

# Bayesian Capture-Recapture Analysis and Model Selection Allowing for Heterogeneity and Behavioral Effects

Sujit K. GHOSH\*

James L. NORRIS†

Institute of Statistics Mimeo Series# 2562

## Abstract

In this paper, we present Bayesian analysis of capture-recapture models for a closed population which allows for heterogeneity of capture probabilities between animals and bait/trap effects. We utilize a flexible discrete mixture model to account for the heterogeneity and behavioral effects. In addition we present a solid model selection criterion. Through illustrations with a real-world motivating example, we demonstrate how Bayesian analysis can be applied in this setting and discuss some major benefits which result, including consideration of informative priors based on historical data.

**Key words:** Bayesian inference, Capture-recapture models, Closed population, Heterogeneity, Gibbs sampling, MCMC, WinBUGS.

---

\*Sujit K. Ghosh is Associate Professor in the Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203 (Email: sghosh@stat.ncsu.edu)

†James L. Norris is Professor in the Department of Mathematics, Wake Forest University Winston-Salem, NC 27109-7388 (Email: norris@wfu.edu)

# 1 INTRODUCTION

Likelihood-based inference is a foundation of statistical applications. For many animal capture-recapture settings, Maximum Likelihood (ML) estimation has been one of the most preferred inferential method. Darroch (1958) considered time differences, Norris and Pollock (1996) considered animal differences, and Pledger (2000) developed a unified theory for closed populations using finite mixtures. ML based inferences for open population models have been developed by Jolly (1965), Seber (1965) and Brownie et al. (1985). Although this ML estimation can be very useful, another likelihood-based method, namely Bayesian analysis, allows one to: (i) obtain probability regions for parameters, (ii) incorporate information from previous similar studies, (iii) conduct rigorous non-nested model selection and (iv) incorporate ecologically meaningful decision structures and loss functions.

Particularly with the recent development and easy availability of Bayesian analysis software (e.g., WinBUGS, Spiegelhalter et al., 2001), Bayesian procedures are attractive. In particular, this software eliminates many of the complex integration difficulties which have limited Bayesian applications.

Most of the previous Bayesian capture-recapture research for closed population have centered on model “ $M_t$ ” which only allows differences in capture probabilities over capture times/periods (t) (see e.g., Castledine, 1981, George and Robert, 1992). In contrast, our proposed model assumes that the animals’ capture probabilities do not vary over capture periods (i.e., no time effects), but such probabilities are allowed to vary with animals (i.e. heterogeneity (h) effect) and bait/trap (b) effects. Such models are denoted by  $M_{bh}$  in capture recapture literature. Basu and Ebrahimi (2001) have independently examined Bayesian analysis for the heterogeneity (only) model  $M_h$  and motivated us to study an extension of their model to include bait/trap effects. In our article, we incorporate the proposed informative priors based on a previous similar study. In addition, to the best of our knowledge the proposed methodology yields

the first rigorous model selection criterion within the capture-recapture setting which is based on posterior predictive distributions. The proposed model selection criterion is adapted from the decision theoretic framework of Gelfand and Ghosh (1998).

Throughout our paper, we emphasize a real-world example for which our proposed methodology is very useful. In Section 2, we present this example and develop a general Bayesian mixture methodology to examine this setting. A posterior predictive *model selection* methodology is developed in Section 3, and detailed results on our example are presented in Section 4. Section 5 discusses the general applicability of our methods to numerous other settings. All WinBUGS code that we have used to fit the proposed models can be obtained from the first author. As an illustration, in the Appendix we include code to fit the most general model  $M_{bh}$ .

## 2 A REAL-WORLD EXAMPLE AND METHODOLOGICAL DEVELOPMENT

### 2.1 Description of the Meadow vole Example

Jim Nichols of Patuxent National Wildlife Research station graciously provided us with data on his monthly *Microtus* (meadow vole) sampling for the year 1999. During July 1999, the usual five consecutive capture days were employed. However, during August 1999 (on the same site), the last two of the five capture days were contaminated by a squirrel infestation. Thus there is very little uncontaminated data available for August 1999, so we would especially like to consider incorporating prior information, i.e., including the use of July 1999's data, when we make inferences about August 1999. Jim Nichols states that there is a similar setting over the July to August 1999 sampling window, see Section 4.2 for details. Thus we model a closed population with no time effects. However, admittedly this modeling is approximate since rainfall and other

factors varied from one day to the next during this period. The meadow voles could have different capture probabilities and an animal can become trap shy (or trap happy) after first capture; to allow all such possibilities, we consider  $M_{bh}$  models, which we describe explicitly in the next section.

## 2.2 General $M_{bh}$ Mixture Model

Consider a capture-recapture experiment with  $k$  similar capture periods. We allow heterogeneity (h) in capture probabilities among individual animals, and we also allow an individuals' future capture probability to be influenced by previous capture; the latter is known as a "bait/trap effect", and is denoted by the subscript  $b$ . Otis *et al.* (1978) and others label this combined heterogeneity and bait/trap model as  $M_{bh}$ . Specifically,  $p_{1m}$  denotes animal  $m$ 's chance of capture (on a given period) if it has not been captured previously, and  $p_{2m}$  denotes animal  $m$ 's chance of capture if it has been previously captured. It is noted that some of the authors have used the notations  $p_m$  and  $c_m$  instead of our  $p_{1m}$  and  $p_{2m}$  notations, respectively. To allow heterogeneity (h) between animals, we assume that each animal's capture probabilities, given by the vector of probabilities  $\mathbf{p}_m = (p_{1m}, p_{2m})^T$  arise randomly from a common but unknown 2-dimensional distribution. Let the cumulative distribution function of this common distribution be denoted by  $G(\cdot, \cdot)$ ; i.e.,  $\Pr[p_{1m} \leq p_1, p_{2m} \leq p_2] = G(p_1, p_2)$ . Norris and Pollock (1995) have examined this setting from a frequentist perspective and have found that the Nonparametric MLE (NPMLE) nearly always has four or less points of support for  $G$  for most of their examples. Thus, we assume that the unknown distribution function  $G$  is discrete with  $r$  points of support, where  $r = 1, 2, 3$ , or  $4$ . To express this explicitly, the probability mass function corresponding to  $G$  is given by  $\Pr[p_{1m} = p_1, p_{2m} = p_2] = dG(\mathbf{p}) = \sum_{i=1}^r \pi_i I(\mathbf{p} = \boldsymbol{\theta}_i)$ , where  $I(\cdot)$  denotes the indicator function and  $\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{21})^T, \dots, \boldsymbol{\theta}_r = (\theta_{1r}, \theta_{2r})^T$  are unknown support points with unknown probabilities  $\pi_1, \dots, \pi_r$  respectively, such that  $\sum_{i=1}^r \pi_i = 1$

and for identifiability of parameters we assume that  $0 \leq \theta_{11} < \dots < \theta_{1r} \leq 1$ . Notice that by choosing  $r$  large, we can approximate a nonparametric Dirichlet process prior for  $G$ ; more discussions on such approximations are presented at the end of Section 2.3.1.

Under the above model  $M_{bh}$ , since the capture probability  $\mathbf{p}_m$  of the  $m$ -th animal is assumed to be the same for each of the  $k$  capture periods, it follows (Norris and Pollock, 1995) that it is sufficient to consider the count statistics

$$\mathbf{f}^{obs} = \{f_{jl}; j = 1, \dots, k, l = 1, \dots, k - j + 1\},$$

where  $f_{jl}$  denotes the number of animals that were first caught on the  $j$ -th capture period and are caught a total of  $l$  times in the  $k$  capture periods. Thus, separately for each fixed value of  $r$ , we consider the multinomial likelihood function given by,

$$L(N, G) \propto \binom{N}{S} \prod_{j=1}^k \prod_{l=1}^{k-j+1} P_{jl}^{f_{jl}} \left( 1 - \sum_{j=1}^k \sum_{l=1}^{k-j+1} P_{jl} \right)^{N-S} \quad (1)$$

where  $S = \sum_{j=1}^k \sum_{l=1}^{k-j+1} f_{jl}$  denotes the total number of distinct animals caught within the  $k$  capture periods and

$$\begin{aligned} P_{jl} &= \binom{k-j}{l-1} \int_0^1 \int_0^1 p_1 (1-p_1)^{j-1} p_2^{l-1} (1-p_2)^{k-j-l+1} dG(p_1, p_2) \\ &= \binom{k-j}{l-1} \sum_{i=1}^r \pi_i \theta_{1i} (1-\theta_{1i})^{j-1} \theta_{2i}^{l-1} (1-\theta_{2i})^{k-j-l+1} \end{aligned}$$

denotes the probability that an animal was first captured at period  $j$  and caught a total of  $l$  times. Note that, the submodel  $M_h$  is obtained from  $M_{bh}$  by assuming that  $\theta_{1i} = \theta_{2i}, \forall i$ . Also the other submodel,  $M_b$  (bait/trap or behavioral effect only) is obtained from  $M_{bh}$  by assuming that there is no heterogeneity and hence the distribution function  $G$  is degenerate at  $\boldsymbol{\theta} = \boldsymbol{\theta}_0 = (\theta_{10}, \theta_{20})$ . In other words, an  $M_{bh}$  model with  $r = 1$ , is the same as an  $M_b$  model. Finally, note that a  $M_h$  model with  $r = 1$  does not allow heterogeneity

or bait/trap effects (i.e., does not allow any effects) and thus is labeled  $M_0$ . We compare all of these submodels to the full model,  $M_{bh}$ , using a rigorous model selection criterion (see Section 2.4). Although all of these submodels are nested within the  $M_{bh}$  models, the submodels themselves might not be nested, e.g.,  $M_b$  and  $M_h$  are not nested. This motivates us to develop model selection criterion which allows for non-nested mixture models.

## 2.3 Bayesian Analysis

In order to complete the model specification from a Bayesian perspective, we specify a joint prior distribution for all parameters of the model. We consider both noninformative and informative prior distributions. To model the prior distributions comparatively insensitive to posterior estimates we specify a class of non-informative (Jeffreys, 1946) priors for all the model parameters. However it is worthwhile to note that the posterior distribution becomes insensitive to the choice of prior distribution when the data overwhelms the prior.

### 2.3.1 Noninformative Prior distributions for the $M_{bh}$ Models

We define a class of non-informative priors for the model parameters. In particular, we specify a non-informative prior for the set of parameters  $(N, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_r, \pi_1, \dots, \pi_r)$  by assuming that a priori  $N, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_r$  and  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_r)^T$  are independent with marginal distributions given by,

$$\begin{aligned}
 p(N) &\propto 1/N^\delta I(N = 1, 2, \dots) \\
 p(\theta_{11}, \dots, \theta_{1r}) &\propto \prod_{i=1}^r \theta_{1i}^{\beta_1-1} (1 - \theta_{1i})^{\beta_1-1} I(\theta_0 < \theta_{11} < \dots < \theta_{1r} \leq 1) \\
 \theta_{2i} &\sim \text{Beta}(\beta_2, \beta_2), \quad i = 1, \dots, r \\
 (\pi_1, \dots, \pi_r) &\sim \text{Dirichlet}\left(\frac{\alpha}{r}, \dots, \frac{\alpha}{r}\right) \tag{2}
 \end{aligned}$$

where  $\delta, \beta_1, \beta_2, \theta_0, \alpha$  are arbitrary but known *positive* numbers. Choosing these fixed numbers (except  $\theta_0$ ) to be close to half yields non-informative priors,

based on the idea of Jeffreys (1946). Notice that a flat prior is obtained by choosing  $\delta = 0, \beta_1 = \beta_2 = 1$  and  $\alpha = r$ . However when  $\delta \leq 1$ , it follows that  $\sum_{N=1}^{\infty} 1/N^\delta = \infty$  and hence results in an improper prior. To see that the above prior specifications yields a proper posterior distribution when  $\delta \leq 1$ , notice that,

$$\begin{aligned} 1 - \sum_{j=1}^k \sum_{l=1}^{k-j+1} P_{jl} &= \int_0^1 (1 - p_1) dG_1(p_1) \\ &= \sum_{i=1}^k \pi_i (1 - \theta_{1i})^k \leq (1 - \theta_0)^k \end{aligned}$$

and hence the product of the likelihood in (??) and the prior in (??) is bounded by the quantity,

$$\binom{N}{S} (1 - \theta_0)^{k(N-S)} \prod_{i=1}^r \theta_{1i}^{\beta_1-1} (1 - \theta_{1i})^{\beta_1-1} \theta_{2i}^{\beta_2-1} (1 - \theta_{2i})^{\beta_2-1} \pi_i^{\alpha-1}$$

which is integrable with respect to the parameters  $N, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_r$  and  $\boldsymbol{\pi}$ . In (2), we have used a special form of the Dirichlet distribution, which can be used to approximate a Dirichlet Process Prior for  $G$ , by letting  $r \rightarrow \infty$ . In fact, Ishwaran and Zarepour (2002) proposed a similar method to construct Dirichlet prior sieves for normal mixture models. In most cases, one may choose  $r = k^*$  for small  $k^*$  or  $r = \lceil \sqrt{k^*} \rceil$  for large  $k^*$  to approximate a Dirichlet process prior for  $G$ , where  $\lceil x \rceil$  denotes the largest integer not exceeding  $x$  and  $k^*$  is the dimension of the sufficient count statistics, e.g., for  $M_{bh}$  models  $k^* = k(k+1)/2$ , where  $k$  is the number of capture periods.

### 2.3.2 Informative Prior Distributions for $M_{bh}$ Models

There may be biologically motivated prior information about  $N$  and  $G$ . In particular, if this prior information is from a (previously run) similar study then we can multiply the likelihood function of the previous study, as derived in (1), with a noninformative prior (or a different prior if applicable) to obtain,

up to a multiplicative constant, the posterior of the similar study and thus the informative prior for the current study. Then this prior is used along with the current study's likelihood to obtain the posterior distribution for the current study. Even when the posterior distribution of the similar study is not available in closed form, we can still appropriately incorporate the similar study's information to obtain the posterior for the current study. This is most easily accomplished by beginning with the, say, noninformative prior for the previous study and then multiply it by the combined likelihood of the similar study and the current study to obtain the posterior of the current study, this uses the sequential version of Bayes theorem. In Section 4.2, we will use this informative prior technique for our meadow vole example.

### 2.3.3 Markov Chain Monte Carlo

The posterior density of  $(N, G)$  is proportional to the joint prior density times the likelihood function (in (1)). Determination of the constant of proportionality and integration over some parameters (say to obtain the marginal posterior distribution of only  $N$ ) are not analytically possible for this Bayesian model. So we use Markov Chain Monte Carlo (MCMC) methods to sample from the joint posterior distribution of  $(N, G)$  given the observed counts  $f_{jl}$ . Computationally, it is rather easy to implement the model in WinBUGS, a free software routine available on the internet (see <http://www.mrc-bsu.cam.ac.uk/bugs>). The software uses Gibbs sampling with necessary Metropolis-Hastings steps to obtain samples from the posterior distributions. Our WinBUGS code presents a user-friendly version of the MCMC method which can be used by a practitioner. Moreover, our code can be easily modified to obtain posterior estimates based on a user-defined prior distribution. One major advantage of Bayesian analysis is that it provides finite sample interval estimates based on the posterior distribution. Theoretically, it is well known that the Bayes estimates have several optimal properties including some of the frequentist properties such as

admissibility, minimaxity etc. (Berger, 1985). For squared error loss and absolute error loss, the Bayes estimates are respectively the mean and the median of the posterior distribution. Shortest length probability intervals can also be obtained from the corresponding posterior distribution. For our application we have obtained the equal-tail 95% posterior intervals by computing the 2.5 and 97.5 percentiles of the posterior distribution of the parameters. For skewed distributions, we would prefer to use robust point estimates based on quantiles rather than moments, which amounts to using robust asymmetric absolute loss functions.

### 3 BAYESIAN MODEL SELECTION FOR NONNESTED MODELS

One major contribution of this article is to use model selection based on the decision-theoretic framework of Gelfand and Ghosh (1998). The proposed model selection method is applicable even for non-nested mixture models.

With increasing power in fitting simulation-based Bayesian models, it becomes very important to find which model(s) fit the data best. One major difficulty with most model selection criteria is that they require the explicit calculation of the dimension of the parameter space. This is troublesome for mixture models which are often used to model heterogeneity. Model selection criteria based on cross-validation or another holdout method is also difficult to adapt here; specifically, given the limited data available for August 1999, it is not clear what and how much sample can be retained for cross-validation. Fortunately, Gelfand and Ghosh (1998) offer a decision-theoretic model selection criterion based on the posterior predictive loss approach. An added advantage of this criterion is its straightforward computation from the MCMC samples.

Related to model determination we also discuss the *deviance* (McCullagh and Nelder, 1989) of each model, which is defined as twice the difference be-

tween the *saturated* log-likelihood and the log-likelihood. For instance, for each of the  $M_{bh}$  models and its submodels, the *deviance* based on the likelihood (1) is given by,

$$dev = 2 \left[ \sum_{l+j < k} f_{jl} \log \frac{f_{jl}}{NP_{jl}} + (N - S) \log \frac{N - S}{N(1 - \sum_{l+j < k} P_{jl})} \right]. \quad (3)$$

Note that the *dev* is a function of the model parameters; its posterior distribution can easily be calculated from the MCMC samples.

A Bayesian version of another popular model selection criterion *AIC* (see Akaike, 1973 and Schwartz, 1978) is often calculated by adding an appropriate penalty (a function of the dimension of the parameters) to the deviance evaluated at the posterior mean of the parameters. This is slightly different from the usual definition of *AIC* in which the deviance is evaluated at the MLE of the parameters instead of the posterior mean. However, this *AIC* model selection criterion is based on asymptotic results whose regularity conditions are not satisfied for mixture models (see McLachlan et al., 2000 (page 203) and Titterton et al., 1985). Similar problems arise when using the *BIC*, as the dimension of the parameters is not well defined for models involving latent variables. Alternatively, we could have used the Deviance Information Criteria (*DIC*) (built within the WinBUGS environment) as proposed by Spiegelhalter et al.(2002) to choose between the models. However use of the *DIC* is not recommended for models involving mixtures of distributions (see WinBUGS manual, Spiegelhalter et al., 2001).

We now discuss the Gelfand and Ghosh (1998) model selection criterion (*GGC*) in more detail. It is based on the finite sample predictive performances of the models. The *GGC* is based on the posterior predictive distributions, and hence, irrespective of the dimensions of the parameter space, it can be used to compare different models that may not be nested. Following Gelfand and Ghosh (1998) we define  $\mathbf{f}^{obs} = (f_{11}, \dots, f_{k1})$  as the observed counts and  $\mathbf{f}^{pred} = (f_{11}^{pred}, \dots, f_{k1}^{pred})$  as the predicted data obtained from the following

posterior predictive distribution,

$$p(\mathbf{f}^{pred}|\mathbf{f}^{obs}) = \int p(\mathbf{f}^{pred}|N, G)p(N, G|\mathbf{f}^{obs})dNdG \quad (4)$$

where  $p(\mathbf{f}^{obs}|N, G)$  denotes the sampling distribution of the data, which is the same likelihood function defined in (??), evaluated at  $\mathbf{f}^{pred}$ , and  $p(N, G|\mathbf{f}^{obs})$  denotes the posterior distribution of the parameters  $(N, G)$  given the observed data  $\mathbf{f}^{obs}$ . Next, to define the *GGC*, we need a loss function that measures discrepancy between  $\mathbf{f}^{obs}$  and  $\mathbf{f}^{pred}$ . One popular choice is the Mean Square Predicted Error (MSPE) on the log-scale,

$$MSPE = \frac{2}{k(k+1)} \sum_{j=1}^k \sum_{l=1}^{k-j+1} \left[ \log(f_{jl}^{pred} + 0.5) - \log(f_{jl} + 0.5) \right]^2 \quad (5)$$

where we have used a stabilizing factor of 0.5 to avoid  $\log(0)$ , when the observed counts ( $f_{jl}$ 's) are zeros. Based on the loss function in (??), we obtain the Gelfand and Ghosh (1998) criteria as,

$$GGC = E \left[ MSPE|\mathbf{f}^{obs} \right] \quad (6)$$

where the expectation is taken with respect to the predictive distribution defined in (??). For our meadow vole example, we will select the model with the smallest *GGC*, i.e., with the smallest posterior mean of MSPE. The computation of *GGC* is done within the `WinBUGS` code by sampling the ‘‘predicted observations’’ from the distribution given in (??). It can be shown that the *GGC* has nice interpretations as the sum of predictive variances (as a penalty term) and goodness-of-fit terms (see Gelfand and Ghosh, 1998 for more details). In this context, another popular loss function is to use a square-root transformation on the counts.

Calibration of *GGC* is a difficult issue as with most of the other model selection criteria. As a rule of thumb we pick the model with the lowest value of *GGC* irrespective of the actual difference. This is because by construction

GGC already achieves a compromise between goodness-of-fit and penalty (see Gelfand and Ghosh, 1998 for justifications).

## 4 RESULTS FOR THE MEADOW VOLE EXAMPLE

In this section we fit the models introduced in Section 2.2 to the August 1999 data. Consider again the meadow vole sampling experiment introduced in Section 2.1. First, we present the sufficient counts (the  $f_{jl}$ 's, described in Section 2.2) based on the uncontaminated first three days of the week of August 1999.

**August data:**  $f_{11} = 25, f_{12} = 30, f_{13} = 43, f_{21} = 10, f_{22} = 10, f_{31} = 4$ .

The sufficient counts for the five days of the week of July 1999 are given by:

**July data:**  $f_{11} = 6, f_{12} = 9, f_{13} = 6, f_{14} = 6, f_{15} = 5, f_{21} = 3, f_{22} = 6,$   
 $f_{23} = 3, f_{24} = 2, f_{31} = 5, f_{32} = 5, f_{33} = 2, f_{41} = 7, f_{42} = 0, f_{51} = 4$ .

Recall that we do not consider any time heterogeneity models ( $M_t, M_{tb}$  etc.) as the samples were collected on three/five days within a week for both July and August of 1999. There could be some unmeasured effects due to severe weather conditions in Maryland during these sampling periods. We assume that such effects are minimal compared to heterogeneity and/or bait effects. We will examine the posterior distribution of  $(N, G)$  for August 1999 using both a noninformative and an informative prior. We use the July data only to construct an informative prior for August data. First, we use the noninformative prior proposed in Section 2.3.1 and use the method presented in Section 2.3.3.

## 4.1 Noninformative Prior

In this section, we assume that we do not have any informative prior information about  $(N, G)$  for August 1999; then we obtain the posterior distribution of  $(N, G)$  after observing the August data. We use a noninformative prior on  $(N, G)$ , as discussed in Section 2.3.1. Using the `WinBUGS1.3` code, we obtain samples from the posterior distributions of the parameters. As WinBUGS does not allow the use of an arbitrary discrete distribution with unbounded support, we have approximated the discrete prior for  $N$  by truncating it at a large number (e.g. we used  $S + 150$  for all our models). We used several graphical techniques such as, trace plots and ACF plots, as well as numerical diagnostics like Gelman-Rubin that are built into WinBUGS to check the convergence of the Gibbs sampler. After some preliminary studies we decided to run three parallel chains with dispersed starting values for each parameter. For each chain, We used a burn-in of 1000 iterations then generated 5000 post-burn-in samples. Thus, we obtained a total of 15000 final (auto-correlated) samples from the posterior distributions. For each fixed  $r = 1, 2, 3$  and 4, this procedure was followed for the full  $M_{bh}$  model and its submodels. Due to space limitations, we do not present the diagnostic plots for all of the models, but we do show some of these plots for our best fitting model. See Figure 1 and Section 4.2.

[Tables 1 and 2 about here]

The posterior summaries for the full model  $M_{bh}$ , are presented in Table 1. Separately for each  $r$  we present the five-number posterior summary of  $N$  consisting of its mean, standard deviation (sd), 2.5-percentile, median and 97.5-percentile. We also present the five-number posterior predictive summary of MSPE and use the model selection method presented in Section 3. In particular, the values of  $GGC$  given by the posterior mean of MSPE have been highlighted using bold. We also present similar summary results for submodels  $M_h$  in Table 2. First, comparing these two tables we see that the  $GGC$  has a

much smaller value for models involving  $h$  and thus a heterogeneity model is preferred over the models  $M_0$  and  $M_b$ . Next, among the heterogeneity models  $M_{bh}$  and  $M_h$ , we see that GGC picks up a  $M_{bh}$  model with  $r = 2$  as the best model. This also suggests that a higher value of  $r$  is not needed for this data set. *Based on this best fitting model that uses only a noninformative prior, our best estimate of the meadow vole population is  $\hat{N} = 124$ , given by the posterior median.* From these two tables, we also see that an  $M_h$  model if used, would produce an inflated estimate of  $N$ , whereas an  $M_b$  model would produce an estimate of  $N$  that is practically the same as that obtained by the best fitting model for these data. For the best fitting model, Table 1 also shows the five-number posterior summaries of the components of  $G$ , namely the  $\theta$ 's and the  $\pi$ 's. In particular, we would conclude that the heterogeneous population of capture probabilities can be estimated by a 2-point distribution having supports at  $\hat{\theta}_1 = (0.705, 0.319)$  and  $\hat{\theta}_2 = (0.864, 0.803)$  with probability  $\hat{\pi}_1 = 0.519$  and  $\hat{\pi} = 0.481$ , respectively. Notice that for all our final estimates (e.g.,  $\hat{N}$ ,  $\hat{\theta}_i$ , etc.) we have used the posterior median, as such estimates are more robust than the mean when the posterior distribution is skewed.

## 4.2 Informative Prior

In this section we use July 1999 data to construct an informative prior for August 1999 data and obtain the posterior estimates of  $(N, G)$ . This is done by using the method described in Section 2.3.3. We discussed with Jim Nichols, the provider of the meadow vole sampling data, if it would be reasonable to consider July 1999 data as being from a very similar setting as the August 1999 data and thus (July, 1999 data) might provide some prior information for August 1999. He responded that given the common season of the year and similar time during the species' life span, and because no major changes had occurred to the environment between the two sampling times that it may well be a very similar setting for the two sampling times. Thus in this section we

examine utilizing July 1999 data as prior information for August 1999.

[Tables 3 and 4 about here]

As our goal is to compare the estimates for only August 1999 data based on noninformative and informative prior, we do not present estimates on July 1999 data. Separately for each  $r$ , we begin with a noninformative prior (as described in Section 2.1) for the July 1999 and multiply it by the combined likelihoods of the July 1999's data and the uncontaminated August 1999 data to obtain the posterior distribution of August's parameters  $(N, G)$ , based on an informative prior incorporating July 1999 data. We use similar diagnostic checks as described in the previous section to monitor the convergence of our MCMC chains. We present the five-number posterior summary results for August 1999 data in Tables 3 and 4 for  $M_{bh}$  and  $M_h$  models respectively. We also perform model selection using the criterion presented in Section 3. By comparing the results from each model presented in Tables 1 and 2 to those presented in Tables 3 and 4, we see that GGC, via mean of MSPE, prefers the later models slightly more. Other features of the models when using informative priors, such as inflated estimates of  $N$  by  $M_h$  models etc., remain very similar to those when using noninformative priors. Also we notice that the use of the informative prior generally reduces the variability of the posterior distribution (though not substantially for these data sets), thus giving a more informative inference. In summary, from the results presented in Tables 1, 2, 3 and 4 we conclude a that a  $M_{bh}$  model with  $r = 2$  and our informed prior is our selected model.

Thus, referring to the  $r = 2$  results of Table 3, our estimate of the meadow vole population for August 1999 data is  $\hat{N}_{\text{best}} = 125$ , given by the posterior median of  $N$ , with a 95% equal-tail posterior interval estimate of [123, 131]. In addition, we estimate that  $G$  is given by a 2-point distribution having supports at  $\hat{\theta}_1 = (0.671, 0.338)$  and  $\hat{\theta}_2 = (0.860, 0.800)$  with probability  $\hat{\pi}_1 = 0.515$  and

$\hat{\pi}_2 = 0.485$ , respectively. This indicates that the meadow vole population is heterogeneous and possibly consists of two groups; (i) group 1 consisting of animals with lower capture probabilities (e.g., 67.1% and 33.8%) and (ii) group 2 consisting of animals with higher capture probabilities (e.g., 86% and 80%). Also notice that  $\hat{\theta}_{1i} > \hat{\theta}_{2i}$  for  $i = 1, 2$ , which suggests that the animals become trap shy after their first capture. Moreover,  $\hat{\theta}_{21} < \hat{\theta}_{22}$  suggests that animals in the lower capture probability group become more trap shy than those in the higher capture probability group. Finally,  $\hat{\pi}_1 > \hat{\pi}_2$  implies that the proportion of animals in the lower capture probability group (i.e., group 1) is slightly larger than those in the group 2 with higher capture probabilities.

[Figure 1 about here]

In addition to numerical summaries based on the posterior distribution of  $N$ , Figure 1 presents some graphical summaries of  $N$ . For each of the three chains, we overlay the trace plots (top left figure), posterior histograms (top right figure), autocorrelations plots (bottom left figure) and the side by side boxplots of posterior samples of  $N$  (bottom right figure) based on the best fitted model. From the trace plots we see that the chains mixed well, the histograms and boxplots provide empirical evidence of the convergence of the posterior samples from three dispersed starting values. The side-by-side boxplots indicates that chain 1 generated slightly more extreme values of  $N$ , but none were close to the upper bound of  $N$ , set to  $S + 150 = 272$ . The autocorrelation function (ACF) plots indicate that for all three chains the correlation dies off at about lag 18.

## 5 DISCUSSION

In this paper, we have demonstrated how Bayesian analysis can be applied to settings which allow heterogeneity of capture probabilities between animals as

well as bait/trap effects. Some of the major benefits of using Bayesian analysis are as follows:

- (i) the ability to obtain rigorous probability intervals for the parameters in contrast to approximate confidence (not probability) intervals available in non-Bayesian analysis. For example, for the meadow vole population we can say that there is a 95% chance that the population size is between 123 and 131,
- (ii) the ability to incorporate even complex information from previous studies through the use of the sequential version of the Bayes' rule. For instance, we have used the July 1999 data to obtain an informative prior for August 1999 data and the subsequent analysis showed the gain in the information as compared to only using a noninformative prior,
- (iii) the ability to use finite sample based model selection criteria even when competing models are not nested, e.g. GGC can be used to compare models like  $M_b$  and  $M_h$ , which are not nested and
- (iv) the ability to allow different estimates based on ecologically meaningful loss functions, e.g., use of the posterior median versus the posterior mean if the associated loss function is absolute error loss. Given the skewed posterior distribution of  $N$ , we would suggest the use of the posterior median instead of the posterior mean for these kind of data.

In addition, for the use of practitioners, we have provided simple code which uses freely available software; this eliminates many of the barriers which obstruct the use of Bayesian analysis. In our analysis we have not considered models that allow time differences. As a possible extension of our models, it would be worthwhile to study more general models, such as  $M_{th}$ ,  $M_{tb}$  and  $M_{tbh}$ , that allows for time differences as well as heterogeneity and bait/trap effects. Again we emphasize that our proposed *GGC* utilizes finite sample

(non-asymptotic) measures for appropriate model selection among non-nested mixture models.

## ACKNOWLEDGMENTS

The authors would like to thank Jim Nichols for providing interesting data sets presented in the motivating example. The authors are grateful to two referees, an associate editor, and the editor for their insightful comments on the article.

## APPENDIX I

Below we present the WinBUGS code for only the  $M_{bh}$  model. Some minor modifications are needed to obtain the code for the submodels. To obtain the multinomial likelihood as defined in (1), we write it as conditional binomial densities.

Gibbs sampling for Mbh Models

```

model{
#The multinomial likelihood as conditional binomials
#kstar=k(k+1)/2, k=#capture periods
for(j in 1:(kstar-1)){freq[j]~dbin(q[j], Ncondl[j])
  freqnew[j]~dbin(q[j], Ncondl[j])
  resid[j]<-log(freqnew[j]+0.5)-log(freq[j]+0.5)
  q[j]<-p[j]/(1-sum(p[(j+1):kstar]))
  Ncondl[j]<-N-sum(freq[(j+1):kstar])}
freq[kstar]~dbin(p[kstar], N)
freqnew[kstar] ~ dbin(p[kstar], N)
resid[kstar]<-log(freqnew[kstar]+0.5)-log(freq[kstar]+0.5)
for(j in 1:k){for(l in 1:(k-j+1)){for(i in 1:r){
cellprob[i,j,l]<-pi[i]*theta[1,i]*pow(1-theta[1,i],j-1)*
pow(theta[2,i],l-1)*pow(1-theta[2,i],k-j-l+1) }

```

```

    p[1+(j-1)*(2*(k+1)-j)/2]<-exp(logfact(k-j)-logfact(k-j-1+1)-
    logfact(1-1))*sum(cellprob[,j,1])} }
#Prior distribution for N: truncated at S+M
#S=#distinct species captured (sum(freq[]))
N <- S+M
M ~ dcat(prob[])
for(j in 1:Mmax){weight[j] <- pow(S+j, -delta)}
for(j in 1:Mmax){prob[j] <- weight[j]/sum(weight[])}
  for(i in 1:r){thetastar[i] ~ dbeta(b[1], b[2])
  theta[2,i] ~ dbeta(b[1], b[2])
  theta[1,i] <- ranked(thetastar[], i)
#Dirichlet prior as ratio of Gamma's:
  gama[i] ~ dgamma(a,1)
  pi[i] <- gama[i]/sum(gama[]) }
#Model Choice using Gelfand and Ghosh (1998) Criteria
  MSPE <- inprod(resid[],resid[])/kstar
}
Data:
Nichols August 1999 data: list(freq=c(25,30,43,10,10,4),
k=3,r=2,a=1,b=c(0.5, 0.5),delta=1,kstar=6,Nmin=122,Mmax=150)

```

Input your initial values to the run the above code in WinBUGS1.3. *We have not tested this code on the current version WinBUGS1.4.*

## REFERENCES

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki (eds), 267-281. Akademiai Kiado, Budapest.
- Basu, S. and Ebrahimi, N. (2001). Bayesian capture-recapture methods for error detection and estimation of population size: heterogeneity and dependence. *Biometrika*, **88**, 269-279.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer-Verlag, New York.
- Brownie, C., Anderson, D. R., Burnham, K. P., and Robson, D. S. (1985). *Statistical inference from band recovery data: a handbook*, 2nd ed., U.S. Fish and Wildlife Service Resources Publication 156, Washington, D.C.
- Castledine, B. (1981). Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika*, **67**, 197-210.
- Darroch, J. N. (1958). The multiple recapture census. I: Estimation of a closed population. *Biometrika*, **36**, 343-359.
- Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika*, **85**, 1-11.
- George, E. I. and Robert, C. P. (1992). Capture-recapture estimation via Gibbs sampling. *Biometrika*, **79**, (4) 677-683.
- Hoeting, J., Madigan, D., Raftery, A.E. and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, **14**, .

- Ishwaran, H. and Zarepour, M. (2002). Dirichlet Prior Sieves in Finite Normal Mixtures. *Statistica Sinica*, **12**, 941-963.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Statistical Society, Series A*, **186**, 453-461.
- Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration - Stochastic model. *Biometrika*, **52**, 225-247.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models, 2nd ed.*, Chapman and Hall, New York.
- McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*, John Wiley and Sons, Inc., New York.
- Norris, J. L. and Pollock, K. H. (1995). A capture-recapture model with heterogeneity and behavioral response. *Environmental and Ecological Statistics*, **2**, 305-313.
- 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 population. *Wildlife Monographs*, **62**, 135 pp.
- Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models under mixtures. *Biometrics*, **56**, 434-442.
- Schwartz, G. (1978). Estimating the dimension of a model, *Annal of Statistics*. **6**, 461-464.
- Seber, G. A. F. (1965). A note on the multiple-recapture census. *Biometrika*, **52**, 249-259.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2001). WinBUGS User Manual, version 1.4. MRC Biostatistics Unit, Cambridge, UK.  
<http://www.mrc-bsu.cam.ac.uk/bugs>

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit, (with discussion and rejoinder). *Journal of the Royal Statistical Society, Series B*, **64**, 583-639.

Titterton, D. M., Smith, A.F.M., and Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York.

Table 1: Posterior summary of parameters under  $M_{bh}$  model for Nichols data (August, 1999) using noninformative prior

$r = 1(M_b)$	mean	sd	2.5%	median	97.5%
$N$	123.9	1.179	123.0	124.0	127.0
$MSPE$	<b>0.340</b>	0.138	0.070	0.295	0.848
$r = 2$	mean	sd	2.5%	median	97.5%
$N$	124.9	2.78	123.0	124.0	132.0
$\theta_{11}$	0.683	0.099	0.425	0.705	0.817
$\theta_{12}$	0.870	0.066	0.751	0.864	0.997
$\theta_{21}$	0.363	0.273	0.002	0.319	0.984
$\theta_{22}$	0.720	0.278	0.021	0.803	0.999
$\pi_1$	0.507	0.188	0.186	0.519	0.829
$\pi_2$	0.490	0.188	0.171	0.481	0.813
$MSPE$	<b>0.124</b>	0.137	0.014	0.087	0.422
$r = 3$	mean	sd	2.5%	median	97.5%
$N$	134.0	17.93	123.0	127.0	194.0
$MSPE$	<b>0.131</b>	0.135	0.016	0.095	0.431
$r = 4$	mean	sd	2.5%	median	97.5%
$N$	141.2	21.75	123.0	132.0	204.0
$MSPE$	<b>0.138</b>	0.134	0.017	0.104	0.426

(Posterior summaries of the  $\theta$ 's and  $\pi$ 's for rest of the models, i.e., with  $r = 1, 3$  and  $4$  can be obtained from the first author)

Table 2: Posterior summary of parameters under  $M_h$  model for Nichols data (August, 1999) using noninformative prior

$r = 1(M_o)$	mean	sd	2.5%	median	97.5%
$N$	127.8	2.845	123.0	127.0	134.0
$MSPE$	<b>0.550</b>	0.038	0.101	0.472	0.530
$r = 2$	mean	sd	2.5%	median	97.5%
$N$	147.1	19.35	127.0	141.0	205.0
$MSPE$	<b>0.292</b>	0.027	0.024	0.213	1.000
$r = 3$	mean	sd	2.5%	median	97.5%
$N$	156.7	23.49	128.0	149.0	214.0
$MSPE$	<b>0.303</b>	0.028	0.022	0.221	1.040
$r = 4$	mean	sd	2.5%	median	97.5%
$N$	158.5	23.00	129.0	153.0	214.0
$MSPE$	<b>0.323</b>	0.027	0.024	0.232	1.090

(Posterior summaries of the  $\theta$ 's and  $\pi$ 's for these models can be obtained from the first author)

Table 3: Posterior summary of parameters under  $M_{bh}$  model for Nichols data (August, 1999) using informative prior

$r = 1(M_b)$	mean	sd	2.5%	median	97.5%
$N$	124.2	1.082	123.0	124.0	127.0
$MSPE$	<b>0.320</b>	0.129	0.058	0.298	0.641
$r = 2$	mean	sd	2.5%	median	97.5%
$N$	125.3	2.22	123.0	125.0	131.0
$\theta_{11}$	0.678	0.078	0.420	0.671	0.805
$\theta_{12}$	0.873	0.059	0.748	0.860	0.971
$\theta_{21}$	0.365	0.261	0.008	0.338	0.985
$\theta_{22}$	0.711	0.269	0.019	0.800	0.981
$\pi_1$	0.503	0.177	0.189	0.515	0.821
$\pi_2$	0.497	0.175	0.172	0.485	0.810
$MSPE$	<b>0.121</b>	0.129	0.028	0.082	0.418
$r = 3$	mean	sd	2.5%	median	97.5%
$N$	132.0	16.13	123.0	128.0	193.0
$MSPE$	<b>0.129</b>	0.127	0.020	0.099	0.425
$r = 4$	mean	sd	2.5%	median	97.5%
$N$	140.2	20.39	123.0	131.0	200.0
$MSPE$	<b>0.132</b>	0.129	0.015	0.102	0.425

(Posterior summaries of the  $\theta$ 's and  $\pi$ 's for rest of the models, i.e., with  $r = 1, 3$  and  $4$  can be obtained from the first author)

Table 4: Posterior summary of parameters under  $M_h$  model for Nichols data (August, 1999) using informative prior

$r = 1(M_o)$	mean	sd	2.5%	median	97.5%
$N$	127.9	2.372	123.0	127.0	134.0
$MSPE$	<b>0.502</b>	0.032	0.112	0.460	0.520
$r = 2$	mean	sd	2.5%	median	97.5%
$N$	146.1	18.61	125.0	139.0	201.0
$MSPE$	<b>0.250</b>	0.021	0.030	0.191	0.986
$r = 3$	mean	sd	2.5%	median	97.5%
$N$	150.5	20.23	124.0	135.0	209.0
$MSPE$	<b>0.283</b>	0.016	0.027	0.229	1.010
$r = 4$	mean	sd	2.5%	median	97.5%
$N$	155.1	21.72	127.0	150.0	211.0
$MSPE$	<b>0.290</b>	0.019	0.025	0.231	1.016

(Posterior summaries of the  $\theta$ 's and  $\pi$ 's for these models can be obtained from the first author)

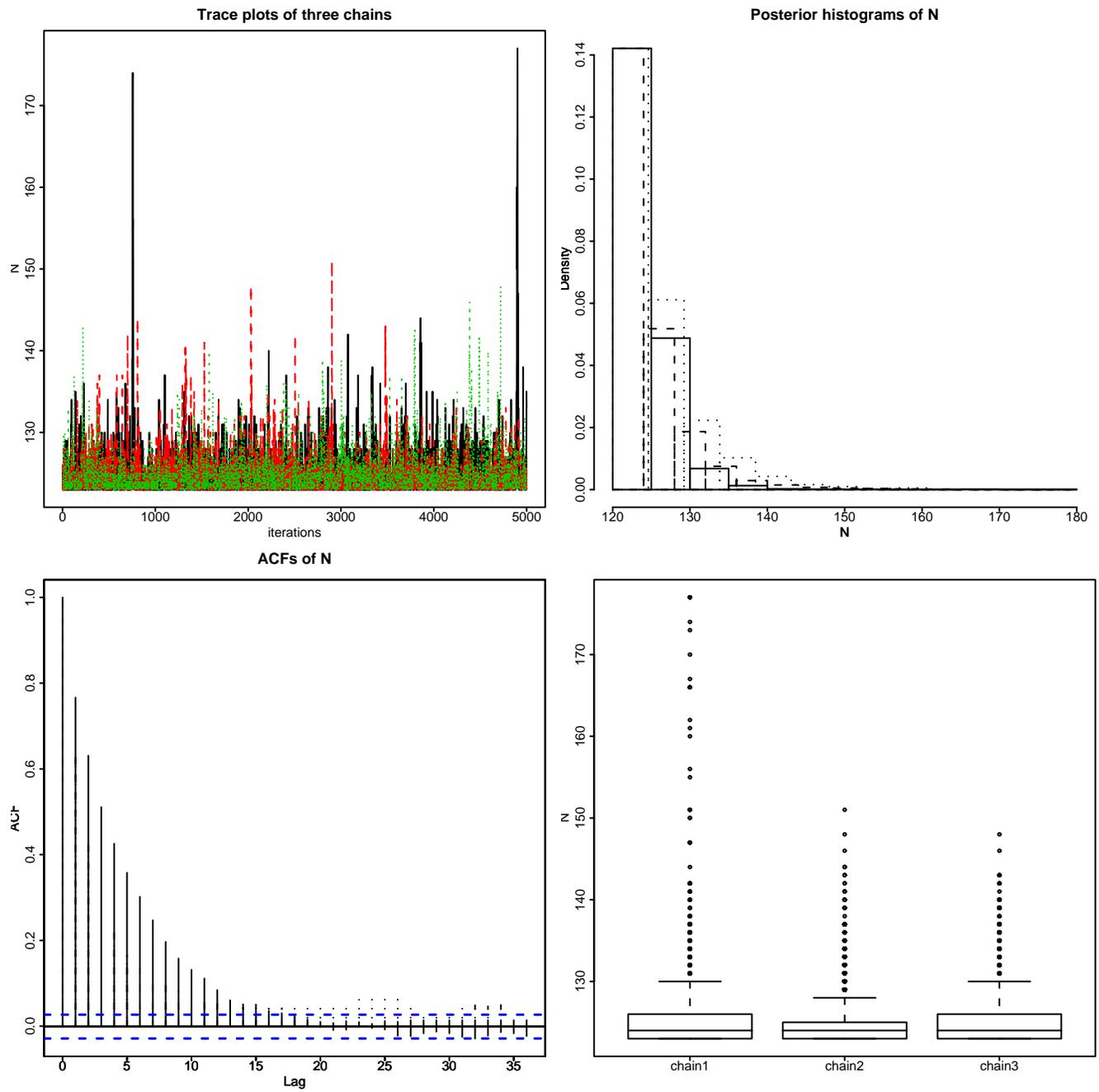


Figure 1: Posterior summary plots of  $N$  for  $M_{bh}$  with  $r = 2$  and informative prior