

# Bayesian analysis of abundance for binomial sighting data with unknown number of marked individuals

Brett T. McClintock  
USGS Patuxent Wildlife Research Center

Jennifer A. Hoeting\*  
Colorado State University

June 2, 2008

## Abstract

The mark-resight method for estimating the size of a closed population can in many circumstances be a less expensive and less invasive alternative to traditional mark-recapture. Despite its potential advantages, one major drawback of traditional mark-resight methodology is that the number of marked individuals in the population available for resighting needs to be known exactly. In real field studies, this can be quite difficult to accomplish. Here we develop a Bayesian model for estimating abundance when sighting data are acquired from distinct sampling occasions without replacement, but the exact number of marked individuals is unknown. By first augmenting the data with some fixed number of individuals comprising a marked “super population,” the problem may then be reformulated in terms of estimating the proportion of this marked super population that was actually available for resighting. This then allows the data for the marked population available for resighting to be modeled as random realizations from a binomial logit-normal distribution. We demonstrate the use of our model to estimate the New Zealand robin (*Petroica australis*)

---

\*Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877

population size in a region of Fiordland National Park, New Zealand. We then evaluate the performance of the proposed model relative to other estimators via a series of simulation experiments. We generally found our model to have advantages over other models when sample sizes are smaller with individually heterogeneous resighting probabilities. Due to limited budgets and the inherent variability between individuals, this is a common occurrence in mark-resight population studies. WinBUGS and R code to carry out these analyses is available from <http://www.stat.colostate.edu/~jah/software>.

## 1 Introduction

For estimating closed population abundance,  $N$ , mark-resight methodology (Neal et al., 1993; White and Shenk, 2001) can often be a less expensive and less invasive alternative to traditional mark-recapture (Otis et al., 1978; Minta and Mangel, 1989). The primary advantage of this approach is that animals only need to be physically captured and marked once. After the initial investment in marking animals, subsequent sighting data from both marked and unmarked individuals are used to estimate  $N$ . Mark-resight can therefore be appealing to researchers because funds are often limited and capture is generally the most expensive aspect of mark-recapture studies. Not only can the financial cost of mark-recapture be daunting for long-term population monitoring, but capture is also the most hazardous aspect for the animals. Mark-resight can substantially reduce stress to animals because they can be observed at a distance with minimal disturbance after the initial capture.

Despite the possible advantages, one major drawback of mark-resight methods is that the number of marked individuals available for resighting usually needs to be known exactly. This can often be quite difficult to accomplish in the field (McClintock et al., 2008a). Arnason, Schwarz and Gerard (1991) developed a model for estimating abundance for sighting data when an unknown number of marked individuals were present in the population of interest, but this model has found little application due to several key limitations (McClintock et al., 2008a). McClintock et al. (2008a) introduced a model, the zero-truncated Poisson-log normal estimator (ZPNE), that addressed these limitations, but they identified several conditions when the distributional assumptions of their model began to break down. In par-

ticular, they found their estimator to be biased when applied to small sample data generated without replacement under logit-normal distributed sighting probabilities. If sampling is without replacement (i.e., any single individual may only be sighted once per distinct sampling occasion) and the number of marked individuals in the population available for resighting is known exactly, McClintock et al. (2008b) found advantages in the use of their logit-normal mixed effects estimator when compared to other estimators. Under such sampling conditions, the logit-normal model would be preferred to ZPNE. However, if the number of marked individuals is not known exactly, then the logit-normal likelihood is of unknown index and the model cannot be implemented.

Royle, Dorazio and Link (2007) proposed a data augmentation method to facilitate Bayesian analysis of multinomial capture-recapture models with unknown population size. This is often the case with mark-capture models (Otis et al., 1978) because the number of individuals in the population that were never encountered is unknown, but interest is in estimating the total population size. Bayesian analyses in these situations have been difficult because the dimension of the parameter space is not fixed and is a function of the unknown  $N$ . Royle, Dorazio and Link (2007) proposed augmenting the data with individuals that were never encountered but comprise part of some “super population” of  $M$  individuals,  $N$  of which make up the actual population size with unknown probability  $\psi$ . They demonstrate that a Bayesian analysis can be applied with appropriate  $M > N$ , and suggest that  $M$  be large enough so that mass of  $N$  is not concentrated near  $M$ . Here, we propose to modify this approach and investigate the utility of data augmentation as a method of estimating abundance with mark-resight data when sampling is without replacement and the number of marked individuals is unknown. After developing our Bayesian binomial logit-normal mixed effects (BBLNE) model in the next section, we implement the model in Section 3 to estimate the New Zealand robin (*Petroica australis*) population size in a region of Fiordland National Park, New Zealand. In Section 4, we compare the performance of our model to ZPNE under a variety of simulated sampling conditions. We conclude with a discussion of the implications of our work for abundance estimation in mark-resight studies.

## 2 The Binomial Logit-Normal Mixed Effects Model

### 2.1 Background

Mark-resight methodology requires some number of individuals to be labeled with field-readable marks prior to the collection of sighting data. For some closed interval(s) of time (i.e., the population is closed to birth, death, immigration and emigration) during which abundance is of interest, sighting data for both marked and unmarked individuals are collected. Using the terminology of the robust design (Kendall, Pollock and Brownie, 1995), closed interval  $j$ ,  $j = \{1, \dots, J\}$ , is referred to as the  $j$ th “primary” sampling interval which is comprised of  $k_j \geq 2$  “secondary” sampling occasions. When sampling is without replacement, then individuals may be seen at most once within each of the  $k_j$  distinct secondary sampling occasions. If individually identifiable marks are used, the data consist of resightings,  $\delta_{sij} = \{0, 1\}$ , for marked individual  $s$  on secondary occasion  $i$  of primary interval  $j$  and the total number of unmarked individual sightings,  $U_{ij}$ . Across all  $k_j$  occasions within the  $j$ th primary interval, marked individual  $s$  is resighted  $\sum_{i=1}^{k_j} \delta_{sij} = y_{sj} = \{0, 1, \dots, k_j\}$  times, and  $U_j = \sum_{i=1}^{k_j} U_{ij}$ .

If  $n_{ij}$ , the number of marked individuals available for resighting during primary interval  $j$ , is known, then the data are completely observable and the logit-normal mixed effects model of McClintock et al. (2008b) is appropriate. Their model assumes: 1) geographic and demographic closure within each primary sampling interval; 2) sampling without replacement within secondary sampling occasions; 3) no loss of marks within primary intervals; 4) no errors in distinguishing marked and unmarked animals; and 5) a random sample of marked individuals, such that resighting probabilities for marked individuals are representative of unmarked individuals. Generally, this last assumption can be reasonably satisfied by using a different method for the selection of marked individuals (e.g., helicopter capture) than is used for the collection of sighting data (e.g., ground surveys on foot). The most general form of the logit-normal estimator assumes  $\delta_{sij}$  is a Bernoulli random variable with success probability  $p_{sij}$ . For simplicity, if one were to ignore temporal variation in sighting probabilities between secondary sampling occasions by constraining  $p_{sij} = p_{sj}$  ( $i = 1, \dots, k_j$ ), this general model becomes the binomial logit-normal mixed effects (BLNE) model. Assuming individ-

ually identifiable marks, the  $y_{sj}$  are then modeled as independent binomial logit-normal random variables with individual resighting probabilities  $p_{sj}$ . For each primary occasion  $j$ , these resighting probabilities are parameterized in terms of a fixed time effect  $\beta_j$  and a random individual heterogeneity effect with mean zero and unknown variance  $\sigma_j^2$ . The  $y_{sj}$  therefore have conditional expectation

$$E(y_{sj} | \sigma_j, Z_{sj}, \beta_j) = k_j p_{sj} = \frac{k_j}{1 + \exp(-(\sigma_j Z_{sj} + \beta_j))},$$

where  $Z_{sj} \sim N(0, 1)$ . This formulation allows parameterizations ignoring individual sighting heterogeneity ( $\sigma_j = 0$ ) and incorporation of individual or temporal covariate information on  $\sigma_j$  and  $\beta_j$ .

McClintock et al. (2008b) proposed a numerical maximum likelihood method to estimate the parameters in the BLNE model. If marks are individually identifiable, the BLNE model can estimate overdispersion due to individual sighting probability heterogeneity or lack of independence between animals (e.g., clustering). Also, if  $n_j$  is not known exactly, then the index of the marked individual binomial likelihood of BLNE is unknown, and the model cannot be used.

When the number of marks is unknown, then  $y_{sj} = 0$  is no longer observable for the marked population and only  $n_j^*$  individuals resighted at least once during primary interval  $j$  are known to be marked and available for resighting, where  $n_j^* \leq n_j$ . When sampling is with or without replacement, the zero-truncated Poisson-log normal (ZPNE) model of McClintock et al. (2008a) may be used when there are an unknown number of individually identifiable marks. McClintock et al. (2008a) used maximum likelihood to estimate the parameters. While the ZPNE model generally performed well when sampling was either with or without replacement, McClintock et al. (2008a) found that the ZPNE model performed less than nominally with simulated small-sample data generated under logit-normal resighting probabilities. Under these conditions, they found the model to slightly overestimate abundance and its variance in the absence of individual heterogeneity and to underestimate abundance and its variance in the presence of moderate to high individual heterogeneity. To address these shortcomings, we now develop a reformulation of BLNE for an unknown number of marks using data augmentation and Bayesian analysis methods.

## 2.2 Bayesian Binomial Logit-Normal Mixed Effects Model

To modify the binomial logit-normal mixed effects (BLNE) model for an unknown number of marks, we extend the data augmentation approach of Royle, Dorazio and Link (2007). Without loss of generality, we focus on a single primary interval,  $J = 1$ , such that the data consist of  $y_s = \{1, 2, \dots, k\}$  resightings across all  $k$  secondary resighting occasions for  $s = 1, \dots, n^*$  marked individuals known to be in the population (because they were observed) and the total number of unmarked individual sightings  $U$  across all  $k$  secondary occasions. If one were to augment the marked resighting data with  $M - n^*$  individuals with  $y_s = 0$ , where  $M$  is some known “super population” of marked individuals, then the problem can instead be formulated in terms of the probability  $\psi$  that any of the  $M$  marked individuals is available for resighting. This requires the introduction of a partially observed, binary latent variable  $q_s$ ,  $s = 1, \dots, M$ , which indicates whether an individual in the marked super population is available for resighting, where  $\sum_{s=1}^M q_s = n$ . For the  $n^*$  marked individuals (all with  $y_s > 0$ ), it is known that  $q_s = 1$ , but  $q_s$  is unknown for the other  $M - n^*$  individuals with  $y_s = 0$ . This approach suggests a Bayesian binomial logit-normal mixed effects (BBLNE) model, where  $[y_s | q_s, p_s] \sim \text{Binomial}(k, q_s p_s)$  for  $s = 1, \dots, M$ . This binomial specification for  $y_s$  has success probability  $q_s p_s$  because the  $M - n$  individuals that are not available for resighting have zero probability of being sighted. Thus the conditional complete-data likelihood is given by

$$p(\mathbf{y} | \mathbf{q}, \mathbf{p}, k) = \prod_{s=1}^M p(y_s | k, q_s, p_s),$$

where  $\mathbf{y} = (y_1, \dots, y_M)$  and  $\mathbf{q}$  and  $\mathbf{p}$  are defined similarly. The partially observed  $q_s$  may then be modeled as independent  $[q_s | \psi] \sim \text{Bernoulli}(\psi)$  for  $s = 1, \dots, M$ . Because in mark-resight studies essentially nothing is known a priori about resighting probabilities, below we propose relatively non-informative proper priors that constrain the parameters to be within their allowable ranges of support. The posterior distribution for the individual sighting heterogeneity model is given by  $p(\mathbf{q}, \mathbf{p}, N, \psi, \beta, \sigma^2 | \mathbf{y}, U, M, k)$ .

### 2.3 Parameter estimation for the heterogeneity and non-heterogeneity models

Since the posterior distribution for the parameters is not available in closed form, we develop a Markov chain Monte Carlo procedure (e.g., Givens and Hoeting, 2005) for parameter estimation. We develop the algorithms for both the no-heterogeneity and individual heterogeneity models below.

For the no individual sighting heterogeneity model with  $\sigma^2 = 0$  and  $p_s = p$ , a simple and efficient Gibbs sampler may be implemented using the univariate conditional densities of  $\psi$ ,  $q_s$ , and  $p$  to sample from their joint posterior distribution. We adopt Uniform(0, 1) priors on  $\psi$  and  $p$ . At the  $t$ th iteration of the Gibbs sampler, we update the parameters from the univariate conditional densities

$$[\psi^{(t)} \mid q_s^{(t-1)}] \sim \text{Beta} \left( 1 + \sum_{s=1}^M q_s^{(t-1)}, 1 + M - \sum_{s=1}^M q_s^{(t-1)} \right), \quad (1)$$

$$[q_s^{(t)} \mid y_s = 0, \mu^{(t)}] \sim \text{Bernoulli}(\mu^{(t)}),$$

and

$$[p^{(t)} \mid y_s, q_s^{(t)}] \sim \text{Beta} \left( 1 + \sum_{s=1}^M y_s q_s^{(t)}, 1 + k \sum_{s=1}^M q_s^{(t)} - \sum_{s=1}^M y_s q_s^{(t)} \right),$$

where

$$\mu^{(t)} = \text{Prob}(q_s^{(t)} = 1 \mid y_s = 0, p^{(t-1)}, \psi^{(t)}) = \frac{(1 - p^{(t-1)})^k \psi^{(t)}}{(1 - p^{(t-1)})^k \psi^{(t)} + (1 - \psi^{(t)})}.$$

For each draw  $t$  from this distribution,  $N^{(t)}$  may be derived using a modified Horvitz-Thompson (Horvitz and Thompson, 1952) estimator:

$$N^{(t)} = \frac{U}{kp^{(t)}} + \sum_{s=1}^M q_s^{(t)}. \quad (2)$$

Although deriving  $N$  in this manner renders our approach as not fully Bayesian, we chose to do so from a practical standpoint. A full Bayesian specification for  $N$  makes the univariate conditional densities for  $\beta$  and  $\mathbf{q}$  intractable and, therefore, very difficult to sample from. The Horvitz-Thompson estimator is

unbiased (Thompson, 1992), and its use here greatly simplifies the problem. In Section 4 we examine the utility and performance of this simplification as a practical means for estimating  $N$ .

The more complex individual sighting heterogeneity model with  $\sigma^2 > 0$ ,  $p_i \neq p_j$  for all  $i, j \in \{1, \dots, M\}$  requires a more complicated Metropolis-within-Gibbs hybrid MCMC algorithm. Here we adopt the logit-normal transformation where  $\log(p_s/(1-p_s)) = \theta_s$  and  $[\theta_s | \beta, \sigma^2] \sim N(\beta, \sigma^2)$ . We assume hyperpriors  $\beta \sim N(0, \tau)$  and  $\sigma^2 \sim \text{Gamma}(a, b)$ . At the  $t$ th iteration of the Gibbs sampler,  $[\beta^{(t)} | \boldsymbol{\theta}^{(t)}, (\sigma^2)^{(t-1)}, \tau]$  is given by

$$N \left( \left( \frac{M}{(\sigma^2)^{(t-1)}} + \frac{1}{\tau} \right) \frac{\sum_{s=1}^M \theta_s^{(t)}}{(\sigma^2)^{(t-1)}}, \left( \frac{M}{(\sigma^2)^{(t-1)}} + \frac{1}{\tau} \right)^{-1} \right),$$

where  $\tau$ ,  $a$ , and  $b$  are fixed constants. The conditional distribution  $[\psi^{(t)} | q_s^{(t-1)}]$  is given in (1) and  $[q_s^{(t)} | y_s = 0, \mu_s^{(t)}] \sim \text{Bernoulli}(\mu_s^{(t)})$  with

$$\mu_s^{(t)} = \text{Prob}(q_s^{(t)} = 1 | y_s = 0, p_s^{(t-1)}, \psi^{(t)}) = \frac{(1 - p_s^{(t-1)})^k \psi}{(1 - p_s^{(t-1)})^k \psi^{(t)} + (1 - \psi^{(t)})}.$$

The univariate conditional densities  $p(\sigma^2 | \boldsymbol{\theta}, \beta, a, b) \propto \prod_{s=1}^M p(\theta_s | \beta, \sigma^2) p(\sigma^2 | a, b)$  and  $p(\theta_s | y_s, q_s, \theta_s, \beta, \sigma^2) \propto p(y_s | q_s, \theta_s) p(\theta_s | \beta, \sigma^2)$  are not of standard form. For  $\theta_s^{(t)} = \log(p_s^{(t)} / (1 - p_s^{(t)}))$ , we adopt a uniform random walk Metropolis-Hastings (MH) step where the proposal distribution is given by  $\theta_s^* = \theta_s^{(t-1)} + \epsilon^*$ , with  $\epsilon^* \sim \text{Unif}(-c, c)$  where  $c$  is chosen by the user.

To facilitate acceptable mixing and convergence for  $\sigma^2$ , we use the transformation  $\xi = \log(\sigma^2)$  and propose a multiple-try MH step (Givens and Hoeting, 2005) with normal random walk proposal  $\xi^* = \xi^{(t-1)} + \epsilon^*$  and  $\epsilon^* \sim N(0, d)$  where  $d$  is chosen by the user. For the multiple-try component of the algorithm, we weight the  $r$  generated candidates by their importance sampling weight,

$$w(\sigma^{2(t-1)}, \sigma^{2*}) = p(\sigma^{2*} | \boldsymbol{\theta}, \beta, a, b) | J(\xi^*) | g(\sigma^{2(t-1)} | \sigma^{2*}) \lambda(\sigma^{2(t-1)}, \sigma^{2*}),$$

where  $J(\xi^*)$  is the Jacobian for the transformation from  $\xi$  to  $\sigma^2$ ,  $g(\sigma^{2(t-1)} | \sigma^{2*})$  is the proposal density, and the weighting function

$$\lambda(\sigma^{2(t-1)}, \sigma^{2*}) = \left[ g(\sigma^{2*} | \sigma^{2(t-1)}) g(\sigma^{2(t-1)} | \sigma^{2*}) \right]^{-1}.$$

We then select a single proposal  $\sigma_j^{2*}$  from the  $r$  candidates with probability proportional to  $w_i(\sigma^{2(t-1)}, \sigma_i^{2*})$ , sample  $r - 1$  random variables  $\sigma_i^{2**}$  ( $i = 1, \dots, r - 1$ ) from  $g(\sigma_i^{2**} | \sigma_j^{2*})$ , and set  $\sigma_r^{2**} = \sigma^{2(t-1)}$ . Thus, the generalized MH ratio for acceptance of  $\sigma_j^{2*}$  is given by

$$R_g = \frac{\sum_{i=1}^r w(\sigma^{2(t-1)}, \sigma_i^{2*})}{\sum_{i=1}^r w(\sigma_j^{2*}, \sigma_i^{2**})}.$$

For each draw from  $p(\mathbf{q}, \boldsymbol{\theta}, N, \psi, \beta, \sigma^2 | \mathbf{y}, U, M, k)$ ,  $N^{(t)}$  is again derived as in equation (2), but with  $p^{(t)} = \text{logit}^{-1}(\beta^{(t)})$ .

### 3 Example: New Zealand Robin

#### 3.1 Example Methods

The New Zealand Department of Conservation conducted a pilot study on New Zealand robin (*Petroica australis*) populations in Fiordland National Park to assess whether or not mark-resight methodology may be a suitable method for monitoring endangered populations of a closely related species inhabiting the Chatham Islands (McClintock et al., 2008b). Between September 2003 and March 2005, robins were captured and banded on two 100 ha study areas (Knobs Flat and Walker Creek). Resighting data were then collected during March 2005 on both sites. For the original analysis of these data (McClintock et al., 2008b), the number of marked birds available for resighting was specified by selecting a “marked” subset from the unknown number of banded birds in the population. This subset included all banded individuals that were seen during a survey immediately prior to the onset of resighting occasions. The unmarked data therefore could include previously banded birds not included in the marked subset. This situation is not ideal, particularly for very rare or elusive species, because every banded observation can contain important information for an analysis. Further, this method for selecting the known number of marked individuals could potentially violate an assumption of the model because the “marked” population probably consisted of individuals with higher sightabilities. This would result in underestimation of abundance. An additional complication is the potential for this selection process to result in undetected individual heterogeneity and, hence, underestimates of uncertainty as well.

In light of these potential issues with the original data, we demonstrate the use of BBLNE using the Knobs Flat study area. Between September 2003 and March 2005, 80 banded birds were released into the population. Of these, 23 banded birds were observed and thus designated as “marked” during the survey immediately prior to the collection of the resighting data. Across  $k = 7$  secondary resighting occasions (without replacement), the 23 individuals in the marked subset were resighted  $y_s = 1$  to 6 times with respective frequencies 7, 4, 5, 1, 5, and 1. For the unbanded and banded birds not included in the marked subset,  $U = 45$ . These data yield  $\hat{N} = 39$  (95% CI 34 – 45) using the BLNE model incorporating individual heterogeneity in Program MARK (White and Burnham, 1999). We suspect both the point estimate and level of uncertainty to be underestimated using this approach. We stress that the potential problem here lies not in the modeling framework itself, but in the appropriateness of the data for the model. If the number of marked individuals is unknown, the data then consist of  $n^* = 23$  individuals with the same frequencies for  $y_s$  because all members of the marked subset were resighted. Further, because none of the banded birds not included in the marked subset were resighted,  $U = 45$  for these data as well. If we then augment these data with  $M - n^* = 80 - 23 = 57$  marked individuals with  $y_s = 0$ , the data for BBLNE are completely specified.

We fit the BBLNE model under both the simple no individual heterogeneity and the more complicated heterogeneity regimes. For the no heterogeneity model, we initiated four independent random start Markov chains using the Gibbs sampler outlined in Section 2, with each chain consisting of 100,000 samples after a burn-in of 20,000. The starting values for the parameters were randomly drawn from their respective prior distributions. Mixing and convergence diagnostics included calculation of the Gelman-Rubin statistic, sample path plots, autocorrelation plots, and pair-wise correlations of parameters (Gelman, 1996). After obtaining sufficient evidence of convergence for these shorter chains, we then ran a single random start Markov chain of 200,000 samples after a burn-in of 20,000 for summarizing the posterior distribution.

We followed a similar procedure for the heterogeneity model after specifying values for the hyperparameters ( $\tau$ ,  $a$ , and  $b$ ) and random walk steps ( $c$  and  $d$ ). We found the algorithm to be sensitive to values for  $\tau$ ,  $a$ , and  $b$ . If  $\tau$  was specified too large with  $a$  and  $b$  too low, the Gibbs step for  $\beta$  too often drew extreme values that resulted in many unreasonably large values for  $N$ . We therefore fixed  $\tau = 0.25$  and  $a = b = 0.5$  to make the probability

of draws from the upper and lower extremes of  $\beta$  (and therefore  $p$ ) less likely under the prior distribution. This is a reasonable prior set-up for the robin problem as we do not expect resighting probabilities to be very near 0 or 1. To achieve adequate acceptance, mixing, and convergence rates we specified  $c = 2$  and  $d = 0.15$ , with each multiple-try MH proposal step consisting of  $r = 4$  candidates. Chain diagnostics were examined as described above, but we also examined the influence of our prior specifications on the marginal posterior distributions of  $\sigma^2$  and  $\beta$ .

To aid in determining whether the data support the hypothesis of individual heterogeneity, we compared the two BBLNE models using the Deviance Information Criterion (Spiegelhalter et al., 2002). Because all individuals have the same  $p$  when there is no heterogeneity, the standard DIC =  $2\hat{D}_{\text{ave}}(\mathbf{y}) - D_{\hat{\theta}}(\mathbf{y})$  was used for this model, where  $\hat{D}_{\text{ave}}(\mathbf{y}) = 1/L \sum_{i=1}^L D(\mathbf{y}, \theta_i)$ ,  $D_{\hat{\theta}}(\mathbf{y}) = D(\mathbf{y}, \hat{\theta})$ ,  $D(\mathbf{y}, \theta) = -2 \log p(\mathbf{y} | \theta)$ , and  $L$  is the chain length. For the heterogeneity model, an approximate DIC was calculated as DIC\* =  $2\hat{D}_{\text{ave}}(\mathbf{y}) - D_{\hat{\theta}^*}(\mathbf{y})$ , where  $D_{\hat{\theta}^*}(\mathbf{y}) = D(\mathbf{y}, \hat{\theta}^*) = -2 \log p(\mathbf{y} | \hat{\theta}^*)$ . For  $\hat{\theta}^*$ , posterior medians were used for  $\hat{N}$  and  $\hat{n}$  (because of their right-skewed distributions), and posterior means were used for  $\hat{\psi}$  and  $\hat{p}_s$ . For the latent  $q_s = 1$ ,  $\hat{n} - n^*$  individuals were randomly sampled from the  $M - n^*$  marked individuals in the super population with probability weights proportional to  $\hat{p}_s$ . This sampling process was repeated 1000 times, and the average of the DIC\* values was used for model comparisons. For comparative purposes, we also evaluated the results from the BBLNE model with greater support (based on minimum-DIC) relative to those obtained from a maximum likelihood analysis with ZPNE using Program MARK. Model selection between the heterogeneity and no heterogeneity ZPNE models was performed using Akaike's Information Criterion adjusted for small sample sizes ( $\text{AIC}_c$ ) (Burnham and Anderson, 2002), the default criterion in Program MARK. We note that DIC and  $\text{AIC}_c$  cannot be compared as a means for model selection between BBLNE and ZPNE.

## 3.2 Example Results and Discussion

For the no heterogeneity BBLNE, diagnostics of the four independent chains indicated satisfactory mixing and convergence. The univariate Gelman-Rubin statistic was less than 1.00001 for  $N$ ,  $p$ ,  $\psi$ , and  $n$ . Serial autocorrelations fell to zero within a lag of five iterations for all parameters. For the

Table 1: Comparison of BBLNE, ZPNE, and BLNE estimates (and 95% intervals) of abundance ( $N$ ), number of marked individuals ( $n$ ), and individual heterogeneity ( $\sigma^2$ ). Estimates for  $\sigma^2$  are on the logit scale for BBLNE and BLNE or on the log scale for ZPNE. Model selection criteria between parameterizations with and without individual heterogeneity were DIC for BBLNE or  $AIC_c$  for ZPNE and BLNE. BLNE assumes a known number of marked individuals ( $n = 23$ ). We note that these model selection criteria cannot be used for comparisons between BBLNE, ZPNE, and BLNE.

Model	$\hat{N}$	$\hat{n}$	$\hat{\sigma}^2$	DIC/ $AIC_c$
BBLNE No Het.	40 (37 – 46)	24 (23 – 26)	NA	101.7
BBLNE With Het.	44 (37 – 59)	26 (23 – 33)	0.99 (0.17 – 2.74)	88.6
ZPNE No Het.	42 (35 – 50)	25 (24 – 26)	NA	92.9
ZPNE With Het.	42 (34 – 51)	25 (24 – 27)	0.06 (0.01 – 0.71)	95.3
BLNE No Het.	39 (35 – 44)	NA	NA	226.9
BLNE With Het.	39 (34 – 45)	NA	0.43 (0.13 – 1.39)	226.5

heterogeneity BBLNE, the diagnostics suggested adequate mixing and convergence for the four independent chains. The Gelman-Rubin statistic was less than 1.0008 for  $N$ ,  $\beta$ ,  $\sigma^2$ ,  $\psi$ , and  $n$ . Serial autocorrelations fell off rapidly for all parameters except  $\sigma^2$ , for which autocorrelations remained around 0.3 after a lag of 100 iterations. This behavior justified a longer chain of 200,000 samples for posterior inferences.

The estimated marginal posterior distributions from the final chain of 200,000 draws for the no heterogeneity BBLNE (Figure 1) had median abundance estimate  $\hat{N} = 40$  with a 95% posterior interval (PI) of 37 – 46 (Table 1). For the number of marked individuals available for resighting, the median was  $\hat{n} = 24$  with a 95% PI of 23 – 26. Means for  $\hat{p}$  and  $\hat{\psi}$  were 0.39 (95% PI 0.31 – 0.47) and 0.30 (95% PI 0.21 – 0.41), respectively. The calculated DIC for this model was 101.7.

For the heterogeneity BBLNE, the estimated marginal posterior distributions (Figure 2) had median  $\hat{N} = 44$  (95% PI 37 – 59) and  $\hat{n} = 26$  (95% PI 23 – 33). Means for  $\hat{p} = \text{logit}^{-1}(\hat{\beta})$  and  $\hat{\psi}$  were 0.36 (95% PI 0.23 – 0.49) and 0.33 (95% PI 0.22 – 0.47), respectively. Individual heterogeneity was

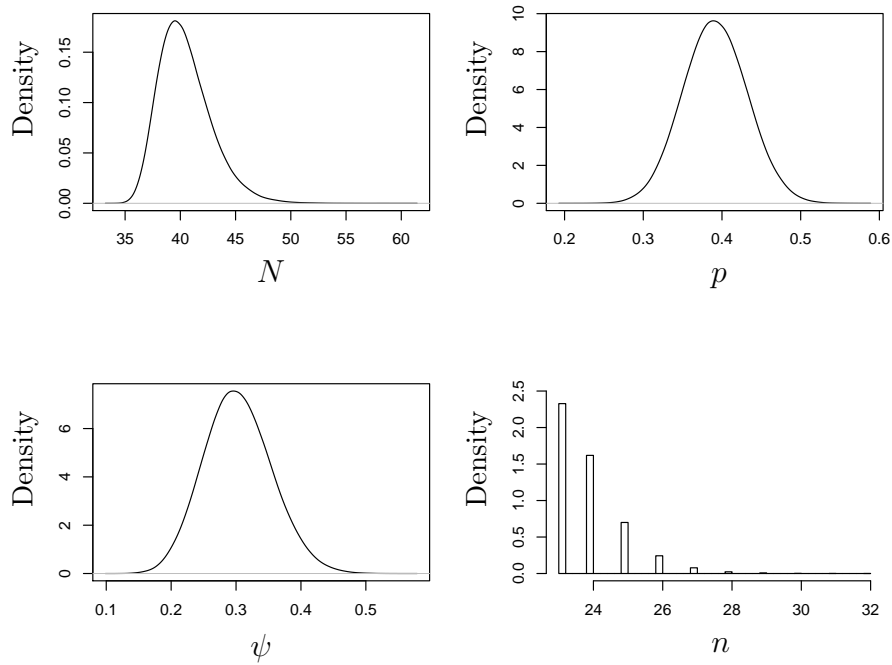


Figure 1: Estimated marginal posterior distributions for abundance ( $N$ ), resighting probability ( $p$ ), proportion of marked super population available for resighting ( $\psi$ ), and the number of marked individuals in the population available for resighting ( $n$ ) when ignoring individual heterogeneity for a New Zealand robin population in Fiordland National Park, New Zealand.

non-negligible with median  $\hat{\sigma}^2 = 0.99$  (95% PI 0.17 – 2.74). Plots of the estimated posterior densities for  $\beta$ ,  $p$ , and  $\sigma^2$  against their respective prior distributions indicate little influence of the priors despite the sensitivity of our algorithm to particularly non-informative specifications for  $p(\beta)$  and  $p(\sigma^2)$  (Figure 2). The estimated DIC = 88.6 was considerably smaller than for the no heterogeneity model. For the analysis of these data using ZPNE in Program MARK, the no heterogeneity model had  $\hat{N} = 42$  (95% CI 35 – 50), and the heterogeneity model had  $\hat{N} = 42$  (95% CI 34 – 51) (Table 1). In this case,  $AIC_c$  favored the no heterogeneity model ( $AIC_c = 92.9$ ) over the heterogeneity model ( $AIC_c = 95.3$ ).

These results provide some degree of validation to our suspicion that the original method of selecting the known “marked” subset for an analysis using BLNE (McClintock et al., 2008b) may have been inappropriate. When more realistically assuming an unknown number of marked individuals, both ZPNE and BBLNE produced larger estimates of  $N$  and its uncertainty. Based on the BBLNE model including heterogeneity, there were an estimated  $\hat{n} - n^* = 3$  marked individuals in this population that were never detected. This is indicative of the potential to overestimate resighting probabilities when selecting a marked population based on sightability, and this would result in BLNE underestimating  $N$ .

It is also important to model individual heterogeneity to avoid underestimation of  $N$  and its uncertainty (Otis et al., 1978; Neal et al., 1993). This was demonstrated by the relatively large differences between the two BBLNE models for these data, from which we found strong evidence of individual heterogeneity. Comparing the DIC-supported BBLNE heterogeneity model with the  $AIC_c$ -supported ZPNE model excluding heterogeneity, the point estimate for  $N$  for the latter was lower with considerably greater precision. This is consistent with our suspicion that ZPNE may underestimate  $N$  and its uncertainty when used for relatively small data sets with heterogeneous resighting probabilities when sampling is without replacement. However, because this single example provides no information on the expected relative performance of BBLNE relative to ZPNE under these conditions, we report on a series of simulation experiments in the next section.

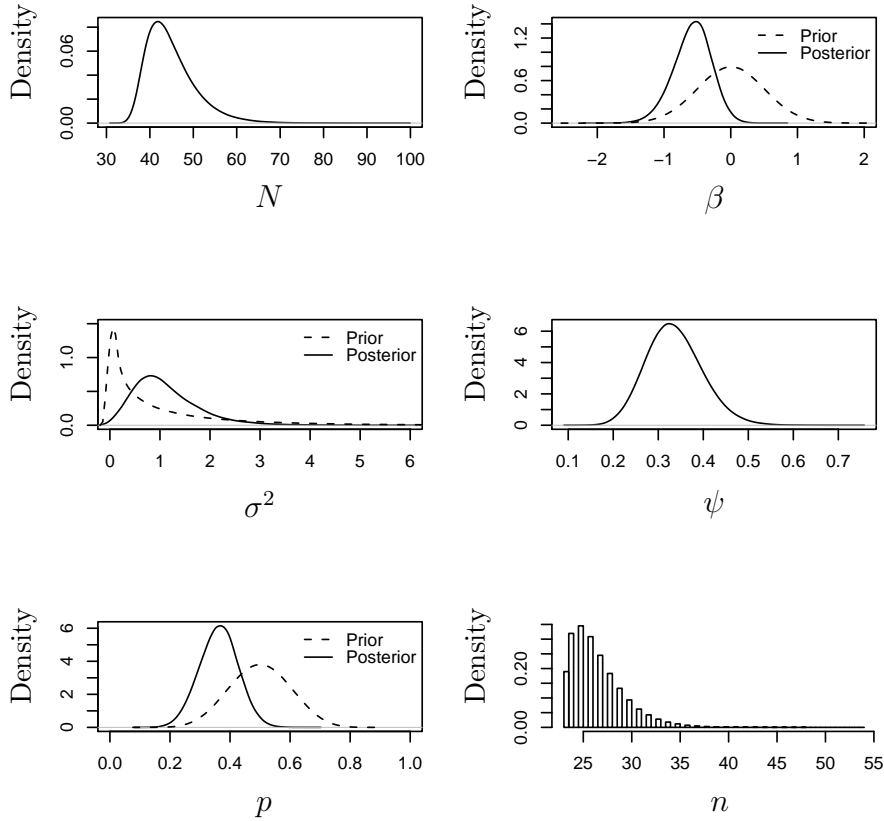


Figure 2: Estimated marginal posterior distributions for abundance ( $N$ ), resighting probability on the logit scale ( $\beta$ ), individual heterogeneity ( $\sigma^2$ ), proportion of marked super population available for resighting ( $\psi$ ), resighting probability ( $p$ ), and the number of marked individuals in the population available for resighting ( $n$ ) when including individual heterogeneity for a New Zealand robin population in Fiordland National Park, New Zealand. Prior distributions are included for  $\beta$ ,  $\sigma^2$ , and  $p$ .

## 4 Simulation Experiments

### 4.1 Simulation Methods

We performed three series of simulation experiments, one with generated “small” sample data ( $n = 25, N = 40, p = 0.30$ ) emulating that of the robin example, one with “medium” sample data ( $n = 50, N = 250, p = 0.40$ ), and one with “large” sample data ( $n = 75, N = 500, p = 0.85$ ). The number of resighting occasions,  $k$ , was set to 7 for all experiments. For each series, three sighting probability scenarios were simulated: 1) no individual heterogeneity ( $\sigma^2 = 0$ ); 2) moderate individual heterogeneity ( $\sigma^2 = 0.75$ ); and 3) high individual heterogeneity ( $\sigma^2 = 1.75$ ). For scenarios without heterogeneity, sighting data were generated as realizations of binomial random variables, where  $y_s \sim \text{Binomial}(k, p)$  ( $s = 1, \dots, N$ ). For the scenarios with heterogeneity, the data were generated as  $y_s \sim \text{Binomial}(k, p_s)$  ( $s = 1, \dots, N$ ), where  $p_s = \text{logit}^{-1}(\theta_s)$ ,  $\theta_s \sim N(\beta, \sigma^2)$ , and  $\beta = \log(p/(1-p))$ . For all data sets,  $U = \sum_{s=n+1}^N y_s$ . With  $\sigma^2 = 0.75$  or  $1.75$ , these input parameters exhibit moderate to severe individual heterogeneity on the inverse logit scale. For each population within the three sample size classes, we generated 500 simulated data sets with an unknown number of marked individuals by excluding the unobservable  $y_s = 0$  from the data.

We analyzed each simulated data set using both BBLNE and ZPNE. For BBLNE, each data set was augmented with  $M - n^*$  marked individuals with  $y_s = 0$ , where  $M = 50$  for the small sample data,  $M = 80$  for the medium sample data, and  $M = 125$  for the large sample data. For the data without individual heterogeneity, the no heterogeneity parameterizations of BBLNE and ZPNE were used. For the data with heterogeneity, the heterogeneity parameterizations of BBLNE and ZPNE were used. For BBLNE, we used the same prior specifications and MCMC algorithm as in the robin example. For each simulated data set, a chain of 100,000 samples was drawn (after a burn-in of 20,000), and the median, standard deviation (as an estimate of SE), and 95% posterior interval for  $N$  were calculated from its estimated marginal posterior distribution. For ZPNE, maximum likelihood estimates of  $N$  were obtained using the NLMIXED procedure in SAS 9.1 (SAS Institute, 2002). We compared the average performance of BBLNE and ZPNE as estimators of  $N$  based primarily on bias/SE =  $E(\hat{N} - N)/SE(\hat{N})$ , percent posterior (BBLNE) or confidence (ZPNE) interval coverage of  $N$ , percent posterior or confidence interval length %IL =  $100(\text{Upper} - \text{Lower})/N$ ,

root mean squared error  $\left(\text{RMSE} = \sqrt{\text{E}(\hat{N} - N)^2 + \text{var}(\hat{N})}\right)$ , and the estimated ratio of the theoretical  $\text{SE} = 1/500 \sum_{i=1}^{500} \hat{\text{SE}}(\hat{N}_i)$  to the empirical  $\text{SE} = \sqrt{1/499 \sum_{i=1}^{500} (\hat{N}_i - \bar{N})^2}$  (Burnham et al., 1987).

## 4.2 Simulation Results and Discussion

As expected, we generally found BBLNE to outperform ZPNE with small sample sizes generated from a binomial logit-normal distribution (Table 2). Despite having very similar percent coverages of  $N$ , BBLNE tended to be less positively biased and more precise than ZPNE when there was no heterogeneity in the data. In this case, BBLNE tended to slightly underestimate the empirical variance (SE ratio = 0.9) while ZPNE tended to overestimate this quantity (SE ratio = 1.1). As expected, ZPNE exhibited a negative bias as an estimator of  $N$  and its uncertainty when heterogeneity was present in the data, particularly when the heterogeneity was extreme. Confidence intervals for ZPNE only included  $N$  in 61% of these simulations. Although better than for ZPNE, coverage for BBLNE was slightly less than the nominal 95% with small sample sizes because it tended to underestimate the uncertainty about  $N$  or exhibited a small negative bias in the presence of individual heterogeneity.

With the medium and large sample data when there was no individual heterogeneity, BBLNE was generally consistent and exhibited greater efficiency than ZPNE (Table 2). With medium sample sizes and moderate heterogeneity, BBLNE and ZPNE performed similarly, with ZPNE exhibiting a slight negative bias with greater efficiency compared to BBLNE. This tendency was reversed with large sample sizes and moderate to high heterogeneity, when BBLNE exhibited a larger bias and greater precision than ZPNE. For these scenarios, coverage for BBLNE was near nominal, and coverage for ZPNE was exceptionally high because it tended to overestimate the uncertainty about  $N$ . However, with large sample data and severe heterogeneity, both models performed quite similarly overall.

Surprisingly, ZPNE performed very similarly for all of the large sample simulations. Examination of the estimated heterogeneity parameter for ZPNE indicates that very little Poisson overdispersion was detectable in these data. In other words, despite moderate to severe heterogeneity on the inverse logit scale, these large sample zero-truncated binomial logit-normal

Table 2: Comparison of BBLNE and ZPNE average bias divided by standard error (Bias/SE), percent interval coverage (Cover), percent interval length (% IL), root mean squared error (RMSE), and the theoretical to empirical SE ratio for abundance estimates from a series of simulation experiments consisting of small, medium, and large sample sizes. Each sample size series included a scenario without individual heterogeneity, with moderate individual heterogeneity, and with high individual heterogeneity.

	No Ind. Het.		Mod. Ind. Het.		High Ind. Het.	
	BBLNE	ZPNE	BBLNE	ZPNE	BBLNE	ZPNE
<b>Small Sample</b>						
Bias/SE	0.2	0.6	-0.7	-0.5	-0.9	-1.5
Cover	89	89	89	82	88	61
% IL	37	47	41	38	44	34
RMSE	3.9	6.1	5.2	4.7	6.2	6.6
SE Ratio	0.9	1.1	1.0	0.9	1.0	0.8
<b>Medium Sample</b>						
Bias/SE	0.1	0.5	0.1	-0.3	0.0	-1.1
Cover	95	98	94	92	97	74
% IL	24	34	45	31	57	29
RMSE	15.2	25.8	29.3	21.7	36.8	29.7
SE Ratio	0.9	1.3	1.2	1.0	1.3	1.0
<b>Large Sample</b>						
Bias/SE	0.0	0.0	-0.2	0.1	-0.4	-0.1
Cover	94	100	94	100	92	100
% IL	6	17	9	17	14	18
RMSE	7.8	22.1	11.6	22.6	20.8	22.9
SE Ratio	1.0	2.7	1.0	1.9	1.1	1.5

data generally fit the standard zero-truncated Poisson distribution with no overdispersion. (McClintock et al., 2008a) also found ZPNE to be robust to data generated from a binomial logit-normal distribution except when sample sizes were particularly small with heterogeneous resighting probabilities. In addition, we found ZPNE to perform poorly with moderate sample sizes under severe heterogeneity. Relative to BBLNE, it appears the potential for bias and inefficiency using the zero-truncated Poisson distribution to model small- to medium-sized binomial logit-normal data may be overcome with large sample (and more Poisson-like) data because ZPNE does not need to invest in additional parameters to estimate the number of marked individuals.

## 5 Conclusions

When sampling is without replacement, we found BBLNE to have advantages over ZPNE when sample sizes are smaller and individual heterogeneity is more severe. For simplicity, we focused on modeling the resighting data as binomial logit-normal by ignoring temporal variation between secondary sampling occasions. As with the most general logit-normal model of McClintock et al. (2008b), BBLNE may be relatively easily extended to allow variation in resighting probabilities between secondary occasions and  $J > 1$  primary intervals by modeling  $\delta_{sij} \sim \text{Bernoulli}(p_{sij})$ , where  $p_{sij} = \text{logit}^{-1}(\theta_{sij})$  and  $\theta_{sij} \sim \text{N}(\beta_{ij}, \sigma_j^2)$  ( $s = 1, \dots, n_j^*$ ,  $i = 1, \dots, k_j$ ,  $j = 1, \dots, J$ ). The univariate conditional densities for all parameters could then be adjusted accordingly. Although not explored here, our framework also allows the incorporation of relevant covariate information about the resighting probability parameters ( $\beta$  and  $\sigma^2$ ) to more efficiently describe the detection process.

Despite our algorithm's sensitivity to particularly non-informative priors on  $\beta$  and  $\sigma^2$ , we found suitable specifications with little influence on posterior inferences. This was further demonstrated in our large sample simulations, where the prior distributions had very little mass in the region of the actual parameter values. However, there remain several potential refinements to our methods that we can identify. Although not explored here, a more sophisticated approach may involve specifying a joint prior on  $\beta$  and  $\sigma^2$  (and possibly  $\tau$ ). We envision such a prior reducing the negative autocorrelation induced between  $\tau$  and  $\sigma^2$  via the univariate conditional distribution for  $\beta$ . As an alternative to our derived Horvitz-Thompson estimator for  $N$ , a complete

Bayesian specification for the model could be explored further, but we found our approach to perform quite well as a practical solution. Instead of using DIC, the different parameterizations for BBLNE could be implemented using reversible jump MCMC methods (Givens and Hoeting, 2005) as a means of model selection and multimodel inference. However, this approach quickly becomes a technical challenge and beyond the intended scope of this work. Another area not explored here is the potential to implement our model in a maximum likelihood framework using an E-M algorithm (Givens and Hoeting, 2005).

The New Zealand robin data we analyzed using BBLNE were originally used as an illustrative example in McClintock et al. (2008b) with little concern for the validity of these data for their model. However, had that analysis been intended to inform management or policy decisions, we believe it generally better to admit and model uncertainty than to make risky assumptions. Prior to the recent introduction of ZPNE, such assumptions were generally necessary to estimate abundance from sighting data with an unknown number of marked individuals. One advantage of ZPNE is that it can be used when sampling is with or without replacement, and this distinction can sometimes be difficult to control (via design) or ascertain in the field. Another advantage is that ZPNE is easily implemented using the freeware package Program MARK (White and Burnham, 1999). However, with BBLNE, a model is now available that can be more consistent and efficient than ZPNE when sampling is without replacement, particularly with small sample sizes exhibiting heterogeneous resighting probabilities. Due to limited budgets and the inherent variability between individuals, this is a common occurrence in mark-resight population studies. For practitioners familiar with the software, code to implement BBLNE in both WinBUGS and R is available at <http://www.stat.colostate.edu/~jah/software>.

## 6 Acknowledgements

We thank M. A. Pryde and the New Zealand Department of Conservation for use of their data. Funding for BTM was provided by the Colorado State University Program for Interdisciplinary Mathematics, Ecology, and Statistics (PRIMES, NSF-IGERT Grant No. DGE-0221595).

## References

- Arnason, A.N., Schwarz, C.J. and Gerrard, J.M. (1991) Estimating closed population size and number of marked animals from sighting data. *Journal of Wildlife Management*, 55, 716–730.
- Burnham, K.P. and Anderson, D.R. (2002) *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. 2nd edition. Springer-Verlag, New York, USA.
- Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C. and Pollock, K.H. (1987) *Design and Analysis Methods for Fish Survival Experiments Based on Release-Recapture*. American Fisheries Society Monograph 5, Bethesda, Maryland, USA.
- Gelman, A. (1996) Inference and monitoring convergence. In *Markov Chain Monte Carlo In Practice*, W.R. Gilks, S. Richardson and D.J. Spiegelhalter, eds. Chapman and Hall/CRC, Boca Raton, Florida, USA, pp. 131–143.
- Givens, G.H. and Hoeting, J.A. (2005) *Computational Statistics*. Wiley, Hoboken, New Jersey, USA.
- Horvitz, D.G. and Thompson, D.J. (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 77, 663–685.
- Kendall, W.L., Pollock, K.H. and Brownie, C. (1995) A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics*, 51, 293–308.
- McClintock, B.T., White, G.C., Antolin, M.F. and Tripp, D. (2008a) Estimating abundance using mark-resight when sampling is with replacement or the number of marked individuals is unknown. *Biometrics*, in press.
- McClintock, B.T., White, G.C., Burnham, K.P. and Pryde, M.A. (2008b) A generalized mixed effects model of abundance for mark-resight data when sampling is without replacement. *Environmental and Ecological Statistics*, in press.
- Minta, S. and Mangel, M. (1989) A simple population estimate based on simulation for capture-recapture and capture-resight data. *Ecology*, 70, 1738–1751.

- Neal, A.K., White, G.C., Gill, R.B., Reed, D.F. and Olterman, J.H. (1993) Evaluation of mark-resight model assumptions for estimating mountain sheep numbers. *Journal of Wildlife Management*, 57, 436–450.
- Otis, D.L., Burnham, K.P., White, G.C. and Anderson, D.R. (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, 62.
- Royle, J.A., Dorazio, R.M. and Link, W.A. (2007) Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, 16, 67–85.
- SAS Institute (2002) SAS OnlineDoc, Ver. 9. SAS Institute, Cary, North Carolina, USA.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B*, 64, 1–34.
- Thompson, S.K. (1992) *Sampling*. Wiley-Interscience, New York, New York, USA.
- White, G.C. and Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked individuals. *Bird Study*, 46, 120–139.
- White, G.C. and Shenk, T.M. (2001) Population estimation with radio-marked animals. In *Radio Tracking and Animal Populations*, J. Millspaugh and J.M. Marzluff, eds. Academic Press, San Diego, California, USA, pp. 329–350.