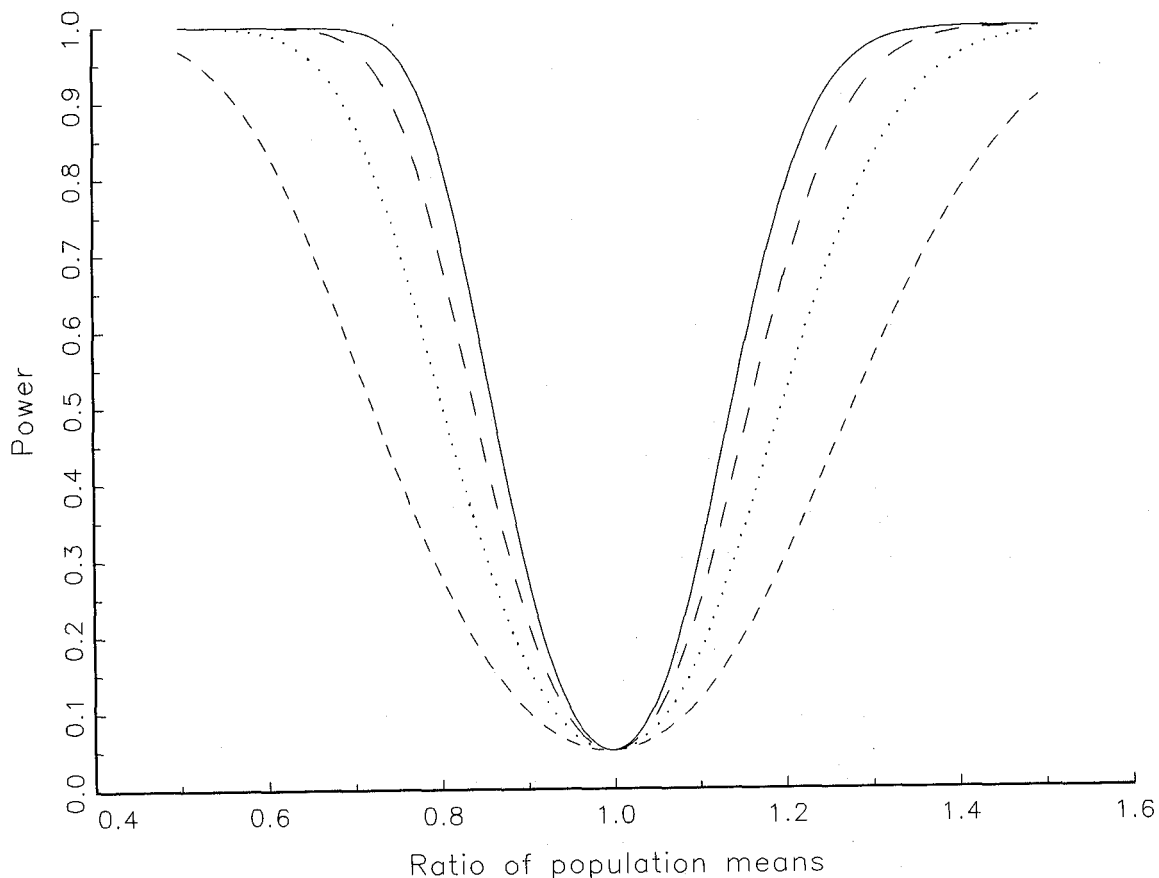


Figure 1--Power of a 2-sided z-test for a difference in means of two populations plotted as a function of the ratio of population means ($1-k$) and detection probability ($p = 1: \text{—}; p = 0.75: \text{---}; p = 0.5: \text{...}; p = 0.25: \text{- - -}$). Detection probabilities are the same in the two populations and 20 point counts are made from each population. The mean and variance of the number of birds present at each point are the same within groups, and in the first group of counts there are an average of 20 birds per site present (μ_1). At the second group of sites there are an average of $(1-k)\mu_2$ birds present per site.

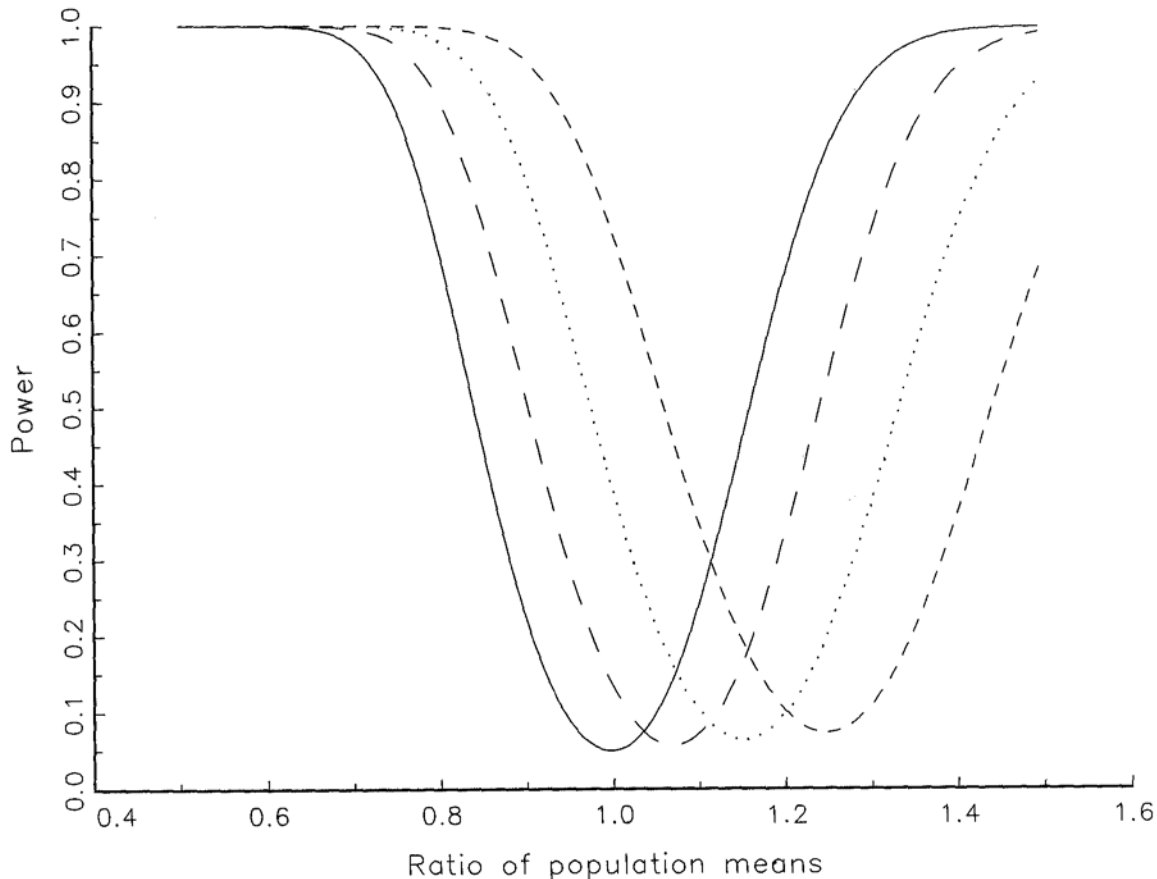


Figure 2--Power of a 2-sided z-test for a difference in means of two populations plotted as a function of the ratio of population means $(1-k)$. Detection probability in the first population $(p_1) = 0.75$, but is different in the second population $(p_2 = 0.75: \text{—}; p_2 = 0.70: \text{- - -}; p_2 = 0.5: \text{...}; p_2 = 0.25: \text{---})$. Twenty point counts are made from each population. The mean and variance of the number of birds present at each point are the same within groups, and in the first group of counts there are an average of 20 birds per site present (μ_1) . At the second group of sites there are an average of $(1-k)\mu_2$ birds present per site.

Clear Statements of Goals Are Necessary for Optimization

The first step in any optimization is to specify the appropriate estimators (or testing procedures) for the goals of the study, because optimal allocation of effort depends on the procedures used. For example, a study that is designed to test for differences in counts among habitats will be optimized differently from a survey designed for analysis of population trends. Of course, in practice, no study would begin without first identifying goals and evaluating if these goals can be met with the proposed study design. Once an estimator is specified, we can use its attributes to evaluate alternative study designs. For example, if we are interested in estimating population trend, we would want to design the study to minimize the mean squared error of the trend estimate.

Once the estimator or testing procedure has been selected, the next step is to obtain an expression that describes performance of the estimation procedure and is also a function of the time spent counting at a point. For example, it may be possible to express the relationship between detection probability and time spent surveying at a point; thus, for a fixed sampling time, the appropriate detection probability can be obtained. Many investigators have empirically described the relationship between point count duration and proportion

of total birds detected (Buskirk and MacDonald, in this volume), although it can also be estimated by using more sophisticated procedures (Dawson and others, in this volume). Finally we need to specify time and other constraints placed on sampling.

Modelling Detection Probabilities as a Function of Count Duration

Detection probabilities play an important role in allocating point count sampling effort. Typically the experimenter is faced with the problem of trying to decide how best to allocate number of points and time spent sampling at a given point. Because the proportion of animals detected increases with time spent sampling, bias and relative variance at a point tends to decrease as count duration increases. Conversely, estimator precision tends to decrease as duration increases because fewer points can be sampled, so between-point components of the total variance increase. To optimally allocate sampling effort, the experimenter needs a model describing these relationships and a criterion for survey performance that can be used to judge the process of trading off time spent at points and the number of points sampled.

To describe the process by which the proportion of birds counted increases with time (denoted as the function $p = f(T_s)$), we assume that this proportion increases from zero to one as the cumulative distribution function of the random variable, time to first detection. One way to model these times until first detection is as independent and identically distributed exponential random variables with mean r . In this case

$$f(t) = 1 - e^{-rt_s} \tag{10}$$

A drawback of this model is that it assumes that detection times are identically distributed. In practice, the parameter r may vary between individuals. For example, in the model described above, it is likely that in the case of unlimited distance methods, birds farther from the observer are counted with lower probability. In the case of fixed distance methods, it seems likely that the probability that one of the N_i birds associated with the sampling point occurs within the fixed radius at the time of the count decreases with distance from the observer.

A more flexible model is the Weibull model, in which the probability that the time to first detection is less than t is given by

$$f(t) = 1 - e^{-(ta)^b} \tag{11}$$

This model reduces to the exponential model when the "shape" parameter $b = 1$.

Given estimates of the parameters, the percentage of animals sighted in the sampling period T_s can be estimated by

$$\hat{f}(T_s) = 1 - e^{-\hat{r}T_s}, \tag{12}$$

under the exponential model, or by

$$\hat{f}(T_s) = 1 - e^{-(T_s \hat{a})^b}, \tag{13}$$

under the Weibull model. In both cases maximum likelihood parameter estimates can be easily obtained.

Optimization Procedure

Using the components described above, we can use standard statistical procedures to estimate a count length that will provide efficient sampling in the context of the goals of the study. We will provide a brief example of this optimization procedure. For a more extensive description of the method and the issues involved in the optimization, see Barker and others (1993).

Consider a survey in which sampling effort is constrained by the time available for sampling. We assume total time surveying can be described by

$$T = (n-1)T_t + nT_s. \tag{14}$$

where T is the total survey time, n is the number of points sampled, T_t is the travel time between points, and T_s is the sampling time spent at each point. We consider three possible goals of the study: estimation of (1) average count; (2) trend; or (3) z -test of difference between two study areas. Performance of these estimators is assessed either by minimizing mean square error (bias² + variance), or by maximizing test power (z -test for population change). Although not associated with a specific hypothesis, many investigators use total birds counted as a criterion for optimization.

Optimal allocation of sampling effort can now proceed using the constraint function (14), an estimate of the function $f(T_s)$ (obtained, say, from pilot data), and an appropriate measure of estimator performance (mean square error, test power, or total count). To obtain solutions we optimize the functions describing mean square error (minimize), test power (maximize), or total expected count (maximize) with respect to n or T_s . These are all functions of both n and T_s , so we use the constraint function to express the additional variable in terms of the variable to be optimized. Note that, as in other sample size allocation procedures, pilot estimates of population sizes and variances must also be used in the procedure.

Mathematical details of the optimization process are provided by Barker and others (1993), and they provide an example of optimization for a hypothetical bird population and several estimation procedures.

Numerical Example

We present a numerical example of the procedure for a bird population with an average of 20 birds present per site (μ), variance of the number of birds among sites of 40, 180 minutes of time available for sampling, travelling time of 10 units between sites, and an exponential parameter of 0.23 (table 1). This latter parameter corresponds to the parameter estimate that Barker and others (1993) obtained for the Hawaiian Thrush (*Myadestes obscurus*) from a published data set (Scott and Ramsey 1981). For this hypothetical population, we estimated allocation of sampling effort that corresponds to optimization criteria of: (1) the minimum mean square error of the count; (2) the maximum expected total count; (3) the maximum power of a 2-sided z -test for a difference in means of 10 percent; and (4) the minimum mean square error of the ratio estimator of population change (table 1). Allocations of point count lengths differ considerably among these criteria. Interestingly, maximum expected count does not coincide with maximum power of the z -test. The explanation lies in the fact that the maximum expected count depends only on the mean number of birds present and the detection probability, whereas test power depends on the variance of the number of birds among sites, as well as the number present, and the detection probability.

Discussion

From a statistical point of view, the dominant feature of point counts is that they do not completely census the population of animals that are associated with each sampling point. Instead, an incomplete count is obtained with the fraction of the population observed at each point unknown and varying. We have considered this sampling characteristic using a binomial sampling model and have demonstrated that it can bias estimates of number of animals present and trends. It also can be a substantial component of the variance of the counts, which is confounded with ecologically meaningful variation. These attributes of point count data suggest that estimates of population parameters based on them should be treated cautiously.

We recognize, however, that point counts are often the only source of data we have for most bird species. Because of their limitations, we believe that point counts are of best use in early-warning monitoring systems, but that population changes estimated from the count data should be used as a basis for further, more specific research. Attempts to use count data in sophisticated modeling procedures often lead to inappropriate results. Elsewhere, we have considered the effects of incomplete counts on modeling of density dependence, and we have shown that existing methods of detecting density dependence from incomplete count data are rendered worthless unless detection probabilities are close to 1, because of greatly elevated type 1 error rates (Barker and Sauer 1991).

If the survey is being used to describe trends for many species, "optimality" of sampling effort is an ill-defined concept because the allocation of sampling effort that leads to optimal performance of estimators or testing procedures is species specific. Thus, how should one allocate effort in such a multi-species program? This process itself must involve trading efficiency among competing species. One approach may be to pick key species from the assemblage associated with the study area and optimize effort with respect to the hardest species to sample. This will lead to a tendency to spend more time sampling at each point.

Table 1--Optimal sampling allocation of number of sites and time spent sampling at each site (T_s) for four optimization criteria. These results apply to a hypothetical study with mean number of animals present at each point of 20, variance of the number of animals among points of 40, total sampling time available of 180 minutes of time, 10 minutes of time required for travel between points, and the function relating detection probability at each site (p) to the time spent sampling at each site (T_s) given by $p = 1 - e^{-rT_s}$ where $r = 0.23$. MSE = Mean Square Error

Optimization criteria	Number of sites	T_s	p
Count MSE	5.24	26.26	0.99
Total count	11.31	6.80	0.79
Power	13.00	4.62	0.66
Ratio	13.06	4.55	0.66

If too little time is spent at each point, bias may dominate estimator performance.

In a single species study, optimization can proceed using the methods we have outlined above. We stress that the critical step in the process is in determining the goals of the study and the analytical means by which this goal is to be met. In the absence of such goals and methods, the notion of optimality is without meaning. If the goal is to estimate some population parameter (e.g., annual growth rate), we suggest that mean square error is an appropriate measure of estimator performance. If the goal is to test a specific hypothesis (e.g., comparing mean count between areas), we suggest that test power is an appropriate performance criterion. Because different tests are different functions of various population parameters, optimal performance in terms of test power is also specific to testing procedures.

We have not attempted to address issues related to estimation of species richness from point counts. Obviously, the observed species richness from a point count is a biased estimate of actual species richness in the same manner that counts are biased estimates of number of animals present. However, because species are the variable of interest, mark-recapture procedures can often be used to estimate species richness.

Ralph, C.J., J.R. Sauer, and Sam Droege, technical editors. 1995. Monitoring bird populations by point counts. Gen. Tech. Rep. PSW-GTR-149. Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture. 197 p.