Mark–Recapture With Occasion and Individual Effects: Abundance Estimation Through Bayesian Model Selection in a Fixed Dimensional Parameter Space

John W. DURBAN and David A. ELSTON

We present a Bayesian mark–recapture method for explicitly communicating uncertainty about the size of a closed population where capture probabilities vary across both individuals and sampling occasions. Heterogeneity is modeled hierarchically using a continuous logistic-Normal model to specify the capture probabilities for both individuals that are captured on at least one occasion and individuals that are never captured and so remain undetected. Inference about how many undetected individuals to include in the model is accomplished through a Bayesian model selection procedure using MCMC, applied to a product space of possible models for different numbers of undetected individuals. Setting the estimation problem in a fixed dimensional parameter space enables the model selection procedure to be performed using the freely available WinBUGS software. The outcome of inference is a full "posterior" probability distribution for the population size parameter. We demonstrate this method through an example involving real mark–recapture data.

Key Words: Capture–recapture; Gibbs Sampling; Markov chain Monte Carlo; Rasch model.

1. INTRODUCTION

Effective conservation and management of animal populations requires knowledge of population size. However, a common problem in the study of natural populations is that not all the animals are detected during surveys, and so complete enumeration of a population is not possible. This has led to extensive development of methods for sampling populations, and estimating population size from sample data using statistical models (Schwarz and Seber 1999; Buckland, Goudie, and Borchers 2000). Mark–recapture is a widely used method in which samples of individuals are repeatedly captured from the target population, noting the

John W. Durban is Research Associate, National Research Council, hosted by the NOAA Fisheries, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115 (E-mail: John.Durban@noaa.gov). David A. Elston is Acting Director, Biomathematics and Statistics Scotland, The Macaulay Institute, Aberdeen, AB15 8QH, UK (E-mail: d.elston@bioss.ac.uk).

©2005 American Statistical Association and the International Biometric Society
Journal of Agricultural, Biological, and Environmental Statistics, Volume 10, Number 3, Pages 291–305
DOI: 10.1198/108571105X58630

presence and absence of individuals distinguished through artificial or natural markings. The aim of the analysis is to estimate the number of individuals in the population that are never captured, and thereby obtain an estimate of the population size N. Accurate estimation of N requires models that adequately describe the major sources of variation in the capture probabilities (Pollock, Nichols, Brownie, and Hines 1990).

A standard assumption of traditional mark–recapture analysis is that capture probabilities are equal for all the animals in the population (Seber 1982). However, it has long been recognized that behavioral differences among individuals are likely to produce variable probabilities of capture, and failure to account for this heterogeneity can lead to negatively biased population estimates (e.g., Carothers 1973a,b; Otis, Burnham, White, and Anderson 1978). Therefore, estimators that model individual variation in capture probabilities are generally required. One approach is to use individual covariates to describe individual heterogeneity (e.g., Huggins 1989; Alho 1990). However, relevant covariate information may not always be available for all captured individuals, and is certainly not available for uncaptured individuals. As an alternative, several models have now been developed that are based on an assumed distributional form for the individual capture probabilities (e.g., Burnham and Overton 1978; Norris and Pollock 1996; Dorazio and Royle 2003), with some also enabling heterogeneous capture probabilities to vary across sampling occasions (e.g., Chao, Lee, and Jeng 1992; Agresti 1994; Coull and Agresti 1999; Fienberg, Johnson, and Junker 1999; Pledger 2000).

These distributional models can be categorized by the form they assume for the individual detection probabilities. One category employs latent class or finite mixture models, where heterogeneity is modeled as a finite mixture distribution, partitioned into two or more classes with homogeneous capture probabilities (Agresti 1994; Norris and Pollock 1996; Coull and Agresti 1999; Pledger 2000). These finite mixture models will be most appropriate for populations that contain homogeneous groups, each with its own probability of capture. If such groups do not exist, then it may be best to assume a different probability of capture for each individual (Dorazio and Royle 2003). Models of this type constitute a second category that can be seen as a special case of the latent class finite-mixture models, with as many classes in the population as there are individuals. Chao et al. (1992) proposed using an approach based on sample coverage to estimate the number of classes (individuals) from the mean and coefficient of variation of capture rates. An alternative approach specifies individual variation in capture probabilities hierarchically using continuous distributions on the latent detection probabilities, such as a beta distribution (Burnham and Overton 1978; Dorazio and Royle 2003) or a Normal distribution on their logits (Coull and Agresti 1999; Fienberg et al. 1999).

Modeling individual heterogeneity hierarchically using standard parametric distributions offers the practical advantage that the number of parameters needed to specify heterogeneity in individual rates of capture does not increase with N, thus helping to ensure that model parameters are well identified (Dorazio and Royle 2003). However, estimation of N under these models can be problematic. Specifically, substantial heterogeneity and small probabilities of detection can make reaching useful conclusions difficult, due to large

uncertainty associated with parameter estimates (e.g., Coull and Agresti 1999).

Conventional approaches for inference from mark–recapture models are not best suited for analyzing and communicating uncertainty. These methods are typically oriented to find a single optimum estimate, such as the maximum likelihood estimate, with statements of uncertainty generally being derived from assumptions about large sample Normality (Buckland, Anderson, Burnham, and Laake 2000). Inference of this type is not best suited for communicating uncertainty to both technical and nontechnical users of population data (Wade 2000). What is therefore needed is a method for parameter estimation that allows for more complete calculation and communication of uncertainty. We follow Fienberg et al. (1999) in adopting a Bayesian method for estimating N when fitting a continuous logistic-Normal model to heterogeneous mark–recapture data. The Bayesian method yields a full probability distribution for the number of animals, rather than point estimates with associated errors, communicating both extent and shape of the associated uncertainty. Model fitting is accomplished using computer-intensive Markov chain Monte Carlo (MCMC) simulation methods (Brooks 1998).

MCMC methods that involve sampling from, rather than maximizing, the likelihood have had a liberating effect on Bayesian data analysis, greatly increasing the range of models that can be fitted in a Bayesian framework for a wide range of applications (Gilks, Richardson, and Spiegelhalter 1996). However, Fienberg et al. (1999) noted that mark–recapture models with individual effects present a problem to the application of standard MCMC algorithms, because the changing number of unobserved individuals leads naturally to arrays of variable dimension whereas standard software requires a constant dimensional parameter space (Brooks 1998). We show how to overcome this problem by data augmentation to specify the correct likelihood using arrays of fixed dimension. The decision of how many unobserved individuals to include is thus cast into a model selection problem. The fixed dimensional parameter space allows model fitting through standard MCMC approaches (Casella and George 1992; Carlin and Chib 1995), which can be implemented using the freely available WinBUGS software (Lunn, Thomas, Best, and Spiegelhalter 2000). We demonstrate this method through application to a published mark–recapture dataset for snowshoe hares (*Lepus americanus*), reported by Cormack (1989). These data were also used by Coull and Agresti (1999) when evaluating the performance of the logistic-normal Rasch model in a maximum likelihood framework, and therefore allows a comparison to the performance of our Bayesian approach.

2. MARK–RECAPTURE DATA MATRIX

Mark–recapture data can be summarized as a matrix $\mathbf{X}$ with elements x_{ij} taking the values 1 or 0 depending whether or not individual i is "captured" in the jth sample. These "captures" may involve any type of detection or sighting by which individuals can be reliably distinguished. It is assumed that the duration of the sampling is sufficiently short that individuals are unlikely to enter or leave the population through births, deaths, and permanent migration. The matrix $\mathbf{X}$ has its number of columns, J, fixed by the number of

samples drawn from the population and its number of rows set by the number of observed individuals, n. In the hare study, $n = 68$ different hares were captured over $J = 6$ consecutive trapping days. The numbers of times each individual was captured (i.e., the row-sums of $\mathbf{X}$) ranged from 1 to 6 (median = 2, mode 1) and the numbers of individuals captured on each occasion (the column-sums of $\mathbf{X}$) ranged from 16 to 32 (median = 25) (Cormack 1989). Clearly, there is scope for individuals to remain undetected in such mark–recapture studies, and to accommodate this Fienberg et al. (1999) augmented $\mathbf{X}$ to include $N - n$ rows of zeros corresponding to the unobserved individuals in a population of size N.

3. MODEL STRUCTURE

The model we employ is based on the Rasch model from the field of educational testing (Rasch 1960), which was introduced to the context of population estimation by Darroch, Fienberg, Glonek, and Junker (1993) and Agresti (1994). This type of model has previously been fitted to the snowshoe hare data, using maximum likelihood approaches (Coull and Agresti 1999). The model is based on the assumption that, conditional on capture probabilities p_{ij}, the observations x_{ij} constitute the outcome of independent Bernoulli trials. The likelihood of the data matrix given by Fienberg et al. (1999) is

$$L(X) = \binom{N}{n} \prod_{i=1}^{N} \prod_{j=1}^{J} p_{ij}^{x_{ij}} \left(1 - p_{ij}\right)^{1-x_{ij}} . \tag{3.1}$$

We only observe data for n rows of the underlying data matrix, where n is the number of individuals that are captured at least once. The remaining $N - n$ rows are all unobserved vectors of 0's. The aim of the modeling is to estimate how many rows of 0's there should be, and therefore estimate N. The n detected individuals are seen as representing a subset from the population of size N, and therefore N needs to be included as an unknown parameter in the likelihood. Specifically, the model includes a factorial term to specify the $\binom{N}{n}$ possible number of ways of partitioning N individuals in the population into n that are detected and $N - n$ that are not.

The Rasch model that we employ assumes that the probabilities of capture vary both across survey occasions and among individuals. These two effects are likely to be particularly important in mark–recapture studies where it is not possible to completely control the coverage of the capture samples, due to individual heterogeneity and perhaps variation in the capture effort across surveys. For the hare example, the requirement for both survey and individual effects is indicated by the variability in the numbers of times each individual was captured and in the numbers of individuals captured on each occasion. Our model construction follows that of Coull and Agresti (1999), based on a logit-linear model for observation probabilities

$$\log\{p_{ij}/(1 - p_{ij})\} = \mu + \theta_i + \alpha_j \tag{3.2}$$

for an overall constant μ, occasion effects $\alpha_j, j = 1 \ldots J$, and individual effects $\theta_i, i =$

$1 \ldots N$. Combining this model to that for the recapture data gives the following likelihood:

$$L\left(X \mid N, \mu, \alpha, \theta\right) \binom{N}{n} \cdot \prod_{i=1}^{N} \prod_{j=1}^{J} \left[\frac{\exp(\mu + \theta_i + \alpha_j)}{1 + \exp(\mu + \theta_i + \alpha_j)}\right]^{x_{ij}} \left[\frac{1}{1 + \exp(\mu + \theta_i + \alpha_j)}\right]^{1 - x_{ij}}. \tag{3.3}$$

This likelihood has four sets of parameters: $\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\theta}$ and N. To allow for unobserved animals, the individual effects, θ_i, have to be modeled as random effects, which we assume are drawn from a Normal distribution with zero mean and variance σ_θ^2. We also use a Normal random effects model for the occasion effects, α_j, with variance σ_α^2, although it is also possible to treat these effects as fixed. We follow Fienberg et al. (1999) in adopting a Bayesian framework for this model, where the overall mean level of capture probability, μ, was assigned a vague Normal prior centred at zero with large variance (=10). To allow nonzero effects to emerge, the variance hyperparameters for the occasion and individual effects, σ_α^2 and σ_θ^2, respectively, were assigned separate inverse gamma prior densities of the form Gamma$^{-1}(a, b)$, where $a = v_0/2$ and $b = v_0\sigma_0^2/2$. These Gamma densities have an average, σ_0^2, and the degree of strength of the prior beliefs is contained in the degrees of freedom parameter v_0. This conveniently allowed us to vary the degrees of freedom to assess sensitivity of the abundance estimate to these prior assumptions. To ensure that the prior for N is flat across the probable range of values, we adopted a truncated Uniform prior, in which the prior distribution for N is proportional to 1, with the lower bound set to be the number of individuals observed (n), and the upper bound is based on prior knowledge of the population (e.g., Fienberg et al. 1999). As an alternative, a truncated Jeffreys' prior (Jeffreys 1946) was obtained simply by formulating both priors for N as proportional to N^{-c}, with the constant c set to 1 for the Jeffreys prior (hence N was proportional to $1/N$) and 0 for the uniform prior (hence N was proportional to 1).

Inference from this Bayesian model can be based on the joint posterior distribution, and inference on single parameters, such as population size, can then be obtained by integrating this joint distribution over all other parameters to obtain the marginal posterior distribution of interest (Gelman, Carlin, Stern, and Rubin 1995). However, for the current model, a closed form analysis of the posterior distribution is intractable (Fienberg et al. 1999) and we use MCMC sampling to accomplish this integration by simulation (Brooks 1998). There are a variety of MCMC methods for sampling from posterior densities. The most straightforward of these is the Gibbs sampler (Casella and George 1992), where values are sequentially generated directly from the full conditional distribution of each parameter given the current value of the other parameters and the data. However, this model presents a problem in the direct application of this kind of MCMC sampling (Fienberg et al. 1999). This difficulty arises because the dimension of the parameter vector of individual effects $\boldsymbol{\theta} = (\theta_1, \ldots \theta_N)$ is determined by another unknown parameter: the number of individuals, N. As a result, although we can condition on $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ when estimating N, we cannot condition on $\boldsymbol{\theta}$. This prohibits the use of conventional Gibbs sampling from full conditional distributions, which requires a constant dimensional parameter space.

4. ABUNDANCE ESTIMATION THROUGH MODEL SELECTION

We have overcome the problem of variable dimensionality by fixing the maximum possible number, M, of unseen individuals. Corresponding changes of fixed dimension must be made, both to the matrix of capture histories (augmenting the observed data by M rows of zeros: $x_{ij} = 0$, for $i = n+1, \ldots, n+M$ and $j = 1, \ldots, J$) and to the vector of individual effects (augmenting the values relating to observed individuals by M values: $\theta_{n+1}, \ldots, \theta_{n+M}$).

We cast the decision of how many, n_0, of the M potentially undetected individuals to include into a model selection problem, with n_0 functioning as a model indicator variable, indicating which capture history data $x_{ij}, i = n+1, \ldots, n+n_0$ should actively augment the data $x_{ij}, i = 1, \ldots, n$, from observed individuals. This formulation is equivalent to the product space approach of Carlin and Chib (1995), with a product space being defined in reference to the pool of all M possible individual parameters for undetected individuals with individual specific parameters $\theta_i = (\theta_{n+1}, \ldots, \theta_{n+M})$, and with the additional indexing variable n_0. Each value for n_0 can then be viewed as a candidate model, which can be described in terms of how many of this pool of M possible individual-specific parameters are included in the model.

Specification of a particular value of n_0 implies no contribution to the likelihood should be made by rows i of the data matrix with $i > n+n_0$. This can be achieved trivially by treating the additional elements of the data matrix as arising from independent Bernoulli trials with $p_{ij} = 0$ so that $x_{ij} = 0$ with probability one for $i = n+n_0+1, \ldots, n+M$ and $j = 1, \ldots, J$. Similarly, although the parent distribution for $\theta_i, i = n+1, \ldots, n+n_0$, is the same as for $\theta_i, i = 1, \ldots, n$, this should not be the parent distribution for $\theta_i, i = n+n_0+1, \ldots, n+M$.

Instead, the approach of Carlin and Chib (1995) for sampling from a product space can be implemented, involving updating moves of two alternate types: sampling from the model indicator, n_0, and then updating the value of the model parameters conditional on the value of the indicator. Under this scheme, individual effects $\theta_i, i = 1, \ldots, n+n_0$, are sampled from their posterior conditional on their inclusion in the model, whereas individual effects $\theta_i\, i = n+n_0+1, \ldots, M$ are sampled from their "pseudo-priors." These pseudo-priors can be thought of as "linking densities," functioning to maintain fixed dimensionality of the product space over all candidate models and their parameters (Carlin and Chib 1995). Gibbs sampling of this kind has been applied to other Bayesian variable selection problems (e.g. Geweke 1996; Kuo and Mallick 1998).

The choice of pseudo-priors is essential for efficient operation of this scheme, and to ensure regular between-model moves. Carlin and Chib (1995) suggested the use of pseudo-priors that are close to the conditional posterior if the parameter were in the model to facilitate switches between models during MCMC sampling. To achieve this, we have found it effective to conduct a pilot run to estimate the posterior distribution of θ_{n+1}. Provided $n_0 > 1$ throughout the pilot run, the sampled values of θ_{n+1} can be used to provide a good estimate of the posterior distribution of the individual-specific parameter for unobserved individuals. To show that choice of the pseudo-priors affects only the computational efficiency, we performed separate sampling runs from the joint model space with different degrees of

variance inflation for the pseudo-priors used in each run.

Prior information in the model indicator n_0 can be specified through the prior distribution on N, as these parameters are linked through the logical relationship $n_0 = N - n$. This formulation has the effect of truncating the prior for N, making it a discrete random variable with support over the interval between n and $n + M$. This is an intuitively logical set-up, because the posterior distribution for N is bounded to be greater than the number, n, of individuals that were actually observed. Additionally, the upper bound, $n+M$, can be specified based on prior information, to avoid an unrealistically large number of candidate models in the product space, and facilitate faster and more efficient computation.

5. APPLICATION TO REAL MARK–RECAPTURE DATA

Based on the summary data presented by Cormack (1989), a binary capture history was compiled for each of the $n = 68$ different snowshoe hares that were captured over $J = 6$ consecutive trapping days, consisting of a row of 1's and 0's to indicate whether or not the individual was captured on a particular occasion. This served as the data to which we fit the Rasch using the WinBUGS software (Lunn et al. 2000). Initially, the occasion and individual effect hyperparameters, σ^2_α and σ^2_θ, respectively, were each assigned Gamma densities with prior averages of $\sigma^2_{0_\alpha}$ and $\sigma^2_{0_\theta} = 1$, but with prior degrees of freedom v_{0_α} and $v_{0_\theta} = 1$ to specify relatively uninformative priors. We performed a goodness-of-fit check of this model for these observed data using a posterior predictive approach (e.g., Gelman, Meng, and Stern 1996). Replicates were drawn from the posterior predictive distribution $p(X^{\text{rep}} \mid X, \mu, \theta, \alpha)$, which was predicted based on posterior estimates of model parameters, and a discrepancy measure D was then calculated for both the predicted X^{rep} and observed data X:

$$\begin{aligned} D(X^{\text{rep}}) &= \Sigma\Sigma \left\{ x^{\text{rep}}_{ij} \times \log(p_{ij}) + (1 - x^{\text{rep}}_{ij}) \times \log(1 - p_{ij}) \right\} \\ D(X) &= \Sigma\Sigma \left\{ x_{ij} \times \log\left(p_{ij}\right) + (1 - x_{ij}) \times \log(1 - p_{ij}) \right\}. \end{aligned} \tag{5.1}$$

There was good agreement between the posterior predictive distribution of $D(X^{\text{rep}})$ and the posterior distribution of $D(X)$ (Figure 1). The exceeding tail area probability, called the posterior predictive p value, was estimated as the proportion of draws above the diagonal line in Figure 1 and took the value .58. Values close to .5 indicate that the realized discrepancy of the data is consistent with what one might expect from replications under the model (Gelman et al. 1996).

The WinBUGS implementation of the mark–recapture model with unobserved individuals was straightforward to set up with the exception of the factorial term $\binom{N}{n}$ in Equation (3.1), which is not included in the likelihood implied by declaring the capture histories, x_{ij}, as the outcome of independent Bernoulli trials. Instead, the factorial term was introduced by another data augmentation step, this time of observing n successes out of N independent trials each with probability $\delta = 10^{-6}$. Choice of δ is arbitrary, so long as it is sufficiently small that the implied likelihood contribution is effectively proportional to $\binom{N}{n}$ over the likely variation of N.

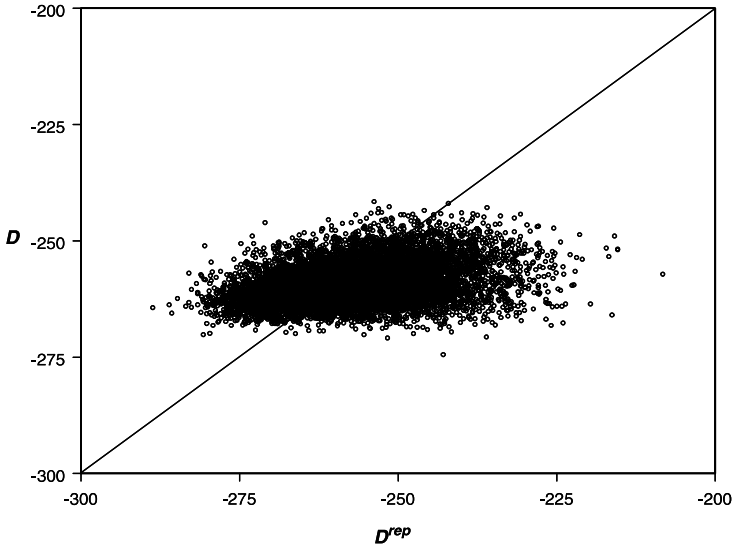

Figure 1. Scatterplot of 10,000 draws from the predictive posterior distribution of the discrepancy measure for the predicted data D(X^{rep}) against the discrepancy measure for the observed data D(X). The posterior predictive p value is the proportion (in this case .58) of draws above the diagonal line.

We began by choosing $M = 200$ for the number of candidate individuals that were not captured, generating a discrete Uniform prior distribution for population size N that was restricted to be between $n = 68$ at the low bound and $M + n = 268$ at the upper bound. Although this value for M was based on knowledge of the interval estimate of Coull and Agresti (1999), it could have been obtained interactively from short MCMC runs to ensure that $n + M$ was greater than the upper bound of the estimated posterior distribution for N, but not so great that it led to unnecessarily slow and inefficient computation. We also conducted a pilot run to monitor an individual effect, θ_i, for an undetected individual $(i = n + 1)$ that was always included in the model $(1 < n_0$ throughout the pilot run). The results of the pilot run were then used to estimate a Normally distributed pseudo-prior for individual effects when individuals were not chosen to be in the model (estimated mean of $-.84$ and standard deviation of .90).

To assess the sensitivity and efficiency of abundance estimation to the choice of pseudo-prior for the individual effects, we performed three separate full runs of the model with different pseudo-prior distributions. Specifically, the mean value remained as estimated from the pilot run, but the standard deviation of the pseudo-prior from the pilot run was inflated by a fixed constant G, for $G = .5$, 1, and 2. For this comparison of pseudo-priors, we kept a fixed specification of the real priors for the variance hyperparameters σ^2_α and σ^2_θ, respectively, for individual effects included in the model, by maintaining prior averages of $\sigma^2_{0_\alpha}$ and $\sigma^2_{0_\theta} = 1$, and prior degrees of freedom v_{0_α} and $v_{0_\theta} = 1$. For each pseudo-prior setup, we employed MCMC sampling to simultaneously simulate three separate sequences of values from the posterior distributions of interest. After a number of "burn-in" iterations required for chain convergence, we stored the outcome of each iteration until a sample

Table 1. Estimates of model parameters using three different pseudo-prior distributions for the individual effects, θ_i, associated with undetected individuals that are not included in the model ($i > n + n_0$). The mean and standard deviation of the pseudo-prior distribution were based on initial pilot runs, and the standard deviation was then inflated by a fixed constant, $G = .5$, 1, or 2. The median (95% probability interval) are given for each marginal posterior distribution.

G	μ	σ^2_α	σ^2_θ	n_0	N
.5	−1.37	.33	1.12	28	96
	(−2.90, −0.56)	(.11, 1.50)	(.30, 3.89)	(9, 106)	(77, 174)
1	−1.39	.33	1.16	29	97
	(−2.96, −.56)	(.11, 1.50)	(.31, 3.85)	(9, 109)	(77, 177)
2	−1.40	.33	1.18	29	97
	(−2.95, −.56)	(.11, 1.53)	(.32, 3.99)	(9, 109)	(77, 177)

of 30,000 realizations was formed (10,000 for each chain). The length of the burn-in was assessed using the method of Gelman and Rubin (1992), as modified by Brooks and Gelman (1998), which is based on summary statistics comparing the variances within and between three different simulated sequences. Convergence was achieved within 10,000 iterations of each chain for all three pseudo-priors.

Changes in the precision of the pseudo-priors had little effect on posterior inference about parameters of the Rasch model (Table 1). For all three pseudo-prior setups, the overall level of the capture probability on the logit scale, μ, corresponded to a mean capture probability of around .20, and the variance of the occasion effects (σ^2_α) was markedly lower than that of the individual effects (σ^2_θ), highlighting the importance of including individual effects in the model. In absolute terms, this difference can be compared by taking $\mu = -1.39$, and adding separately Gaussian random variables with variances either .33 or 1.16 to the linear predictor. This induces coefficients of variation of .42 and .69 in the implied distributions of observation probabilities, respectively. The posterior estimates for the number of unseen individuals, and accordingly, the total population size, were also extremely similar for the three pseudo-prior distributions.

Although changes in the precision of the pseudo-priors had little effect on posterior inference about parameters, they did markedly affect the efficiency of sampling from the posterior, as noted by Carlin and Chib (1995). Specifically, the autocorrelation between sampled values of N increased with departures from the unmodified pseudo-prior $(G = 1)$ constructed from the pilot run, indicating that moves between models (different values of n_0) were less frequent. For an example, the lag =10 autocorrelation was .77 at $G = 1$, but .84 at $G = 2$ and .82 at $G = .5$. Hence more iterations would be needed when using these modified pseudo-priors to adequately sample from the posterior distribution. This supports the suggestion of Carlin and Chib (1995) that pseudo-priors for omitted individual effects should be chosen to closely match to the conditional posterior if the parameter were in the model, to ensure efficient operation of this MCMC model selection approach to abundance estimation.

To assess the sensitivity of the model selection procedure to the upper bound of the possible number of individuals, we performed separate full runs of the model with $M =$

200, 400, and 600 potentially unseen individuals, using discrete Uniform prior distributions for population size N that were restricted to be between $n = 68$ at the lower bound and $n + M = 268$, 468, and 668 at the upper bound, respectively. For consistency, we adopted the unmodified $(G = 1)$ pseudo-prior for all three runs, and the same Gamma priors for the variance hyperparameters σ^2_α and σ^2_θ. The different numbers of candidate unseen individuals that were sampled in each scenario was reflected by the speed of the

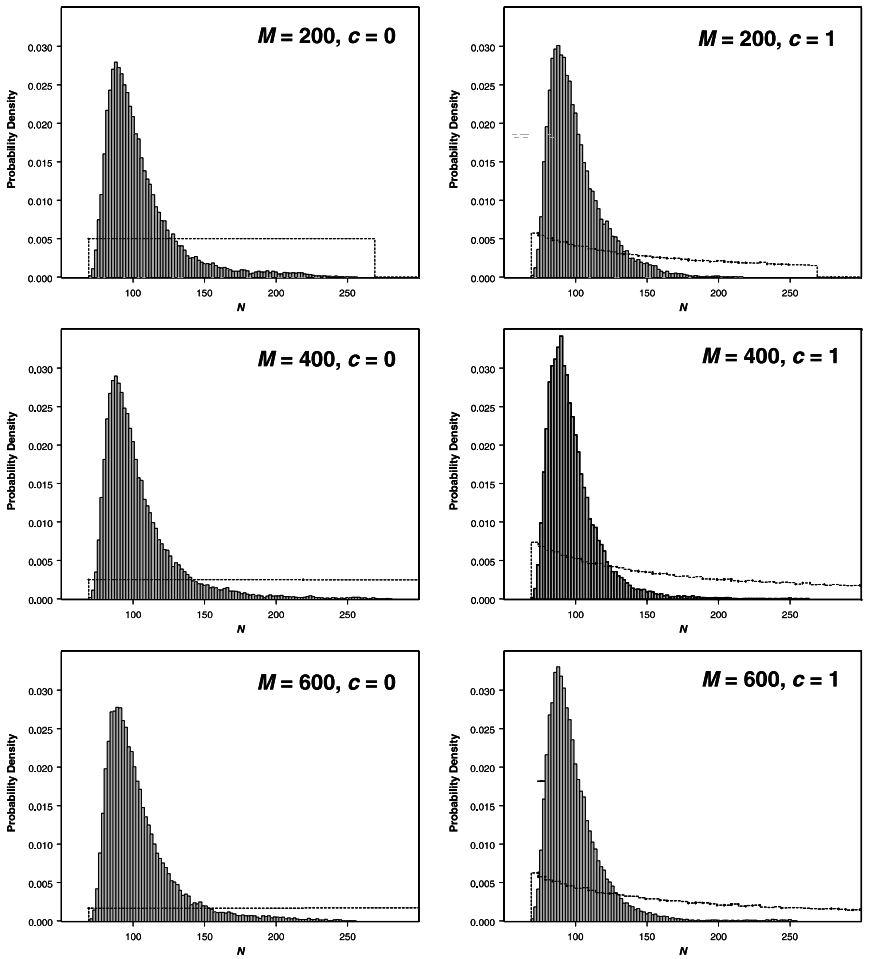

Figure 2. Marginal posterior probability distributions for snowshoe hare population size, N, estimated using a range of prior distributions, illustrated using dashed lines. Prior distributions were formulated using $M = 200$, 400, and 600 potentially unseen individuals, and the prior densities for N were therefore distributed between the number of observed individuals $n = 68$ at the lower bound and $n + M = 268$, 468, and 668 at the upper bound, respectively. Both discrete Uniform and Jeffreys priors were adopted by formulating the prior for N as proportional to N^{-c}, with $c = 1$ for the Jeffreys prior and $c = 0$ for the Uniform prior. For clarity, distributions are plotted over the parameter space defined by the $M = 200$ prior.

Table 2. Summary statistics for the posterior distribution for population size, *N*, estimated for a variety of prior forms for the variance hyperparameters for the occasion and individual effects, σ^2_α and σ^2_θ, respectively. In each case, we maintained values of $\sigma^2_{0_\alpha}$ and $\sigma^2_{0_\theta} = 1$ as the average of these variance terms, but the degree of strength of the prior beliefs was changed by using values of .5, 1, and 5 for the degrees of freedom parameters v_{0_α} and v_{0_θ}. The marginal posterior distributions for *N* are summarized by the most frequently sampled value (mode) in the MCMC sample, along with the 2.5, 50 (median) and 97.5 percentile values of the distribution.

v_{0_α}	v_{0_θ}	*Mode*	*2.5%*	*Median*	*97.5%*
1	.5	86	76	95	186
1	1	88	77	97	177
1	5	88	78	96	165
.5	1	88	77	97	176
5	1	86	77	96	171

MCMC sampling. The 30,000 iterations (10,000 for each of the three chains) were accomplished in approximately 2,000 seconds for $M = 200$, 4,000 seconds for $M = 400$, and 6,000 seconds for $M = 600$. However, the estimated posterior distributions were very similar under the three scenarios, despite the widely differing prior ranges for N (Figure 2). The mode of the posterior distribution was 88 (median =97) in all three cases, and the 95% posterior probability intervals were virtually identical at $[77, 177]$, $[77, 175]$, and $[77, 176]$ for $M = 200$, 400, and 600, respectively. Note that the posterior distributions were highly skewed to the right, with asymmetrical 95% intervals about the median. This lack of effect of the prior range on posterior inference is unsurprising given that the addition to the range of the prior for N is unsupported by the data. However, the time penalty of the wider priors suggests it is worth making preliminary runs to establish the minimum range necessary for the prior for N.

We also explored how inference about N might be sensitive to the specifications of the forms assumed for the real priors. For consistency, we kept fixed values for the pseudo-priors with $G = 1$, but in turn varied the priors assigned to the population size, N, and the variance of occasion and individual effects, σ^2_θ and σ^2_α, respectively. First, keeping the prior for σ^2_θ and σ^2_α fixed as before $(\sigma^2_{0_\alpha}, \sigma^2_{0_\theta} = 1, v_{0_\alpha}, v_{0_\theta} = 1)$, we adopted a truncated Jeffreys prior for N (Jeffreys 1946) in place of the truncated uniform prior. With $M = 200$ there was very little change in the median of the posterior distribution when changing the prior from the truncated uniform (median $= 97$) to the Jeffreys prior (median $= 94$), but the 95% posterior interval reduced from $[77, 177]$ to $[76, 142]$. This shrinking of the upper tail of the posterior distribution is not surprising, as the Jeffreys prior is an asymmetrical distribution that concentrates greater mass on the lower values in the truncated range. As with the uniform prior before, there was little effect of increasing the prior range using $M = 400$ and $M = 600$ with the Jeffreys prior (Figure 2).

Finally, keeping the Uniform prior for N with $M = 200$, we changed the priors for σ^2_θ and σ^2_α by sequentially adopting a range of values for the prior degrees of freedom, $v_{0_\alpha}, v_{0_\theta} = 5, 1, .5$, to reflect informative priors centred on 1, relatively uninformative and very uninformative formulations, respectively (Table 2). Changes in the precision of the prior for either of these variance hyperparameters had relatively little effect on the posterior

distribution of N. There was very little change in the mode or median of the sampled values, but informative priors centered on 1 generally produced a smaller upper tail in the posterior sample, whereas the upper bound was slightly increased as the prior became less informative. This effect appears greater for changes in the prior for the variance of the individual effects compared to that for the occasion effects.

6. DISCUSSION

Bayesian statistical inference has been suggested as a suitable approach for wildlife mark–recapture studies (Pollock 1991) and is increasingly being advocated for analyzing and communicating uncertainty in ecological data analysis (Ellison 1996; Wade 2000). We have shown how a Bayesian hierarchical formulation of the Rasch model (e.g., Fienberg et al. 1999) can be used to make sensible inferences about population size in animal mark–recapture where capture probabilities vary over time and individuals. Inference based on Bayesian posterior probability distributions facilitates communication of both the extent and shape of the associated uncertainty. In our example, estimates from the posterior distribution for population size were similar to the maximum likelihood estimate of 92 produced from the same data by Coull and Agresti (1999). When using a uniform prior distribution for N, the Bayesian posterior distribution had 95% intervals (77–177) that were larger than these authors' profile likelihood confidence intervals (75–154). As with the likelihood confidence intervals, this marginal distribution was positively skewed indicating greater uncertainty about the upper bound of population size than the lower bound. Such skewness is intuitively expected in mark–recapture studies, given that the number of individuals that are actually captured forms a lower bound on population size. This skewness also emphasizes the danger of conventional approaches to interval estimation that rely on asymptotic normality.

Fortunately, modern computational techniques have broadened the range of methodology used in mark–recapture modeling (Buckland et al. 2000). Interval estimation through the use of profile likelihood (e.g., Cormack 1992), bootstrap resampling (e.g., Buckland and Garthwaite 1991) and Bayesian methods for calculation of the posterior distribution (e.g., Smith 1991) now offer more realistic and flexible approaches for inference. The recent development of MCMC approaches for approximation of posterior distributions offers a useful addition to this toolkit, enabling Bayesian inference from relatively complex models with intractable posterior distributions. For example, George and Robert (1992), Lee and Chen (1998), Fienberg et al. (1999), and King and Brooks (2001), have all applied MCMC methods in a mark–recapture context. We have followed these authors in using the Bayesian MCMC approach for fitting the Rasch model to mark–recapture data.

The successful application of MCMC techniques for simulating solutions from the Bayesian Rasch model is not automatic. Fienberg et al. (1999) noted that direct implementation of MCMC is problematic in this model, since the dimension of the vector of individual effects depends on the unknown number of individuals in the population. To overcome this, we constructed a constant parameter dimension by introducing a fixed number of undetected individuals, with associated augmented data and unknown individual

effects. Each possible number of undetected individuals could then be viewed as a different model within a product model space, and we adopted the Gibbs sampling MCMC approach of Carlin and Chib (1995) for sampling from this product model space to estimate the probability of each candidate population size. Fienberg et al. (1999) suggested that this type of abundance estimation through model selection could be achieved using the "reversible jump" MCMC (RJMCMC) approach of Green (1995) for sampling over separately defined alternative models with variable dimensions. However, the product model space approach that we have adopted has the advantage that it can be sampled using standard MCMC methods, such as the Gibbs sampler (George and Casella 1992), which can be implemented using the freely available WinBUGS software (code can be downloaded from http://www.abdn.ac.uk/zoology/lighthouse/research/JABES_2005.shtml). This software package offers the advantage that modifications and extensions can easily be made to the model, with a minimum of programming time (Gilks, Thomas, and Spiegelhalter 1994). For example, this utility could be exploited for fitting different forms for the variation in individual effects (e.g., Pledger 2000) or for including dependence between capture occasions through hierarchical extensions to the model (e.g., Fienberg et al. 1999).

This ease of implementation does not come without a price. In particular, when a large number of candidate models for unobserved individuals is introduced into the product model space, this approach can become computationally expensive, requiring many hours or even days to complete a satisfactory number of iterations. This is because the parameters of every model must be updated at every iteration, regardless of whether they are in the currently chosen model or not. This is not the case with the RJMCMC approach of Green (1995), where only the parameters of the chosen model are updated at a given iteration. However, there is currently no stand-alone software available to implement RJMCMC, which therefore requires substantially greater investment in initial programming time. We suggest that the run time for our approach should be minimized by careful specification of the candidate number of undetected individuals, based on pilot runs and available prior information on the upper bound of possible abundance.

ACKNOWLEDGMENTS

The first author was supported by a postgraduate scholarship from the Carnegie Trust for the Universities of Scotland, and completed most of this work while at the University of Aberdeen, Department of Zoology, Lighthouse Field Station, Cromarty, Ross-Shire IV11 8YJ. The second author was funded by the Scottish Executive Environment and Rural Affairs Department.

[Received May 2004. Revised April 2005.]

REFERENCES

Agresti, A. (1994), "Simple Capture-Recapture Models Permitting Unequal Catchability and Variable Sampling Effort," *Biometrics*, 50, 494–500.

Alho, J. M. (1990), "Logistic Regression in Capture-Recapture Models," *Biometrics*, 46, 623–635.

Brooks, S. P. (1998), "Markov Chain Monte Carlo Method and its Application," *The Statistician*, 47, 69–100.

Brooks, S. P., and Gelman, A. (1998), "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455.

Buckland, S. T., Anderson, D. R., Burnham, K. P, and Laake, J. L. (1993), *Distance Sampling, Estimating Abundance of Biological Populations*, London: Chapman and Hall.

Buckland, S. T., Goudie, I. B. J., and Borchers, D. L. (2000), "Wildlife Population Assessment: Past Developments and Future Directions," *Biometrics*, 56, 1–12.

Buckland, S. T., and Garthwaite, P. H. (1991), "Quantifying Precision of Mark-Recapture Estimates Using the Bootstrap and Related Methods," *Biometrics*, 47, 255–268.

Burnham, K. P., and Overton, W. S. (1978), "Estimation of the Size of a Closed Population when Capture Probabilities Vary," *Biometrika*, 65, 625–633.

Carlin, B. P., and Chib, S. (1995), "Bayesian Model Choice via Markov Chain Monte Carlo," *Journal of the Royal Statistical Society*, Series B, 57, 473–484.

Carothers, A. D. (1973a), "Capture-Recapture Methods Applied to a Population With Known Parameters," *Journal of Animal Ecology*, 42, 125–146.

——— (1973b), "The Effects of Unequal Catchability on Jolly-Seber Estimates," *Biometrics*, 29, 79–100.

Casella, G., and George, E. I. (1992), "Explaining the Gibbs Sampler," *The American Statistician*, 46, 167–174.

Chao, A., Lee, S. M., and Jeng, S. L. (1992), "Estimating Population Size for Capture–Recapture Data when Capture Probabilities Vary by Time and by Individual Animal," *Biometrics*, 48, 201–216.

Cormack, R. M. (1989), "Log-Linear Models for Capture-Recapture," *Biometrics*, 45, 395–413.

——— (1992), "Interval Estimation for Mark–Recapture Studies of Closed Populations," *Biometrics*, 48, 567–576.

Coull, B. A., and Agresti, A. (1999), "The Use of Mixed Logit Models to Reflect Heterogeneity in Capture-Recapture Studies," *Biometrics*, 55, 294–301.

Darroch, J. N., Fienberg, S. E., Glonek, G. F. V., and Junker, B.W. (1993), "A Three-Sample Multiple-Recapture Approach to Census Population Estimation with Heterogeneous Catchability," *Journal of the American Statistical Association*, 88, 1137–1148.

Dorazio, R. M., and Royle, A. J. (2003), "Mixture Models for Estimating the Size of a Closed Population When Capture Rates Vary Among Individuals," *Biometrics*, 59, 351–364.

Ellison, A. M. (1996), "An Introduction to Bayesian Inference for Ecological Research and Environmental Decision-Making," *Ecological Applications*, 6, 1036–1046.

Fienberg, S. E., Johnson, M. S., and Junker, B. W. (1999), "Classical Multilevel and Bayesian Approaches to Population Size Estimation Using Multiple Lists," *Journal of the Royal Statistical Society*, Series A, 162, 383–405.

Gelman, A., Carlin, J. B., Stern H. S., and Rubin. D. B. (1995), *Bayesian Data Analysis*, London: Chapman and Hall.

Gelman, A., Meng, X. L., and Stern, H. S. (1996), "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies," *Statistica Sinica*, 6, 733–807.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences" (with discussion), *Statistical Science*, 7, 457–511.

George, E. I., and Robert, C. P. (1992), "Capture-Recapture Estimation via Gibbs Sampling," *Biometrika*, 79, 677–683.

Geweke, J. (1996), "Variable Selection and Model Comparison in Regression," in *Bayesian Statistics 5*, eds. J. M. Bernado, J. O. Berger, A. P. Dawid, and A. F. M. Smith, New York: Oxford University Press, pp. 609–620.

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994), "A Language and Program for Complex Bayesian Modeling," *The Statistician*, 43, 169–177.

——— (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.

Green, P. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732.

Huggins, R. M. (1989), "On the Statistical Analysis of Capture Experiments," *Biometrika*, 76, 133–140.

Jeffreys, H. (1946), "An Invariant Form for the Prior Probability in Estimation Problems," in *Proceedings of the Royal Statistical Society*, Series A, 186, pp. 453–461.

King, R., and Brooks, S. P. (2001), "On the Bayesian Analysis of Population Size," *Biometrika*, 88, 317–336.

Kuo, L., and Mallick, B. (1998), "Variable Selection for Regression Models," *Sankhya*, Ser. B, 60, 65–81.

Lee, S. M., and Chen, C. W. S. (1998), "Bayesian Inference on Population Size for Behavioural Response Models," *Statistica Sinica*, 8, 1233–1248.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), "WinBUGS—A Bayesian Modeling Framework: Concepts, Structure, and Extensibility," *Statistics and Computing*, 10, 325–337.

Norris, J. L., and Pollock, K. H. (1996), "Nonparametric MLE Under Two Closed Capture-Recapture Models With Heterogeneity," *Biometrics*, 52, 639–649.

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, 1–135.

Pollock, K. H. (1991), "Modeling Capture, Recapture, and Removal Statistics for Estimation of Demographic Parameters for Fish and Wildlife Populations: Past, Present and Future," *Journal of the American Statistical Association*, 86, 225–238.

Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. E. (1990), "Statistical Inference for Capture-Recapture Experiments," *Wildlife Monographs*, 107.

Pledger, S. (2000), "Unified Maximum Likelihood Estimates for Closed Capture-Recapture Models Using Mixtures," *Biometrics*, 56, 434–442.

Rasch, G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*, Chicago: University of Chicago Press.

Schwarz, C. J., and Seber, G. A. F. (1999), "Estimating Animal Abundance, Review III," *Statistical Science*, 14, 427–456.

Seber, G. A. F. (1982), *The Estimation of Animal Abundance and Related Parameters* (2nd ed.), London: Charles Griffin and Company Ltd.

Smith, P. J (1991), "Bayesian Analysis for a Multiple Capture-Recapture Model," *Biometrika*, 78, 399–407.

Wade, P. R. (2000), "Bayesian Methods in Conservation Biology," *Conservation Biology*, 14, 1308–1316.