Biometrics 74, 369–377
March 2018

DOI: 10.1111/biom.12734

On the Reliability of N-Mixture Models for Count Data

Richard J. Barker,[1,*] **Matthew R. Schofield**,[1] **William A. Link**,[2] **and John R. Sauer**[2]

[1]Department of Mathematics and Statistics, University of Otago, P. O. Box 56 Dunedin 9016, New Zealand
[2]U.S. Geological Survey, Patuxent Wildlife Research Center, Maryland 20708, U. S. A.
[*]*email:* rbarker@maths.otago.ac.nz

Summary. N-mixture models describe count data replicated in time and across sites in terms of abundance N and detectability p. They are popular because they allow inference about N while controlling for factors that influence p without the need for marking animals. Using a capture–recapture perspective, we show that the loss of information that results from not marking animals is critical, making reliable statistical modeling of N and p problematic using just count data. One cannot reliably fit a model in which the detection probabilities are distinct among repeat visits as this model is overspecified. This makes uncontrolled variation in p problematic. By counter example, we show that even if p is constant after adjusting for covariate effects (the "constant p" assumption) scientifically plausible alternative models in which N (or its expectation) is non-identifiable or does not even exist as a parameter, lead to data that are practically indistinguishable from data generated under an N-mixture model. This is particularly the case for sparse data as is commonly seen in applications. We conclude that under the constant p assumption reliable inference is only possible for relative abundance in the absence of questionable and/or untestable assumptions or with better quality data than seen in typical applications. Relative abundance models for counts can be readily fitted using Poisson regression in standard software such as R and are sufficiently flexible to allow controlling for p through the use covariates while simultaneously modeling variation in relative abundance. If users require estimates of absolute abundance, they should collect auxiliary data that help with estimation of p.

Key words: Ancillary statistic; Capture recapture; Log linear model; N-mixture models; Partial likelihood.

1. Introduction

Strategies for inference about abundance N from count data under imperfect detection include: (i) capture–recapture modeling in which auxiliary data supplement the counts in order to allow direct inference about detection rates p; (ii) N-mixture modelling in which no such auxiliary data are collected but instead the model is structured in order to allow inference about N and p (which is assumed constant after controlling for covariates, hereafter, "constant p"); or (iii) index models in which inference is made about relative abundance assuming constant p.

N-mixture models were developed by Royle (2004) as an alternative to estimating abundance using tools such as capture–recapture that can be difficult, expensive, and impractical (Royle, 2004; Dennis et al., 2015). They are popular among field biologists and the original model of Royle (2004) has been extended, for example, to model zero-inflation, extra-Poisson variation in abundance, and to relax the assumption of population closure between visits (see Dénes et al., 2015, for a recent review).

The idea that we can estimate both N and p without marking animals seems appealing. However, this economy of field effort imposes a cost on the analysis. Recaptures of marked animals provide auxiliary data for estimation of p; we can model these data in terms of p without reference to N. In N-mixture modeling, we do not have such auxiliary data and instead rely on the constant p assumption along with a specific choice of hierarchical model on N to ensure identifiability.

In choosing among competing methodologies, it is important that users are fully-informed about comparative weaknesses and strengths of different approaches. Here, we describe the N-mixture model and highlight the information loss that makes reliable statistical modeling of N and p problematic using count data alone.

1.1. *N-Mixture Model*

N-mixture models describe count data obtained from repeated visits to multiple sites. They exploit the fact that given replicate counts $n_1, \ldots, n_J$ and the model $n_j \overset{iid}{\sim} \text{Binomial}(N, p)$ $(j = 1, \ldots, J; J > 1)$ it is possible to estimate both parameters N and p. N-mixtures extend this model to collections of counts $\{n_{ij}\}$ sampled across sites $(i = 1, \ldots, I)$, with $N_i \overset{iid}{\sim} \text{Poisson}(v)$ or more heavily parameterized alternatives. The assumptions of (i) conditionally independent binomial replicates each with the same abundance N_i, and (ii) constant detection are both critical. The model for N helps with estimation, but as we show below the reliance on a constant p and a family of distributions to model the collection $\{N_i\}$ makes the resulting inference non-robust. In particular, scientifically plausible alternative models for the counts n_{ij} may not feature N_i (or its expectation) as a parameter, or N_i may feature but be non-identifiable.

The binomial count assumption underpinning N-mixture models has critical limitations:

(i) It is a strong assumption that there exists a closed population of N_i animals to be sampled at site i each visit. When the sampling area is not well defined, prescribing such a population may result in the introduction of individual heterogeneity in p that cannot be controlled for by observable covariates.
(ii) The binomial assumption requires that individuals do not get counted multiple times in a single replicate. Enforcing this without marking may be difficult. A Poisson model, conditional on N, may be preferable to the binomial and allows for sampling with replacement.
(iii) Uncontrolled variation in p among visits leads to model specification problems as we show in Section 2.
(iv) Estimation of the binomial index N from repeated counts is practically difficult and relies on second moment assumptions (Carroll and Lombard, 1985). Such difficulties flow through to N-mixture models for N (Dennis et al., 2015).

1.2. *Structure of the Remainder of the Article*

In Section 2, we consider repeated binomial counts as incomplete data for a standard closed population capture–recapture model. We show the extent of information loss when individuals in the population are not marked. In Section 3, we compare N-mixture models to an alternative that describes the counts as observations from a double Poisson distribution. This double Poisson model has the same first two moments as the N-mixture model and describes a situation where replicate counts do not represent samples from a clearly defined population as required for the N-mixture model.

In Section 4, we compare N-mixture modeling to Poisson regression and show that it is parameters describing relative differences in N and p that can be reliably estimated under both models. We conclude that N-mixture models offer little practical benefit over Poisson regression for count data with inference restricted to relative abundance.

2. A Capture-Recapture Perspective

We can think of the replicate binomial count model for a single site that underpins N-mixtures as a model for incomplete capture–recapture data (a latent multinomial *sensu* Link et al., 2010). Let ω_j be an indicator of capture or recapture on occasion j, and write $\boldsymbol{\omega} = (\omega_1\ \omega_2\ \ldots\ \omega_J)$ as a capture history for an animal over J occasions. Using Ω to denote the complete set of capture histories, including the null history, and $x_{\boldsymbol{\omega}}$ ($\boldsymbol{\omega} \in \Omega$) to denote the count of the number of individuals with capture history $\boldsymbol{\omega}$, the observable counts in a three-sample example are $x_{111}, x_{110}, \ldots, x_{001}$. The inference problem amounts to inferring x_{000}, the number of individuals with the null history 000.

In the N-mixture approach, we forgo marking and cannot identify recaptures. The observable data at a particular site in our three sample example, are now $n_1 = x_{111} + x_{101} + x_{110} + x_{100}$, $n_2 = x_{111} + x_{110} + x_{011} + x_{010}$, and $n_3 = x_{111} + x_{101} + x_{011} + x_{001}$ (cf Dennis et al., 2015).

This perspective makes clear what we are giving up when we move from capture–recapture experiments to instead modeling collections of counts. We have taken a capture–recapture problem, with all of its associated difficulties, and replaced it with one in which the capture summaries x_ω are now latent. The summaries $n_1, \ldots, n_J$ represent a reduction of the original capture–recapture data but they are not sufficient statistics for any of the closed population capture–recapture models in the series M_0, …, M_{tbh} (Otis et al., 1978), even assuming constant p.

The capture–recapture perspective also helps us to understand the likelihood underlying the N-mixture problem. Under model M_t (Otis et al., 1978), we can write the model as

$$[\{x_{\boldsymbol{\omega}}\}|N, \boldsymbol{p}] \propto \underbrace{\prod_{j=2}^{J} \frac{\binom{M_j}{m_j}\binom{N-M_j}{n_j-m_j}}{\binom{N}{n_j}}}_{\mathcal{L}_1=[\{m_j\}|\{n_j\},N]} \times \underbrace{\prod_{j=1}^{J} \binom{N}{n_j} p_j^{n_j}(1-p_j)^{N-n_j}}_{\mathcal{L}_2'=[\{n_j\}|N,\boldsymbol{p}]} \quad (1)$$

where

m_j = the number of marked animals caught in sample j ($j = 2, \ldots, J$),
M_j = the number of marked animals in the population immediately before the time of the jth sample ($j = 2, \ldots, J$).

The term $\mathcal{L}_1$ in (1) corresponds to an extension of the simple hypergeometric sampling model for a two-sample capture–recapture study conditional on the captures $\{n_j\}$ (Seber, 1982). Under model M_t, we can treat $\mathcal{L}_1$ as a partial likelihood and this contains virtually all the information about abundance that we can extract from this model (Schofield and Barker, 2016).

Anticipating the N-mixture model, we now consider model M_0 in which $p_1 = p_2 = \ldots = p_J = p$. We can write the model as

$$[\{x_{\boldsymbol{\omega}}\}|N, p] \propto \underbrace{\prod_{j=2}^{J} \frac{\binom{M_j}{m_j}\binom{N-M_j}{n_j-m_j}}{\binom{N}{n_j}}}_{\mathcal{L}_1=[\{m_j\}|\{n_j\},N]} \times \underbrace{\frac{\prod_{j=1}^{J}\binom{N}{n_j}}{\binom{JN}{t}}}_{\mathcal{L}_2=[\{n_j\}|N,t]} \times \underbrace{\binom{JN}{t} p^t(1-p)^{JN-t}}_{\mathcal{L}_3=[t|N,p]} \quad (2)$$

where $t = \sum_j n_j$.

We make two observations:

(i) Through $\mathcal{L}_2$, the statistics $n_1, \ldots, n_J|t$ now contain some information about N. It is this information that the N-mixture model exploits.
(ii) The product $\mathcal{L}_2 \times \mathcal{L}_3$ corresponds to the likelihood function for the model $\{n_j\}|N \overset{iid}{\sim} \text{Binomial}(N, p)$. Relative to $\mathcal{L}_1$ the residual information provided by the

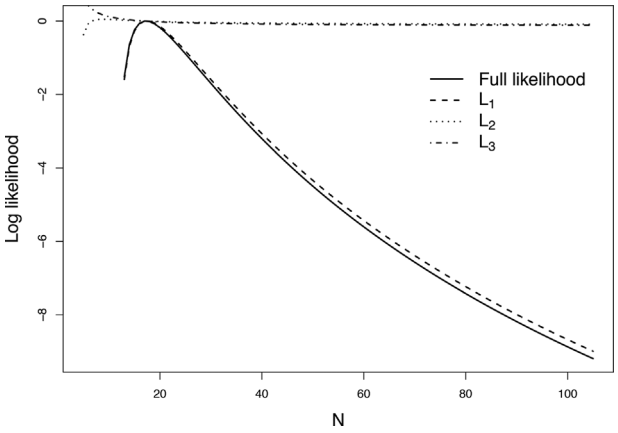

Figure 1. Contributions of the terms $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ to the full likelihood scaled so that the full likelihood is equal to $\mathcal{L}_1 \times \mathcal{L}_2 \times \mathcal{L}_3$. The data were simulated for $k = 5$ sample capture–recapture study with a closed population of 20 individuals and $p = 0.25$ each occasion.

counts $n_1, \ldots, n_J$ through the terms $\mathcal{L}_2$ and $\mathcal{L}_3$ is small (Figure 1).

For N-mixture modeling the data $m_2, \ldots, m_J$ in (1) and (2) are missing. Summing across the missing data in (1) leads to

$$[n_1, \ldots, n_J | N, \boldsymbol{p}] = \prod_{j=1}^{J} \binom{N}{n_j} p_j^{n_j} (1 - p_j)^{N - n_j} = \mathcal{L}_2'$$

and reliable inference about N and $\boldsymbol{p}$ is not possible as the model is over-specified.

Under M_0, summing across the missing data in (2) leads to

$$\mathcal{L}_2 \times \mathcal{L}_3 = \frac{\prod_{j=1}^{J} \binom{N}{n_j}}{\binom{JN}{t}} \times \binom{JN}{t} p^t (1 - p)^{JN - t} \qquad (3)$$

The corresponding model $n_j | N \overset{iid}{\sim} \text{Binomial}(N, p)$ forms a central part of the N-mixture likelihood and is identifiable. However:

(i) Identifiability depends critically on the assumption $p_1 = p_2 = \ldots = p_J$ (or equivalently that the p_j are deterministic functions of time-specific covariates). In capture–recapture studies, this assumption is not required for identifiability and moreover, there is some robustness to departure from this assumption owing to the dominance of the term $\mathcal{L}_1$ in (2) evident in Figure 1.

(ii) Identifiability also depends critically on the assumption of binomial variation in n_j. If instead, we assume that $n_j \overset{iid}{\sim} \text{Poisson}(Np)$ then

$$[n_1, \ldots, n_J | N, p] = \frac{t!}{\prod_i n_j!} \left(\frac{1}{J}\right)^t \times \frac{e^{-JNp} (JNp)^t}{t!}. \qquad (4)$$

Under the Poisson model the n_j given their total t are ancillary statistics (i.e., their distribution is fully known) and the only estimable quantity is $\lambda = Np$. Contrast this with capture–recapture in which modeling the counts as Poisson does not lead to non-identifiability of abundance (Cormack, 1992; Schofield and Barker, 2016).

The Poisson model has to be considered seriously given that it arises as the limiting case of a binomial with $N \to \infty$ and $p \to 0$ while holding $\lambda = Np$ fixed. For sparse data (i.e., small $\mathrm{E}[n_j]$), we would expect the Poisson and binomial models to be near indistinguishable. This also explains in part the instability of the model (3) that arises when the sample mean and variance are close in value (Dennis et al., 2015). Given that N-mixture modeling was developed with sparse data in mind (Royle, 2004), we would anticipate that a simple Poisson model is a plausible alternative representation of the data but one in which N is no longer identifiable. Importantly, this fundamental non-identifiability of the Poisson model cannot be addressed simply through replication in space or time.

2.1. *N-Mixtures and the Multivariate Poisson Distribution*

The N-mixture model exploits spatial replication, in which we now have site-specific abundances denoted N_i, and describes the collection $\{N_i\}$ as exchangeable random variables drawn from a known distribution. Modeling $N_i \overset{iid}{\sim} \text{Poisson}(\nu)$, the capture–recapture model for the complete data $\{x_{i\boldsymbol{\omega}}\}$ can be written as

$$\begin{aligned}[\{x_{i\boldsymbol{\omega}}\}, \{N_i\} | \boldsymbol{p}] &= \prod_i \text{Multinomial}(\{x_{i\boldsymbol{\omega}}\}; N_i, \boldsymbol{\pi}) \times \text{Poisson}(N_i; \nu) \\ &= \prod_i \prod_{\boldsymbol{\omega}} \frac{e^{-\nu\pi_{\boldsymbol{\omega}}} (\nu\pi_{\boldsymbol{\omega}})^{x_{i\boldsymbol{\omega}}}}{x_{i\boldsymbol{\omega}}!} \\ &= \prod_i \prod_{\boldsymbol{\omega}} \text{Poisson}(\{x_{i\boldsymbol{\omega}}\}; \nu\pi_{\boldsymbol{\omega}}), \end{aligned} \qquad (5)$$

where $\boldsymbol{\pi} = \{\pi_{\boldsymbol{\omega}}\}$ represent simple functions of the capture histories and $\boldsymbol{p}$. This is a complete data likelihood for the N-mixture model. Deriving an observed data likelihood, $[\{n_{ij}\} | \nu, p]$, from (5) by summing across latent quantities while holding the observation $\{n_{ij}\}$ fixed leads to a multivariate Poisson model (Dennis et al., 2015). This is the marginal model for $\{n_{ij}\}$ for N-mixtures when $N_i \overset{iid}{\sim} \text{Poisson}(\nu)$ and has the property that the marginal mean and variance are identical (see supplementary materials).

3. Binomial versus Poisson Models

If we extend the simple Poisson model (4) across sites, we have

$$[\{n_{ij}\} | \{\lambda_i\}] = \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{e^{-\lambda_i} \lambda_i^{n_{ij}}}{n_{ij}!}. \qquad (6)$$

Like (5), model (6) has the property that the marginal mean and variance are identical (see supplementary materials).

The main differences are that (i) the covariances among counts at the same site are zero in model (6) and non-zero in N-mixture model (5), and (ii) that conditional on N_i, the N-mixture model allows the counts to have their variance less than their mean while in the Poisson model they are equal. Below we introduce the double Poisson model (Efron, 1986) that allows both under- and over-dispersion.

Under the constant p assumption with $\lambda_i = pN_i$, we can use model (6) to estimate λ_i/λ_h, the abundance at site i relative to site h. We return to this point in Section 4. Alternatively, given the potentially large dimension of $\{\lambda_i\}$, we might wish to model them further by way of summary or to obtain other benefits such as shrinkage for improved estimation (James and Stein, 1961).

The use of a hierarchical model on the expected count is exploited in the N-mixture for a different reason than providing a parsimonious summary. It is necessary to help with identifiability, and is accomplished through modeling $N_i \overset{iid}{\sim} \text{Poisson}(\nu)$. However, the ability to tease apart N_i and p in the N-mixture model depends entirely on the covariance structure introduced in this representation.

In their excellent analysis of the N-mixture model, Dennis et al. (2015) show, among other things, that excessively high and even infinite estimates of abundance can readily occur, particularly when p and J are small. When p is small, the binomial variance is approximately equal to the mean, and the binomial and Poisson models are near indistinguishable leading to non-identifiability of abundance.

3.1. *Can We Discriminate Among Models?*

As an alternative to the conditional binomial model in the N-mixture formulation, we might instead suppose that given our inability to identify animals, we have sampling with replacement. Conditional on N_i the counts can be modeled as Poisson with mean pN_i ($p > 0$). But then N_i is not identifiable as we have shown in Section 2. Or we might instead suppose that counts at sites are simply counts, perhaps over- or under-dispersed Poisson random variables that are unbounded but related by a common site effect governing their mean. In this formulation there may be no sensible notion of an abundance N_i.

Given that N does not feature in (6), or is not identifiable in other formulations, an important question then, is to what extent are we able to discriminate among different hierarchical representations of the count data $\{n_{ij}\}$? We answer this by counter-example using simulation to compare two models under three scenarios. The two models are

Model 1: N-mixture: $n_{ij}|N_i \overset{iid}{\sim} \text{Binomial}(N_i, p)$ where $N_i \overset{iid}{\sim} \text{Poisson}(\nu)$;

Model 2: Double Poisson-Lognormal (DPLN): $n_{ij}|\lambda_i, \theta \overset{iid}{\sim} \text{DoublePoisson}(\lambda_i, \theta)$ where $\lambda_i \overset{iid}{\sim} \text{Lognormal}(\mu, \sigma^2)$ with restriction $\theta = \mu/(\mu - \sigma^2)$.

The double Poisson distribution, fully-described by Efron (1986), is a member of the exponential family that introduces a second parameter $\theta > 0$ to control the variance independently of the mean. The parameter θ gives the double Poisson considerable flexibility, allowing both under-dispersion ($\theta > 1$) and over-dispersion ($\theta < 1$). It includes as a special case the simple Poisson model ($\theta = 1$); we can match first and second moments with the binomial distribution.

To facilitate comparison with the N-mixture model, we restricted the two families (N-mixture and DPLN) to have identical first and second moments (see supplementary materials) by using the constraint $\theta = \mu/(\mu - \sigma^2)$ for the DPLN model. We do not advocate for this restricted version when model fitting, but rather we chose it to simplify comparison by ensuring that the two families of models had similar moments controlled by the same number of parameters in their marginal distributions.

Our three scenarios were $\text{E}[n_{ij}] = 3$, corresponding to $\nu = 30$, $p = 0.1$ in the N-mixture model (scenario A), $\text{E}[n_{ij}] = 9$, corresponding to $\nu = 30$, $p = 0.3$ in the N-mixture model (scenario B), and $\text{E}[n_{ij}] = 15$, corresponding to $\nu = 30$, $p = 0.5$ in the N-mixture model (scenario C). Parameters for the restricted double Poisson model were chosen to match first and second moments (Table 1).

Our scenarios are ordered by decreasing realism. Royle (2004) developed N-mixture modeling with sparse data in mind and this is reflected in applications. For example, the mallard data used in the R package `unmarked` (Swiss Breeding Bird Survey data from Kery et al., 2005) has 239 sites with three visits with a mean value for n_{ij} of 0.24 birds per visit per site and a median of 0. The great tit example data in Chapter 6 of Kéry and Royle (2016), also Swiss Breeding Bird Survey data, has 263 sites with three visits and a mean count of 6.45 birds per site per visit with a median of 4; one third of the great tit counts are 0. The simulation assessment of a simple N-mixture model early in Chapter 6 of Kéry and Royle (2016) has 150 sites, two visits per site and a mean value for n_{ij} of 1.0 animals per visit per site.

Table 1

Parameters used in simulations for the N-mixture and double Poisson models with corresponding marginal moments. The moments are common to both models. Covariances not included in the table are all zero.

	N-mixture		Double poisson		Moments		
Scenario	ν	p	μ	σ^2	$\text{E}[n]$	$\text{Var}(n)$	$\text{Cov}(n_{ij}, n_{ik})(j \neq k)$
A	30	0.1	3	0.3	3	3	0.3
B	30	0.3	9	2.7	9	9	2.7
C	30	0.5	15	7.5	15	15	7.5

We simulated data 1000 times under each model and for each scenario, generating data across 50 sites, with $J = 5$ replicates at each site. We fitted each model under the three scenarios; this leads to six comparisons, and judged each on the basis of the marginal likelihood (the code is available in the supplementary materials). With each model having two parameters in the marginal distribution, twice the difference in maximum log-likelihood corresponds to ΔAIC.

When data are simulated under the N-mixture model the double Poisson model was either preferred or within 1 AIC unit of the N-mixture model in 82% (scenario A), 59% (scenario B), or 41% (scenario C) of cases indicating an inability to discriminate between the two models (Figure 2) even when the expected counts are much larger than typically seen in application and with $J = 5$. Similarly, when the data were generated according to a double Poisson distribution, the N-mixture model was either preferred or within 1 AIC unit of the double Poisson model in 77% (scenario A), 51% (scenario B), or 42% (scenario C) of cases. Even in the best scenario we prefer the non-generating model about 25% of the time. On the basis of model fit we appear unable to discriminate unequivocally among competing models, some of which do not feature N. This is particularly the case when the data are sparse.

4. Relative Abundance: Poisson Models as an Alternative to N-Mixtures

Intuitively, and under the constant p assumption, the replicate counts $n_{i1}, \ldots, n_{ik}$ provide good information on the expected count $\mathrm{E}[n_{ij}] = \mu_i$, in the sense that the sample mean $\bar{n}_i$ is a minimum variance unbiased estimator for μ_i. This is in consequence of the Gauss–Markov theorem and is subject only to mild conditions on the data generating model. Thus, when p is constant or controlled by covariates, count data can provide reliable information on relative abundance. These are precisely the conditions assumed in N-mixture modeling. However, reasonable alternative approaches exist that do not require the strong assumptions associated with the binomial component of the N-mixture model.

We believe that Poisson modeling of the counts is a straightforward alternative to N-mixture modeling if the aim is to extract the information that the $I \times J$ counts reliably contain; *viz*, the information about relative abundances among sites after controlling for p.

In this approach, we can model

$$n_{ij} \sim \text{Poisson}(\mu_{ij}).$$

For now, we assume no variation in p associated with visits (indexed by j) and model μ_{ij} as

$$\mu_{ij} = pN_i = e^{\alpha_p + \alpha_N + \boldsymbol{z}_i' \boldsymbol{\beta}_N}, \tag{7}$$

where $\boldsymbol{z}_i$ are site-specific covariates and $\boldsymbol{\beta}_N$ site-covariate effects. We also relax the restriction that $p < 1$; without marking it is difficult to avoid multiple counting of individuals except when there are very few individuals present. Therefore, there is no reason to suppose that $\mathrm{E}[n_i]$ is strictly less than N_i. We let $t_i = \sum_j n_{ij}$ be the site total and $T = \sum_i t_i$ be the total count. We can then factorize the Poisson model as

$$\begin{aligned}[\{n_{ij}\}|\{\mu_{ij}\}] &= \prod_i \prod_j \frac{e^{-\mu_{ij}} \mu_{ij}^{n_{ij}}}{n_{ij}!} \\ &= \underbrace{\prod_i \frac{t_i!}{\prod_j n_{ij}!} \prod_j \left(\frac{1}{J}\right)^{n_{ij}}}_{[\{n_{ij}\}|\{t_i\}]} \\ &\quad \times \underbrace{\frac{T!}{\prod_i t_i!} \prod_i \left(\frac{\xi_i}{\sum_h \xi_h}\right)^{t_i}}_{[\{t_i\}|T, \{\boldsymbol{\beta}_N\}]} \times \underbrace{\frac{e^{-\phi} \phi^T}{T!}}_{[T|\phi]}\end{aligned} \tag{8}$$

where

$$\frac{\xi_i}{\sum_h \xi_h} = \frac{N_i e^{\alpha_p}}{\sum_h N_h e^{\alpha_p}} = \frac{N_i}{\sum_h N_h} = \frac{e^{\boldsymbol{z}_i' \beta_N}}{\sum_h e^{\boldsymbol{z}_h' \beta_N}}.$$

and

$$\phi = J e^{\alpha_p} \sum_i N_i.$$

In (8) the parameters α_p and α_N are wholly confounded in ϕ. We can obtain maximum likelihood estimators (MLEs) for the parameters $\boldsymbol{\beta}_N$ from the partial likelihood given by the term $[\{t_i\}|T, \boldsymbol{\beta}_N]$. These $\boldsymbol{\beta}_N$ represent the information we can reliably extract; they describe relative abundances.

It is straight-forward to model covariates of detection in this framework. If we use $\boldsymbol{w}_{ij}$ for site- and visit-specific covariates with associated parameters $\boldsymbol{\beta}_p$ then the model (7) becomes

$$\mu_{ij} = e^{\alpha_p + \alpha_N + \boldsymbol{w}_{ij}' \boldsymbol{\beta}_p + \boldsymbol{z}_i' \boldsymbol{\beta}_N}. \tag{9}$$

Correspondingly, we can modify the partitioning in (8) to give

$$\begin{aligned}[\{n_{ij}\}|\{\mu_{ij}\}] &= \underbrace{\prod_i \frac{t_i!}{\prod_j n_{ij}!} \prod_j \left(\frac{e^{\boldsymbol{w}_{ij}' \boldsymbol{\beta}_p}}{\sum_h e^{\boldsymbol{w}_{ih}' \boldsymbol{\beta}_p}}\right)^{n_{ij}}}_{[\{n_{ij}\}|\{t_i\}, \{\boldsymbol{w}_{ij}\}, \boldsymbol{\beta}_p]} \\ &\quad \times \underbrace{\frac{T!}{\prod_i t_i!} \prod_i \left(\frac{\xi_i}{\sum_h \xi_h}\right)^{t_i}}_{[\{t_i\}|\boldsymbol{\beta}_p, \boldsymbol{\beta}_N]} \times \underbrace{\frac{e^{-\phi} \phi^T}{T!}}_{[T|\phi]}\end{aligned} \tag{10}$$

where

$$\xi_i \propto e^{\boldsymbol{z}_i' \boldsymbol{\beta}_N} \sum_j e^{\boldsymbol{w}_{ij}' \boldsymbol{\beta}_p}$$

and

$$\phi = \sum_i \xi_i.$$

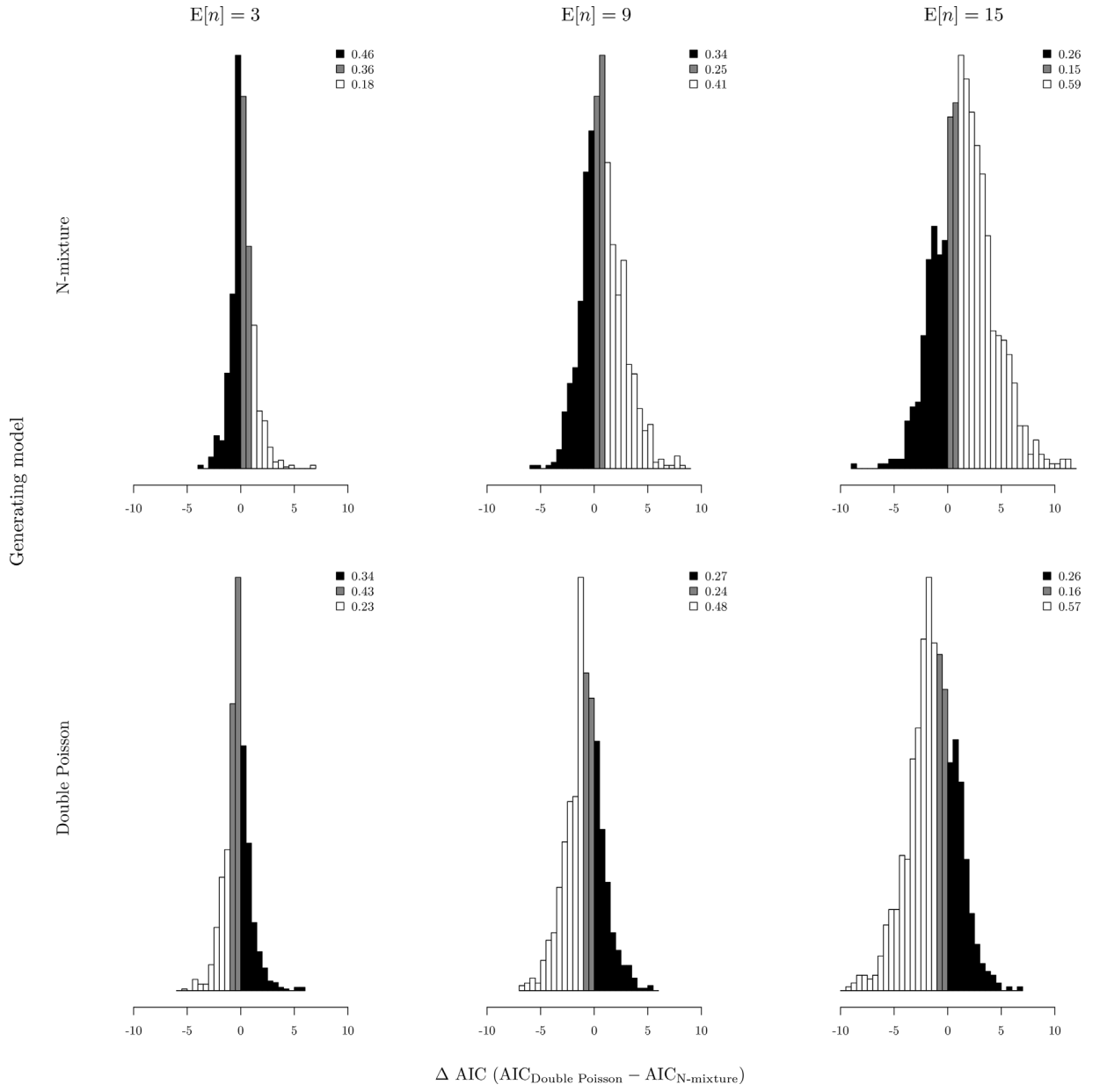

Figure 2. Comparative model fit as measured by ΔAIC for the N-mixture reference model compared to a 2-parameter double Poisson–Lognormal model. Both models were fitted to data generated under each model and according to three scenarios: $E[n_{ij}] = 3$ (scenario A), $E[n_{ij}] = 9$ (scenario B), and $E[n_{ij}] = 15$ (scenario C). Each case represents results from 1000 simulated data sets. Positive values of ΔAIC favor the N-mixture model. Black shading indicates that the non-generating model was preferred. Grey shading indicates that the non-generating model is within 1 AIC unit; the legend indicates the corresponding proportions. The top row of figures data were generated under the N-mixture model; the bottom row data under our restricted double Poisson-Lognormal model.

Now the first two terms allow us to model detectability through $\boldsymbol{\beta}_p$ and relative abundances through $\boldsymbol{\beta}_N$, but α_p and α_N are confounded in ϕ.

The model in (9) is in the standard form for Poisson regression. As with the simpler models we have examined, the baseline levels of abundance α_N and detection α_p are confounded; all we can model is $\alpha = \alpha_N + \alpha_p$. As this is a standard Poisson regression, we can employ the full range of model fitting procedures, including diagnostics such as deviance based goodness-of-fit testing. Moreover, we can investigate random effect extensions to allow for a wider range of models. That is, ignoring detection, we have a full spectrum that ranges from a saturated model, in which all sites have different $E[n_{ij}]$, to a null model, in which they are all the same. Between these two endpoints, we may have covariate and random effects, with random effect

modeling used for partial pooling to improve inference about relative abundance rather than as an operational necessity to allow identifiability of N.

4.1. *Example: Alder Flycatcher Data*

We analyze a subset of the alder flycatcher (ALFL) data of Chandler et al. (2009) (see supplementary materials for the source of these data; only this subset is available at this source). The ALFL data were collected from $I = 50$ sites with $J = 3$. Associated with the visits are two covariates: time (t) and date (d). Associated with the sites are covariates: extent of woody cover (w) and plant vertical structure (s). We standardized all covariates.

The N-mixture model for the ALFL data is:

$$\begin{aligned} n_{ij} &\sim \text{Binomial}(N_i, p_{ij}), \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \\ N_i &\sim \text{Poisson}(v_i), \quad i = 1, \ldots, I, \\ \text{logit}(p_{ij}) &= \alpha_p + \beta_1 t_{ij} + \beta_2 d_{ij} \\ \log(v_i) &= \alpha_v + \gamma_1 w_i + \gamma_2 s_i. \end{aligned}$$

This model was fitted in the R package `unmarked`.

To compare, we also fitted three other models. The first was a standard Poisson regression, modeled in terms of the confounded mean $\mu_{ij} = p_{ij} v_i$, with

$$\begin{aligned} n_{ij} &\sim \text{Poisson}(p_{ij} v_i), \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \\ \log(p_{ij}) &= \alpha_p + \beta_1 t_{ij} + \beta_2 d_{ij} \\ \log(v_i) &= \alpha_v + \gamma_1 w_i + \gamma_2 s_i. \end{aligned}$$

Combining, we have

$$\begin{aligned} n_{ij} &\sim \text{Poisson}(\mu_{ij}), \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \\ \log(\mu_{ij}) &= \alpha + \beta_1 t_{ij} + \beta_2 d_{ij} + \gamma_1 w_i + \gamma_2 s_i, \end{aligned}$$

where $\alpha = \alpha_p + \alpha_v$. This model was fitted using `glm` in R.

For the second model, we fitted a random effects Poisson regression where a site-specific intercept was modeled by a normal distribution. This model was fitted using `glmer` from the `lme4` package in R.

The third model was a random-effects double Poisson model

$$\begin{aligned} n_{ij} &\sim \text{DoublePoisson}(\mu_{ij}, \theta_{ij}), \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \\ \log(\mu_{ij}) &= m_{ij} + \varepsilon_i, \\ m_{ij} &= \alpha_i + \beta_1 t_{ij} + \beta_2 d_{ij} + \gamma_1 w_i + \gamma_2 s_i, \\ \alpha_i &\sim N(\mu_\alpha, \sigma_\alpha^2) \end{aligned}$$

and with $\log(\log(\theta_{ij})) = m_{ij}$. The log–log link function was chosen to mimic the N-mixture model in having Poisson-like behavior for small μ and under-dispersion otherwise. Like the N-mixture model, the dispersion and the mean parameters are functionally related, and together controlled by the same covariates. We fitted this model using `integrate` to find the marginal likelihood and `optim` to maximize it in R. The code for fitting all models is available in the supplementary materials.

Regardless of the type of model (N-mixture vs Poisson), we are led to similar inference about the site-specific effects γ_1 and γ_2 (Figure 3). The explanation is that γ_1 and γ_2 describe how the relative abundance depends on the covariate. It is this information that can be reliably extracted from the N-mixture model but we can simply do this via Poisson regression.

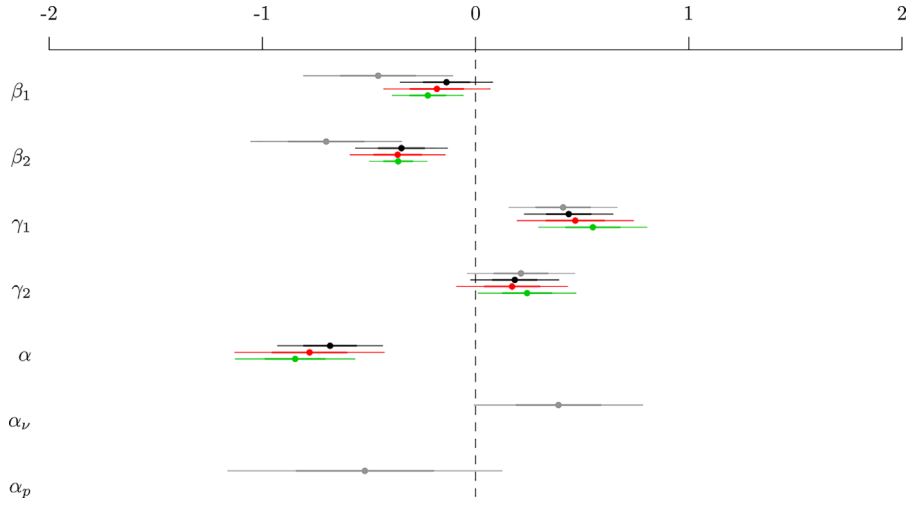

Figure 3. Comparison of model parameter estimates after fitting the N-mixture model (gray), Poisson regression model (black), random effects Poisson regression model (red), and the double Poisson regression model (green) to the ALFL data by maximum likelihood. The parameters β_1 and β_2 represent the visit-specific covariates "time" and "date"; γ_1 and γ_2 the effect of the site-specific covariates "woody cover" and "plant structure." Plotted are the maximum likelihood estimates (enlarged dot), ± one standard deviation (thick horizontal lines) and ± two standard deviations (thin horizontal lines).

Table 2
Model fitting summaries for the four models fitted to the alder flycatcher data.

Model	Log-likelihood	No. par.	AIC
N-mixture	−131.11	6	274.21
Poisson regression	−139.57	5	289.14
Random-effects poisson regression	−139.12	6	290.24
Random-effects double poisson	−130.76	6	273.52

Provided we can get a reliable estimate of the expected count at a site then we should be able draw reasonable inferences about relative abundance. The similarity of inference about relative abundance is despite what appears to be clear differences among some of the four models in terms of AIC, with the double-Poisson model preferred (Table 2).

Superficially, inference about the visit-specific covariates β_1 and β_2 and the confounded intercept $\alpha = \alpha_p + \alpha_v$ appear to differ. However, this is due to the different link functions adopted for p: logit for the binomial model, log for the Poisson. Inference about α_p and α_v is very different. They are completely confounded under Poisson regression. Their identifiability under the N-mixture model is entirely determined by an assumption of repeated binomial counts instead of repeated Poisson counts.

Overall, Figure 3 shows that while the baseline abundance and detection parameters suffer from model sensitivity (as we have highlighted in the sections above), the relative changes in abundance are identifiable, and we are led to similar inference about relative abundance regardless of our choice of model.

5. Discussion

The appeal of N-mixture modeling is that we can infer properties of populations from repeated count data without the expense associated with marking. However, as we show in Section 2, the expedient of substituting auxiliary data from recaptures by a hierarchical model involves a critical loss of information. A consequence is that even under ideal conditions, it is difficult to discriminate between N-mixture models and other reasonable alternatives for count data as we show in Section 3. Importantly, for other choices of hierarchical model for N_i such as a negative binomial or zero-inflated Poisson, we will be able to identify alternative descriptions that also have N_i missing or confounded and that will fit the data equally well (see supplementary materials for binomial/negative-binomial example). Thus, we expect the problems that we have identified for the basic N-mixture model to carry over to more complicated implementations.

This is problematic for N-mixture modeling because we are attempting to decompose the model for $\mathrm{E}[n_{ij}]$ into two factors N_i and p without auxiliary information for making inference about p, such as is provided by recaptures of marked animals. While the N-mixture model allows us to do this, scientifically plausible alternatives, such as our counter example, do not. The problem is not an inability to fit the N-mixture model, but rather an inability of the data to discriminate among models that have vastly different implications for inference about N (Gelman et al., 2014, p. 190).

Only by assuming that (i) these alternatives do not hold, and (ii) p is constant or fully explained by covariates, can we make inference about N. While strong assumptions are often made in statistical modeling, in the case of N-mixture modeling there will be little residual information to allow checking of these assumptions and that indicate that alternative models outside the N-mixture family should be considered. This will be especially the case when the data are sparse.

A commonly stated advantage of the N-mixture model is that it allows for covariates of detectability to be modeled as well as abundance. There is a misconception that alternatives to N-mixtures do not allow simultaneous modeling of variation due to variation in abundance and variation in detectability (Joseph et al., 2009; Dénes et al., 2015). As some of us have shown previously (Link and Sauer, 1997), and is clear from the partitioning of equation (10), random effect Poisson models do allow for simultaneous modeling of detectability and variation in abundance. It is true, that with the N-mixture model we can have the same covariate for both site effects and detectability. While this may sound appealing, we offer a note of caution. The ability to use identical covariates for the two aspects of the model depends entirely on the validity of the assumed conditional binomial structure for the replicate counts and the assumption that variation in detectability is fully explained by this covariate.

Our message is that fundamentally count data under imperfect detection can only be reliably used as indices. Having controlled for detectability using covariates, counts through ratios of their expectations provide reliable information about relative abundance, but not absolute abundance N or p. Although our simulations show that with increasing $\mathrm{E}[n_{ij}]$, we can begin to discriminate among models, the values required are much larger than seen in typical application. Even with $\mathrm{E}[n_{ij}] = 15$ in our simulations, we couldn't clearly differentiate between models. If users have discretionary effort and they require reliable information about absolute abundance, our advice is that this effort should be spent on acquiring additional data, such as recaptures of marked individuals to allow direct estimation of p.

6. Supplementary Materials

Supplementary material referred to in Section 2.1 is available with this article at the *Biometrics* website on Wiley Online Library along with R code for carrying out the simulations in Section 3 and for the analysis of the flycatcher data in Section 4.1.

Acknowledgements

We thank Murray Efford and Jim Nichols for comments on drafts of this manuscript. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Carroll, R. J. and Lombard, F. (1985). A note on N estimators for the binomal distribution. *Journal of the American Statistical Association* **80**, 423–426.

Chandler, R. B., King, D. I., and DeStefanao, D. I. (2009). Scrub-shrub bird habitat associations at multiple spatial scales in beaver meadows in Massachusetts. *The Auk* **126**, 186–197.

Cormack, R. M. (1992). Interval estimation for mark-recapture studies of closed populations. *Biometrics* **48**, 567–576.

Dénes, F. V., Silveira, L. F., and Beissinger, S. R. (2015). Estimating abundance of unmarked animal populations: Accounting for imperfect detection and other sources of zero inflation. *Methods in Ecology and Evolution* **6**, 543–556.

Dennis, E. B., Morgan, B. J. T., and Ridout, M. S. (2015). Computational aspects of N-mixture models. *Biometrics* **71**, 237–246.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* **81**, 709–721.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis (Third edition)*. Boca Raton, Florida, USA: Chapman and Hall/CRC.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, California: University of California Press.

Joseph, L. N., Elkin, C., Martin, T. G., and Possingham, H. P. (2009). Modeling abundance using N-mixture models: The importance of considering ecological mechanisms. *Ecological Applications* **19**, 631–642.

Kéry, M. and Royle, J. A. (2016). *Applied Hierarchical Modeling in Ecology, Volume 1*. Amsetrdam: Elsevier.

Kéry, M., Royle, J. A., and Schmid, H. (2005). Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications* **15**, 1450–1461.

Link, W. A. and Sauer, J. R. (1997). Estimation of population trajectories from count data. *Biometrics* **53**, 488–497.

Link, W. A., Yoshizaki, J., Bailey, L. L., and Pollock, K. H. (2010). Uncovering a latent multinomial: analysis of mark-recapture data with misidentification. *Biometrics* **66**, 178–185.

Otis, D. L., Burnham, K. P., White, G. C. anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **62**, 1–135.

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics* **60**, 108–115.

Schofield, M. R. and Barker, R. J. (2016). 50-year-old curiosities: Ancillarity and inference in capture-recapture models. *Statistical Science* **31**, 161–174.

Seber, G. A. F. (1982). *Estimating Animal Abundance and Related Parameters. Second edition.* London, UK: Charles Griffin and Co..

Received March 2016. *Revised March* 2017. *Accepted April* 2017.