

# Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models

Weidong Gu\*, Robert K. Swihart

*Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907-1159, USA*

Received 25 October 2002; received in revised form 16 April 2003; accepted 2 May 2003

## Abstract

Presence–absence data are used widely in analysis of wildlife–habitat relationships. Failure to detect a species' presence in an occupied habitat patch is a common sampling problem when the population size is small, individuals are difficult to sample, or sampling effort is limited. In this paper, the influence of non-detection of occurrence on parameter estimates of logistic regression models of wildlife–habitat relationships was assessed using analytical analysis and simulations. Two patterns of non-detection were investigated: (1) a random distribution of non-detection among occupied patches; and (2) a non-random distribution of non-detection in which the probability of detecting a species in an occupied patch covaried with measurable habitat variables. Our results showed that logistic regression models of wildlife–habitat relationships were sensitive to even low levels of non-detection in occupancy data. Both analytic and simulation studies show that non-detection yields bias in parameter estimation of logistic regression models. More importantly, the direction of bias was affected by the underlying pattern of non-detection and whether the habitat variable was positively or negatively related to occupancy. For a positive habitat coefficient, a random distribution of non-detection yielded negative bias in estimation, whereas linkage of the probability of non-detection to habitat covariates produced positive bias. For a negative habitat coefficient, the pattern was reversed, with a random distribution of non-detection leading to positive bias in estimation. A release–recapture livetrapping study of small mammals in central Indiana, USA, was used to illustrate the magnitude of non-detection in a typical field sampling protocol with varying levels of sampling intensity. Estimates of non-detection error ranged from 0 to 23% for seven species after 5 days of sampling. We suggest that for many sampling situations, relationships between probability of detection and habitat covariates need to be established to correctly interpret results of wildlife–habitat models.

© 2003 Published by Elsevier Ltd.

**Keywords:** Detection probability; Habitat-based model; Logistic regression model; Misclassification; Parameter estimation; Patch occupancy model; Presence–absence; Release–recapture

## 1. Introduction

The understanding of wildlife–habitat relationships is essential in conservation planning. Ideally, conservation plans are formed based on an understanding of population vital rates (i.e., birth, death and dispersal rates), and the influences of environmental variables upon them. However, obtaining precise quantitative information about these processes is costly and not always an option for conservation managers. In the absence of

detailed information about population dynamics, habitat-based models often focus on relationships between habitat variables and either abundance or occurrence of a species. Compared to abundance data, presence–absence data are easier to collect and much less costly. Not surprisingly, substantial efforts have been made to examine the relationship between species occupancy and environmental and habitat covariates (e.g., Hinsley et al., 1995; Mladenoff et al., 1995; Buckland et al., 1996; Bergeron et al., 1997; Wisser et al., 1998; Odom et al., 2001; Scott et al., 2002). The relationship between occupancy and habitat area has been used to justify the preservation of larger tracts of habitats as a primary goal of conservation (Mortberg, 2001). Occupancy-based habitat models also have been used to predict

\* Corresponding author. Present address: Medical Entomology Program, Illinois Natural History Survey, Champaign, IL 61820, USA. Tel.: +1-217-333-1186; fax: +1-217-333-2359.

E-mail address: [wgu@inhs.purdue.edu](mailto:wgu@inhs.purdue.edu) (W. Gu).

habitat selection patterns of unstudied populations and occurrence of species in poorly sampled areas (Pauley et al., 1993, Edwards et al., 1995, Cowley et al., 2000). In addition to these statistical models, presence–absence data have been used in dynamic metapopulation models to draw inferences about ecological processes, e.g., extinction and colonization probability (Hanski 1994, 1998).

Many studies have focused on improvement of parameter estimation of occupancy-based habitat models. For logistic regression models of habitat suitability, spatial autocorrelation in explanatory variables has been incorporated to enhance precision of parameter estimation (Besag, 1974, Augustin et al., 1996, Gu et al., 2001, Klute et al., 2002). Palmgren and Ekholm (1987) and Copas (1988) derived methods for robust parameter estimation when a binary response variable was subject to misclassification error. In their studies, misclassification was assumed to occur randomly among cases. Additionally, falsely recorded occurrences of an absent species and falsely recorded absences of a species that is present were treated equally in their estimation procedures. Moilanen (2002) demonstrated that false zeros in occupancy can significantly affect estimation of all parameters of an incidence function model. In his study, non-detection was assumed to occur randomly among occupied habitat patches. Recently, estimation of the proportion of patch occupancy with the detection probability less than one has been developed. Habitat covariates can be incorporated in estimating the probability of site-specific detection (MacKenzie et al., 2002).

In ecological studies of species occupancy, falsely recorded occurrences of an absent species seldom occur except in cases of identification or transcription errors. Most ecological surveys, however, have inevitable sampling error associated with non-detection of species occurrence because the probability of detecting a species at a site depends on population density and efficiency of sampling approaches and effort. Therefore, underestimation of occupancy could be common for rare species, species that are cryptic due to low per capita probabilities of detection, or in situations characterized by insufficient sampling effort. This is particularly relevant because ecological studies with a conservation focus deal primarily with rare species and often rely on rapid assessment methods or indicator taxa to infer occupancy (e.g., Kerr et al., 2000). Under these circumstances, failure to detect a species that is present might be substantial.

Moreover, the detection probability of occupancy can be influenced by features of the local habitat or surrounding landscape (Nupp and Swihart, 1996; Mancke and Gavin, 2000; Odell and Knight, 2001). Therefore, non-detection is more likely to happen in habitats characterized by small population size and other features that may interfere with detection of species. Ecological

variables are related to the probability of detecting a species in a habitat patch if they affect (1) the local density of the species, (2) behaviors of individuals that influence detection (e.g., movements), or (3) the efficiency of a sampling method. For example, more densely populated patches exhibit reduced detection probability (Kery, 2002). Structural differences in vegetation among patches, even when patches are comparable in animal density, can lead to differences in mean home-range sizes and thus increased detection probability associated with patches in which vegetative structure leads to more extensive movement and greater detectability of each individual visually or with sampling devices such as traps. Consider patches with comparable population density but varying in natural shelters such as cavities or burrows. Sampling that relies on the use of nest boxes or other structures providing shelter could result in enhanced detection in patches with fewer natural cavities or burrows. Consequently, non-detection will be more likely in habitats with covariates that are associated with low probabilities of detection.

Surprisingly, little is known about the impact of non-detection of species occurrence on parameter estimation of wildlife habitat models. In this paper, we examined the impact of non-detection of species occurrence on parameter estimation of logistic regression models of wildlife–habitat relationships that use presence–absence data as the response variable. First, a case study of release–recapture for seven species of small mammals in central Indiana, USA, was used to illustrate the extent of potential non-detection of occurrence in a typical sampling protocol. Second, we developed an analytical model to examine consequences of random distribution of non-detection for rare species. Finally, simulation models were introduced for more realistic situations in which species occurrence was common or prevalent and the probability of non-detection of species occurrence was linked to habitat covariates. Two mechanisms generating non-detection of species occurrence were investigated. In the first mechanism, non-detection was assumed to occur randomly among occupied habitat patches, whereas in the second case non-detection was assumed to be due to particular features of the habitat, so that the distribution of non-detection was non-random.

## 2. Methods

### 2.1. Estimating probability of non-detection of species occurrence: a case study

As mentioned above, non-detection of species occurrence is unavoidable in field studies and may be substantial. We chose a release–recapture data set on small

mammals to illustrate a method for estimating the magnitude of non-detection error induced by sampling, and the extent to which non-detection can vary from species to species and as a function of sampling effort. Mammals were livetrapped for 5 days in forest patches on the Indian Pine Natural Resources Area and the Davis Purdue Agricultural Center in central Indiana. The study areas were predominantly agricultural, with remnant forest vegetation. During 1992–1996, 42 forest patches were trapped, some multiple times, for a total of 86 occasions. To maintain conceptual and terminological consistency with the simulation portion of our study, we hereafter refer to the 86 occasions as “patches”, i.e. a patch may be re-counted if it was revisited in different years. Details of the sampling protocol are reported elsewhere (Nupp and Swihart, 1998; 2000).

The total number of sampled patches occupied,  $T$ , for a species often is estimated by tallying the number of patches in which the species was detected. We obtained a better estimate of  $T$  by using closed-population release–recapture methods (Otis et al., 1978; White and Burnham, 1999) that estimate the conditional daily probability,  $d$ , of detecting a species in a patch given that the species occurs there. Traditionally, release–recapture studies on closed populations have focused on estimation of abundance (Otis et al., 1978). Recently, these methods have been expanded to parameter estimation of species richness (Boulinier et al., 1998) and turnover rates (Nichols et al., 1998a, b), in a manner analogous to estimation of population abundance and vital rates, respectively. In our study, we recorded whether any individual of species  $k$  was detected in patch  $i$  during day 1, 2, ..., 5 of sampling; failure to detect any individual on a day resulted in an entry of zero for that day. For each species, we used the model selection procedure in CAPTURE (Rexstad and Burnham, 1991; Boulinier et al., 1998) to obtain estimates of  $d$  and  $T$ . Models incorporating temporal variation in  $d$  were not considered, because sampling was not done simultaneously at all sites. To assess the importance of sampling effort, we computed estimates of  $d$  and  $T$  using the entire 5 days of trapping. For all species the heterogeneity model ( $M_h$ ) was selected by CAPTURE and thus was used to estimate  $d$  and  $T$ . The probability of non-detection for  $j$  days of sampling was then estimated for species  $i$  as  $\prod_{m=1}^j (1 - \hat{d}_{mi})$ .

## 2.2. Analytical model of non-detection of species occurrence on parameter estimation of logistic regression models

Typically, logistic regression models of wildlife habitat relationships relate the log odds of probability of occurrence ( $p_i$ ) of patch  $i$  to a linear predictor of  $m$  habitat covariates:

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m \quad (1)$$

where  $\text{logit}(p_i) = \log(p_i / (1-p_i))$ . Parameter estimation is accomplished using the method of maximum likelihood for generalized linear models.

To introduce non-detection error in model analysis, we used the following notations. We defined  $p_i$  as the probability of detecting the focal species in a habitat patch  $i$ ,  $P_i$  as the probability of species occurrence in patch  $i$ , and  $d$  as the conditional probability of detecting the species given its presence. Assuming that a species is never recorded as present at a site when it is not there,  $p_i$  is always less than or equal to  $P_i$ , i.e.  $p_i = dP_i$ . The implicit assumption in almost all current habitat models is that  $p_i$  equates to  $P_i$ , i.e.  $d = 1$ . If  $d < 1$ ,  $p_i$  is a probabilistic measure of species occurrence and obviously subject to sampling variation.

For small  $p_i$ ,  $\text{logit}(p_i)$  can be approximated with  $\log(p_i)$  because  $1-p_i$  approximates 1, i.e.  $\text{logit}(p_i) \approx \log(p_i) \approx \log(dP_i)$ . Therefore, it follows

$$\log(P_i) + \log(d) = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m \quad (2)$$

Clearly,  $\log(d)$  is an offset of the intercept of  $\log(P_i)$  against the linear predictor of habitat covariates. The offset is always negative except for the special case when  $d = 1$ . Thus, when  $d$  is a constant less than unity, parameter estimates of wildlife–habitat models will be biased. The direction of bias depends on the sign of the coefficient,  $\beta_k$ . If  $\beta_k > 0$ , a constant  $d < 1$  tends to underestimate the parameter, whereas an overestimate results if  $\beta_k$  is negative.

The above analysis provides an intuitive interpretation of the effect of non-detection of species occurrence on parameter estimation. Unfortunately, this analytical analysis becomes intractable when the approximation of  $\text{logit}(p)$  with  $\log(p)$  is inappropriate or when  $d$  is site-specific and related to habitat covariates. For these situations, simulations are required to understand the effect of non-detection of species occurrence on parameter estimation. In the following sections, we describe a simulation study which explicitly takes into account non-detection of species occurrence in parameter estimation of logistic wildlife habitat models.

## 2.3. Simulation study

A hypothetical landscape was constructed of 50 habitat patches characterized by two habitat covariates. For each habitat patch, the two habitat covariates (e.g., surrogates for habitat size and some measurable attribute of habitat quality such as percent canopy closure) were specified as log-normal distributed variates (Table 1). The probability ( $P_i$ ) of occupancy in habitat

Table 1  
Coefficients for specifying habitat patches and patch occupancy based on logistic regression models

Parameter	Value		
<i>Habitat covariate</i>			
$X_1 = \exp(v)$	$v \sim \text{normal} (\mu = 0.8, \sigma = 0.8)$		
$X_2 = \exp(v)$	$v \sim \text{normal} (\mu = 0.5, \sigma = 0.5)$		
<i>Logistic regression models</i>			
	$b_0$	$b_1$	$b_2$
Rare ( $P = 0.14$ )	-5	0.8	0.4
Common ( $P = 0.54$ )	-5	2.2	1.1
	-1.5	2.2	-2.1
Abundant ( $P = 0.72$ )	-5	2.6	1.3

patch  $i$  was calculated as a logistic function of the two habitat variables:

$$\text{logit}(P_i) = b_0 + b_1X_1 + b_2X_2, \quad (3)$$

where  $b_j$  ( $j = 0, 1, 2$ ) are coefficients, and  $X_1$  and  $X_2$  are the habitat covariates.

Three habitat specifications were generated by choosing coefficients in Eq. (3) to represent various degrees of patch occupancy. Rare, common, and abundant species were designated at 14, 54 and 72% occupancy, respectively. Proportionality of coefficients was maintained across all three sets, with  $b_2 = b_1/2$  used to distinguish the relative importance of the habitat covariates on  $p_i$  (Table 1). In addition, for the common hypothetical species we tested a set of coefficients with a negative effect of  $X_2$  on occupancy (Table 1). Actual patch occupancy was determined by comparing the calculated probability of occupancy,  $P_i$ , to a uniform [0–1] random number. If  $P_i$  was greater than the generated random number, the patch was assigned as occupied, and vice versa. Because of stochasticity in assigning patch occupancy in the above procedure, habitat patches with small estimated probabilities of occupancy might be assigned as occupied whereas patches with large calculated probability of occupancy might be assigned as unoccupied. The habitat covariates and occupancy data specified by the four sets of coefficients (Table 1) were saved in separate files. Simulations were conducted on each of these data files as described in the following sections.

In the first case of non-detection that we considered, detection probability had no relation to the habitat covariates that were measured, thus providing a stochastic analog to the preceding analytical model. False negative misclassification could occur with an equal probability in any of the occupied patches. In our simulations, this was accomplished by comparing a generated uniform [0–1] random number with a constant representing the level of non-detection error, e.g.

0.05. If the random number was smaller than the constant, the occupied patch was reassigned as unoccupied. This procedure repeated for all occupied patches.

In the second case we considered the situation in which detection probability tends to covary with one or more of the measured habitat variables, because these habitat variables influence a species' probability of detection in a patch (i.e.,  $d$ ). In this case, we modeled detection probability in the  $i$ th patch ( $d_i$ ) using a sigmoid function of the calculated probability of occupancy ( $P_i$ ):

$$d_i = \frac{P_i^2}{P_i^2 + \delta^2} \quad (4)$$

where  $\delta$  is a constant on the interval  $[0, +\infty]$  influencing how  $d_i$  changes with  $P_i$ . A large value of  $\delta$  results in a low detection probability for a patch, whereas  $\delta = 0$  results in  $d_i = 1$  and thus no non-detection happens. Different values of  $\delta$  were used to obtain various degrees of non-detection errors. An occupied patch with  $d_i$  smaller than a uniform [0–1] random number was misclassified as unoccupied. Eq. (4) indicates that probability of detection, as a function of probability of occupancy, is dependent on habitat covariates. In a similar procedure as described for randomly distributed non-detection error, the calculated  $d_i$  was then compared with a generated uniform [0–1] random number to assign on which occupied patches non-detection occurring. For a given level of non-detection error, different occupancy data were generated in each trial run because of stochasticity in assigning non-detection to each occupied habitat patch. We conducted 1000 trials for each given non-detection error. Stochasticity prevented an identical pattern of non-detection error from being produced in each trial using a given value of  $\delta$  in Eq. (4). Thus, levels of non-detection error were categorized into groups comparable to their counterparts generated under the first scenario of randomly assigned non-detection error.

For occupancy data from each trial, parameter estimation was conducted using the maximum likelihood method with Fisher's method of scoring (Collett, 1991). Percent relative bias ( $R_i$ ) of parameter  $i$  was defined as

$$R_i = 100 \left( \frac{\hat{B}_i - \tilde{B}_i}{|\tilde{B}_i|} \right), \text{ where } \hat{B}_i \text{ and } \tilde{B}_i \text{ are estimates of}$$

parameter  $i$  in the presence and absence of non-detection error, respectively. We used  $\tilde{B}_i$  instead of true values of the coefficients  $B_i$ , which were used in generating the original occupancy, because stochasticity in generation of occupancy resulted in various occupancy patterns for the same set of habitat coefficients. Means and 95% confidence limits of  $R_i$  were obtained from 1000 randomizations.



### 3. Results

#### 3.1. Empirical estimation of non-detection error for small mammals

Seven species of small mammals accounted for 7593 captures of 4548 individuals. In descending order of individuals captured, species were *Peromyscus leucopus* (3082), *Tamias striatus* (1442), *Blarina brevicauda* (559), *Sciurus niger* (239), *Sciurus carolinensis* (107), *Glaucomys volans* (79), and *Tamiasciurus hudsonicus* (39). Probabilities of false negative misclassification varied substantially among species (Fig. 1). For all species, a second or third day of sampling produced enhanced detection, some of which were dramatic. After 5 days of sampling, all species except *Tamiasciurus hudsonicus* exhibited probabilities of false negative misclassification <0.05 (Fig. 1). *T. hudsonicus*, the species detected least frequently during the study, retained an elevated probability of non-detection error (0.23) despite additional sampling effort.

#### 3.2. Simulated effects of non-detection error on parameter estimation

For the models in which both habitat covariates ( $X_1$  and  $X_2$ ) were positively correlated with patch occupancy, sampling-induced non-detection error introduced bias into estimators and reduced precision in the corresponding coefficients, i.e.  $b_1$  (Fig. 2) and  $b_2$  (Fig. 3). Absolute levels of bias tended to increase with occupancy. For habitat covariate  $X_1$ , which had twice as much influence on occupancy as  $X_2$ , the relative bias was less than that for the coefficient for variable  $X_2$ .

Interestingly, the direction of bias was opposite for the two patterns of non-detection error. Randomly occur-

ring non-detection among occupied habitat patches generally yielded underestimates of coefficients, i.e. a negative relative bias (Figs. 2a–c and 3a–c). These simulation results were consistent with predictions from our analytical model; a constant level of non-detection error should lead to negative bias for positive coefficients. In contrast, non-random distribution of non-detection error yielded positive bias (Figs. 2d–f and 3d–f).

For the model in which occupancy was positively correlated with  $X_1$  but negatively correlated with  $X_2$ , patterns of bias and precision for  $b_1$  mirrored those depicted in Figs. 2 and 3. However, the directions of bias in  $b_2$  were reverse for the two scenarios of non-detection error (Fig. 4). That is, random non-detection produced positive bias in  $b_2$  (Fig. 4a), whereas linking the probability of non-detection error to the habitat covariates resulted in negatively biased estimates of  $b_2$  (Fig. 4b). The latter result also is compatible with predictions of the analytical model that non-detection of occurrence tends to yield positive bias in habitat covariates which are negatively correlated with  $\logit(p_i)$ .

### 4. Discussion

Failure to detect species when they are present at a site is not uncommon in field surveys. For example, using double-observer surveys for avian point counts, Nichols et al. (2000) showed that a single observer missed varying proportions of the birds actually present. Non-detection of occurrence is most common for species with relatively low detection probability (Dettmers et al., 1999). For rare species with low detection probability, non-detection error can be substantial. For example, in an experiment on monitoring one of the world's rarest tortoises (*Psammobates geometricus*), no observer detected >50% of the tortoise models placed along transects in natural habitat (Gardner et al., 1999). Likewise, the detectability of a threatened and cryptic bird (*Brachyramphus marmoratus*) in individual forest stands was as low as 12% (Stauffer et al., 2002). Even for species that are relatively common (e.g. *T. hudsonicus* in our study), large non-detection error can be non-trivial if sampling effort is limited or a sampling method is used that results in low detectability for the species. Our field study of small mammals in central Indiana showed that non-detection errors were problematic in several species when only 2 or 3 days of sampling were conducted, and non-trivial for *T. hudsonicus* even after 5 days (Fig. 1). In situations where detection probabilities are known beforehand, Stauffer et al. (2002) provide a binomial model that incorporates estimates of  $d$  for purposes of power analysis at patch and regional scales. Their model provides a useful approach to determining the level of sampling intensity in terms of repeated visits to a site and number of sampling sites. Based on prob-

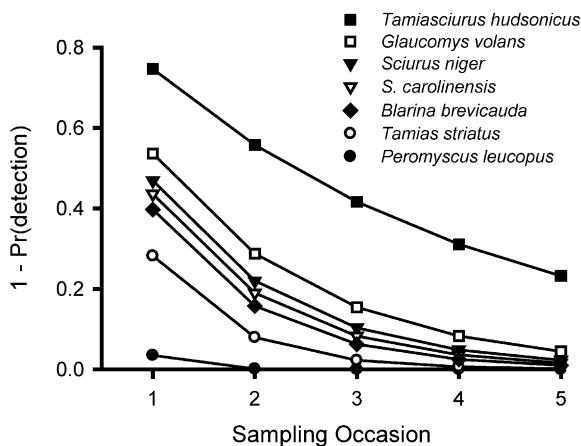


Fig. 1. Estimates of the probability of failing to detect species occupancy of a patch, given that the species occurred in the patch during the sampling period. Estimates were derived from live-trapping data collected on seven species of forest-dwelling small mammals in central Indiana during five-day sampling periods ( $n = 86$ ) from 1992 to 1996.

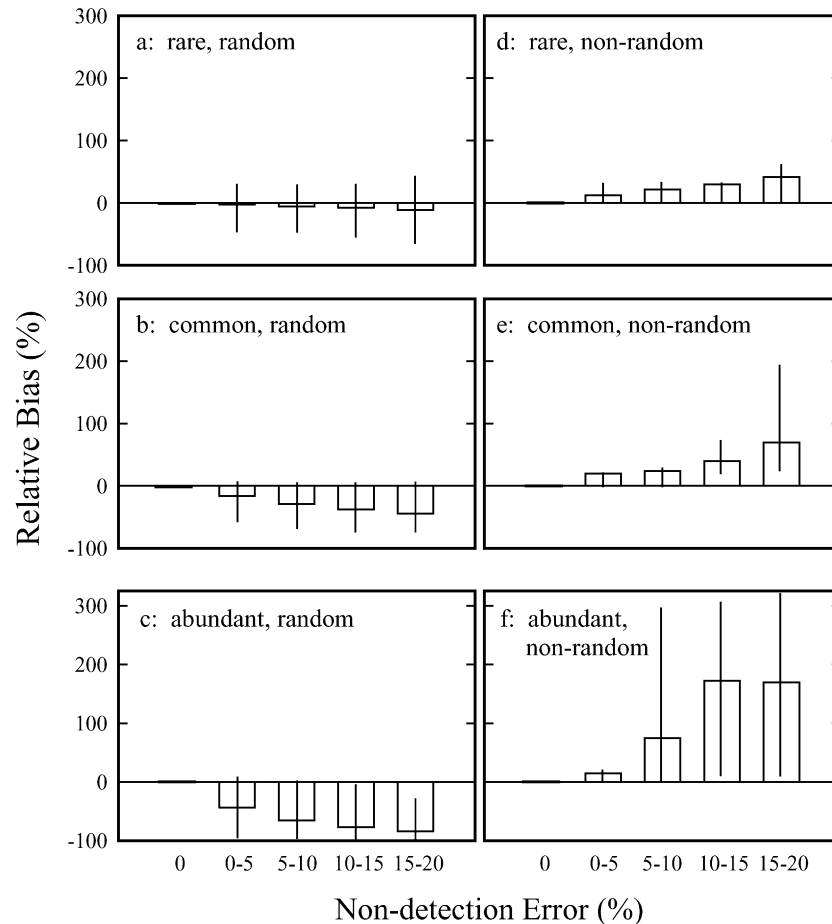


Fig. 2. Effect of non-detection of species occurrence on parameter estimation of habitat covariate  $X_1$  using logistic regression, under various levels of patch occupancy (upper, middle and lower panels represent species occupancy of 14, 54 and 72%, respectively). Panels a–c are parameter estimates in the cases of random non-detection, whereas panels d–f are parameter estimates for the case where the probability of non-detection covaried with  $X_1$  and  $X_2$ . Boxes represent mean relative biases. Vertical bars represent 95% confidence limits from 1000 randomizations.

abilistic models and empirical data, Kery (2002) estimated the minimum number of times that a site must be visited to infer the absence of the particular species of European snakes. The results suggested that some species might be more widespread than previously thought.

The vast majority of wildlife–habitat models that use presence–absence as a response variable have assumed that if a species occurs at a site, it will be detected. This assumption has the effect of equating detection probability to one. As our field study and the examples in the previous paragraph demonstrate, there are many instances in which this assumption of inerrancy is unfounded. Unfortunately, the assumption has seldom been tested in ecological studies, even though resultant inferences about the relationship between species occupancy and habitat variables depend on it. An increased reliance on the use of presence–absence data for decision making in habitat conservation and management highlights the need to evaluate the performance of habitat-occupancy models (Tyre et al., 2001; Cinchy et al., 2002). Uncertainties associated with predictions of occupancy models have been attributed to temporal

variation in animal–landscape relationships, stochastic variation in environmental variables, and assumptions about model structure (Gutzwiller and Barrow, 2001). We extend these findings by demonstrating that failures to detect a species at occupied sites can lead to poorly formulated habitat models and thus to erroneous prediction of a species' response to habitat change.

Our simulation studies demonstrated that bias in parameter estimation is influenced by whether non-detection errors occur independently of characteristics of occupied patches. In situations where non-detection error was randomly distributed among occupied patches, biases in parameters followed the analytical predictions given by Eq. (2), i.e., underestimates of coefficients resulted for habitat variables that were positively related to occupancy. Under this scenario, important habitat variables may be disguised as unimportant solely due to non-detection error. The reduced power of the model is due to loss of information introduced by randomly failing to detect the focal species among occupied patches. Random non-detection error becomes costly if the model containing the errors is

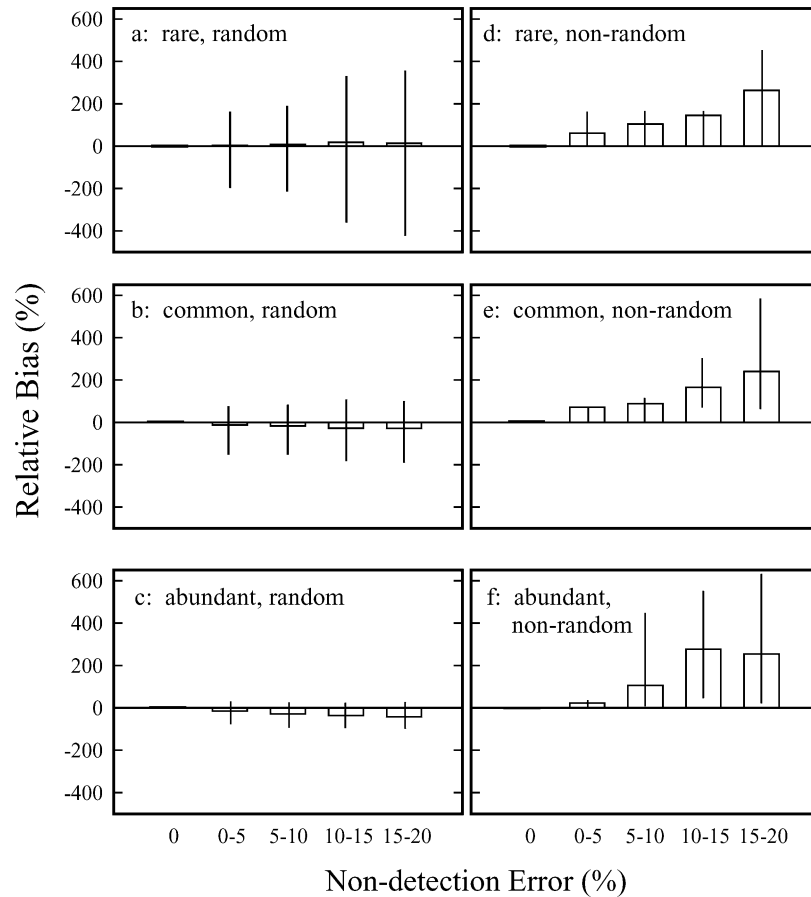


Fig. 3. Effect of non-detection of species occurrence on parameter estimation of habitat covariate  $X_2$  using logistic regression, under various levels of patch occupancy (upper, middle and lower panels represent species occupancy of 14, 54 and 72%, respectively). Panels a–c are parameter estimates in the cases of random non-detection, whereas panels d–f are parameter estimates for the case where the probability of non-detection covaried with  $X_2$ . Boxes represent mean relative biases. The bars represent 95% confidence limits from 1000 randomizations.

poorly formulated and thus does not reveal important habitat variables, leading to “good” habitats being excluded from management consideration.

We also considered the case in which detectability of a species at a site is not random, but instead is related to habitat characteristics of the site. Our simulations showed that models derived from these data tend to overstate the importance of the habitat variables that are positively related to detection probability. Therefore, the conservation programs based on this model may be biased toward sites characterized by habitat covariates favoring detection, even though the sites may be sinks (Pulliam, 1988). For example, suppose that visual surveys are used to detect a species of rabbit in bottomland forest remnants. Further, suppose that the species preferentially uses stumps and fallen logs as vantage points when available. Under these circumstances, the density of stumps and fallen logs at a habitat may favor detection of the rabbit species due to its greater visibility. If this is the case, then the chance of non-detection in patches occupied by the rabbit will be smaller in patches that also contain high densities of stumps and fallen logs. Note that the relationship of

stumps and logs to detectability does not imply that they influence occupancy. However, logistic presence–absence models derived from this data set will tend to overstate the role of the stumps and logs as factors influencing occupancy (and by association, habitat suitability).

Few studies have been conducted to examine factors affecting probability of detection (but see McKenzie et al., 2002; Royle and Nichols, 2003). Kery (2002) investigated the variables affecting probability of detection of snake populations. Based on known species occupancy and population size in habitat patches, he found that population size and sampling season were the two major factors influencing probability of detection. His study highlighted the need to investigate the factors relating to the probability of detection in sampling studies. Recently, MacKenzie et al. (2002) demonstrated how a robust measurement of the effect of non-detection error on patch occupancy models for closed populations in which the probabilities of occupancy and detection are differentiated. An important advantage of the likelihood-based method of MacKenzie et al. (2002) is its ability to model both detection probability and occupancy as

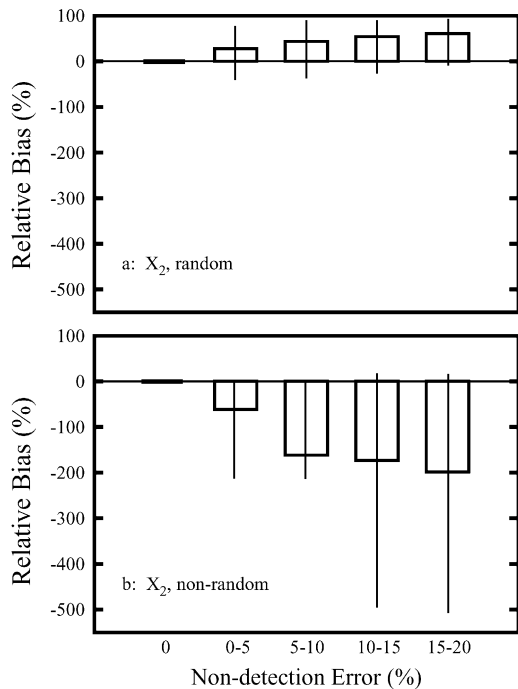


Fig. 4. Effect of non-detection of species occurrence on parameter estimation of habitat models in which  $X_2$  covaries negatively with the probability of patch occupancy. Overall species occupancy is 54%. Patterns for  $b_1$  match those in Fig. 2b and e and are not shown here. Random non-detection error yielded overestimation (a), whereas linkage of the probability of non-detection to the covariates resulted in underestimation of  $b_2$  (b).

functions of habitat covariates. Using their approach, a logistic function is appropriate for linking habitat covariates to estimates of detectability and occupancy. As highlighted in our study and the work by MacKenzie et al. (2002), estimating and correcting for non-detection error requires multiple sampling occasions. Multiple occasions should become a standard component of presence–absence sampling protocols, with designs that allow sampling to be completed in a period of time short enough to ensure that local extinction or colonization events do not occur. Estimating the relationships between the probability of detection and habitat covariates could assist in identifying the habitat patches where more sampling effort is needed to reduce non-detection error. Our results emphasize the importance of sound sampling design in ecological studies aimed at developing wildlife–habitat models in situations where the potential exists for non-detection error.

### Acknowledgements

We thank N. A. Slade, J. E. Moore, S. M. Knapp, and M. S. Miller for reviewing the earlier version of the manuscript. J. Nichols and an anonymous reviewer provided useful comments on the manuscript. T. E. Nupp and numerous research technicians were instru-

mental in completing the release–recapture component of this study. Funding was provided by National Research Initiative Competitive Grants Program/United States Department of Agriculture award 93-37101-8702, by the National Science Foundation under Grant Number SES-0119908, and by the John S. Wright Fund in support of the Upper Wabash Ecosystem Project. This is manuscript number 16693 of the Purdue Agricultural Research Programs.

### References

- Augustin, N.H., Muggleston, M.A., Buckland, S.T., 1996. An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology* 33, 339–347.
- Bergeron, Y., Leduc, A., Li, T., 1997. Explaining the distribution of Pinus's in a Canadian boreal insular landscape. *Journal of Vegetation Science* 8, 37–44.
- Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, B* 36, 192–236.
- Boulinier, T., Nicholds, J.D., Sauer, J.R., Hines, J.E., Pollock, K.H., 1998. Estimating species richness: the importance of heterogeneity in species detectability. *Ecology* 79, 1018–1028.
- Buckland, S.T., Elston, D.A., Beaney, S.J., 1996. Predicting distributional change, with application to bird distributions in northeast Scotland. *Global Ecology and Biogeography Letters* 5, 66–84.
- Cinchy, M., Haydon, D.T., Smith, A.T., 2002. Pattern does not equal process: what does patch occupancy really tell us about metapopulation dynamics? *The American Naturalist* 159, 351–362.
- Collett, D., 1991. *Modelling Binary Data*. Chapman & Hall, London.
- Copas, J.B., 1988. Binary regression models for contaminated data (with discussion). *Journal of Royal Statistical Society Series B* 50, 225–265.
- Cowley Jr., M., Wilson, R.J., Leon Cortes Jorge, L., Gutierrez, D., Bulman, C.R., Thomas, C.D., 2000. Habitat-based statistical models for predicting the spatial distribution of butterflies and day-flying moths in a fragmented landscape. *Journal of Applied Ecology* 37, 60–72.
- Dettmers, R., Buehler, D.A., Bartlett, J.G., Klaus, N.A., 1999. Influence of point count length and repeated visits on habitat model performance. *Journal of Wildlife Management* 63, 815–823.
- Edwards Jr., T.C., Deshler, E.T., Foster, D., Moisen, G.G., 1995. Adequacy of wildlife habitat relation models for estimating spatial distributions of terrestrial vertebrates. *Conservation Biology* 10, 263–270.
- Gardner, S., Baard, E.H.W., le-Roux, N.J., 1999. Estimating the detection probability of the geometric tortoise. *South African Journal of Wildlife Research* 29, 62–71.
- Gu, W., Kuusinen, M., Konttinen, Tuija, Hanski, I., 2001. Spatial pattern in the occurrence of the lichen *Lobaria pulmonaria* in managed and virgin boreal forests. *Ecography* 24, 139–150.
- Gutzwiller, K.J., Barrow Jr., W.C., 2001. Bird–landscape relations in the Chihuahuan desert: coping with uncertainties about predictive models. *Ecological Applications* 11, 1517–1532.
- Hanski, I., 1998. Metapopulation dynamics. *Nature* 396, 41–49.
- Hanski, I., 1994. A practical model of metapopulation dynamics. *Journal of Animal Ecology* 63, 151–162.
- Hinsley, S.A., Bellamy, P.E., Newton, I., Sparks, T.H., 1995. Habitat and landscape factors influencing the presence of individual breeding bird species in woodland fragments. *Journal of Avian Biology* 26, 94–104.
- Kerr, J.T., Sugar, A., Packer, L., 2000. Indicator taxa, rapid biodiversity assessment, and nestedness in an endangered ecosystem. *Conservation Biology* 14, 1726–1734.



- Kery, M., 2002. Inferring the absence of a species—a case study of snakes. *Journal of Wildlife Management* 66, 330–338.
- Klute, D.S., Lovallo, M.J., Tzilkowski, W.M., 2002. Autologistic regression modeling of American woodcock habitat use with spatially dependent data. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, DC, pp. 335–343.
- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A., Langtimm, C.L., 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83, 2248–2255.
- Mancke, R.G., Gavin, T.A., 2000. Breeding bird density in woodlots: effects of depth and buildings at the edges. *Ecological Applications* 10, 598–611.
- Mladenoff, D.J., Sickley, T.A., Haight, R.G., Wydevan, A.P., 1995. A regional landscape analysis and prediction of favorable gray wolf habitat in the northern Great Lakes region. *Conservation Biology* 9, 279–294.
- Moilanen, A., 2002. Implications of empirical data quality to meta-population model parameter estimation and application. *Oikos* 96, 516–530.
- Mortberg, U.M., 2001. Resident bird species in urban forest remnants: landscape and habitat perspectives. *Landscape Ecology* 16, 193–203.
- Nichols, J.D., Boulinier, T., Hines, J.E., Pollock, K.H., Sauer, J.R., 1998a. Inference methods for spatial variation in species richness and community composition when not all species are detected. *Conservation Biology* 12, 1390–1398.
- Nichols, J.D., Boulinier, T., Hines, J.E., Pollock, K.H., Sauer, J.R., 1998b. Estimating rates of local species extinction, colonization, and turnover in animal communities. *Ecological Applications* 8, 1213–1225.
- Nichols, J.D., Hines, J.E., Sauer, J.R., Fallon, F.W., Fallon, J.E., Heglund, P., 2000. A double-observer approach for estimating detection probability and abundance from point counts. *Auk* 117, 393–408.
- Nupp, T.E., Swihart, R.K., 1996. Effect of forest patch area on population attributes of white-footed mice (*Peromyscus leucopus*) in fragmented landscapes. *Canadian Journal of Zoology* 74, 467–472.
- Nupp, T.E., Swihart, R.K., 1998. Influence of forest fragmentation on population attributes of white-footed mice and eastern chipmunks. *Journal of Mammalogy* 79, 1234–1243.
- Nupp, T.E., Swihart, R.K., 2000. Landscape-level correlates of small mammal assemblages in forest fragments of farmland. *Journal of Mammalogy* 81, 512–526.
- Odell, E.A., Knight, R.L., 2001. Songbird and medium-sized mammal communities associated with exurban development in Pitkin County, Colorado. *Conservation Biology* 15, 1143–1150.
- Odom, R.H., Ford, W.M., Edwards, J.W., Stihler, C.W., Menzel, J.M., 2001. Developing a habitat model for the endangered Virginia northern flying squirrel (*Glaucomys sabrinus fuscus*) in the Allegheny Mountains of West Virginia. *Biological Conservation* 99, 245–252.
- Otis, D.L., Burnham, K.P., White, G.C., Anderson, D.R., 1978. Statistical inference for capture data on closed animal populations. *Wildlife Monographs* 62, 1–135.
- Palmgren, J., Ekholm, A., 1987. Exponential family non-linear models for categorical data with errors of observation. *Applied Stochastic Modeling and Data Analysis* 3, 111–124.
- Pauley, G.R., Peek, J.M., Zager, P., 1993. Predicting white-tailed deer habitat use in northern Idaho. *Journal of Wildlife Management* 57, 904–913.
- Pulliam, H.R., 1988. Sources, sinks, and population regulation. *American Naturalist* 132, 652–661.
- Rexstad, E., Burnham, K., 1991. User's Guide for Interactive Program CAPTURE. Colorado Cooperative Fish and Wildlife Research Unit, Fort Collins, Colorado.
- Royle, J.A., Nichols, J.D., 2003. Estimating abundance from repeated presence absence data or point counts. *Ecology* 84, 777–790.
- Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, DC.
- Stauffer, H.B., Ralph, C.J., Miller, S.L., 2002. Incorporating detection uncertainty into presence-absence surveys for marbled murrelet. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington, DC, pp. 357–365.
- Tyre, A.J., Possingham, H.P., Lindenmayer, D.B., 2001. Matching observed pattern with model process: can territory occupancy provide information about life history parameters? *Ecological Applications* 11, 1722–1737.
- White, G.C., Burnham, K.P., 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study* 46 (supp), 120–138.
- Wiser, S.K., Peet, R.K., White, P.S., 1998. Prediction of rare-plant occurrence: a southern appalachian example. *Ecological Applications* 8, 909–920.