

## Large scale wildlife monitoring studies: statistical methods for design and analysis

Kenneth H. Pollock<sup>1\*</sup>, James D. Nichols<sup>2</sup>, Theodore R. Simons<sup>3</sup>,  
George L. Farnsworth<sup>3</sup>, Larissa L. Bailey<sup>3</sup> and John R. Sauer<sup>2</sup>

<sup>1</sup>*Department of Statistics, Box 8203, North Carolina State University, Raleigh, NC 27695-8203, U.S.A.*

<sup>2</sup>*U.S. Geological Survey, Biological Resources Division, Patuxent Wildlife Research Center, Laurel, MD 20708, U.S.A.*

<sup>3</sup>*NC Cooperative Fish and Wildlife Research Unit, Department of Zoology, North Carolina State University, Raleigh, NC 27695-7617, U.S.A.*

### SUMMARY

Techniques for estimation of absolute abundance of wildlife populations have received a lot of attention in recent years. The statistical research has been focused on intensive small-scale studies. Recently, however, wildlife biologists have desired to study populations of animals at very large scales for monitoring purposes. Population indices are widely used in these extensive monitoring programs because they are inexpensive compared to estimates of absolute abundance. A crucial underlying assumption is that the population index ( $C$ ) is directly proportional to the population density ( $D$ ). The proportionality constant,  $\beta$ , is simply the probability of 'detection' for animals in the survey. As spatial and temporal comparisons of indices are crucial, it is necessary to also assume that the probability of detection is constant over space and time. Biologists intuitively recognize this when they design rigid protocols for the studies where the indices are collected. Unfortunately, however, in many field studies the assumption is clearly invalid. We believe that the estimation of detection probability should be built into the monitoring design through a double sampling approach. A large sample of points provides an abundance index, and a smaller sub-sample of the same points is used to estimate detection probability. There is an important need for statistical research on the design and analysis of these complex studies. Some basic concepts based on actual avian, amphibian, and fish monitoring studies are presented in this article. Copyright © 2002 John Wiley & Sons, Ltd.

**KEY WORDS:** absolute abundance; capture–recapture sampling; detection probability; distance sampling; long-term monitoring; relative abundance; removal sampling; wildlife monitoring studies

### 1. INTRODUCTION

The focus of much early wildlife statistics work was estimation of absolute animal abundance on one (or perhaps a few) small areas. The methods used were very labor intensive and expensive (e.g. capture–recapture). Gradually focus has shifted to estimating population parameters on more areas. (For example, a very important reference is Burnham *et al.* (1987) on comparison of treatments using capture–recapture methods). The focus of much contemporary wildlife research is on even

---

\*Correspondence to: K. H. Pollock, Department of Statistics, Box 8203, North Carolina State University, Raleigh, NC 27695-8203, U.S.A.

larger scales with the objective of monitoring large populations or even communities. The goal of this article is to consider the principles of design for large-scale wildlife monitoring studies that incorporate estimating detection probability. We shall also link the research back to more basic research on estimation of absolute animal abundance.

Two important long-term monitoring studies in the United States are the Mourning Dove Survey (e.g. Dolton, 1996) and the Breeding Bird Survey (Peterjohn *et al.*, 1996). These surveys are based on counts of birds seen or heard from points along secondary roads and they have been in continuous operation for many years. These surveys have two fundamental problems: (i) non-random placement of survey points (only roadside habitat) and (ii) no estimation of probability of detection of the birds.

With the rise of conservation biology and human population growth putting so much pressure on wildlife populations, many new inventory and long-term monitoring studies are currently being designed. For example, we are involved in projects considering amphibian and avian community inventory and monitoring in the southeastern United States, and stream fish community monitoring across the U.S.

In this article we shall emphasize important design issues for large-scale wildlife monitoring programs of species of animals and the estimation of the probability of detection for individuals of those species. We shall not cover the very important topic of how to design inventory programs for animal communities in geographic regions of concern that have not been studied. Clearly these regions will have to be inventoried first before sensible monitoring programs could be designed. We shall also not cover the important topic of monitoring and estimating species richness (and related parameters) of communities using the methods developed by Nichols and co-workers (see, for example, Boulinier *et al.*, 1998; Nichols *et al.*, 1998a, 1998b).

The article begins with a general discussion of monitoring objectives. We proceed to a discussion of the properties of relative abundance indices that form the basis of most long-term monitoring studies. We then review methods of estimating absolute abundance. This is followed by a section on how to design estimation of detection probability into the monitoring study using a double sampling approach. This enables us to convert estimates of relative abundance to absolute abundance. We give a very brief discussion of other design issues and then a final discussion of important concepts covered and some directions for future research.

## 2. MONITORING OBJECTIVES

Specification of objectives is an essential first step in the development of any monitoring program. Why are inferences about variation in abundance sought and for what purpose(s) are these inferences to be used? Exactly how are the data resulting from the monitoring to be used to develop these inferences? Objectives of monitoring programs can be classified generally as either scientific or management. When monitoring programs are viewed as serving scientific purposes, then it is important to specify a priori hypotheses of interest. The a priori hypothesis is the key feature distinguishing science from other methods of learning and obtaining knowledge, and monitoring programs established for scientific purposes should be designed to test or distinguish among competing hypotheses. Certainly, it is possible to use monitoring data to develop hypotheses and to accidentally gain insights into processes not considered during program establishment. However, in times of limited resources, we believe that serious scientific monitoring programs should be established with a priori hypotheses in mind, as such programs are much more likely to be efficient in contributing to our knowledge base.

Many current programs for animal population monitoring have been implemented for the purpose of providing information used in management. In some cases management can involve annual or periodic decisions about actions designed to bring about changes in animal populations. In other cases management may be more passive, involving implementation of management actions only when abundance is perceived to be too low or declining, or perhaps too large or increasing too rapidly.

Monitoring programs designed for use in population management typically serve two specific roles in the management process (e.g. Nichols *et al.*, 1995). One role is to provide periodic assessments of system state at the potential decision points in the management process. The manager would like to know population size at time  $t$  before deciding on the appropriate management action to be applied during the interval  $t$  to  $t + 1$ . The other role of monitoring is to help the manager develop knowledge of the manner in which the system is likely to respond to the various possible management alternatives. This role of monitoring in management is thus very similar to the objectives of scientific monitoring programs. The manager must know how the population is likely respond to different actions. In an uncertain world, this knowledge is typically expressed as hypotheses or models about the dynamics of the managed system. The development of these models is nearly always based on data about past responses of the system to various management actions, and monitoring programs provide such data. Various scientific processes can be used to develop models from historical data, including retrospective analyses, constrained or true manipulative experimentation and adaptive management (e.g. Nichols, 1991). All of these approaches require monitoring data for the population of interest.

The design specifics of a monitoring program will depend very heavily on objectives. It is thus important to specify these objectives a priori and to then tailor the monitoring program to these objectives. The initial step in developing a monitoring program is thus to specify the objectives of the program and the manner in which monitoring data are to be used to meet those objectives.

In the absence of specific objectives, it is difficult to be very specific about estimation methods likely to be useful in monitoring. However, most monitoring programs face two important sources of variation that must be dealt with in program design: spatial variation and detectability (e.g. Thompson, 1992; Lancia *et al.*, 1994). Spatial variation arises in the typical situation where the investigator(s) cannot apply survey or monitoring techniques over the entire area to which inference is to be drawn. In such situations, sample areas must be selected from the entire area of interest and monitoring effort applied to these sample areas. Results from these sampled areas are then used to draw inference about the entire area of interest. Detection probability refers to the near-universal situation in animal population monitoring in which survey methods do not detect all animals present in the sampled area. Even on the areas that are selected as part of the spatial sample, we cannot enumerate all animals present. Monitoring programs must thus incorporate methods for estimating or removing effects of detectability, so that estimated changes in animal abundance, or a related quantity, reflect true changes in the quantity of interest and not differences in detectability. In the sections that follow, we consider the kinds of count statistics used in monitoring programs and discuss the issues of detectability and spatial variation.

### 3. ASSUMPTIONS OF RELATIVE ABUNDANCE INDICES

The development in this section is extended from work in Lancia *et al.* (1994). Counts used in long-term monitoring studies can be viewed as random variables with expectation given by

$$E(n_{ijk}) = \beta_{ijk} N_{ijk} \quad (1)$$

where:

$n_{ijk}$  = count of animals detected at a point  $(i, j, k)$  in space and time

$\beta_{ijk}$  = probability of 'detection' (seen or caught) of a counted animal at a point  $(i, j, k)$  in space and time

$N_{ijk}$  = population size at a point  $(i, j, k)$  in space and time.

It is convenient to re-express the above equation in terms of local densities,

$$E(C_{ijk}) = \beta_{ijk} D_{ijk} \quad (2)$$

where:

$C_{ijk}$  = count per unit area of animals detected at a point  $(i, j, k)$  in space and time

$D_{ijk}$  = population density at a point  $(i, j, k)$  in space and time.

Thus we have two spatial and temporal processes—one for detection probability and one for population density. The assumptions for the use of such counts as indices are that there is a linear relationship between the count and the population size and that the detectability is constant over space and time, or at least over the dimension used in comparisons. Consider two different points (1,1,1) and (2,2,2) and approximate the expected value of the ratio of the counts as:

$$E(C_{111}/C_{222}) \approx \beta_{111} D_{111}/\beta_{222} D_{222}$$

so that

$$E(C_{111}/C_{222}) \approx D_{111}/D_{222}$$

Therefore the count ratio estimates the ratio of the local population densities if, and only if,  $\beta_{111} = \beta_{222}$ . Trends in relative abundance over more than two time points are based on regression models, but they still require the same assumption of equal detectability over time points.

Another way of stating the problem facing us is that changes in the detection probability and changes in the population size are confounded (i.e. not separable) unless detectability is constant (or at least has constant expectation). We believe that detection probability is rarely constant in practice. In our salamander monitoring program temporal comparisons based on counts on the same plot are likely to be problematic due to soil moisture changes affecting the detectability of salamanders. In dry weather fewer salamanders are near the soil surface. In our bird monitoring studies spatial comparisons based on counts in different habitats are likely to be problematic because the detectability of birds often varies by habitat. In very dense vegetation both sight and sound detection of birds is curtailed.

There are basically three approaches used to try and deal with failure of the detectability assumptions:

1. When designing the monitoring study, use a detailed protocol to try and keep detectability constant over space and time.
2. Measure covariates that may influence detectability (but not the true density) and use a model adjustment to make detectability constant over space and time.
3. Acknowledge that detectability is not constant over space and time and that not all covariates can be measured, modeled or even perceived, and therefore estimate detection probability spatially and temporally as part of the monitoring design.

We believe that, while 1 and 2 should be used to reduce differences in detectability over time and space, 3 is the only scientifically rigorous approach and the rest of our manuscript is predicated on that assumption.

The fundamental equations used by almost all estimates of absolute abundance and density are

$$\hat{N}_{ijk} = n_{ijk} / \hat{\beta}_{ijk}$$

or

$$\hat{D}_{ijk} = C_{ijk} / \hat{\beta}_{ijk}$$

(3)

They are based on equations (1) and (2) presented earlier. We would like to estimate absolute abundance ( $N_{ijk}$ ) and the probability of detection ( $\beta_{ijk}$ ) at a particular point in space and time. There are many methods of doing this, including capture–recapture methods, removal methods, and distance methods, which we review in the next section.

#### 4. ABSOLUTE ABUNDANCE ESTIMATION: A BRIEF REVIEW

##### 4.1. Capture–recapture models

Let us begin with the simplest capture–recapture model, the Lincoln–Petersen model (Seber, 1982, p. 59). As an example, consider that, as part of a large monitoring study on amphibians in a national park, we have selected one point in space where we want to estimate detectability of salamanders, using a mark and recapture method. We denote:

$n_1$  as the number of salamanders captured and marked in sample 1;  
 $m_2$  as the number of marked salamanders recaptured in sample 2; and  
 $(n_2 - m_2)$  as the number of unmarked salamanders captured in sample 2.

Assuming equal catchability of animals in each sample, a closed population, and no mark loss, we have:

$$\hat{N} = n_1 n_2 / m_2$$

$$\hat{\beta}_1 = m_2 / n_2$$

$$\hat{N} = n_1 / \hat{\beta}_1$$

where  $N$  is the population size and  $\beta_1$  is the probability of detection at time 1. (The equation for estimating population size can be converted to density by dividing by the area of the plot). The purpose of the second sample can be viewed as to enable the estimation of detectability at time 1. To obtain precise estimates one needs to have high detection probabilities (Seber, 1982, p. 64).

Pollock (1991, 2000) and Buckland *et al.* (2000) provide a brief review of more general capture–recapture models with many key references. They first consider closed models and then open models. All of these models can be used to adjust count statistics by their detection probabilities but the details are not shown in this brief review.

The equal catchability assumption may be violated in two ways: (i) heterogeneity which occurs when different animals have inherently different capture probabilities; and (ii) trap response when the probability of capture depends on the animals' prior capture history. Often one samples the population more than twice and, each time, every unmarked animal caught is uniquely marked, previously marked animals have their captures recorded, and usually all animals are released back into the population. This more extensive sampling enables sophisticated modeling that permits unequal catchability due to heterogeneity and trap response. Otis *et al.* (1978) consider a set of eight closed models where capture probabilities vary due to time ( $t$ ), heterogeneity ( $h$ ), and trap response ( $b$ ) in all possible combinations ( $M_0$ ,  $M_b$ ,  $M_h$ ,  $M_{bh}$ ,  $M_t$ ,  $M_{tb}$ ,  $M_{th}$  and  $M_{tbh}$ ). They also provided a computer program, CAPTURE, to compute the estimates and select among models.

The heterogeneity models that use a distribution of capture probabilities have caused theoretical difficulty for statisticians. An early *ad hoc* approach was based on the 'jackknife' method (Burnham and Overton, 1978) and proved quite useful. Later, Chao and her colleagues in a series of papers used a method based on sample coverage (Lee and Chao, 1994). Others have suggested use of log linear models (Cormack, 1989). Maximum likelihood estimation where the heterogeneity is modeled as a finite mixture distribution (usually with two or three support points) (Norris and Pollock, 1996; Pledger, 2000) is a recent development. Another approach to modeling heterogeneity uses covariates (Huggins, 1989; Alho, 1990). The original approach to model selection for this series of models in Otis *et al.* (1978) does not work well. We suspect that the finite mixture approaches will aid in solving this problem (Pledger, 2000).

Capture–recapture studies often last a long time, rendering the assumptions of closed models invalid. Therefore there has been a need for the development of models that allow for additions (births and immigrants) and deletions (deaths and emigrants) to the population. The first general open model was developed independently by Jolly (1965) and Seber (1965). Their model, which requires equal catchability and equal survival rates of all animals at each sampling period, enables one to estimate population sizes, survival rates, and birth numbers for almost all samples. Detailed treatments are in Seber (1982) and Pollock *et al.* (1990). Recently there has been an emphasis on integration of recruitment in the likelihood (Schwarz and Arnason, 1996; Pradel, 1996). Since the original Jolly–Seber papers there has been much research on modeling survival rates allowing for multiple strata (e.g. sex, age, location). An important reference is Lebreton *et al.* (1992). An interesting development has been the shift to fitting large numbers of models, which has necessitated development of model selection criteria based on the AIC (Akaike Information Criteria). The alternative approach of averaging over some reasonable models is also being considered. The recent book by Burnham and Anderson (1998) on model selection is very important.

Another focus of research has been to develop combinations of different sampling methods. One example uses a 'robust' design that combines both open and closed models in one analysis (Pollock, 1982). Many papers have used this design for many reasons, including to allow for unequal catchability, to separate recruitment from immigration, and to estimate temporary emigration (Pollock *et al.*, 1990; Kendall *et al.*, 1995, 1997). Temporary emigration is likely to be important in estimating detectability of salamanders in long-term monitoring studies (Bailey, 2001). An extremely useful program for capture–recapture modeling is MARK (White and Burnham, 1999).

#### 4.2. Removal models

Sometimes it is possible to remove animals permanently (harvested species, pest species) or temporarily (fish and salamanders can be removed and later returned to the population). Here the

animals are not marked. We begin by considering the simplest case of a two-sample removal model (Seber, 1982, p. 318). As an example, consider a study to monitor freshwater fish in streams. We want to estimate detection probability at one point in space using an electro-fishing removal study. We denote by  $n_1$  and  $n_2$  the fish captured and removed in samples 1 and 2, respectively.

We assume a closed population and equal detection probability of fish at both time points and obtain the following estimates:

$$\begin{aligned}\hat{N} &= n_1^2 / (n_1 - n_2) \\ \hat{\beta}_1 &= (n_1 - n_2) / n_1 \\ \hat{N} &= n_1 / \hat{\beta}_1\end{aligned}$$

where  $N$  is the population size and  $\beta_1$  is the probability of detection at time 1. Now the assumption of equal catchability is stronger than in the capture–recapture model because we also require equal capture probabilities for the two samples. The equations for population size in the Lincoln–Petersen capture–recapture model and in the two-sample removal model here have a similar structure of a count divided by an estimate of a detection probability.

Removal models for more than two samples can be fit using the CAPTURE program using Model  $M_b$  and Model  $M_{bh}$  (Otis *et al.*, 1978). Model  $M_{bh}$  is likely to be particularly important as heterogeneity of capture probabilities is almost always present. If effort is unequal in the different samples, then catch-effort models can be used (Seber, 1982, p. 296). These are often used in fisheries problems.

#### 4.3. Distance methods

Another estimation technique is based on distance methods. They have the advantage of providing an estimate of detection probability without catching animals for mark and recapture or removal. They can only be used if the animals to be monitored have high detectability. For large mammals, *line transect methods* are frequently used. For birds, typically *point counts* (and sometimes *line transects*) are used. If the distance from the animal to the point is measured, then the theory of variable circular plots (or line transects) can be used to estimate the probability of detection (Buckland *et al.*, 1993). The key feature of the method is the fitting of a detection function  $g(y)$ , where  $y$  is the distance from the point, using general non-parametric fitting methods. The crucial assumptions are:

- (i) detection at the point is one ( $g(0) = 1$ );
- (ii) no movement before detection;
- (iii) no measurement error; and
- (iv) independent sightings.

For line transect sampling the estimate of population size can again be expressed as a count divided by detection probability:

$$\hat{N} = n / \hat{\beta}$$

and

$$\hat{\beta} = 2L / (\hat{f}(0)A)$$

where  $L$  is the transect length,  $A$  is the area size and  $\hat{f}(0)$  is the estimate of the probability density for the detected distances evaluated at 0.

These are extremely useful methods, and we refer the interested reader to the large literature for more detail. The book by Buckland *et al.* (1993) is a good starting point. There is also a software package, DISTANCE. Recently there has been a lot of interest in combining distance and capture–recapture methods (Alpizar and Pollock, 1996; Borchers *et al.*, 1998) to enable the assumption (i) that  $g(0) = 1$  to be relaxed. Movement before detection of birds is also an important assumption violation that affects the validity of estimates of detection. Measurement error (assumption iii) can be reduced by providing good training for the observers carrying out the surveys.

#### 4.4. Other avian methods

Some biologists have been reluctant to measure distance when using point counts for terrestrial forest birds. Therefore, there has been recent interest in finding practical alternatives that allow estimation of detection probability. Nichols *et al.* (2000) developed a method for use with avian point counts with primary and secondary observers. Our research group is also considering use of the Lincoln–Petersen model with two independent observers. Both methods require the assumption of no heterogeneity of detection probabilities between animals. There may also be matching errors that are assumed small.

Farnsworth *et al.* (2001) investigated a method where the time of detection of individual birds is recorded. We realized that the removal method described in Section 3.2 could be used here. The assumptions are that: birds do not move into or out of the area during the count; detection rate is constant during the count period; and individual birds can be distinguished.

We are currently developing comparisons between these methods (distance, multiple observers and time depletion) and also considering how method combinations might provide more robust estimators.

#### 4.5. Potential methods for different taxa

Methods of estimating detectability (catchability, sightability, observability) that are useful depend on the taxa being sampled as part of long-term monitoring studies. For example, small mammals typically need to be caught, and typically capture–recapture and removal sampling using live or kill traps are used. For large visible mammals aerial surveys and the line transect method are often used to estimate detection probability.

In our study on salamanders in the Great Smoky Mountains (Bailey, 2001), we are evaluating the feasibility of capture–recapture and removal methods. We are using the robust capture–recapture design because we were particularly concerned with vertical temporary emigration. For some frog species call counts may be used with perhaps multiple observers to estimate detectability.

In stream monitoring for fish, we believe multiple pass removal sampling based on electro-fishing will be useful (Seber, 1982, p. 318). In marine systems catch-effort and change-in-ratio methods could be used to estimate catchability of fish (Gould and Pollock, 1997; Udevitz and Pollock, 1992).

For terrestrial birds, point counts are the preferred method of monitoring because large areas can be covered relatively inexpensively. Birds seen or heard in a fixed or variable radius plot around the observer are counted and the counts used as an index of abundance. The possible methods of estimating detection probability are distance methods, multiple observers and time depletion. We are currently comparing these methods. Occasionally mist nets are used when monitoring birds on relatively small areas using mark and recapture methods (e.g. DeSante *et al.*, 1995).



For all taxa the methods used need to be reasonably inexpensive because the monitoring is large scale and detection probability will have to be estimated across both space and time.

## 5. MONITORING DESIGN APPROACHES

Sections 3 and 4 have concerned issues of detectability at specific points in space and time. However, most monitoring programs will involve sampling at multiple points in space and time. In this section, we consider conceptual plans that deal with detectability over multiple sample locations. This section, thus considers spatial variation in detectability. The problem of allocating sample locations over space, in a manner permitting inference about a larger area of interest, is discussed briefly in Section 6.

### 5.1. *The double sampling concept*

We present some elementary concepts on how to design a monitoring study in a manner that accounts for potential variation in detection probabilities. Our choices may be broken down into three conceptual approaches:

*Plan 1.* Count data are collected at all sample points in space and time. No detection probabilities are estimated.

*Plan 2.* Count data are collected at all sample points in space and time. Detection probabilities are estimated at some subset of the sample points in space and time.

*Plan 3.* Count data are collected at all points in space and time. Detection probabilities are estimated at all the sample points in space and time.

Although Plan 1 is commonly used, it is not scientifically defensible. Plan 3 may be possible in some cases (i.e. perhaps with point counts of birds) but there may be extreme practical difficulties in estimating detection probabilities at every space–time point. For example, in our salamander study we had about 100 spatial points where count indices were taken, but only 15 spatial points, where the capture–recapture study could be run to estimate detection probabilities, due to the high cost. Therefore we suspect that Plan 2 will be the most sensible for most species. A double sampling approach will be used and there is the need to devise optimal or near optimal designs. What fraction of the resources should be allocated to the large count index sample and what fraction to the smaller detection probability sample, so that the variance in population trends is minimized for a fixed cost? In addition, a decision should be made as to how the more expensive detection sample points should be placed in both space and time. These are very complex questions that require substantial research for particular taxa and sampling methods. For example, results for fish counts will likely be quite different from results for salamander plot capture counts.

### 5.2. *Allocation of relative sampling effort: double sampling*

It is beyond the scope of this publication to consider complex spatial models of detectability. For simplicity we assume that we have a simple random sample without replacement of study plots in  $L$  habitat strata monitored over  $K$  times. We assume that the population of plots is so large that the finite population correction can be ignored in variance calculations. Each point will have a count (index) collected and potentially also a more detailed study to estimate detection probability carried out. Not all detection probabilities will be estimable under Plan 2 because of the cost, so we shall have to develop a simple model of the process to make any progress.

Within a year we assume that detectability varies between strata but not within strata, so that we can represent detectability as  $\beta_1, \beta_2, \dots, \beta_L$  for the  $L$  different strata in a particular year. (Actually we can assume more generally that the  $\beta$ s within a stratum  $h$  are random from a distribution with mean  $\beta_h$ ). Now we concentrate on the estimation of the density ( $D_h$ ) and detectability ( $\beta_h$ ) for a particular stratum  $h$ . We assume there is a random sample of  $n'$  plots where the count is taken and a sub-sample of  $n$  plots where the count is taken but also the detectability is estimated. The stratum density estimate is given by

$$\hat{D}_h = \bar{C}_h / \hat{\beta}_h$$

where

$$\bar{C}_h = \sum_{j=1}^{n'} C_{hj} / n'$$

is the count mean over the whole sample of  $n'$  plots and

$$\hat{\beta}_h = \sum_{j=1}^n \hat{\beta}_{hj} / n$$

is the mean detectability estimate over the sub-sample of  $n$  plots. An approximate large sample variance of the stratum density estimate (using a standard Taylor series approach) is given by

$$\text{Var}(\hat{D}_h) = (D_h^2) [(\tau_c^2)/n' + (\tau_\beta^2)/n]$$

Note that the  $\tau$ s are the coefficients of variation for the count data and the detectability estimates, with both expressed on a plot basis. These coefficients of variation will include both measurement variability and variability caused by the stochastic nature of the true density and detectability parameters within a stratum. The stratum densities could be used to make comparisons among strata within a year.

An important question is how to allocate the sampling effort between collecting data on the count index (the  $n'$  sample) and collecting the more detailed data to do the detectability estimation (the  $n$  sub-sample). This topic requires more research but one simple approach is to minimize the variance of the density estimate subject to a total fixed cost as follows. We assume the cost function

$$C = c_c n' + c_\beta n$$

where  $c_c$  and  $c_\beta$  are the costs per sampling unit for the count data and the detectability data, respectively. The ratio of  $n$  to  $n'$  is given by

$$(n/n') = (\tau_\beta^2 c_c / \tau_c^2 c_\beta)^{1/2}$$

and this can be re-expressed as

$$(n/n') = (\delta/\gamma^2)^{1/2}$$

Note that  $\delta = c_c/c_\beta$ , the relative cost ratio, and  $\gamma = \tau_c/\tau_\beta$ , the relative coefficient of variation ratio. These equations can be derived using similar approaches to those used by Thompson (1992) for other double sampling problems.

### 5.3. Sampling allocation examples

We do not know of any field studies that have used this approach but we hope this will change in the future. We now present two oversimplified examples based on our knowledge of salamander monitoring (using counts plus an intense capture–recapture sampling method to estimate detection probability) and stream fish monitoring (using counts plus removal sampling to estimate detection probability). These examples are presented purely for illustration.

For our salamander monitoring study we found that the relative cost ratio was quite small because the capture–recapture study was quite expensive, so for illustration let us use  $\delta = 1/16$ , which is likely to be very close to the reality. We found the count data quite variable, so let us use  $\tau_c = 1$  and  $\tau_\beta = 0.25$  for  $\gamma = 4$ . This means that  $(n/n') = 1/16$ , so that if we monitored 96 plots for counts we would do the detailed detectability estimation on only six of them.

In contrast, for a stream fish monitoring study using two pass removal sampling to estimate the detectability, the relative cost ratio would be much larger with, say,  $\delta = 1/2$  close to the reality. This is simply based on the argument that two removal passes would cost about twice the cost of one removal pass. The count data are again likely to be quite variable compared to the individual detectability estimates, so let us use  $\gamma = 2$  (although I do not have direct field evidence for this value). For this illustration  $(n/n') = (1/8)^{1/2} = 0.35$ , so that if we monitored 100 streams for the one pass count we would take 35 streams for the additional passes to carry out the detailed detectability estimation. If one decided to use three or four removal passes the relative costs would be 1/3 or 1/4 and these values would have the advantage of giving more precise estimates of detectability. For any of these cases the sampling fraction devoted to estimating detectability would be much higher than for the salamander example, because the relative cost ratio is so much larger.

While these results are purely for illustration, the general principles will apply to future real study designs. To our knowledge, no one has suggested this double sampling approach before and it seems to fill an important gap in the literature. We note that there is also the question about how to allocate the resources when there are multiple years of study. The simplest and most conservative approach is to use the equation above in each stratum and each year separately, but this assumes that detectability is different in different years. If the detectability did not change very much over the years one would be able to apply fewer resources to estimating detectability. We defer this question to a future paper. In addition we emphasize that other sampling designs (unequal probability sampling, adaptive sampling) would have different variance structures for the estimate of density.

## 6. OTHER DESIGN PRINCIPLES

Other design principles are relevant in addition to those discussed above. Variation in detection probability between habitat strata and between years was discussed in the previous section. This

discussion was predicated on a reasonable selection of sample locations that permit reasonable inference to the entire area of interest (the spatial variation noted at the end of Section 2). This topic is the subject of books on sampling (e.g. Thompson, 1992) and will not be elaborated here. We note that such approaches as simple random sampling, stratified sampling, cluster sampling, systematic sampling, dual-frame sampling, general unequal probability sampling, and adaptive sampling are well-developed and may be useful in different monitoring situations (see Thompson, 1992; Haines and Pollock, 1998). An interesting consideration regarding use of these approaches to sampling space in animal monitoring programs is the need to consider both the underlying variable of interest (e.g. animal abundance) *and* detection probability. For example, it is common to stratify an area of interest by habitat, or some other predictor of animal abundance, as a means of reducing variances of resulting estimates. However, because of the importance of estimating detection probability, predicted variation in this quantity (which may also be a function of habitat) should be considered as a basis for stratification as well.

Random selection of sampling units at one level or another (e.g. within strata) is an important element of most of the above-listed approaches to sampling space. We believe that the importance of randomization is not appreciated in many animal monitoring programs. As noted in the Introduction, the Mourning Dove and Breeding Bird surveys conducted each year in North America are roadside surveys. As such they provide information about bird populations found in the portion of North America along roads, but do not permit inference about areas distant from roads. Such surveys would benefit greatly from the creation of off-road strata. We recognize that such strata would likely require different survey methods (e.g. travel between point counts would be on foot; thus fewer points could be visited in a morning), and would likely not be sampled nearly as intensively (fewer routes), at least initially, as the on-road strata. Nevertheless, addition of off-road strata would permit estimation of abundance for entire areas of interest rather than for roadside subsets of such areas.

Development of adequate sampling designs requires many additional considerations. For example, the size of spatial sampling units, the number of units selected for survey efforts, and the placement and size of the plots on which counts are actually conducted within selected units must be determined. Temporal considerations are important as well, including the duration of the sampling within each major time period (e.g. 20 days during June of each year) and the temporal spacing of samples (e.g. annually). Consideration should also be given to the time horizon for the entire program, although we realize that some programs have been designed with the thought of continuation into the indefinite future. Finally, the joint consideration of spatial and temporal allocation of samples is also important. If the program objective is to provide the best estimate of abundance over a large area of interest each year, then random selection of sampling units might occur each year. If estimation of trend (changes in abundance over time) is of primary importance, then the re-sampling of some fraction of sampling units each year will likely be recommended.

## 7. DISCUSSION

We believe that the design of a monitoring program for natural animal populations requires substantial thought about program objectives (why?) and implementation (how?) relative to those objectives. Objectives can usually be classified as either scientific or management, and clear specification of objectives facilitates development of an efficient program design. Monitoring program designs must incorporate two primary sources of variation in animal counts: spatial variation and detectability. Space should be sampled in a manner that permits inference about the entire area of interest, based on

the selected sampling units. In addition, monitoring programs should include the collection of data, permitting estimation of detection probability for the selected count statistics.

The above recommendations are extremely general. We provide them because we are surprised at how few animal monitoring programs appear to have been developed with these simple principles in mind. However, we fully recognize that the development of a good monitoring program involves many detailed considerations that go far beyond these general suggestions.

These detailed considerations will depend on the specifics of the objectives, areas and taxa of interest. We shall briefly discuss some of these now. One very important issue is optimal allocation in the double sampling design between the large scale monitoring, using the index of abundance, and the detectability estimation on a small sub-sample of plots. Another issue is balancing cost versus complexity in choice of methods to estimate detection probability. The more robust methods are complex and hence cost more per plot than simple methods. (For example, it would be much cheaper to use the Lincoln–Petersen method to estimate detectability than to use a more general closed capture–recapture model, but the more general closed model would allow unequal catchability of animals due to time, heterogeneity, and trap response.) Modeling the detection process spatially and temporally is another important issue that has not been explored much in research to this point in time. Count methods of monitoring population trends and estimating detectability are usually much cheaper than capture and marking methods for many taxa (especially birds). We believe, however, that there is a need for comparative studies as marking studies do provide a lot more information, albeit at a higher cost. In fact, the theme of trade-off between cost and complexity and robustness of estimators is a recurring theme in many of our comments and will certainly arise in all future designs of monitoring studies.

The reason for our emphasis on the double sampling approach is that many methods of estimating detectability are very expensive compared to traditional count indices (i.e. capture–recapture and removal methods). We do emphasize, however, that variable circular plots and line transects use distance to estimate detectability. Then the distance data can be collected routinely at the same time as the count data and at little additional cost. In those cases detection probability would be estimated everywhere and there would be no need for the double sampling design.

The choice of which species to monitor has also not been considered. Costs dictate that certain ‘indicator’ species from a community will need to be chosen for detailed monitoring. This is a topic of considerable discussion by conservation biologists. There is a trade-off between species that are easier to monitor versus those that would be more important biologically.

We have emphasized important design issues for large-scale wildlife monitoring programs of species of animals and the estimation of the probability of detection for individuals of those species. We have not covered the very important topic of how to design inventory programs for animal communities in geographic regions of concern that have not been studied. Clearly these regions will have to be inventoried first before sensible monitoring programs could be designed. Many biologists emphasize compiling species lists using purposive sampling for such studies. However, we hope that some idea of spatial distribution and abundance could also be obtained in these studies. Therefore statistical sampling, perhaps using stratified systematic random sampling or adaptive sampling, should be useful (Thompson, 1992).

We also have not covered the important topic of monitoring and estimating species richness (i.e. the number of species in a community) and related parameters of communities using the methods developed by Nichols and co-workers (see, for example, Boulmier *et al.*, 1998; Nichols *et al.*, 1998a, 1998b). However, we note that our own involvement in large-scale monitoring efforts for amphibians and stream fish have convinced us that estimation of species richness both spatially and temporally is a

very important component of monitoring animal communities. Many biologists tend to ignore the fact that species are missed in both inventory and monitoring studies. The estimation of the number of species missed and its precision is very important.

#### ACKNOWLEDGEMENTS

This article owes substantially to various colleagues with whom we are working on a variety of monitoring projects. They are Susan Shriner and Julie McIntyre from North Carolina State University and Dr Mike Meador of the Water Resources Division of the U.S. Geological Survey. Funding came from the Biological Resources Division of the U.S. Geological Survey and the U.S. Environmental Protection Agency.

#### REFERENCES

- Alho JM. 1990. Logistic regression in capture–recapture models. *Biometrics* **46**: 623–635.
- Alpizar-Jara R, Pollock KH. 1996. A combination line transect and capture–recapture sampling model for multiple observers in aerial surveys. *Journal of Environmental and Ecological Statistics* **3**: 311–327.
- Bailey LL. 2001. PhD. Dissertation, North Carolina State University, in preparation.
- Borchers DL, Zucchini W, Fewster RM. 1998. Mark–recapture models for line transect surveys. *Biometrics* **54**: 1207–1220.
- Boulinier T, Nichols JD, Sauer JR, Hines JE, Pollock KH. 1998. Estimating species richness: the importance of heterogeneity in species detectability. *Ecology* **79**: 1018–1028.
- Buckland ST, Anderson DR, Burnham KP, Laake JL. 1993. *Distance Sampling: Estimating Abundance of Biological Populations*. Chapman & Hall: London.
- Buckland ST, Goudie IBI, Borchers DL. 2000. Wildlife population assessment: past developments and future directions. *Biometrics* **56**: 1–12.
- Burnham KP, Anderson DR. 1998. *Model Selection and Inference: A Practical Information Theoretic Approach*. Springer-Verlag: New York.
- Burnham KP, Overton WS. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**: 625–633.
- Burnham KP, Anderson DR, White GC, Brownie C, Pollock KH. 1987. *Design and Analysis Methods for Fish Survival Experiments based on Release–Recapture*. American Fisheries Society Monographs 5, Bethesda, Maryland. (ISSN 0362-1715).
- Cormack RM. 1989. Loglinear models for capture–recapture. *Biometrics* **45**: 395–413.
- DeSante DF, Burton KM, Saracco JF, Walker BL. 1995. Productivity indices and survival rate estimates from MAPS, a continent-wide programme of constant-effort mist-netting in North America. *Journal of Applied Statistics* **22**: 935–947.
- Dolton DD. 1996. *Mourning Dove Breeding Population Status 1995*. U.S. Fish and Wildlife Service, Washington, DC.
- Farnsworth G, Pollock KH, Nichols JD, Simons TR, Hines JE, Sauer JR. 2001. A removal model for estimating the detection probability during point counts divided into time intervals. *The Auk*. (In press.)
- Gould WR, Pollock KH. 1997. Catch-effort maximum likelihood estimation of population parameters. *Canadian Journal of Fisheries and Aquatic Science* **54**: 890–897.
- Haines DE, Pollock KH. 1998. Estimating the number of active and successful bald eagle nests: an application of the dual frame method. *Environmental and Ecological Statistics* **5**: 245–256.
- Huggins RM. 1989. On the statistical analysis of capture experiments. *Biometrika* **76**: 133–140.
- Jolly GM. 1965. Explicit estimates from capture–recapture data with both death and immigration—stochastic model. *Biometrika* **52**: 225–247.
- Kendall WL, Pollock KH, Brownie C. 1995. A likelihood-based approach to capture–recapture estimation of demographic parameters under the robust design. *Biometrics* **51**: 293–308.
- Kendall WL, Nichols JD, Hines JE. 1997. Estimating temporary emigration using capture–recapture data with Pollock’s robust design. *Ecology* **78**: 563–578.
- Lancia RA, Nichols JD, Pollock KH. 1994. Estimating the number of animals in wildlife populations. In *Research and Management Techniques for Wildlife and Habitats, 5th edn*, Bookout TA (ed.). The Wildlife Society: Bethesda, MD; 215–253.
- Lebreton J-D, Burnham KP, Clobert J, Anderson DR. 1992. Modelling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs* **62**: 67–118.

- Lee S-M, Chao A. 1994. Estimating population size via sample coverage for closed capture–recapture models. *Biometrics* **50**: 88–97.
- Nichols JD. 1991. Extensive monitoring programs viewed as long-term population studies: the case of North American Waterfowl. *Ibis* **133** Suppl. 1: 89–98.
- Nichols JD, Boulinier T, Hines JE, Pollock KH, Sauer JR. 1998a. Estimating rates of local species extinction, colonization and turnover in animal communities. *Ecological Application* **8**: 1213–1225.
- Nichols JD, Boulinier T, Hines JE, Pollock KH, Sauer JR. 1998b. Inference methods for spatial variation in species richness and community composition when not all species are detected. *Conservation Biology* **12**: 1390–1398.
- Nichols JD, Hines JE, Sauer JR, Fallon FW, Fallon JE, Heglund PJ. 2000. A double-observer approach for estimating detecting probability and abundance from avian point counts. *Auk* **117**: 393–408.
- Nichols JD, Johnson FA, Williams BK. 1995. Managing North American Waterfowl in the face of uncertainty. *Annual Review of Ecology and Systematics* **26**: 177–199.
- Norris JL, Pollock KH. 1996. Nonparametric MLE under two closed capture–recapture models with heterogeneity. *Biometrics* **52**: 639–649.
- Otis DL, Burnham KP, White GC, Anderson DR. 1978. Statistical inference from capture data on closed animal populations. *Wildlife Monograph* **62**: 135.
- Peterjohn BG, Sauer JR, Link WA. 1996. The 1994 and 1995 summary of the North American Breeding Bird Survey. *Bird Populations* **4**: 48–66.
- Pledger S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics* **56**: 434–442.
- Pollock KH. 1982. A capture–recapture design robust to unequal probability capture. *Journal of Wildlife Management* **46**: 752–757.
- Pollock KH. 1991. Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future. *Journal of the American Statistical Association* **86**: 225–238.
- Pollock KH. 2000. Capture–recapture models. *Journal of the American Statistical Association* **95**: 293–296.
- Pollock KH, Nichols JD, Brownie C, Hines JE. 1990. Statistical inference for capture–recapture experiments. *Wildlife Society Monographs* **107**: 1–97.
- Pradel R. 1996. Utilization of capture–mark–recapture for the study of recruitment and population growth rate. *Biometrics* **52**: 703–709.
- Schwarz CJ, Arnason AN. 1996. A general methodology for the analysis of capture–recapture experiments in open populations. *Biometrics* **52**: 860–873.
- Seber GAF. 1965. A note on the multiple recapture census. *Biometrika* **52**: 249–259.
- Seber GAF. 1982. *The Estimation of Animal Abundance and Related Parameters* (2nd edn). Charles W. Griffin: London.
- Thompson SK. 1992. *Sampling*. Wiley: New York.
- Udevitz MS, Pollock KH. 1992. Change-in-ratio methods for estimating population size. In *Wildlife 2001: Populations*, McCullough DR, Reginald HB (eds). Elsevier Applied Science: New York; 90–101.
- White GC, Burnham KP. 1999. Survival estimation from populations of marked animals. *Bird Study* **46**: 120–138.