

REVIEWS AND
SYNTHESES

Zero tolerance ecology: improving ecological inference by modelling the source of zero observations

Tara G. Martin,^{1*} Brendan A. Wintle,² Jonathan R. Rhodes,³ Petra M. Kuhnert,⁴ Scott A. Field,⁵ Samantha J. Low-Choy,⁶ Andrew J. Tyre^{7†} and Hugh P. Possingham¹

Abstract

A common feature of ecological data sets is their tendency to contain many zero values. Statistical inference based on such data are likely to be inefficient or wrong unless careful thought is given to how these zeros arose and how best to model them. In this paper, we propose a framework for understanding how zero-inflated data sets originate and deciding how best to model them. We define and classify the different kinds of zeros that occur in ecological data and describe how they arise: either from ‘true zero’ or ‘false zero’ observations. After reviewing recent developments in modelling zero-inflated data sets, we use practical examples to demonstrate how failing to account for the source of zero inflation can reduce our ability to detect relationships in ecological data and at worst lead to incorrect inference. The adoption of methods that explicitly model the sources of zero observations will sharpen insights and improve the robustness of ecological analyses.

Keywords

Bayesian inference, detectability, excess zeros, false negative, mixture model, observation error, sampling error, zero-inflated binomial, zero-inflated Poisson, zero inflation.

Ecology Letters (2005) 8: 1235–1246

INTRODUCTION

A distinguishing characteristic of many ecological data sets, whether comprised of data measuring binary presence/absence, counts of abundance, proportional occupancy rates or continuous population densities, is their tendency to contain a large proportion of zero values (Clarke & Green 1988, Fig. 1). When this number of zeros is so large that the data do not readily fit standard distributions (e.g. normal, Poisson, binomial, negative-binomial and beta), the data set is referred to as ‘zero inflated’ (Heilbron 1994; Tu 2002). Zero inflation is often the result of a large number of ‘true

zero’ observations caused by the real ecological effect of interest. For example, the study of rare organisms or events will often lead to the collection and analysis of data with a high frequency at zero (Welsh *et al.* 1996). However, the term can also be applied to data sets with excess zeros caused by ‘false-zero’ observations because of sampling or observer errors in the course of data collection. Failure to account for either source of zero inflation will cause bias in parameter estimates and their associated measures of uncertainty (Lambert 1992; MacKenzie *et al.* 2002).

The presence of zero inflation due to excess true zeros, a special case of overdispersion (McCullagh & Nelder 1989;

¹The Ecology Centre, School of Life Sciences, The University of Queensland, Brisbane, Qld 4072, Australia

²Environmental Science, School of Botany, University of Melbourne, Vic. 3010, Australia

³School of Geography, Planning and Architecture, The University of Queensland, Brisbane, Qld 4072, Australia

⁴CSIRO Mathematical and Information Sciences, Cleveland, Qld, Australia

⁵School of Earth and Environmental Sciences, University of Adelaide, North Terrace, SA 5005, Australia

⁶School of Mathematical Sciences, Queensland University of Technology, Brisbane, Qld 4001, Australia

⁷School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, USA

*Correspondence and Present address: CSIRO Sustainable Ecosystems, QBP, 306 Carmody Road, Qld 4067, Australia. E-mail: tara.martin@csiro.au

†A contribution of the University of Nebraska Agricultural Research Division, Lincoln, NE 68583, USA.

Journal series no. 14657

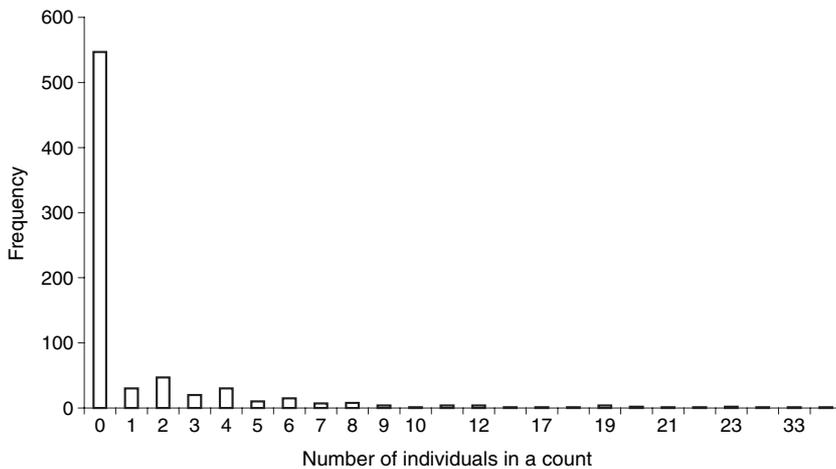


Figure 1 Example of a typical zero-inflated data set. Frequency of counts for 31 bird species across eight sites and three grazing treatments ($n = 744$) from Martin *et al.* (2005). Over 70% of the data set is represented by zero counts, which is more than expected if a Poisson distribution is assumed for the species' abundances.

Hinde & Demétrio 1998; Poortema 1999), creates problems with making sound statistical inference by violating basic assumptions implicit in the use of standard distributions (Mullahy 1986; Cameron & Trivedi 1998). One common violation is a misrepresentation of the variance–mean relationship of the error structure (Barry & Welsh 2002). In ecology, transformations are often used to overcome such problems. However, the difficulty with this approach for zero-inflated data sets is that, while the transformation may normalize the distribution of the non-zero values, no transformation will spread out the zero values. The high frequency of zero values is simply replaced by an equally high frequency of the value to which zero is transformed (Hall 2000).

The presence of zero inflation as a result of false zeros may or may not violate the distributional assumptions, but will lead to uncertainty regarding parameter estimates because it is no longer possible to determine whether a difference in the number of individuals surveyed over time and space is because of a change in the size of the population or because of a change in the detection probability of the individuals (MacKenzie *et al.* 2002).

Zero inflated count data and the application of models which cope with zero inflation are found in a wide range of disciplines including epidemiology (Böhning *et al.* 1999; Lewsey & Thomson 2004), medicine (Campbell *et al.* 1991; Ghahramani *et al.* 2001; Cheung 2002), occupational health (Lee *et al.* 2002; Carrivick *et al.* 2003; Wang *et al.* 2003; Yau *et al.* 2004) and econometrics (Freund *et al.* 1999).

The ecological literature has seen a recent upsurge of interest in techniques for dealing with excess zero values. Zero-inflated models have been applied in a range of ecological scenarios, including data sets with zero inflation caused by true zero (Welsh *et al.* 1996, 2000; Barry & Welsh 2002; Podlich *et al.* 2002; Kuhnert *et al.* 2005; Martin *et al.*

2005) and false-zero observations (Kery 2002; MacKenzie *et al.* 2002, 2003, 2004; Tyre *et al.* 2003; Wintle *et al.* 2004).

In this paper, we propose a framework for understanding how zero-inflated data sets originate and to decide which of the many available models to apply in any given case. In doing so, we aim to bring these models to the attention of a broader ecological readership and help ecologists navigate the growing number of zero-inflated modelling approaches at their disposal. First, we define the different kinds of zeros that occur in ecological data and describe how they arise. We then describe the approaches used to model the two types of data typical of that collected in ecological studies; presence/absence and count. The use of a selection of these models is then illustrated through two detailed examples in which the data are subject to different kinds of zero inflation. Finally we discuss the potential gains in ecological understanding made by applying such models.

Sources of zeros in ecological data

Zero values occur in one of four ways; two of which can be defined as 'true zero counts' and two as 'false-zero counts' (Table 1). The first kind of true zero arises from a low frequency of occurrence, which can be the result of range of ecological processes and life-history strategies (Gaston 1994) or the result of a strong ecological effect that leads to sites having no organisms present. For example, a species may be absent because of demographic processes, competition, or poor habitat quality (e.g. because of disturbances or unsuitable vegetation structure). These zeros are true zeros resulting from the real ecological effect that we are trying to determine. Secondly, a zero may occur simply by chance, because the species does not saturate its entire suitable habitat (e.g. because of local extinctions caused by demographic stochasticity).

Table 1 Four ways zeros arise in ecological data

Type of zero	Definition
True zero	Species does not occur at a site because of the ecological process, or effect under study (e.g. habitat unsuitable) Species does not saturate its entire suitable habitat by chance
False zero	Species occurs at a site, but is not present during the survey period Species occurs at a site and is present during the survey period, but the observer fails to detect it (particularly common for rare or cryptic species)

The first kind of false zero is caused by failing to record a species that inhabits a site because, although it occupies the site, it was not present at the time of the survey. This can be caused by using a sampling area that is small and/or length of visit that is short, relative to the temporal and spatial scale of movements of the species (Tyre *et al.* 2003). The second kind occurs when the species occupies a site and is present at the time of sampling, but the observer does not detect the species. These errors are common for cryptic or secretive species (MacKenzie *et al.* 2002).

It is worth noting that the type of zero represented by a particular observation depends on the study objective. For example, in the case where a species may be temporarily absent from a study site, if the aim is to quantify where the species is instantaneously, its absence would not constitute a false zero (i.e. the species really was not there when surveyed). However, if we were interested in what areas were being used by the species over a longer time frame, then its absence would constitute a false zero.

Aside from the categories defined above, a large number of zeros can arise in ecological data in other ways, when observations are obtained from outside the environmental range of a species, referred to by Austin & Meyers (1996) as 'naughty naughts'. The solution to this problem involves a reduction or filtering of data sets to exclude the 'naughty

naughts' from outside the species' range (Austin & Meyers 1996; Elith & Burgman 2002) or to simply avoid their collection through thoughtful sampling design.

Choosing an appropriate zero-inflated model

When considering how to model zero-inflated data sets, it is important to take into account which kinds of zeros are present (Table 1). In this section, we outline the recommended modelling approaches when the data set are dominated by either true zeros, false zeros or a combination of the two for presence/absence and count data (Table 2).

Zero-inflated Poisson (ZIP) and binomial (ZIB) models fitted to data without covariates have a long history (Johnson & Kotz 1969). Lambert (1992) provides the general form of ZIP regression with covariates to model defects in a manufacturing process. Models specifically for zero-inflated count data have been under development by Heilbron (1994), Welsh *et al.* (1996, 2000), Faddy (1998), Hall (2000), Dobbie & Welsh (2001), Barry & Welsh (2002) and Wang (2003) and applied using a Bayesian philosophy to statistical inference by Angers & Biswas (2003), Martin *et al.* (2005) and Kuhnert *et al.* (2005). The development of zero-inflated models for continuous data, such as fish stock assessment has also received attention (i.e. log-normal, delta log-normal and delta-Gamma models) (Aitchison 1955; Stefansson 1996; Syrjala 2000) and has been developed further by Fletcher *et al.* (2005).

No zero-inflation

In the absence of zero-inflation, a standard single distribution model such as the binomial or Poisson is used. McCullagh & Nelder (1989) provide a full discussion of the sampling distributions and models for this type of data (Table 2).

Zero-inflation due to true zeros

When true zeros lead to an excess of zeros, zero-inflated models such as two-part (also known as conditional or hurdle models) or mixture models are recommended (Lambert 1992; Welsh *et al.* 1996). The negative binomial

Table 2 Four scenarios of zero occurrences in ecological data and the modelling approach recommended for presence/absence and count data, where zero inflation can be caused by false zeros, true zeros or a combination of both

Zero inflation	Modelling approach	Key references
None	Single distribution models (e.g. binomial)	McCullagh & Nelder (1989)
True zeros	Zero-inflated mixture models, ZIB or ZIP with point mass at zero, or hurdle models	Lambert (1992), Welsh <i>et al.</i> (1996) and Hall (2000)
False zeros	Zero-inflated mixture models (e.g. ZIB or ZIP)	MacKenzie <i>et al.</i> (2002, 2003) and Tyre <i>et al.</i> (2003)
Both	Mixture of two or more distributions	None found

The zero-inflated models are based on the binomial distribution for presence/absence data, and on the Poisson or negative-binomial model for count data. ZIP, zero-inflated Poisson; ZIB, zero-inflated binomial.

has also been advocated for modelling data sets with many zeros because of its ability to account for overdispersion (Warton 2005). However, Welsh *et al.* (1996) and Hall (2000) demonstrated that the excess number of zeros often exceeds those expected under a negative-binomial distribution.

For count data, a two-part modelling approach has appeared in the ecological literature, whereby the first part is a binary outcome model (i.e. Bernoulli), and the second part is a truncated count model (e.g. Poisson or negative binomial) (Cameron & Trivedi 1998). This approach assumes that zeros arise from a single process and a set of covariates. One of its computational benefits is that it is possible to fit these models in two parts, for example, fitting zeros using a logistic regression separately from fitting non-zeros using a truncated Poisson (e.g. Welsh *et al.* 1996; Dobbie & Welsh 2001). Using this approach, one is able to estimate the probability that a species is present and then given it is present, estimate the relative mean number of individuals.

Mixture models are combinations of probability distributions chosen for their ability to represent two or more real ecological processes. The ZIP mixture model used to model count data is a mixture of a point mass at zero and a Poisson distribution. With this approach, zeros may arise from one of two processes and their related covariates, a zero-process from which only zero values are observed and a Poisson process in which non-zero and a proportion of the zero values, appropriate to the Poisson distribution are observed (Lambert 1992). The interpretation of mixture model parameters is less straight forward than the two-part model. For example, to get the true estimate of relative mean abundance from the ZIP one must multiply the estimated relative mean number of individuals at a site by the probability that the relative mean number of individuals at a site is generated through a Poisson distribution.

Where there is zero inflation and overdispersion caused by large counts of individuals (e.g. flocking birds), the use of a zero-inflated negative binomial (ZINB) mixture model has been shown to be appropriate (Welsh *et al.* 2000).

Zero inflation due to false zeros

If false zeros are present in the data a zero-inflated mixture modelling approach is required (MacKenzie *et al.* 2002; Tyre *et al.* 2003) because we are interested in modelling two processes, a process leading to true zeros and a process leading to false zeros.

A recent set of articles highlight the problem of false zeros in ecological data sets that are collected for the purpose of assessing site occupancy (Kery 2002; MacKenzie *et al.* 2002) and making inferences about species–habitat relationships, or the effects of anthropogenic activities, on species distributions (e.g. Tyre *et al.* 2003; Gu & Swihart

2004). Failing to take account of false-zero observations in analyses may have substantial impacts on the ability to accurately infer relationships between site occupancy and habitat attributes or management actions (MacKenzie *et al.* 2003; Field *et al.* 2005; Rhodes *et al.* 2005). The zero-inflated binomial (ZIB) model and its extensions provide an appropriate framework for analysing data that are collected for these purposes and which are likely to contain false-zero observation error (MacKenzie *et al.* 2002; Tyre *et al.* 2003; Wintle *et al.* 2004, 2005).

Zero inflation due to both excess true zeros and false zeros

In the literature there has been no formal discussion of how to model data sets that contain both excess true zeros and false zeros. Using a Bayesian framework one approach would be to incorporate information on the contribution of false zeros to the data (e.g. detection probability) as an informative prior in a zero-inflated model.

Uncertainty regarding the source of zero inflation

In some cases it is not possible to determine the source of zero observations. One way of dealing with this uncertainty is to use a truncated distribution whereby the zeros are eliminated completely and only the occurrences are modelled. For example, Baum & Myers (2004) were unable to determine whether the absence of sharks in bycatch data sets were a result of true zeros (e.g. there were no sharks in the bycatch) or the fisherman's failure to record the sharks in the bycatch. They dealt with this uncertainty by using a truncated negative-binomial model to estimate trends in shark numbers from only the non-zero shark catches.

Modelling zero inflation in ecological data

In this section, we present two examples that deal with zero inflation: (i) generated by excessive numbers of true zeros in count data and (ii) arising from false zeros in presence/absence data. Both examples are illustrated through Bayesian inference using simulation-based Markov Chain Monte Carlo (Ellison 2004).

In the first example, we illustrate the use of the ZIP and ZINB mixture models and compare their performance with standard Poisson and negative-binomial models, in an examination of the impact of livestock grazing on the relative mean abundance of four Australian woodland birds, where zero inflation is a result of an ecological process leading to an excess of true zeros. The second example demonstrates the use of the ZIB mixture model in making inferences about the suitability of habitat in a highly fragmented landscape for four woodland bird species. It specifically accounts for zero inflation resulting from false zeros generated through the sampling process.

Modelling the impact of grazing on bird assemblages with zero-inflated count data caused by excess true zeros

Zero inflated mixture models

Using the mixture modelling approach, $p(x_i)$ represents the probability that an observation i is generated through either the Poisson distribution or the negative binomial, irrespective of whether the observation is a zero or non-zero value. Equations 1 and 2 show expressions for the two models under investigation.

ZIP mixture model

$$\begin{aligned} \Pr(Y_i = 0|x, z) &= 1 - p(x_i) + p(x_i) \exp(-\lambda(z_i)), \\ \Pr(Y_i = r|x, z) &= p(x_i) \frac{\exp(-\lambda(z_i)) \lambda(z_i)^r}{r!}, \quad r = 1, 2, \dots, \end{aligned} \quad (1)$$

where

$$\begin{aligned} \text{logit}(p(x_i)) &= \alpha_0 + \beta_0(x_i) \\ \log(\lambda(z_i)) &= \alpha_1 + \beta_1(z_i) \end{aligned}$$

In both equations, $\lambda(z_i)$ represents the mean number of individuals at site i and it can be expressed as a function of the explanatory variables, z through a log transformation. Similarly, $p(x_i)$ can be expressed as a function of the explanatory variables, x , using a logit transformation where x does not necessarily have to be the same set of covariates as those represented by z . Here, the parameters α_0 and α_1 represent constant terms in each regression component and β_0 and β_1 are vectors, representing the coefficients estimated for each explanatory variable fitted in the model.

ZINB model

$$\begin{aligned} \Pr(Y_i = 0|x, z) &= 1 - p(x_i) + p(x_i) \frac{1}{1 + \phi \lambda(z_i)}^{1/\phi}, \\ \Pr(Y_i = r|x, z) &= p(x_i) \frac{\Gamma(r + 1/\phi)}{r! \Gamma(1/\phi)} \\ &\times \frac{\phi^r \lambda(z_i)^r}{(1 + \phi \lambda(z_i))^r (1 + \phi \lambda(z_i))^{1/\phi}}, \quad r = 1, 2, \dots, \end{aligned} \quad (2)$$

where

$$\begin{aligned} \text{logit}(p(x_i)) &= \alpha_0 + \beta_0(x_i) \\ \log(\lambda(z_i)) &= \alpha_1 + \beta_1(z_i) \end{aligned}$$

Equation 2 has an additional parameter, ϕ which allows estimation of an overdispersion parameter in situations where large counts have been recorded or alternatively, a large number of zeros have been observed. In both models, if $p(x_i)$ is equal to 1, we default to the usual Poisson and

negative-binomial models for count data. See Lambert (1992); Welsh *et al.* (1996) and Dalrymple *et al.* (2003) for more details.

The species, study site and data collection

Martin *et al.* (2005) and Kuhnert *et al.* (2005) examined the impact of livestock grazing on the relative abundance of 31 woodland birds in subtropical Australia. Bird count data was collected across three broad levels of grazing (low, moderate and high) in eucalypt woodland habitat. Eight replicate sites of each grazing regime were sampled. Sites were visited on two separate days and over two seasons, giving a total of 24 sites and 96 site visits.

For comparisons of relative mean abundance estimates to be valid, detection or capture probabilities of individuals are assumed to be equal (e.g. across different sites). In this study this assumption was justified by the open vegetation structure of the sites and conspicuous behaviour of the birds examined (Martin *et al.* 2005).

Using data from four of the bird species investigated by Martin *et al.* (2005) and Kuhnert *et al.* (2005), we compared the relative mean abundance estimates and credible intervals from fitting Poisson, negative-binomial, ZINB mixture and a ZIP mixture models. To get an estimate of relative mean abundance from the ZIP mixture that could be compared with the relative mean abundance estimate from the Poisson model, the ZIP mixing probability $p(x)$, the probability that the number of individuals at a site has a Poisson distribution, was multiplied by $\lambda(z)$ the mean of the estimate given that it was generated from a Poisson distribution.

Models were fitted using the Bayesian statistical modelling freeware package, WinBUGs (Spiegelhalter *et al.* 2003). The deviance information criterion (DIC) was calculated to compare the fit of the four models (Spiegelhalter *et al.* 2002). From a Bayesian perspective the DIC is analogous to Akaike's information criterion (Akaike 1973), in that its intent is to assess the models in terms of their fit and complexity (Burnham & Anderson 2002). The DICs computed by WinBUGs were checked using the formula recommended by Celeux *et al.* (2003).

We modelled four bird species separately treating grazing as a fixed effect. Convergence was achieved after a burn-in of 10 000 iterations and estimates were obtained after a further 30 000 iterations. Convergence of the Markov chains was examined using the CODA package (Best *et al.* 1995).

An examination of the frequency of counts for the four bird species under investigation revealed that data for three of the species (brown thornbill *Acanthiza pusilla*, noisy miner *Manorina melanocephala* and superb fairy-wren *Malurus cyaneus*) were zero inflated (Fig. 2). This zero-inflation was a result of species showing strong preferences for particular grazing levels and an avoidance of others.

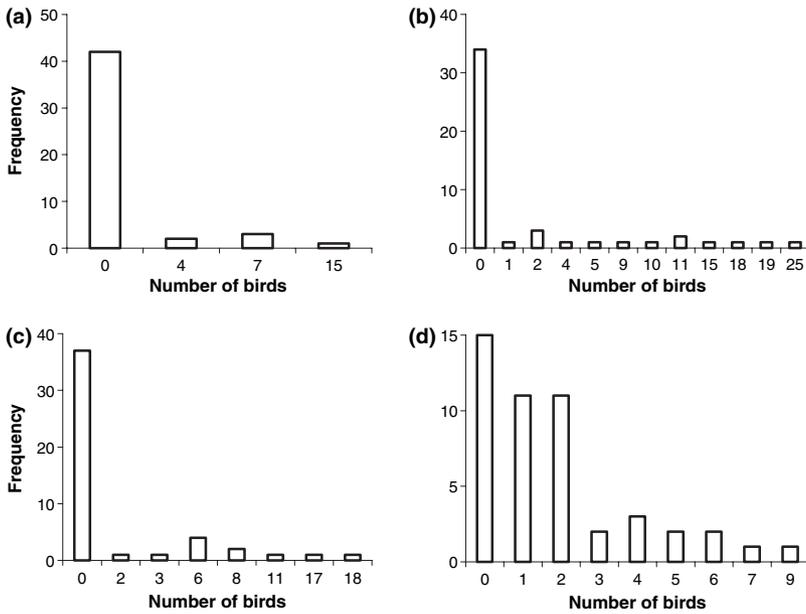


Figure 2 Frequency of counts of four woodland bird species: (a) brown thornbill, (b) noisy miner, (c) superb fairy-wren and (d) rufous whistler across 24 sites visited twice in summer and twice winter.

On the contrary, the data for the rufous whistler *Pachycephala rufiventris* was more consistent with properties of the Poisson distribution.

For purpose of illustration, counts for each species were pooled across visits and modelled across grazing levels. Season was not a significant contributor to the model and was not included as a factor. Although informative priors were used in the full study (Martin *et al.* 2005), here we considered non-informative normal priors with a mean of zero and precision parameters equal to 0.0001. In this example, the mixing probability $p(x)$ was fixed across grazing levels, however, one could allow $p(x)$ to vary by modelling grazing as a covariate (see Appendix S1 for code). Full details on the study design, data collection, analyses using both mixture and two-part zero-inflated models and results for all species are described in Martin *et al.* (2005) and Kuhnert *et al.* (2005).

Results

Comparison of model fit as determined by the DIC of all four models showed that the ZIP performed best for the brown thornbill, which coincidentally was the most zero-inflated species (Fig. 2). The negative binomial performed best for the noisy miner, and the ZINB performed best for the superb fairy-wren and rufous whistler (Table 3). The standard Poisson had the poorest fit for all four bird species. The DICs for the rufous whistler were only marginally different amongst the four models. This is a result of the data for this species exhibiting properties more consistent to that of the Poisson distribution, i.e. the mixing probability $p(x)$ from the ZIP that an observation came from a Poisson distribution was closer to 1 (Table 3).

Comparing the estimates from the negative-binomial, ZINB and ZIP mixture with the Poisson model revealed that the 95% credible intervals from the negative-binomial, ZINB and ZIP mixture were much broader than those using the

Model	Brown thornbill	Noisy miner	Superb fairy-wren	Rufous whistler
Poisson	123.5	245.1	267.1	195.4
Negative binomial	67.7	137.9	121.0	180.8
ZINB mixture	–	141.1	105.9	177.0
ZIP mixture	60.8	167.3	120.9	189.6
ZIP \hat{p}	0.341	0.479	0.337	0.822
95% CI	(0.132–0.586)	(0.249–0.774)	(0.185–0.513)	(0.649–0.983)

Table 3 Deviance information criterion (DIC) for the Poisson, negative-binomial, zero-inflated negative binomial (ZINB) and zero-inflated Poisson (ZIP) mixture models for four woodland bird species

Estimates of the mixing probability \hat{p} and 95% credible interval in brackets from the ZIP are reported in the last row of the table, where \hat{p} is the probability that an observation is generated through the Poisson distribution. A dash (–) denotes the model could not be fit.

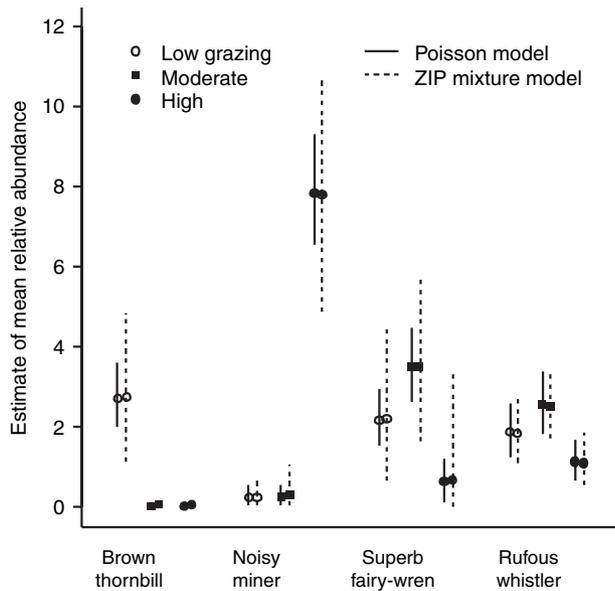


Figure 3 Estimates of relative abundance and 95% credible interval of four woodland bird species across three grazing levels (low, moderate and high) for the Poisson model and the zero-inflated Poisson mixture model.

standard Poisson model for the three species whose frequency were most zero inflated as illustrated for the ZIP and Poisson in Fig. 3. The superb fairy-wren was predicted to be significantly less abundant under high grazing than either low or moderate under the Poisson model, whereas under the negative-binomial, ZINB and ZIP model there was no substantial difference in relative mean abundance estimates across the three grazing levels. Conversely, estimates from the four models did not vary substantially for the rufous whistler, the species whose distribution was least zero inflated (Fig. 2). In general, the Poisson model was over-confident regarding the uncertainty (smaller credible intervals) and in the case of the superb fairy-wren led to a significantly different conclusion regarding the impact of high grazing on its relative mean abundance.

Modelling influences on woodland bird patch occupancy when patch occupancy observations are subject to false-zero errors

To illustrate the use of the ZIB model, we analyse site occupancy data and investigate influences of habitat type and landscape metrics (patch area and connectivity), on site occupancy rates for four woodland bird species in the Mt Lofty Ranges (MLR) in south-eastern Australia. We compare the inference resulting from the application of the standard logistic regression model with that resulting from the use of a generalized ZIB model.

The ZIB model

Under imperfect detection, site occupancy data are best thought of as realizations of two binomial processes acting simultaneously at two different time scales (MacKenzie *et al.* 2002; Tyre *et al.* 2003). The first process influences p , the probability of a site being occupied over a relatively long-time period. The second process influences the detectability q , the probability of observing the species in a particular visit (or survey) at a site, given that it is present over the longer time period. The survey period may be comprised of 1, 2, ..., v visits. The outcome of the two processes is a finite mixture distribution known as the ZIB mixture model (Hall 2000). Failure to detect the species can occur because the species is absent (occurring with probability $1 - p$) or it is present and remains undetected over the v visits [arising with a probability $p(1 - q)^v$]. When the species is present at the site and detected, the actual number of observations is drawn from a binomial distribution. Thus, ignoring the influence of covariates, the ZIB model is:

$$\begin{aligned} \Pr[Y = 0] &= 1 - p + p(1 - q)^v, & y = 0 \\ \Pr[Y = y] &= p \binom{v}{y} q^y (1 - q)^{v-y}, & y \geq 1, \end{aligned} \quad (3)$$

where y is the number of detections in v visits to a site, and p and q are defined as above. The model may be easily generalized to allow covariates to influence p and q as in a logistic regression. Tyre *et al.* (2003) presents a maximum likelihood implementation of that extension in R (R Development Core Team 2005) and MacKenzie *et al.* (2002) do so in PRESENCE, and Wintle *et al.* (2005) present a Bayesian version using WinBUGS (Spiegelhalter *et al.* 2003). Note that the maximum likelihood version of the ZIB model cannot be estimated unless two or more visits are undertaken on at least some of the survey sites.

The species, study site and data collection

The MLR of South Australia is a highly fragmented landscape with only 14% of its original 686 000 ha area now covered by native vegetation. The MLR is an area of national conservation significance with numerous bird species threatened by loss and fragmentation of habitat (Paton *et al.* 1994; Garnett & Crowley 2000). The bird community is the subject of a multispecies recovery plan and planning for large-scale reinstatement of habitat is a high research priority for the region (Westphal *et al.* 2003). In order to target management and restoration efforts most effectively, it would be useful to investigate how occupancy rates of various species depend on local habitat and landscape characteristics.

To this end we modelled the effect of habitat type, patch area and landscape connectivity on occupancy levels of four MLR bird species of conservation concern: the scarlet robin,

Table 4 Five alternative models for explaining the probability of woodland bird species occurrence, where Habitat defines the type of vegetation at the survey location, 'Area' is the area of the patch in which the survey site is located and 'Connectivity' is a connectivity measure calculated according to Moilanen & Nieminen (2002, eqn 3a)

Model 1	$\text{logit}[\text{Pr}(Y = 1)] = \beta_0$
Model 2	$\text{logit}[\text{Pr}(Y = 1)] = \beta_0 + \beta_1 \times \text{Habitat}$
Model 3	$\text{logit}[\text{Pr}(Y = 1)] = \beta_0 + \beta_2 \times \text{Habitat} + \beta_3 \times \text{Area}$
Model 4	$\text{logit}[\text{Pr}(Y = 1)] = \beta_0 + \beta_2 \times \text{Habitat} + \beta_3 \times \text{Connectivity}$
Model 5	$\text{logit}[\text{Pr}(Y = 1)] = \beta_0 + \beta_2 \times \text{Habitat} + \beta_3 \times \text{Area} + \beta_4 \times \text{Connectivity}$

The variable 'Habitat' is a binary variable where a value of 0 indicates stringy-bark eucalyptus woodland vegetation and a 1 indicates gum-bark eucalyptus woodland vegetation.

Petroica multicolor, buff-rumped thornbill *Acanthiza reguloides*, white-throated tree creeper, *Cormobates leucophaeus*, and rufous whistler *Pachycephala rufiventris*. Three-repeat surveys (20 min–2 ha active timed searches; Loyn 1986; Field *et al.* 2002) were conducted at each of 155 forest and woodland sites during the main breeding season (September to December) in 2003. To model the effect of habitat, sites were classified by major habitat type as either 'stringybark' (canopy dominated by *Eucalyptus obliqua*, *Eucalyptus baxteri*) or 'gum' (*Eucalyptus leucoxyton*, *Eucalyptus viminalis*, *Eucalyptus fasciculosa*, *Eucalyptus goniacalyx*). To model landscape characteristics, the area of each patch containing a survey site was obtained from a GIS, and connectivity was calculated according to Moilanen & Nieminen (2002). A subset of possible combinations of habitat, area and connectivity variables yielded five candidate models (Table 4).

A generalized ZIB model for woodland bird occupancy data

The standard approach to modelling the influence of landscape and habitat attributes on the probability of occupancy (p) at a given site is to use a logistic regression (McCullagh & Nelder 1989) such that:

$$p = \frac{\exp(\alpha + \beta'X)}{1 + \exp(\alpha + \beta'X)}, \quad (4)$$

where α and the vector β are the regression coefficients and the vector X represents the values of the independent environmental variables influencing p . This model assumes that the observations, Y , are realizations of independent Bernoulli trials with event probabilities p . However, because our data contain multiple (3) visits to 155 sites in the model fitting data set, it is possible to embed eqn 4 in eqn 3, allowing simultaneous estimation of regression coefficients β and the detection probability parameter q . The combination of eqns 3 and 4 may be thought of as a generalization of the ZIB model that allows unbiased estimation of habitat model coefficients β .

The generalized ZIB model and the standard logistic regression model were fitted to each of the five candidate models (Table 4) using WinBUGS. Non-informative normal

priors with a mean of zero and precision parameters equal to 0.0001 were used (see Appendix S2 for code). DIC statistics were calculated for each model and used to compare the five competing models (Spiegelhalter *et al.* 2002). Convergence was achieved after a burn-in of 10 000 iterations. Estimates were obtained after a further 30 000 iterations. Convergence of the Markov chains was examined using CODA package (Best *et al.* 1995). For the purpose of this paper we were primarily interested in the difference in inference obtained under the two types of model.

Results

The four bird species showed varying responses to woodland vegetation types and landscape attributes. Best models, determined on the basis of DIC included the variable 'Habitat', with white-throated treecreepers strongly preferring stringybark woodland, scarlet robins showing a similar but weaker preference for stringybark and both buff-rumped thornbills and rufous whistlers displaying a moderate preference for gum woodland (Table 5). On the basis of our results, only one of the species, the scarlet robin, was strongly influenced by habitat area (Table 5 and Fig. 4a) and only one species, the white-throated tree creeper, was strongly influenced by patch connectivity (Table 5). Single visit detection probabilities (q) for all species ranged from c 0.24 (rufous whistler) to 0.61 (white-throated tree creeper) (Table 5).

According to model DICs, the best standard logistic model was always the same as the best generalized ZIB model in terms of which variables were most important. This may be the result of assuming ' q ' was equal across covariates, hence the model likelihood for the ZIB was proportional to the logistic regression likelihood. An alternative approach is to model ' q ' as a function of covariates, allowing factors that affect occupancy to be teased apart from those that affect detectability (MacKenzie 2005).

Regardless, both the magnitude of the effect and their credible intervals were always greater in the ZIB model (Table 5). In other words, using the logistic regression failed to account for the zeros generated by false absences,

Table 5 The favoured model (M) for each species with habitat and landscape coefficient posterior estimates and 95% credible interval (CI) for the standard logistic (binomial) and generalized zero-inflated binomial (ZIB) model

Species	M	Variable	Posterior coefficient estimate and 95% CI	
			Standard logistic	Generalized ZIB
Scarlet Robin ($q = 0.336$)	3	Habitat	-1.146 (-1.909 to 0.424)	-1.61 (-3.218 to -0.061)
		Area	0.180 (0.022 to 0.344)	0.258 (0.047 to 0.568)
Rufous whistler ($q = 0.243$)	2	Habitat	0.909 (0.058 to 1.817)	1.133 (0.079 to 2.796)
White-throated treecreeper ($q = 0.611$)	4	Connectivity	0.167 (-0.018 to 0.354)	0.189 (-0.059 to 0.449)
		Habitat	-2.932 (-3.903 to -2.066)	-3.674 (-6.339 to -2.374)
Buff-rumped thornbill ($q = 0.311$)	2	Habitat	1.438 (0.694 to 2.233)	1.876 (0.863 to 5.77)

The favoured model presented for each species is the best of the five competing models (Table 4) on the basis of deviance information criterion values and q is the detection probability.

resulting in a consistent underestimation of both the mean and variance of model effects. This result corroborates the findings of Tyre *et al.*'s (2003) simulation study. Inference based on standard analyses could therefore be erroneous, and, if used for conservation planning purposes, lead to misdirected management actions. For example, if a set of occupancy models were used to underpin multispecies habitat reconstruction planning (e.g. Westphal *et al.* 2003), mis-specification of the type, amount and connectivity of habitat required for each species could result in suboptimal allocation of reconstruction effort across the landscape.

CONCLUSION

Understanding how zeros arise and what types of zeros occur in ecological data are more than just semantics; failing to model zeros correctly can lead to impaired ecological understanding. In this paper, we have categorized the different sources of zeros in ecological data and discussed the statistical tools available for modelling zero inflated presence/absence and count data. We have demonstrated that failing to properly account for zero inflation as a result of false-zero and excess true-zero observations can lead to substantially different parameter and precision estimates.

Excess true zeros are a result of ecological processes. This may be due to species rarity (Gaston 1994) or the result of habitat condition. In our first example the number of birds counted (i.e. abundance) is of direct interest and the mixture component of the ZIP represents a trigger point in habitat condition where the probability of the number of birds present goes from zero to a Poisson distribution. The ecological interpretation is subtly (but importantly) different to the case where the entire zero-class was modelled separately as in a two-part modelling approach.

In the second example, we are interested in the presence or absence of the bird species. As this is simply a binary response, it cannot be zero inflated as a result of 'true zeros' and the zero inflation occurs through the sampling process

of the number of times the species is not detected when it is present, hence 'false zeros'.

Situations where one may have both 'true' and 'false zeros' occur when extra zeros may be included in the data set from the ecological and sampling processes (e.g. there may be a set of minimal conditions for any members of the species to be present at the site, and then when the population is surveyed, the number of individuals actually present is undercounted because of imperfect detection).

Our worked examples show that the choice of modelling approach influences the predictive performance and hence ecological inference. The degree to which inference changes necessarily depends on the amount of zero inflation present in the data. In the examples presented here, the magnitude of zero inflation for three of the four bird species examined in the grazing effects study was substantial and not accounting for zero inflation as a result of excess true zeros led to an underestimation of the variance and hence overconfidence in the certainty of the estimates or increased risk of type I error. Similarly, in the habitat modelling bird example, when the probability of detection is < 1 , or in other words when the data set contains false zeros, not accounting for the resultant zero inflation leads to both an underestimation of the size of effects and their variance or increased risk of type II error (Tyre *et al.* 2003; Wintle *et al.* 2004).

To the extent that management decisions are made within a hypothesis-centred statistical framework, higher type I and type II error rates, respectively, could result in financial losses because of erroneous interventions (type I) or failures to take action (type II) (Mapstone 1995; Dayton 2001; Field *et al.* 2004). Under a scientific framework that relies on model-based inference (Burnham & Anderson 2002), biased estimation of ecological effects can lead to erroneous model selection, predictions and conclusions regarding ecological effects. Given the considerable focus on modelling rare species and/or rare events in ecology and conservation biology it is imperative that the source of zero observations

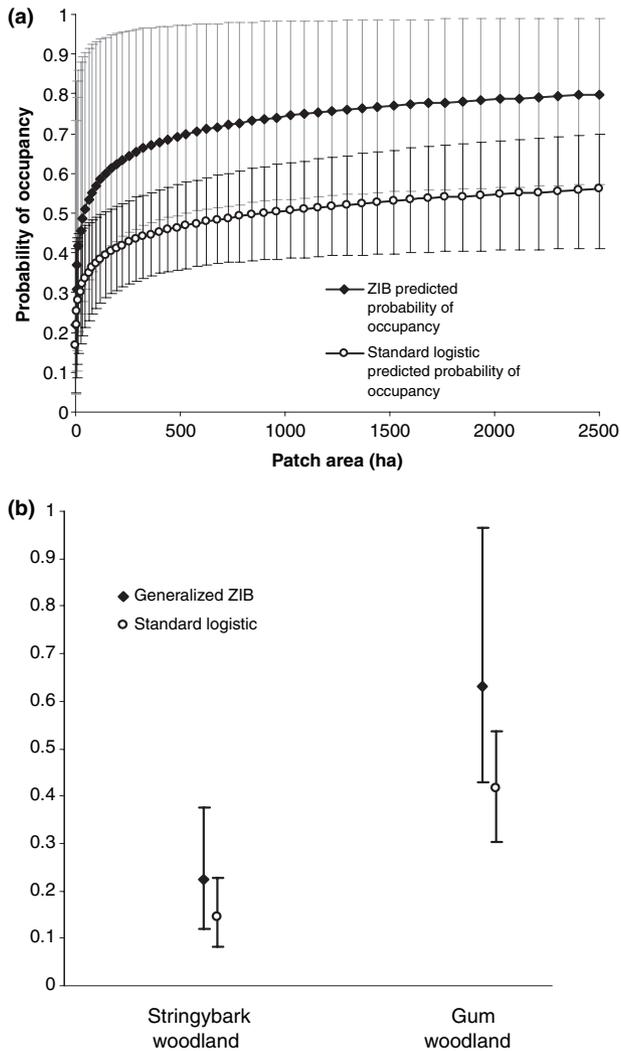


Figure 4 Predicted probabilities of occupancy of (a) the scarlet robin and (b) the buff-rumped thornbill in the Mt Lofty Ranges based on patch area and habitat coefficient estimates obtained from the generalized zero-inflated binomial and standard logistic models. The probability of scarlet robin site occupancy increases with the area (ha) of contiguous vegetation surrounding a site. The probability of buff-rumped thornbill site occupancy is approximately three times higher in gum woodlands than stringybark woodlands under both types of model.

be considered and modelled accordingly, or we risk making incorrect inferences about the species and events that are of greatest conservation concern.

ACKNOWLEDGEMENTS

We are grateful to Darryl MacKenzie and three anonymous referees whose comments improved this paper. This paper developed in part out of the Optimal Monitoring Work-

shop, sponsored by The Ecology Centre, University of Queensland, ARC grant to H. P. Possingham and M. A. McCarthy. Financial support was provided by Birds Queensland, Birds Australia Stuart Leslie Research Award, APA to TM, APAI to JRR, ARC grant (LP0347473) to BW, and logistical support by CSIRO Sustainable Ecosystems, Rangelands and Savannas Programme.

REFERENCES

- Aitchison, J. (1955) On the distribution of a positive random variable having a discrete probability mass at the origin. *J. Am. Stat. Assoc.*, 50, 901–908.
- Akaike, H. (1973) Information theory and an extension of the maximum likelihood. In: *2nd International Symposium on Information Theory*. (eds B.N. Petrov and F. Csaki). Akademia Kaido, Budapest, pp. 267–281.
- Angers, J.-F. & Biswas, A. (2003) A Bayesian analysis of zero-inflated generalized Poisson model. *Comput. Stat. Data Anal.*, 42, 37–46.
- Austin, M.P. & Meyers, A.J. (1996) Current approaches to modelling the environmental niche of *Eucalypts*: implications for management of forest biodiversity. *For. Ecol. Manage.*, 85, 95–106.
- Barry, S.C. & Welsh, A.H. (2002) Generalized additive modelling and zero inflated count data. *Ecol. Modell.*, 157, 179–188.
- Baum, J.K. & Myers, R.A. (2004) Shifting baselines and the decline of pelagic sharks in the Gulf of Mexico. *Ecol. Lett.*, 7, 135–145.
- Best, N.G., Cowles, M.K. & Vines, S.K. (1995) *CODA: Convergence Diagnostics and Output Analysis Software for Gibbs Sampling Output*, Version 0.3. MRC Biostatistic Unit, Cambridge, UK.
- Böhning, D., Dietz, E. & Schlattmann, P. (1999) The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *J. R. Stat. Soc. A*, 162, 195–209.
- Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multi-model Inference: A Practical Information-theoretic Approach*, 2nd edn. Springer, New York, NY.
- Cameron, A.C. & Trivedi, P.K. (1998) *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.
- Campbell, M.J., Machin, D. & D’Arcangues, C. (1991) Coping with extra Poisson variability in the analysis of factors influencing vaginal ring expulsions. *Stat. Med.*, 10, 241–254.
- Carrivick, P.J.W., Lee, A.H. & Yau, K.K.W. (2003) Zero-inflated Poisson modeling to evaluate occupational safety interventions. *Saf. Sci.*, 41, 53–63.
- Celeux, G., Forbes, F., Robert, C. & Titterton, M. (2003), *Deviance Information Criteria for Missing Data Models*, Technical report 4859. Unité de recherche INRIA Rhône-Alpes, Rhône-Alpes, France.
- Cheung, Y.B. (2002) Zero-inflated models for regression analysis of count data: a study of growth and development. *Stat. Med.*, 21, 1461–1469.
- Clarke, K.R. & Green, R.H. (1988) Statistical design and analysis for a ‘biological effects’ study. *Mar. Ecol. Prog. Ser.*, 46, 213–226.
- Dalrymple, M.L., Hudson, I.L. & Ford, R.P.K. (2003) Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Comput. Stat. Data Anal.*, 41, 491–504.
- Dayton, P.K. (2001) Reversal of the burden of proof in fisheries management. *Science*, 279, 821–822.

- Dobbie, M.J. & Welsh, A.H. (2001) Modelling correlated zero-inflated count data. *Aust. N. Z. J. Stat.*, 43, 431–444.
- Elith, J. & Burgman, M.A. (2002) Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. In: *Predicting Species Occurrences: Issues of Scale and Accuracy* (eds Scott, J.M., Heglund, P.J., Morrison, M., Raphael, M., Haufler, J. & Wall, B.). Island Press, Covello, CA, pp. 303–314.
- Ellison, A.M. (2004) Bayesian inference in ecology. *Ecol. Lett.*, 7, 509–520.
- Faddy, M.J. (1998) Stochastic models for analysis of species abundance data. In: *Statistics in Ecology and Environmental Modelling* (eds Fletcher, D.J., Kavalieris, L. & Manly, B.F.J.). University of Otago Press, Dunedin, pp. 33–40.
- Field, S.A., Tyre, A. J. & Possingham, H.P. (2002) Estimating bird species richness: how should repeat surveys be organized in time? *Aust. Ecol.*, 27, 624–629.
- Field, S.A., Tyre, A.J., Jonzén, N., Rhodes, J.R. & Possingham, H.P. (2004) Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecol. Lett.*, 7, 669–675.
- Field, S.A., Tyre, A.J. & Possingham, H.P. (2005) Optimizing allocation of monitoring effort under economic and observational constraints. *J. Wildl. Manage.*, in press.
- Fletcher, D., Mackenzie, D.I. & Villouta, E. (2005) Modelling skewed data with many zeros: a simple approach combining ordinary and logistic regression. *Environ. Ecol. Stat.*, 12, 45–54.
- Freund, D.A., Kniesner, T.J. & LoSasso, A.T. (1999) Dealing with the common econometric problems of count data with excess zeros, endogenous treatment effects, and attrition bias. *Econ. Lett.*, 62, 7–12.
- Garnett, S.T. & Crowley, G.M. (2000) *The Action Plan for Australian Birds*. Environment Australia, Canberra, Australia.
- Gaston, K.J. (1994) *Rarity*. Chapman and Hall, London.
- Ghahramani, M., Dean, C.B. & Spinelli, J.J. (2001) Simultaneous modelling of operative mortality and long-term survival after coronary artery bypass surgery. *Stat. Med.*, 20, 1931–1945.
- Gu, W. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biol. Conserv.*, 116, 195–203.
- Hall, D.B. (2000) Zero-inflated Poisson binomial regression with random effects: a case study. *Biometrics*, 56, 1030–1039.
- Heilbron, D.C. (1994) Zero-altered and other regression models for count data with added zeros. *Biom. J.*, 36, 531–547.
- Hinde, J. & Demétrio, C.G.B. (1998) Overdispersion: models and estimation. *Comput. Stat. Data Anal.*, 27, 151–170.
- Johnson, N.L. & Kotz, S. (1969) *Distributions in Statistics: Discrete Distributions*. Houghton Mifflin, Boston, MA.
- Kery, M. (2002) Inferring the absence of a species – a case study of snakes. *J. Wildl. Manage.*, 66, 330–338.
- Kuhnert, P.M., Martin, T.G., Mengersen, K. & Possingham, H.P. (2005) Assessing the impacts of grazing levels on bird density in woodland habitat: a Bayesian approach using expert opinion. *Environmetrics*, 16, 1–31.
- Lambert, D. (1992) Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics*, 34, 1–14.
- Lee, A.H., Stevenson, M.R., Wang, K. & Yau, K.K.W. (2002) Modeling young driver motor vehicle crashes: data with extra zeros. *Accid. Anal. Prev.*, 34, 515–521.
- Lewsey, J.D. & Thomson, W.M. (2004) The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community. Dent. Oral. Epidemiol.*, 32, 183–189.
- Loyn, R. (1986) The 20 minute search – a simple method for counting forest birds. *Corella*, 10, 58–60.
- MacKenzie, D.I. (2005) Is a resource unit really unused? The effect of, and dealing with, detecting a species imperfectly. *J. Wildl. Manage.*, in press.
- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255.
- MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G. & Franklin, A.B. (2003) Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84, 2200–2207.
- MacKenzie, D.I., Bailey, L.L. & Nichols, J.D. (2004) Investigating species co-occurrence patterns when species are detected imperfectly. *J. Anim. Ecol.*, 73, 546–555.
- Mapstone, B. (1995) Scalable decision rules for environmental impact studies: effect size, type I, and type II errors. *Ecol. Appl.*, 5, 401–410.
- Martin, T.G., Kuhnert, P.M., Mengersen, K. & Possingham, H.P. (2005) The power of expert opinion in ecological models using Bayesian methods: impact of grazing on birds. *Ecol. Appl.*, 15, 266–280.
- McCullagh, P., & Nelder, J.A. (1989) *Generalised Linear Models*, 2nd edn. Chapman and Hall, London.
- Moilanen, A., & Nieminen, M. (2002) Simple connectivity measures in spatial ecology. *Ecology* 83, 1131–1145.
- Mullahy, J. (1986) Specification and testing of some modified count data models. *J. Econom.*, 33, 341–365.
- Paton, D.C., Carpenter, G. & Sinclair, R. (1994) A second bird atlas of the Adelaide region. Part 1: changes in the distribution of birds: 1974–75 vs. 1984–85. *South Aust. Ornithol.* 31, 151–193.
- Podlich, H.M., Faddy, M.J. & Smyth, G.K. (2002) A general approach to modeling and analysis of species abundance data with extra zeros. *J. Agric. Biol. Environ. Stat.*, 7, 324–334.
- Poortema, K. (1999) On modelling overdispersion of counts. *Stat. Neerl.*, 53, 5–20.
- R Development Core Team (2005) R: A language and environment for statistical computing. In: *R Foundation for Statistical Computing*, Vienna Austria. ISBN 3-900051-07-0, URL Available at: <http://www.r-project.org>.
- Rhodes, J.R., Tyre, A.J., Jonzén, N., McAlpine, C.A. & Possingham, H.P. (2005) Optimizing presence/absence surveys for detecting population trends. *J. Wildl. Manage.*, in press.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Stat. Soc. Ser. B*, 64, 583–640.
- Spiegelhalter, D.J., Thomas, A., Best, N.G. & Lunn, D. (2003) *WinBUGS Version 1.4*. Imperial College and MRC Biostatistics Unit, London.
- Stefansson, G. (1996) Analysis of ground fish survey abundance data: combining the GLM and delta approaches. *ICES J. Mar. Sci.*, 53, 577–588.
- Syrjala, S.E. (2000) Critique on the use of the delta distribution for the analysis of trawl survey data. *ICES J. Mar. Sci.*, 57, 831–842.
- Tu, W. (2002) Zero-inflated data. In: *Encyclopedia of Environmetrics* (eds El-Shaarawi, A.H. & Peiegorsch, W.W.). John Wiley and Sons, Chichester.

- Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Paris, K. & Possingham, H.P. (2003) Improving precision and reducing bias in biological surveys by estimating false negative error rates in presence-absence data. *Ecol. Appl.*, 13, 1790–1801.
- Wang, P. (2003) A bivariate zero-inflated negative binomial regression model for count data with excess zeros. *Econ. Lett.*, 78, 373–378.
- Wang, K., Lee, A.H., Yau, K.K.W. & Carrivick, P.J.W. (2003) A bivariate zero-inflated Poisson regression model to analyze occupational injuries. *Accid. Anal. Prev.*, 35, 625–629.
- Warton, D.I. (2005) Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics*, 16, 275–289.
- Welsh, A.H., Cunningham, R.B., Donnelly, C.F. & Lindenmayer, D.B. (1996) Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecol. Modell.*, 88, 297–308.
- Welsh, A.H., Cunningham, R.B. & Chambers, R. (2000) Methodology for estimating the abundance of rare animals: seabird nesting on North East Herald Cay. *Biometrics*, 56, 22–30.
- Westphal, M.I., Pickett, M., Getz, W.M. & Possingham, H.P. (2003) The use of stochastic dynamic programming in optimal landscape reconstruction for metapopulations. *Ecol. Appl.* 13, 543–555.
- Wintle, B.A., McCarthy, M.A., Parris, K.M. & Burgman, M.A. (2004) Precision and bias of methods for estimating point survey detection probabilities. *Ecol. Appl.*, 14, 703–712.
- Wintle, B.A., Kavanagh, R.P., McCarthy, M.A. & Burgman, M.A. (2005) Estimating and dealing with detectability in occupancy surveys for forest owls and arboreal marsupials. *J. Wildl. Manage.*, in press.
- Yau, K.K.W., Lee, A.H. & Carrivick, P.J.W. (2004) Modeling zero-inflated count series with application to occupational health. *Comput. Methods. Programs. Biomed.*, 74, 47–52.

SUPPLEMENTARY MATERIAL

The following supplementary material is available online for this article from <http://www.Blackwell-Synergy.com>:

- Appendix S1** WinBUGS code for the ZIP mixture model.
Appendix S2 WinBUGS code for the generalized ZIB model used when accounting for excess zeros generated by false zeros.

Editor, Marti Anderson

Manuscript received 19 May 2005

First decision made 27 June 2005

Second decision made 2 August 2005

Manuscript accepted 16 August 2005