

ESTIMATING PATCH OCCUPANCY WHEN PATCHES ARE INCOMPLETELY SURVEYED

Leonard A. STEFANSKI, Matt RUBINO and George HESS *

Abstract - Institute of Statistics Mimeo Series #2543

In this paper we describe a method for statistical modeling of the probability of patch occupancy by a species as a function of patch size and other covariates when patches may be incompletely surveyed. The model is appropriate for study designs in which each patch: is visited once; is surveyed at a set of access sites from which an observer can detect the species within a measurable portion of the patch; and is surveyed until the species is detected or until all access sites have been visited. The model explicitly accounts for incomplete patch coverage, providing unbiased estimates of patch occupancy relative to the “naive” modeling approach that ignores incomplete patch coverage. The statistical model is motivated by and illustrated with data from a study of Barred owl habitat.

KEY WORDS: Barred owl habitat; biodiversity monitoring; detection probability; logistic regression; metapopulation dynamics; negative binomial distribution; patch occupancy; Poisson distribution; statistical modeling.

Institute of Statistics Mimeo Series No. 2543

April 2003

*Leonard A. Stefanski is Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695–8203. Matt Rubino is Biologist / GIS Analyst, North Carolina Gap Analysis Project, North Carolina State University, Raleigh, NC 27695-7617. George Hess is Associate Professor, Department of Forestry, North Carolina State University, Raleigh, NC 27695–8002. Email addresses: stefanski@stat.ncsu.edu, matt_rubino@ncsu.edu, george_hess@ncsu.edu.

1 Introduction

Population estimates have long been based on counts of individual organisms. However, the probability of detecting (counting) an individual organism, given that it is present, is almost always less than one (e.g., Otis et al. 1978; Lancia et al. 1994). Uncorrected counts of individuals therefore almost always underestimate the true size of a population. Researchers have developed a variety of statistical modeling approaches to estimate population size when the detection probability for individuals is less than one (e.g., Seber 1982; Nichols et al. 2000; Williams et al. 2002; Thompson, 1992, Ch 16). These models, with varying levels of sophistication, arrive at population estimates by adjusting the counts (number of detections) to account for the probability of detecting individual organisms. In the simplest forms, such as capture-recapture methods, the count is divided by a single detection probability to provide a corrected population estimate (see Seber 1982; Lancia et al. 1994). More complex models use continuous detectability functions that estimate detectability as a function of distance from the observer (e.g., Buckland et al. 1980; Lancia et al. 1994; Buckland et al. 2001; Williams et al. 2002).

Analogously, the probability of detecting a species in a habitat patch, given that at least one individual of the species is present in the patch, is usually less than one. Thus, the proportion of patches in which a species is detected underestimates the true proportion of occupied patches. Researchers have begun to address the issue of estimating patch occupancy when detection probability is less than one. For example, McKenzie et al. (2002) developed a method to estimate patch occupancy that uses a sampling scheme in which patches are visited repeatedly during the season in which the species is most detectable. Barbraud et al. (2003) developed a method to estimate patch occupancy for bird colonies that uses a sampling scheme in which patches are visited from multiple times each year over a number of years.

In this paper, we describe a method for estimating patch occupancy by a species when:

- each patch is visited only once;
- each patch is surveyed at a predefined set of access sites from which an observer can detect

the species within a known portion of the patch;

- each patch is surveyed until the species is detected or until all access sites have been visited;
- patches may be surveyed incompletely, even if all access sites have been visited.

Under these conditions, observers might not detect the species in a patch for one of three reasons:

1. the species was not present in the patch;
2. the species was present in unsurveyed portions of the patch, and therefore was not detected;
3. the species was present in the portion of the patch surveyed, but was not detected during the survey (we assume that this does not happen, i.e., that the species will be detected, if present in the surveyed portion).

Our estimates of patch occupancy use information about the portions of the patch surveyed at each access site in addition to the detection status information.

We developed this method as part of an ongoing project to create a wildlife conservation plan for an approximately one million hectare suburbanizing region in the United States (Hess et al. 2000; Hess and King 2001; Rubino and Hess 2003). The ability to quickly develop and test habitat models for a number of species is central to our approach. The habitat models we have been developing identify habitat patches predicted to be suitable for a species. Our objective in model verification is to uncover major flaws quickly and inexpensively using a single survey and simple presence/absence approaches. We are interested in identifying errors of commission (predicting a species is present when it is not) and omission (not predicting a species is present when it is) in our habitat model. Patch occupancy of habitat and non-habitat patches is a measure of these errors.

Our method will be of use to those interested in metapopulation dynamics and biodiversity monitoring programs. Patch occupancy is a key measure in both of these endeavors, in which patches are almost always incompletely surveyed and species detectability is almost always less than one. The portion of patches occupied is a state variable in many metapopulation models (e.g.,

Levins 1969; Hanski and Gilpin 1997). Change in the occupancy status of patches through time can be used to estimate metapopulation extinction and recolonization rates (e.g., Barbraud et al. 2003, Mackenzie et al. 2003). In biodiversity monitoring programs, researchers often survey a portion of an area and use that information to make inferences about a larger area (e.g., Yoccoz et al. 2001). Clearly, getting the best estimate for patch occupancy is an important goal.

Patch occupancy has been used to define minimum patch sizes for wildlife habitat. For example, Robbins et al. (1989) used logistic models of forest patch occupancy for 26 area-sensitive bird species in the Middle Atlantic United States. They reported the patch size at which the probability of species occurrence was at a maximum, as well as the patch size for which there was at least a 50% probability of finding a species. These patch sizes have been used widely to define minimum habitat requirements for the 26 area-sensitive bird species documented. The method we present can improve estimates of species presence upon which such recommendations are based.

2 Survey Protocol

In developing our statistical model, we made the following assumptions:

- Surveys for a species are conducted on a sample of discrete habitat patches of a range of sizes.
- Some of the patches surveyed are non-habitat and not expected to contain the species. These are included to estimate errors of omission in the habitat model.
- Because of limited access to the patches, most patches are surveyed incompletely.
- Each patch is surveyed from a set of access sites from which the species can be detected. Access sites are locations within or adjacent to the patch.
- It is possible to measure the total area of the patch.
- The areas of the portions of a patch surveyed from the access sites, and also the areas of overlap surveyed from multiple sites can be measured.

- Access sites may be visited in any order, and the order in which they are visited is recorded.
- The patch survey is considered complete once the species is detected, or after all access sites have been visited.

We carried out such a survey to test a barred owl (*Strix varia*) habitat model (Rubino and Hess 2003). We visited a sample of patches delineated by the habitat model. To investigate errors of omission in the habitat model, we also visited a number of non-habitat patches predicted by our habitat model to be unsuitable for barred owls. Most patches were on private lands limiting accessibility. Therefore access sites were established along public roads surrounding the patches (Figure 1). We used a vocalization call-back technique in which surveyors broadcast a recording of owl calls in order to elicit responses from territorial individuals (Rubino and Hess 2003). Because barred owl vocalizations are audible for approximately 800 meters (Smith 1978; Bosakowski et al. 1987), access sites were approximately 1.6 km apart. This spacing was designed to minimize overlapping sampling effort and maximize patch area coverage per unit effort.

A patch was surveyed until a response was heard, or until all access sites had been visited. Once a barred owl was detected, sampling was discontinued and the patch was designated as positive for owl presence. If there was no response at any of the access sites, the patch was scored negative for barred owl presence. Listening coverage for a patch was almost always incomplete, especially for larger patches, (Figure 2). For an incompletely surveyed patch, a negative score does not imply that the patch was empty, only that no owls were present in the surveyed portion.

Rubino and Hess (2002) modeled the relationship between patch size and owl detection using logistic regression based on data from 95 surveyed patches. This “naive” approach does not use information about detection probability within patches to estimate the true proportion of patches occupied. The methods presented here use information about the portion of each patch surveyed to improve estimates of patch occupancy. Coverage data were available for 94 of the 95 patches Hess and Rubino surveyed. Because our methods require coverage data, we used only the 94 patches with coverage information to develop our model. To facilitate comparison, we reproduced Rubino and Hess’ logistic regression model using data on 94 of the patches (Figure 2, solid curve).

A global positioning system was used to obtain the coordinates of each access site during the survey. We downloaded these coordinates into a geographic information system and quantified the proportion of the survey patch measured from each access site. We accomplished this by overlaying the coordinates on the habitat map and assuming an 800m detection radius for each access site. The cumulative proportions of the patch surveyed at each access site, and the final status of the patch (owl detected or not), served as the foundation for our statistical model of patch occupancy.

3 Methods: Statistical Modeling and Analysis

3.1 Single Patch Outcome Probabilities

The patch observation process is determined by a sequence of access sites and the portions of patch covered from each site. Observation of a patch continues until species detection or all access sites have been visited. For a given patch, let $C_1 < C_2 < \dots < C_m$ denote the cumulative coverage proportions afforded by the m access sites in the order that the sites are visited. For example, if 2% of the patch is covered from the first access site, 3% from the second not covered by the first, and 6% from the third not covered from the first or second, then $C_1 = 0.02$, $C_2 = 0.05$, and $C_3 = 0.11$.

The observation of a patch results in one of $m + 1$ possible outcomes, $AP_1, \dots, AP_m, AP_m^{(0)}$.

- AP_1 : The species is detected at the first access site and the site provided a coverage of C_1 . This outcome is summarized by $(1, C_1)$, where 1 indicates detection and C_1 indicates the proportion of patch surveyed.
- AP_2 : The species is detected at the second access site which provided additional coverage of $C_2 - C_1$. The outcome is summarized by the vector $(0, 1, C_1, C_2)$. The first two elements $(0, 1)$ indicate no detection at the first site, detection at the second site. The second two elements (C_1, C_2) give the corresponding cumulative proportions of the patch surveyed.
- AP_j : The species is first detected at the j th access site which provided a coverage proportion of $C_j - C_{j-1}$ not covered by the previous $j - 1$ sites. The outcome is summarized by the

vector $(0, \dots, 0, 1, C_1, \dots, C_j)$. The $j - 1$ leading zeros indicate the failure to detect the species at the first $j - 1$ sites; (C_1, \dots, C_j) indicate the corresponding cumulative proportions of the patch covered.

If all access sites are visited, there are two possible outcomes at the final site.

- AP_m : The species is first detected at the m th access site, summarized $(0, \dots, 0, 1, C_1, \dots, C_m)$ where the 1 appears in the m th component.
- $AP_m^{(0)}$: The species is not detected at any of the m sites, summarized $(0, \dots, 0, C_1, \dots, C_m)$.

Under these assumptions, we derived the probability of each outcome. First, we determined conditional probabilities given that there are k individuals of the species in the patch.

Unconditional probabilities are obtained by summing over an assumed distribution for the number of individuals per patch. We refer to the successive new portions of patch surveyed at each access site as subpatches. The mathematical assumptions used to develop our model are stated in this section. Their reasonableness, and the likely consequences of violations, are discussed in detail in Section 5

When the species does not occupy the patch ($k = 0$), the only possible outcome (assuming the probability of a false detection is 0) is $AP_m^{(0)}$. Thus $\Pr(AP_j) = 0$, $j = 1, \dots, m$, and $\Pr(AP_m^{(0)}) = 1$.

Probabilities when $k > 0$ are derived under the assumptions that individuals are uniformly and independently distributed throughout the patch, and when present in a surveyed area, individuals are detected with probability one. Event AP_1 occurs if and only if there is at least one individual in the first subpatch (with proportion C_1) surveyed. This is calculated as the complement of the probability that there are no individuals in the first subpatch, and under our assumptions is $\Pr(AP_1) = 1 - (1 - C_1)^k$. Here, $1 - C_1$ is the proportion of the patch not surveyed (i.e., all but the first subpatch). Raising this the k th power gives the probability that all k individuals in the patch are in the unsurveyed portion (thus $(1 - C_1)^k$ is the probability of no individuals in the first subpatch); and subtraction from 1 results in the probability that not all individuals are in the

unsurveyed portion, which is equivalent to having at least one individual in the surveyed portion (the first subpatch).

Event AP_2 occurs if and only if there are no individuals in the first subpatch and there is at least one in the second subpatch. The probability $\Pr(AP_2)$ can be factored as the conditional probability of at least one individual in the second subpatch given no individuals in the first subpatch, and the probability of no individuals in the first subpatch. The probability of no individuals in the first subpatch was determined previously and is $(1 - C_1)^k$.

The conditional probability of at least one individual in the second subpatch given no individuals in the first subpatch is calculated using the same reasoning that was used to calculate $\Pr(AP_1)$. The area surveyed from the first access site is “removed” from consideration and the remaining area is regarded as a “new” patch, with the understanding that proportions must be normalized by the factor $1 - C_1$. The conditional probability of no individuals in the second subpatch given no individuals in the first subpatch is thus $\{(1 - C_2)/(1 - C_1)\}^k$, and hence the conditional probability of at least one individual in the second subpatch given none in the first subpatch is $1 - \{(1 - C_2)/(1 - C_1)\}^k$. Multiplying by the probability of no individuals in the first subpatch, $(1 - C_1)^k$, results in $\Pr(AP_2) = (1 - C_1)^k - (1 - C_2)^k$. Similar reasoning shows that $\Pr(AP_j) = (1 - C_{j-1})^k - (1 - C_j)^k$ for $j = 1, \dots, m$. Finally, the probability of no individuals in the total surveyed area is calculated as the probability that all k individuals are in the unsurveyed area, and thus $\Pr(AP_m^{(0)}) = (1 - C_m)^k$. The probabilities for all cases can be written

$$\Pr(AP_j) = (1 - C_{j-1})^k - (1 - C_j)^k, \quad j = 1, \dots, m, \quad \Pr(AP_m^{(0)}) = (1 - C_m)^k, \quad k \geq 0, \quad (1)$$

where by definition $C_0 = 0$.

Assuming that the distribution of the number of individuals, K , in a patch is given by a parametric mass function $p(k; \lambda) = \Pr(K = k; \lambda)$ depending on the parameter λ , unconditional probabilities are obtained as

$$\Pr(AP_j) = \sum_{k=0}^{\infty} \{(1 - C_{j-1})^k - (1 - C_j)^k\} p(k; \lambda), \quad j = 1, \dots, m$$

$$\Pr(\text{AP}_m^{(0)}) = \sum_{k=0}^{\infty} (1 - C_m)^k p(k; \lambda).$$

These probabilities can be written in terms of the probability generating function of K ,

$$\mathcal{P}(t, \lambda) = E(t^K), \quad 0 \leq t \leq 1,$$

$$\Pr(\text{AP}_j) = \mathcal{P}(1 - C_{j-1}, \lambda) - \mathcal{P}(1 - C_j, \lambda), \quad j = 1, \dots, m,$$

$$\Pr(\text{AP}_m^{(0)}) = \mathcal{P}(1 - C_m, \lambda).$$

Note that the probabilities depend only on whether an individual was observed ($Y = 1$) or not ($Y = 0$), and on the last two cumulative proportions, denoted (C_{L-1}, C_L) , leading to the concise representation for the probability of the outcome (C_{L-1}, C_L, Y) ,

$$\begin{aligned} \Pr(C_{L-1}, C_L, Y) &= \{\mathcal{P}(1 - C_{L-1}, \lambda) - \mathcal{P}(1 - C_L, \lambda)\}^Y \{\mathcal{P}(1 - C_L, \lambda)\}^{1-Y} \\ &= \mathcal{P}(1 - C_L, \lambda) \left\{ \frac{\mathcal{P}(1 - C_{L-1}, \lambda)}{\mathcal{P}(1 - C_L, \lambda)} - 1 \right\}^Y \end{aligned}$$

for $Y = 0, 1$, and $0 \leq C_{L-1} < C_L \leq 1$.

Different choices for the distribution of K (the number of individuals per patch) result in different versions of the models. We continue development of the model in parallel for the cases in which K is Poisson(λ) and Negative Binomial(r, λ). The Poisson distribution is the most commonly used model for count data and thus it is a natural choice; the Negative Binomial is included for comparison and to illustrate the generality of the modeling approach. For the Poisson distribution, $\mathcal{P}(t, \lambda) = \exp\{-\lambda(1 - t)\}$ and

$$\Pr(C_{L-1}, C_L, Y) = \exp(-\lambda C_L) \{\exp(\lambda(C_L - C_{L-1})) - 1\}^Y.$$

For the Negative Binomial distribution, $\mathcal{P}(t, \lambda) = \{(1 - \lambda)/(1 - \lambda t)\}^r$ and

$$\Pr(C_{L-1}, C_L, Y) = \left\{ \frac{1 - \lambda}{1 - \lambda(1 - C_L)} \right\}^r \left\{ \left[\frac{1 - \lambda(1 - C_L)}{1 - \lambda(1 - C_{L-1})} \right]^r - 1 \right\}^Y.$$

3.2 Incorporating Predictor Variables

The number of individuals in a patch will depend on patch characteristics and such dependencies are incorporated through the parameter λ . We consider patch size, although the modeling extends readily to other predictor variables. For example, if the mean number of individuals is thought to be proportional to patch size, then when $K \sim \text{Poisson}(\lambda)$, $\lambda = \beta X$ is appropriate, where β is an unknown regression parameter and X is patch size. More generally, $\lambda = \lambda(X, \boldsymbol{\beta})$, some function of patch size X and parameter vector $\boldsymbol{\beta}$.

Logistic regression is a common model for relating patch occupancy to habitat characteristics such as patch size (Rubino and Hess, 2003; Scott et al., 2002). In our modeling framework, patch occupancy corresponds to $K > 0$, and thus the probability of patch occupancy is $\Pr(K > 0) = 1 - \Pr(K = 0) = 1 - \mathcal{P}(0, \lambda(X, \boldsymbol{\beta}))$. It is possible to choose $\lambda(X, \boldsymbol{\beta})$ so that the induced model for patch occupancy is logistic. This facilitates comparison of logistic models fit under the assumption of perfect detection (sometimes called “naive” models) to our models that explicitly account for imperfect detection due to incomplete coverage. For the model with K distributed Poisson,

$$\lambda(X, \boldsymbol{\beta}) = \ln \{1 + \exp(\beta_0 + \beta_1 X)\} \quad (2)$$

induces a logistic model for patch occupancy, $\Pr(K \geq 1) = F(\beta_0 + \beta_1 X)$ where F is the logistic distribution function, $F(t) = 1 / \{1 + \exp(-t)\}$. For the model with K distributed Negative Binomial,

$$\lambda(X, \boldsymbol{\beta}) = 1 - [1 / \{1 + \exp(\beta_0 + \beta_1 X)\}]^{1/r} \quad (3)$$

induces a logistic model for patch occupancy.

3.3 Likelihood Methods

When a sample of n patches are studied resulting in data $\{(X_i, C_{L-1,i}, C_{L,i}, Y_i), \quad i = 1, \dots, n\}$, the log-likelihood for $\boldsymbol{\beta}$ is $\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathcal{L}_i(\boldsymbol{\beta})$ where, with $\lambda_i = \lambda(X_i, \boldsymbol{\beta})$,

$$\mathcal{L}_i(\boldsymbol{\beta}) = Y_i \ln \{\mathcal{P}(1 - C_{L-1,i}, \lambda_i) - \mathcal{P}(1 - C_{L,i}, \lambda_i)\} - (1 - Y_i) \ln \{\mathcal{P}(1 - C_{L,i}, \lambda_i)\}.$$

The maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ is found by numerical maximization of $\mathcal{L}(\boldsymbol{\beta})$. Inference is generally based on the large-sample, normal approximation to the distribution of $\hat{\boldsymbol{\beta}}$. The asymptotic-distribution variance of $\hat{\boldsymbol{\beta}}$ can be estimated by either the inverse sample information matrix $\hat{V}_{\text{INFO}} = n^{-1}\hat{A}^{-1}$ or the so-called sandwich variance matrix $\hat{V}_{\text{SAND}} = n^{-1}\hat{A}^{-1}\hat{B}(\hat{A}^{-1})^T$, where

$$\hat{B} = \frac{1}{n} \sum_{i=1}^n \dot{\mathcal{L}}_i(\hat{\boldsymbol{\beta}})\dot{\mathcal{L}}_i(\hat{\boldsymbol{\beta}})^T, \quad \hat{A} = -\frac{1}{n} \sum_{i=1}^n \ddot{\mathcal{L}}_i(\hat{\boldsymbol{\beta}}), \quad (4)$$

$\dot{\mathcal{L}}_i(\boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta}) \mathcal{L}_i(\boldsymbol{\beta})$, and $\ddot{\mathcal{L}}_i(\boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta}^T) \dot{\mathcal{L}}_i(\boldsymbol{\beta})$. The variance estimate \hat{V}_{INFO} is preferred when the assumed model is correct, whereas the alternative estimate \hat{V}_{SAND} is robust to model departures (Stefanski and Boos, 2002).

3.4 Computation and Software

Fitting the proposed models to data involves standard statistical computations. We have written programs in GAUSS (Aptech Systems) for doing the necessary calculations. These programs were used to compute the estimates reported in the next section. The GAUSS code, and the used in the following section, are available from the first author upon request.

4 Results: Statistical Modeling and Analysis

We illustrate our statistical models using data from the Barred owl study of Rubino and Hess (2003). They presented a logistic regression model of *observed patch occupancy*, i.e., their analysis did not account for incomplete patch coverage. Observed patch occupancy can differ from *true patch occupancy*. Our models are designed to provide improved estimates of true patch occupancy by explicitly accounting for incomplete coverage. We compare the standard logistic model (ignoring incomplete coverage) to models from Section 3. In our models we use $X = \ln(\text{patch size} + 1)$ to facilitate comparison to the results presented by Rubino and Hess (2003).

Examination of the data revealed some large patches with no observed occupancy ($Y = 0$) and coverages much less than 100%. It is not unreasonable to expect that had these large patches been surveyed completely, owls would have been detected in some or all of them. It follows that the

effect of these points on the fitted naive logistic regression model is to bias downward the estimated probability of patch occupancy for large patch sizes, which tend to have a smaller proportion of their areas surveyed. Consequently, the logistic model of observed patch occupancy is biased low for large patch sizes. The models in Section 3 explicitly account for incomplete coverage and should result in estimates of patch occupancy parameters that are not biased due to incomplete coverage.

We estimated the occupancy curve as function of $\ln(\text{patch size} + 1)$ based on several choices for the distribution of the number of individuals per patch, K (Table 1). The naive model is standard logistic regression of observed patch occupancy, Y , on $X = \ln(\text{patch size} + 1)$ and coincides with the results reported by Rubino and Hess (2003), with the previously noted exception that we used data on only ninety-four patches. The Poisson and Negative Binomial ($r = 1, 2, 4$) are the models described in Section 3 with $\lambda(X, \boldsymbol{\beta})$ chosen so that the induced models for patch occupancy are logistic. The adjusted-coverage models have intercepts ($\hat{\beta}_0$) that are less than the intercept of the logistic model, and slopes ($\hat{\beta}_1$) that are greater than that of the logistic model. These differences are consistent with the expectation that the coverage-adjusted models have higher estimated occupancy probabilities for large patches than does the naive logistic model. This fact is illustrated in Figure 2, which displays the estimated patch occupancy models for the naive, Poisson and Negative Binomial ($r = 1$) models. All of these models have two parameters so that $\text{AIC} = -2(\text{Log-likelihood}) + 2$. The Negative Binomial models fit the barred owl data marginally better than the Poisson model.

For additional comparison we fit a standard logistic model to data that were modified to represent the data that might have been obtained with more complete patch coverage. There are five large patches (patch size $> 934\text{ha}$) in the data set with no owl observed ($Y = 0$) and low coverage proportions (< 0.56). Under the supposition that these patches were truly occupied by owls, and that $Y = 0$ only because of the low coverage values, we set $Y = 1$ for these five patches and refit the standard logistic regression model to the modified data. The estimates from this model appear in Table 1 under the heading of a modified logistic model, and the corresponding occupancy curve is shown in Figure 2. The similarity of the modified-data estimates to the coverage-adjusted model

estimates is noteworthy.

5 Discussion

We have presented a statistical model to estimate patch occupancy when patches are incompletely surveyed. Our motivation for developing this method was to improve our ability to validate a habitat model being used for conservation planning. Compared to the naive occupancy model (i.e., proportion of patches in which we detected barred owls) the incomplete coverage model indicates that our habitat model has fewer errors of omission at larger patch sizes.

For metapopulation modeling and environmental monitoring applications, our incomplete coverage model provides a better estimate of patch occupancy than simply counting the number of patches in which a species is detected. In metapopulation modeling, the proportion of occupied sites is a state variable; in fact, it describes completely the state of the system in the simplest metapopulation models. Naive approaches that do not account for detection probability less than one underestimate the proportion of occupied sites. Similarly, monitoring programs require inferences about areas that cannot be completely surveyed. Naive approaches would again result in underestimates of the occurrence of species.

Where patch occupancy has been used to define minimum patch sizes for wildlife habitat (e.g., Robbins et al. 1989), the naive approach misrepresents the minimum patch size. For the barred owl, and for all but the lowest occupancy probabilities, the naive approach overestimates the patch size required for a given probability of occupancy (Figure 2). This has important implications for conservation reserve planning initiatives — patches previously considered too small to serve as conservation habitat might be adequate. If one were to choose a minimum patch size for barred owl habitat based on our data, the incomplete coverage models suggest smaller patch sizes than does the naive model. For example, for the patch size at which there is a 50% chance of finding a barred owl (Robbins et al. 1989) our incomplete coverage model (Neg Bin ($r = 1$)) results in a smaller estimated patch size ($\approx 164\text{ha}$) than does the naive model ($\approx 288\text{ha}$).

Our model relies on a number of assumptions, all of which are likely to be violated to some extent in applications. The assumptions are of three types: (1) assumptions on the dispersion of individuals in a patch; (2) assumptions on the subpatch detection method; and (3) statistical modeling assumptions. We review the key assumptions and give a qualitative assessment of the effects of violations. Future work on the statistical model will study the impact of model violations quantitatively, and describe extensions of the basic model that address certain model violations.

The key assumptions on the species under study is that individuals are independently and uniformly distributed throughout the patch. This assumption is not necessarily violated for species whose members occur in pairs, packs, flocks, or other family or social units. In such cases the model applies with the understanding that “individual” means a family or social unit. For example, for the barred owl data a detection is regarded as a detection of a nesting pair.

The uniformity assumption is likely to be reasonable when habitat quality does not vary excessively throughout a patch. For the barred owl study, patches were identified in such a way that ensured a fair degree of habitat uniformity throughout a patch.

The appropriateness of the independence assumption depends on whether individuals tend to aggregate or be self-avoiding. Self-avoiding species will, on average, be more evenly dispersed than would be expected under the independence assumption. This means that, relative to the case of independence, not as much area would have to be surveyed (on average) before an individual is encountered. Individuals will be more likely to be detected at one of the first few access sites relative to the case when individuals are independently dispersed, and observed population density will likely be higher than what would be predicted by a model that assumes independence. This suggests that estimates of occupancy probabilities derived from a model that assumes independence might be biased high. The opposite (occupancy probabilities biased low) will likely be true of gregarious species. With regard to the barred owl study, mating pairs compete for a limited resource (habitat patches) within the larger landscape matrix of the study area (Research Triangle) and are therefore territorial. As a result, there is a tendency for pairs to avoid one another. Thus pairs may be more evenly distributed throughout patches and co-occur less

frequently than would be the case under independence assumption. The impact of this possible violation of the assumptions is not known at this time. Future work is planned to investigate the robustness of our model to violations of the independence assumption.

The assumptions on the subpatch survey detection method are threefold. The first two are that the probability of a false detection is zero, and the probability of detection is one when an individual is in a surveyed subpatch. Violations of the latter are more common and hence more problematic than are violations of the former, and result in underestimation of occupancy probabilities. Our basic model can be modified to handle both types of errors provided the error rates of the detection method are known or can be estimated. The modifications are technically involved and will be presented in a forthcoming paper. The required error rates can be estimated by multiple surveys of the same access sites, or experimentally by repeatedly surveying subpatches that are known to contain individuals (to estimate the probability of detection) or known not to contain individuals (to estimate the probability of false detection). The third assumption is that subpatch coverage areas can be determined with acceptable accuracy. Subpatch areas are accumulated over access sites and divided by total patch area to obtain C_{L-1} and C_L , the patch coverage proportions required in our model. Thus errors in the subpatch areas are propagated to C_{L-1} and C_L . For the barred owl study, coverage areas were determined by GIS based on GPS-determined locations that are accurate to within seven-to-ten meters, resulting in very accurate subpatch coverage area measurements and thus errors in C_{L-1} and C_L are not likely to be problematic. This may not be the case in other applications and future work will investigate the effect of errors in C_{L-1} and C_L .

The key modeling assumption relates to the distribution for the latent variable, K , the number of individuals in a patch. There is very little information in a single data set on which to base the choice. However, the models fit to the barred owl data suggest that the choice is not too critical. The Negative Binomial distribution is heavier tailed and more dispersed than the Poisson distribution, yet the parameter estimates and patch occupancy curves displayed in Table 1 and Figure 2 are similar, indicating a reasonable measure of insensitivity to the model for K .

6 Acknowledgements

We thank Jim Nichols for encouragement and advice in the initial stages of our work, and for comments on the manuscript. The manuscript also benefitted from the thoughtful comments of Cavell Brownie and Ken Pollock.

Literature Cited

- Barbraud, Christophe, James D. Nichols, James E. Hines, And Heinz Hafner 2003. Estimating Rates Of Local Extinction And Colonization In Colonial Species. *Oikos*, in press.
- Burnham, K.P., D.R. Anderson, and J.L. Laake. 1980. Estimation of a density from line transect sampling of biological populations. *Wildlife Monographs* 72.
- S.T. Buckland, D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers and L. Thomas. 2001. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, New York.
- Hanski, I. A. and M. E. Gilpin (editors). 1997. *Metapopulation biology: ecology, genetics, and evolution*. Academic Press, New York, New York, USA.
- Hess, G. R., K. Dixon and M. Woltz. 2000. *State of open space 2000: the status of the Triangle's green infrastructure*. Triangle Land Conservancy, Raleigh, North Carolina, USA.
- Hess, G. R. and T. J. King. 2002. Planning for wildlife in a suburbanizing landscape I: selecting focal species using a Delphi survey approach. *In Press. Landscape and Urban Planning* 58(1): 25-40.
- Lambeck, R. J. 1997. Focal species: a multi-species umbrella for nature conservation. *Conservation Biology* 11(4):849-856.
- Lambeck, R.J. 2002. Focal species and restoration ecology: Respose to Lindenmayer et al. *Conservation Biology*, in press.
- Lancia, R. A., J. D. Nichols, and K. H. Pollock. 1994. Estimating the number of animals in wildlife populations. Pages 215-253 in T. Bookhout, editor. *Research and management techniques for wildlife and habitats*. The Wildlife Society, Bethesda, Maryland, USA.
- Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* 15:237-240.
- Lindenmayer, D.B., A.D. Manning, P.L. Smith, H.P. Possingham, J. Fischer, L. Oliver, and M.A.

- McCarthy. 2002. The focal-species approach and landscape restoration: A critique. *Conservation Biology*, in press.
- MacKenzie, D.I. James D. Nichols, Gideon B. Lachman Sam Droege, J. Andrew Royle and Catherine A. Langtimm. 2003; Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. *Ecology*, 83:2248-2255.
- Nichols, J. D., J. E. Hines, J. R. Sauer, F. W. Fallon, J. E. Fallon, and P. J. Heglund. 2000. A double-observer approach for estimating detection probability and abundance from point counts. *Auk* 117: 393-408.
- 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* 62.
- Robbins, C.R., D.K. Dawson, and B.A. Dowell. 1989. Habitat area requirements of breeding forest birds of the middle Atlantic states. *Wildlife Monographs* 103: 1-34.
- Rubino, M. J. 2001. Identifying barred owl habitat in the North Carolina Piedmont: using GIS in focal species conservation planning. Thesis, North Carolina State University, Raleigh, North Carolina, USA.
- Rubino, M.J. and G.R. Hess. 2003. Planning open spaces for wildlife II: Modeling and verifying focal species habitat. *Landscape and Urban Planning*, in press.
- Scott, J.M., P.J. Heglund, M. L. Morrison, M.G. Raphael, W.A. Wall, and F.B. Samson (editors). 2002. Predicting species occurrences: Issues of scale and accuracy. Island Press, Covello, California, USA.
- Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. Second Edition, MacMillan Press, New York, New York, USA.
- Simberloff, D. 1998. Flagships, umbrellas, and keystones: is single-species management passe in the landscape era? *Biological Conservation* 83(3):247-257.
- Stefanski, L. A. and Boos, D. (2002). The Calculus of M-estimation. *The American Statistician*, 56:29-38.
- Thompson, S. K. 1992. *Sampling*. Wiley, New York, New York, USA.
- Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. *Analysis and Management of Animal Populations*. Academic Press, San Diego, California, USA.

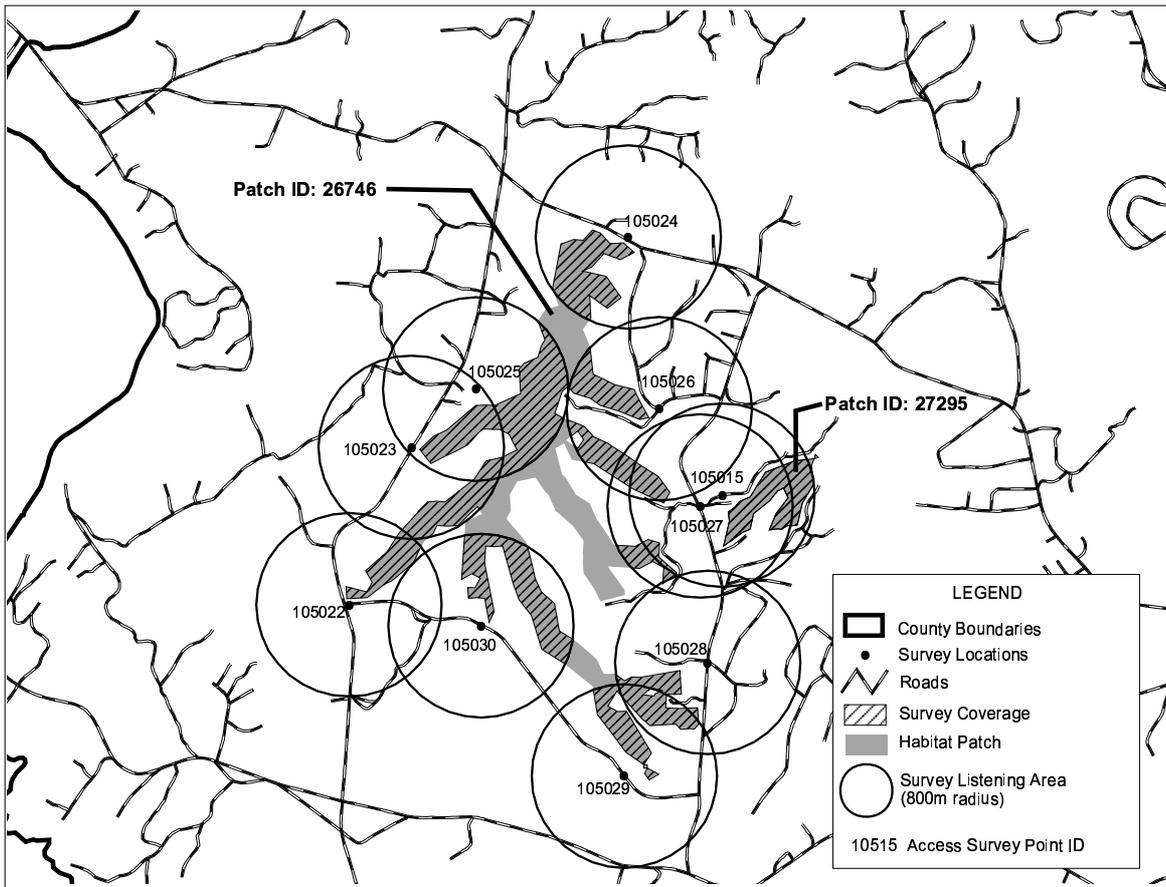


Figure 1: Portion of study area showing Patches Nos. 26746 and 27295 (gray), access sites along public roads, survey listening areas, and subpatch coverages (lined). Access site 105015 provided $\approx 99\%$ coverage of Patch No. 27295; the other access sites combined provided $\approx 75\%$ of Patch No. 26746.

Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$-2(\text{Log-likelihood})$
Logistic (naive)	-3.40 (0.82)	0.60 (0.15)	97.34
Poisson	-4.07 (0.96)	0.84 (0.18)	207.27
Neg Bin ($r = 1$)	-3.83 (0.84)	0.75 (0.14)	198.72
Neg Bin ($r = 2$)	-4.05 (0.92)	0.82 (0.16)	201.61
Neg Bin ($r = 4$)	-4.09 (0.94)	0.83 (0.17)	204.10
Logistic (modified)	-4.13 (0.97)	0.79 (0.17)	87.02

Table 1: Parameter estimates, standard errors (in parentheses) and log-likelihoods for the naive logistic model; several models of the distribution of the number of individuals per patches (K) from Section 3; and the logistic model fit to modified data in which the five largest patches in which no owls were detected were assumed to have owls present (but not detected because of low coverage).

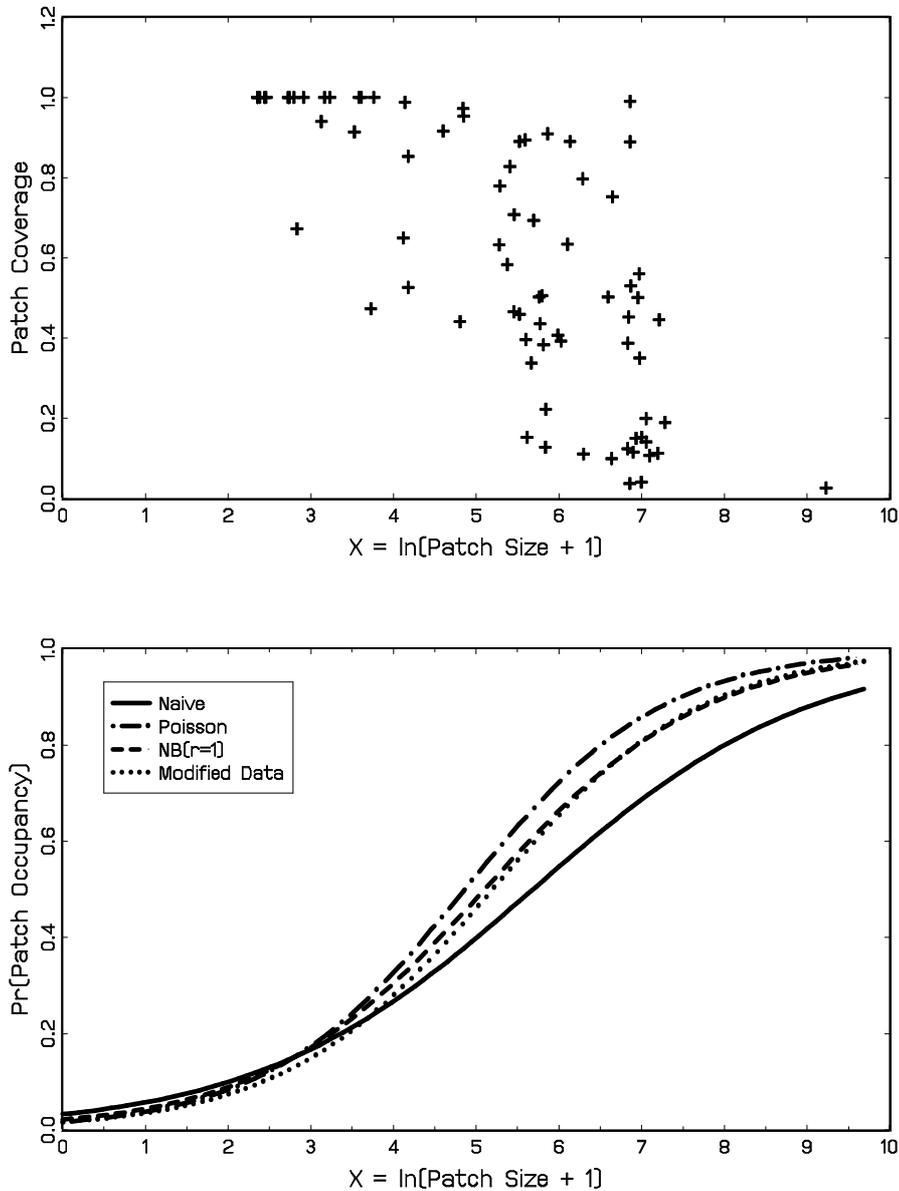


Figure 2: **Top.** Patch coverage fractions by transformed patch size (non-habitat patches excluded). **Bottom.** Estimated probability of patch occupancy. Naive, logistic regression with no adjustment for incomplete coverage; Poisson, coverage-adjusted Poisson model; $NB(r = 1)$, coverage-adjusted Negative Binomial model with $r = 1$; Modified Data, logistic regression with modified data.