

## IMPROVING PRECISION AND REDUCING BIAS IN BIOLOGICAL SURVEYS: ESTIMATING FALSE-NEGATIVE ERROR RATES

ANDREW J. TYRE,<sup>1,5</sup> BRIGITTE TENHUMBERG,<sup>1</sup> SCOTT A. FIELD,<sup>1,2</sup> DARREN NIEJALKE,<sup>3</sup> KIRSTEN PARRIS,<sup>4</sup>  
AND HUGH P. POSSINGHAM<sup>1</sup>

<sup>1</sup>*The Ecology Centre, University of Queensland, St. Lucia, Queensland 4072, Australia*

<sup>2</sup>*Department of Applied and Molecular Ecology, Waite Campus, University of Adelaide PMB 1, Glen Osmond, South Australia 5064, Australia*

<sup>3</sup>*Western Mining Corporation Ltd. (Olympic Dam), PO Box 150, Roxby Downs, South Australia 5725, Australia*

<sup>4</sup>*Australian Research Centre for Urban Ecology, Royal Botanic Gardens Melbourne, c/o School of Botany, University of Melbourne, Victoria 3010, Australia*

**Abstract.** The use of presence/absence data in wildlife management and biological surveys is widespread. There is a growing interest in quantifying the sources of error associated with these data. We show that false-negative errors (failure to record a species when in fact it is present) can have a significant impact on statistical estimation of habitat models using simulated data. Then we introduce an extension of logistic modeling, the zero-inflated binomial (ZIB) model that permits the estimation of the rate of false-negative errors and the correction of estimates of the probability of occurrence for false-negative errors by using repeated visits to the same site. Our simulations show that even relatively low rates of false negatives bias statistical estimates of habitat effects. The method with three repeated visits eliminates the bias, but estimates are relatively imprecise. Six repeated visits improve precision of estimates to levels comparable to that achieved with conventional statistics in the absence of false-negative errors. In general, when error rates are  $\leq 50\%$  greater efficiency is gained by adding more sites, whereas when error rates are  $> 50\%$  it is better to increase the number of repeated visits. We highlight the flexibility of the method with three case studies, clearly demonstrating the effect of false-negative errors for a range of commonly used survey methods.

**Key words:** *biological surveys; false-negative errors; habitat effects; presence–absence data; zero-inflated binomial (ZIB) model.*

### INTRODUCTION

Presence–absence data are widely collected for a range of taxa—as part of biological surveys, ecological monitoring programs, or other sampling regimes. The results of such samples are used to assess the efficacy of management actions, to look for species declines or reductions in range, or to model the habitat of a species. A ubiquitous, but underappreciated, problem with such survey data is the presence of false-negative errors. That is, a visit to a site fails to record a species when it is in fact present. The probability of such an error is likely to vary among species, with weather or other local conditions, observer experience, survey methodology, and many other factors.

There is a growing concern with quantifying such measurement error. For example, Lindenmayer et al. (2001) tested the efficacy of spotlighting for arboreal marsupials by simultaneously using radio tracking to locate animals. They found that spotlighting missed a large proportion of the animals involved, although it

is widely used to estimate abundances of arboreal marsupials. Craig and Roberts (2001) did a similar comparison using radio-tagged birds and timed area searches. They tallied the number of opportunities that searchers had to count radio-tagged individuals known to be in the area from radiolocations, and then calculated a sighting probability as the number of actual sightings divided by the number of opportunities. Repeated pit-fall trapping has also been used to look for optimum configurations of effort for sampling arid-zone reptiles and small mammals (Moseby and Read 2001, Read and Moseby 2001). B. A. Wintle, M. A. Burgman, and R. P. Kavanagh (*unpublished manuscript*) show how to use survival analysis to determine detection probabilities.

False negatives are not the only type of error that can occur in a presence–absence survey. False positives, or recording a species as present when it is in fact absent, are also possible. We consider this kind of error as less prevalent for two reasons. First, false positives reflect an error in identifying an organism after it has been seen, and this sort of mistake should decline rapidly even with modest amounts of observer experience, training, and good survey protocols. Second, it is common practice to not record a species if there is any doubt about its identity. This converts many false

Manuscript received 28 February 2002; revised 22 November 2002; accepted 10 December 2002; final version received 3 April 2003. Corresponding Editor: F. C. James.

<sup>5</sup> Present address: School of Natural Resource Sciences, 202 Natural Resources Hall, University of Nebraska, Lincoln, Nebraska 68583-0819 USA. E-mail: atyre2@unl.edu

positives into false-negative results, and thus will tend to inflate the rate of false negatives.

In this paper we introduce a method using zero-inflated binomial models (Hall 2000) of repeated surveys at each site to estimate false-negative rates, and also accommodate the effects of other variables. First, we demonstrate the effect that even modest false-negative error rates can have on ecological conclusions drawn from presence/absence data. Second, we detail the method, and test its ability to recover parameter estimates from simulated data. Third, we apply the method to three case studies: woodland birds, forest dwelling frogs, and aquatic invertebrates.

EFFECTS OF FALSE NEGATIVES ON HABITAT MODELING

A typical scenario for a wildlife ecologist interested in habitat modeling is to visit a large number of sites, record the presence or absence of one or more species at each site using a standard survey method, and then to measure some number of habitat or landscape attributes at that site. The results of the study are subjected to logistic regression, and a statistical habitat model constructed (Austin et al. 1990, Lindenmayer et al. 1990, 1995, Morrison et al. 1992). This model identifies habitat or landscape features that are associated with the presence of the species. Habitat models can aid management decisions about preservation or reconstruction of habitats. If an ecologist makes false-negative errors, what effect does this have on the ability of a statistical habitat model to detect significant predictors of suitable habitat?

We explore the effect of false negatives on habitat models with a “virtual ecologist” (Tyre et al. 2001). First, we simulate data from a straightforward habitat model with a single habitat variable positively affecting the probability of detecting a species:

$$Y_i \sim \text{Bernoulli}(p_i) \quad \ln\left(\frac{p_i}{1-p_i}\right) = a + b \times h_i$$

$$b > 0 \quad i = 1 \dots 100 \quad h \sim \mathcal{N}(0, 1) \quad (1)$$

where  $Y_i$  is the true occupancy of site  $i$ ,  $p_i$  is the probability the site is occupied,  $h_i$  is the value of the habitat variable in site  $i$ ,  $a$  is the “intercept,” and  $b$  is the slope of the true habitat model. When  $b = 0$  there is no effect of the habitat variable on the probability that a site is occupied. In all cases, we assume the habitat variable is normally distributed with a mean of 0 and a variance of 1, and that 100 sites are sampled. Therefore, the proportion of occupied sites varies from run to run depending on the random values of the habitat variable and the particular value for the slope. We then use a virtual ecologist to collect a sample from this simulated habitat while making errors at a fixed rate. False-negative errors are introduced by randomly converting observations of occupied sites to unoccupied sites with a probability  $1 - q$ , the false-negative rate. A logistic

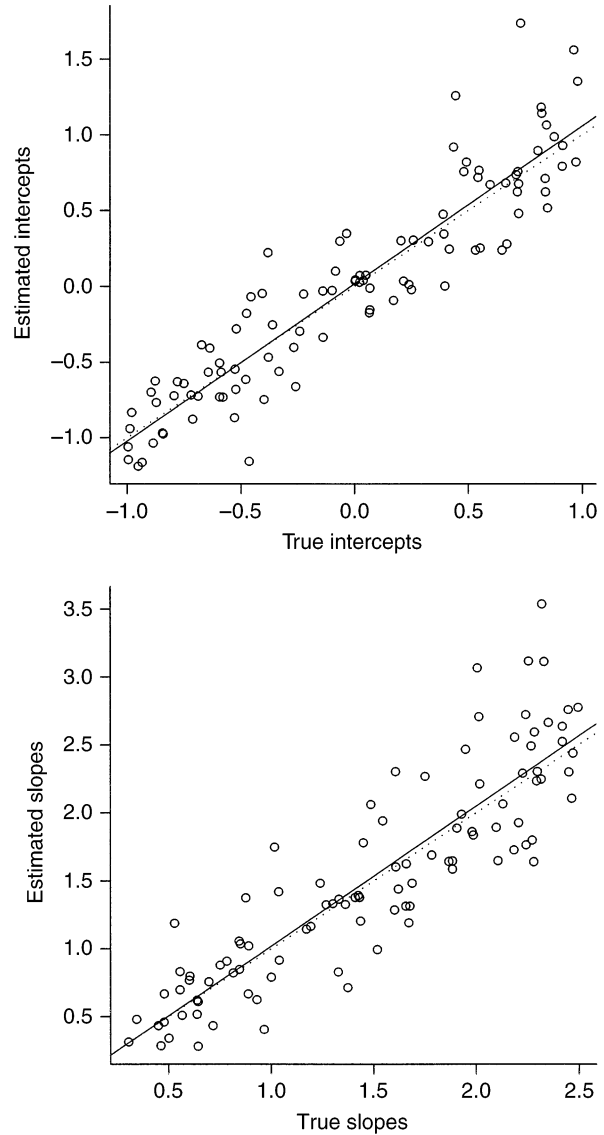


FIG. 1. Accuracy of logistic regression at recovering the true values of a statistical habitat model when there is no observer error. Each point is a pair of randomly selected parameters, the dotted line is the 1:1 line, and the solid line is a least-squares fit.

regression model is fitted to the sampled data, and the success of the exercise evaluated by comparing the estimated effect of the habitat variable  $\hat{b}$  on site occupancy with the true, known effect  $b$ . We vary both  $a$  and  $b$  in the true habitat model over a wide range of values to look for regions where the model is particularly sensitive to false-negative errors.

When there is no observer error, both the slope and the intercept of the habitat–occupancy relationship are accurately estimated by logistic regression (Fig. 1). However, when false negatives occur 20, 40, or 60% of the time, both the intercept and slope are underestimated, and the effect on the slope is greater for larger

slopes (Fig. 2A and B). That is, the stronger the effect a habitat variable has on occupancy, the more the magnitude of that effect is underestimated in the presence of modest amounts of false-negative errors. This effect is well known from more general analyses of errors in logistic regression (Collett 1991).

The effect on the intercept can be understood intuitively by considering that false negatives decrease the overall proportion of presences in the data set; the smaller the proportion of presences the lower the intercept. The effect of false-negative errors on the estimated slope is subtler. When the true slope  $b$  is small, the true presences will tend to be distributed throughout the range of the habitat variable, and randomly "missing" presences will likewise occur throughout the full range of the habitat variable. In the extreme ( $b = 0$ ), false negatives should have no effect on the estimate of the habitat-quality effect; they will only reduce the estimated intercept. However, when there is a strong habitat-quality effect ( $b \gg 0$ ), then the true presences (and thus the false-negative errors) will tend to be concentrated at one end of the range of the habitat variable. This dilutes the apparent effect of the habitat variable on the probability of a presence.

It is conceivable that false-negative errors could lead to an overestimation of the effect of habitat. Consider a bird that uses two discrete habitats, one open habitat and one closed, and that it marginally prefers the open habitat. However, it is more readily seen, and its calls carry farther, in the open habitat, reducing the rate of false-negative errors (Verner 1985). In this instance the false-negative errors are concentrated in the less preferred habitat, and the difference between these two habitats would be *overestimated*. The solution to both problems is to estimate the false-negative rate, and how it varies with habitat or other variables. We describe a method for doing this in the next section.

#### ESTIMATING FALSE-NEGATIVE RATES FROM REPEATED SURVEYS

We assume that the result for any given survey (species present or not present) is the outcome of two binomial processes acting simultaneously at two different time scales: (1) the probability that the species is indeed utilizing the site ( $p$ ) over a longer time period; and (2) the probability that the species is present within the site and observed in any given survey visit within that time period, given that it utilizing the site ( $q$ ). Thus the survey results follow a "finite mixture distribution" with a mixing probability ( $p$ , the probability a site is occupied), and two binomial components, one with a probability of success equal to 0. This is the zero-inflated binomial (ZIB) model described by Hall (2000), who used an expectation-maximization algorithm in combination with the iteratively re-weighted least-squares algorithm to find parameter estimates. We use straightforward maximum-likelihood methods to estimate the unknown parameters (Hilborn and Mangel

1997); both approaches reach identical results, and require some programming to implement.

The probability  $q$  is more than the probability that a species is observed. For species whose home ranges exceed the survey area it also includes the likelihood that the individuals are present within the survey area at the time it is conducted. Species with large home ranges may be in parts of their home range outside the survey site, and therefore would not be recorded as present even though they utilize the survey site as part of their home range. For species that have small home ranges, or are not very mobile over the length of time between repeated surveys,  $q$  will mostly represent a pure probability of observation.

To begin with, assume that  $n$  sites have been visited  $m$  times each and that a species has a probability of occupying a site,  $p$ , which is constant for all sites throughout the landscape. A species also has a species-typical probability  $q$  of being observed during any one visit. The false-negative error rate is  $1 - q$ . After the  $m$  visits are complete, the number of observations of a species at a given site is  $y$  ( $y \leq m$ ). Given our assumptions the likelihood of  $y$  observations from  $m$  visits to a site is

$$L(y|\hat{p}, \hat{q}) = \hat{p} \binom{m}{y} \hat{q}^y (1 - \hat{q})^{m-y} \quad y > 0 \quad (2)$$

$$L(y|\hat{p}, \hat{q}) = (1 - \hat{p}) + \hat{p}(1 - \hat{q})^m \quad y = 0. \quad (3)$$

Equation 2 is the probability that the species was observed  $y$  times conditional on it being present, and is the likelihood for an observation where  $y > 0$ . Equation 3 is the probability that it was not there plus the probability that it was there but was not observed in  $m$  visits. The negative logarithms of these likelihoods are summed over all sites, and this value numerically minimized to find the maximum-likelihood estimates of the two parameters,  $\hat{p}$  and  $\hat{q}$ .

In the equations above, we assume that  $p$ ,  $q$ , and  $m$  are identical across all sites. However, we are generally interested in identifying covariates that influence  $p$  ( $q$  may also vary between sites). Following the procedure used for generalized linear modeling we allow covariates to linearly influence the log odds of  $\hat{p}$ :

$$\ln\left(\frac{\hat{p}_i}{1 - \hat{p}_i}\right) = \hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \dots + \hat{\beta}_k X_{k,i} \\ i = 0 \dots n \quad j = 0 \dots k \quad (4)$$

where  $X_i$  is the value of the covariate at site  $i$ ,  $\beta_j$  is a parameter describing the influence of covariate  $j$ , or the position of the intercept. Covariates can influence  $\hat{q}$  in the same manner, and need not be the same variables as for  $\hat{p}$ .

The method can be easily implemented in readily available spreadsheet software using built-in optimization functions (e.g., Solver in Microsoft Excel). We implemented the method in S-plus 2000 for Windows,

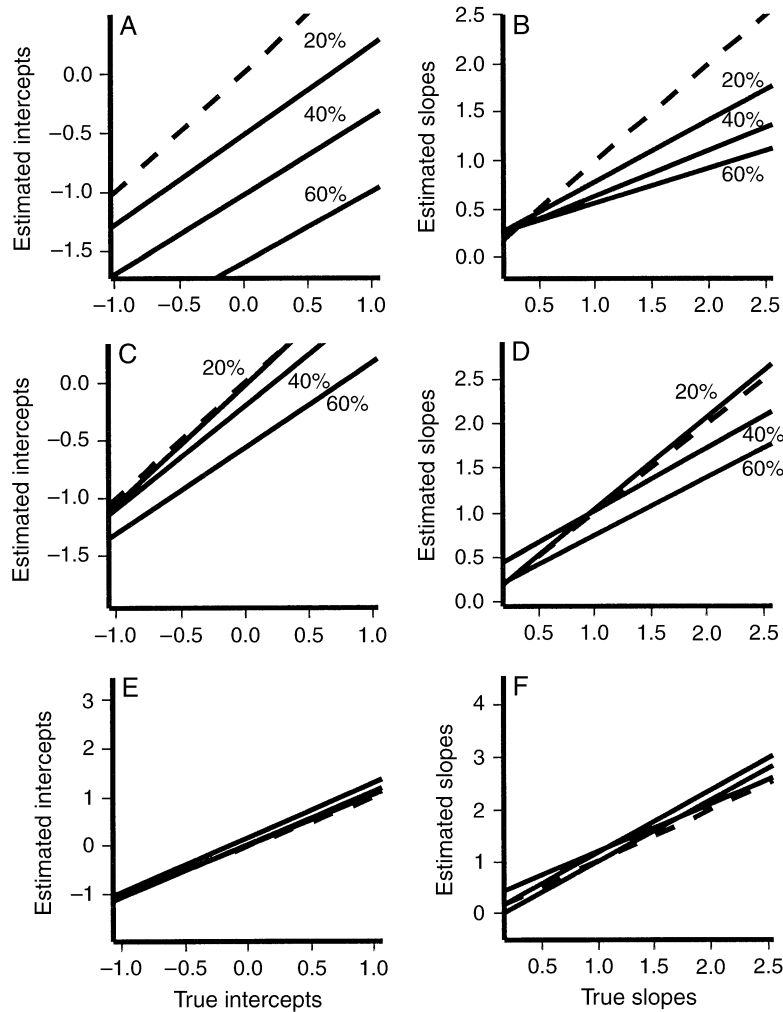


FIG. 2. Effect of 20%, 40%, and 60% false-negative observations on estimated parameters of a statistical habitat model. (A, B) One visit to 100 sites analyzed with logistic regression. (C, D) Three visits to 100 sites analyzed with logistic regression. (E, F) Three visits to 100 sites analyzed with the zero-inflated binomial (ZIB) model. Points have been omitted for clarity. The dashed line is the 1:1 line, while each of the three solid lines is a least-squares fit line.

(Insightful Corporation, Seattle, Washington, USA), using the nonlinear minimization routine `nlminb` to maximize Eqs. 2 and 3, and in R 1.6.0 using the nonlinear minimization routine `optim`. An R package for carrying out the analyses, including documentation, is available as a Supplement. We calculate confidence intervals numerically with likelihood profiles (Hilborn and Mangel 1997) by searching for parameter values that give total log likelihood 1.97 greater than the likelihood at the maximum-likelihood estimate.

We compared alternative models using Akaike's Information Criterion (AIC), which evaluates models based on both the likelihood of the model and the number of estimated parameters:

$$AIC = 2L + 2K \tag{5}$$

where  $L$  is the negative log likelihood of the model and  $K$  is the number of parameters. Generally a model with

a lower AIC value is considered to be better. However it is possible to go beyond a simple ranking of models by calculating Akaike weights, which are the probability that each model out of a set of models is closest to the correct one (Burnham and Anderson 1998). First calculate a  $\Delta_i$  value for each model by subtracting the AIC of the best model from the AIC of all models. The Akaike weight  $w_i$  for model  $i$  is then

$$w_i = \frac{e^{-\Delta_i/2}}{\sum_j e^{-\Delta_j/2}} \tag{6}$$

and this weighted distance between model  $i$  and the best model can be interpreted as the probability that model  $i$  is correct. Beginning with the best model, a 95% confidence set of models is constructed by adding models to the set until the sum of their Akaike weights exceeds 0.95.

TABLE 1. Precision of parameter estimates under different methods, for different error rates.

Sampling plan†			Parameter‡	Error rates§			
Method	<i>m</i>	<i>n</i>		0%	20%	40%	60%
Logistic	1	100	intercept	0.054	0.079	0.066	0.086
Logistic	3	100	intercept	0.479	0.437	0.326	0.244
ZIB	3	100	intercept	...	0.072	0.124	0.271
ZIB	3	200	intercept	...	0.030	0.028	0.099
ZIB	6	100	intercept	...	0.088	0.071	0.083
Logistic	1	100	slope	0.174	0.099	0.087	0.127
Logistic	3	100	slope	0.681	0.579	0.352	0.312
ZIB	3	100	slope	...	0.123	0.452	0.315
ZIB	3	200	slope	...	0.071	0.095	0.161
ZIB	6	100	slope	...	0.122	0.119	0.116
ZIB	3	100	<i>q</i>	...	0.051	0.052	0.066
ZIB	3	200	<i>q</i>	...	0.025	0.022	0.029
ZIB	6	100	<i>q</i>	...	0.026	0.019	0.017

† Abbreviations are: *m*, number of visits per site; *n*, number of sites visited; ZIB, zero-inflated binomial model.

‡ Intercept and slope values are the variances of residuals from least-squares fits between estimated and true parameter values. Precision of *q* is simply the variance of the estimate for each error rate.

§ The ZIB method was not fitted to data with no errors.

Background theory on mixture distributions suggests that at least three replicate visits are required to ensure  $\hat{p}$  and  $\hat{q}$  can be estimated (Titterton et al. 1985). In fact, these conditions are not as strict for our circumstance, because the component identity of some of the sites (the ones where the species is sighted at least once) is known. This is also why straightforward maximum likelihood works, and we do not require more sophisticated algorithms commonly employed for mixture estimation.

For an initial test of the method, we repeat the exercise performed above for logistic habitat modeling with a single covariate. We perform the calculation for three and six replicate visits to 100 sites, and for three visits to 200 sites. For comparison, we also analyze the three-visits, 100-sites scenario using logistic regression after “collapsing” all the observations across visits so that a presence is recorded if the species was observed at least once in the three visits. The last two scenarios have double the sampling effort of the first scenario, but the effort is distributed either among extra visits to the same sites or new sites.

Carrying out three repeat visits to 100 sites, collapsing the results to presence/absence, and using logistic regression eliminates much or all of the bias, especially if error rates are relatively low (Fig. 2C and D). Applying the ZIB method to this scenario corrects for the bias at all error rates (Fig. 2E and F), to the extent that there is no discernable difference in the average estimates across three increasing false-negative error rates. Doubling the sampling effort obviously improves the situation more (scenarios not shown), regardless of whether the extra effort is put into new sites or more visits to the same sites.

Bias is only one measure of the accuracy of a statistical estimate; we must also consider the precision

of the estimates, or the amount of scatter around the true value. We evaluate the precision by looking at the residual variance around the least-squares regression lines (Table 1). Although the logistic-regression estimates are increasingly biased with increasing error rates, the precision of the estimates does not change. Analyzing repeated-visit data with logistic regression, while not as biased as not having repeat-survey data, produces estimates with generally larger variances than the ZIB estimates. The precision of ZIB estimates does decrease with increasing error rates when only three replicate visits at 100 sites are conducted. Precision of the ZIB estimates is constant and comparable to logistic-regression estimates without error when the sampling effort is doubled to either six repeated visits or 200 sites. When error rates are low (<50%, roughly), precision is improved by adding more sites. When error rates are higher, precision is improved by increasing the number of visits to the same sites. Extra survey effort, distributed across either new sites or more visits, yields more-precise estimates of the error rate, at all levels of error considered.

To further illuminate the effect on the variance of the estimates of having more sites or more visits we carried out an additional set of simulations with  $a = 0$ ,  $b = 1$  and  $q = 0.8$  or  $q = 0.4$ . We simulated 100 data sets for three, six, or nine visits, and varied the number of sites systematically from 50 to 200. The ZIB estimates for all parameters were unbiased over all sampling plans (not shown), consistent with the results shown above. When  $q = 0.8$  the variance of  $\hat{b}$  decreases with the number of sites only; adding more visits beyond 3 does not reduce the variance (Fig. 3A). Adding more visits does decrease the variance of  $\hat{q}$  (Fig. 3B), and doubling the number of sites or the number of visits appears to have roughly similar effects. When  $q = 0.4$



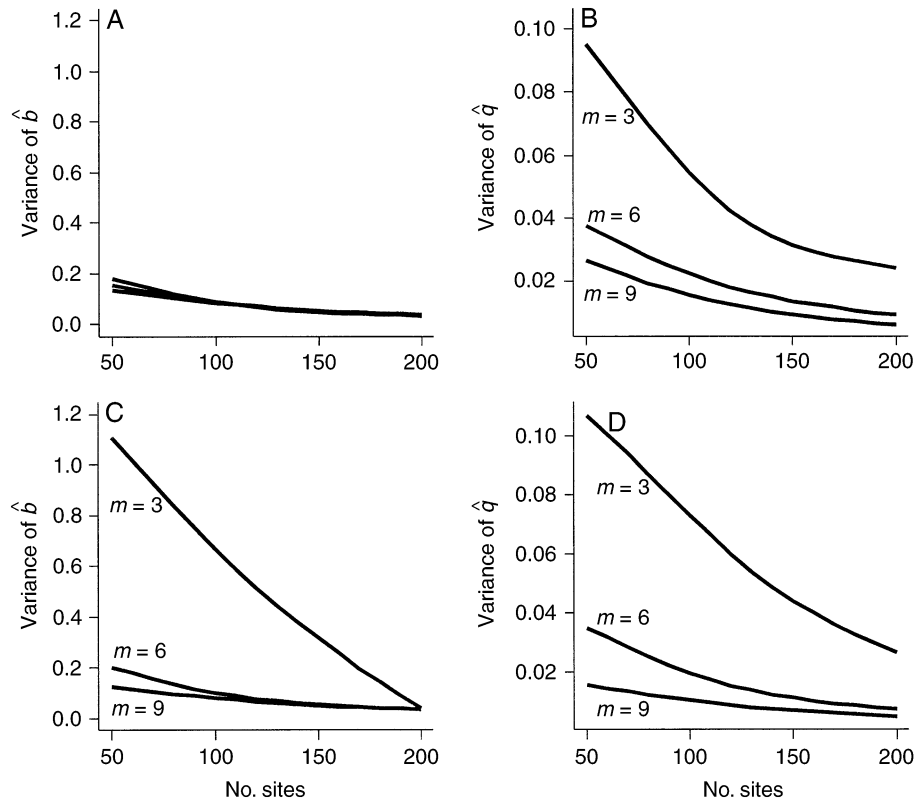


FIG. 3. Variance of parameter estimates from the ZIB method for different numbers of visits ( $m$ ) as a function of the number of sites for (A, B)  $a = 0$ ,  $b = 1$ , and  $q = 0.8$  or (C, D)  $a = 0$ ,  $b = 1$ , and  $q = 0.4$ , where  $a$  is the intercept,  $b$  is the slope of the true habitat model, and  $q$  is the species-typical probability of being observed during any one visit. Lines are smooth splines fit through simulation results.

the variance of  $\hat{b}$  is also affected strongly by the number of visits (Fig. 3C); the variance is much higher with three visits than with six for nearly all numbers of sites. The variance of  $\hat{q}$  varies similarly to that observed when  $q$  is higher (Fig. 3D). When the number of sites is relatively low ( $<100$ ), it appears that better estimates would be obtained by doubling the number of visits to existing sites rather than increasing the number of sites.

We now show how these methods may be applied to improve survey reliability in some typical ecological scenarios. We consider three different study systems: woodland birds, forest-dwelling frogs, and mound-spring invertebrates.

#### CASE STUDIES

##### Woodland birds

The study was conducted in remnant patches of stringybark woodland in the high-rainfall areas of the Southern Mount Lofty Ranges, South Australia (Field and Tyre 2002). The patches were of relatively uniform habitat type, in which the tree cover was wholly or mostly ( $>90\%$ ) messmate stringybark, *Eucalyptus obliqua* (32 patches), or brown stringybark, *E. baxteri* (two patches). This is the most abundant and widespread class of native vegetation remaining in the

Southern Mount Lofty Ranges, enabling us to obtain data from a wide range of patch sizes (4–1686 ha) evenly spread over a broad geographic area ( $\sim 150 \times 40$  km, from Morialta Conservation Park in the north ( $34^\circ 54.22' S$ ,  $138^\circ 43.44' E$ ) to Deep Creek Conservation Park in the south ( $35^\circ 36.01' S$ ,  $138^\circ 13.70' E$ ). Thirty-four sites within this area were selected. Surveys were undertaken over spring/summer between November and February in 1999 and 2000, starting no earlier than 0545 hours and finishing no later than 1400 hours (Australian Central Standard Daylight Saving Time). Migrant species to the region such as kingfishers and cuckoos typically arrive well before this starting date and remain until the autumn, justifying our assumption that the bird community was relatively unchanging for the duration of the study.

We used the 20-min–2-ha search method (Loyn 1986). Sites were circumscribed by pacing out a 2-ha area and only those birds using the habitat within this area were recorded. This included birds seen perching within the site, or heard calling from a stationary position within the site for part or all of the 20-min survey period. Birds flying over or through the site but not alighting were not included, with the exception of swallows, martins, woodswallows, and birds of prey that were actively foraging.

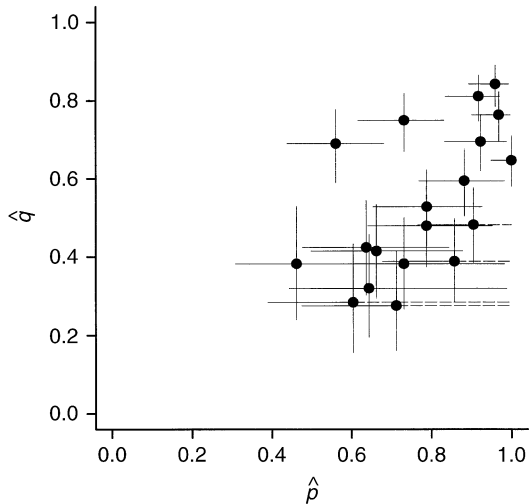


FIG. 4. Estimates of  $p$  and  $q$  for the Mount Lofty Ranges (South Australia) bird-survey data. Bars represent likelihood profile 95% confidence limits. Some upper limits for  $p$  are indicated with dashed lines; in these cases the upper limit is indistinguishable from 1.

In both years sites were visited three times on different days, with no more than 90 d separating the first and last visits to a site. We used the ZIB (zero-inflated binomial) method to find  $\hat{p}$  and  $\hat{q}$  for the 19 species observed at 10 or more sites in both years.

In an initial test, we combined all the data for both years, and estimated  $p$  and  $q$  for each of the 19 species, along with 95% likelihood profile confidence limits for both parameters (Fig. 4). The parameter  $\hat{p}$  is generally larger than  $\hat{q}$ , and the estimates of both parameters are generally more precise (smaller confidence limits) when  $\hat{p}$  is larger. It is clear that many species are subject to significant false-negative error rates under this survey method, a point that must be kept in mind when these survey data are analyzed.

A numerical problem arises when searching for the upper confidence limit of  $\hat{p}$  when  $\hat{q}$  is relatively low; the likelihood does not converge to an upper estimate before the numerical calculation overflows at an intermediate step. In these cases the logit of the probability is generally  $>20$  (i.e.,  $p > 1-10^{-6}$ ) before overflow occurs, and we interpret this as an effective upper confidence limit of 1 (shown in Fig. 4 as dashed lines for upper confidence limits). It appears to be difficult to distinguish between two cases when  $\hat{q}$  is relatively low. First, the species occurs nearly everywhere (high  $\hat{p}$ ) but is hard to detect (low  $\hat{q}$ ). Second, the species is less common (lower  $\hat{p}$ ) but is easier to detect (higher  $\hat{q}$ ). In both cases the most common result is to not observe the species at all out of  $m$  visits. The solution is to have more repeated visits to sites to improve the estimates of  $\hat{q}$ .

We examined the use of covariates for  $\hat{p}$  in this case by considering three additional candidate models, a

year effect (comparing 1999 vs. 2000), a landscape metric (the total area of forest within a 2-km radius of the survey point), and both effects together. The landscape metric was chosen on the basis of prior work in the region, with historical data suggesting that this metric has the strongest effect on bird species occurrence at particular sites (Westphal et al. 2003). There is tremendous model-selection uncertainty (Table 2), the model without covariates is often the best model, and never outside a 95% confidence set for model adequacy. This is more likely to reflect the relatively small sample size ( $\frac{1}{3}$  as many samples as our simulations), rather than a problem with the underlying model.

For comparison, we carried out a traditional logistic regression analysis after transforming the data into pure presence/absence by treating any site where the species was observed at least once in that year as a presence. This transformation discards information about how many times the species was observed (i.e., about  $q$ ), and is a commonly used technique to deal with sampling error.

From a model-selection perspective, the logistic analysis gave similar Akaike weights (to within 0.05) as the ZIB model for 14 of the 19 species with no changes in the rank order of the models. For the five species where the results differed, three had the lowest  $\hat{q}$ , between 0.28 and 0.31.

Coefficients for year or land-metric effects were also very similar, except where the effects were very strong, and then the logistic analysis produced much smaller estimates (e.g., Fig. 5). Three of these species had  $\hat{q} < 0.5$ . This was precisely the effect noted in the simulated data, where large effects of a habitat variable or other covariate were underestimated in the presence of substantial false-negative errors. Although these results may appear to suggest that logistic modeling is not much worse than using the ZIB method, recall that the demonstrations of the effect of errors on logistic regression with simulated data used only a single visit, while here we have assumed that three visits have been carried out to each site regardless of which method was used to analyze the data. Overcoming sampling errors requires extra effort; the ZIB method squeezes the most information out of such additional effort.

As with all statistical modeling, caveats abound. The largest negative estimated year effect was  $-20$  for Grey Fantails, a widely distributed common bird with  $\hat{q} \sim 0.85$ . However, this apparent catastrophic decline is purely a consequence of them appearing in every site in 1999, and in only 31 of 34 sites in 2000! These estimates suffered from considerable numerical problems with the likelihood failing to converge; it is worth noting that a normal GLM (generalized linear model) implementation would also struggle with such a data set.

#### *Forest-dwelling frogs*

Parris and McCarthy (1999) and Parris (2001, 2002) studied the habitat requirements of 14 species of

TABLE 2. Akaike weights and best-fit parameter estimates for the four models considered for the bird survey data.

Species	Model				Best-fit parameters†			
	No covariates	Year effect‡	TLA effects§	Year + TLA	$\beta_0$	$\beta_1$	$\beta_2$	$\hat{q}$
Adelaide Rosella ( <i>Platycercus elegans</i> var. <i>adelaidae</i> )	<b>0.4137</b>	0.221	0.24	0.127	0.92			0.81
Blackbird ( <i>Turdus merula</i> )	0.1165	0.063	<b>0.5</b>	0.322	4.51		-2.6	-0.1
Brown thornbill ( <i>Acanthiza pusilla</i> )	<b>0.4501</b>	0.166	0.28	0.104	0.88			0.59
Crescent honeyeater ( <i>Phylidronis melanops</i> )	0.2393	<b>0.449</b>	0.13	0.186	4.74	-3.77		-0.5
Eastern spinebill ( <i>Acanthorhynchus tenuirostris</i> )	0.1456	0.054	0.21	<b>0.596</b>	15	8.16	-10	-1.4
Golden Whistler ( <i>Pachycephala pectoralis</i> )	<b>0.4612</b>	0.17	0.27	0.1	0.73			0.38
Grey Currawong ( <i>Strepera versicolor</i> )	0.0809	0.088	<b>0.43</b>	0.406	0.68		0.88	-0.8
Grey Fantail ( <i>Rhipidura fuliginosa</i> )	0.1437	<b>0.398</b>	0.12	0.341	22.1	-19.7		1.69
Grey Shrike-Thrush ( <i>Colluricincla harmonica</i> )	0.123	0.052	<b>0.6</b>	0.226	1.52		-0.9	0.12
Red Wattlebird ( <i>Anthochaera carunculata</i> )	<b>0.513</b>	0.212	0.2	0.08	0.66			0.41
Scarlet Robin ( <i>Petroica multicolor</i> )	0.095	0.039	0.11	<b>0.759</b>	10.8	15	15	-1.2
Silvereye ( <i>Zosterops lateralis</i> )	<b>0.4927</b>	0.236	0.18	0.088	0.64			0.42
Striated Pardalote ( <i>Pardalotus striatus</i> )	<b>0.4769</b>	0.182	0.25	0.094	0.73			0.75
Striated Thornbill ( <i>Acanthiza lineata</i> )	<b>0.5344</b>	0.197	0.2	0.072	1			0.65
Superb Fairy Wren ( <i>Malurus cyaneus</i> )	<b>0.3062</b>	0.135	0.26	0.303	0.97			0.76
White-browed Scrubwren ( <i>Sericornis frontalis</i> )	<b>0.5081</b>	0.214	0.2	0.081	0.79			0.48
White-throated Treecreeper ( <i>Cormobates leucophaea</i> )	<b>0.5225</b>	0.208	0.19	0.077	0.92			0.7
Yellowfaced Honeyeater ( <i>Lichenostomus chrysops</i> )	0.282	<b>0.448</b>	0.1	0.166	0.69	-0.88		0.8
Yellow-tailed Black Cockatoo ( <i>Calyptorhynchus funereus</i> )	<b>0.5032</b>	0.211	0.2	0.085	0.46			0.38

Notes: The best model is shown in boldface type. Considerable model-selection uncertainty exists, and the no-covariates model is within the 95% confidence set of adequate models for all species. Parameter estimates for the best model are shown on a logit scale.

† Best-fit parameters:  $\beta_0$  = intercept,  $\beta_1$  = coefficient for the year effect, and  $\beta_2$  = coefficient for the total-land-area effect where these are part of the best model;  $\hat{q}$  is the logit of the probability of a species being observed in a single visit.

‡ Year effect = comparing 1999 vs. 2000.

§ TLA = total land area of stringybark forest within 2 km of the survey site.

stream-breeding frogs in the forests of southeast Queensland, Australia. As part of their work, two different methods for sampling frogs (nocturnal searches and automatic recording of advertisement calls) were simultaneously tested with between three and nine repeated surveys at 21 sites. Parris et al. (1999) argued that nocturnal searches were superior to automatic call recorders because on average they detected approximately twice as many species of frogs at a site. Here we fit the ZIB model to quantitatively evaluate the differences between these two sampling methods for two of the species.

This example highlights one limitation of the method—if there are insufficient numbers of positive records then adequate parameter estimates are not available. Only 2 of the 14 species, the great barred frog

*Mixophyes fasciolatus* and the cascade treefrog *Litoria pearsoniana*, had >7 out of 21 positive records for both survey methods. There were insufficient records of the other 12 species for one or both survey methods, and we do not report on them here. We fitted two observation-error models to the data, one with a parameter distinguishing the survey method, and one without, and used AIC to evaluate the evidence that the survey methods differed in their ability to detect the two frog species.

Nocturnal searches had  $\hat{q} = 0.41$  for *M. fasciolatus* and 0.51 for *L. pearsoniana*, while call recorders had  $\hat{q} = 0.34$  and 0.50, respectively. The differences appear nonexistent for *L. pearsoniana*, and comparisons with a null model using AIC indicate that there is no strong evidence for a difference between the methods for ei-



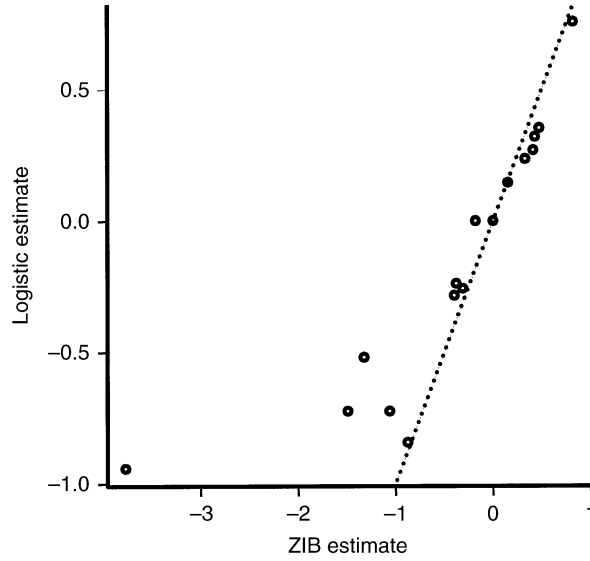


FIG. 5. Estimated coefficient for the year effect from the ZIB (zero-inflated binomial) model and a traditional logistic regression analysis. The dotted line is the one-to-one correspondence. Grey Fantails at  $(-20, -8)$  have been eliminated for clarity.

ther species (Table 3). Both these species of frogs are persistent callers with distinctive advertisement calls; thus, they are more likely to be detected with call recorders than other species in the study area that were often found to be present at a site but not calling (Parris et al. 1999).

*Mound-spring invertebrates*

Natural spring outflows from the Great Artesian Basin in central Australia are home to several species of endemic fish, invertebrates, and plants (e.g., Ponder 1986, Kodric-Brown and Brown 1993). There is anecdotal evidence of local extinctions and recolonizations suggesting the possibility of classical metapopulation dynamics. There are literally hundreds of springs, clustered into groups of a few to hundreds of spring vents. Efforts are currently underway to survey sufficient numbers of these springs to build up a time series of species presence/absence within groups of springs to fit metapopulation models. Detailed monitoring of a small number of springs has revealed large

TABLE 4. Point estimates ( $\hat{q}$ ) and 95% profile confidence limits of  $\hat{q}$  for the three mound-spring endemic crustacean species (Great Artesian Basin, central Australia).

Species	$\hat{q}$	95% CI	
		Lower	Upper
Amphipod	0.66	0.51	0.78
Isopod	0.96	0.88	0.99
Ostracod	0.72	0.63	0.82

fluctuations in abundance of invertebrates from year to year within springs, including apparent disappearances.

These detailed sampling methods are too time consuming for a large-scale survey effort, so a less intensive method has been employed at the spring-group scale to collect the large volumes of data required. This method involves collecting substrate from a range of microhabitats within each spring, and combining them into a single sample jar for processing.

In the 2000 round of surveys, a subset of springs were repeat sampled using the same methodology on the same day to estimate the rate at which single samples would miss individual species given that they were in fact present. This information is critical to interpreting a time series of presence-absence data in a metapopulation context, because false negatives increase the apparent rate of transitions in the data (A. J. Tyre, *unpublished manuscript*). In this case we were not interested in the estimates of  $\hat{\rho}$ , but only in  $\hat{q}$ . Between one and six samples were taken in each spring vent. Preliminary work had indicated false-negative rates were likely to be  $<10\%$ .

The estimates of  $\hat{q}$  and profile confidence intervals for the three crustacean species are given in Table 4. These estimates are considerably lower than expectations based on previous work, especially for the amphipods, and have major implications for survey design. For example, with  $\hat{q} = 0.66$ , the probability of getting a false negative at a site over all surveys is  $(1 - \hat{q})^n$ , and to ensure that this is less than 0.01 requires five repeated surveys at each site.

DISCUSSION

*Interpreting  $\hat{q}$*

The single most important concept underlying this statistical model is the distinction between a survey at

TABLE 3. AIC (Akaike information criteria) model selection results for the Queensland, Australia, rainforest frog data set.

Species	Model†	Log-likelihood	$K‡$	AIC	$\Delta AIC$	Akaike weights
<i>Mixophyes fasciolatus</i>	null	86.41	2	176.83	0	0.6
	method	85.82	3	177.63	0.8	0.4
<i>Litoria pearsoniana</i>	null	78.05	2	160.09	0	0.73
	method	78.04	3	162.07	1.98	0.27

† The null model incorporates no covariates, while the method model has a dummy variable describing which survey method was used.

‡  $K$  = the number of estimated parameters in the model.

an occupied site where the species is not recorded, and an unoccupied site where the species cannot be recorded. In the narrowest sense occupancy means that the species is present inside the survey site during the time the site is being surveyed. This “narrow sense” interpretation of occupancy is true for the mound-spring invertebrates, because their daily movement “ecological neighborhoods” (*sensu* Addicott et al. 1987) are much smaller than the spring wetlands that the surveys cover. It will even be true for some of the bird species sampled, such as Superb Fairy Wrens, which have daily movement neighborhoods similar to the 2-ha area used in the bird survey. For such species,  $q$  is the probability that the survey method detects the species, conditional on the species occupying the site. More cryptic and difficult-to-detect species will have smaller values for  $q$ . This observability will be influenced by many species- and site-dependent factors. For example, the observability of birds will be influenced by the likelihood of vocalization at a particular time of day.

For many other organisms the daily movement neighborhoods are much larger than the survey sites; this is particularly true for birds or larger mammals. For these species even if a survey site lies within the daily movement neighborhood there is no guarantee that the individuals in that area will be within the survey site while it is being sampled. The larger the species neighborhood, or the slower they move around it, the less likely they are to be present within the site when it is sampled. For these species,  $q$  contains an additional component representing this probability of being in the survey site at the time the site is being visited. For these species,  $q$  can be thought of as the probability the survey method records the species in a site known to be “occupied” by a species. Here “occupied” means the site is part of the species home range and is being used at least some of the time within the season the surveys were being conducted, and not that the species was actually on the site when the survey was conducted.

#### *Assumptions*

As with any statistical procedure, we have made a number of assumptions. Some of these are almost certainly violated in some of our examples.

The main assumption is that the “occupancy” of a site does not change during the sequence of repeated surveys. This assumption gets at the heart of what we mean by occupancy vs. observability, and cannot be relaxed. However, it can be readily violated, especially in species that undergo seasonal changes in habitat use or migrations. In the future we would like to examine the possibility of including transitions in occupancy at a site directly in the model, in the fashion of a meta-population model or simple site-occupancy model (e.g., Clark and Rosenzweig 1994).

A second assumption in the current implementation is that  $q$  is consistent at a site between repeated surveys. In effect, this precludes the inclusion of observer effects on  $q$  and temporal variation such as that associated with weather. Site-level effects can be included, and this could include observer effects if the same observer carried out all visits to a site. Weather probably affects observability in both the frog and bird examples above. If  $q$  is variable, it probably has the same effect as extra-binomial variation in a normal logistic regression or binomial GLM, leading to underestimates of the confidence limits on parameters. Incorporating variation in  $q$  between replicate surveys within a site should be straightforward in future implementations, by making the likelihood conditional on the total number of positive observations, but calculating it for each replicate survey independently in the manner of a logistic regression where the maximum number of sites is always 1.

A final assumption, not an atypical one for statistical analyses, is that the outcome of individual repeated surveys must be independent. Normally, this assumption means the absence of spatial and temporal autocorrelations between observations. In the present case, this assumption is subtle, because it is conceivable that an observer might react differently once a species has been detected at the site. For example, once a rare species is recorded at a site, there may be a tendency to listen or look harder for that species in future visits. In some cases this is not a problem. For example, precise species identifications of mound-springs invertebrates are not always possible until the samples have been returned to the laboratory for examination with a microscope. It may well be a problem for bird surveys.

#### *What else can we do with it?*

One extension would be to permit more than two component distributions. The present two components can be thought of as habitat ( $p > 0$ ) and non-habitat ( $p = 0$ ). It is conceivable that there could be one or more intermediate components, poor habitat, with  $p_{\text{good}} > p_{\text{poor}} > 0$ . This would probably require a more sophisticated fitting method able to cope with non-identified components, as well as more data. This is different from merely fitting covariates to  $\hat{p}$ , because it is inevitable that there are unmeasured differences between sites, and fitting component distributions to them would be a way to pick up this variation.

#### *When is it worth the extra effort?*

Nearly all ecological studies sample many fewer sites than is thought desirable, simply due to logistical constraints. Based on the simulations presented in this paper, it appears that up to 6 times the effort may be required to correct for the presence of false-negative errors in presence-absence data. The cost of such an increase in surveying effort is simply enormous, and probably unrealistic on an ongoing basis. The maximum-likelihood estimates of both  $q$  and the logistic

regression coefficients will converge to their true values as  $n$  (the number of sites visited)  $\rightarrow \infty$  when  $m$  (the number of visits to each site) is fixed. In contrast, only the maximum-likelihood estimates of  $q$  converge to their true values as  $m \rightarrow \infty$  when  $n$  is fixed. Therefore it is not entirely clear whether it is better to increase  $m$  at the expense of  $n$ ; the precise nature of the trade-off also depends on the relative costs of adding new sites vs. revisiting sites, on the species- and method-specific values of  $q$ , and on the objectives of the study. Calculating optimum combinations of  $n$  and  $m$  is the subject of our current research.

We can make two clear recommendations for preliminary work to determine the extent to which false negatives are present. First, the minimum number of replicate visits to a site to obtain a useable estimate of  $q$  is three. Obviously this should be applied to as many sites as practicable. As an indication of a good number, we obtained reasonably good estimates of  $q$  in the bird example with 34 sites and three visits to each site. The smaller  $q$  is the more sites will be required to get a good estimate, in the sense of a small confidence limit. However, in preliminary work the goal is mainly to indicate the potential magnitude of the problem, and in this case a precise estimate will not be required if  $q$  is relatively small. It is also more difficult to obtain good estimates of  $q$  when  $p$  (the probability that a species occupies a site) is small, and this is more problematic. If the only purpose of the preliminary work is to obtain estimates of  $q$  to ascertain the potential magnitude of the false-negative problem, then all the sites selected should have the species present. This reduces the problem of estimating  $q$  to the estimation of a simple binomial probability, and will ensure the best possible estimate of  $q$  is obtained with the least effort.

The second recommendation is based on the limited number of scenarios presented in Table 1 and Fig. 2. It appears that when false-negative rates are low (e.g., <50%), it will be better to increase the number of sites rather than the number of visits. As false-negative rates increase, the variance of parameter estimates is reduced more by increasing the number of visits, especially when the number of sites is small.

We have clearly demonstrated that realistic rates of false-negative errors can have a dramatic impact on estimated relationships between habitat variables and the occupancy of a site. In all three of our case studies, we were surprised by the magnitude of false errors revealed by our estimates and repeated surveys. We think it is critical that all studies relying on presence-absence data conduct repeated surveys on at least a subset of sites to obtain an estimate of  $\hat{q}$ ; only then can confident conclusions be drawn from such data.

#### ACKNOWLEDGMENTS

A reviewer pointed out the different nature of convergence for the maximum-likelihood estimates, and both reviewers' comments improved the presentation. This work was supported by an Australian Research Council Strategic Partner-

ships in Industry and Training grant to H.P. Possingham and Western Mining Corporation Ltd.

#### LITERATURE CITED

- Addicott, J. F., J. M. Aho, M. F. Antolin, D. K. Padilla, J. S. Richardson, and D. A. Soluk. 1987. Ecological neighborhoods: scaling environmental patterns. *Oikos* **49**:340–346.
- Austin, M. P., A. O. Nicholls, and C. R. Margules. 1990. Measurement of the realised qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs* **60**:161–177.
- Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York, New York, USA.
- Clark, C. W., and M. L. Rosenzweig. 1994. Extinction and colonization processes: parameter estimates from sporadic surveys. *American Naturalist* **143**:583–596.
- Collett, D. 1991. Modelling binary data. Chapman and Hall, London, UK.
- Craig, M. D., and J. D. Roberts. 2001. Evaluation of the impact of time of day, weather, vegetation density and bird movements on outcomes of area searches for birds in eucalypt forests of south-western Australia. *Wildlife Research* **28**:33–39.
- Field, S. A., H. P. Possingham, and A. J. Tyre. 2002. Estimating bird species richness: How should repeat surveys be organised in time? *Austral Ecology* **27**:624–629.
- Hall, D. B. 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* **56**:1030–1039.
- Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, New Jersey, USA.
- Kodric-Brown, A., and J. A. Brown. 1993. Highly structured fish communities in Australian desert springs. *Ecology* **74**:1847–1855.
- Lindenmayer, D. B., R. B. Cunningham, C. F. Donnelly, R. D. Incoll, M. L. Pope, C. R. Tribolet, K. L. Viggers, and A. H. Welsh. 2001. How effective is spotlighting for detecting the greater glider (*Petauroides volans*)? *Wildlife Research* **28**:105–109.
- Lindenmayer, D. B., R. B. Cunningham, M. T. Tanton, A. P. Smith, and H. A. Nix. 1990. Habitat requirements of the mountain brushtail possum and the greater glider in the montane ash-type eucalypt forests of the central highlands of Victoria. *Australian Wildlife Research* **17**:467–478.
- Lindenmayer, D. B., K. Ritman, R. B. Cunningham, J. D. B. Smith, and D. Horvath. 1995. A method for predicting the spatial distribution of arboreal marsupials. *Wildlife Research* **22**:445–456.
- Loyn, R. 1986. The 20-minute search—a simple method for counting forest birds. *Corella* **10**:58–60.
- Morrison, M. L., B. G. Marcot, and R. W. Mannan. 1992. Wildlife habitat relationships: concepts and applications. University of Wisconsin Press, Madison, Wisconsin, USA.
- Moseby, K. E., and J. L. Read. 2001. Factors affecting pitfall capture rates of small ground vertebrates in arid South Australia. II. Optimum pitfall trapping effort. *Wildlife Research* **28**:61–71.
- Parris, K. M. 2001. Distribution, habitat requirements and conservation of the cascade treefrog (*Litoria pearsoniana*, Anura: Hylidae). *Biological Conservation* **99**:285–292.
- Parris, K. M. 2002. The distribution and habitat requirements of the great barred frog *Mixophyes fasciolatus*. *Wildlife Research*.
- Parris, K. M., and M. A. McCarthy. 1999. What influences the structure of frog assemblages at forest streams? *Australian Journal of Ecology* **24**:495–502.
- Parris, K. M., T. W. Norton, and R. B. Cunningham. 1999. A comparison of techniques for sampling amphibians in

- the forests of south-east Queensland, Australia. *Herpetologica* **55**:271–283.
- Ponder, W. F. 1986. Mound springs of the Great Artesian Basin. Pages 403–420 *in* D. De Deckker and W. D. Williams, editors. *Limnology in Australia*. Dr. W. Junk, Dordrecht, The Netherlands.
- Read, J. L., and K. E. Moseby. 2001. Factors affecting pitfall capture rates of small ground vertebrates in arid South Australia. I. The influence of weather and moon phase on capture rates of reptiles. *Wildlife Research* **28**:53–60.
- Titterton, D. M., A. F. M. Smith, and U. E. Makov. 1985. *Statistical analysis of finite mixture distribution*. John Wiley and Sons, Chichester, UK.
- Tyre, A. J., H. P. Possingham, and D. B. Lindenmayer. 2001. Matching observed pattern with ecological process: Can territory occupancy provide information about life history parameters? *Ecological Applications* **11**: 1722–1738.
- Verner, J. 1985. Assessment of counting techniques. Pages 247–302 *in* R. F. Johnston, editor. *Current ornithology*. Volume 2. Plenum Publishing.
- Westphal, M. I., S. A. Field, A. J. Tyre, D. C. Paton, M. Possingham, and H. P. Possingham. 2003. Effects of landscape pattern on bird species distribution in the Mt. Lofty ranges of SA. *Landscape Ecology*, *in press*.

#### SUPPLEMENT

An R package for fitting zero-inflated binomial (ZIB) models to biological survey data is available in ESA's Electronic Data Archive: *Ecological Archives* A013-025-S1.