

# Modeling abundance using $N$ -mixture models: the importance of considering ecological mechanisms

LIANA N. JOSEPH,<sup>1,3</sup> CHÉ ELKIN,<sup>1</sup> TARA G. MARTIN,<sup>2</sup> AND HUGH P. POSSINGHAM<sup>1</sup>

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

<sup>2</sup>*CSIRO Sustainable Ecosystems, 306 Carmody Road, St Lucia, Queensland 4067 Australia*

**Abstract.** Predicting abundance across a species' distribution is useful for studies of ecology and biodiversity management. Modeling of survey data in relation to environmental variables can be a powerful method for extrapolating abundances across a species' distribution and, consequently, calculating total abundances and ultimately trends. Research in this area has demonstrated that models of abundance are often unstable and produce spurious estimates, and until recently our ability to remove detection error limited the development of accurate models. The  $N$ -mixture model accounts for detection and abundance simultaneously and has been a significant advance in abundance modeling. Case studies that have tested these new models have demonstrated success for some species, but doubt remains over the appropriateness of standard  $N$ -mixture models for many species. Here we develop the  $N$ -mixture model to accommodate zero-inflated data, a common occurrence in ecology, by employing zero-inflated count models. To our knowledge, this is the first application of this method to modeling count data. We use four variants of the  $N$ -mixture model (Poisson, zero-inflated Poisson, negative binomial, and zero-inflated negative binomial) to model abundance, occupancy (zero-inflated models only) and detection probability of six birds in South Australia. We assess models by their statistical fit and the ecological realism of the parameter estimates. Specifically, we assess the statistical fit with AIC and assess the ecological realism by comparing the parameter estimates with expected values derived from literature, ecological theory, and expert opinion. We demonstrate that, despite being frequently ranked the "best model" according to AIC, the negative binomial variants of the  $N$ -mixture often produce ecologically unrealistic parameter estimates. The zero-inflated Poisson variant is preferable to the negative binomial variants of the  $N$ -mixture, as it models an ecological mechanism rather than a statistical phenomenon and generates reasonable parameter estimates. Our results emphasize the need to include ecological reasoning when choosing appropriate models and highlight the dangers of modeling statistical properties of the data. We demonstrate that, to obtain ecologically realistic estimates of abundance, occupancy and detection probability, it is essential to understand the sources of variation in the data and then use this information to choose appropriate error distributions.

**Key words:** *AIC; ecological realism; excess zeros; model choice; negative binomial; overdispersion; Poisson regression; South Australian birds; zero-inflated Poisson; zero inflation.*

## INTRODUCTION

Estimating the abundance of a species across its distribution is a valuable tool for biodiversity management and ecological studies. For example, abundance estimates may be used to inform conservation planning (Gaston and Rodrigues 2003) and to investigate ecological questions such as understanding the processes that drive population size (e.g., Meents et al. 1983) and source–sink dynamics (e.g., Pulliam 1988, Tittler et al. 2006). The logistical constraints of collecting detailed abundance data across a species distribution have led to the development of a range of predictive techniques for

modeling abundance. Often these techniques model the relationship between survey data and environmental variables to obtain predictions of site and total abundances (Pearce and Ferrier 2001). However, predictions of abundances across species distributions are often unsuccessful (Pearce and Ferrier 2001). Successful development of robust modeling techniques is important, for both ecological and conservation applications.

Identifying and modeling the sources of variation in data are key components of successfully predicting abundance. Variation in the number of individuals recorded in count data at each site will often be the product of both ecological processes (true variation) and sampling error (false variation [Martin et al. 2005b]). True variation in abundance results from ecological mechanisms such as the reaction of a species to environmental gradients, metapopulation dynamics,

Manuscript received 21 December 2007; revised 16 June 2008; accepted 30 June 2008. Corresponding Editor: D. F. Parkhurst.

<sup>3</sup> E-mail: l.joseph@uq.edu.au

interactions with other species (e.g., competition or predation) or by chance (i.e., environmental and demographic stochasticity [Martin et al. 2005b]). Conversely, false variation in count data results from incorrect identification and imperfect detection, i.e., a surveyor miscounting or failing to observe an individual that occupies the survey site. Failing to observe a species may occur in two ways: the species may be present but the surveyor failed to observe it or the species utilizes the site but was not present at the time of the survey (Tyre et al. 2003, Martin et al. 2005b). A mobile species may often be absent from a site that it utilizes when it is visiting a proportion of its home range that lies outside the survey site. Other species (e.g., perennial plants or fungi) may not be visible during the survey. A model of abundance needs to simultaneously model true variation and false variation in a manner that accurately represents underlying ecological mechanisms and observation error.

#### *Modeling true variation*

Traditionally, site abundance has been estimated by modeling variation that arises through environmental gradients while ignoring the other forms of true variation and detection error (e.g., Sánchez-Zapata and Calvo 1999). Poisson regression is a popular approach for modeling the true variation in count data (see Nelder and Wedderburn 1972, McCullagh and Nelder 1989) as it assumes that variation can be described with environmental variables and the error is Poisson distributed. However, if extra-Poisson variation exists due to a deviation from randomness these models will be inappropriate. Non-randomness may be caused by the omission of relevant explanatory variables, demographic and environmental stochasticity or nonindependence of individuals as may occur when organisms are observed in pairs or groups.

Abundance data sets often have a large number of zeros that result from surveying unoccupied sites. If a data set contains more zeros than expected from a standard distribution then these data are referred to as zero inflated. Zero inflation is a special case of extra-Poisson true variation and often precludes the use of Poisson regression models. When a data set contains extra-Poisson variation due to zero inflation or other problems mentioned above, a negative binomial distribution is often used (e.g., Lee et al. 2002, Kuhnert et al. 2005). The negative binomial accounts for extra-Poisson variation by allowing the mean abundance estimate to vary stochastically. However, while a negative binomial can accommodate a limited amount of zero inflation, a high number of zeros can exceed its capabilities (Welsh et al. 1996, Hall 2000).

Astute sampling of only the sites that are likely to be occupied may reduce zero-inflation (Austin and Meyers 1996). However, this technique may be impractical in many situations such as when surveys are designed for multiple species or we don't know what sites are likely to

be occupied. Instead, the number of unoccupied sites may be reduced after the data have been collected, by truncating the data set. This process may inadvertently remove zeros that arise through other processes, such as detection error, and may compromise parameter estimation. Additionally, this process removes information about the species' probability of occupancy and makes it impossible to examine environmental variables that determine presence. It may be preferable to model occupancy and abundance simultaneously. Zero-inflated Poisson (ZIP) or zero-inflated negative binomial (ZINB) regression methods (see Lambert 1992, Zorn 1996, Ridout et al. 1998, Barry and Welsh 2002, Kuhnert et al. 2005, Martin et al. 2005a, b) are modeling techniques that allow for occupancy and abundance to be estimated simultaneously. These models predict the areas where individuals are absent and estimate abundance when they are present without the need to truncate the data.

#### *Modeling false variation*

Standard count regressions ignore false variation (i.e., detection error). These models produce an index of abundance that is assumed to be related to true abundance through a proportionality constant (Pollock et al. 2002). Because the proportionality constant remains fixed, the assumption is made that detection error is constant in space and time. This assumption is universally untrue as the proportionality constant and detection error vary spatially and temporally with habitat, time of day and observer, among other things. If variation in detection error is ignored, the abundance prediction will be incorrect. This will result in the model having poor predictive abilities and will compromise scientific inference.

Successfully accounting for detection error has been a limiting stage in the development of adequate Poisson regression models (Pearce and Ferrier 2001). There has been much interest in dealing with detectability for presence-absence data (MacKenzie et al. 2002, 2003, Tyre et al. 2003), but until recently no practical techniques were available to simultaneously model true variation and false variation in abundance data (Anderson 2001, Rosenstock et al. 2002). The *N*-mixture regression method uses spatially and temporally replicated count data to model abundance and detection error simultaneously (Royle 2004, Royle et al. 2005). It is a mixture between the binomial distribution and a standard count model (Poisson or negative binomial, see *Methods* [Royle 2004, Royle et al. 2005]). The binomial distribution and explanatory variables (e.g., time of day or season) are employed to estimate detection probability. The binomial distribution requires an estimate for the actual number of individuals; this is modeled as a random variable from a count model such as the Poisson or the negative binomial. Simulations (Royle 2004) and empirical studies (Dodd and Dorazio 2004, Kéry et al. 2005, Royle et al. 2005) have demonstrated that this model is robust for many scenarios.

The  $N$ -mixture model can be a mixture of the binomial and any other distribution used to model count data. The  $N$ -mixture with a Poisson distribution is an appropriate model to use to estimate abundance and detection probability simultaneously when the count data are randomly distributed. In other cases, the  $N$ -mixture negative binomial may be appropriate. The negative binomial form is only useful when the true variation arises from the omission of covariates and a limited proportion of excess zeros; it is not useful when there is a violation of the independence assumption (i.e., IID; Doob 1934) or when data are greatly zero-inflated. Non-independence of data may occur when individuals are clumped (e.g., flocks of birds). Because the independence assumption is made in both the binomial and the count model components of the  $N$ -mixture model, simply replacing the Poisson with a model that can accommodate dependent data will not suffice. When data are not independent, the  $N$ -mixture model, in the forms that have been described to date, is not appropriate. In the case where the data are zero inflated, the negative binomial may be less appropriate than a zero-inflated Poisson model.

In this paper, we develop Royle's (2004)  $N$ -mixture model to accommodate zero-inflated data by using zero-inflated count models (zero-inflated Poisson and zero-inflated negative binomial). To our knowledge, this is the first application of this approach to modeling abundance. The methodology presented accomplishes three goals: (1) it removes error due to false variation (detection error), (2) it predicts whether a species will occur in an area and, if the species does occur, (3) it provides an accurate abundance estimate through modeling the true variation that results from biological responses to ecological gradients. We employ four variants of the  $N$ -mixture model: Poisson, zero-inflated Poisson, negative binomial, and zero-inflated negative binomial. We compare models using the Akaike information criterion (AIC) and assess their ecological realism by comparing the parameter estimates with expected values derived from literature, ecological theory and expert opinion. We present a novel interpretation of the results obtained from the  $N$  mixtures. Specifically, we explain that the site abundance predicted with the  $N$ -mixture model is an estimate of the number of individuals that utilize a site, not the true site density, and is more appropriately referred to as a relative density estimate. We provide examples of the implementation of the novel combination of modeling methods and the new interpretation of parameter estimates with survey data of the birds of the Mount Lofty ranges in South Australia.

#### METHODS

We conducted a survey of woodland birds in the Mount Lofty Ranges (34°58' S, 138°42' E), an area of national conservation significance in South Australia, during the spring and summer of 2000 (25 August 2000 to

19 December 2000). The Mount Lofty Ranges lie in a Mediterranean climatic region that consists of eucalypt woodlands and low shrublands. There are two main vegetation types: "stringybark" forest (dominated by *Eucalyptus obliqua* and *E. baxteri*) and "gum" woodland (dominated by *E. leucoxylon* and/or *E. fasciculosa*). We surveyed 109 sites and visited each nine times; 48 of the sites (44%) were located in stringybark woodland and the other 61 sites (56%) were in gum woodland. We used active, time-area searches (e.g., 20 minutes in 2-ha plots [Loyn 1986]). Individual birds were recorded only if they were seen or heard within the 2-ha plot. For statistical purposes it was necessary to assume that the populations were closed (i.e., no birth, death, immigration, or emigration) over the survey period. We believe this assumption to be reasonable, as surveys were made during the first half of the breeding season (Possingham and Possingham 2000). Six bird species (out of 111 identified species) were selected to encompass a range of abundances, detection probabilities and site occupancies: the White-throated Treecreeper (*Cormobates leucophaeus*), Gray Fantail (*Rhipidura fuliginosa*), Mistletoebird (*Dicaeum hirundinaceum*), Rufous Whistler (*Pachycephala rufiventris*), Golden Whistler (*Pachycephala pectoralis*), and Scarlet Robin (*Petroica multicolor*).

#### Statistical model

Four variants of Royle et al.'s  $N$  mixture—the Poisson (P), zero-inflated Poisson (ZIP), negative binomial (NB), and zero-inflated negative binomial (ZINB)—are employed to simultaneously model detection probability, abundance and, in the case of ZIP and ZINB, probability of occupancy, of the six bird species.

Environmental covariates are used to inform the abundance estimates, detection error estimates and probability of occupancy estimates (only for zero-inflated forms). The distribution that models each parameter can utilize unique covariates (e.g., vegetation type, time of day of survey) that can be selected based on the mechanism being modeled (i.e., observer detection error, likelihood of occurrence and abundance). Below, we present the  $N$ -mixture model with three main components—(1) the estimation of detection probability, (2) abundance, and (3) probability of occupancy—and highlight statistical issues associated with each.

#### Estimation of detection probability

The  $N$ -mixture models the number of observed individuals,  $n_{it}$ , recorded at  $i = 1, 2, \dots, R$  locations during  $t = 1, 2, \dots, T$  sampling occasions. Counts,  $n_{it}$ , are binomial random variables with index  $N_i$  (the actual number of individuals present during the survey) and  $p$  (the probability of detecting each individual). The likelihood for the data from site  $i$  is

$$L(N_i, p | \{n_{i1}, \dots, n_{iT}\}) = \prod_{t=1}^T \text{Bin}(n_{it}; N_i, p) \quad (1)$$

where

$$\text{Bin}(n_{it}; N_i, p) = \frac{N_i!}{n_{it}!(N_i - n_{it})!} p^{n_{it}} (1 - p)^{N_i - n_{it}} \quad (2)$$

The joint likelihood of the data from replicate samples at  $R$  locations in space is the product of all the site binomials (Eq. 1):

$$L(\{N_i\}, p | \{n_{it}\}) = \prod_{i=1}^R \left[ \prod_{t=1}^T \text{Bin}(n_{it}; N_i, p) \right] \quad (3)$$

conditional on  $\{N_i\} = (N_1, N_2, \dots, N_R)$  and detection probability  $p$ . The actual number of individuals,  $N_i$ , in a survey period has a distribution  $f(N_i; \theta)$ . The prior distribution on  $N$  can be any distribution suitable to model count data, typically a Poisson or a negative binomial. The integrated likelihood is

$$L(p, \theta | \{n_{it}\}) = \prod_{i=1}^R \left\{ \sum_{N_i=\max_i n_{it}}^{\infty} \left[ \prod_{t=1}^T \text{Bin}(n_{it}; N_i, p) \right] f(N_i; \theta) \right\} \quad (4)$$

where  $\theta$  is the parameter(s) of  $f(N_i; \theta)$  (e.g.,  $\theta = \lambda$  for the Poisson model).

Detection probability,  $p$ , can depend on covariates such as time of day, season, or observer. This variation can be modeled using standard generalized linear regression techniques. In this case, the response variable is binary (individual bird is detected or not detected). Hence, the expected value may be modeled using the logit link function (McCullagh and Nelder 1989). The logistic regression is

$$\text{logit}(p_{it}) = \alpha_0 + \sum_j \alpha_j x_{it} \quad (5)$$

where  $p_{it}$  is the probability that the individual will be detected at site  $i$  at time  $t$ ,  $\alpha_0$  is the intercept coefficient, the  $x_i$  are the predictor variables, and  $\alpha_j$  is the predictor coefficient for the  $j$ th predictor.

#### Estimation of abundance

Abundance is commonly modeled using either the Poisson or the negative binomial distribution. The Poisson distribution is a natural choice for modeling count data because it assumes that events occur at random in space. If the discrete random variable,  $N$  (the number of individuals at the site), is Poisson distributed, it has a density

$$\Pr(N = x | \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, \dots, \infty \quad (6)$$

where the parameter lambda,  $\lambda$ , is the mean number of events occurring in a unit area or, in this case, the number of individuals present at the site at the time of the survey.

The negative binomial accounts for deviation from randomness by allowing the mean  $\mu$  (analogous to the  $\lambda$  of the Poisson distribution) to vary stochastically

through the inclusion of an explicit dispersion parameter,  $\delta$ :

$$\Pr(N = x | \mu, \delta) = \frac{(x + \delta - 1)!}{(\delta - 1)! x!} \left( \frac{\delta}{\delta + \mu} \right)^{-\delta} \left( \frac{\mu}{\mu + \delta} \right)^x, \quad x = 0, 1, 2, \dots, \infty \quad (7)$$

where  $N$  is the number of individuals present at the site at the time of the survey.

Environmental covariates that influence abundance are incorporated into the abundance models by again using generalized linear regression. As the Poisson or negative binomial error distribution is used for the mean site abundance, a log-linear transformation is appropriate:

$$\log(\theta_i) = \gamma_0 + \sum_j \gamma_j x_i \quad (8)$$

where  $\theta_i = \lambda_i$  or  $\mu_i$ , is the mean abundance at site  $i$  for the Poisson and negative binomial distribution, respectively,  $\gamma_0$  is the intercept coefficient, the  $x_i$  are the predictor variables, and  $\gamma_j$  is the predictor coefficient for the  $j$ th predictor.

#### Estimation of probability of occupancy

While Poisson and negative binomial models may do a good job of modeling abundance at sites that are occupied, they do not adequately model a large number of true zeros that will arise from sampling unoccupied sites. These zeros can be incorporated into the abundance models by estimating the probability of occupancy through the use of zero-inflated forms of the abundance models. There are two types of zero-inflated count models used in the ecological literature: the conditional models (Mullahy 1986, King 1989) and the mixture models (Lambert 1992). They are both a combination of a Bernoulli process (to determine occupancy) and a Poisson or negative binomial process (to determine the site abundance). However, in the conditional model, the abundance distribution in occupied sites is truncated at zero, whereas in the mixture model it is not (for a more detailed description of the differences between the models, see Zorn 1996). A truncated-at-zero distribution requires the assumption that all the zeros in the data arise from the process driving occupancy and that none of the zeros arise from the Poisson or negative binomial process that drives site abundance. This would imply that non-zero abundance is a certainty in suitable sites and zero abundance could not occur by chance. As this is unlikely, the mixture model, which describes some zeros as part of the Poisson distribution and others as the point mass, is more appropriate for describing the distribution of site abundance data.

In the case of the Poisson mixture, the probability density function is as follows:

$$\Pr(N_i = 0|\lambda, \psi) = \psi + (1 - \psi)e^{-\lambda}$$

$$\Pr(N_i > 0|\lambda, \psi) = (1 - \psi) \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots, \infty$$
(9)

where *N* is the actual site abundance,  $\psi$  is the probability that the site is occupied, and  $\lambda$  is the mean abundance (Lambert 1992, Martin et al. 2005a).

Similarly, the mixture negative binomial probability density function is

$$\Pr(N_i = 0|\mu, \delta, \psi) = \psi + (1 - \psi) \left( \frac{1}{1 + \mu/\delta} \right)^\delta$$

$$\Pr(N_i > 0|\mu, \delta, \psi) = (1 - \psi) \frac{(x + \delta - 1)!}{x!(\delta - 1)!} \frac{(\mu/\delta)^x}{(1 + \mu/\delta)^x (1 + \mu/\delta)^\delta}$$

$$x = 1, 2, \dots, \infty$$
(10)

where *N* is the actual site abundance,  $\psi$  is the probability that the site is occupied,  $\mu$  is the mean abundance, and  $\delta$  is the dispersion parameter of the negative binomial.

Environmental covariates that affect occupancy were included in the probability of occupancy component of these models. The covariates that influence occupancy may or may not be the same covariates that are incorporated into the abundance component of the model. The probability that the site is occupied ( $\psi$ ) is a binary process (occupied or unoccupied); hence, a logistic regression is appropriate to predict occupancy:

$$\text{logit}(\psi_i) = \beta_0 + \sum_j \beta_j x_i$$
(11)

where  $\beta_0$  is the intercept coefficient, the  $x_i$  are the predictor variables, and  $\beta_j$  is the predictor coefficient for the *j*th predictor.

*Covariates*

We selected five covariates to include in the models: rainfall, solar radiation, area of vegetation cover within a 5-km buffer, Julian day of survey, and time of day of survey. These five covariates were chosen based on expert opinion and a preliminary model-fitting exercise where we assessed different covariate combinations using the Akaike information criteria (AIC, Burnham and Anderson 1998). We used two covariates to model detectability (Julian day of year and time of day) and the same three covariates for each of the abundance and occupancy components of the model (rainfall, solar radiation, 5-km buffer). In addition to running the models with the covariates, we also conducted null model runs in which no covariates were used.

*Model fit and selection using AIC*

The models were fit by minimizing the negative log likelihood using the minimization procedure (fminunc.m)

TABLE 1. Estimates of ecologically realistic abundance, probability of detection, and probability of occupancy for six bird species of the Mount Lofty Ranges, South Australia derived from published surveys, expert opinion, and ecological theory.

Species	Abundance (no. birds/ha <sup>2</sup> )	Probability of detection	Probability of occupancy
Rufous Whistler	4	>0.1	0.56
Mistletoebird	4	>0.1	0.56
Scarlet Robin	6	>0.1	0.5
White-throated Treetreeper	6	>0.1	0.9
Gray Fantail	10.5	>0.1	1
Golden Whistler	2.6	>0.1	0.9

*Notes:* Abundance is the maximum abundance reported for each species in the literature. The probability of detecting each species is suspected to be at least once in 10 surveys. Probability of occupancy was estimated from the known distributions of vegetation types and the expected vegetation preference of each species.

in the software package MATLAB (Version 7.0.1; The MathWorks, Natick Massachusetts, USA). To enhance convergence of the numerical optimization algorithm, all covariates were transformed into standard normal deviates by first subtracting the arithmetic mean and then dividing by the standard deviation. Following standard maximum likelihood theory, the asymptotic variance-covariance matrix of the parameter estimates was derived by finding the inverse of the Hessian matrix (i.e., the observed Fisher Information Matrix). Standard errors were calculated as the square root of the diagonal of the covariance matrix (De Groot and Schervish 2002). Ninety-five percent confidence intervals (CI) were obtained using the standard method (i.e., mean  $\pm$  1.96 SE).

The best model for each species was selected by comparing the AIC values (Burnham and Anderson 1998). Models with the smallest AIC value are considered the best model. However, it is insufficient to simply select the model with the lowest AIC value because values of AIC may vary due to model uncertainty. Instead, using the difference between model AIC and the minimum AIC ( $\Delta$ ) is recommended:

$$\Delta_i = \text{AIC}_i - \min \text{AIC}$$
(12)

Models having  $\Delta_i \leq 2$  can be considered equally superior (Burnham and Anderson 1998).

*Assessment of ecological realism*

For each species, we predicted the abundances, detection probabilities, and site occupancies using estimates derived from the literature and expert opinion (Table 1). To assess ecological realism of model results, we compared the parameters estimated by the best statistical models (as determined by AIC) to these predictions. The predicted estimates are approximations only and, hence, it is not statistically sensible to perform standard statistical comparative tests. Instead, we used the predictions as a gauge and employed threshold

criteria to assess the ecological realism of the model estimates. Only models that passed the thresholds for all three parameters were deemed ecologically realistic.

We determined a plausible abundance from our models by selecting the maximum density reported by Australian surveys (mean number of studies = 14.6, CV = 0.4) published in the Handbook of Australian, New Zealand and Antarctic Birds (Higgins et al. 2001, 2006, Higgins and Peter 2002). We selected the maximum value, as this provided an upper estimate to compare to the modeled estimates. We converted the published density (birds/ha) into an approximate estimate of the number of birds in 2 ha (i.e., the units that are predicted by the statistical models) by doubling the values. The abundance estimates produced by the *N*-mixture models are estimates of the number of species that utilize a site rather than true density (see *Discussion* for further explanation). Therefore, the parameter estimates may be slightly greater than the predictions from the literature. Also, because the density estimates from the literature were for different regions (sites located throughout Australia), we expect some inconsistency with modeled estimates. However, despite these caveats, it is possible to assume that considerably larger estimates would not be ecologically plausible. Hence, we assume that the density estimated with the statistical models should not be more than 10 birds/2 ha greater than the prediction.

Obtaining a priori estimates of the probability of detecting a species is possible though eliciting expert knowledge. Because estimates provided by experts are approximate, we asked a group of experts to provide a lower bound on the probability of detecting an individual if it was present in the survey rather than a point estimate. The experts believe that the probability of detection should be greater than 0.1 for all species (i.e., an individual that is present in the survey site will be detected by an observer during the survey period at least once in 10 surveys). For the species considered here, estimates of detection probability that are less than 0.1 are assumed to be ecologically unlikely.

The predictions of probability of occupancy of each species were based on the proportions of survey sites located in each of the vegetation types and prior knowledge of vegetation preferences of each of the species. Forty-eight (44%) survey sites were located in stringybark forest and 61 (56%) were located in gum woodlands. Two of the species, Mistletoebird and Rufous Whistler, have a well-known preference for gum woodlands. Given that surveys were conducted in both stringybark forests and gum woodlands, we expected the count data of these two species to be zero inflated. The Scarlet Robin has declined significantly from much of its former range in the Mount Lofty Ranges (Ford and Howe 1980, Possingham et al. 2004). Therefore, we expected that this species was likely to be absent from greater than 50% of the sites (i.e., 54 sites). The other three species were found throughout the study region, hence were unlikely to have excess zeros (true zeros) that would arise through unoccupied sites. It is

possible for sites to be unoccupied for reasons other than vegetation preference; hence, it is realistic to have more zeros than predicted based on proportion of sites in suitable vegetation types. However, it is extremely unlikely to have fewer zeros than predicted. Therefore, we assumed the models were not ecologically realistic if the number of zero-counts estimated by the statistical models were considerably fewer than predicted (i.e., >10 sites fewer).

## RESULTS

### *Model selection using AIC*

Based on the comparison of AIC values, models that incorporated covariates always produced better fits than the null models. Covariate models had AIC values that were 4 to 93 points less than the comparable null models, with most of the best covariate models having AIC values that were at least 20 points less. For the remainder, we present only the results from the models that used covariates. We graph the estimates for mean abundance per site for (1) the species that were likely to be absent in many sites and, therefore, have zero-inflated data (Fig. 1) and (2) the species that were likely to occur in most sites (Fig. 2).

AIC values and the parameter estimates (mean site abundance, probability of detection, mean probability of occupancy and dispersion parameter,  $\delta$ , are presented for each of the *N*-mixture models (P, ZIP, NB, and ZINB) in Table 2. According to AIC, the zero-inflated negative binomial model was the most parsimonious model for all six species. For two species (the White-throated Treecreeper and the Golden Whistler), one other model was equally as parsimonious (the zero-inflated Poisson and the negative binomial, respectively).

### *Assessment of ecological realism*

The Poisson and the zero-inflated Poisson variant of the *N*-mixture generated ecologically realistic parameter estimates for five of the six species (Table 3). Estimates from both of these models were unrealistic for the scarlet robin. The Poisson model estimated a probability of occupancy that was much larger than expected (i.e., 1 instead of <0.50). The ZIP estimated probability of detection that was too low (i.e., 0.06 instead of >0.1). In contrast, the estimates generated by both the negative binomial and zero-inflated negative binomial were realistic for only two species. Specifically, realistic estimates were generated for species having data unlikely to be zero inflated. The negative binomial models performed poorly at estimating abundance and detection error for most of the species; however, they successfully predicted occupancy for most species.

## DISCUSSION

### *Inconsistency between AIC and ecological realism*

Ecologically unreasonable predictions of abundance and detection probabilities were obtained from the models that were ranked as the best models based on

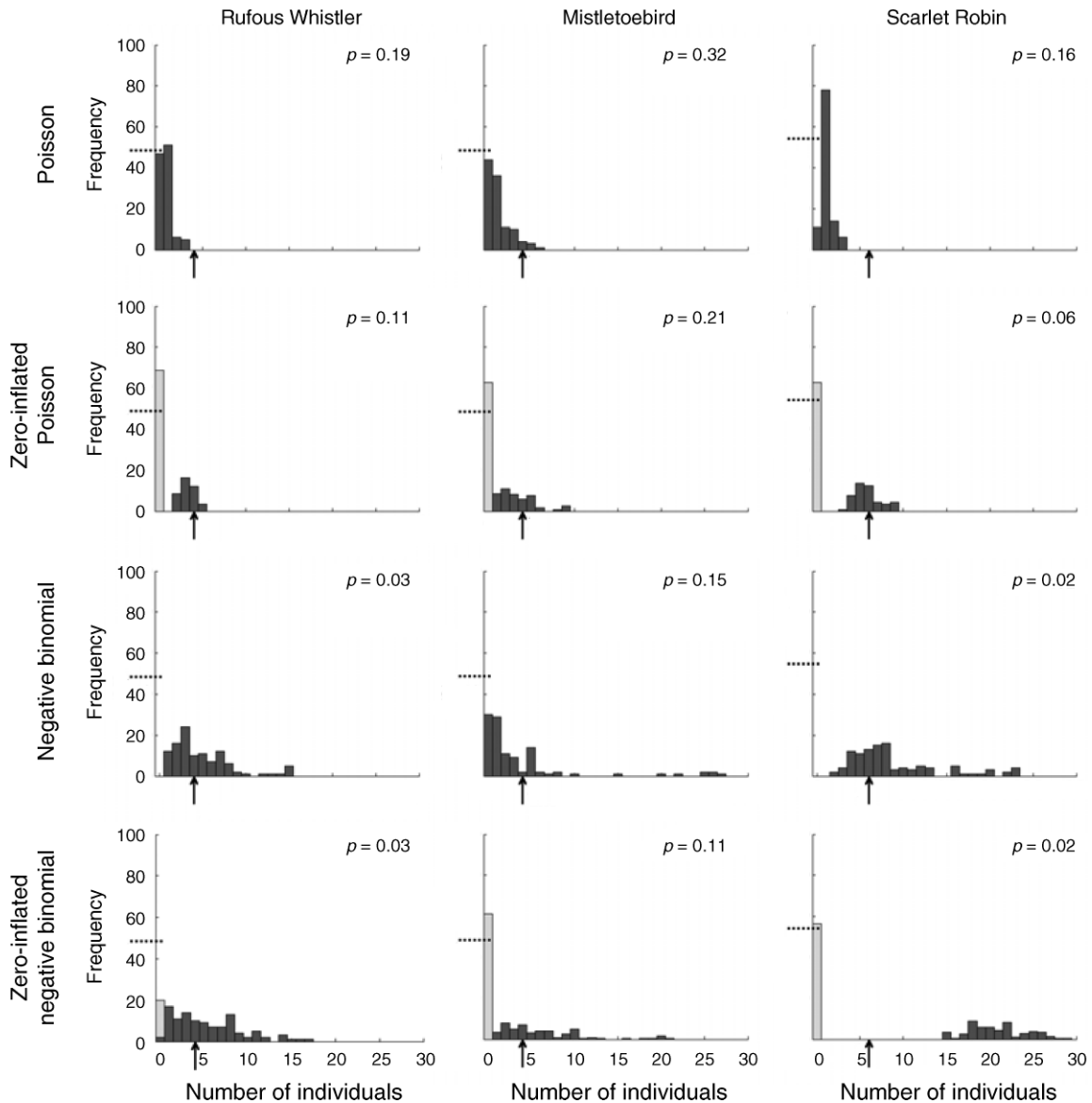


FIG. 1. Histograms of the modeled actual abundance predicted by the four *N*-mixture count models for the species with known vegetation preferences or low levels of occupancy and, consequently, zero-inflated data. The abundance predicted with the binomial component of the zero-inflated models ( $N = 0$ ) is presented as pale gray bars, whereas the abundance predicted by the Poisson or negative binomial component of the zero-inflated model ( $N = 0, 1, \dots, \infty$ ) is presented as dark gray bars. Dotted lines on the *y*-axes indicate the approximate number of zero counts (i.e., unoccupied sites) that are expected given vegetation preferences (i.e., ecologically realistic estimates). Arrows on the *x*-axes point to approximate, ecologically realistic abundance derived from published surveys. The probability of detection, *p*, is stated for each species.

the Akaike information criteria. The zero-inflated negative binomial variant of the *N*-mixture model was always the best model according to the AIC (although for two of the species either the zero-inflated Poisson or the negative binomials were indistinguishable). The negative binomial variant was often ranked second best according to AIC. However, parameter estimates for both of the negative binomial variants were ecologically unreasonable for all three of the birds that were absent from a large proportion of the sites (i.e., Rufous Whistler, Mistletoebird, and Scarlet Robin) and one

that was expected to be almost everywhere (i.e., Golden Whistler). For these species, the negative binomial variants predicted site abundances that were often higher (e.g., 20 birds/2 ha) and detection probabilities that were appreciably lower (e.g., 0.02) than reasonable. These models suggest that the sites are frequented by a large number of birds, all with a low detection probability. This behavior is very unlikely for these species, especially at the time of year when the surveys were conducted (i.e., during the breeding season, when discrete territories are maintained).

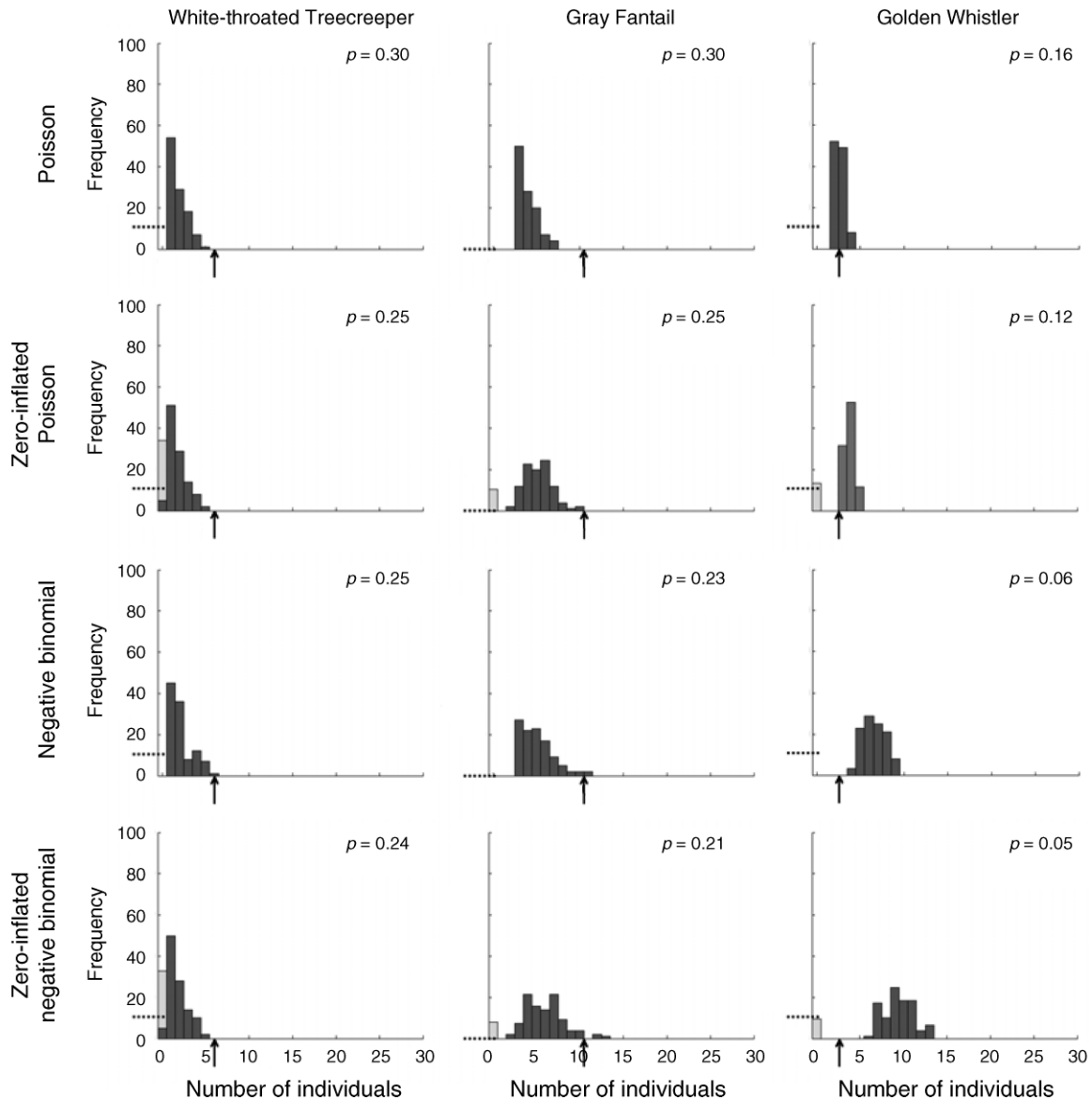


FIG. 2. Histograms of the modeled actual abundance predicted by the four  $N$ -mixture count models for species that are thought to utilize all vegetation types and, hence, for which the data are unlikely to be zero-inflated. The abundance predicted with the binomial component of the zero-inflated models ( $N = 0$ ) is presented as pale gray bars, whereas the abundance predicted by the Poisson or negative binomial component of the zero-inflated model ( $N = 0, 1, \dots, \infty$ ) is presented as dark gray bars. Dotted lines on the  $y$ -axes indicate the approximate number of zero counts (i.e., unoccupied sites) that are expected given known vegetation preferences (i.e., ecologically realistic values). Arrows on the  $x$ -axes point to approximate, ecologically realistic abundance derived from published surveys. The probability of detection,  $p$ , is stated for each species.

In contrast, for two of the three species that were expected to occupy most sites (White-throated Trecreeper and Gray Fantail), the two forms of the negative binomial  $N$  mixture generated reasonable parameter estimates. The fact that the negative binomial variants of the  $N$ -mixture models perform poorly when data are zero-inflated but perform better when excess zeros were absent suggests that these variants are unable to successfully model extra-Poisson variation due to excess true zeros.

Importantly, the Poisson and ZIP variant of the  $N$ -mixture model generated ecologically reasonable parameter estimates for five of the six species. Given that the Poisson variant is unlikely to appropriately model zero-inflated data, it is surprising that this model generated good estimates for two of the species that were expected to have zero-inflated data (i.e., Rufous Whistler and Mistletoebird). This is likely to occur because a Poisson model provides a good approximation of the true values when the true maximum



TABLE 2. *N*-mixture model results for six bird species of the Mount Lofty Ranges using Poisson (P), zero-inflated Poisson (ZIP), negative binomial (NB), and zero-inflated negative binomial (ZINB) distributions.

Bird species and distribution	AIC	Abundance, $\lambda$	Probability of detection, $d$	Probability of occupancy, $\psi$	Delta, $\delta$
<b>Rufous Whistler</b>					
P	730.34	0.79 (0.70–0.95)	0.19 (0.15–0.25)		
ZIP	706.21	3.30 (1.83–6.31)	0.11 (0.06–0.19)	0.37 (0.30–0.48)	
NB	699.90	4.85 (0.52–51.17)	0.03 (0.00–0.29)		0.30 (0.16–0.56)
<b>ZINB</b>	<b>693.38</b>	<b>6.33 (0.95–60.63)</b>	<b>0.03 (0.00–0.21)</b>	<b>0.83 (0.20–0.80)</b>	<b>0.39 (0.21–0.73)</b>
<b>Mistletoebird</b>					
P	1148.90	1.25 (1.08–1.63)	0.32 (0.27–0.37)		
ZIP	1099.98	3.46 (2.64–4.87)	0.21 (0.15–0.28)	0.43 (0.36–0.53)	
NB	1104.54	3.51 (3.60–3.89)	0.15 (0.08–0.27)		0.06 (0.34–1.07)
<b>ZINB</b>	<b>1094.76</b>	<b>6.14 (2.84–15.77)</b>	<b>0.11 (0.04–0.29)</b>	<b>0.44 (0.37–0.56)</b>	<b>3.42 (1.15–10.79)</b>
<b>Scarlet Robin</b>					
P	869.80	1.09 (0.70–1.87)	0.16 (0.12–0.21)		
ZIP	831.72	5.84 (2.44–15.17)	0.06 (0.03–0.14)	0.43 (0.31–0.53)	
NB	826.25	8.87 (3.11–31.49)	0.02 (0.01–0.04)		0.36 (0.22–0.58)
<b>ZINB</b>	<b>823.89</b>	<b>20.79 (8.36–55.92)</b>	<b>0.02 (0.01–0.04)</b>	<b>0.48 (0.31–0.61)</b>	<b>1.65 (0.49–5.52)</b>
<b>White-throated Treecreeper</b>					
P	1532.45	1.78 (1.33–2.43)	0.30 (0.25–0.35)		
<b>ZIP</b>	<b>1518.69</b>	<b>2.75 (1.96–4.10)</b>	<b>0.25 (0.19–0.31)</b>	<b>0.73 (0.63–0.78)</b>	
NB	1525.97	2.11 (1.40–3.30)	0.25 (0.20–0.32)		2.65 (1.07–6.61)
<b>ZINB</b>	<b>1520.27</b>	<b>2.82 (1.93–4.44)</b>	<b>0.24 (0.18–0.31)</b>	<b>0.74 (0.62–0.79)</b>	<b>15.71 (0.55–447.69)</b>
<b>Gray Fantail</b>					
P	2484.73	3.88 (3.43–4.43)	0.30 (0.26–0.34)		
ZIP	2466.85	5.17 (4.38–6.27)	0.25 (0.20–0.30)	0.90 (0.77–0.96)	
NB	2469.40	5.09 (4.15–6.49)	0.23 (0.17–0.30)		3.76 (1.93–7.31)
<b>ZINB</b>	<b>2463.76</b>	<b>5.97 (4.59–8.06)</b>	<b>0.21 (0.15–0.29)</b>	<b>0.93 (0.83–0.93)</b>	<b>8.15 (2.89–23.01)</b>
<b>Golden Whistler</b>					
P	1573.44	2.62 (2.09–3.36)	0.16 (0.12–0.20)		
ZIP	1569.11	3.82 (2.53–5.88)	0.12 (0.08–0.18)	0.88 (0.75–0.94)	
<b>NB</b>	<b>1563.52</b>	<b>6.55 (1.81–24.37)</b>	<b>0.06 (0.02–0.22)</b>		<b>2.03 (0.97–4.27)</b>
<b>ZINB</b>	<b>1564.19</b>	<b>9.42 (1.35–67.32)</b>	<b>0.05 (0.01–0.28)</b>	<b>0.91 (0.69–0.98)</b>	<b>2.88 (1.06–7.81)</b>

Note: The best models (according to AIC) are indicated in bold; 95% confidence intervals appear in parentheses.

abundance is relatively close to zero (i.e., mean of approximately five) and the number of sites with a count of zero is relatively high (i.e., approximately 40 sites are unoccupied). Imagine site abundance displayed in a histogram similar to those presented in Fig. 1 and Fig. 2. Imagine that the number of unoccupied sites (e.g., abundance = 0) is reasonably high (e.g., 40–50) and the mean abundance of all occupied sites is quite low (e.g., ~5). A Poisson model will fit this distribution well. The Poisson will slightly underestimate the true abundance; however, the underestimated parameters remain ecologically reasonable. Conversely, if the number of unoccupied sites is reasonably high but the mean abundance of occupied sites is further from zero (e.g., ~10), the Poisson will not fit the data as well. In this case, the Poisson model will generate parameter estimates that are not reasonable.

For one species (the Scarlet Robin), the ZIP variant of the *N*-mixture model estimated the probability of occupancy to be less than expected (i.e., 0.06 instead of >0.1). We suspect the estimates would be improved if variation in the probability of detection were modeled appropriately. Specifically, the probability of detection is likely to vary with many variables, including the

observer, time of day, time of year and weather. A binomial distribution was used to model detection probability with only two covariates to model detection probability: time of day and time of year. Therefore, there is likely to be extra-binomial variation in the detection probability response variable. The beta-binomial distribution is an alternative to the binomial and accounts for the extra-binomial variation (n.b. the beta binomial may also account for extra-binomial variation that results from violation of the independence assumption). The beta-binomial model assumes that the responses follow an independent Bernoulli process, and that the Bernoulli parameter itself is a random variable that varies among groups according to a beta distribution. Further research is required to determine if the *N*-mixture model with a beta binomial distribution modeling the detection probability will generate ecologically reasonable parameter estimates. Given the large number of components of a beta-binomial *N*-mixture model, care must be taken to ensure that this model does not overfit the data.

In most cases, both of the Poisson variants of the *N*-mixture model produced parameter estimates that are ecologically sensible. In contrast, the parameter esti-

TABLE 3. Assessment of the ecological realism of parameter estimates for six bird species of the Mount Lofty Ranges for the four variants of the  $N$ -mixture model: Poisson (P), zero-inflated Poisson (ZIP), negative binomial (NB), and zero-inflated negative binomial (ZINB).

Bird species and distribution	Abundance	Probability of detection	Probability of occupancy	Ecologically realistic models
<b>Rufous Whistler</b>				
<b>P</b>	×	×	×	×
<b>ZIP</b>	×	×	×	×
<b>NB</b>	O	O	O	O
<b>ZINB</b>	O	O	O	O
<b>Mistletoebird</b>				
<b>P</b>	×	×	×	×
<b>ZIP</b>	×	×	×	×
<b>NB</b>	O	×	O	O
<b>ZINB</b>	O	×	×	O
<b>Scarlet Robin</b>				
<b>P</b>	×	×	O	O
<b>ZIP</b>	×	O	×	O
<b>NB</b>	O	O	O	O
<b>ZINB</b>	O	O	×	O
<b>White-throated Treecreeper</b>				
<b>P</b>	×	×	×	×
<b>ZIP</b>	×	×	×	×
<b>NB</b>	×	×	×	×
<b>ZINB</b>	×	×	×	×
<b>Gray Fantail</b>				
<b>P</b>	×	×	×	×
<b>ZIP</b>	×	×	×	×
<b>NB</b>	×	×	×	×
<b>ZINB</b>	×	×	×	×
<b>Golden Whistler</b>				
<b>P</b>	×	×	×	×
<b>ZIP</b>	×	×	×	×
<b>NB</b>	×	O	×	O
<b>ZINB</b>	O	O	×	O

Notes: Crosses (×) indicate that parameter estimates are realistic, and circles (O) indicate that parameters are not realistic based on criteria described in *Methods*. Models that produced ecologically reasonable estimates for all three parameters are deemed ecologically realistic models and are highlighted in bold.

mates generated by the negative binomial variants were often ecologically unrealistic. This result directly conflicts with AIC model selection that commonly selected the negative binomial variants in preference to the Poisson variants. The AIC is used to select models on their statistical fit to data and, clearly, is unable to assess ecological realism.

Statistical models like the negative binomial variants of the  $N$  mixture, that model statistic properties of data and not ecological mechanisms, are susceptible to unexpected and erroneous parameter estimation such as we describe here. The process by which negative binomial models accommodate zero-inflation (or any form of extra-Poisson variation) is to allow the mean site abundance ( $\lambda$ ) to vary stochastically. This process will lead to the simultaneous prediction of excess zeros and excessively large values. When data are modeled with a traditional negative binomial distribution (i.e., not within an  $N$ -mixture model), the extent of large values that can be predicted by a parsimonious model will be restricted by the data. However, for an  $N$ -mixture model, the large estimates of site abundance can be

‘counteracted’ with excessively small detection probabilities by modeling the observed site abundance as a large number of individuals present each with very low probability of detection. The AIC cannot discriminate among models that use different components of the  $N$  mixture (i.e., the detection error or the abundance estimators) to model the error and may, as is the case in our study, find a “best” model that generates unrealistic estimates. The ZIP variant of the  $N$ -mixture model is an appropriate alternative that models ecological mechanism and successfully estimates ecologically realistic parameters. This variant of the  $N$ -mixture models excess zeros as either detection error or unoccupied sites in a manner that is similar to the ecological mechanism underlying the data.

Our distrust of the negative binomial  $N$ -mixture model is corroborated by results presented in Kéry et al. (2005), who compared estimates of site abundance predicted with the  $N$ -mixture model (fitting Poisson and negative binomial distributions) with those estimated from territorial mapping methods for six bird species. They found that the negative binomial form of the  $N$ -mixture

model commonly estimated mean site abundance that was approximately double the estimates from the territory mapping method. Specifically, the ratio of site abundances of the two methods ( $N_{N\text{-mixture}}/N_{N\text{-mapping}}$ ) ranged up to 8.86. For all species, the  $N$  mixture under the Poisson distribution and the territory mapping technique gave very similar estimates for site abundance (mean  $N_{N\text{-mixture}}/N_{N\text{-mapping}}$   $N$  was 1.18, ranging from 0.93 to 1.45). The problem with these comparisons was that the AIC and goodness-of-fit statistics frequently selected the negative binomial form of the  $N$  mixture as the best model (i.e., for five of the six species). Kéry et al. (2005) expressed suspicion that the negative binomial distribution may not have been the best choice for representing the over-dispersion in abundance relative to the Poisson. Given the conclusions drawn from our analysis, we suspect that the negative binomial distribution was inappropriate. We suggest that estimates of site abundance given under the Poisson distribution and the territorial mapping technique are more trustworthy than those produced under the negative binomial distribution.

#### *Ecologically sensible model choice*

Our results demonstrate the importance of carefully considering sources of variation and selecting ecologically sensible models before fitting data to models. An inappropriate error distribution in the  $N$ -mixture model will give misleading parameter estimates. Conversely, properly representing ecological mechanisms in models will assist in appropriate model selection (see Austin and Meyers 1996). We have demonstrated that the extra-Poisson variation in three of our six species occurred due to an ecological mechanism that is easily recognized and modeled. Zero-inflation can be successfully modeled by allowing an extra parameter to estimate the site suitability before estimating conditional density. This process is ecologically meaningful and appropriate, unlike the process represented by the negative binomial model. Our results demonstrate the dangers of using statistical models that lack an underlying ecological mechanism and are instead based only on statistical properties of the data.

The  $N$  mixture is a powerful model for estimating the abundance and detection probability of species. The zero-inflated Poisson  $N$ -mixture model improves the capabilities of the modeling technique by allowing the simultaneous estimating of probability of site occupancy. These models can be used to simultaneously describe true variation that represents important ecological mechanisms and false variation that is due to observation error. If a species is likely to be present in all survey sites and true and false variation exists in abundance estimates, a standard Poisson  $N$ -mixture model to estimate abundance and detectability may be sufficient. However, if the species is absent from a reasonable proportion of sites (e.g., >30%), the zero-inflated Poisson  $N$ -mixture model is more appropriate.

#### *Interpretation of site abundance for mobile species*

It is useful to be able to make predictions about the potential abundance across a species distribution; however,  $N$ -mixture estimates of abundance should be used with care. By extrapolating the abundance that is estimated with the  $N$ -mixture model across a species' distribution, we make the implicit but invalid assumption that the abundance estimate corresponds to density (i.e., number of individuals per unit area). In actuality, the mean abundance estimated (i.e.,  $\theta_i = \lambda_i$  or  $\mu_i$ ) by the  $N$ -mixture model ( $\hat{N}$ ) is the number of individuals that have territories that at least partially overlap with the survey area. For mobile species, the actual numbers of individuals that utilize a site and are available for observation ( $N$ ) come from an area that is generally much larger than the actual survey area. The  $N$ -mixture abundance estimate (i.e., the number of individuals that utilize a site) will be consistently larger than the true density and in some cases the overestimation will be substantial.

#### CONCLUSION

By examining the parameters estimated using standard and zero-inflated forms of the Poisson and negative binomial  $N$ -mixture models, we demonstrate that the zero-inflated Poisson model appears an appropriate and meaningful model. On the other hand, the negative binomial forms are, in many cases, not appropriate and may give misleading results. We demonstrate that to obtain ecologically realistic estimates of abundance and detection probability it is essential to understand the sources of variation in the data and then use this information to choose appropriate error distributions. There is little ecological reason for using a negative binomial  $N$ -mixture model. Given that the negative binomial variant produces ecologically unrealistic estimates, we recommend that it not be used to predict abundance and detection probability when using the  $N$ -mixture model. The zero-inflated Poisson will be particularly applicable when excess zeros in the data are due to unoccupied sites. We have demonstrated that the zero-inflated Poisson will produce sensible estimates of detection probability and abundance even when data are not zero inflated. Additionally, using the zero-inflated Poisson model, by properly modeling error, will allow researchers to make best use of other features of  $N$ -mixture models such as the ability to include ecological correlates relevant to occupancy and abundance.

#### ACKNOWLEDGMENTS

We thank Max Possingham, Matthew Watts, Matthew Ward, Michelle Ensbey, Jane Elith, David Pavlacky, and Norbert Menke for discussions and/or technical assistance. This work was supported by the Applied Environmental Decision Analysis research hub which is funded by the Commonwealth Environment Research Facilities (CERF) program, an Australian Government initiative which promotes world class, public good research, an Australian Postgraduate Award, the Coop-

erative Research Centre for Tropical Rainforest Ecology and Management Scholarship, and a Birds Australia Stuart Leslie Award for L. N. Joseph, NSERC for T. G. Martin, and ARC Discovery and ARC Linkage grants for H. P. Possingham.

## LITERATURE CITED

- Anderson, D. R. 2001. The need to get the basics right in wildlife field studies. *Wildlife Society Bulletin* 29:1294–1297.
- Austin, M. P., and J. A. Meyers. 1996. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *Forest Ecology and Management* 85:95–106.
- Barry, S. C., and A. H. Welsh. 2002. Generalized additive modelling and zero inflated count data. *Ecological Modelling* 157:179–188.
- Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York, New York, USA.
- De Groot, M., and M. Schervish. 2002. Probability and statistics. Third edition. Addison Wesley, Boston, Massachusetts, USA.
- Dodd, C. K., and R. M. Dorazio. 2004. Using counts to simultaneously estimate abundance and detection probabilities in a salamander community. *Herpetologica* 60:468–478.
- Doob, J. L. 1934. Probability and statistics. *Transactions of the American Mathematics Society* 36:759–775.
- Ford, H. A., and R. W. Howe. 1980. The future of birds in the Mount Lofty Ranges. *South Australian Ornithologist* 28:85–89.
- Gaston, K. J., and A. S. L. Rodrigues. 2003. Reserve selection in regions with poor biological data. *Conservation Biology* 17:188–195.
- Hall, D. B. 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 56:1030–1039.
- Higgins, P. J., and J. M. Peter. 2002. Handbook of Australian, New Zealand and Antarctic Birds. Volume 6. Pardalotes to Spangled Drongo. Oxford University Press, Melbourne, Australia.
- Higgins, P. J., J. M. Peter, and S. J. Cowling. 2006. Handbook of Australian, New Zealand and Antarctic Birds. Volume 7. Boatbill to Starlings. Oxford University Press, Melbourne, Australia.
- Higgins, P. J., J. M. Peter, and W. K. Steele. 2001. Handbook of Australian, New Zealand and Antarctic Birds. Volume 5. Tyrant-Flycatchers to Chats. Oxford University Press, Melbourne, Australia.
- Kéry, M., J. A. Royle, and H. Schmid. 2005. Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications* 15:1450–1461.
- King, G. 1989. Even count models for interventional relations: generalisations and applications. *International Studies Quarterly* 33:123–147.
- Kuhnert, P. M., T. G. Martin, K. Mengersen, and H. P. Possingham. 2005. Assessing the impacts of grazing levels on bird density in woodland habitat: a Bayesian approach using expert opinion. *Environmetrics* 16:717–747.
- Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34:1–14.
- Lee, A. H., M. R. Stevenson, K. Wang, and K. K. W. Yau. 2002. Modeling young driver motor vehicle crashes: data with extra zeros. *Accident Analysis and Prevention* 34:515–521.
- Loyn, R. 1986. The 20 minute search: a simple method for counting forest birds. *Corella* 10:58–60.
- MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* 84:2200–2207.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy when detection probabilities are less than one. *Ecology* 83:2248–2255.
- Martin, T. G., P. M. Kuhnert, K. Mengersen, and H. P. Possingham. 2005a. The power of expert opinion in ecological models using Bayesian methods: impact of grazing on birds. *Ecological Applications* 15:266–280.
- Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham. 2005b. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters* 8:1235–1246.
- McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Chapman and Hall, London, UK.
- Meents, J. K., J. Rice, B. W. Anderson, and R. D. Ohmart. 1983. Nonlinear relationships between birds and vegetation. *Ecology* 64:1022–1027.
- Mullahy, J. 1986. Specification and testing of some modified count data models. *Journal of Econometrics* 33:341–365.
- Nelder, J. A., and R. W. M. Wedderburn. 1972. Generalized linear models. *Journal of the Royal Statistical Society: Series A* 135.
- Pearce, J., and S. Ferrier. 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biological Conservation* 98:33–43.
- Pollock, K. H., J. D. Nichols, T. R. Simons, G. L. Farnsworth, L. L. Bailey, and J. R. Sauer. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* 13:105–119.
- Possingham, M., S. A. Field, and H. P. Possingham. 2004. Species richness and abundance of birds in Mt Lofty Ranges Stringybark habitat: 1999–2000 survey. *South Australian Ornithologist* 34:153–169.
- Possingham, M. L., and H. P. Possingham. 2000. The birds of the South East of South Australia and their habitats. Pages 180–200 in R. Collier, J. Hatch, B. Matheson, and R. Russell, editors. *Birds, birders and bird watching: 1899–1999: a celebration of one hundred years of the SAOA*. South Australian Ornithological Association, Adelaide, Australia.
- Pulliam, H. R. 1988. Sources, sinks, and population regulation. *American Naturalist* 132:652–661.
- Ridout, M., C. G. B. Demetrio, and J. Hinde. 1998. Models for count data with many zeros. Pages 179–190 in *Invited paper presented at the Nineteenth International Biometric Conference*, Cape Town, South Africa.
- Rosenstock, S. S., D. R. Anderson, K. M. Gieson, T. Leukering, and M. F. Carter. 2002. Landbird counting techniques: current practices and an alternative. *Auk* 119:46–53.
- Royle, J. A. 2004. *N*-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60:108–115.
- Royle, J. A., J. D. Nichols, and M. Kery. 2005. Modelling occurrence and abundance of species when detection is imperfect. *Oikos* 110:353–359.
- Sánchez-Zapata, J. A., and J. F. Calvo. 1999. Raptor distribution in relation to landscape composition in semi-arid Mediterranean habitats. *Journal of Applied Ecology* 36: 254–262.
- Tittler, R., L. Fahrig, and M.-A. Villard. 2006. Evidence of large-scale source-sink dynamics and long-distance dispersal among wood thrush populations. *Ecology* 87:3029–3036.
- Tyre, A. J., B. Tenhumberg, S. A. Field, D. Niejalke, K. Parris, and H. P. Possingham. 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications* 13:1790–1801.
- Welsh, A. H., R. B. Cunningham, C. F. Donnelly, and D. B. Lindenmayer. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling* 88:297–308.
- Zorn, C. J. 1996. Evaluating zero-inflated and hurdle Poisson specifications. *Midwest Political Science Association* April 18–20:1–16.