ecological modelling 199 (2006) 153–163

available at www.sciencedirect.com

journal homepage: www.elsevier.com/locate/ecolmodel

# Comparing species abundance models

*Joanne M. Potts*[a,*], *Jane Elith*[b]

[a] *Arthur Rylah Institute for Environmental Research, Department of Sustainability and Environment,*
*123 Brown St, Heidelberg, Victoria 3084, Australia*
[b] *School of Botany, The University of Melbourne, Parkville, Victoria 3010, Australia*

## ARTICLE INFO

*Article history:*
Published on line 13 July 2006

*Keywords:*
Zero-inflation
Over-dispersion
Poisson regression
Abundance data
Model evaluation

## ABSTRACT

Five regression models (Poisson, negative binomial, quasi-Poisson, the hurdle model and the zero-inflated Poisson) were used to assess the relationship between the abundance of a vulnerable plant species, *Leionema ralstonii*, and the environment. The methods differed in their capacity to deal with common properties of ecological data. They were assessed theoretically, and their predictive performance was evaluated with correlation, calibration and error statistics calculated within a bootstrap evaluation procedure that simulated performance for independent data.

The hurdle model performed best, with the highest correlations between the observed and predicted abundances. This model was also well calibrated, giving the closest agreement between observed and predicted abundances. The negative binomial was the worst performing model. It had weaker correlations than the other models and resulted in a strong, inconsistent bias in predictions. The standard Poisson model which accommodates neither zero-inflation nor over-dispersion gave accurate estimates of regional population abundance, but at the individual population level they were inconsistent and biased.

The strong performance of the hurdle model, coupled with theoretical properties that suit it for these data and for the ecology of this species, suggest that it is a useful alternative to other modelling methods. The gains in performance have practical advantages where predictions are used by conservation planners to understand population dynamics or to assess the relative risks of alternative management scenarios.

## 1. Introduction

Environmental managers require estimates of species abundance in a broad range of situations: classifying species according to the IUCN Red List for threatened species (IUCN, 2001); conducting population viability analyses (Possingham et al., 2001); managing fire regimes (e.g. the endangered shrub *Grevillea caleyi*, Regan et al., 2003); monitoring (e.g. population changes of pest species over time, Hone, 1999); and reintroducing or translocating animals (Lubow, 1996). Obtaining such estimates can be resource demanding because surveys are expensive and time-consuming, especially if the species is rare or occurs in remote locations.

Mathematical models that quantify the relationship between a species' abundance and environmental characteristics may be used to complement survey work. Predictions of abundance can then be made at unsurveyed locations and used to guide management decisions. The choice of a mathematical model should be governed by knowledge of the species and characteristics of the available data. Accommodating characteristics of the data in a model can increase its complexity and thus decrease the ease with which it is

---

[*] *Corresponding author*. Tel.: +61 3 9450 8701; fax: +61 3 9450 8799.
E-mail address: Joanne.Potts@dse.vic.gov.au (J.M. Potts).

0304-3800/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2006.05.025

developed, interpreted and understood. There is a trade-off between complicated models that account appropriately for characteristics of the data and simpler models that are easier to develop but may be sub-optimal.

This study explores these trade-offs using data collected on an Australian threatened plant species, *Leionema ralstonii* (F. Muell.) Paul G. Wilson (Rutaceae). The species is protected under both state and federal legislation due to its small geographical range (~40,000 ha in the southeast corner of New South Wales) and population size (~18,100 individual plants distributed across 71 discrete populations on rocky outcrops, NSW National Parks and Wildlife Service, 2003). Predictions of species abundance were required to guide management of the species and as inputs for a population viability analysis (Potts et al., submitted for publication).

## 2. Technical review

This paper focuses on regression methods within a generalised linear model framework (McCullagh and Nelder, 1989). These types of models are used frequently to quantify the relationship between species abundances and environmental characteristics (e.g. Wintle et al., 2005; Leathwick et al., this issue). Regression models are typically described in terms of their systematic component in which the response is linked to the environmental data, and their stochastic structure that describes the error distribution (Venables and Ripley, 2002). In order to focus on the link function separately to the rest of the systematic component, in this paper we use the terms model structure and model specification. Model structure includes both the choice of environmental characteristics (the explanatory variables) assumed to affect species abundance (the response variables) and the shape of the modelled responses (linear, quadratic and so on). Model specification defines how these variables are related using a 'link' function. The choice of 'link' function allows the response variable to be non-linearly related to the explanatory variables.

When the response variable is count data (as is the case when working with abundance observations), the response variable can be linked to the explanatory variables using a log transformation (McCullagh and Nelder, 1989):

$$\log(p) = \ln(p) = \beta_0 + \beta_1 X_1 + \ldots + \beta_N X_N \tag{1}$$

where $p$ is the probability of an event occurring, $X_N$ the $N$th independent variable and $\beta_N$ is the regression coefficient. In our example, an event is the mean rate at which individuals occur on each outcrop (termed $\mu$). This model is referred to as a standard Poisson regression and is the simplest and most commonly specified model for count data.

This model specification assumes equi-dispersion, meaning if $Y$ is Poisson distributed, then the expectation of $Y$ is equal to the variance of $Y$. Since the variance is not constant, the regression is intrinsically heteroskedastic (i.e. the variance increases with increasing mean). Violating the assumption of equi-dispersion has similar consequences to violating the assumption of homoskedasticity in linear regression (Cameron and Trivedi, 1998). The standard errors of the predictions are biased because the different populations have different variances.

If the variance exceeds (or is less than) the mean, then the data are said to be over- (or under-) dispersed (Cox, 1983). An indication of the magnitude of over- or under-dispersion can be obtained by comparing the sample mean and variance of the dependent count variable. Over-dispersion can be reduced using explanatory variables. When working with ecological data the equi-dispersion assumption is commonly violated, especially if the data are zero-inflated (Cameron and Trivedi, 1998).

Zero-inflated data contain substantially more zeros that the specified distribution suggests (Tu, 2002). They occur because the data generating process adds an additional mass at zero, inflating the probability of observing a zero above that which is consistent with the specified distribution. It may therefore be a mis-specification to assume that the zero and non-zero observations come from the same source. Visual inspection of a histogram of the observed data might suggest a spike of zero observations if zero-inflation is present.

Count data for rare species commonly are zero-inflated. The species may be observed absent at many sites because of true negative or false negative observations (Martin et al., 2005b). We can think about these in terms of the source of the error (i.e. the uncertainty). True negative observations are attributable to structural zeros (i.e. unsuitable habitat) or environmental process (i.e. suitable but unoccupied habitat because the species does not saturate its environment). The latter are also known as stochastic zeros. False negatives are attributable to experimental design (i.e. survey site is utilised by the species, but not during the survey period) or observer error (i.e. species is present but not detected). If not modelled properly, the presence of excess zeros can violate the distributional assumptions of the analysis, lead to invalid scientific inferences and create computational difficulties (Tu, 2002). Zero-inflation may cause over-dispersion, but it is possible for either of these two features to occur independently in any data set. Formal statistical tests are available for both equi-dispersion and zero-inflation (see Cox, 1983; Böhning, 1994; van den Broek, 1995; Ridout et al., 2001; Hall and Berenhaut, 2002).

If a data set is zero-inflated and/or violates the equi-dispersion assumption, the standard Poisson regression is still commonly used (Cameron and Trivedi, 1998). We believe this is because the Poisson model is easy to implement and available in a number of statistical packages. Incorrectly specifying a Poisson distribution in the presence of zero-inflation and/or over-dispersion has two important consequences. Firstly, it will result in incorrect predictions at each site, although the average prediction across all sites will be consistent with that observed (Cameron and Trivedi, 1998; Barry and Welsh, 2002). Secondly, it will cause overly optimistic conclusions about the statistical significance of the explanatory variables (i.e. reduced standard errors of the coefficients). This means that under common model-building procedures such as stepwise selection, incorrect variables are more likely to be retained (Fitzmaurice, 1997). Both of these consequences are important for environmental managers, as either the predictions and/or the model structure may influence decision making.

Appropriate alternatives to standard Poisson regression are applied in medicine, econometrics and statistics. For example, quasi-likelihood GLMs allow for non-standard response distributions, relying on correct specification of mean and variance relationship (i.e. $\mathrm{var} = \Phi\mu$, Wedderburn, 1974). Here, the quasi-Poisson model specifies the dispersion parameter, $\Phi$ (=variance/mean), such that $\Phi$ will be greater than one if over-dispersion is present (Wedderburn, 1974; McCullagh and Nelder, 1989; Nelder, 2000). Other models, collectively called mixture models, separately account for the features of the data by combining (or mixing) two distributions (Tu, 2002). These are discussed below.

### 2.1. *Negative binomial*

The negative binomial (NB) distribution was derived from a Poisson-gamma mixture (Greenwood and Yule, 1920; McCullagh and Nelder, 1989) but is now considered a distribution in its own right. This model is the standard parametric model used to account for over-dispersion, as the variance function ($\omega_i$) is a multiple of the mean:

$$\omega_i = \mu_i + \alpha\mu_i^p \tag{2}$$

Restricted cases occur when $p = 1$ and $p = 2$ (called the NB1 and NB2 models, respectively) and where $p = 0$ (the Poisson model). The NB model does not explicitly account for zero-inflation.

### 2.2. *Zero-inflated models*

The foundation for the entire class of zero-inflated models is the mixing of a common parametric distribution (e.g. a Poisson distribution) with one degenerate at zero (i.e. all observations are zero) (Tu, 2002). If the parametric distribution is specified as a Poisson, a zero-inflated Poisson (ZIP) model results (Umbach, 1981; Dietz and Böhning, 2000). This was developed by Lambert (1992) to model the number of defects produced during a manufacturing process, and is defined as:

$$Y_i \sim \begin{cases} 0, & \text{with probability}\, p_i \\ \text{Poisson}\,(\lambda_i), & \text{with probability}\; 1 - p_i \end{cases} \tag{3}$$

where $Y_i$ is the number of defects on the *i*th product, $i = 1, \ldots, n$. This model structure implies the zero observations may arise with probability $p_i$ from the distribution degenerate at zero (thus implying the zero is a true negative observation). Alternatively, the zero observation may arise from the Poisson distribution predicting a zero observation with probability $1 - p_i$. This implies the zero observation is a false negative observation, that is, the habitat is suitable but unoccupied (Fig. 1, Lewsey and Thomson, 2004). The ZIP model will only account for zero-inflation, it will not explicitly account for over-dispersion. However, if over-dispersion is caused by the numerous zero observations, the level of over-dispersion may be reduced once zero-inflation is correctly accounted for.

Zero-inflation can affect other parametric distributions. For example, Cheung (2002) analysed fine motor development in children using a zero-inflated negative binomial distribution for count data, and Hall (2000) demonstrated the zero-inflated binomial distribution to model the maximum number (upper bound) of adult insects on a plant.

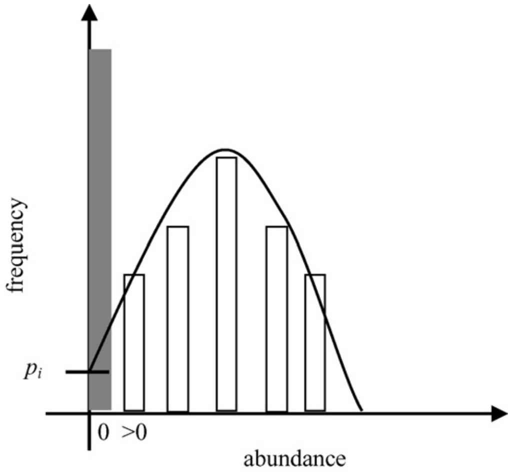

**Fig. 1 – Typical histogram of zero-inflated data, shaded bar represents zero observations, non-shaded bars represent positive observations. Overlayed is the probability distribution function of a standard Poisson distribution. Notice the number of observed zeros is much greater than the proportion expected ($p_i$). The proportion of zero observations below $p_i$ are those expected to be absent, as predicted by the Poisson model (i.e. false negatives). The proportion of zero observations above $p_i$ are those expected to be absent, as predicted by the distribution degenerate at zero (i.e. true negatives).**

Zero-inflated Poisson models can be difficult to implement in standard statistics packages and may require additional programming. Some of the few published examples of zero-inflated models include assessing the impacts of grazing on bird densities (Kuhnert et al., 2005; Martin et al., 2005b), prevention in dental epidemiology (Böhning et al., 1999), evaluation of an occupational injury prevention program (Yau and Lee, 2001) and investigation of manual handling injuries and their relationship to exposure (Lee et al., 2001).

### 2.3. *Hurdle model*

The hurdle model, developed by Cragg (1971), consists of two parts. The first part is a binary (presence/absence) outcome model (e.g. a logistic regression):

$$\mathrm{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_N X_N \tag{4}$$

where $p$ is the probability of an event occurring, $X_N$ the Nth independent variable and $\beta_N$ is the regression coefficient (McCullagh and Nelder, 1989). The second part is a count model that can either accommodate zero observations (e.g. a Poisson model, Eq. (1), Zorn, 1996) permitting the stochastic process resulting in unoccupied sites, even when habitat is suitable (Tyre et al., 2001) or not (e.g. a truncated-at-zero Poisson distribution, Barry and Welsh, 2002). The positive observations arise from crossing the zero-hurdle or probability threshold. It is important to note the probability threshold does not need to be zero (Cameron and Trivedi, 1998). This approach recognises the possibility that the mechanisms that determine presence can be different to those that determine abundance (Ridout et al., 1998).

In contrast to the zero-inflated models, the hurdle model assumes that the zero observations are all true negatives. That is, a zero observation is observed because the species can never occur there (i.e. unsuitable habitat). Ecological examples in the literature of this type of model are more common. They include modelling the abundance of the Leadbeater's Possum in montane ash forests of south-east Australia (Welsh et al., 1996; Faddy, 1998; Podlich et al., 2002), estimating the number of seabird nests on north-east Herald Cay (Welsh et al., 2000), modelling the number of *Eucalyptus mannifera* stems across temperature and rainfall gradients (Barry and Welsh, 2002) and assessing the impacts of grazing on bird densities (Kuhnert et al., 2005; Martin et al., 2005a,b).

The choice of which model specification to use should be governed by characteristics of the data. In situations of both zero-inflation and over-dispersion there are theoretical grounds for preferring models that deal with both these characteristics. However, the question is whether the more complex models improve predictive performance. We will explore the empirical properties of these models and in doing so, look for generalisations about their suitability for rare plant habitat modelling.

## 3. Methods

### 3.1. Target species and distribution

*L. ralstonii* (F. Muell.) Paul G. Wilson (Rutaceae) is an open woody shrub that grows to approximately 2 m high. It is endemic to a 40,000 ha area of south-east New South Wales (Fig. 2), and is confined to dry rocky habitats. Even though this species was first recorded in the 1860s (Albrecht, 1986), knowledge about its distribution and ecology is still limited. It is categorised as threatened because it occurs over a limited geographic extent (NSW National Parks and Wildlife Service, 2003). It inhabits rocky outcrops and can extend a short distance into the ecotonal forest around them. It is thought to be restricted to outcrops because they offer a refuge from competition, predation, fire, human activities or climatic change in the surrounding landscape.

There are approximately 200 outcrops varying in size from 2500 to 250,000 $m^2$ in the study region. Until recently only about 20% of these had been surveyed, because many are not easily accessible and there are limited resources to conduct surveys. Government agencies require information on the distribution of the species over the whole area to help them manage the species in the context of a multiple-landuse region. The outcrops are currently not threatened by direct habitat destruction because they are unsuitable for agriculture and forestry. However, surrounding land use may increase the risk of fire (from burning of logged sites), may alter patterns of soil and water movement and may increase the risk of disease spread.

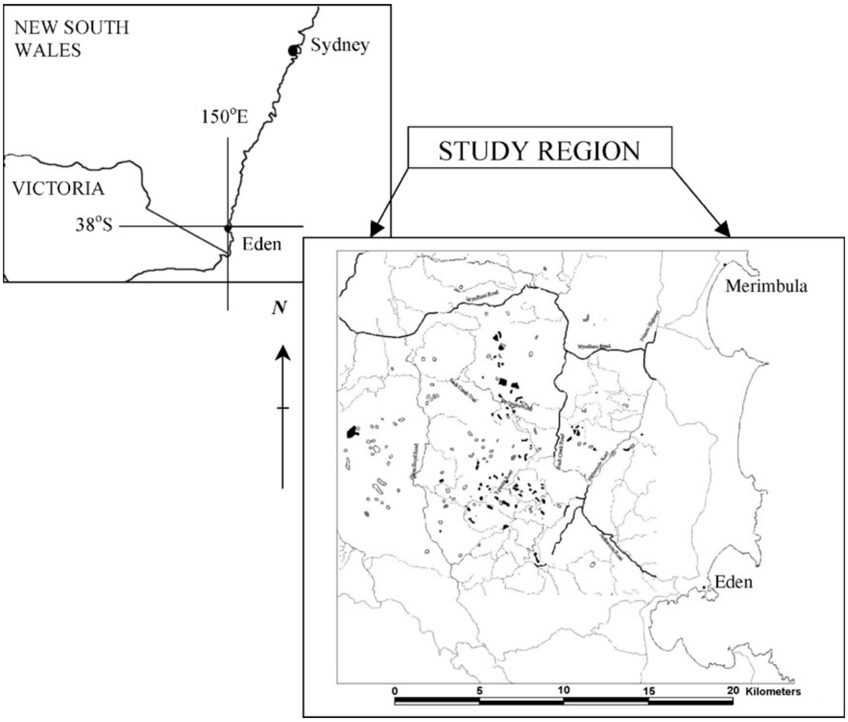

**Fig. 2 – Distribution map of *Leionema ralstonii*. Solid black and grey outcrops indicate species presence and absence, respectively. Black outline outcrops have not been surveyed.**

### 3.2. *Data set*

This species is known to be restricted to rocky outcrops (NSW National Parks and Wildlife Service, 2003), so we first created and validated a map of outcrops through Aerial Photo Interpretation and ground truthing (Elith, 2002). Two hundred and ten outcrops were mapped in the region. Over the course of a 3-year period, we visited 85 previously unsurveyed outcrops and recorded species abundance (Elith, 2002; Potts and Elith, unpubl. data). A further 29 outcrops had previously been surveyed by other researchers (Albrecht, 1986; Binns and Kavanagh, 1990; NSW National Parks and Wildlife Service, 2003), and we revisited five of these (covering a range of survey year, surveyor and geographic position) to compare our abundance estimates with the existing ones. At the end of this work, a total of 114 outcrops had been surveyed for species abundance, and *L. ralstonii* occurred on 76 of them.

We used a pilot study to identify a counting method that gave a reasonable compromise between accuracy and time. Searching can be difficult because the outcrops can be large and steep, with the plants scattered around them and at their base (sometimes in dense patches). Two people spent at least 20 min 10,000 $\mathrm{m}^{-2}$ of outcrop. Searching was directed first towards the more exposed areas and to those patches with plants commonly found in association with the target species. Once the shrub was found, individuals were counted if scattered over the area, or, if in a patch, the number in a typical quarter of the patch was counted and extrapolated to the whole patch. After a maximum of 5 h, if the whole outcrop had not been surveyed, an estimate was made of the unvisited proportion, and the count estimate was adjusted for unsurveyed areas. In an effort to reduce error, we first worked together on several outcrops and recorded counts in the same areas and compared estimates. If there were large discrepancies we did more counting and less estimation until we were confident of reasonable estimates. At the end of this procedure we were satisfied that the estimates of numbers were consistent and reasonably accurate.

The environmental data used in the modelling comprised variables that were previously found to be important predictors of the distribution and abundance of rocky outcrop species (Elith, 2002; Potts and Elith, unpubl. data, Table 1). The rainfall variable was constructed using long-term average annual rainfall data available in the program ANUCLIM (CRES, 2002), interpolated to the region via a DEM with 25 m grid cells. Rainfall varies across the study area and is correlated with elevation and denser vegetation. This appears to affect whether the outcrops form the dry, open habitat that the species seems to prefer. Outcrop area is important because larger outcrops can maintain larger populations of *L. ralstonii*. The number of neighbouring outcrops is a useful predictor, probably because locally frequent outcrops increase the likelihood that seed will reach the site.

Once all data were collected, exploratory data analysis identified characteristics that needed to be accounted for in the model specification process. The presence of over-dispersion in the *L. ralstonii* data set was assessed by computing the ratio between the mean and variance of the data. If the data were over-dispersed, the variance would be much greater than the mean. Zero-inflation was investigated by visually inspecting a histogram of the data.

### 3.3. *Model fitting*

Model fitting involves determining both model structure and specification. Here, model structure (choice of explanatory variables) was fixed for each model (Table 1) so that differences between the models and their predictions could be attributed to differences in model specification. The number of variables is small because the species is rare, and there are not enough data to support many variables (Harrell, 2001). The particular variables used in this study have been found previously to be related to the distribution of plant species in these rocky outcrops (Elith, 2002).

To investigate the consequences of incorrect model specification, five models were chosen: Poisson, negative binomial, quasi-Poisson, hurdle and the zero-inflated Poisson (Table 2). The alternative models represent different degrees of accessibility for environmental modellers and accommodate different features (i.e. over-dispersion and zero-inflation). All models were implemented in R (R Development Core Team, 2004), see Appendix A for relevant packages and for specific settings.

### 3.4. *Model evaluation*

There were insufficient data for this species to have an evaluation data set that was independent of the training (modelling) data, so a bootstrap evaluation procedure proposed by Efron and Tibshirani (1997) was used to assess predictive accuracy. Bootstrapping provides a realistic estimate of the predictive performance of a model, without incurring the expenses of collecting a completely new model-testing data set. It involves resampling the modelling data and conducting a series of model building and testing simulations that provide an estimate of the optimism arising from in-sample validation. The estimate of optimism is used to provide an adjusted estimate of the model evaluation statistics (see below). The bootstrapping version implemented here is believed to provide the least biased estimate of predictive performance of any of the model evaluation methods that are

**Table 1 – The definitions, and form, of the variables included in the models**

| Variable | Definition | Form |
|---|---|---|
| Logarea | Natural log of the area of outcrop ($\mathrm{m}^2$) | Linear |
| Rainfall | Mean annual rainfall at outcrop | Quadratic |
| Lroutcrop | Number of outcrops with *L. ralstonii* present in a 400 m radius | Quadratic |

**Table 2 – Five model specifications were used to investigate the consequences on predictions when working with zero-inflated and over-dispersed data**

| Number | Model specification | Explanation |
|---|---|---|
| 1 | Standard Poisson | Estimates mean and variance from the data, therefore does not account for zero-inflation or over-dispersion. |
| 2 | Negative binomial | Estimates the mean from the data, and sets the variance to be a multiple of the mean, accommodating for over-dispersion. |
| 3 | Quasi-likelihood | Accommodates for over-dispersion by specifying a relationship between the mean and variance, using a dispersion parameter, as determined from the data. |
| 4 | Hurdle model | Removes effect of zero-inflation in the presence/absence model and over-dispersion in the non-zero observations using a quasi-likelihood. The second stage of this model accommodates zero observations, permitting the stochastic process resulting in unoccupied sites, even when habitat is suitable (Tyre et al., 2001). |
| 5 | Zero-inflated Poisson | Removes effect of zero-inflation by having a zero distribution, possibly reducing effect of over-dispersion. Estimates mean and variance of non-zero observations from data. |

based on re-sampling, including cross-validation (Hastie et al., 2001). Cross-validation (Efron and Tibshirani, 1997) provides an alternative approach to model evaluation, and might be more feasible with methods or data sets that create large computational loads, and more appropriate with larger data sets. However, its estimates of error rates with independent data can be less precise than those derived from bootstrapping (Steyerberg et al., 2001), which can be thought of as a smoothed version of cross-validation (Efron and Tibshirani, 1997). The bootstrapping method (the 0.632+ bootstrap) is detailed in Appendix B. Here we have used a range of evaluation statistics, to clarify different aspects of the predictive performance:

*Pearson's correlation coefficient*, $r$, provides an indication of how closely the observed and predicted values agree in relative terms, though a perfect correlation (=1) does not imply exact predictions—all predictions may be biased in a consistent direction.

*Spearman's rank correlation*, $\rho$, provides an indication of similarity between the ranks of the observed and predicted values. It also varies from −1 to +1. The use of ranks means that, as long as the order of the predictions is correct, the statistic will be high.

*Model calibration* was assessed by fitting a simple linear regression between the observed and predicted values (e.g. observed = $m$(predicted) + $b$), providing information on the bias and consistency of the predictions. This is comparable to calibration analyses for binomial data described by Pearce and Ferrier (2000). The intercept term in this model ($b$) provides an indication of the bias, and the gradient of the fitted line ($m$) provides an indication of the distribution (spread) of predictions over their numerical range compared with the spread of the observations. That is, for a perfectly calibrated model, $b$ should equal zero and $m$ should equal one (line A, Fig. 3). In a model with consistent bias, $m$ equals one, and $b$ will be a non-zero number (line B, Fig. 3). In a model with both bias and predictions that are spread over a larger range than the observations, $b$ will not equal zero and $m$ will be less than (or greater than) one (line C, Fig. 3). This will result in larger errors between the observed and predicted values at the extremes of the observations.

*Root mean square error* (RMSE) and the *average error* (AVE) depend on the sample size ($n$), and the discrepancy between the observed ($y_i$) and predicted ($\hat{y}_i$) values.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{5}$$

$$\mathrm{AVE}_{\mathrm{error}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{6}$$

All the statistics outlined above were calculated across all outcrops. To better understand what was happening at the individual population level, ranks were used to compare the relative performance of the different methods at each outcrop. That is, for each outcrop the predictions from the five methods were ranked from best (score = 1, lowest discrepancy between observed and predicted values) to worst (score = 5, highest discrepancy). This provided an indication of how accurate each method was in their predictive ability. If a method was consistently accurate, its average rank score over all outcrops was low.

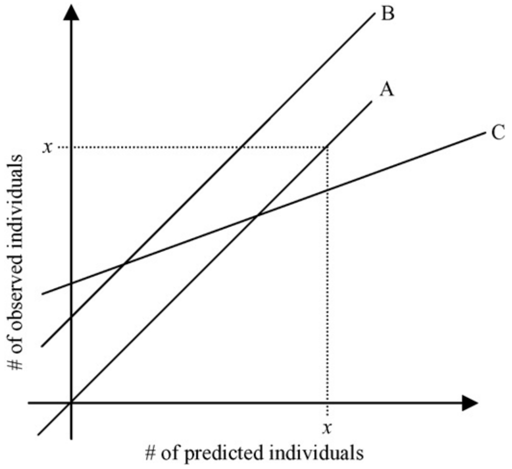

**Fig. 3 – A graph of observed values against predicted values (A) model with perfect calibration (i.e. observed = predicted); (B) model with consistent bias (i.e. observed = predicted + constant, line is parallel to line A); and (C) model with inconsistent bias.**

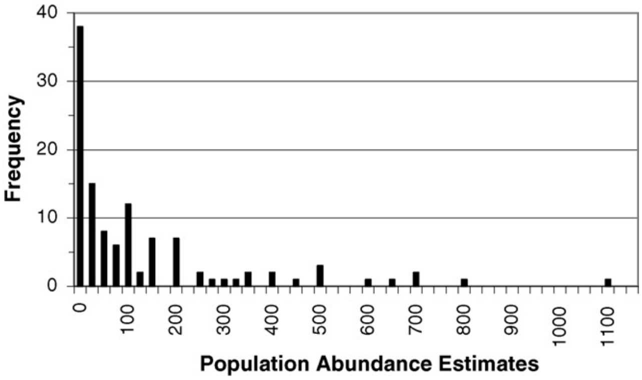

**Fig. 4 – A histogram of the observed population abundance estimates. The spike at an abundance of zero indicates that there are many zero observations, suggesting the standard Poisson distribution may not be appropriate.**

## 4. Results

### 4.1. *Survey data*

The 114 outcrops had from zero to 1100 individuals (Fig. 4). The spike at zero indicated zero-inflation. The data were overdispersed, because the variance was much greater than the mean of the data ($\mathrm{var}(Y) = 38{,}666 \gg E(Y) = 125$).

### 4.2. *Model fitting*

The between-model effects of model specification may be assessed by examining the coefficients and their standard errors (Table 3). Coefficients are most similar for the Poisson, quasi-Poisson and NB models; however, their standard errors increase progressively as they more realistically model the over-dispersion present in the data. These standard errors are derived parametrically, and so will underestimate the true error around the estimate if the model is misspecified for the amount of overdispersion. Small standard errors do not indicate a more accurate coefficient, but a false view of the uncertainty of the estimate. We have confirmed this with bootstrap estimates of the standard errors. Interestingly, the parametric standard errors for the ZIP model are approximately equal to those of the Poisson model. This suggests that accommodating the extra zeros is not dealing with the over-dispersion.

**Table 4 – The evaluation statistics for the five model specifications**

| Model | $r$ | $\rho$ | Model calibration | | RMSE | $\mathrm{AVE}_{\mathrm{error}}$ |
|---|---|---|---|---|---|---|
| | | | $b$ | $m$ | | |
| Poisson | 0.52 | 0.65 | 24.49 | 0.82 | 184.83 | 0.00 |
| NB | 0.38 | 0.66 | 72.08 | 0.28 | 348.71 | 64.12 |
| Quasi | 0.52 | 0.66 | 24.21 | 0.84 | 184.65 | 0.00 |
| Hurdle model | 0.58 | 0.66 | 7.32 | 0.86 | 173.59 | 12.78 |
| ZIP | 0.40 | 0.28 | 1.01 | 1.04 | 192.18 | −15.63 |

$r$ is the Pearson's correlation coefficient. $\rho$ is the Spearman's rank correlation coefficient. $b$ is the intercept and $m$ is the gradient of the fitted line: observed = $m$(predicted) + $b$. RMSE is the root mean square error and $\mathrm{AVE}_{\mathrm{error}}$ is the average error. See text for equations.

The models contain some variables that are not statistically significant, but these significance tests are affected by the standard error estimates. For example, the statistically significant parameters in the quasi-Poisson model were logarea, and lroutcrop as a quadratic. In contrast, the ZIP model also included rainfall as a quadratic among the statistically significant parameters. If models had been developed with stepwise procedures some of the final models would have included fewer terms.

### 4.3. *Model evaluation*

The evaluation statistics, derived from the bootstrap analysis (Appendix B), indicated that the hurdle model performed best (Table 4). It had the highest or equal highest correlation between observed and predicted values for both the Pearson's and Spearman's rank correlations (0.58 and 0.66, respectively). This indicates that predictions and observations were relatively similar in magnitude and similarly ordered. This is reflected in the model calibration, that indicates a relatively small but consistent bias ($b = 7.32$, $m = 0.86$). Despite having the smallest RMSE (=173.59), the $\mathrm{AVE}_{\mathrm{error}}$ was median (=12.78).

The worst performing model was the NB (Table 4). Although it had a weak correlation ($r = 0.38$) its Spearman's rank was equal highest ($\rho = 0.66$). The poor linear correlation is reflected

**Table 3 – Parameter coefficients for each of the five different model specifications**

| Model | Parameter coefficients | | | | | |
|---|---|---|---|---|---|---|
| | Intercept | Logarea | Rainfall, 1 | Rainfall, 2 | Lroutcrop, 1 | Lroutcrop, 2 |
| Poisson | −1.57(0.08)* | 0.64(0.01)* | −0.75(0.12)* | −3.33(0.16)* | 6.25(0.13)* | −2.72(0.12)* |
| Quasi | −1.57(0.98) | 0.64(0.10)* | −0.75(1.50) | −3.33(2.08) | 6.25(1.67)* | −2.72(1.49)* |
| NB | −1.60(1.57) | 0.61(0.17)* | 0.89(2.32) | −5.17(2.48)* | 15.96(2.30)* | −11.34(2.28)* |
| Two-step | | | | | | |
| Pres/abs | −3.23(2.15) | 0.47(0.24)* | 3.79(3.21) | −2.17(2.94) | 22.31(4.85)* | −6.78(4.58) |
| Abund | 0.26(0.87) | 0.51(0.09)* | −1.33(1.08) | −2.13(1.37) | 0.12(0.97) | 0.61(0.93) |
| ZIP | 0.29(0.08) | 0.51(0.01)* | −2.12(0.12)* | −3.18(0.17)* | −0.84(0.18)* | 1.12(0.15)* |

Asterisk (*) denotes statistically significant parameters at the $p = 0.05$ level.

in the model calibration which indicates a strong and inconsistent bias ($b = 72.08$, $m = 0.28$). The RMSE and $\text{AVE}_{\text{error}}$ values were the highest of all models tested (348.71 and 64.12, respectively).

The ZIP model had a low $r$ (=0.40) and a low $\rho$ (=0.28), despite having the best model calibration of all models tested ($b = 1.01$, $m = 1.04$). This is because the amount of error around the predictions was high, but on average were accurate (RMSE = 192.18 and $\text{AVE}_{\text{error}} = -15.63$).

The other two models (Poisson and quasi-Poisson) were comparable in performance, largely because they had the same parameter coefficients (Table 3). They both had median correlations ($r = 0.52$) and high Spearman's ranks ($\rho = 0.65$ and 0.66, respectively) indicating the predictions were relatively dis-similar in value but similar in rank to the observations. This is reflected by medium biases ($b = 24.49$ and 24.21, respectively) that were relatively consistent ($m = 0.82$ and 0.84, respectively). Interestingly, these models (that is, Poisson and quasi-Poisson) had small RMSE and $\text{AVE}_{\text{error}}$ because, when averaged across all locations, their mean predictions were relatively accurate (Cameron and Trivedi, 1998).

Table 5 contains the rank predictions of the 114 outcrops. A good model will have a high rank (score close to 1) at a majority of the 114 outcrops. The hurdle model performed best with 58 (=39 + 19) of the 114 outcrops ranked in within the top two of the five closest predictions. This model also had the lowest number of outcrops (49 = 23 + 26) with a low rank (score = 4 or 5). This is consistent with the evaluation statistics that suggested the hurdle model was the most accurate and consistent. Note the NB model has 52 (=24 + 28) outcrops with the top two predictions, very close to that achieved by the hurdle model. However, the number of top ranking predictions of the hurdle model compared to the NB model is greater (39 and 24, respectively). The NB also has 61 outcrops in the bottom two predictions, indicating that any one prediction may be very close to, or very far from, the observed population size. This is consistent with the evaluation statistics that reported a large and inconsistent bias ($b = 72.08$, $m = 0.28$). The worst performing models were the Poisson and quasi-Poisson. That is, they had the lowest proportion of outcrops with high ranks (20 = 19 + 1) and a high proportion of outcrops with low ranks (55 = 50 + 5).

**Table 5 – The rank predictions for the five model specifications**

| Score | Poisson | NB | Quasi | Hurdle model | ZIP |
|---|---|---|---|---|---|
| 1 | 19 | 24 | 19 | 39 | 31 |
| 2 | 1 | 28 | 1 | 19 | 9 |
| 3 | 39 | 1 | 39 | 7 | 16 |
| 4 | 50 | 11 | 50 | 26 | 22 |
| 5 | 5 | 50 | 5 | 23 | 36 |

Cell $C_{i,j}$ represents the number of outcrops within row $i$ (i.e. rank i, $i = 1, \ldots, 5$) that were specified according to model $j$ ($j$ = Poisson, NB, quasi, hurdle, ZIP). Each column contains 114 outcrop predictions. A score of 1 implies of the five predictions for that particular outcrop, model $j$ resulted in the smallest discrepancy. A score of 5 had the largest discrepancy of the five predictions. For example, the Poisson model has 19 outcrops where the prediction was the closest to that observed, of all five predictions.

## 5. Discussion

Accurate predictions ensure decision-making by environmental managers is well informed. The aim of this study was to investigate the consequences of model specification in the presence of zero-inflation and over-dispersion, two commonly encountered problems when working with rare species. This study confirmed that model specification may have substantial impacts on model prediction.

The standard Poisson model is the most commonly available GLM and most straightforward to apply. However, this model does not accommodate either zero-inflation or over-dispersion, and tended to produce biased estimates of abundance on individual outcrops. This model may be adequate in some limited situations, because the errors cancelled one another out over all sites, providing a reasonable estimate of total abundance over the whole region.

In comparison to the standard Poisson, the quasi-Poisson, the negative binomial, the hurdle and the zero-inflated Poisson were more complicated to implement, but they can accommodate for either or both of the zero-inflation and over-dispersion. These models also allow more accurate prediction at the individual population level, estimates that are required for example, to target fieldwork at areas most likely to contain the species.

The hurdle model performed the best and was simple to interpret and implement. The model has another useful feature that was not explored in our investigation: the model for the binomial component can have different predictor variables. This can be useful because it may be ecologically sensible for cases where a different process might be driving the presence of a species to that which is affecting its abundance (Ridout et al., 1998).

Mixture models recognise the extra zero observations are from a different data generating process to the non-zero observations. However, this then assumes that only one data-generating process is causing the zero observations, which is often incorrect (Martin et al., 2005b). Even the ZIP model, specifically developed to accommodate the extra zero observations, does not explicitly inform you of which type of zero you are dealing with (only that any one zero observation is either a true negative (occurring if the habitat is unsuitable) or a false negative (occurring if the habitat is unsuitable but occupied), with probability $p_i$ or $1 - p_i$, respectively).

The tests presented here are important but comprise a small part of all those that might be attempted. Alternative model structures could be tested, such as the zero-inflated negative binomial model (Cheung, 2002). The differences between the models could be explored in conjunction with model selection (e.g. Fitzmaurice, 1997). The models were implemented using parametric methods, but there is scope for using non-parametric models. For example, the hurdle model could comprise generalised additive models (GAMs) rather than GLMs. The advantage of non-parametric methods is that they have additional flexibility for modelling biologically plausible relationships (e.g. Yee and Mitchell, 1991; Bio et al., 1998; Leathwick, 1998).

The evaluation of the models presented here included a bootstrap assessment of model performance. This is a useful tool in circumstances where an independent data set is not available for evaluation (a common situation). Evaluation on training data (i.e. on the data used to fit the model) is flawed (Leathwick et al., this issue), and should not be used if predictive performance is important. Bootstrap evaluations approximate true predictive performance more reliably than other resampling methods (Steyerberg et al., 2001). Here they enabled us to make the most efficient use of a small data set.

## 6. Conclusions and recommendations

When using mathematical models to quantify the relationship between a response variable (here, abundance estimates) and explanatory variables (here, environmental characteristics), it is imperative that correct model structure and model specification is used. The choices are plentiful, and should be governed by expert knowledge of the species and characteristics of the data set. It is recognised that model mis-specification influences the statistical significance of the explanatory variables, and consequently model selection methods (Cameron and Trivedi, 1998). However, the consequences for model predictions have been given little attention, despite their importance in the decision-making process of environmental managers.

This study focussed on a range of alternative models for abundance data, and demonstrated important differences between them. Our data were characterised by zero-inflation and over-dispersion, a common problem when working with rare species. The best performing model was the hurdle model, and the worst was the negative binomial. The statistical significance of the explanatory variables changed with changing model specification. Under the conditions of zero-inflation and over-dispersion, we recommend the hurdle model because it performed well consistently, and is relatively easy to interpret and implement.

## Acknowledgements

This work arose from research completed as part of the requirements for the Bachelor of Science (Honours) qualification within the School of Botany, at the University of Melbourne under the co-supervision of Professor Mark Burgman. David Keith, Department of Environment and Conservation (DEC, NSW), prompted our interest in *L. ralstonii* and kindly provided advice and data. State Forests (NSW) supplied GIS data for environmental variables and supported our field work, as did Max Beukers (DEC, NSW). We thank Matthew Carragher and Hedley Grantham for assistance with fieldwork, and Charles Todd, Graeme Newell, and the anonymous reviewers for discussion and comments on the manuscript.

## Appendix A

| Model | Computer packages implemented in R (R Development Core Team 2003) |
|---|---|
| Standard Poisson | “stats” |
| Negative binomial | “mass” |
| Quasi-likelihood | “stats” |
| Hurdle model | “stats” |
| Zero-inflated Poisson | “vgam” (Yee, 2005) |

## Appendix B. Method for 0.632+ bootstrap evaluation, after Steyerberg et al. (2001) and Efron and Tibshirani (1997).

1. Develop model on all *n* observations.
2. Calculate the statistic(s) of choice for evaluation on the same data (i.e. the training data)—call this $\text{Stat}_{\text{app}}$ because it is the apparent value of the statistic.
3. Take a bootstrap sample: i.e. a sample of size *n* with replacement, of rows of the data matrix. Keep track of which sites are in the bootstrap sample, and which are excluded.
4. Fit the model on the bootstrap sample (using the same methods as used on the full set). Compute the statistic on the bootstrap data set (observations versus fitted values) and call it $\text{Stat}_{\text{boot}}$.
5. Also compute the statistic on a version of the bootstrap data where the observations are randomised ($\text{Stat}_{\text{permute}}$).
6. Use the bootstrap model to predict to the excluded data set, and calculate the statistic on these predictions: $\text{Stat}_{\text{excl}}$.
7. Use $\text{Stat}_{\text{boot}}$, $\text{Stat}_{\text{permute}}$ and $\text{Stat}_{\text{excl}}$ to calculate the amount of overfitting, the relative overfitting rate and weights that are then used to make a best estimate of predictive performance, $\text{Stat}_{\text{best\_est}}$. This statistic puts most emphasis on predictions to the excluded data, particularly when the model is overfitted (i.e. when $\text{Stat}_{\text{boot}} - \text{Stat}_{\text{excl}}$ is large). For details of this step see Steyerberg et al. (2001).
8. Measure how optimistic the fit on the bootstrap sample was: $O = \text{Stat}_{\text{boot}} - \text{Stat}_{\text{best\_est}}$.
9. Repeat steps 3–8 100–200 times. Calculate an average optimism, O. Use this to correct $\text{Stat}_{\text{app}}$ for its optimism: $\text{Stat}_{\text{app}} - O$. This is a near unbiased estimate of the expected value of the external predictive performance of the process which generated $\text{Stat}_{\text{app}}$.

## REFERENCES

Albrecht, D.E., 1986. An Assessment of the Conservation Significance of Rhyolite Outcrops in Nullica State Forest. National Herbarium of Victoria.

Barry, S.C., Welsh, A.H., 2002. Generalized additive modelling and zero inflated count data. Ecol. Mod. 157, 179–188.

Binns, D., Kavanagh, R.P., 1990. Flora and Fauna Survey, Nullica State Forest (Part), Eden District, Eden Region. Forest

Resources Series No. 10. Forestry Commission of New South Wales, Sydney.

Bio, A.M.F., Alkemande, R., Barendregt, A., 1998. Determining alternative models for vegetation response analysis—a non-parametric approach. J. Veg. Sci. 9, 5–16.

Böhning, D., 1994. A note on a test for Poisson overdispersion. Biometrika 81, 418–419.

Böhning, D., Dietz, E., Schlattmann, P., Mendonca, L., Kirchner, U., 1999. The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. J. Roy. Statist. Soc. Ser. A Stat. 162, 195–209.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press.

Cheung, Y.B., 2002. Zero-inflated models for regression analysis of count data: a study of growth and development. Stat. Med. 21, 1461–1469.

Cox, D.R., 1983. Some remarks on overdispersion. Biometrika 70, 269–274.

Cragg, J.G., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 39, 829–844.

CRES, 2002. http://cres.anu.edu.au/outputs/anuclim.html.

Dietz, E., Böhning, D., 2000. On estimation of the Poisson parameter in zero-modified Poisson models. Computat. Statist. Data Anal. 34, 441–459.

Efron, B., Tibshirani, R., 1997. Improvements on cross-validation: the.632+ bootstrap method. J. Am. Statist. Assn. 92, 548–560.

Elith, J., 2002. Predicting the Distribution of Plants. Ph.D. Thesis. The University of Melbourne, Melbourne, Australia, 301 pp.

Faddy, M., 1998. Stochastic models for analysis of species abundance data. In: Fletcher, D.J., Kavalieris, L., Manly, B.F.J. (Eds.), Statistics in Ecology and Environmental Monitoring. University of Otago Press, pp. 33–40.

Fitzmaurice, G.M., 1997. Model selection with overdispersed data. The Statistician 46, 81–91.

Greenwood, M., Yule, G.U., 1920. An enquiry into the nature of frequency distributions representative of multiple happenings with particular reference of multiple attacks of disease or of repeated accidents. J. R. Statist. Soc. 83, 25–279.

Hall, D.B., 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics 56, 1030–1039.

Hall, D.B., Berenhaut, K.S., 2002. Score tests for heterogeneity and overdispersion in zero-inflated Poisson and binomial regression models. Can. J. Statist. 30, 415–430.

Harrell, F.E., 2001. Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis. Springer Series in Statistics. Springer-Verlag, New York, p. 568.

Hastie, T., Tibshirani, R., Friedman, J.H., 2001. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer-Verlag, New York.

Hone, J., 1999. On rate of increase (r): patterns of variation in Australian mammals and the implications for wildlife management. J. Appl. Ecol. 36, 709–718.

IUCN, 2001. The IUCN Red List of Threatened Species. 3.1, Gland, Switzerland.

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, 717–747.

Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34, 1–4.

Leathwick, J., Elith, J., Hastie, T., this issue. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecol. Mod., doi:10.1016/j.ecolmodel.2006.05.022.

Leathwick, J.R., 1998. Are New Zealand's Nothofagus species in equilibrium with their environment? J. Veg. Sci. 9, 719–732.

Lee, A.H., Wang, K., Yau, K.K.W., 2001. Analysis of zero-inflated Poisson data incorporating extent of exposure. Biomet. J. 8, 963–975.

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. Comm. Dent. Oral Epidemiol. 32, 183–189.

Lubow, B.C., 1996. Optimal translocation strategies for enhancing stochastic metapopulation viability. Ecol. Appl. 6, 1268–1280.

Martin, T.G., Kuhnert, P.M., Mengersen, K., Possingham, H.P., 2005a. The power of expert opinion in ecological models using Bayesian methods: impact of grazing on birds. Ecol. Appl. 15, 266–280.

Martin, T.G., Wintle, B., Rhodes, J., Kuhnert, P.M., Field, S.A., Low Choy, S., Tyre, A.J., Possingham, H.P., 2005b. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecol. Lett. 8, 1235–1246.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Monographs on Statistics and Applied Probability. Chapman and Hall, London, 261.

Nelder, J.A., 2000. Quasi-likelihood and pseudo-likelihood are not the same thing. Appl. Stat. 27, 1007–1011.

NSW National Parks and Wildlife Service, 2003. Draft NSW and National Recovery Plan: Recovery Plan for Threatened Flora of Rocky Outcrops in South Eastern New South Wales, Threatened Species Unit, National Parks and Wildlife Service, Hurstville, NSW.

Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Mod. 133, 225–245.

Podlich, H.M., Faddy, M.J., Smyth, G.K., 2002. A general approach to modeling and analysis of species abundance data with extra zeros. JABES 7, 324–334.

Possingham, H.P., Lindenmayer, D.B., McCarthy, M.A., 2001. Population Viability Analysis, Encyclopædia of Biodiversity. Academic Press, pp. 831–843.

Potts, J.M., Elith, J., Burgman, M.A. A population viability analysis for *Leionema ralstonii*, a threatened plant of New South Wales, submitted for publication.

R Development Core Team, 2004. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Regan, H.M., Auld, T.D., Keith, D.A., Burgman, M.A., 2003. The effects of fire and predators on the long-term persistence of an endangered shrub, *Grevillea caleyi*. Biol. Cons. 109, 73–83.

Ridout, M., Demetrio, C., Hinde, J., 1998. Models for count data with many zeros. In: International Biometric Conference, Cape Town, pp. 1–13.

Ridout, M., Hinde, J., Demétrio, C.G.B., 2001. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. Biometrics 57, 219–223.

Steyerberg, E.W., Harrell, F.E., Borsboom, G.J.J.M., Eijkemans, M.J.C., Vergouwe, Y., Habbema, J.D.F., 2001. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J. Clin. Epidemiol. 54, 774–781.

Tu, W., 2002. Zero-inflated data. In: El-Shaarawi, A.H., Piegorsch, W.W. (Eds.), Encyclopedia of Environmetrics. John Wiley and Sons, Chichester, pp. 2387–2391.

Tyre, A.J., Possingham, H.P., Lindenmayer, D.B., 2001. Matching observed pattern with ecological process: can territory occupancy provide information about life history parameters? Ecol. Appl. 11, 1722–1738.

Umbach, D., 1981. On inference for a mixture of a Poisson and a degenerate distribution. Commun. Statist.—Theory Methods A10, 299–306.

van den Broek, J., 1995. A score test for zero inflation in a Poisson distribution. Biometrics 51, 738–743.

Venables, W.N., Ripley, B.D., 2002. Modern and Applied Statistics with S. Statistics and Computing. Springer, United States of America.

Wedderburn, R.W.M., 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika 61, 439–447.

Welsh, A.H., Cunningham, R.B., Chambers, R.L., 2000. Methodology for estimating the abundance of rare animals: Seabird nesting on north east Herald Cay. Biometrics 56, 22–30.

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. Mod. 88, 297–308.

Wintle, B.A., Elith, R.J., Potts, J.M., 2005. Fauna habitat modelling and mapping in an urbanising environment: a review and case study in the Lower Hunter Central Coast region of NSW. Aust. Ecol. 30, 719–738.

Yau, K.K.W., Lee, A.H., 2001. Zero-inflated Poisson regression with random effects to evaluate an occupied injury prevention programme. Stat. Med. 20, 2907–2920.

Yee, T., 2005. VGAM software [online]. Available from http://www.stat.auckland.ac.nz/~yee/VGAM/ [accessed 17th January 2005].

Yee, T.W., Mitchell, N.D., 1991. Generalized additive models in plant ecology. J. Veg. Sci. 2, 587–602.

Zorn, C.J.W., 1996. Evaluating zero-inflated and hurdle Poisson specifications. Midwest Political Science Association [online]. Available from http://polmeth.wustl.edu/retrieve.php?id=345 [accessed 26th June, 2006].