

ESTIMATING SPECIES OCCURRENCE, ABUNDANCE, AND DETECTION PROBABILITY USING ZERO-INFLATED DISTRIBUTIONS

SETH J. WENGER^{1,3} AND MARY C. FREEMAN²

¹*University of Georgia River Basin Center, 110 Riverbend Road, Athens, Georgia 30602 USA*

²*U.S. Geological Survey, Patuxent Wildlife Research Center, Athens, Georgia 30602 USA*

Abstract. Researchers have developed methods to account for imperfect detection of species with either occupancy (presence–absence) or count data using replicated sampling. We show how these approaches can be combined to simultaneously estimate occurrence, abundance, and detection probability by specifying a zero-inflated distribution for abundance. This approach may be particularly appropriate when patterns of occurrence and abundance arise from distinct processes operating at differing spatial or temporal scales. We apply the model to two data sets: (1) previously published data for a species of duck, *Anas platyrhynchos*, and (2) data for a stream fish species, *Etheostoma scotti*. We show that in these cases, an incomplete-detection zero-inflated modeling approach yields a superior fit to the data than other models. We propose that zero-inflated abundance models accounting for incomplete detection be considered when replicate count data are available.

Key words: abundance estimation; Cherokee darter; detectability; *Etheostoma scotti*; negative binomial; occupancy model; Poisson; presence–absence.

INTRODUCTION

Species abundance and site occupancy (occurrence) of species are both useful measures of population status, and therefore of considerable interest to ecologists. Both measures may be confounded when species detection is less than perfect (Bayley and Peterson 2001, MacKenzie et al. 2002, Gu and Swihart 2004). Models have been developed to address this problem in occupancy estimation, based on observed presences and absences of species from replicated samples (MacKenzie et al. 2002, Royle and Nichols 2003), and in abundance estimation, using replicated count samples (Royle 2004, Royle et al. 2005). In the latter model, occupancy is a derived parameter based on locations where abundance is greater than zero.

However, there are cases in which the occupancy pattern and the local abundance distribution of a species arise from two distinct processes, which may be operating at different temporal or spatial scales. Consider a species that has been extirpated from a portion of its range by historic land use activities, with re-colonization limited by movement barriers. The relationship between historic land use and species occurrence may be well represented by an occupancy model. Where the species does occur, however, its

abundance may be predictable by other covariates. In such cases, it is reasonable to consider models in which species abundance is modeled as the product of two processes: (1) species presence and (2) species abundance when present. Such an approach may also be useful when abundances exhibit a threshold effect, such that a species is either not present, or present at moderate to high abundances. In both cases, the abundance may be best represented by a bimodal, zero-inflated distribution (Welsh et al. 1996).

Zero-inflated distributions have been proposed as appropriate models for describing the spatial distribution of rare species because of their ability to account for extra absences in the data (Welsh et al. 1996, Ridout et al. 1998, Cunningham and Lindenmayer 2005, Martin et al. 2005). A zero-inflated distribution can be viewed as a two-part model, in which (1) the probability of species presence and (2) the abundance, given presence, are modeled from the same data. Such models have been used to estimate species abundances (e.g., Cunningham and Lindenmayer 2005), but usually without addressing incomplete detection, which can be of particular concern for rare species. This omission can be rectified by adding a third part to the model that estimates per-individual detection probability based on repeated samples at a site. The incomplete-detection occupancy model of MacKenzie et al. (2002) also uses a zero-inflation term to model occupancy, but lacks an abundance term because it applies to presence–absence data rather than count data.

Manuscript received 9 July 2007; revised 24 January 2008; accepted 24 January 2008; final version received 15 February 2008. Corresponding Editor: F. He.

³ E-mail: swenger@uga.edu

In this paper, we show how the N-mixture models of Royle (2004) and Royle et al. (2005) can be used to simultaneously model occurrence (occupancy), abundance, and detection by specifying a zero-inflated distribution for the abundance. Royle (2004) and Royle et al. (2005) developed models using Poisson and negative binomial distributions for abundance, but noted that the approach can be employed with other distributions. The specification of a zero-inflated distribution is especially interesting and useful, because the resulting model is effectively a marriage of Royle et al.'s N-mixture abundance model and the zero-inflated binomial occupancy model of MacKenzie et al. (2002); the first accounts for abundance and incomplete detection, the second accounts for occupancy and incomplete detection, while our proposed formulation simultaneously estimates all three parameters. After introducing the models, we apply them to two data sets: one for Mallard Ducks (*Anas platyrhynchos*) published in a supplement to Kery et al. (2005), and the other for the Cherokee darter (*Etheostoma scotti*), a federally threatened fish species. We then discuss the general applicability, advantages and limitations of the approach.

BACKGROUND: N-MIXTURE MODELS FOR ABUNDANCE ESTIMATION

Royle (2004) and Royle et al. (2005) introduced a class of models for data sets where the study organism is counted at R sites, $i = 1, 2, \dots, R$, with multiple counts made at time $t = 1, 2, \dots, T$ at each site, with such counts denoted y_{it} . These counts are viewed as realizations of a binomial process with index N_i (abundance) and detection probability p_{it} , which we denote as

$$y_{it} \sim \text{binomial}(N_i, p_{it}). \quad (1)$$

Note that p_{it} is the per-individual detection probability, which is also referred to as capture efficiency. This model assumes that the population at any given site is "closed" across sampling counts, i.e., there is no change in abundance from count to count, which allows estimation of both abundance and detection probability. Detection probability can be assumed to be constant or it can be modeled as a function of covariates. For the latter, generalized linear modeling can be employed, for example with the logit link:

$$\text{logit}(p_{it}) = a_0 + a_1 x_{it} \quad (2)$$

where x_{it} represents a covariate on detection as measured at site i on visit t (multiple covariates are possible and will often be necessary). Thus, this portion of the model is simply logistic regression.

Estimation of abundance is facilitated by assuming the observed counts are drawn from a statistical distribution, denoted generally as $f(N_i; \theta)$. Perhaps the simplest useful distribution is the Poisson, which has only one parameter, the mean (λ):

$$K_i \sim \text{Poisson}(\lambda_i). \quad (3)$$

Covariates can be added to explain variation in λ using a log link:

$$\log(\lambda_i) = b_0 + b_1 y_i \quad (4)$$

where y_i is the value of a covariate at site i . Note that unlike covariates on detection, covariates on abundance (λ_i) are assumed to be constant across all t visits. Again, multiple covariates are allowed, although not explicitly shown in the equation above. Many alternative distributions to the Poisson are possible; Royle (2004) also considered the negative binomial, which can be parameterized as a Poisson with a free variance parameter that allows for overdispersion. In this formulation, covariates may be placed on the mean of the negative binomial distribution, and the variance or "size" parameter provides an estimate of unexplained deviation from the mean. In the next section, we discuss how zero-inflated versions of the Poisson and negative binomial distributions can be used for $f(N_i; \theta)$.

ZERO-INFLATED DISTRIBUTIONS TO MODEL OCCURRENCE AND ABUNDANCE

Zero-inflated mixture distributions (Martin et al. 2005) or zero-modified distributions (Ridout et al. 1998), such as the zero-inflated Poisson (ZIP), are mixtures of two probability distributions, one with a point mass only at zero. Employing such a zero-inflated distribution for $f(N_i; \theta)$ provides a simple method for simultaneously modeling both presence (or occurrence or occupancy; the three terms are equivalent in this context) and abundance. The ZIP distribution has two parameters, a probability that the species is present (ψ) and the mean abundance of the species, if present (λ). The probability mass function ($\Pr(Y = y)$) for a ZIP distribution is given by

$$(1 - \psi) + (\psi) \times \exp(-\lambda) \quad y = 0 \quad (5a)$$

$$(\psi) \times \exp(-\lambda) \times \lambda^y / y! \quad y > 0 \quad (5b)$$

(Ridout et al. 1998). Alternatively, we can represent the mixture distribution as follows, using the pseudo-code notation from the previous section:

$$N_i = \text{pres}_i \times K_i \quad (6a)$$

$$\text{pres}_i \sim \text{Bernoulli}(\psi_i) \quad (6b)$$

$$K_i \sim \text{Poisson}(\lambda_i) \quad (6c)$$

where pres_i is a binary value indicating whether or not the species is present at site i , and K_i is the realized abundance at site i , given presence.

We will usually wish to model ψ as a function of covariates, which we can do with a logit link:

$$\text{logit}(\psi_i) = c_0 + c_1 z_i \quad (7)$$

where z_i is the value of a covariate at site i . As was the case for abundance, covariates on presence (ψ_i) are assumed to be constant across all t site visits, and multiple covariates are possible. The covariates used for presence may be the same or different from the ones used on abundance. For example, we might wish to model the abundance of a species as a function of elevation and current land use, and presence of the species as a function of elevation and historic land use.

We can write the entire general model as follows:

$$y_{it} \sim \text{binomial}(N_i, p_{it}) \quad (8a)$$

$$N_i = \text{pres}_i \times K_i \quad (8b)$$

$$\text{pres}_i \sim \text{Bernoulli}(\psi_i) \quad (8c)$$

$$K_i \sim \text{Poisson}(\lambda_i) \quad (8d)$$

$$\text{logit}(p_{it}) = a_0 + a_1 x_{it} \quad (8e)$$

$$\log(\lambda_i) = b_0 + b_1 y_i \quad (8f)$$

$$\text{logit}(\psi_i) = c_0 + c_1 z_i. \quad (8g)$$

It is straightforward to adapt this model to use zero-inflated distributions other than the Poisson, such as the zero-inflated negative binomial (ZINB). This is the same as the above formulation, but Eq. 8d is replaced by

$$K_i \sim \text{negative binomial}(\lambda_i, \alpha) \quad (9)$$

where λ_i is the mean at site i and α is the “size parameter” or variance of the distribution. As for the Poisson distribution, the mean of the negative binomial distribution may accept covariates. However, whereas the variance of the Poisson distribution is equal to the mean, the variance of the negative binomial may be greater (though never less) than the mean. In practice this means that if there is significant residual variation after including the covariates, a ZINB model will tend to provide a better fit than a ZIP model.

The parameters of this model can be estimated using various approaches, including maximum likelihood and Bayesian Markov chain Monte Carlo (MCMC) methods. The likelihood is of the same form as given for the N -mixture model in Royle (2004), but a zero-inflated Poisson distribution is substituted for the Poisson distribution. In the Supplement, we provide code for maximum-likelihood using the `nlm` package in R (R Development Core Team 2005), based on code by A. Royle given in Kery et al. (2005), and code for Bayesian estimation in WinBUGS (Spiegelhalter et al. 2003). The code in WinBUGS is extensible to complex structures, such as mixed-effects models, although there may be

complications with the use of negative binomial models (see Supplement). The coding in `nlm` is somewhat less intuitive and not as readily extensible to complex model structures, but has a great speed advantage. In addition, we have developed a user-friendly interface to allow users with only basic knowledge of R to run ZIP and ZINB models for any dataset, without the need for coding in `nlm` (a tutorial is also included in the Supplement). In tests we found that models converged well using maximum likelihood and Bayesian approaches even with relatively small data sets (30 sites with three samples per site; S. J. Wenger and M. C. Freeman, *unpublished data*).

The model inherits the assumptions of abundance models (Royle 2004) and basic occupancy models (MacKenzie et al. 2002). The first of these is that the sampled population is closed to immigration, emigration and state changes between samples, so that the same number of individuals is available for counting each time. For mobile species this is likely to be violated to some degree. A violation will tend to reduce estimates of detection probability and inflate estimates of abundance, since changes in counts will be interpreted as non-detections rather than changes in population. A second assumption is that individuals are independent and equally available for capture. If this assumption is greatly violated (as can happen with some schooling fishes, for example), there can be excessive variation among counts at a site and detection probability will again be underestimated. In such circumstances it may be appropriate to estimate the number of groups rather than individuals. Finally, there are assumptions associated with the selected distribution used to model abundance; the choice of distribution is essentially a prior assumption about the structure of the unexplained error in abundance among sites. For example, if a Poisson distribution is used where there is excessive unexplained variation (overdispersion), parameter estimates may be biased.

APPLICATION 1: MALLARD DUCK DATA

Our first illustration of the zero-inflated abundance model uses the data set for Mallard Ducks (*Anas platyrhynchos*) reported by Kery et al. (2005). These data, from the national breeding bird monitoring program in Switzerland consist of 235 sites, of which 191 were sampled three times, 42 were sampled two times, and two were sampled only once. Mallard presence was observed at 40 of the sites with counts ranging from 1 to 12, with a mean count of 1.3 for sites where the species was observed. Kery et al. (2005) fitted Poisson and negative binomial mixture models to the Mallard data using covariates for elevation, forest cover, monitoring route length, effort, date, and interactions of these. They selected best-supported models based on Akaike's Information Criterion (AIC; Burnham and

TABLE 1. Results of the Mallard data study.

Distribution	Covariates			AIC
	Abundance	Occurrence	Detection	
ZINB	route length, elevation	route length, forest cover	date, date ² , date × elevation, date ² × elevation	455.0
NB	route length, elevation, forest cover	NA	date, date ² , date × elevation, date ² × elevation	472.7
ZIP	route length, elevation	route length, forest cover	effort, date, date ² , date × elevation, date ² × elevation	474.2
Poisson	route length, elevation, forest cover	NA	effort, date, date ² , date × elevation, date ² × elevation	506.8

Notes: The best supported covariate combination for each distribution is shown along with the corresponding AIC score. Distributions are zero-inflated negative binomial (ZINB), negative binomial (NB), zero-inflated Poisson (ZIP), and Poisson. Included covariates are listed according to the term to which they are applied (abundance, occurrence, or detection). Covariates and AIC scores for NB and Poisson models are from Kery et al. (2005). NA indicates “not applicable,” as these models do not have a separate term for occurrence and therefore cannot have covariates for occurrence/presence.

Anderson 2002). We repeated their analysis using ZIP and ZINB distributions as assumptions on abundance, rather than Poisson and negative binomial distributions. We fitted a selected subset of covariate combinations likely to result in a good fit based on the results reported by Kery et al. (2005). We ranked all models by AIC and compared the best-supported models with the best-supported models reported by Kery et al. (2005). We used AIC rather than AIC adjusted for small sample size (AIC_c) for consistency with the previously reported values.

Results of Mallard Duck study

Based on AIC, we found that the best-supported model was one with a ZINB assumption on abundance (Table 1). In general, models with a zero-inflated abundance distribution ranked above their non-zero-inflated counterparts, and models using negative binomial distributions ranked above those using Poisson distributions. The best-supported zero-inflated models included the same covariates as the best-supported non-zero-inflated models, but a better fit was obtained by placing some covariates on occurrence rather than abundance, or on both occurrence and abundance (Table 1). For example, Mallard presence/absence was better explained by forest cover, while Mallard abundance was better explained by elevation. Both occurrence and abundance were correlated with route length, indicating that increasing the length of the route driven by observers increased both the probability of encountering habitat for Mallards and the abundance of Mallards, when encountered. Mean detection probabilities were essentially the same (about 50% per individual, under mean covariate values) for the best ZINB and the best negative binomial model.

APPLICATION 2: CHEROKEE DARTER DATA

For the second illustration we used count data for the Cherokee darter (*Etheostoma scotti*), a federally protected stream fish endemic to the Etowah River Basin in Georgia, USA. Accurate prediction of species abun-

dances across environmental gradients can inform conservation strategies, which was our interest in building predictive models for the Cherokee darter. Gaps in the Cherokee darter's distribution within the Etowah Basin suggested that the species had been extirpated from portions of its native range (Bauer et al. 1995, Burkhead et al. 1997), effectively creating two sources of non-occurrence: extirpation and unsuitable local conditions. We hypothesized that a zero-inflated abundance model would therefore provide a better fit to count data for this species than models without zero-inflation.

Data were collected during 276 monitoring events at 215 sites in small and medium-sized streams between 1999 and 2003 using backpack electrofishing and kick seining. During each collection, captured Cherokee darters were counted and returned to the stream. A subset of sites was sampled more than once; those collections made at the same site within two consecutive years were included in the analysis under the assumption that the population was closed during that period (this is a long period to assume closure, which means that detection might be somewhat underestimated, as discussed previously). The subset comprised 54 sites sampled a second time and seven sites sampled a third time. Methods and effort at each site were intended to be consistent, although we hypothesized that there was variation in the effectiveness of collectors associated with the three major institutions performing the collections. Collector institution was therefore considered a covariate on detection. Cherokee darter counts ranged from 0 to 145, with a mean of 10 and standard deviation of 19; Cherokee darters were collected at 115 of the 215 sites.

Six potential site-level covariates of abundance and occurrence, including a variable of management concern (effective impervious area, EIA) were recorded from mapped data using a Geographic Information System (GIS). The covariates were identified through exploratory analyses of 24 potential variables in relation to

TABLE 2. Covariates on abundance and occurrence for Cherokee darter models.

Covariate set	Covariates	
	Abundance	Occurrence
A	surficial geology, EIA	elevation, bedrock geology, impoundments
B	surficial geology, EIA	elevation, bedrock geology, Little River system
C	surficial geology, EIA	elevation, bedrock geology, impoundments, Little River system
D	surficial geology, EIA, elevation, bedrock geology	elevation, bedrock geology, impoundments
E	surficial geology, EIA, elevation, bedrock geology	elevation, bedrock geology, impoundments, Little River system
F	surficial geology, EIA, elevation, bedrock geology, impoundments	NA
G	surficial geology, EIA, elevation, bedrock geology, Little River system	NA
H	surficial geology, EIA, elevation, bedrock geology, impoundments, Little River system	NA

Note: "EIA" indicates effective impervious area.

darter abundance and occurrence, using linear and logistic regression, respectively. The five variables exhibiting strongest association with fish abundance or occurrence, in addition to EIA, were: elevation, bedrock geology, surficial (Quaternary) geology, sub-basin reservoir area (area inundated by reservoirs in the tributary system to which the site belongs), and occurrence in the Little River tributary system (where the Cherokee darter appears to be widely extirpated [Burkhead et al. 1997]). EIA was measured as the proportion of the area within 1 km of the collection site that was impervious and drained by storm sewers. These six variables were not strongly intercorrelated (i.e., $r < 0.4$).

Five sets of model covariates were formulated to represent alternative hypothesized effects of site variables on fish abundance, occurrence, or both (Table 2). Based on the results of the exploratory analysis, our strategy was to include two variables (Quaternary geology and EIA) in all models as covariates on abundance and two variables (elevation and bedrock geology) in all models as covariates on occurrence, and to test alternative combinations of the remaining variables as covariates on occurrence and/or abundance. Each of five sets of covariates (Table 2, covariate sets A–E) was modeled using ZIP and ZINB assumptions on abundance. Variations of these covariate sets were also fit using Poisson and negative binomial models without zero-inflation terms, in which cases all covariates were placed on the abundance term (Table 2, covariate sets F–H). A total of 16 models were evaluated. Akaike's Information Criterion for small sample size (AIC_c) was calculated for each model, with the sample size set equal to the number of sites. ΔAIC_c was calculated by subtracting the lowest (best) AIC_c score from every AIC_c score.

Results of Cherokee darter study

Based on AIC_c , the best-supported model was the zero-inflated negative binomial abundance model with covariate set B (Table 3). With this covariate set,

occurrence was explained by elevation, bedrock geology, and whether or not the collection was made within the Little River system; abundance was explained by surficial geology and EIA. There was a clear separation of models based on the assumed abundance distribution (Table 3). All the negative binomial models provided a much better fit than the Poisson models, while the zero-inflated models performed better than models with the same distribution but no zero-inflation term. Collector identity had a strong influence on detection in all models; mean per-individual detection probability varied from as low as 9% to as high as 22%, depending on the collector.

DISCUSSION

Our results for the two data sets show that zero-inflated abundance models can be useful in estimating species occurrence and abundance from replicate count data while accounting for incomplete detection. In both

TABLE 3. Results of the Cherokee darter study.

Distribution	Covariate set	ΔAIC_c
ZINB	B	0
ZINB	D	2
ZINB	C	2
ZINB	E	5
ZINB	A	15
NB	G	59
NB	H	61
NB	F	75
ZIP	E	1766
ZIP	D	1812
ZIP	B	1824
ZIP	C	1825
ZIP	A	1841
Poisson	H	3750
Poisson	G	3766
Poisson	F	4115

Note: Values for ΔAIC_c are shown for alternative covariate combinations applied to each distribution (see Table 2 for covariates in each set).

cases, models with a zero-inflation term provided a better fit to the data than N-mixture abundance models without a zero-inflation term. Based on these results and on results from studies of simulated data (S. J. Wenger and M. C. Freeman, *unpublished data*), we believe zero-inflated abundance models will prove useful for many types of data sets.

In the past, zero-inflated models have been recommended as appropriate for rare species (Welsh et al. 1996, Cunningham and Lindenmayer 2005), although studies have shown that a negative binomial model can also provide a good fit to apparently zero-inflated data sets (Warton 2005). We believe that rarity alone is not sufficient grounds for selecting a zero-inflated model, and we agree with Warton (2005) that negative binomial and other distributions without zero-inflation terms should also be considered as alternatives. Using class covariates (e.g., soil types for a rare plant) to model abundances may adequately account for absences, without the need for a zero-inflation term. It is important to remember that the choice of the distribution $f(N_i; \theta)$ is a specification of the error distribution, not the distribution of the observed counts, so zero-inflation in counts is not necessarily evidence of the need for a zero-inflated modeling distribution.

A zero-inflated distribution is likely to be better supported than a non-inflated distribution under at least two circumstances. First, if a continuous covariate is a useful predictor of occurrence but not of abundance, then applying that covariate to the occurrence term of a zero-inflated abundance model will likely yield a better fit than applying that covariate to the abundance term in a model that doesn't account for zero inflation. However, if the covariate can be converted to a class variable, it may be a useful covariate for the latter model type. Second, if abundance shows a threshold relationship to a continuous covariate, a better fit should be obtained by adding that covariate to both the occurrence and abundance terms, rather than just abundance. The zero-inflation term can thus account for non-linear relationships. Modeling techniques such as neural networks (e.g., Olden and Jackson 2002) and maximum entropy (see review of this and other methods by Elith et al. [2006]), are even better suited to handling nonlinearities, but as yet these approaches have not been extended to account for incomplete detection of species.

An additional benefit to using zero-inflated distributions is that the separate occurrence and abundance terms can have heuristic value in representing the different mechanisms that gave rise to observed patterns of species abundance. These may be used to represent factors operating on different temporal or spatial scales. This was previously explored by Cunningham and Lindenmayer (2005) in a paper examining the use of zero-inflated distributions for modeling rare species. However, the approach used in that study did not

account for incomplete detection, which can be a major concern with rare species, so we see the approach presented here as a significant advance.

As a general guideline, when presented with a set of count data collected by repeat visits to sites, we recommend formulating a candidate set of models that includes both zero-inflated and non-zero-inflated incomplete detection models. The models should include covariates reflecting a priori hypotheses to the extent possible, following a general information theoretic framework (Burnham and Anderson 2002). The best-supported models can be selected based on AIC, cross validation, or other methods.

Our formulation of the zero-inflated distribution was as a simple mixture based on a full Poisson or negative binomial distribution. Other authors (Welsh et al. 1996, Cunningham and Lindenmayer 2005, Warton 2005) have suggested using a truncated Poisson or truncated negative binomial distribution rather than the full distribution. This produces a two-part, conditional model in which occurrence and abundance are truly separate and orthogonal, and eliminates the possibility that a species may be predicted as present (occupancy = 1) but with a mean abundance of zero. On the other hand, if mean abundance is moderate or high, the two approaches are functionally equivalent (Welsh et al. 1996). We found the simple mixture easy to construct in the software programs we used (WinBUGS and R), and note that conceptually it is not illogical to envision a species as potentially present based on the covariates that govern its distribution (occurrence), but absent at a patch because the covariates that govern its abundance are unfavorable. Therefore, we see both formulations as valid alternatives.

In their review of the use of zero-inflated distributions in ecology, Martin et al. (2005) noted that "in the literature there has been no formal discussion of how to model data sets that contain both excess true zeros and excess false zeros" (false zeros being cases of non-detection). While the occupancy models of MacKenzie et al. (2002) actually do this for presence-absence data using a zero-inflated binomial model, it is true that until now there has not been an approach that accounts for excess zeros and non-detection for count data. The method outlined here—a specific case of the general model of Royle (2004)—fills this gap. We suggest that these models are worthy of consideration under almost any circumstances where repeat count data are available.

ACKNOWLEDGMENTS

This work was supported in part by grants from the U.S. Fish and Wildlife Service. We thank Jim Peterson, Larissa Bailey, Jim Nichols, and two anonymous reviewers for extremely helpful comments on earlier drafts. We thank Bud Freeman for the use of the Cherokee darter data set. We thank Darryl Mackenzie for additional suggestions and help with WinBUGS modeling code.

LITERATURE CITED

- Bauer, B. H., D. A. Etnier, and N. M. Burkhead. 1995. *Etheostoma (Ulocentra) scotti* (Osteichthyes: Percidae), a new darter from the Etowah River system in Georgia. *Bulletin of the Alabama Museum of Natural History* 17:1–16.
- Bayley, P. B., and J. T. Peterson. 2001. An approach to estimate probability of presence and richness of fish species. *Transactions of the American Fisheries Society* 130:620–633.
- Burkhead, N. M., S. J. Walsh, B. J. Freeman, and J. D. Williams. 1997. Status and restoration of the Etowah River, an imperiled southern Appalachian ecosystem. Pages 375–444 in G. W. Benz and D. E. Collins, editors. *Aquatic fauna in peril: the southeastern perspective*. Special publication 1. Southeast Aquatic Research Institute, Lenz Design and Communications, Decatur, Georgia, USA.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer-Verlag, New York, New York, USA.
- Cunningham, R. B., and D. B. Lindenmayer. 2005. Modeling count data of rare species: Some statistical issues. *Ecology* 86:1135–1142.
- Elith, J., et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151.
- Gu, W. D., and R. K. Swihart. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation* 116:195–203.
- Kery, M., J. A. Royle, and H. Schmid. 2005. Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications* 15:1450–1461.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255.
- Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters* 8:1235–1246.
- Olden, J. D., and D. A. Jackson. 2002. A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology* 47:1976–1995.
- R Development Core Team. 2005. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ridout, M., C. G. B. Demetrio, and J. Hinde. 1998. Models for count data with many zeros. Pages 179–192 in *Proceedings of XIXth International Biometric Society Conference*, Cape Town, South Africa. International Biometric Society, Washington, D.C., USA.
- Royle, J. A. 2004. *N*-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60:108–115.
- Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. *Ecology* 84:777–790.
- Royle, J. A., J. D. Nichols, and M. Kery. 2005. Modelling occurrence and abundance of species when detection is imperfect. *Oikos* 110:353–359.
- Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn. 2003. *WinBUGS*, version 1.4. MRC Biostatistics Unit, Cambridge, UK.
- 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., R. B. Cunningham, C. F. Donnelly, and D. B. Lindenmayer. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling* 88:297–308.

SUPPLEMENT

Codes, notes, and tutorial for running zero-inflated abundance incomplete detection models in WinBUGS and R (*Ecological Archives* E089-166-S1).