ADVANCED REVIEW

A review of N-mixture models

Lisa Madsen[1] [iD] | J. Andrew Royle[2]

[1]Department of Statistics, Oregon State University, Corvallis, Oregon, USA

[2]USGS Patuxent Wildlife Research Center, Laurel, Maryland, USA

Correspondence
Lisa Madsen, Department of Statistics, Oregon State University, Corvallis, OR, USA.
Email: madsenl@oregonstate.edu

Edited by: David Scott, Review Editor and Co-Editor-in-Chief

Abstract

N-mixture models were born in 2004 of the necessity to model animal population size from point counts with imperfect detection of individuals, where capture-recapture methods are infeasible. Initially developed for applications where population size was assumed constant, N-mixture models were extended in 2011 to include population dynamics, allowing application to populations whose size fluctuates during the study. A further extension in 2014 accommodates populations with multiple "states" such as age class or sex. More recent extensions model spatial movement of animals among habitat patches or the spatial spread of infectious disease in a human population. The core idea underlying this class of models is a hierarchical structure, where the observation model is defined conditional on the model for true abundance. This hierarchy allows researchers to incorporate information about observation and abundance processes, while permitting distinct inferences about elements affecting detection and those affecting abundance. Another benefit of the hierarchical approach is the ability to accommodate many existing sampling protocols such as removal sampling and distance sampling. One drawback to N-mixture models is that since they estimate both abundance and detection from replicated but unmarked counts, model parameters may not be clearly identifiable. A second drawback is that when observed counts are large, calculating the N-mixture likelihood is computationally infeasible. This difficulty motivated an approximate likelihood based on the normal approximation to the binomial. The normal approximation provides a diagnostic of parameter estimability based on the closed-form expression of the Fisher information matrix for a multivariate normal likelihood.

This article is categorized under:
    Data: Types and Structure > Image and Spatial Data
    Statistical Learning and Exploratory Methods of the Data Sciences > Modeling Methods

KEYWORDS

abundance estimation, dynamic N-mixture model, multi-state N-mixture model, spatial N-mixture model

1 | INTRODUCTION

Royle (2004b) introduced N-mixture models for estimating the size N of a population given imperfect detection of individuals when capture-recapture methods are infeasible due to constraints on resources or due to the nature of the population. The earliest applications were estimating avian abundance from point counts. These data are often sparse: few observations and perhaps a large percentage of 0's. Subsequent applications include small vertebrates (Ficetola et al., 2018), insects (Della Rocca et al., 2020), salamanders (Dodd Jr & Dorazio, 2004; McKenny et al., 2006), feral horses (Schoenecker et al., 2021), deer (Keever et al., 2017), and disease modeling (DiRenzo et al., 2019).

Royle's original model assumed the data consisted of independent detections of individuals from a closed meta-population consisting of a number of independent sites. The closed population assumption implies that each local population is not subject to gains (recruitment or immigration) or losses (mortality or emigration). Hence, the local population sizes do not change over the sampling period. Population closure is usually regarded as an assumption that might hold under a model of "instantaneous" sampling which is seldom reasonable. In practice, deviations from closure result when repeated samples take place over longer periods of time. This has led to specialized models for the "components of detection" (Nichols et al., 2009), and distinctions between the notions of instantaneous population and super-population, the latter of which consists of all individuals exposed to sampling. The linkage between the two is established under formal models of the temporary emigration process (Chandler et al., 2011; Kéry & Royle, 2021).

Extensions of the N-mixture model have relaxed the original assumptions to allow for general models of non-closure in the form of explicit population dynamics (Dail & Madsen, 2011; Hostetler & Chandler, 2015), spatial structure (Brintz et al., 2018; Zhao et al., 2017), multiple demographic states (DiRenzo et al., 2019; Zipkin, Thorson, et al., 2014), and other considerations that provide more realism to the model.

In Section 2, we describe the details of the original N-mixture model for binomial observations, including the maximum likelihood and Bayesian estimation methods. We discuss concerns about parameter identifiability and reference several investigations into the nature of the model that leads to non-identifiability of the parameters. Section 3 details an open-population extension to the original model and references computing resources for implementing the extended model. Section 5 describes a multi-state extension to the open-population model. The multiple states occur when the sampled population can be partitioned into sub-populations, and the size of each sub-population is of interest. Section 4 describes the asymptotic approximation to the open-population model and its associated diagnostic for parameter identifiability. Section 6 details two spatial N-mixture models. The first of these models wildlife populations, and the second models the spread of disease. In each model, spatial dependence among sites is modeled according to the process being modeled.

2 | THE ORIGINAL BINOMIAL N-MIXTURE MODEL

The basic model assumes we have observed counts $n_{i,t}$ of distinct individuals at $i = 1, ..., R$ sites during $t = 1, ..., T$ sampling occasions, where the populations within each of the R sites are assumed closed with respect to births, deaths, and migration during the T occasions. Under this closure assumption, and with the additional assumptions of independent detections and constant detection probability p,

$$n_{i,t} \sim \text{Binomial}(N_i, p) \tag{1}$$

with N_i the size of the population at site i. The parameters of interest are the site population sizes N_i and the total population size $\sum_{i=1}^{R} N_i$. The objective of Royle (2004b) was a model with good performance given sparse data (small R and T, small p, and perhaps some $N_{i,t} = 0$). Prior attempts at estimating binomial index N when both it and success probability are unknown produced estimators that were unstable under these conditions (Carroll & Lombard, 1985; Olkin et al., 1981). Carroll and Lombard (1985) offered an approach that put a prior distribution on p, then integrated p from the joint likelihood and maximized the resulting marginal or "integrated likelihood" for N.

Royle (2004b) took a similar approach but integrated the N_i, rather than p, from the joint likelihood. Specifically, he modeled the N_i as independent identically distributed (iid) random variables with density $f(N_i; \boldsymbol{\theta})$, typically Poisson or negative binomial, then integrated (summed) them from the joint likelihood. The resulting likelihood is

$$L(p,\boldsymbol{\theta}) = \prod_{i=1}^{R} \left\{ \sum_{N_i=\max_t n_{i,t}}^{\infty} \left(\prod_{t=1}^{T} \mathrm{Bin}(n_{i,t};N_i,p) \right) f(N_i;\boldsymbol{\theta}) \right\}, \quad (2)$$

where $\boldsymbol{\theta}$ represents the vector of parameters of the distribution of $N_1,...N_R$ (e.g., Poisson mean λ or negative binomial mean μ and over dispersion parameter α), and

$$\mathrm{Bin}(n_{i,t},N_i;p) = \binom{N_i}{n_{i,t}} p^{n_{i,t}} (1-p)^{N_i - n_{i,t}}$$

represents the binomial likelihood with index N_i and success probability p, evaluated at n_{it}. The lower bound of the sum in (2) is the largest count observed at site i, since if $N_i < n_{i,t}$ for some t, $\mathrm{Bin}(n_{i,t};N_i,p) = 0$.

2.1 | Covariates

Royle (2004b) suggested that both Poisson mean λ and detection probability p could be modeled using covariates such as landscape characteristics. This induces a more flexible model and also allows inference about the relationship between abundance and landscape, as well as separate inference about detection. Dodd and Dorazio (2004) used elevation as a covariate for λ to model abundance of the salamander *P. jordani* at nine sites in Great Smoky Mountains National Park in North Carolina and Tennessee. Section 8.3 of Royle and Dorazio (2008) includes an example analysis of willow tit (*Parus montanus*) data from the Swiss Breeding Bird Survey where the detection probability at site i on sampling occasion j as modeled as a function of sampling intensity and date:

$$\mathrm{logit}\left(p_{i,j}\right) = \alpha_0 + \alpha_1 \mathrm{intensity}_{i,j} + \alpha_2 \mathrm{date}_{i,j} + \alpha_3 \mathrm{date}_{i,j}^2,$$

and the mean abundance at site i is modeled as a function of site characteristics:

$$\log(\lambda_i) = \beta_0 + \beta_1 \mathrm{elevation}_i + \beta_2 \mathrm{elevation}_i^2 + \beta_3 \mathrm{forest\ cover}_i + \beta_4 \mathrm{length}_i.$$

Royle and Dorazio (2008) applied both maximum likelihood and Bayesian analyses to estimate the model parameters. These methods are explained below.

2.2 | Maximum likelihood estimation

The integrated likelihood (2) is approximated by summing over N_i to a large finite bound K, then numerically maximized with respect to $\boldsymbol{\theta}$ and p to obtain maximum likelihood estimates (MLEs) $\widehat{\boldsymbol{\theta}}$ and $\widehat{p}$. Sensitivity to choice of K should be evaluated to ensure a stable likelihood solution has been obtained (Kéry, 2018). Since the sites' population sizes $N_1,...,N_R$ are independent and identically distributed (iid), and since their common expectation can be expressed in terms of parameter vector $\boldsymbol{\theta}$, we can estimate $E(N_i)$ as $\widehat{E}(N_i) \cdot R$, where $\widehat{E}(N_i)$ is $E(N_i)$ with $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. For example, if the N_i are assumed iid Poisson(λ), then $\boldsymbol{\theta} = \lambda$, and $\widehat{E}(N) = \widehat{\lambda} \cdot R$, where here N denotes the total abundance over all sites, $N = \sum_{i=1}^{R} N_i$.

Standard errors of MLEs are usually based on the observed Hessian matrix, the matrix of second derivatives of the negative log integrated likelihood. This matrix is a sample version of the Fisher Information matrix, the inverse of which is the asymptotic variance–covariance matrix of the vector of MLEs. Numerical optimization algorithms return an approximation of the Hessian at the optimum parameter values. The diagonal entries of the inverse of this matrix are taken as variance estimates.

The pcount function in the R (R Core Team, 2021) package unmarked (Fiske & Chandler, 2011) implements numerical optimization of the integrated likelihood. The user may specify fixed-effect covariates for detection p and/or mean abundance λ. The prior distribution of the N_i may be either Poisson, negative

binomial, or zero-inflated Poisson. The function gives the user the ability to customize the numerical optimization by choosing the finite bound K used to approximate the infinite sums in (2), among other adjustments.

2.3 | Bayesian estimation

A "plug in" empirical Bayes approach provides an alternative estimator of $E(N_i)$ (Royle, 2004b). Bayes' rule gives

$$\Pr(N_i = k | n_{i,1}, ..., n_{i,T}; \theta, p) = \frac{\Pr(n_{i,1}, ..., n_{i,T} | N_i = k; p) \Pr(N_i = k; \theta)}{\sum_{k=0}^{\infty} \Pr(n_{i,1}, ..., n_{i,T} | N_i = k; p) \Pr(N_i = k; \theta)}, \quad (3)$$

which provides the probability distribution of N_i in terms of the observed counts $n_{i,1}, ..., n_{i,T}$ and parameters p and θ. Plugging in ML estimates $\widehat{p}$ and $\widehat{\theta}$ yields an estimated posterior distribution of N_i, which can then be used for inference.

Adopting a fully Bayesian analysis is another option. With this approach, in addition to specifying a Poisson or negative binomial prior distribution on the N_i, one must provide prior distributions for all model parameters, that is, p and θ. As a simple example, let

$$n_{i,t} \sim \text{independent Binomial}(N_i, p)$$

$$N_i | \lambda \sim \text{iid Poisson}(\lambda)$$

$$\log(\lambda) \sim N(0, 1000)$$

$$\text{logit}(p) \sim N(0, 1000)$$

where the large normal variances in the last two lines give "vague" priors on $\log(\lambda)$ and $\text{logit}(p)$, reflecting a lack of prior information about these parameters.

Bayesian estimation usually employs Markov chain Monte Carlo simulation (MCMC), implemented by software such as BUGS (Gilks et al., 1994), JAGS (Plummer, 2013) or Nimble (de Valpine et al., 2017), to obtain a large sample from the joint posterior distribution of the parameters. Inference is based on this sample, with point estimates taken as the posterior means or medians and limits of $(1-\alpha)100\%$ interval estimates chosen so that $(1-\alpha)$ of the posterior sample is within these limits.

MCMC methods are computationally intensive. Integrated Nested Laplace Approximation (INLA) is a computationally efficient alternative which can be used for Bayesian models that may be formulated as latent Gaussian process models. Meehan et al. (2020) detailed how to fit an N-mixture model using the R package INLA (Lindgren & Rue, 2015). The current version of INLA allows covariates at the site and sampling occasion (e.g., year) level, but if there are multiple surveys per sampling occasion, any survey-level covariates must be averaged to the site or site-sampling occasion level. Meehan et al. (2020) compared results and computing time to analyses using unmarked and JAGS, and finds the results comparable and the computing times 10 and 500 times faster than unmarked and JAGS, respectively. The R package ubms (Kellner et al., 2022) is another option for Bayesian analysis which uses an unmarked-like formula interface but STAN (Carpenter et al., 2017) as the back-end computing engine.

Bayesian analysis has two advantages over ML estimation. First, in the Bayesian approach, it is easy to incorporate random effects into the covariate models of detection and expected abundance. In principle, this is feasible with ML estimation, but random and fixed effects must be handled separately. The R package unmarked (Fiske & Chandler, 2011) implements empirical Bayes but not ML estimation of random effects. The other advantage to the Bayesian approach is that it yields posterior distributions for the site-specific abundance parameters N_i, so estimates of these quantities are immediately available, whereas with the maximum likelihood approach, the estimated quantities are $E(N_i)$ not N_i.

2.4 | Alternative formulations

Motivated by concerns that the MLEs may depend on the choice of K used to approximate the infinite sums in (2), Dennis et al. (2015) gave an expression for the likelihood that does not include infinite sums. They showed simulation

results that demonstrate MLEs of expected abundance are sometimes infinite, particularly when the number of sampling occasions is small ($T \leq 3$).

Haines (2016) rewrote likelihood (2) in terms of generalized hypergeometric functions to yield an alternative closed-form expression for the likelihood that does not depend on infinite sums.

In their Appendix B, Meehan et al. (2020) employed recursive formulations of the binomial and Poisson probability mass functions to obtain an expression for $\text{Bin}(n_{i,t}; N_i, p)f(N_i; \boldsymbol{\theta})$ in likelihood (2) when the $N_i \sim$ Poisson. Their formulation requires approximating infinite sums with finite sums, but it is faster to compute and more numerically stable. They remarked that since the negative binomial mass function has a similar recursive expression, a similar computational improvement is available when the $N_i \sim$ Negative Binomial.

2.5 | Identifiability concerns

If population sizes $N_1,...,N_R$ are iid Poisson(λ) and there is only one sampling occasion ($T = 1$), then the observed counts are distributed as Poisson($p\lambda$), and it not possible to estimate p and λ separately–the two parameters are *not identifiable* (but see Sólymos et al. (2012)). When the number of sampling occasions is $T \geq 2$, the parameters are identifiable, but the likelihood may be too flat to optimize numerically. This is the behavior observed by Dennis et al. (2015) where, for $T = 2$ and $T = 3$, a large proportion of simulated data sets generated infinite MLEs of expected abundance. They speculated that including site-specific covariates for abundance and detection could yield a more well-defined likelihood surface. However, Knape and Korner Nievergelt (2015) showed that for single-visit designs, including covariates does not yield an identifiable model when both expected abundance and expected detection are modeled using log links. Researchers should be aware of potential non-identifiability, even when their models include covariates.

Kéry (2018) investigated the identifiability of Poisson, zero-inflated Poisson (ZIP), negative binomial, and multinomial (see Section 2.6 below) models and found few problems for the Poisson and ZIP models for real data from a bird monitoring program, provided the sample sizes were sufficient. However, negative binomial frequently exhibited unstable estimates.

Dennis et al. (2015) developed diagnostics to assess model identifiability. For Poisson N-mixture models, their diagnostic is the average pairwise sample covariance of the observed counts. A negative value for this diagnostic indicates an unstable MLE. For the negative binomial N-mixture model, their diagnostic is based on constraints that must be imposed to yield valid moment estimators for detection and for the expectation and variance of abundance.

Haines (2016) showed that when the $N_i \sim$ iid Poisson(λ), the product of the MLEs of λ and p is equal to $\overline{n}_c \equiv \sum_i \sum_t n_{it}/(RT)$. Replacing λ in the likelihood function with $p/\overline{n}_c$ yields a likelihood that has been "concentrated" in detection probability p. The Poisson N-mixture model diagnostic of Dennis et al. (2015) relates to the shape of this concentrated likelihood, provided observed counts satisfy $\sum_{i=1}^{R}\sum_{t=1}^{T}(n_{i,T} - n_{i,t}) > 0$. In this case, if the diagnostic is positive (indicating a stable MLE), the concentrated likelihood for p has at least one maximum in the interval $[0,1]$.

Barker et al. (2018) argued that the information lost by not marking individuals makes N-mixture model identifiability depend critically on the binomial assumption. In particular, they assert that since the Poisson model is the limit of the binomial when $N \to \infty$ and $p \to 0$ with product $Np = \lambda$ held constant, the Poisson should be a reasonable alternative model. However, this is exactly the situation in which N and p are not identifiable from point counts $n_{i,t}$. To accommodate both under- and over-dispersion, they advocate using a double Poisson model (Efron, 1986). A simulation study comparing this model to the basic N-mixture model with the same first two moments suggests that the models cannot be reliably distinguished by comparing AICs.

2.6 | Multinomial N-mixture models

Royle (2004a) and Dorazio et al. (2005) placed the binomial N-mixture model within a larger class of N-mixture models where observed counts $n_{i,t}$ are modeled as multinomial rather than binomial (see Kéry and Royle (2016), Ch. 7). They showed that this larger class could accommodate several common sampling protocols for estimating population size when detection of individuals is imperfect. These protocols include capture-recapture, multiple-observer sampling, and distance sampling (Royle et al., 2004) in addition to the simple point-count sampling which is our focus. Royle and Nichols (2003) considered a specialized observation model in which only binary presence/absence data are collected instead of counts of unique individuals.

The multinomial classes of N-mixture models have a number of advantages that make them broadly practical. Many common survey protocols that produce multinomial observations are in widespread use for surveying multiple taxa, and data collected from such efforts are naturally modeled using the multinomial N-mixture models. Furthermore, these models are less sensitive to the likelihood instability and identifiability issues described previously, because many of the multinomial observation model parameterizations provide direct information about detection probability. Kéry (2018) found that multinomial N-mixture models were generally well-behaved from an estimation standpoint, exhibiting few unstable cases.

Because there are few if any novel considerations in extending the binomial N-mixture model to multinomial observation models, we do not include further details here. The extensions that have been developed for the binomial N-mixture model generally apply directly to the multinomial models for example, see Chandler et al. (2011). Indeed, several of them have recently been implemented in the unmarked package (Kellner et al., 2023) including the extension to open populations described in the next section.

3 | RELAXING THE CLOSED-POPULATION ASSUMPTION

The closed-population assumption is unrealistic in many situations. When population size changes within the study period, parameter estimates may be unacceptably biased. Dail and Madsen (2011) augmented the basic N-mixture model to account for population dynamics allowing immigration and emigration, births and deaths.

Rather than assuming a constant abundance N_i at site i, the dynamic model allows distinct $N_{i,t}$, the true abundance at site i during sampling occasion t. Thus, the data model in (1) becomes

$$n_{i,t} \sim \text{Binomial}(N_{i,t}, p),$$

and instead of assuming N_i iid with density $f(N_i; \boldsymbol{\theta})$, we need a joint prior for $N_{i,1}, ..., N_{i,T}$. The dynamic model retains the assumption that the R sites are independent. Within a site, however, the $N_{i,t}$ will be dependent, and their joint distribution needs to reflect population dynamics.

To model population dynamics, Dail and Madsen (2011) defined random variables $S_{i,t}$ ("survivors") and $G_{i,t}$ ("gains") for site i and sampling occasions $t = 2, ..., T$, whose distributions, conditional on site abundance at the previous sampling occasion, are

$$S_{i,t} | N_{i,t-1} \sim \text{Binomial}(N_{i,t-1}, \phi) \tag{4}$$

$$G_{i,t} | N_{i,t-1} \sim \text{Poisson}\{\gamma(N_{i,t-1})\}.$$

Survivors $S_{i,t}$ are individuals that occupied site i at time $t-1$ and remained at time t. The binomial probability ϕ is the probability of survival. For populations that may be subject to movement of individuals in and out, the interpretation of ϕ is usually described as "apparent" survival which is the component of the population that remained and survived. $G_{i,t}$ models the number of individuals new to site i at time t (which may include immigrants, depending on the context). The Poisson mean $\gamma(N_{i,t-1})$ is a (possibly constant) function of abundance at time $t-1$.

The abundance $N_{i,t}$ at time t given the abundance $N_{i,t-1}$ at time $t-1$ is the sum of the survivors and gains:

$$N_{i,t} | N_{i,t-1} = (S_{i,t} + G_{i,t}) | N_{i,t-1}.$$

Assuming $S_{i,t}$ and $G_{i,t}$ are conditionally independent given $N_{i,t-1}$, the conditional distribution of $N_{i,t} | N_{i,t-1}$ can be expressed by calculating probabilities

$$\Pr(N_{i,t} = k | N_{i,t-1} = j) \equiv P_{j,k} = \sum_{c=0}^{\min(j,k)} \Pr(S_{i,t} = c | N_{i,t-1} = j) \Pr(G_{i,t} = k - c | N_{i,t-1} = j) \tag{5}$$

for $j, k = 1, ..., \infty$. Each term in the sum in (5) is the probability that there are c survivors and $k - c$ gains to site i between time $t-1$ and time t, given $N_{i,t-1} = j$. Summing over the possible values of c gives the total (conditional) probability $\Pr(N_{i,t} = k | N_{i,t-1} = j)$.

The distribution of $N_{i,t} \,|\, N_{i,1}, \ldots, N_{i,t-1}$ is the same as that of $N_{i,t} \,|\, N_{i,t-1}$, that is, $N_{i,1}, \ldots, N_{i,T}$ is a first-order *Markov chain*. The joint probability mass function (pmf) of $N_{i,1}, \ldots, N_{i,T}$ can be written as the product of the pmf of $N_{i,1}$ and all the conditional pmfs of $N_{i,t} \,|\, N_{i,t-1}$ for $t = 2, \ldots, T$. Assuming $N_{i,1} \sim \text{Poisson}(\lambda)$, the joint pmf of $N_{i,1}, \ldots, N_{i,T}$ is

$$\Pr(N_{i,1} = k_1, N_{i,2} = k_2, \ldots, N_{i,T} = k_T) = \frac{e^{-\lambda}\lambda^{k_1}}{k_1!} \prod_{t=2}^{T} P_{k_{t-1}, k_t}.$$

Now, the dynamic joint likelihood, analogous to (2), is

$$L(p, \lambda, \gamma, \phi) = \prod_{i=1}^{R} \left[\sum_{N_{i,1}=n_{i,1}}^{\infty} \cdots \sum_{N_{i,T}=n_{i,T}}^{\infty} \left\{ \left(\prod_{t=1}^{T} \text{Bin}(n_{i,t}; N_{i,t}, p) \right) \frac{e^{-\lambda}\lambda^{N_{i,1}}}{N_{i,1}!} \prod_{t=2}^{T} P_{N_{i,t-1}, N_{i,t}} \right\} \right]. \tag{6}$$

As with (2), the infinite sums in (6) must be approximated by summing to a finite value, chosen to be large relative to the observed counts. Dail and Madsen (2011) set $K = 200$ for their two data examples, where the maximum observed counts were $n_{i,t} = 12$ and $n_{i,t} = 5$. Note that increasing K requires more computing effort for each evaluation of (6).

Dail and Madsen (2011) listed three forms of mean recruitment $\gamma(N_{it-1})$: constant recruitment $G_{i,t} \,|\, N_{i,t-1} \sim \text{Poisson}(\gamma)$, "autoregressive" recruitment $G_{i,t} \,|\, N_{i,t-1} \sim \text{Poisson}(\gamma \cdot N_{i,t-1})$, and constant abundance $G_{i,t} \,|\, N_{i,t-1} \sim \text{Poisson}((1-\phi)\lambda)$. Hostetler and Chandler (2015) proposed model extensions to accommodate population growth including exponential $N_{it} \,|\, N_{i,t-1} \sim \text{Poisson}(\exp(r)N_{i,t-1})$ and Ricker $N_{i,t} \,|\, N_{i,t-1} \sim \text{Poisson}(N_{i,t-1} \exp\{r(1 - \log(N_{i,t-1}+1)/\log(K+1))\})$. These population growth models avoid the discrete convolution in (5). Bellier et al. (2016) augmented the model further to allow density-dependent survival as well as recruitment.

The basic N-mixture model with likelihood (2) is a special case of the dynamic model when $\gamma = 0$ and $\phi = 1$. The nesting of these two models allows a formal hypothesis test of population closure. Because $\gamma = 0$ and $\phi = 1$ are on the boundary of the parameter space, the usual asymptotic likelihood ratio test (Lehmann, 1998) is not available, and one may use the asymptotic test given in Self and Liang (1987).

The open population model may be fit to data using the pcountOpen function in the unmarked R package. Like the pcount function described in Section 2.2, pcountOpen allows covariate formulas for initial expected abundance and detection. The open-population model has additional parameters describing survival and recruitment which also may be modeled using covariates. The function implements the population dynamics enhancements of Hostetler and Chandler (2015), allowing the user to specify Ricker or Gompertz-logistic density-dependent population growth in addition to exponential, autoregressive, or constant recruitment dynamics.

Since estimation of N and p is problematic under the closed-population model when there are few sampling occasions, a natural question arises: does the dynamic model have similar estimability problems, and if so, under what circumstances?

The simulation study of Bellier et al. (2016) showed that when the data generation mechanism matches the model, parameters are generally well-estimated, and that accurate estimates of abundance and detection probability are obtained, even when the model is misspecified.

The results of Knape and Korner-Nievergelt (2015) suggest that when p varies over time, even if it's modeled using covariates, the parameters may not be identifiable without sampling under the "robust design," where the population is assumed closed within specific intervals, called *primary periods*, and sites are surveyed multiple times within a primary period. Zhao and Royle (2019) showed, also by way of a simulation study, that time-varying p is identifiable without the robust design if the parameter is correctly modeled as a random effect. However, for fixed-effect $p_t's$, estimates of the detection and dynamics parameters are biased. Their results showed that replicate surveys are necessary to fully identify the model if detection varies with time.

4 | THE ASYMPTOTIC APPROXIMATION

When some of the $n_{i,t}$ are larger than about 30, the approximation of likelihoods (2) or (6) by replacing infinite sums with sums to a large finite upper bound K becomes computationally intractable. Brintz et al. (2018) developed an asymptotic approximation to these likelihoods that avoids the infinite sums.

The approximation relies on the normal approximation to the binomial. In particular, when $N_{i,t}$ is sufficiently large relative to p, the marginal distribution of $n_{i,t} \sim \text{Bin}(N_{i,t},p)$ may be approximated by a normal distribution with mean $\mu_{i,t} = E(n_{i,t})$ and variance $\sigma_{i,t} = \text{var}(n_{it})$. Brintz et al. (2018) proposed approximating the joint distribution of $n_{i,1},...,n_{i,T}$ as multivariate normal (MVN) with mean vector $\boldsymbol{\mu_i} = [\mu_{i,1},...\mu_{i,T}]'$ and $T \times T$ variance–covariance matrix $\boldsymbol{\Sigma_i}$. Since sites $i = 1,...R$ are independent, we get the following approximate likelihood:

$$L(\lambda,p,\phi,\gamma|\{n_{i,t}\}) \approx \prod_{i-1}^{R} MVN(\boldsymbol{\mu_i},\boldsymbol{\Sigma_i}) = MVN(\boldsymbol{\mu},\boldsymbol{\Sigma}), \tag{7}$$

where $\boldsymbol{\mu} = [\boldsymbol{\mu_i'},...\boldsymbol{\mu_R'}]'$ and $\boldsymbol{\Sigma}$ is an $RT \times RT$ block-diagonal matrix with diagonals $\boldsymbol{\Sigma_i}$.

The elements of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the expressions in λ, p, ϕ, and γ induced by the dynamic model described in Section 3, and derived by the laws of total expectation, variance, and covariance.

4.1 | Example

As a simple example, suppose initial site abundances are distributed as Poisson with mean λ and gains at time t are Poisson with constant mean γ. The complete model for site i is

$$n_{i,t} \mid N_{i,t} \sim \text{Bin}(N_{i,t},p)$$

$$N_{i,1} \sim \text{Poisson}(\lambda)$$

$$S_{i,t} \mid N_{i,t-1} \sim \text{Bin}(N_{i,t-1},\phi)$$

$$G_{i,t} \mid N_{i,t-1} \sim \text{Poisson}(\gamma)$$

$$N_{i,t} \mid N_{i,t-1} = (S_{i,t} + G_{i,t}) \mid N_{i,t-1},$$

where $S_{i,t}$ and $G_{i,t}$ are conditionally independent given $N_{i,t-1}$. Under this model, the $N_{i,t}$ are marginally Poisson, since $S_{i,t} \sim \text{Poisson}[\phi E(N_{i,t-1})]$ and $N_{i,t} = S_{i,t} + G_{i,t}$ is the sum of independent Poisson random variables. Therefore, $n_{i,t}$ is also marginally Poisson with mean $pE(N_{i,t})$.

When $t = 1$,

$$\mu_{i1} = E[E(n_{i,1}|N_{i,1})]$$

$$= \lambda p,$$

and for $t > 1$, $\mu_{i,t}$ can be defined inductively as

$$\mu_{i,t} = E[E(n_{i,t}|N_{i,t},N_{i,t-1},...,N_{i,1})]$$

$$= E[E(pN_{i,t}|N_{i,t-1},...,N_{i,1})]$$

$$= E[E(p(S_{i,t}+G_{i,t})|N_{i,t-1},...,N_{i,1})]\}$$

$$= p(\mu_{i,t-1}\phi + \gamma).$$

Since $n_{i,t} \sim \text{Poisson}$, $\sigma_{i,t}^2 = \mu_{i,t}$.

Assuming $n_{i,t}$ and n_{i,t^*}, for $t < t^*$, are conditionally independent given $\{N_{i,t^*-1}\} \equiv N_{i,1}, \ldots N_{i,t^*-1}$ (that is, all correlation among the $n_{i,t}$ is entirely due to correlation among the $N_{i,t}$), an off-diagonal element of S is

$$\mathrm{cov}(n_{i,t}, n_{i,t^*}) = E[\mathrm{cov}(n_{i,t}|n_{i,t^*}, \{N_{i,t^*-1}\})] + \mathrm{cov}[E(n_{i,t} \mid \{N_{i,t^*-1}\}, E(n_{i,t^*}|\{N_{i,t^*-1}\})]$$

$$= p^2 \mathrm{cov}(N_{i,t}, N_{i,t^*}).$$

An inductive derivation shows that

$$\mathrm{cov}(N_{i,t}, N_{i,t^*}) = \frac{\phi^{t^*-t}\mu_{i,t}}{p},$$

so

$$\sigma_{t,t^*} \equiv \mathrm{cov}(n_{i,t}, n_{i,t^*}) = p\phi^{t^*-t}\mu_{i,t}.$$

4.2 | Estimability diagnostic

With the normal approximation, it is possible to diagnose parameter configurations leading to parameter confounding by examining the Fisher information matrix, since this matrix has a closed-form expression. The inverse Fisher information matrix is the asymptotic variance–covariance matrix for the vector of MLEs, so a lack of invertibility of this matrix reveals parameter confounding.

Let $\boldsymbol{\theta}$ denote the vector of model parameters. The element in the jth row and kth column of the Fisher information matrix is

$$I_{j,k} = \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_j} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_k} + \frac{1}{2}\mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\right).$$

Brintz et al. (2018) suggested the determinant of the Fisher information matrix as a measure of parameter identifiability. When the determinant is close to 0, the matrix is nearly singular. Note that the sample version of this quantity is unavailable from data unless optimization of the approximate likelihood (7) is successful.

5 | MULTI-STATE EXTENSIONS

An important extension of the dynamic model described in Section 3 applies to stratified populations (Zipkin, Sillett, et al., 2014; Zipkin, Thorson, et al., 2014) in which every individual belongs to one of a number of mutually exclusive groups or sub-populations. Typical examples are age, sex, geographic location, or disease status. In this case, the population size parameter is multi-variate, having values indexed by the categorical group membership. For example, in a common age-structured system, we might partition the population into juveniles and adults, and then $\mathbf{N}_{i,t} = (N_{juv,i,t}, N_{ad,i,t})$. In class-structured populations, counts can be made for each class, or possibly for combinations of classes. Thus, while the state variable $\mathbf{N}$ is a vector (one for each class), so too is the observed vector of counts $\mathbf{y}$ for each site and sample occasion. Such multistate dynamic models are analogous to the classical multi-state capture-recapture models (Arnason, 1972, 1973; Hestbeck et al., 1991; Schwarz et al., 1993) except that the latter are models for individual-level encounter history data, whereas the former are models for observed class frequencies subject to binomial sampling.

To illustrate a multi-state dynamic model we consider a two-class structured population having class-specific survival probabilities, and allowing that individuals may transition between the two states. Let $S_{c,i,t}$ be the number of survivors of class c, for survey site i and year t, and let $T_{c,c',i,t}$ be the number of individuals that transition from class c to class c' in the interval from $t-1$ to t. Then, the 2-state model structure has the form

$$S_{1,i,t} \mid N_{1,i,t-1} \sim \text{Binomial}(N_{1,i,t-1}, \phi_1)$$

$$S_{2,i,t} \mid N_{2,i,t-1} \sim \text{Binomial}(N_{2,i,t-1}, \phi_2)$$

$$T_{1,2,i,t} \mid S_{1,i,t} \sim \text{Binomial}(S_{1,i,t}, \theta_{1,2})$$

$$T_{2,1,i,t} \mid S_{2,i,t} \sim \text{Binomial}(S_{2,i,t}, \theta_{2,1}), \tag{8}$$

where ϕ_c is the survival probability in class c and $\theta_{c,c'}$ is the transition probability from class c to class c'.

Thus, the number of surviving individuals is expressed conditional on the population size of the previous (primary) occasion, $t-1$, and the number of individuals that transition from one state to the other is conditional on the number of surviving individuals of the origin state. In addition to the survival and state-transition model components above, we also may allow for recruitment, for example using models such as (8) so that $R_{c,i,t}$ is the number of recruits entering class c between $t-1$ and t.

As before, the population size of class c in year t is derived as the sum of survivors, new recruits, and individuals that transition into that class from every other class, less individuals transitioning out of that class. For the 2-state model there are thus two population growth equations:

$$N_{1,i,t} = R_{1,i,t} + S_{1,i,t} - T_{1,2,i,t} + T_{2,1,i,t}$$

and

$$N_{2,i,t} = R_{2,i,t} + S_{2,i,t} + T_{1,2,i,t} - T_{2,1,i,t}.$$

DiRenzo et al. (2018) applied the multi-state dynamic model to studies of the fungal pathogen, *Batrachochytrium dendrobatidis* (Bd) in tropical amphibian communities in Panama. In their study, sample units were defined by 20 m stream and trail segments. Observers counted frogs and recorded whether they were infected with the Bd fungus (class 1) or not (class 2). Individuals transition between the two states: infected frogs may become uninfected and vice versa. In addition, it seems reasonable to expect that survival and even recruitment might differ between the infected and uninfected classes. All these features are readily accommodated in the multi-state dynamic framework. This example is presented in Kéry and Royle (see 2021, Ch. 2) who provide an R code template for simulating data and fitting models in JAGS.

Size-structured multi-state data represent another canonical example that is well-suited for the multi-state dynamic model. Zipkin, Thorson, et al. (2014) analyzed data from surveys of the Northern dusky salamander (*Desmognathus fuscus*) in two National Park Service units (Chesapeake and Ohio National Historic Park in Maryland, USA, and Rock Creek Park, D.C., USA). Here again the sample units were stream segments which were searched by observers who recorded all individuals found and measured their snout-to-vent length (SVL). The model of Zipkin, Thorson, et al. (2014) is a size-structured model with three stages, first year juveniles, second-year juveniles, and adults. Note that only two classes are observable: total number of juveniles (individuals $\leq 35\,\text{mm}$ SVL) and adults, but the model is described in terms of two classes of juveniles in order to account for reproductive lag (individuals do not start reproducing until their third year of life). In this model, survival probability is assumed to vary between juveniles and adults. Thus, define ϕ_1 and ϕ_2 as the apparent annual survival for juveniles and adults, respectively. The 3-component survival model is

$$S_{1,i,t} \mid N_{1,i,t-1} \sim \text{Binomial}(N_{1,i,t-1}, \phi_1)$$

$$S_{2,i,t} \mid N_{2,i,t-1} \sim \text{Binomial}(N_{2,i,t-1}, \phi_1)$$

$$S_{3,i,t} \mid N_{3,i,t-1} \sim \text{Binomial}(N_{3,i,t-1}, \phi_2).$$

Unlike the model of DiRenzo et al. (2018) for Bd infected frogs, there are no stage transition probabilities to estimate here because the stage transitions are deterministic.

6 | SPATIAL MODELS

The models in Sections 2 and 3 assume independence of observations on different sites, as well as independence of the latent abundance variables N_i or $N_{i,t}$. For some applications, this is unreasonable, and modeling spatial dependence among sites is desirable. The nature of the spatial dependence will depend on the particular application.

6.1 | A spatial model for wildlife populations

Zhao et al. (2017) observed that the dynamic model of Section 3 does not distinguish immigration from reproduction nor emigration from a lack of survival. They argued that when "sites" are habitat patches (i.e., consistent with a classic metapopulation; Hanski, 1999), the movement of animals among patches is an important component of the system, and that movement between patches is more likely when the patches are adjacent. They augmented the dynamic model by modeling true survival, reproduction, immigration, and emigration.

As with the dynamic model, they modeled the true population size at time $t = 1$ as $N_{i,1} \sim \text{Poisson}(\lambda)$ and true survival for site i between sampling occasions $t-1$ and t as

$$S_{i,t} \mid N_{i,t-1} \sim \text{Binomial}(N_{i,t-1}, \phi).$$

This is identical to the definition of the survival random variable in (4), but Zhao et al. (2017) distinguished emigrants from non-survivors by defining emigrants from site i between $t-1$ and t as

$$E_{i,t} \mid S_{i,t} \sim \text{Binomial}(S_{i,t}, \kappa),$$

where κ is the rate of emigration.

They defined immigrants to patch i between $t-1$ and t as

$$I_{i,t} \mid \{E_{1,t}, ..., E_{R,t}\} \sim \text{Poisson}\left(\sum_{j=1}^{R} w_{i,j} E_{j,t}\right),$$

where the $w_{i,j}$ are weights which describe the spatial pattern of movement. Zhao et al. (2017) modeled migration between adjacent patches by specifying $w_{i,j} = 0$ unless patches i and j are adjacent, in which case $w_{i,j} = 1/n_i^{adj}$ where n_i^{adj} denotes the number of patches j adjacent to patch i.

Finally, they modeled reproduction in patch i between $t-1$ and t by defining

$$R_{i,t} \mid N_{i,t-1} \sim \text{Poisson}(\gamma N_{i,t-1}),$$

where γ is the rate of reproduction.

The population size in patch i at time t is the sum of the true survivors, immigrants, and births, less the number of emigrants:

$$N_{i,t} = S_{i,t} + I_{i,t} + R_{i,t} - E_{i,t},$$

and imperfectly-detected observed counts are

$$n_{i,t} \sim \text{Binomial}(N_{i,t}, p).$$

Since wildlife populations at the scale of typical sample units may be small, the asymptotic approximation is inappropriate, whereas the conditional definitions of $S_{i,t}$, $I_{i,t}$, $R_{i,t}$, and $E_{i,t}$ make Bayesian estimation a natural choice (see Section 2.3). Zhao et al. (2017) employed this estimation approach, and their appendix includes JAGS code to fit the model.

Howell et al. (2020) devised a similar spatially-explicit model allowing distance-based movement of frogs among ponds in a metapopulation system, but with a simpler model of net population growth without distinguishing survival from recruitment. Zhao et al. (2022) developed a multi-species extension of this "net growth" model.

6.2 | A spatial model for disease surveillance

Brintz et al. (2021) adapted the asymptotic approximation to model the number of cases $N_{i,t}$ of a communicable disease in site i at time t, where observed counts $n_{i,t}$ may be much too large to approximate (6) using finite sums. The spatial aspect of infection requires a spatial model, but the spread of disease cases is different from animal movement, so the model of Section 6.1 is not appropriate.

In the context of disease surveillance, asymptomatic cases or a lack of thorough testing may result in imperfect detection.

Instead of modeling reproduction and migration, Brintz et al. (2021) redefined and reinterpreted the "survivor" and "gain" random variables (4) of the dynamic model.

The number of surviving disease cases at site i and time t has the familiar definition

$$S_{i,t} \mid N_{i,t-1} \sim \text{Binomial}(N_{i,t-1}, \phi), \tag{9}$$

but here $S_{i,t}$ represents those infected at both time $t-1$ and time t.

The distribution of the "gain" random variable is defined by conditioning on the $N_{i,t}$ at all sites i and all previous sampling occasions, so define notation $\{N_{\alpha,t_0}\}$ to be the set of all $N_{i,t}$ for $i = 1,...,R$ and all $t = 1...,t_0$. Then the number of new cases at site i and time t is

$$G_{i,t} \mid \{N_{\alpha,t-1}\} \sim \text{Poisson}\left(\gamma \widetilde{N}_{i,t-1}\right), \tag{10}$$

where

$$\widetilde{N}_{i,t-1} = \text{pop}_{i,t-1} \sum_{r=1}^{R} \frac{A_{i,r} N_{r,t-1}/\text{pop}_{r,t-1}}{A_{i\cdot}}. \tag{11}$$

In (11), $\text{pop}_{i,t-1}$ denotes the population of site i at time $t-1$, and matrix $\boldsymbol{A}$ is a neighborhood matrix comprised of entries $A_{i,j} = 1$ if sites i and j are neighbors and 0 otherwise. $A_{i\cdot}$ denotes the sum of the ith row of $\boldsymbol{A}$, the size of site i's neighborhood. The quantity $N_{i,t}/\text{pop}_{i,t}$ is the proportion of the population infected, that is, disease *prevalence*, so $G_{i,t} \mid \{N_{\alpha,t-1}\}$ in (10) models the new cases at site i and time t conditional on the average prevalence of the disease in the neighborhood of site i at time $t-1$.

Initial disease counts are

$$N_{i,1} \sim \text{Poisson}\left(\beta \text{pop}_{i,1}\right), \tag{12}$$

where β represents the initial prevalence of the disease. This definition neglects any spatial dependence among the $N_{i,1}$.

Equations (9), (10), and (12) imply expressions for $E_{(n,t)}$, $\mathrm{var}(n_{i,t})$, and $\mathrm{cov}(n_{i,t}, n_{i',t'})$ in terms of model parameters p, β, γ, and ϕ. Brintz et al. (2021) employed an asymptotic approximation similar to (7), but because of the spatial dependence, variance–covariance matrix $\boldsymbol{\Sigma}$ is no longer block-diagonal. The asymptotic approximation is appropriate when observed case counts $n_{i,t}$ are moderate or large, as they will be when the disease of interest is not exceptionally rare.

Brintz et al. (2021) maximized the approximate likelihood to obtain MLEs $\hat{p}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\phi}$. Instead of estimating $E(N_{i,t})$, they employed a parametric bootstrap to simulate the $N_{i,t}$ according to the model described in (9), (10), and (12), with MLEs serving as parameter values. The simulated distributions of the $N_{i,t}$ yield point and interval estimates in the same way that the simulated posterior distributions of the $N_{i,t}$ are used in Bayesian inference.

7 | CONCLUSION

Since their introduction in 2004, binomial N-mixture models have been extended and adapted to accommodate diverse modeling challenges including open populations, multi-state population structure, and populations with spatial dynamics. N-mixture models are appropriate for estimating the size of closed populations or modeling variation in population size when detection of individuals in the population is imperfect and capture-recapture sampling is infeasible, thus limiting available data to counts of unmarked individuals as opposed to more informative encounter history data that capture-recapture produces. Because N-mixture models rely on replicated counts to model both population size and detection probability, parameters may not be readily identifiable and can be sensitive to parametric model assumptions. However, the models have been used widely and successfully and new adaptations and extensions continue to be developed.

An important reason for their widespread adoption is the core idea underlying the N-mixture model, that is, their formulation as a hierarchical model that integrates distinct observation and ecological components, allows investigators to test explicit hypotheses about mechanisms that influence detectability and abundance separately (Kéry, 2008). Moreover, the hierarchical formulation applies to a large number of existing sampling protocols beyond the binomial counting model such as "removal" sampling, distance sampling, and surveys based on multiple observer systems. Finally, N-mixture models improve our ability to study complex ecological systems such as disease in wild populations, predator–prey systems, and invasive species, where often explicit mechanisms are understood to operate on true abundance (e.g., density dependence), and inference about those mechanisms may be compromised in the presence of unaccounted-for imperfect detection. Thus, being able to address such hypotheses from observation data without the need to uniquely mark individuals increases their scope and relevance greatly.

The future of N-mixture models includes expanded application and technical developments focused on increasing complexity of model components, such as spatial and temporal dynamics as well as increasing ecological sophistication, such as their application to multiple species systems [e.g., Zhao et al. (2022)] to model predator–prey dynamics or competition. We also envision increased application of N-mixture models to problems of "data integration," in which observations obtained by multiple monitoring programs can be analyzed jointly using individual component models having shared parameters (Schaub & Kery, 2021).

AUTHOR CONTRIBUTIONS
Lisa Madsen: Conceptualization (supporting); methodology (supporting); writing – original draft (lead); writing – review and editing (equal). **J. Andrew Royle:** Conceptualization (lead); methodology (lead); writing – original draft (supporting); writing – review and editing (equal).

CONFLICT OF INTEREST STATEMENT
The authors have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Lisa Madsen https://orcid.org/0000-0001-5501-6163

RELATED WIREs ARTICLE

A survey of software for fitting capture-recapture models

REFERENCES

Arnason, N. (1972). Parameter estimates from mark-recapture experiments on two populations subject to migration and death. *Researches on Population Ecology*, *13*(2), 97–113.

Arnason, N. (1973). The estimation of population size, migration rates and survival in a stratified population. *Researches on Population Ecology*, *15*(2), 1–8.

Barker, R. J., Schofield, M. R., Link, W. A., & Sauer, J. R. (2018). On the reliability of N-mixture models for count data. *Biometrics*, *74*(1), 369–377.

Bellier, E., Kéry, M., & Schaub, M. (2016). Simulation-based assessment of dynamic n-mixture models in the presence of density dependence and environmental stochasticity. *Methods in Ecology and Evolution*, *7*(9), 1029–1040.

Brintz, B., Fuentes, C., & Madsen, L. (2018). An asymptotic approximation to the N-mixture model for the estimation of disease prevalence. *Biometrics*, *74*(4), 1512–1518.

Brintz, B. J., Madsen, L., & Fuentes, C. (2021). A spatially explicit N-mixture model for the estimation of disease prevalence. *Statistical Modelling, 23*(1), 31–52. page 1471082X211020872.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–29.

Carroll, R. J., & Lombard, F. (1985). A note on N estimators for the binomial distribution. *Journal of the American Statistical Association*, *80*(390), 423–426.

Chandler, R. B., Royle, J. A., & King, D. I. (2011). Inference about density and temporary emigration in unmarked populations. *Ecology*, *92*(7), 1429–1435.

Dail, D., & Madsen, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*, *67*(2), 577–587.

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with nimble. *Journal of Computational and Graphical Statistics*, *26*(2), 403–413.

Della Rocca, F., Milanesi, P., Magna, F., Mola, L., Bezzicheri, T., Deiaco, C., & Bracco, F. (2020). Comparison of two sampling methods to estimate the abundance of Lucanus cervus with application of n-mixture models. *Forests*, *11*(10), 1085.

Dennis, E. B., Morgan, B. J., & Ridout, M. S. (2015). Computational aspects of N-mixture models. *Biometrics*, *71*(1), 237–246.

DiRenzo, G. V., Campbell Grant, E. H., Longo, A. V., Che-Castaldo, C., Zamudio, K. R., & Lips, K. R. (2018). Imperfect pathogen detection from non-invasive skin swabs biases disease inference. *Methods in Ecology and Evolution*, *9*(2), 380–389.

DiRenzo, G. V., Che-Castaldo, C., Saunders, S. P., Grant, E. H. C., & Zipkin, E. F. (2019). Disease-structured N-mixture models: A practical guide to model disease dynamics using count data. *Ecology and Evolution*, *9*(2), 899–909.

Dodd, C. K., Jr., & Dorazio, R. M. (2004). Using counts to simultaneously estimate abundance and detection probabilities in a salamander community. *Herpetologica*, *60*(4), 468–478.

Dorazio, R. M., Jelks, H. L., & Jordan, F. (2005). Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations. *Biometrics*, *61*(4), 1093–1101.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, *81*(395), 709–721.

Ficetola, G. F., Barzaghi, B., Melotto, A., Muraro, M., Lunghi, E., Canedoli, C., Parrino, E. L., Nanni, V., Silva-Rocha, I., Urso, A., Carretero, M. A., Salvi, D., Scali, S., Scarì, G., Pennati, R., Andreone, F., & Manenti, R. (2018). N-mixture models reliably estimate the abundance of small vertebrates. *Scientific Reports*, *8*(1), 1–8.

Fiske, I., & Chandler, R. (2011). Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, *43*(10), 1–23.

Gilks, W. R., Thomas, A., & Spiegelhalter, D. J. (1994). A language and program for complex Bayesian modelling. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *43*(1), 169–177.

Haines, L. M. (2016). Maximum likelihood estimation for N-mixture models. *Biometrics*, *72*(4), 1235–1245.

Hanski, I. (1999). *Metapopulation ecology*. Oxford University Press.

Hestbeck, J. B., Nichols, J. D., & Malecki, R. A. (1991). Estimates of movement and site fidelity using mark-resight data of wintering Canada geese. *Ecology*, *72*(2), 523–533.

Hostetler, J. A., & Chandler, R. B. (2015). Improved state-space models for inference about spatial and temporal variation in abundance from count data. *Ecology*, *96*(6), 1713–1723.

Howell, P. E., Hossack, B. R., Muths, E., Sigafus, B. H., & Chandler, R. B. (2020). Informing amphibian conservation efforts with abundance-based metapopulation models. *Herpetologica*, *76*(2), 240–250.

Keever, A., McGowan, C., Ditchkoff, S., Acker, P., Grand, J., & Newbolt, C. (2017). Efficacy of N-mixture models for surveying and monitoring white-tailed deer populations. *Mammal Research.*, *62*, 413–422.

Kellner, K. F., Fowler, N. L., Petroelje, T. R., Kautz, T. M., Beyer, D. E., Jr., & Belant, J. L. (2022). Ubms: An r package for fitting hierarchical occupancy and n-mixture abundance models in a bayesian framework. *Methods in Ecology and Evolution*, *13*(3), 577–584.

Kellner, K. F., Smith, A. D., Royle, J. A., Kery, M., Belant, J. L., and Chandler, R. B. (2023). The unmarked r package: Twelve years of advances in occurrence and abundance modeling in ecology. *Methods in Ecology and Evolution.* https://doi.org/10.1111/2041-210X.14123

Kéry, M. (2008). Estimating abundance from bird counts: Binomial mixture models uncover complex covariate relationships. *The Auk*, *125*(2), 336–345.

Kéry, M. (2018). Identifiability in N-mixture models: A large-scale screening test with bird data. *Ecology*, *99*(2), 281–288.

Kéry, M., & Royle, J. (2016). *Applied hierarchical modelling in ecology—Modeling distribution, abundance and species richness using R and BUGS*. Elsevier/Academic Press.

Kéry, M., & Royle, J. A. (2021). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models*. Academic Press.

Knape, J., & Korner-Nievergelt, F. (2015). Estimates from non-replicated population surveys rely on critical assumptions. *Methods in Ecology and Evolution*, *6*(3), 298–306.

Lehmann, E. (1998). *Elements of large-sample theory, chapter 7*. Springer.

Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, *63*(19), 1–25.

McKenny, H. C., Keeton, W. S., & Donovan, T. M. (2006). Effects of structural complexity enhancement on eastern red-backed salamander (plethodon cinereus) populations in northern hardwood forests. *Forest Ecology and Management*, *230*(1–3), 186–196.

Meehan, T. D., Michel, N. L., & Rue, H. (2020). Estimating animal abundance with N mixture models using the R-INLA package for R. *Journal of Statistical Software*, *95*(1), 1–26.

Nichols, J. D., Thomas, L., & Conn, P. B. (2009). Inferences about landbird abundance from count data: Recent advances and future directions. In *Modeling Demographic Processes in Marked Populations*, (pp. 201–235). Springer.

Olkin, I., Petkau, A. J., & Zidek, J. V. (1981). A comparison of n estimators for the binomial distribution. *Journal of the American Statistical Association*, *76*(375), 637–642.

Plummer, M. (2013). JAGS: Just another Gibbs sampler. http://mcmc-jags.sourceforge.net.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Royle, J. A. (2004a). Generalized estimators of avian abundance from count survey data. *Animal Biodiversity and Conservation*, *27*(1), 375–386.

Royle, J. A. (2004b). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, *60*(1), 108–115.

Royle, J. A., Dawson, D. K., & Bates, S. (2004). Modeling abundance effects in distance sampling. *Ecology*, *85*(6), 1591–1597.

Royle, J. A., & Dorazio, R. M. (2008). Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities. Elsevier.

Royle, J. A., & Nichols, J. D. (2003). Estimating abundance from repeated presence–absence data or point counts. *Ecology*, *84*(3), 777–790.

Schaub, M., & Kery, M. (2021). *Integrated population models: Theory and ecological applications with R and JAGS*. Academic Press.

Schoenecker, K. A., King, S. R., Ekernas, L. S., & Oyler-McCance, S. J. (2021). Using fecal DNA and closed-capture models to estimate feral horse population size. *The Journal of Wildlife Management.*, *85*, 1150–1161.

Schwarz, C. J., Schweigert, J. F., & Arnason, A. N. (1993). Estimating migration rates using tag-recovery data. *Biometrics*, *49*, 177–193.

Self, S. G., & Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*(398), 605–610.

Sólymos, P., Lele, S., & Bayne, E. (2012). Conditional likelihood approach for analyzing single visit abundance survey data in the presence of zero inflation and detection error. *Environmetrics*, *23*(2), 197–205.

Zhao, Q., Fuller, A. K., & Royle, J. A. (2022). Spatial dynamic n-mixture models with interspecific interactions. *Methods in Ecology and Evolution*, *13*(10), 2209–2221.

Zhao, Q., & Royle, J. A. (2019). Dynamic n-mixture models with temporal variability in detection probability. *Ecological Modelling*, *393*, 20–24.

Zhao, Q., Royle, J. A., & Boomer, G. S. (2017). Spatially explicit dynamic N-mixture models. *Population Ecology*, *59*(4), 293–300.

Zipkin, E. F., Sillett, T. S., Grant, E. H. C., Chandler, R. B., & Royle, J. A. (2014). Inferences about population dynamics from count data using multistate models: A comparison to capture–recapture approaches. *Ecology and Evolution*, *4*(4), 417–426.

Zipkin, E. F., Thorson, J. T., See, K., Lynch, H. J., Grant, E. H. C., Kanno, Y., Chandler, R. B., Letcher, B. H., & Royle, J. A. (2014). Modeling structured population dynamics using data from unmarked individuals. *Ecology*, *95*(1), 22–29.