# Faster asymptotic solutions for N-mixtures on large populations

M. R. P. Parker[1,*], J. Cao[1,†], L. L. E. Cowen[2,†], L. T. Elliott[1,†]

[1]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia, Canada

[2]Department of Mathematics and Statistics, University of Victoria, Victoria, British Columbia, Canada

† Authors listed in alphabetical order.

*Corresponding author email: mrparker909@gmail.com

**Abstract**

We derive an asymptotic likelihood function for open-population $N$-mixture models and show that it has favourable computational complexity and accuracy when compared to the traditional likelihood function for large population sizes. We validate our asymptotic model with simulation studies, and apply our model to estimate the population size of Ancient Murrelet chicks, comparing against results obtained using the traditional $N$-mixture likelihood and an alternative asymptotic model based on the multivariate normal distribution. For the Ancient Murrelet case study, our asymptotic model computes twice as fast as the traditional models, eleven times faster when parallel processing is used, and provides higher precision estimates than the asymptotic multivariate normal model. We provide an open source implementation of our methods in the *quickNmix* R package.

Keywords: asymptotic approximation; N-mixture models; population abundance estimation; Synthliboramphus antiquus; unmarked.

# 1 Introduction

$N$-mixtures are likelihood based models which estimate population size using observed counts of unmarked individuals over several sampling locations and sampling occasions. $N$-mixture models were originally proposed as closed population models in which the population size is assumed to be constant over the sampling occasions (Royle, 2004). Later, the models were extended to include open population modelling (often referred to as dynamic $N$-mixtures), in which the population size may change over the sampling occasions (Dail and Madsen, 2011). A recent overview of $N$-mixture models is given in Madsen and Royle (2023). $N$-mixture models are used frequently in ecological studies where under-counting is expected (Belant et al., 2016; Veech and Cave, 2021). These models have recently been extended to study transient populations (Kwon et al., 2018), the use of auxiliary populations (Parker et al., 2020), and for applications to wildlife disease analytics (DiRenzo et al., 2019). An asymptotic $N$-mixtures model has been previously developed using a multivariate normal approximation (Brintz et al., 2018). Traditionally $N$-mixture models are used for estimating population abundances when the observed counts are small and the expected true population size is also small. There are two reasons for this traditional restriction: 1) Numerical accuracy must be dealt with when large populations are considered, 2) Computation times become intractable for large populations. The first issue can be dealt with using well known high precision computational techniques, such as the use of scaling in hidden Markov models (Zucchini and MacDonald, 2009, p. 48), or the use of the numerically precise log-sum-exp technique (Parker et al., 2023). We develop an asymptotic likelihood function based on the original likelihood model (rather than on the multivariate normal approximation used in Brintz et al. 2018) to solve the second issue.

Our paper has five major contributions: 1) We develop an asymptotic likelihood function for open-population N-mixture models, 2) We show that the asymptotic model computes

substantially faster than the traditional models, 3) We show through simulation studies that the asymptotic model provides nearly identical accuracy and precision compared to the traditional models, 4) We apply several competing methods to a moderately sized population of Ancient Murrelet chicks, and compare results as well as computational efficiency of the methods, 5) We provide an open source R package, *quickNmix* which is available on CRAN, to facilitate future use of the asymptotic model.

The remainder of this manuscript is laid out as follows. In Section 2.1, we give a brief overview of traditional $N$-mixture models. Next we introduce our asymptotic likelihood in Section 2.2. We discuss the approximation error associated with using the asymptotic likelihood in Section 2.3. Model extensions are discussed in Section 2.4, and a method for estimating standard errors is overviewed in Section 2.5. In Section 2.6 we overview some common problems which $N$-mixture models are known to exhibit. We discuss the R package *quickNmix* in Section 2.7, which we have made available for application to other populations of interest. Traditional $N$-mixture models are implemented via the function *pcountOpen* in the R package *unmarked* (Fiske and Chandler, 2011), and we compare our asymptotic model against the *unmarked* traditional implementation using simulation studies in Sections 3.1 and 3.2, and using a case study of Ancient Murrelet (*Synthliboramphus antiquus*) chicks in Section 4. We validate the asymptotic model for large abundance using simulation studies in Section 3.3.

# 2 Methods

## 2.1 N-mixture Models Overview

The dynamic $N$-mixture model considers $U$ independent sampling sites, from which samples are observed on $M$ discrete sampling occasions. Then, the population size at site $i$ and time $t$ is $N_{it}$, where $i \in \{1, 2, ..., U\}$ and $t \in \{1, 2, ..., M\}$. Initial population abundance at

sampling occasion 1, $N_{i1}$, is assumed to be a random variable with mean $\lambda$. Usually the Poisson distribution is assumed so that $N_{i1} \sim \text{Pois}(\lambda)$. The observed counts $n_{it}$ are assumed to be under-counted observations of $N_{it}$, with probability of detection $p$, and assuming the binomial distribution, $n_{it} \sim \text{Binom}(N_{it}, p)$. Population dynamics for $t > 1$ are modelled by summing those recruited into the population with those that have survived to the next time period, $N_{it} = G_{it-1} + S_{it-1}$, with parameters $\gamma$ and $\omega$ (respectively the recruitment rate and the survival probability). Under Poisson and binomial assumptions, the recruitment $G_{it}$ is $G_{it} \sim \text{Pois}(\gamma)$, while the survival $S_{it}$ is $S_{it} \sim \text{Binom}(N_{it}, \omega)$. Thus, the model has four estimable parameters: $\lambda$, $\gamma$, $\omega$, and $p$. The population sizes $N_{it}$ are treated as confounding variables, and are integrated from the likelihood function by summing over possible values of $N_{it}$ up to a sufficiently large upper bound $K$.

The original dynamic $N$-mixture likelihood for open populations (Dail and Madsen, 2011) is shown in (1), with the transition probability function $P_{a,b}$ shown in (2). We denote the likelihood function by $\mathscr{L} = \Pr(\{n_{it}\}|\lambda, \gamma, \omega, p)$:

$$\mathscr{L} = \prod_{i=1}^{U} \left[ \sum_{N_{i1}=n_{i1}}^{K} \cdots \sum_{N_{iM}=n_{iM}}^{K} \left\{ \left( \prod_{t=1}^{M} \text{Binom}(n_{it}; N_{it}, p) \right) \text{Pois}(N_{i1}; \lambda) \prod_{t=2}^{M} P_{N_{it-1}, N_{it}} \right\} \right] \quad (1)$$

$$P_{a,b} = \sum_{c=0}^{m=\min\{a,b\}} \text{Binom}(c; a, \omega) \text{Pois}(b - c; \gamma) \quad (2)$$

## 2.2 Asymptotic Likelihood

We derive an asymptotic likelihood from the original likelihood in equation (1) by considering the asymptotic distributions of each component distribution in the transition probability function (2). Both the binomial distribution and the Poisson distribution have limiting distributions which are normal. For the binomial distribution: $\text{Binom}(N, p) \xrightarrow{d} \text{N}(Np, Np(1-$

$p$)) as $Np$ and $N(1-p)$ become large. For the Poisson distribution: $\text{Pois}(\lambda) \xrightarrow{d} \text{N}(\lambda, \lambda)$ as $\lambda$ becomes large. Let $g(x; \mu, \sigma^2)$ be the probability density function of a normal random variable with mean $\mu$ and variance $\sigma^2$. The product of two normal densities is proportional to another normal density as shown in (3) (see the Supplemental Material for a derivation, see the Supplemental Material from Vinga and Almeida (2004) for a multidimensional derivation).

$$g(c; a\omega, a\omega(1-\omega)) \cdot g(b-c; \gamma, \gamma) = W \cdot g(c; \mu_*, \sigma_*^2) \tag{3}$$

Here, $W$ is a proportionality constant

$$W = \frac{\sigma_*}{\sqrt{2\pi\gamma a\omega(1-\omega)}} \exp\left(-\frac{1}{2}\frac{(a\omega+\gamma-b)^2}{a\omega(1-\omega)+\gamma}\right),$$

the effective mean is

$$\mu_* = \frac{a\omega\gamma + (b-\gamma)a\omega(1-\omega)}{a\omega(1-\omega)+\gamma},$$

and the effective variance is

$$\sigma_*^2 = \frac{a\omega(1-\omega)\gamma}{a\omega(1-\omega)+\gamma}.$$

Using (3), we can approximate (2) using the following (4) when $a\omega$, $a(1-\omega)$, and $\gamma$ are all large enough.

$$P_{a,b} \approx P_{a,b}^{\text{asymptotic}} = \sum_{c=0}^{m=\min\{a,b\}} W \cdot g(c; \mu_*, \sigma_*^2) \tag{4}$$

The summation in (4) can be further approximated by the integral

$$P_{a,b}^{\text{I}} = W \cdot \int_{x=0}^{m=\min\{a,b\}} g(x; \mu_* - 0.5, \sigma_*^2)\text{d}x.$$

Here the subtraction of 0.5 from the mean is a continuity correction due to the switch from a discrete sum to a continuous integration. Then, the integral reduces to the CDF of the normal distribution:

$$\int_{x=0}^{m=\min\{a,b\}} g\left(x; \mu_* - 0.5, \sigma_*^2\right) \mathrm{d}x = \Phi\left(\frac{m - \mu_* + 0.5}{\sigma_*}\right) - \Phi\left(\frac{-\mu_* + 0.5}{\sigma_*}\right).$$

Here $\Phi(x)$ is the CDF of a standard normal random variable. Using the integral approximation with continuity correction to approximate $P_{a,b}$ leads to the following (5).

$$P_{a,b} \approx P_{a,b}^{\mathrm{I}} = W \cdot \left(\Phi\left(\frac{m - \mu_* + 0.5}{\sigma_*}\right) - \Phi\left(\frac{-\mu_* + 0.5}{\sigma_*}\right)\right) \tag{5}$$

We note a few caveats to (5) for approximating $P_{a,b}$. When $\sigma_* = 0$, the approximation fails. However, this only happens when at least one of the following are true: $a = 0$, $\omega = 0$, $\omega = 1$, or $\gamma = 0$. For the first and second cases, we are either transitioning away from a state with population size 0 ($a = 0$), so that there is no survival term, or there are no survivals from the previous state ($\omega = 0$). This means that no approximating is necessary, as the convolution in $P_{a,b}$ collapses to a single computation: $P_{a,b} = \mathrm{Pois}(b; \gamma)$. The third and fourth cases similarly indicate that all population change is due to either recruitments ($\omega = 1$), so that $P_{a,b} = \mathrm{Pois}(b - a; \gamma)$ (and $b - a \geq 0$), or due to population survival ($\gamma = 0$), so that the convolution collapses to $P_{a,b} = \mathrm{Binom}(b; a, \omega)$.

Comparing (2) to (5), we have reduced the number of density calculations necessary from $2m$ to two. Thus we have reduced the complexity of computing of $P_{a,b}$ from $\mathcal{O}(m)$ to $\mathcal{O}(1)$. This provides a large improvement in complexity when calculating the full transition probability matrix $M_K$, which in practice is the function $P_{a,b}$ calculated $(K+1)^2$ times for each iteration of the optimizer (once for each combination of $a \in \{0, 1, 2, ..., K\}$ and

$b \in \{0, 1, 2, ..., K\}$). Here, $K$ is the upper bound on the summations in the likelihood function (1), so that when the population size is large, $K$ must necessarily also be large.

## 2.3 $P_{a,b}$ Approximation Error

An alternative to the asymptotic approximation to $P_{a,b}$ shown in (4) exists in solving explicitly and in closed form the original $P_{a,b}$. The difficulty in finding a closed form solution lies in calculating the normalizing constant $D$ for the distribution, shown in (6).

$$D = \sum_{c=0}^{\min\{a,b\}} \binom{a}{c} \omega^c (1-\omega)^{-c} \gamma^{-c} / (b-c)!. \tag{6}$$

If a closed form for $D$ exists, the resulting equation would be an exact solution with all of the computational benefits of the asymptotic approximation, for which a similar integral approximation to (5) could be applied. In the absence of a closed form solution, the computational cost of calculating $D$ precludes its use.

We investigated the error structure associated with making the asymptotic approximation $P^I_{a,b}$. In Figure 1, we illustrate the error structure of the asymptotic transition probability matrix $M^I_K$, computed using (5). Figure 1 (top left) shows the transition probability matrix calculated using $P_{a,b}$. Figure 1 (top right) shows the asymptotic transition probability matrix calculated using $P^I_{a,b}$. Figure 1 (bottom left) shows the difference between the two matrices, $M_K - M^I_K$. For this comparison we chose $K = 100$, $\gamma = 45$, and $\omega = 0.5$. We note that other parameter values will give similar results, with $\omega$ determining the slope of the diagonal structure, $\gamma$ determining the $b$-axis intercept of the diagonal structure, and $K$ determining the size of the matrix. The two matrices $M_K$ and $M^I_K$ are nearly identical. However, a triangular region of noticeable error is visible near $a = 1$, and $b = \gamma = 45$ (see Supplemental Figure 1 for an enlarged view of the error region). Fortunately, the triangular error region

can be computed exactly at negligible computing cost. This is because when $a$ is small, the convolution in the computation of $P_{a,b}$ has very few terms. When calculating $M_K^I$, we will refer to calculating the triangular region exactly as the "small $a$ correction."

## 2.4 Model Extensions

The asymptotic model described in Section 2.2 can be easily extended to add parameter covariates. Parameters in the likelihood (1) and (5) can be replaced by corresponding covariate summations. For example, suppose we would like to include covariates for the probability of detection parameter $p$. We consider the set of covariates $\{x_j\}$, with $j \in \{1, 2, ..., J\}$. Then we define $\beta_0$ to be the baseline probability of detection, and $\beta_j$ to be the coefficients for each covariate $x_j$ (the additive effect size on $\beta_0$ due to the covariate $x_j$). In this case, the parameter replacement in the likelihood function would be $p \to \beta_0 + \sum_{j=1}^{J} \beta_j x_j$.

In practice, it is necessary to limit the range of the parameter values during likelihood optimization. The parameters $p$ and $\omega$ are probability parameters, and so must take values between zero and one. This can be guaranteed using the logit transformation, for example so that $\text{logit}(p) = \beta_0 + \sum_{j=1}^{J} \beta_j x_j$. Likewise, the parameters $\lambda$ and $\gamma$ must take non-negative values, so that a log transform is appropriate. For example, $\log(\lambda) = \beta_0 + \sum_{j=1}^{J} \beta_j x_j$.

For large values of $K$, the transition probability matrix will be computationally time consuming to calculate. This can be partially alleviated by breaking the matrix into rows, which can be calculated independently in parallel. When $K$ is small, this parallel computing solution will be much slower than computing the matrix in serial, due to the overhead of using parallel computing. However, as $K$ grows, the benefit of using parallel computing to calculate $M_K^I$ increases. The R package *optimParallel* (Gerber and Furrer, 2019) provides an alternative method of utilizing parallelization to improve compute times. This method uses parallel computing to approximate the Hessian matrix during likelihood optimization. This method has the disadvantage that the number of cores which can be utilized effectively in

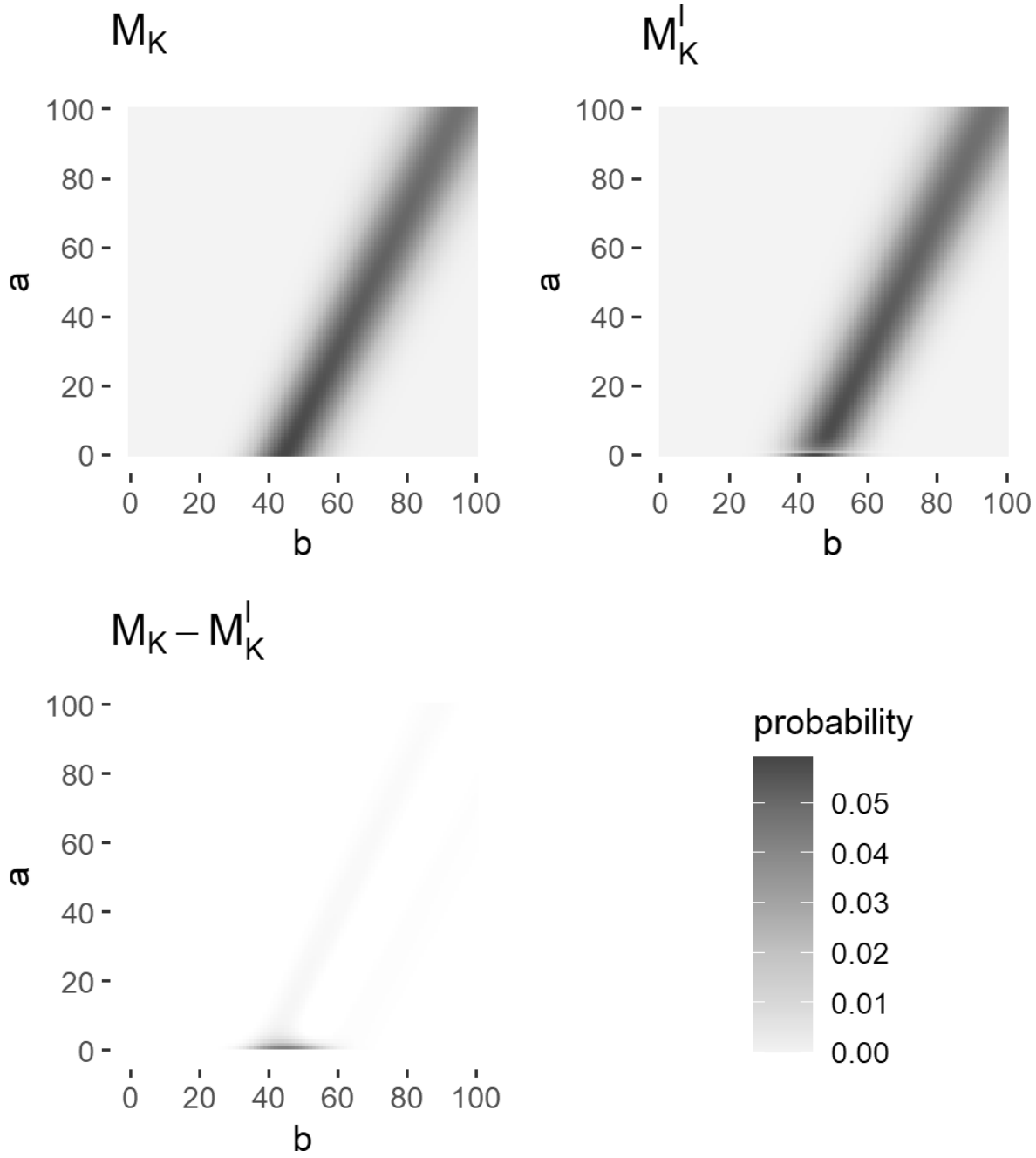


Figure 1: **Top Left:** Transition probability matrix $M_K$. **Top Right:** Asymptotic transition probability matrix $M_K^I$. **Bottom Left:** The error matrix associated with using $M_K^I$ over $M_K$. See Supplemental Figure 1 for an enlarged view of the error region.

parallelization is determined by the number of parameters in the model, so that the benefit of additional compute cores cannot be realized.

When considering models with different covariate structures, model selection can be implemented using techniques such as AIC (Akaike, 1974) or BIC (Schwarz, 1978). Goodness of fit can be difficult to assess for $N$-mixture type models. This is an active area of research, and some recent progress has been made (see for example: Knape et al., 2018; Duarte et al., 2018; Costa et al., 2021).

## 2.5 Standard Error Estimation

Estimated standard errors (ESEs) are useful for understanding the reliability of parameter estimates. Likelihood optimization, such as through use of the BFGS algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970), produces maximum likelihood estimates (MLEs) of parameter values at a local likelihood function maximum. The MLEs define a point on the likelihood manifold, and the second derivative matrix of the log-likelihood function can be estimated at the MLE point. The negative of the second derivative matrix is known as the observed Fisher Information matrix (negative Hessian matrix). The diagonal entries of the inverse Hessian matrix are asymptotic estimates for the variance of the MLE parameter estimates, so that their square roots are ESEs for the parameter estimates (see for example: Efron and Hinkley, 1978).

Maximum likelihood estimates have the advantage of asymptotic normality (Bain and Engelhardt, 1992, p. 316), which allows asymptotic confidence intervals to be constructed using the estimated parameter values, normal distribution quantiles, and the ESEs for the estimated parameters. For example, let $\alpha$ be a parameter, with MLE $\hat{\alpha}$, and ESE $\hat{\sigma}_\alpha$. An asymptotic 95% confidence interval for $\alpha$ would be: $\hat{\alpha} \pm 1.96 \times \hat{\sigma}_\alpha$. We use this method to calculate our parameter confidence intervals in Section 4.

## 2.6 $N$-Mixture Model Problems

There are many known issues with $N$-mixture models, and many arguments against their use (see for example: Barker et al., 2018; Link et al., 2018). Any issues with dynamic $N$-mixture models are also likely to be issues for the asymptotic models we propose in this paper. However, there is currently no replacement for $N$-mixture models without additional data demands, and a plethora of urgent applications for such models, ensuring that $N$-mixtures continue to see wide spread use in ecological monitoring, disease analytics, pest management, and more (Manica et al., 2019; Zhao, 2021; Parker et al., 2021b). Alternative models, such as those in the capture-recapture literature (Cormack, 1964; Jolly, 1965; Schwarz and Seber, 1999), require additional data such as capture histories, which in practice can be costly or impractical to collect. When only count data are available, $N$-mixture models can be used, but care must be taken that model assumptions are not violated (Fogarty and Fleishman, 2021), "infinite abundance" estimates are checked for (Dennis et al., 2015), there is high data quality (Link et al., 2018), sufficient count sizes have been collected at each sampling occasion (Barker et al., 2018), and that sufficient sampling occasions are used (Dennis et al., 2015). We note that the work of Dennis et al. (2015) is focused on the closed population models, and we assume in this work that the dynamic models require at least as many time replicates. When supplemental data exists beyond simple counts, or when it is feasible to collect such additional data, more reliable estimates can often be obtained by using alternative models (such as encounter histories for capture-recapture models). An important extension to $N$-mixture models is the robust design of secondary sampling occasions for which the closed population assumption is used within sampling seasons (Zhao and Royle, 2019; Costa et al., 2021). This extension requires additional data collection, which is not always feasible. However, when such data is collected, model estimates can be improved over the traditional models.

## 2.7 R Package: quickNmix

Novel code for fitting the asymptotic model is contained in our R package: *quickNmix* (Parker et al., 2021a). The package is available for download from CRAN, or from github: www.github.com/mrparker909/quickNmix. The package allows for site and time dependent parameter covariates to be incorporated into the model fitting, and also allows for parallel computing to calculate the transition probability matrix more efficiently for large $K$, and efficient parallel computing for large numbers of parameters using the R package *optimParallel* (Gerber and Furrer, 2019).

# 3 Simulations

We conducted three distinct simulation studies, varying the value $\lambda$ over the set $\{100, 500, 1000\}$. The purpose of the first two simulations is to compare our asymptotic model directly against R package *unmarked*. The purpose of the third simulation is to illustrate the efficacy of the new asymptotic model in the large population regime where *unmarked* becomes untenable due to large computation times and numerical precision issues. For all simulations, we used the small $a$ correction when computing the asymptotic likelihood probability transition matrix. We apply the small $a$ correction (see Section 2.3) to the triangular region of the matrix determined by $a < (b - 0.25\gamma)/2$ and $a < (1.75\gamma - b)/2$.

We note that for the two simulation sets which compare against *unmarked*, the stochastic nature of the simulations can cause some generated populations to approach $K$ in size. This leads to parameter estimates which could have been improved by increasing $K$. However, computing time is largely determined by the choice of $K$. Thus, we chose to keep $K$ constant within each simulation set, in order to appropriately compare the computation times.

We chose to design the simulations to match the real world data in our Ancient Murrelet case study, Section 4. For this reason we use 6 sampling sites and 17 sampling occasions

for all simulated data sets. For each set of simulations, we ran 100 iterations for each combination of parameter values. For the first two simulations, the parameters were chosen from: $\gamma \in \{3, 6, ..., 30\} \times \frac{\lambda}{100}$, $\omega \in \{0.25, 0.30, 0.35, ..., 0.75\}$, and $p = 0.75$. For the third simulation set the parameters where chosen from: $\gamma \in \{25, 150, 300\}$, $\omega \in \{0.25, 0.50, 0.75\}$, and $p = 0.75$. We illustrate the comparison between the asymptotic model and the *unmarked* model using the computation times, the ratio of negative log-likelihood (nll) functions, and the distributions of the estimated parameters.

For each iteration of the simulations, a random population/observation pair ($\{N_{it}\}$, $\{n_{it}\}$) was generated using the $N$-mixture distributions laid out in Section 2.1. The same generated population was then used in fitting both the asymptotic model, and the *unmarked* model. Simulations 1-3 were run on Westgrid, using Cedar environment 2016.4 and R version 3.5.0 (R Core Team, 2020), while simulation 4 was run on an AMD Ryzen 9 3900X with 24 logical processors. The R package *optimParallel* (Gerber and Furrer, 2019) was used to decrease compute time for our asymptotic model in simulation 4.

## 3.1 Simulation 1: Initial population size $\lambda = 100$

For the $\lambda = 100$ simulation, we chose to use $K = 300$ as the upper bound on summations. We expected the asymptotic approximation to improve with increasing population sizes. As such, $\lambda = 100$ would be considered a "small" population size where the traditional $N$-mixture model would be more appropriate. This simulation set is intended to show that the asymptotic model performs adequately compared to *unmarked* even in the relatively small population scenario.

We compared the computation times for the asymptotic model and the *unmarked* model using boxplots (Supplemental Figure 2). Here, *unmarked* outperformed the asymptotic model in computation time. However, this is due to the specific implementation of the algorithm in the *unmarked* package, which is optimized using C++, rather than a benefit of the

algorithm complexity (as we will see with the subsequent simulations). Plots of the empirical distributions of the estimated parameter values (Supplemental Figure 3), along with the ground truth parameter distributions show that the asymptotic and the *unmarked* models produce similar estimates. However, the distribution of $\lambda$ estimates is skewed towards larger values of $\lambda$ for the asymptotic models than for the *unmarked* models, and the distribution of $p$ estimates is skewed towards smaller values. This discrepancy is small and diminishes for larger true values of $\lambda$ (as will be seen in Section 3.2). Supplemental Figure 4 illustrates the ratio of the likelihood function values for the *unmarked* and asymptotic models for 12,100 simulations. We include the case $\gamma = 0$, illustrating that the surfaces are nearly identical, except when $\gamma$ is small. The difference between the two likelihood surfaces for small $\gamma$ is dependent on the value of $\omega$: when $\omega > 0.65$ the likelihood surfaces are nearly identical even when $\gamma$ is small. See Section 5 for a discussion of the small $\gamma$ problem. When $\gamma \geq 6$, we see that the two surfaces are essentially identical, with the inter-quartile range of the likelihood ratios decreasing for either increasing $\gamma$ or increasing $\omega$.

## 3.2 Simulation 2: Initial population size $\lambda = 500$

For the $\lambda = 500$ simulations, we chose to use $K = 800$ as the upper bound for summations. This set of simulations is intended to illustrate the effectiveness of using the asymptotic models over the *unmarked* models when population sizes become large. In this large $\lambda$ scenario, the asymptotic model outperforms *unmarked* in computation time (Supplemental Figure 5), being roughly twice as fast for model fitting.

Further, both the asymptotic models and the *unmarked* models produce similar parameter estimates. A plot of the empirical distributions of the estimated parameter values, along with the ground truth parameter values (Figure 2) indicates that the asymptotic and the `unmarked` models perform similarly. The skewness in the asymptotic parameter estimates for $\lambda$ and $p$ when compared with the *unmarked* parameter estimates, which was evident in

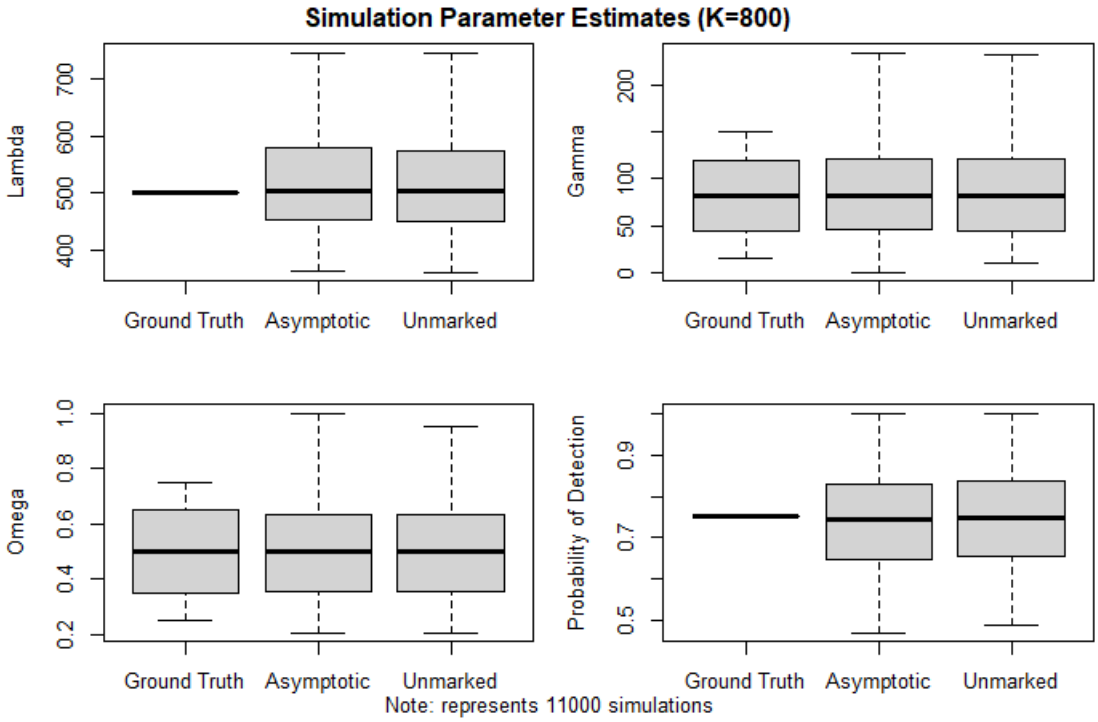


Figure 2: Parameter ground truths and parameter estimates from fitting asymptotic and *unmarked* (traditional) $N$-mixture models with $K = 800$. Estimated parameters are $\lambda$ (initial mean site abundance), $\gamma$ (importation rate), $\omega$ (survival probability), and $p$ (probability of detection). A total of 11,000 simulations are represented.

Section 3.1, is not evident with $\lambda$ increased from 100 to 500.

## 3.3 Simulation 3: Initial population size $\lambda = 1000$

For the $\lambda = 1000$ simulations, we chose to use $K = 2000$. For $K$ this large, it becomes impractical due to computation times to compare the asymptotic models against the *unmarked* models. For this reason we only consider simulations for the asymptotic model.

Supplemental Figure 7 shows the empirical distributions of the estimated parameter values. The estimated parameter densities are seen to be very similar to the true parameter densities, with the distribution medians closely matched. No bias and no skewness are evident in the estimates.

### 3.4 Simulation 4: Comparing against asymptotic MVN

Brintz et al. (2018) developed an alternative asymptotic approximation that assumes a multivariate normal distribution on the random vector of the unobserved population sizes for each sampling occasion (in contrast, our approximation targets the terms in the convolution—the tightest bottleneck in the asymptotic complexity). We will refer to the multivariate normal approximation method as AsymMVN. The advantage of AsymMVN over both the traditional $N$-mixture models, and our asymptotic approximation, is computational efficiency. AsymMVN has computational complexity $\mathcal{O}(1)$ in terms of population upper bound $K$, while our approach has $\mathcal{O}(K^2)$, and the traditional model has $\mathcal{O}(K^3)$. To compare the accuracy and precision of the three methods, we chose ground truth parameter values to be similar to the estimates obtained for the Ancient Murrelet population detailed in Section 4, such that $\lambda = 250$, $\gamma = 10$, $\omega = 0.8$, $p = 0.5$, $M = 17$, $U = 6$, and $K = 600$.

We generated 100 population observation pairs (see Section 2.1), and used each set of observations to fit three separate models: the traditional method (*unmarked*), our asymptotic model with likelihood optimized using the R package *optimParallel* (AsymP; Gerber and Furrer, 2019), and also the multivariate normal approximation (AsymMVN). Our asymptotic model estimates similar mean parameter values as the traditional $N$-mixture model, exhibits less parameter uncertainty than AsymMVN, and mean parameter values that are closer to the ground truth than those of AsymMVN (Table 1).

## 4 Application: Ancient Murrelet Chicks

Ancient Murrelet seabirds are a species of special concern, due to population declines and colony collapse due to excessive predation; see for example Gaston et al. (2009) and Major et al. (2012). East Limestone Island, Haida Gwaii, BC, is home to a colony of Ancient Murrelet seabirds. We use the Ancient Murrelet chick count data from Parker et al. (2020),

Table 1: Results from simulation study 4. Shown are the mean estimates for each parameter ($\lambda$, $\gamma$, $\omega$, $p$) and mean computation time in seconds. AsymMVN excludes the17 simulations where that method failed due to non-invertible matrices. The mean standard deviation for each parameter estimate is shown in parenthesis (calculated using the estimated Hessian matrices).

| **Method** | **mean**($\hat{\lambda}$) | **mean**($\hat{\gamma}$) | **mean**($\hat{\omega}$) | **mean**($\hat{p}$) | **mean**(**time**) |
|---|---|---|---|---|---|
| **Ground Truth** | 250 | 10 | 0.8 | 0.5 | |
| **Traditional** | 276.79 (87.73) | 10.95 (3.68) | 0.80 (0.01) | 0.48 (0.11) | 1069.18 (83.63) |
| **AsymP** | 279.57 (87.09) | 11.10 (3.60) | 0.80 (0.01) | 0.48 (0.11) | 682.80 (113.80) |
| **AsymMVN** | 326.10 (330.86) | 12.74 (11.80) | 0.80 (0.01) | 0.47 (0.13) | 1.66 (0.58) |

which was collected at the East Limestone Island colony by the Laskeek Bay Conversation Society during the years 1990 to 2006. The Ancient Murrelet chick count data is plotted in Figure 3. The data consists of 17 years worth of annual chick counts taken during the hatching period (from early May to late June), at 6 separate trapping regions set up on the island. The trapping funnels were set up identically each year, so that the methodology is consistent across sampling occasions. Due to geographic features such as ridge lines, new fledglings from a particular burrow are extremely likely to use the same funnel from year to year, making the sites spatially distinct. We chose an upper bound on summations of $K = 600$. We verified the choice of $K$ by testing the change in parameter estimates at $K = 800$, and saw no change; thus we confirmed that the model had converged.

In Table 2 we compare the traditional method (*unmarked*), our asymptotic model (Asym), our asymptotic model with likelihood optimized using the R package *optimParallel* (AsymP; Gerber and Furrer, 2019), and also the multivariate normal approximation from Brintz et al. (2018; AsymMVN). Comparing parameter estimates between the traditional method and the Asym parameter estimates, the two models perform similarly. The estimated standard errors are also closely matched. The ratio of negative log-likelihoods (*unmarked* nll / asymptotic nll) is 0.9989; together with the similar estimated standard errors, this shows that the likelihood surfaces near their maximum values are nearly identical between the two models. The

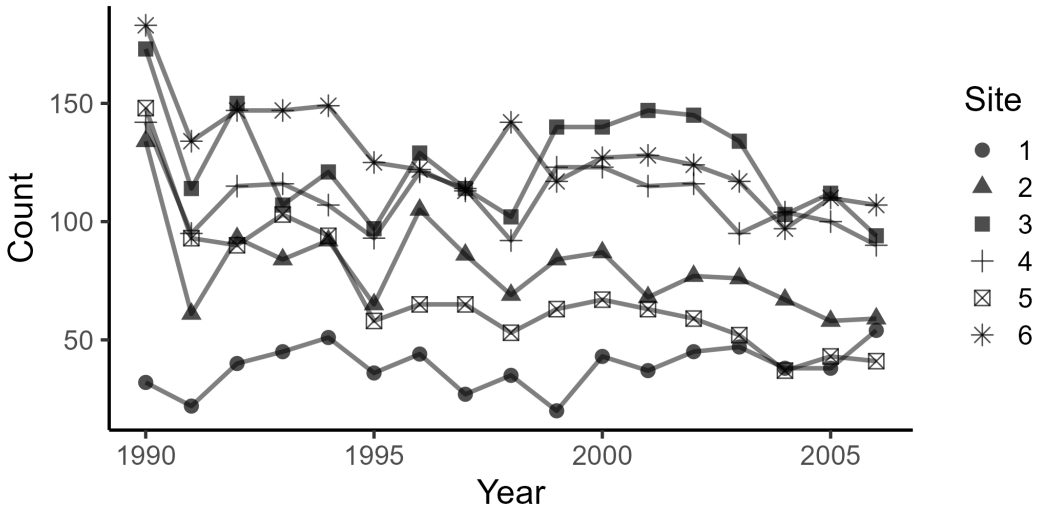


Figure 3: Time series data collected by the Laskeek Bay Conservation Society on annual Ancient Murrelet chick counts from the year 1990 to 2006. The data is collected for six sampling sites on East Limestone Island. The locations of these six sites on the island are provided in Parker et al. (2020).

advantage of using the asymptotic model is apparent in the decreased computation time, which in this example is about 2.25 times faster than the traditional model implemented in *unmarked*. We show that further computing time gains are possible by using the R package *optimParallel* (Gerber and Furrer, 2019), which was used to optimize our asymptotic likelihood 11 times faster than *unmarked* through parallel computing with 9 compute cores. We also compare Asym against AsymMVN, which was by far the fastest method, computing 540 times faster than *unmarked*. The parameter estimates are similar between all four methods. However, for this case study the standard error estimates are substantially larger for the AsymMVN method. Thus, there is a trade off between computation speed and precision of estimates when using the AsymMVN method. The case study was run using a 4.0 GHz AMD Ryzen 9 3900X CPU, and using R version 3.5.0 (R Core Team, 2020).

Table 2: Results from fitting dynamic $N$-mixture models to the Ancient Murrelet chick count data using the *unmarked* (traditional) model, and the asymptotic models. AsymMVN represents the multivariate normal approximation from Brintz et al. (2018), Asym represents our asymptotic approximation, and AsymP represents our asymptotic approximation run in parallel using *optimParallel* (Gerber and Furrer, 2019). For this population, there are $M = 17$ sample times, $U = 6$ sites, and we chose to use $K = 600$ as the upper bound on summations. The table includes computation time in seconds, the four parameter estimates log and logit transformed ($\log(\hat{\lambda})$, $\log(\hat{\gamma})$, $\text{logit}(\hat{\omega})$, $\text{logit}(\hat{p})$), the negative log-likelihood (nll) evaluated at the parameter estimates, and parameter standard error estimates in parentheses. The standard error estimates were calculated using the estimated Hessian matrix.

| | `unmarked` | **Asym** | **AsymP** | **AsymMVN** |
|---|---|---|---|---|
| **computation time (s)** | 928.69 | 410.10 | 83.45 | 1.72 |
| **log($\hat{\lambda}$)** | 5.494 (0.068) | 5.487 (0.067) | 5.487 (0.067) | 5.608 (3.932) |
| **log($\hat{\gamma}$)** | 1.991 (0.216) | 2.125 (0.207) | 2.125 (0.207) | 2.125 (1.708) |
| **logit($\hat{\omega}$)** | 2.727 (0.163) | 2.621 (0.165) | 2.621 (0.165) | 2.704 (3.652) |
| **logit($\hat{p}$)** | -0.134 (0.126) | -0.114 (0.127) | -0.114 (0.127) | -0.086 (2.309) |
| **nll** | 607.79 | 608.44 | 608.44 | 577.15 |

# 5 Discussion

We have developed new asymptotic solutions to the open-population $N$-mixture models. We have verified the models with simulation studies comparing the asymptotic model against the popular *unmarked* implementation of the traditional models. We have found that the asymptotic models perform well, providing excellent parameter estimates comparable in accuracy to the traditional models. We have also shown that the computation times are much improved when comparing the asymptotic model to the traditional models when population sizes are large.

For our simulations, we did not use parameter covariates in the models. However, it is straightforward to include parameter covariates. When time covariates are included for either $\gamma$ or $\omega$, the matrix $M_K$ will need to be recalculated for each time point. The effect of this on model fitting is to greatly increase computing times (proportional to $M - 1$). This increase occurs for both the traditional $N$-mixture models, and for the asymptotic models.

However, since the asymptotic model computes the matrix $M_K$ much more efficiently than the traditional models, the asymptotic models will be faster in comparison to the traditional models than is indicated in the simulations with constant parameters.

In Section 3.3, we compared the computation times for the asymptotic model using boxplots (Supplemental Figure 6). For $K = 2000$, model fitting takes on the order of 5 to 10 hours per model. Comparing against *unmarked* would be computationally infeasible. However, we would expect model fitting to take on the order of 10 times longer for the *unmarked* models (around 50 to 100 hours per model). Similar to the previous simulations from Sections 3.1 and 3.2, computation time for the asymptotic model for a given value of $K$ is seen to be dependent on the true parameter value $\omega$. Larger values of $\omega$ tend to decrease computation time. The computation time for fixed $K$ is largely influenced by the number of iterations taken by the optimization algorithm when optimizing the likelihood function. Since the optimization algorithm used is Quasi-Newton like (linear extrapolation based on first derivative approximations with an approximate Hessian correction), it is probable that smaller values of $\omega$ lead to either rougher likelihood surfaces, or to larger curvatures in the likelihood surfaces.

The asymptotic model has several shortcomings. When $\gamma = 0$, or when $\hat{\gamma} \approx 0$, parameter identifiability is an issue. In this scenario, $\hat{\lambda}$ increases, bounded only by choice of $K$, and $\hat{p}$ shrinks in proportion to $\hat{\lambda}$. This indicates that $\lambda$ and $p$ are not identifiable when $\gamma \approx 0$. Fortunately, this deficiency is easily diagnosed: if the fitted model gives estimated parameter $\hat{\gamma}$ close to zero, and the estimated parameter $\hat{\lambda}$ close to $K$ even for large $K$, then we can conclude that the asymptotic model is a poor choice given the observed data. A second shortcoming is evident in (5), which only holds when $a\omega$, $a(1-\omega)$, and $\gamma$ are all large enough (where $a$ denotes population size). This indicates that the approximation will only be good if the probability of survival $\omega$ is not near the boundary (not close to 0 or 1). As $\omega$ nears the boundary values, the approximation would require larger and larger population

sizes $a$ to achieve similar accuracy. For this reason, models which find estimates for $\omega$ close to either 0 or 1 should be considered suspect.

We used an integral approximation with continuity correction for the sum in (4). However, other approximations exist for calculating a sum using an integral. As an example of another approximation method, the Euler-Maclaurin formula is shown in (7) (see for details Mollin 2009, Chapter 5.1).

$$\sum_{x=x_0}^{x_1} f(x) = \int_{x_0}^{x_1} f(x)\mathrm{d}x + \frac{f(x_0)+f(x_1)}{2} + \sum_{k=1}^{\lfloor q/2 \rfloor} \frac{B_{2k}}{(2k)!}\left[f^{(2k-1)}(x_1) - f^{(2k-1)}(x_0)\right] + R_q. \tag{7}$$

Here $B_i$ is the $i^{\text{th}}$ Bernoulli number, and $R_q$ is the remainder (or error term). Note that $q$ is a chosen stopping point for the approximation. We found that the Euler-Maclaurin formula boundary correction $\frac{f(x_0)+f(x_1)}{2}$ performed slightly worse than the continuity correction for this application. We also found that using additional terms in the formula (increasing $q$) caused the approximation to diverge. For these reasons we chose to use the continuity correction instead of the Euler-Maclaurin formula.

Our asymptotic likelihood function has favourable performance in computation time when compared against the popular *unmarked* implementation, and produces both parameter estimates and standard error estimates which are similar to those produced by traditional $N$-mixtures methods. $N$-mixture models have far reaching applications both within ecology, and beyond. These new methods allow much larger populations to be studied than previously possible while still retaining the favourable standard error estimates (compared to the MVN approximation) of the traditional $N$-mixture models. Some caution should be employed when choosing to use the asymptotic likelihood, since the distributional assumptions made for the asymptotic approximations of (2) may be invalid for some extreme populations (such as when $\gamma$ is very small). Applying these models to future large population studies will be simplified by use of our asymptotic $N$-mixtures R package, *quickNmix*.

# Acknowledgements

We would like to thank the Editor, the Associate Editor and two anonymous referees for their careful review of our manuscript and their many insightful comments. Their comments are very helpful for us to improve our work. We would also like to acknowledge the Micheal Smith Foundation for Health Research and the Victoria Hospitals Foundation for support through a COVID-19 Research Response grant, as well as a Canadian Statistical Sciences Institute Rapid Response Program—COVID-19 grant to LC that supported this research. We also acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) for providing PGS-D support for MP [funding reference number 569754]. JC was supported by an NSERC discovery grant (RGPIN-2023-04057) and the Canada Research Chairs program. LTE was supported by NSERC grant numbers RGPIN/05484-2019 and DGECR/00118-2019. We thank the many Laskeek Bay Conservation Society volunteers and staff who have been counting Ancient Murrelet chicks since 1990. This research was enabled in part by support provided by WestGrid (www.westgrid.ca) and Compute Canada/Calcul Canada (www.computecanada.ca). We also thank Jordan Bell and Paul Turner for helpful discussion.

# Author Contributions

All authors contributed to the methodology, and all authors contributed to writing the manuscript. The authors have no conflict of interest to report.

# References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723. Conference Name: IEEE Transactions on Automatic

Control.

Bain and Engelhardt (1992). *Introduction to Probability and Mathematical Statistics*. Brooks/Cole, Cengage Learning, 2nd edition.

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

Belant, J. L., Bled, F., Wilton, C. M., Fyumagwa, R., Mwampeta, S. B., and Jr, D. E. B. (2016). Estimating lion abundance using N-mixture models for social species. *Scientific Reports*, 6:35920.

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

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms: 2. The new algorithm. *IMA Journal of Applied Mathematics*, 6(3):222–231.

Cormack, R. M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika*, 51(3/4):429–438. Publisher: [Oxford University Press, Biometrika Trust].

Costa, A., Salvidio, S., Penner, J., and Basile, M. (2021). Time-for-space substitution in N-mixture models for estimating population trends: a simulation-based evaluation. *Scientific Reports*, 11(1):4581.

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

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

DiRenzo, G. V., Che-Castaldo, C., Saunders, S. P., Grant, E. H. C., and 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.

Duarte, A., Adams, M. J., and Peterson, J. T. (2018). Fitting N-mixture models to count data with unmodeled heterogeneity: Bias, diagnostics, and alternative approaches. *Ecological Modelling*, 374:51–59.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 65(3):457–483.

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

Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322.

Fogarty, F. A. and Fleishman, E. (2021). Bias in estimated breeding-bird abundance from closure-assumption violations. *Ecological indicators*, 131:108170.

Gaston, A. J., Bertram, D. F., Boyne, A. W., Chardine, J. W., Davoren, G., Diamond, A. W., Hedd, A., Montevecchi, W. A., Hipfner, J. M., Lemon, M. J. F., Mallory, M. L., Rail, J.-F., and Robertson, G. J. (2009). Changes in Canadian seabird populations and ecology since 1970 in relation to changes in oceanography and food webs. *Environmental Reviews*.

Gerber, F. and Furrer, R. (2019). optimParallel: An R package providing a parallel version of the L-BFGS-B optimization method. *The R Journal*, 11(1):352–358.

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109):23–26.

Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, 52(1/2):225–247.

Knape, J., Arlt, D., Barraquand, F., Berg, A., Chevalier, M., Pärt, T., Ruete, A., and Żmihorski, M. (2018). Sensitivity of binomial N-mixture models to overdispersion: The importance of assessing model fit. *Methods in Ecology and Evolution*, 9(10):2102–2114.

Kwon, E., Houghton, L. M., Settlage, R. E., Catlin, D. H., Karpanty, S. M., and Fraser, J. D. (2018). Estimating transient populations of unmarked individuals at a migratory stopover site using generalized N-mixture models. *Journal of Applied Ecology*, 55(6):2917–2932.

Link, W. A., Schofield, M. R., Barker, R. J., and Sauer, J. R. (2018). On the robustness of N-mixture models. *Ecology (Durham)*, 99(7):1547–1551.

Madsen, L. and Royle, J. A. (2023). A review of n-mixture models. *Wiley interdisciplinary reviews. Computational statistics*, 15(6):e1625.

Major, H. L., Lemon, M. J. F., and Hipfner, J. M. (2012). Habitat as a potential factor limiting the recovery of a population of nocturnal seabirds: Ancient Murrelet Habitat Use. *The Journal of Wildlife Management*, 76(4):793–799.

Manica, M., Caputo, B., Screti, A., Filipponi, F., Rosà, R., Solimini, A., della Torre, A., Blangiardo, M., and Pocock, M. (2019). Applying the N-mixture model approach to estimate mosquito population absolute abundance from monitoring data. *The Journal of applied ecology*, 56(9):2225–2235.

Mollin, R. A. (2009). *Advanced Number Theory with Applications*. Chapman and Hall/CRC.

Parker, M. R. P., Cowen, L. L. E., Cao, J., and Elliott, L. T. (2023). Computational efficiency and precision for replicated-count and batch-marked hidden population models. *JABES*, 28:43–58.

Parker, M. R. P., Elliott, L. T., Cowen, L. L. E., and Cao, J. (2021a). *quickNmix: Asymptotic N-Mixture Model Fitting*. R package version 1.0.3. https://CRAN.R-project.org/package=quickNmix.

Parker, M. R. P., Li, Y., Elliott, L. T., Ma, J., and Cowen, L. L. E. (2021b). Underreporting of COVID-19 in the Northern Health Authority region of British Columbia. *Canadian Journal of Statistics*, 49(4):1018–1038.

Parker, M. R. P., Pattison, V., and Cowen, L. L. E. (2020). Estimating population abundance using counts from an auxiliary population. *Environmental and Ecological Statistics*, 27(3):509–526.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

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

Schwarz, C. J. and Seber, G. A. F. (1999). Estimating animal abundance: Review III. *Statistical Science*, 14(4):427–456. Publisher: Institute of Mathematical Statistics.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464. Publisher: Institute of Mathematical Statistics.

Shanno, D. F. (1970). Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24(111):647–656. Publisher: American Mathematical Society.

Veech, J. A. and Cave, T. (2021). Using road surveys and N-mixture models to estimate the abundance of a cryptic lizard species. *Journal of Herpetology*, 55(1):46–54.

Vinga, S. and Almeida, J. S. (2004). Rényi continuous entropy of DNA sequences. *Journal of Theoretical Biology*, 231(3):377–388.

Zhao, Q. (2021). A simulation study of the age-structured spatially explicit dynamic N-mixture model. *Ecological research*, 36(4):744–754.

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

Zucchini, W. and MacDonald, I. L. (2009). *Hidden Markov Models for Time Series: An Introduction Using R*. Chapman and Hall/CRC.