

Insights Into the Latent Multinomial Model Through Mark-Resight Data on Female Grizzly Bears With Cubs-of-the-Year

Megan D. HIGGS, William A. LINK, Gary C. WHITE,
Mark A. HAROLDSON, and Daniel D. BJORNLI

Mark-resight designs for estimation of population abundance are common and attractive to researchers. However, inference from such designs is very limited when faced with sparse data, either from a low number of marked animals, a low probability of detection, or both. In the Greater Yellowstone Ecosystem, yearly mark-resight data are collected for female grizzly bears with cubs-of-the-year (FCOY), and inference suffers from both limitations. To overcome difficulties due to sparseness, we assume homogeneity in sighting probabilities over 16 years of bi-annual aerial surveys. We model counts of marked and unmarked animals as multinomial random variables, using the capture frequencies of marked animals for inference about the latent multinomial frequencies for unmarked animals. We discuss undesirable behavior of the commonly used discrete uniform prior distribution on the population size parameter and provide OpenBUGS code for fitting such models. The application provides valuable insights into subtleties of implementing Bayesian inference for latent multinomial models. We tie the discussion to our application, though the insights are broadly useful for applications of the latent multinomial model.

Key Words: Bayesian; Discrete uniform; Greater Yellowstone Ecosystem (GYE); Mark-recapture; Population size.

1. INTRODUCTION

Current methods for estimating the number of grizzly bears and setting mortality limits in the Greater Yellowstone Ecosystem (GYE) rely on estimating the number of female

Megan D. Higgs (✉) is Assistant Professor of Statistics, Department of Mathematical Sciences, Montana State University, P.O. Box 172400, Bozeman, MT 59717-2400, USA (E-mail: higgs@math.montana.edu). William A. Link is Mathematical Statistician, Patuxent Wildlife Research Center, U.S. Geological Survey, 12100 American Holly Drive, Laurel, MD 21163, USA. Gary C. White is CWB Professor Emeritus, Department of Fish, Wildlife, and Conservation Biology, 239 Wagar, Colorado State University, Fort Collins, CO 80523, USA. Mark A. Haroldson is Supervisory Wildlife Biologist, Northern Rocky Mountain Science Center, Interagency Grizzly Bear Study Team, U.S. Geological Survey, 2327 University Way, Ste 2, Bozeman, MT 59715, USA. Daniel D. Bjornlie is Large Carnivore Biologist, Large Carnivore Section, Wyoming Game and Fish Department, Lander, WY 82520, USA.

© 2013 International Biometric Society

Journal of Agricultural, Biological, and Environmental Statistics, Accepted for publication

DOI: [10.1007/s13253-013-0148-8](https://doi.org/10.1007/s13253-013-0148-8)

Published online: 11 June 2013

grizzly bears with cubs-of-the-year (FCOY). The method currently employed for obtaining this estimate relies on a convenience sample of ground sightings combined with a more rigorous sample of sightings from aerial surveys. Individuals are not uniquely identifiable, meaning the number of repeat sightings of the same FCOY is unknown. Under the current method, the sightings for a particular year are clustered into sightings thought to be from the same bear using a set of rules relying primarily on distance and time between sightings (Knight, Blanchard, and Eberhardt 1995). Application of this rule set provides an estimated minimum number of FCOY per year and estimates of frequencies of sightings of the different FCOY, which are then used as data to obtain an FCOY population size estimate using the Chao2 (Chao 1989), as recommended by Cherry et al. (2007). It is recognized that this approach produces estimates that are biased low (Schwartz et al. 2008). Additionally, the uncertainty in the estimates, obtained via bootstrap confidence intervals, is likely understated because it does not account for uncertainty inherent in identifying unique individuals from the rule-based clustering method. Therefore, the Interagency Grizzly Bear Study Team (IGBST) is actively pursuing alternative methods to obtain more reliable inference for FCOY abundance. In this paper, we describe a modeling framework and specific methods for utilizing only the data obtained from the aerial surveys conducted each year as part of a traditional mark-resight design. The challenge in implementing this approach is the sparseness of the data on marked animals within each year, as a result of a low number of marked FCOY each year *and* low probabilities of resighting.

Mark-resight population estimators use resighting information on a sample of marked animals, assumed to have the same resighting probabilities as unmarked animals, to estimate the number of unmarked animals in the population (e.g. White 1996). The mark-resight method depends on marking (e.g., using radio or GPS collars) a known number of individuals and then carrying out surveys to record sightings of marked and unmarked animals over a sampling period small enough to assume a closed population. This method is clearly safer, cheaper, and less invasive than traditional mark-recapture methods, which is particularly important when dealing with grizzly bears in an extensive geographic area such as the GYE.

While we do have individually identifiable marks (collars), the sparseness of the data on marked animals introduces additional challenges for implementing models for mark-resight data. This problem is not uncommon in mark-resight or mark-recapture designs where the amount of data from marked animals may be inadequate to compute a useful population estimate. In response to this problem for mark-recapture designs, McKelvey and Pearson (2001) suggest simply using the minimum number known alive as an index, though others argue for aggregation and/or borrowing of information in order to make inference. Models originally developed for estimating ungulate populations borrow information across surveys (e.g., Samuel et al. 1987; Anderson and Lindzey 1996), where a model of sightability is built using marked animals, and then applied to observations of individuals in populations where no marked animals exist, assuming the model captures the range of sightability conditions to which it will be applied. White (2005) and Litt and Steidl (2009) propose strategies for aggregating data in order to fit more complex models. Using software such as program MARK (White and Burnham 1999), data with low num-

bers of marked individuals might be pooled when appropriate to achieve useful estimates (White 2005).

For populations with few marked animals that are rarely sighted, such as FCOY within the GYE, broad assumptions of homogeneity may be needed. We build a model of sightability using marked animals from the same spatially defined population, but combined over many years of sampling, such that resighting data of marked animals from all available years collected under a consistent sampling design are combined to make inference about sightability. This common sightability model is then used for yearly inference based on the varying yearly numbers of sightings of unmarked FCOY. Of course, this strategy of aggregation over sampling occasions (years) assumes homogeneity in sightability of marked animals over time and is not recommended if resighting probabilities or sampling scheme are thought to have changed during the period of interest. For FCOY, we are interested in assessing yearly population size over 16 years during which two aerial surveys were conducted each year, assuming homogeneity in sighting probabilities over the 16 years.

More complex models for mark-resight designs focus on individual-level modeling using available covariates to describe heterogeneity in sighting probabilities among individuals. For example, McClintock et al. (2009b) describe a logit-normal modeling framework to estimate population size when each individual can be sighted at most once (without replacement surveys) and McClintock et al. (2009a) provide a Poisson log-normal model for sampling with replacement. However, these models are designed for studies with individual level covariates, and while this is attractive, it is not always possible to obtain meaningful covariates, and this is indeed the situation for the FCOY. These methods also rely on a normal approximation for the likelihood function, which is not needed for our method.

We appeal to a conceptually simple latent multinomial model for inference, as others have for mark-recapture designs (e.g. Link et al. 2010; Bonner and Holmberg 2013; McClintock et al. 2013). This model connects the logical multinomial distribution capturing the sighting frequencies of marked animals to an analogous, though only partially observed, latent multinomial distribution for the sighting frequencies for unmarked animals. The conceptual connection between the two multinomial distributions is not new, though it is not yet common to directly use the connection for inference, which we accomplish using Bayesian inference and Markov chain Monte Carlo methods, in the same spirit as Link et al. (2010). For most mark-recapture models, recognizing the latent multinomial structure is not necessary to complete the analysis, but as Link and Barker (2010) describe, it provides “conceptual simplicity and an omnibus framework for analysis.” The direct use of the multinomial distribution is natural here because we lack individual covariates leading to the explicit models of individual heterogeneity of detection probabilities, such as those described in McClintock et al. (2009a), McClintock et al. (2009b), and McClintock and Hoeting (2010), and it allows us to proceed with inference without appealing to the normal approximation.

The challenge of using the latent multinomial model to obtain a posterior distribution for a population size within an MCMC framework comes from ensuring the sampling algorithm is adequately exploring the set of latent multinomial frequencies consistent with

Table 1. Data for the entire region are provided in the left side of the table, giving the number of marked (collared) FCOY available for resighting in each year, the number of marked animals sighted zero, one, and two times (Y_0, Y_1, Y_2), and the total number of sightings of unmarked FCOY S . The right side of the table provides the same information excluding marked FCOY available on moth sites and sightings of unmarked animals within 500 m of moth sites (nm = no moth).

| Year | marked | Y_0 | Y_1 | Y_2 | S | marked _{nm} | $Y_{0,nm}$ | $Y_{1,nm}$ | $Y_{2,nm}$ | S_{nm} |
|-------|--------|-------|-------|-------|-----|----------------------|------------|------------|------------|----------|
| 1997 | 6 | 4 | 2 | 0 | 16 | 6 | 4 | 2 | 0 | 4 |
| 1998 | 4 | 2 | 2 | 0 | 26 | 4 | 2 | 2 | 0 | 7 |
| 1999 | 6 | 5 | 1 | 0 | 7 | 6 | 5 | 1 | 0 | 7 |
| 2000 | 7 | 7 | 0 | 0 | 16 | 7 | 7 | 0 | 0 | 11 |
| 2001 | 9 | 4 | 5 | 0 | 32 | 8 | 4 | 4 | 0 | 19 |
| 2002 | 5 | 5 | 0 | 0 | 65 | 5 | 5 | 0 | 0 | 30 |
| 2003 | 4 | 3 | 1 | 0 | 25 | 4 | 3 | 1 | 0 | 7 |
| 2004 | 4 | 2 | 2 | 0 | 35 | 4 | 2 | 2 | 0 | 20 |
| 2005 | 3 | 3 | 0 | 0 | 22 | 3 | 3 | 0 | 0 | 14 |
| 2006 | 8 | 7 | 0 | 1 | 43 | 7 | 7 | 0 | 0 | 24 |
| 2007 | 6 | 3 | 3 | 0 | 45 | 5 | 3 | 2 | 0 | 20 |
| 2008 | 5 | 3 | 1 | 1 | 42 | 5 | 3 | 1 | 1 | 20 |
| 2009 | 6 | 6 | 0 | 0 | 28 | 6 | 6 | 0 | 0 | 14 |
| 2010 | 3 | 3 | 0 | 0 | 38 | 3 | 3 | 0 | 0 | 24 |
| 2011 | 3 | 2 | 1 | 0 | 28 | 3 | 2 | 1 | 0 | 16 |
| 2012 | 6 | 3 | 3 | 0 | 36 | 5 | 3 | 2 | 0 | 12 |
| Total | 85 | 62 | 21 | 2 | 504 | 81 | 62 | 18 | 1 | 249 |

the observed data. We present an easily programmed Gibbs sampling algorithm, as well as show how these latent multinomial analyses can be conducted in popular software such as OpenBUGS (Lunn et al. 2009) or WinBUGS (Lunn et al. 2000). Difficulties in using such software for this problem has probably contributed to the rare use of the latent multinomial model in practice, despite its conceptual simplicity. The coding trick we present is not only useful here, but is applicable to other discrete latent variable models.

We also contribute to the understanding of appropriate prior distributions for the latent multinomial model for unknown population sizes by examining unfortunate behavior of the commonly used discrete uniform prior.

The problem of estimating the number of female grizzly bears with cubs-of-the-year in the Greater Yellowstone Ecosystem from limited data is clearly an important ecological problem and, from a statistical perspective, it provides a canvas for improving our understanding and ability to implement inference based on the latent multinomial model.

2. MARK-RESIGHT DATA FOR FEMALE GRIZZLIES WITH CUBS-OF-THE-YEAR

Before discussing the modeling framework, we provide details of the sampling design for our application. The Interagency Grizzly Bear Study Team (IGBST) has conducted two intensive aerial surveys per annum since 1997 to record sightings of FCOY (collared and uncollared) obtained without the aid of telemetry. The first round of flights occurs between late May and early August, depending on location, and the second round occurs in late

August through September. Each flight lasts approximately two hours and generally occurs before 10:00 a.m. In 1997, the Grizzly Bear Recovery Zone (USFWS 1993) was divided into 23 bear observation areas, each intended to be surveyed twice each year. In 1998, 11 additional units were delineated, primarily within 10 miles of the Grizzly Bear Recovery Zone boundary, and three existing areas were split into two observation areas, for a total of 37 bear observation units. In 2007, an additional 9 bear observation areas were added in the southern portion of the Yellowstone Ecosystem, and two more were added in 2009 on the Wind River Indian Reservation when the tribes joined the IGBST, for a total of 48 bear observation units. Despite the increase in area surveyed, the collared FCOY have remained in or within 10 miles of the original recovery zone surveyed in 1997 and out of a total 507 sightings of unmarked FCOY as of 2012, only 43 were obtained outside the recovery zone. This consistency in the area over which FCOY are sighted helps to justify the assumption of homogeneity in sighting probabilities over time. The IGBST has maintained a collared (marked) sample of between 3 and 9 FCOY per year, many of which are never resighted in the aerial surveys (Table 1).

A fundamental assumption, providing the logical basis for mark-resight inference, is that marked animals are representative of unmarked animals in terms of their sighting probabilities. Specifically, this means the expected proportion of marked animals resighted 0, 1, or 2 times should be the same as the proportions we expect for the N_u unmarked animals available in the population. For approximately eight weeks in the summer, many Yellowstone grizzly bears feed on cutworm moths in remote, high elevation areas. While on or near the moth sites, bears are easily sighted during the aerial surveys; the flights correspond to when FCOY are actively using the cutworm moths as a food source. In fact, approximately half of the sightings of unmarked FCOY during the sixteen years were within 500 meters of identified moth sites. However, due to the remoteness and ruggedness of the moth sites, there are very few FCOY collared near enough to the moth sites to be resighted on them. This creates a clear violation of the assumption that the expected sighting rates for marked FCOY are equal to those for the unmarked FCOY. The implications of this violation on inference are serious, resulting in overestimation of the number of FCOY in the population because unmarked animals are actually sighted far more frequently than marked, solely due to the moth sites.

While we would like to be able to include an indicator or distance to moth sites as a covariate to account for this heterogeneity, because there are essentially no marked FCOY available to be sighted around moth sites it is not feasible with the current data. In summer 2012, a combination of ground and flight surveys around moth sites made researchers confident very few FCOY available within the 500 m buffer around identified moth sites were being missed during the aerial surveys. Therefore, researchers are looking into other methods to combine moth site counts with results from the mark-resight analysis based on data excluding moth sites. For marked bears, the data change very little when excluding FCOY available around moth sites because there are so few collared FCOY in the region. However, this does remove one of the only two FCOY over the sixteen years seen on both survey flights (Table 1), changing the observed proportion of marked FCOY sighted twice from 0.023 to 0.012. Our goal is to make inference for the population of FCOY not using the moth-sites by excluding moth-site data from both the marked and unmarked FCOY.

3. METHODS

Latent multinomial models have recently received attention in the analysis of mark-recapture data with misidentification (Link et al. 2010; Bonner and Holmberg 2013; McClintock et al. 2013). In these applications, an observation S (counts of captures or sightings) is obtained as a known function f of an unobserved multinomial random variable Z , based on the study design. Thus, while one usually cannot determine the realized value of Z , one can evaluate whether specific values Z_* are consistent with an observation, by checking whether $S = f(Z_*)$. Maximum likelihood for a parameter vector θ can be conducted, in principle, using

$$\mathcal{L}(\theta|S) = \sum_{Z_*: S=f(Z_*)} \Pr(Z_*|\theta). \quad (3.1)$$

However, Bayesian methods based on

$$[Z, \theta|S] \propto \Pr(Z|\theta) \times \mathbf{I}(S = f(Z)) \times [\theta]$$

are more easily implemented, and have the added benefit of allowing easy and reliable inference for Z .

We present a simple, yet novel, example of the latent multinomial model. We provide advice for choosing prior distributions and details for programming a Gibbs sampler, as well as code for implementing the model within OpenBUGS. We hope these contributions will make the general methods for discrete latent variable models more widely accessible.

3.1. THE LATENT MULTINOMIAL MODEL

Our methods apply to counts of sightings of unmarked individuals over K surveys within a sampling period, where individuals are sighted an unknown number of times and the goal is to estimate the number of unmarked animals. We assume auxiliary data are available, typically from marked individuals, providing information about the frequencies we expect individuals to be sighted.

We model the sighting frequencies for marked and unmarked animals in year t as $(K + 1)$ -cell multinomial random variables $Y_t = (Y_{0,t}, Y_{1,t}, \dots, Y_{K,t})'$ and $Z_t = (Z_{0,t}, Z_{1,t}, \dots, Z_{K,t})'$. Vector Y_t is observed, with known index m_t (the number of marked animals); vector Z_t is latent, with unknown index $N_{u,t}$ (the number of unmarked animals). The total number of sightings of unmarked individuals in a sampling period S_t may contain duplicate sightings because unmarked animals are not individually identifiable; we may see as many as S_t distinct individuals, or (if K divides S_t) as few as S_t/K distinct individuals. The count S_t is connected to the latent multinomial Z_t by the known function

$$S_t = Z_{1,t} + 2Z_{2,t} + \dots + KZ_{K,t} = \sum_{k=0}^K kZ_{k,t}. \quad (3.2)$$

Information from sightings of marked individuals informs $N_{u,t}$ under the assumption that the expected sighting rates are the same for marked and unmarked animals. We further assume the expected rates are constant over time. Thus, multinomials Y_t and Z_t have

common parameter $\boldsymbol{\pi} = (\pi_0, \pi_1, \dots, \pi_K)'$, not depending on t , and we may aggregate the information from marked individuals into a single multinomial $\mathbf{Y} = \sum_t \mathbf{Y}_t$ with index $M = \sum_t m_t$. The complete data likelihood (CDL) is

$$[\mathbf{Z}_t, \boldsymbol{\pi}, N_{u,t} | S_t, \mathbf{Y}, M] \propto [\mathbf{Z}_t | \boldsymbol{\pi}, N_{u,t}] [\mathbf{Y} | \boldsymbol{\pi}, M] \mathbf{I}\left(S_t = \sum_{k=0}^K k Z_{k,t}\right).$$

The main challenge in the analysis based on the CDL lies in defining the set of vectors of \mathbf{Z}_t consistent with an observed S_t . If we wish to sample the full conditional distribution $[\mathbf{Z}_t | \boldsymbol{\pi}, N_{u,t}, S_t, \mathbf{Y}, M]$ using a random walk algorithm, we must ensure all elements in the feasible set (i.e., those values of \mathbf{Z}_t consistent with the observation S_t) are reachable and that the algorithm can efficiently move among elements in the feasible set. For small values of $K > 1$, defining the feasible set is not particularly difficult, but it does add a challenge to programming, and the challenge gets larger as K increases. The general idea behind finding feasible sets for mark-recapture problems, as well as useful notation, are provided in Link et al. (2010) and Link and Barker (2010).

For $K = 2$, the only admissible vectors $\mathbf{Z}_t = (Z_{0,t}, Z_{1,t}, Z_{2,t})'$ are those such that $S_t = Z_{1,t} + 2Z_{2,t}$. It is straightforward to enumerate the possible pairs (Z_1, Z_2) consistent with an observed value of S . If S is even, there are $(S/2) + 1$ possible pairs, namely $\{(0, (S/2)), (2, (S/2) - 1), (4, (S/2) - 2), \dots, (S, 0)\}$, and if S is odd, there are $(S - 1)/2 + 1$ elements in the feasible set, namely $\{(1, (S - 1)/2), (3, ((S - 1)/2) - 1), (5, ((S - 1)/2) - 1), \dots, (S, 0)\}$.

A Gibbs sampling algorithm based on the CDL is relatively straightforward to implement; we provide details in Section 3.3. We also recognize the appeal of being able to implement the analysis in readily available software such as OpenBUGS (Lunn et al. 2009), WinBUGS (Lunn et al. 2000), or JAGS (Plummer 2003). Programming the latent multinomial model requires some creativity to ensure the feasible set of \mathbf{Z} is properly sampled; we provide details for OpenBUGS in Section 3.4. The implementation uses a computational trick we believe to be of general interest, allowing Gibbs sampling subject to prescribed constraints on parameter values.

Alternative strategies for inference use a likelihood function defined in terms of S (as (3.1)), and avoid sampling of the latent variable. Further details for a Gibbs sampler based on (3.1) are provided in Appendix B. Yet another strategy utilized in recent work (e.g., McClintock et al. 2009a, 2009b) uses a normal approximation for the likelihood based on S with expectation and variance obtained through the connection to the latent multinomial \mathbf{Z} . This normal approximation can easily be implemented in OpenBUGS, WinBUGS, or JAGS and example code is provided in Appendix B. While the approximation works well for our application, more investigation into its behavior is warranted before proposing its general use.

Bayesian analysis using the CDL requires the further specification of priors for $\boldsymbol{\pi}$ and N_u , to which we now turn our attention.

3.2. PRIORS FOR N_u AND π

The Dirichlet family of priors is conjugate for the parameter vector π of an observed multinomial random variable such as \mathbf{Y} ; given a Dirichlet prior with parameter vector \mathbf{a} , denoted $\mathcal{D}(\mathbf{a})$, the posterior distribution for π is $\mathcal{D}(\mathbf{a} + \mathbf{Y})$. For our analysis of the FCOY data, we chose to use $\mathbf{a} = (1/2, 1/2, 1/2)'$, the Jeffreys prior. Analysis of the frequencies for only the marked animals $\mathbf{Y} = (62, 18, 1)'$ with this prior leads to the posterior $\mathcal{D}(62.5, 18.5, 1.5)$. Using this as an informative prior for π in the analysis of S_t 's alone is equivalent to using the Jeffreys prior and including \mathbf{Y} as another source of data in the analysis.

The selection of an objective prior on a binomial (or multinomial) sample size such as N_u boils down to two choices, a discrete uniform (DU) prior over a sufficiently large range or the scale prior suggested by Jeffreys $[N] \propto 1/N$ where $N \in \{0, 1, \dots\}$. Considering the more general form $[N] \propto 1/N^c$, the DU corresponds to $c = 0$ and the scale prior corresponds to $c = 1$. The DU does hold an initial allure by assigning equal weight to all practically possible values of N_u , though the scale prior has more attractive analytical properties (Berger, Bernardo, and Sun 2008; Link 2013). Lunn et al. (2013) briefly discuss priors distributions on the positive integers, providing both options as reasonable alternatives, likening the scale prior to the commonly used prior for σ^2 in the normal models, $[\sigma^2] \propto 1/\sigma^2$. Berger, Bernardo, and Sun (2008) explain that Jeffreys suggested this prior for any unbounded, positive parameter, regardless of whether the parameter space is continuous or discrete. The discrete uniform prior is widely used, but we believe the scale prior is superior because it avoids paradoxical and unacceptable behavior manifested by the discrete uniform, described in Appendix A.

For our application, the effect of the discrete uniform prior is profound when simultaneously analyzing the $T = 16$ years, as it changes the posterior distribution for π_0 , and therefore $N_{u,t}$. Intuitively, S_t contains no information about the number of animals never sighted $Z_{0,t}$, and therefore it is clearly undesirable behavior for S_t to influence our inference about π_0 and it is particularly disturbing for inference to depend on we utilize the training data (number of years we analyze simultaneously). Hence, we use and recommend the scale prior $N_{u,t} \propto 1/N_{u,t}$. Note, S_t does contain information (albeit little) regarding π_1 and π_2 through its informing of the feasible set of $(z_{1,t}, z_{2,t})$ pairs such that $S_t = z_{1,t} + 2z_{2,t}$.

For many potential users of such models, the ability to implement the scale $1/N_u$ prior in common software is important, and this is possible through introduction of an additional nuisance parameter ψ . If we let $W|\psi \sim \text{Geometric}(\psi)$ and $\psi \sim \text{Beta}(0, 1)$, then $[W + 1] \propto 1/(W + 1)$, and we get our desired prior by letting $N_u = W + 1$. While we cannot set the parameter of the Beta distribution equal to zero, we can make it sufficiently small that the derivation does work in practice. To implement this in OpenBUGS over multiple years, we set $N_{u,t} = W_t + 1$ and then specify

$$[W_t|\psi_t] = \text{Geometric}(\psi_t) = \text{Negative Binomial}(\psi_t, 1),$$

where $\psi_t \sim \text{Beta}(0.001, 1)$ (Figure 1). When programming one's own Gibbs sampler, the prior is clearly easy to implement.

3.3. GIBBS SAMPLING ALGORITHM USING \mathbf{Z}

The Gibbs sampling algorithm samples in turn from full (or complete) conditional distributions of the parameters and latent variables ($\boldsymbol{\pi}$, N_u , and \mathbf{Z}) (for ease of notation, we suppress dependence on t for now). Since S is simply a deterministic function of the latent variable \mathbf{Z} , when we explicitly use \mathbf{Z} in the algorithm, the distribution of S conditional on \mathbf{Z} is simply given by a binary function indicating if \mathbf{Z} falls within the feasible set of values defined by the observed data, $[S|\mathbf{Z}, N_u, \boldsymbol{\pi}] = I(S = \sum_{k=0}^K kZ_k)$, as introduced in Section 3.1. Similarly, the values of \mathbf{Z} and N_u are linked by $N_u = \sum_{k=0}^K Z_k$. Therefore, we consider the joint full conditional distribution of (\mathbf{Z}, N_u) within the Gibbs sampling algorithm, similar to that described in Link and Barker (2010):

$$[\mathbf{Z}, N_u | \boldsymbol{\pi}, S, \mathbf{Y}, M] \propto [S|\mathbf{Z}, N_u, \boldsymbol{\pi}][\mathbf{Z}|N_u, \boldsymbol{\pi}][N_u]$$

$$= I\left(S = \sum_{k=0}^K kZ_k\right) \left\{ \frac{N_u!}{\prod_{k=0}^K Z_k!} \prod_{k=0}^K \pi_k^{Z_k} \right\} I\left(N_u = \sum_{k=0}^K Z_k\right) p(N_u). \quad (3.3)$$

To complete the Gibbs sampler, we also need draws from the full conditional distribution of $\boldsymbol{\pi}$, which is the product of two multinomial distributions with different indexes, but sharing the same vector of probabilities, multiplied by the prior for $\boldsymbol{\pi}$,

$$[\boldsymbol{\pi} | S, \mathbf{Z}, N_u, \mathbf{Y}, M] \propto [\mathbf{Z}|N_u, \boldsymbol{\pi}][\mathbf{Y}|M, \boldsymbol{\pi}][\boldsymbol{\pi}]. \quad (3.4)$$

The following steps describe implementation of a Gibbs sampling algorithm for years $t = 1, \dots, T$ using the prior distributions for N_u and $\boldsymbol{\pi}$ discussed in Section 3.2.

1. Construct the feasible set of $(Z_{1,t}, Z_{2,t})$ pairs for S_t for all t .
2. Initialize $(Z_{1,t}, Z_{2,t})^{\text{curr}}$ from the feasible set of latent frequencies for all t .
3. Initialize $Z_{0,t}^{\text{curr}}$ (and thus $N_{u,t}^{\text{curr}}$) at a reasonable values for all t .
4. Draw $\boldsymbol{\pi}^{\text{curr}}$ from its full conditional distribution under use of the $\mathcal{D}(a_0, a_1, a_2)$ prior:

$$[\boldsymbol{\pi}] = \mathcal{D}\left(Y_0 + a_0 + \sum_{t=1}^T Z_{0,t}, Y_1 + a_1 + \sum_{t=1}^T Z_{1,t}, Y_2 + a_2 + \sum_{t=1}^T Z_{2,t}\right).$$

5. Update \mathbf{Z}_t (and thus $N_{u,t}$) using Metropolis–Hastings steps to sample from the full conditional distribution for all t .
 - (a) Draw candidate values $(Z_{1,t}, Z_{2,t})$ by sampling from the feasible set. This can be done independent of current values by simply selecting a pair with equal probability, or a more efficient sampler can be constructed by randomly selecting a direction to move within the feasible set, centered on the current pair, and with move distance chosen by tuning the algorithm. For our application with relatively small feasible sets, moving a single step within the feasible set is adequate for most years. Candidate values falling outside the feasible set are formally rejected (i.e. algorithm stays at the current value).

- (b) Draw a candidate value for $Z_{0,t}$ using a discrete uniform distribution centered on the current value $Z_{0,t}^{\text{curr}}$, with the width for each year obtained by tuning the algorithm. If the chosen candidate value is negative, it is rejected. A non-symmetric proposal distribution can also be used, ensuring candidate values are never negative and consistent with the right skewness we expect in the posterior distributions for $N_{u,t}$. For example, a negative binomial distribution with mean equal to $Z_{0,t}^{\text{curr}}$ and size adjusted relative to S_t can nicely explore the parameter space. With this proposal distribution, however, the asymmetry must be accounted for in the M–H ratio by including the ratio $\Pr(Z_{0,t} = Z_{0,t}^{\text{curr}} | Z_{0,t}^{\text{cand}}) / \Pr(Z_{0,t} = Z_{0,t}^{\text{cand}} | Z_{0,t}^{\text{curr}})$.
 - (c) Set $N_{u,t}^{\text{cand}} = \sum_{k=0}^2 Z_{k,t}^{\text{cand}}$.
 - (d) Evaluate the joint complete conditional distribution for $(\mathbf{Z}_t, N_{u,t})^{\text{cand}}$ using (3.3). Our candidate proposal methods ensure $S_t = \sum_{k=0}^K k Z_{k,t}$ and $N_{u,t} = \sum_{k=0}^K Z_{k,t}$. Therefore, the numerator of the M–H ratio for each t simply evaluates the Multinomial p.m.f. with parameters $N_{u,t}^{\text{cand}}$ and $\boldsymbol{\pi}^{\text{curr}}$ at $\mathbf{Z}_t^{\text{cand}}$ which is multiplied by $1/N_{u,t}^{\text{cand}}$. The denominator has the same form, though uses $N_{u,t}^{\text{curr}}$ and $\mathbf{Z}_t^{\text{curr}}$. If the asymmetric proposal distribution is used, the M–H ratio must be multiplied by the ratio in (b).
6. Let $r_t = \min(1, \text{M–H ratio}_t)$ and accept candidate values with probability r_t by drawing u_t from a $\text{Unif}(0, 1)$ distribution and accepting candidate values for which $u_t < r_t$.
 7. Repeat steps 4 through 6 until adequate convergence.

3.4. IMPLEMENTATION IN OPENBUGS

The implementation in OpenBUGS requires some creativity and we now give attention to the code in Figure 1. First, note the indices for the elements of $\boldsymbol{\pi}$, \mathbf{Z} , and \mathbf{Y} are shifted from 0, 1, 2 to 1, 2, 3. To simplify the programming, we parameterize \mathbf{Z} using a binomial decomposition to express $[\mathbf{Z}|N_u]$ as the product of two binomial distributions and an indicator variable:

$$[\mathbf{Z}|N_u] = [Z_2|N_u] \times [Z_1|Z_2, N_u] \times I(Z_0 = N_u - Z_2 - Z_1),$$

where $[Z_2|N_u] = \text{Binomial}(N_u, \pi_2)$, $[Z_1|Z_2, N_u] = \text{Binomial}((N_u - Z_2), \frac{\pi_1}{(1-\pi_2)})$, and $Z_0 = N_u - Z_2 - Z_1$.

The trick to implementing the latent multinomial model in OpenBUGS lies in restricting samples of \mathbf{Z}_t to the feasible set defined by the observed S_t . Ideally, we would simply like to define $\text{theta}[t] = \text{good}[t]$ (see Figure 1) as the indicator of whether the latent vector \mathbf{Z}_t is consistent with our observation S_t , i.e., $\theta_t = I(S_t = Z_{1,t} + 2Z_{2,t})$, and then treat θ_t as the success parameter of a Bernoulli trial ($\text{constraint}[t]$) for which we have observed a success. Our specification of a successful Bernoulli trial would then rule out $\theta_t = 0$, and hence leads to immediate rejection of values \mathbf{Z}_t which are inconsistent with the data S_t . However, direct implementation of this idea does not allow for movement of the

```

model{
  pi[1:3] ~ ddirich(a[1:3])
  pi.c[2] <- pi[2]/(1-pi[3])

  for (t in 1:T){
    #Binomial decomposition of Z
    z[t,3] ~ dbin(pi[3],N.u[t])
    z[t,2] ~ dbin(pi.c[2],N.z[t,2])
    N.z[t,2] <- N.u[t] - z[t,3]
    z[t,1] <- N.u[t] - z[t,3] - z[t,2]

    #Prior for N.u : [N] propto 1/N
    psi[t] ~ dbeta(0.001,1)
    q[t] ~ dnegbin(psi[t],1)
    N.u[t] <- q[t] + 1

    #Latent structure 'z' agreement with observations 'S'
    constraint[t] <- 1
    constraint[t] ~ dbern(theta[t])
    theta[t] <- step(1.5 - abs(z[t,2]+2*z[t,3]-S[t]))
    good[t] <- equals(z[t,2]+2*z[t,3],S[t])
  }
}

```

Figure 1. OpenBUGS code for the latent multinomial model with $K = 2$ and T years of data. Values for T , $a[1]$, $a[2]$, $a[3]$ and $S[t]$ must be specified as data; initial values for $\psi[t]$ are usually needed.

\mathbf{Z} 's around their parameter space. Any proposed change to $Z_{1,t}$ without a corresponding change to $Z_{2,t}$ leads to immediate rejection of the candidate value.

Instead, we define $\text{theta}[t]$ as an indicator for the event that $Z_{1,t} + 2Z_{2,t}$ is within 1 of the desired value S_t . This change allows the chain for \mathbf{Z}_t to move around the parameter space, though it also results in accepting \mathbf{Z}_t 's that are not in the feasible set. However, keeping track of $\text{good}[t]$ allows us to easily post-process our samples to select only the $N_{u,t}$'s for which $\text{good}[t]$ is equal to 1.

4. RESULTS

4.1. RESULTS FOR FCOY ABUNDANCE

We used the Gibbs sampling algorithm described in Section 3.3 to make inference for the 16 years of data on females with cubs-of-the-year in the Greater Yellowstone Ecosystem, implementing samplers in both R (R Development Core Team 2009) and Gauss (Table 2).

The model forces a linear relationship between S_t and the location of the distribution of $N_{u,t}$, with an additional 5 sightings of unmarked FCOY shifting the center of distribution by approximately 20 animals. As expected, the variance increases as S_t increases, making

Table 2. Posterior means, medians, and 0.025 and 0.975 quantiles of the posterior distributions of $N_{u,t}$ and π obtained from the Gibbs sampler under the $1/N$ prior for each $N_{u,t}$ and the Dirichlet(0.5, 0.5, 0.5) prior for π . Therefore, there is no sharing of information about N_u among years through prior distributions. The sampler was programmed in R and run for two million iterations, saving every tenth iteration. The results are obtained after removing the first 5000 saved observations, for a total of 195,000 samples used to compute the summaries in the table.

| Year | S | Mean | Median | $Q_{0.025}$ | $Q_{0.975}$ |
|---------|----|-------|--------|-------------|-------------|
| 1997 | 4 | 16.1 | 14 | 5 | 37 |
| 1998 | 7 | 28.1 | 26 | 11 | 57 |
| 1999 | 7 | 28.2 | 26 | 11 | 57 |
| 2000 | 11 | 44.3 | 42 | 21 | 83 |
| 2001 | 19 | 76.4 | 73 | 41 | 133 |
| 2002 | 30 | 121.8 | 116 | 70 | 200 |
| 2003 | 7 | 28.2 | 26 | 11 | 57 |
| 2004 | 20 | 80.5 | 77 | 44 | 138 |
| 2005 | 14 | 56.3 | 53 | 28 | 101 |
| 2006 | 24 | 96.4 | 92 | 54 | 164 |
| 2007 | 20 | 80.4 | 77 | 44 | 138 |
| 2008 | 20 | 80.5 | 77 | 44 | 138 |
| 2009 | 14 | 56.3 | 53 | 28 | 101 |
| 2010 | 24 | 96.5 | 92 | 54 | 164 |
| 2011 | 16 | 64.2 | 61 | 33 | 114 |
| 2012 | 12 | 48.4 | 46 | 23 | 89 |
| π_0 | — | 0.760 | 0.761 | 0.662 | 0.845 |
| π_1 | — | 0.220 | 0.221 | 0.140 | 0.318 |
| π_2 | — | 0.020 | 0.014 | 0.001 | 0.056 |

it more difficult to sample the upper tail of the posterior distribution. Care should be taken in assessing convergence in the tails of the distribution to get the reported upper quantile for $N_{u,t}$, beyond simply relying on common convergence criteria such as the Gelman and Rubin Statistic (\hat{R}) (Gelman et al. 2004).

As desired, the posterior distribution for π is very close to what we obtain using only the training data from marked FCOY, where the posterior distribution is Dirichlet(62.5, 18.5, 1.5). Recall from Section 3.2 we are actually using this as the prior for π when making inference on $N_{u,t}$. Therefore, the learning about π from information in S_t is very minimal, and is based only on the use of the observed S_t in defining the feasible set of $(Z_{1,t}, Z_{2,t})$ pairs. For our application, the learning is barely noticeable, but in general will depend on the values in π .

4.2. EVALUATION OF ASSUMPTIONS

As previously discussed, we are assuming the data from each year are informing a single, common sightability model (i.e. homogeneity of π over years). That is, the proportion of marked bears sighted 0, 1, or 2 times remains constant (or close to it) over all years included in the aggregation, 1997 to 2012. Of course, even if a common sightability model holds, we expect yearly variation in the observed resighting frequencies based on the number of collared FCOY available and random variability. Under our assumed model, the

observed resighting frequencies should be consistent with random draws from a multinomial distribution with probabilities obtained from the aggregated data and multinomial sample sizes equal to the yearly number of collared FCOY. The number of marked FCOY available in any year (excluding moth sites) ranges from three to eight. The single FCOY sighted two times within one year came in 2008 with five marked FCOY available, and the zeros for other years are not unusual under the homogeneity assumption. The 18 sightings included in Y_1 are nicely spread over the 16 years, with six years contributing no resightings, four years contributing one resighting, five contributing two resightings, and 2001 contributing four resightings from the largest number of FCOY available. If, for example, all contributions to Y_1 were the result of large counts in only two or three years, we would be very hesitant to assume π was constant over time. However, we do not see any obvious indications that the homogeneity assumption is unreasonable.

For our application, there is a small chance an animal could be sighted more than one time during a single round of flights. However, no collared FCOY has ever been observed more than one time during a single flight, suggesting sightings with a single round of flights are of unique FCOY. We also assume a closed population within the time frame of the two aerial surveys in each year. We know this has been violated by temporary emigration from the sampling area surveyed by two collared females over the sixteen years, though in general feel this assumption is reasonably satisfied due to the typical size of home ranges for FCOY. We also have information about mortalities that can be incorporated for the marked animals.

An assumption of independence among years is also violated because some females are collared and available in multiple years. Out of the 81 “marked animals” considered for the analysis, there are 64 different individuals, with 50 of the 64 available in only one year, 12 available in two years, one each available for three and four years. However, the years are generally not consecutive because grizzly bears typically have cubs every three years. Some FCOY sighted in one year are not sighted in another and the sighting probability is in general so low we do not feel this violation of independence is severely affecting inference. The probabilities of resighting are very close to those obtained using only the 50 individuals collared in only a single year.

5. INVESTIGATION USING SIMULATED DATA

In this section, we briefly investigate the variability in results for N_u obtained over many realizations of data generated under the latent multinomial model with known N_u and π . We set $N_u = 50$ because a population of FCOY of this size is small enough to trigger possible changes in management strategies. Starting from $M = 81$, $N_u = 50$, and $\pi = (0.765, 0.222, 0.013)$, we generated 300 sets of resighting frequencies for marked and unmarked animals under the model described in Section 3.1 and obtained posterior distributions for $N_{u,t}$ using the Gibbs sampling algorithm (Figure 2). Thus, we are capturing variability due to different stochastic realizations of data from marked and unmarked animals under a constant π . While the distribution of posterior means is centered on the true value, the variability among posterior distributions obtained under the different realizations is relatively large, with posterior means for N_u ranging from 23.4 to 90.3.

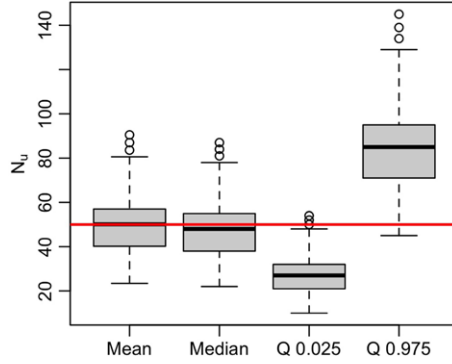


Figure 2. Box-plots of the posterior means, medians, 0.025 quantiles, and 0.975 quantiles over 300 realizations simulated under $N_u = 50$ (horizontal line), $M = 81$, and $\pi = (0.765, 0.222, 0.013)$.

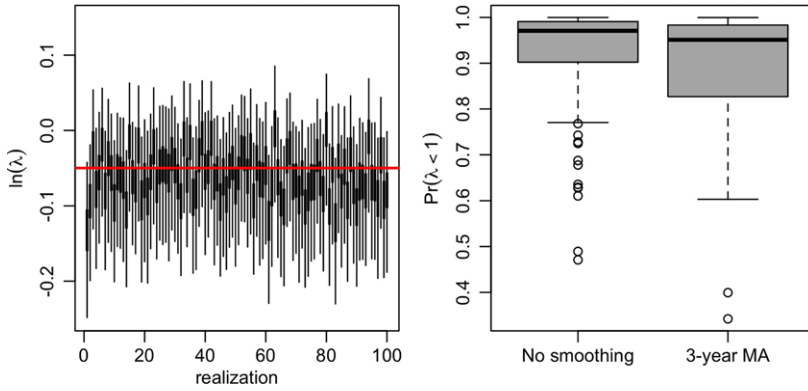


Figure 3. (Left) Box-plots of posterior distributions of the trend coefficient obtained from 100 realizations of 11 years of simulated data starting at $N_u = 70$ and declining 5 % per year (rounded to nearest integer). The horizontal line represents the data generating value of $\ln(\lambda) = -0.05$ ($\lambda = 0.95$). (Right) Probability $\lambda < 1$ calculated from the 100 posterior distributions using the unsmoothed distributions of $N_{u,t}$ and those obtained using the 3-year moving average.

6. CONSIDERATIONS FOR TREND DETECTION

Researchers hoping to use these data to inform management decisions are interested in the yearly posterior distributions, but also desire a less-noisy picture of trend over time. For example, managers are required to think about mortality limits and from a practical perspective these limits should not vary widely between adjacent years. A desirable property of any estimator of abundance is the ability to detect changes in trajectory of population size over time. Therefore, the substantial variability in the posterior distributions of N_u demonstrated in Section 5 has important implications for inference about FCOY in the GYE, making it difficult to accurately characterize trend of the population over time to inform management decisions regarding the population of grizzly bears in the GYE. For the population of female grizzlies with cubs-of-the-year, a decrease in the population of 5 % per year ($\ln(\lambda) = -0.05$) is considered large enough to raise a red flag. Using the same strategy described above, we generated 100 realizations of 11 years of data under a

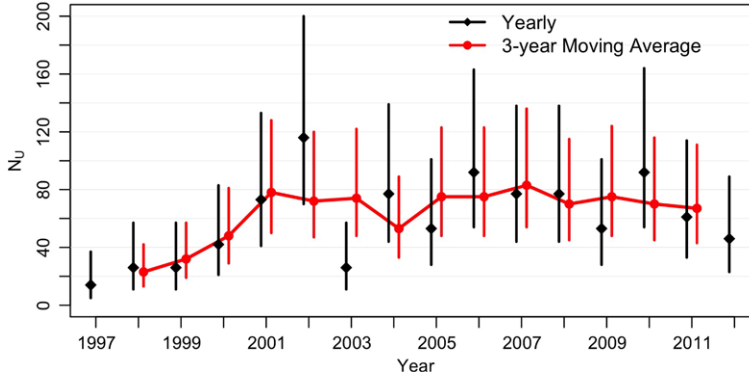


Figure 4. 95 % posterior intervals and medians for N_u assuming independent years with objective priors (from results in Table 2) and 95 % posterior intervals for 3-year moving averages connected by the line.

5 % reduction in population size each year, starting with 70 FCOY and ending with 42. Posterior distributions of the trend coefficient obtained as the slope over the 11 years of $\ln(N_{u,t})$ extend well above zero *and* well below the data-generating value of -0.05 , making it difficult for most realizations to distinguish between a flat trend and the practically meaningful decreasing trend (Figure 3). However, for most of the 100 posterior distributions the $\Pr(\lambda < 1)$ calculated from the posterior draws is greater than 0.90.

The current method for smoothing the results over time is to fit a linear trend model and a quadratic trend model to the Chao2 estimates over all available years (1983–2012), and then use the model averaged estimate as the smoothed estimate for the current year (Harris et al. 2007). While there are many strategies for smoothing varying greatly in their sophistication, here we choose to investigate a simple moving average. We do not present this as the best, or only, method, but as a simple, transparent method easily tied to grizzly bear reproduction. Female grizzly bears typically reproduce every three years after two years of caring for their cubs, naturally leading to consideration of a three year moving average as a method of investigating trend over time. The three year moving average is just a simple function of $N_{u,t}$,

$$N_{u,t}^{3\text{yr}} = \frac{1}{3} \sum_{j=-1}^1 N_{u,t-j} \quad \text{for } t = 2, \dots, T-1, \quad (6.1)$$

and can therefore be calculated for each posterior draw to quantify uncertainty in the moving averages. The posterior means and 95 % posterior intervals are compared to the original results in Figure 4. In terms of detecting the 5 % decrease in population size per year, the 3-year moving averages do not noticeably improve our ability to detect the negative trend when considered over all realizations. For many realizations it does pull the posterior distribution of the trend coefficient closer to the true value, but for sets of more unusual realizations the moving average actually pulls the distribution further from the truth than the raw posterior distributions for $N_{u,t}$. If this method is in fact adopted by the IGBST, further investigation into trend detection will be conducted and we may pursue more model-based methods, such as implementing an autoregressive time series model or specifying

priors for $N_{u,t}$ depending on $N_{u,t-1}$. However, a first-order autoregressive model does not smooth the results to the extent desired by the researchers, and it is difficult to specify an appropriate prior distribution based on the previous year's posterior distribution resulting reasonable and biologically meaningful results. It is clear from this preliminary examination that other sources of information are needed to decrease the variability among years currently hindering detection of trends.

7. DISCUSSION

With this paper, we provide timely insights and suggestions for implementing the latent multinomial model in terms of prior distributions and computational details. The insights are applicable to the fitting of latent multinomial models beyond our specific application, and highlight some of the subtleties of fitting this conceptually simple model.

For comparison, we provide estimates and 95 % confidence intervals using the Poisson log-normal mixed effects model (McClintock et al. 2009a) and the estimator of Bowden and Kufeld (1995), both fit using the aggregated sightability data (Figure 5). Bowden and Kufeld (1995) incorporate individual heterogeneity from the radio-marked sample by computing the variation in number of detections across animals to account for extra-binomial variation, and we assume the 81 marked animals were part of the population when re-sightings were taking place, and then subtract out the marked animals from the estimate and confidence interval bounds to produce an estimate of unmarked population size for the annual aerial surveys. While the posterior means and medians from the latent multinomial model differ little from the point estimates obtained from the other methods, the methods do vary in their posterior/confidence intervals. Bowden's method leads to smaller lower and upper bounds, whereas the Poisson log-normal model is consistent with the latent multinomial in the lower bound, but has an upper bound extending beyond the upper bound of the latent multinomial model. Therefore, while the other methods provide useful point estimates, they may not be as adequate for quantifying uncertainty around those estimates.

While this method provides a useful foundational piece for inference, the IGBST is currently investigating other sources of information and new data collection methods to improve estimation of the FCOY population size. Future improvements or extensions could also include more explicit incorporation of other variables affecting sightability, such as time of day and land cover. However, the challenge of this application remains the very low sightability of animals. In fact, during radio-telemetry flights flown for the purpose of locating collared bears, only about 10 % of bears located via telemetry are actually visually confirmed. A logical extension, and planned future work, is to better connect the data and process models by explicitly modeling yearly variation and trend over time with pertinent covariates and connecting this to the demographic models used to make inference about other sex and age classes.

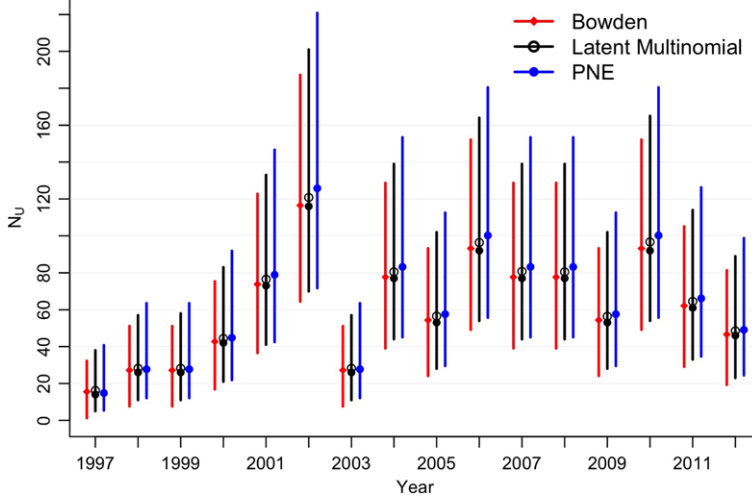


Figure 5. Point estimates and 95 % intervals for the Bowden method (left), the latent multinomial (center) with posterior means (open circles) and medians (closed circles), and the Poisson log-normal mixed effects model (right).

APPENDIX A: COMPARISON OF DISCRETE UNIFORM AND SCALE PRIORS FOR N_u

For our application, we analytically compare the posterior distribution of π under the discrete uniform and scale priors for N_u , where any differences will manifest as different posterior distributions for the parameters of interest $N_{u,t}$. To compare the two candidate priors on N_u , we begin by noting a characteristic of the Dirichlet prior for π . If $D \sim \mathcal{D}(a_0, a_1, a_2)$, then

$$D \stackrel{d}{=} \left(\frac{X_0}{X_0 + X_1 + X_2}, \frac{X_1}{X_0 + X_1 + X_2}, \frac{X_2}{X_0 + X_1 + X_2} \right),$$

where X_i are independent gamma random variables with shape parameters a_i and constant rates. We can rewrite this as

$$\begin{aligned} D &\stackrel{d}{=} \left(\frac{X_0}{X_0 + X_1 + X_2}, \frac{X_1}{X_1 + X_2} \cdot \frac{X_1 + X_2}{X_0 + X_1 + X_2}, \frac{X_2}{X_1 + X_2} \cdot \frac{X_1 + X_2}{X_0 + X_1 + X_2} \right) \\ &= (\theta_0, \theta_1(1 - \theta_0), (1 - \theta_1)(1 - \theta_0)), \end{aligned}$$

and then it can be shown that $\theta_0 \sim \beta(a_0, a_1 + a_2)$ and $\theta_1 \sim \beta(a_1, a_2)$ are independent random variables. Therefore, we may reparameterize the Dirichlet in terms of only θ_0 and θ_1 , denoting the two beta priors as $b_0(\theta_0)$ and $b_1(\theta_1)$.

Now, consider the posterior distribution obtained based on analysis of a single S_t , suppressing notational dependence on t . The joint posterior for θ_0 , θ_1 , N_u , Z_1 , and Z_2 is proportional to

$$\frac{N_u!}{(N_u - z_1 - z_2)!z_1!z_2!} \theta_0^{N_u - z_1 - z_2} [\theta_1(1 - \theta_0)]^{z_1} [(1 - \theta_1)(1 - \theta_0)]^{z_2}$$

$$\times \frac{1}{N_u^c} b_0(\theta_0) b_1(\theta_1) \mathbf{I}(S = z_1 + 2z_2).$$

Then, if we marginalize over N_u (summing over all possible values for $N_u = z_1 + z_2, z_1 + z_2 + 1, z_1 + z_2 + 2, \dots$) we obtain a simple result for either $c = 0$ or $c = 1$:

$$[\theta_0, \theta_1, Z_1, Z_2 | S] \propto \frac{(z_1 + z_2 - c)!}{z_1! z_2!} (1 - \theta_0)^{c-1} \theta_1^{z_1} (1 - \theta_1)^{z_2} b_0(\theta_0) b_1(\theta_1) \mathbf{I}(S = z_1 + 2z_2). \quad (8.1)$$

Factorization of $[\theta_0, \theta_1, Z_1, Z_2 | S]$ into $h_0(\theta_0)[\theta_1, Z_1, Z_2 | S]$, with $h_0(\theta_0) \propto (1 - \theta_0)^{c-1} \times b_0(\theta_0)$, demonstrates the posterior distribution for θ_0 does not depend on S . For $c = 1$ it is the same as the original beta prior $\beta(a_0, a_1 + a_2)$, as originally specified, but for $c = 0$ the posterior for θ_0 becomes $\beta(a_0, a_1 + a_2 - 1)$. The paradoxical result is that if we use $c = 0$ (the DU prior) to analyze more than one year (more than one S_r), the resulting inferences for $N_{u,t}$'s will differ from what would have been obtained if analyzing the sampling periods separately. Recall the information from the marked animals (training data) remains constant over all years.

This strange behavior of the posterior distribution for θ_0 , resulting from the choice $c = 0$, gets worse as the number of sampling periods we analyze simultaneously increases. For T , years of data each with an independent discrete uniform prior on $N_{u,t}$, the posterior for θ_0 becomes $\beta(a_0, a_1 + a_2 - T)$. The information content of a $\beta(r, s)$ random variable is often measured by the sum $r + s$, and therefore the use of independent DU priors for $N_{u,t}$'s leads to a paradoxical loss of knowledge; indeed, if $T \geq a_1 + a_2$, the posterior is improper. Similar unacceptable behavior of the discrete uniform prior for N_u has been noted elsewhere in analysis of binomial data (Link 2013). For our application, the effect of the discrete uniform priors is profound as we have $a_1 + a_2 = 20$ (combining the prior with Y_1 and Y_2) and $T = 16$, changing our desired $\beta(62.5, 20)$ prior into a $\beta(62.5, 4)$ posterior! Use of the scale prior ($c = 1$) avoids such problems.

APPENDIX B: INFERENCE WITHOUT DIRECTLY SAMPLING Z

B.1. GIBBS SAMPLING ALGORITHM

The Gibbs sampling algorithm can also be accomplished without direct sampling of Z using the likelihood defined by marginalizing over Z , or summing over all possible Z in the feasible set defined by the observed S , as in (3.1). The full conditional distributions of the unknown parameters, N_u and π , then depend on this likelihood function and are sampled from using the Metropolis–Hastings algorithm, where we compare evaluations of $[S | N_u, \pi, M, Y]$ at the observed value of S under candidate to current values of N_u and π , also incorporating the prior distributions. The feasible set of (Z_1, Z_2) pairs consistent with each observed S must still be defined to obtain the likelihood function, and we more explicitly write out (3.1) here as

$$[S | N_u, \pi] = \sum_{Z_*: S = Z_{*1} + 2Z_{*2}} \left(\frac{N_u!}{Z_{*0}! Z_{*1}! Z_{*2}!} \right) \pi_0^{Z_{*0}} \pi_1^{Z_{*1}} \pi_2^{Z_{*2}}. \quad (9.1)$$

```

model{
  p[1:K] ~ ddirich(a[1:K])

  Y[1:3] ~ dmulti(p[1:3],N.m)

  for (t in 1:T){
    E.S[t] <- N.u[t]*(p[2] + 2*p[3]) #Expected value of S
    Var.S[t] <- N.u[t]*(p[2]*(1-p[2]) + 4*p[3]*(1-p[3])
                      -2*2*p[2]*p[3])
    prec.S[t] <- pow(Var.S[t], -1)
    S[t] ~ dnorm(E.S[t], prec.S[t])

    #Prior for N.u : [N] propto 1/N
    psi[t] ~ dbeta(0.001,1)
    q[t] ~ dnegbin(psi[t],1)
    N.u[t] <- q[t] + 1
  }
}

```

Figure 6. JAGS or BUGS code for the normal approximation to S with $K = 2$ for $T = 16$ years of data with independent objective scale priors for $N_{u,t}$.

B.2. NORMAL APPROXIMATION

For those who would rather not program their own Gibbs sampling algorithm, the use of approximations to enable more automatic fitting in available software, is appealing. For this setting, the normal distribution can be a useful approximation to the distribution of $S|N, \pi, Y, M$ (obtained by marginalizing over Z), although S is defined on the non-negative set of integers which is clearly inconsistent with the support of a Gaussian random variable. However, this normal approximation has been used recently in a mark-resight context in McClintock et al. (2009a, 2009b), where McClintock et al. (2009a) use a normal distribution left-truncated at zero to solve the non-negative problem, but the integer-valued support and upper bound on S are still ignored. While we prefer the direct analysis with no approximations, the convenience of the approximation is appealing and it does work well for our application. However, we do not recommend its broad use before testing it for a particular application, as we expect its appropriateness to depend on π and K .

The approximation is implemented by finding the expectation and variance of S_t , easily derived from the connection to the latent multinomial Z_t , and then parameterizing a normal distribution with mean and variance equal to the expectation and variance. Dropping

notational dependence on t , the expected value is

$$E(S) = E\left(\sum_{k=0}^K k Z_k\right) = \sum_{k=0}^K k E(Z_k) = N_u \sum_{k=0}^K k \pi_k, \quad (9.2)$$

and the variance of is

$$\begin{aligned} \text{var}(S) &= \text{var}\left(\sum_{k=0}^K k Z_k\right) \\ &= \sum_{k=0}^K k^2 \text{var}(Z_k) + 2 \sum_{k \neq k'} k k' \text{cov}(Z_k, Z_{k'}) \\ &= \sum_{k=0}^K k^2 N_u \pi_k (1 - \pi_k) - 2 \sum_{k \neq k'} k k' N_u \pi_k \pi_{k'} \\ &= N_u \left(\sum_{k=0}^K k^2 \pi_k (1 - \pi_k) - 2 \sum_{k \neq k'} k k' \pi_k \pi_{k'} \right). \end{aligned}$$

The variance was obtained using the following properties of the multinomial distribution: $\text{var}(Z_k) = N_u \pi_k (1 - \pi_k)$ and $\text{cov}(Z_k, Z_{k'}) = -N_u \pi_k \pi_{k'}$ for $k \neq k'$. Then, for our application with $K = 2$,

$$E(S) = N_u (\pi_1 + 2\pi_2) = 2N_u \left(\frac{\pi_1}{2} + \pi_2 \right) \quad \text{and}$$

$$\text{var}(S) = N_u [\pi_1 (1 - \pi_1) + 4\pi_2 (1 - \pi_2) - 4\pi_1 \pi_2].$$

To implement the normal approximation, we simply use the expected value and variance of S as the mean and variance:

$$S_{\text{Norm}}^* \sim N(E(S), \text{var}(S)).$$

This is easily implemented in software such as WinBUGS, OpenBUGS, and JAGS (Plummer 2003), with example code provided in Figure 6.

ACKNOWLEDGEMENTS

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We especially thank C.C. Schwartz, for his efforts in data collection and program direction during his tenure as Study Team Leader during 1997–2011. We thank S. Cherry for providing a critical review of the manuscript as part of the USGS Fundamental Science Practices. We thank D. Stradley, R. Stradley, S. Monger, S. Ard, G. Lust, C. Tyrrel, A. Spencer, D. Stinson, T. Schell, K. Overfield, J. Martin, N. Cadwell and G.E. Hyatt, who have flown grizzly bear radio-tracking and observation flights in the Greater Yellowstone Ecosystem. We also thank the many people from member agencies of the Study Team who were responsible for grizzly bear capture and radio-collaring, and those individuals who provided us with sightings of females with cubs-of-the-year. We also thank Jim Clark for his insightful and helpful review.

[Received March 2012. Accepted May 2013.]

REFERENCES

- Anderson, C., and Lindzey, F. G. (1996), "Moose Sightability Model Developed from Helicopter Surveys," *Wildlife Society Bulletin*, 24, 247–259.
- Berger, J., Bernardo, J., and Sun, D. (2008), "Reference priors for discrete parameter spaces," Technical report, Department of Statistical Science, Duke University.
- Bonner, S. J., and Holmberg, J. (2013, in press), "Mark-recapture with multiple non-invasive marks," *Biometrics*.
- Bowden, D. C., and Kufeld, R. C. (1995), "Generalized Mark-Sight Population-Size Estimation Applied to Colorado Moose," *The Journal of Wildlife Management*, 59, 840–851.
- Chao, A. (1989), "Estimating Population Size for Sparse Data in Capture-Recapture Experiments," *Biometrics*, 45, 427–438.
- Cherry, S., White, G. C., Keating, K. A., Haroldson, M. A., and Schwartz, C. C. (2007), "Evaluating Estimators for Numbers of Females with Cubs-of-the-Year in the Yellowstone Grizzly Bear Population," *Journal of Agricultural, Biological, and Environmental Statistics*, 12, 195–215.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis* (2nd ed.), Boca Raton: Chapman & Hall/CRC.
- Harris, R. B., White, G. C., Schwartz, C. C., and Haroldson, M. A. (2007), "Population Growth of Yellowstone Grizzlies: Uncertainty, Correlation, and Future Monitoring," *Ursus*, 18, 167–177.
- Knight, R. R., Blanchard, B. M., and Eberhardt, L. L. (1995), "Appraising Status of the Yellowstone Grizzly Bear Population by Counting Females with Cubs-of-the-Year," *Wildlife Society Bulletin*, 23, 245–248.
- Link, W. A. (2013, in press), "A cautionary note on the discrete uniform prior for the binomial N," *Ecology*.
- Link, W. A., and Barker, R. J. (2010), *Bayesian Inference with Ecological Applications*, London: Elsevier.
- Link, W. A., Yoshizaki, J., Bailey, L. L., and Pollock, K. H. (2010), "Uncovering a Latent Multinomial: Analysis of Mark-Recapture Data with Misidentification," *Biometrics*, 66, 178–185.
- Litt, A. R., and Steidl, R. J. (2009), "Improving Estimates of Abundance by Aggregating Sparse Capture-Recapture Data," *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 228–247.
- Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2013), *The BUGS Book*, Boca Raton: CRC Press, Taylor & Francis Group.
- Lunn, D. J., Spiegelhalter, D. J., Thomas, A., and Best, N. (2009), "The BUGS Project: Evolution, Critique, and Future Directions," *Statistics in Medicine*, 28, 3049–3067.
- Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), "WinBUGS—a Bayesian Modelling Framework: Concepts, Structure, and Extensibility," *Statistics and Computing*, 10, 325–337.
- McClintock, B. T., Conn, P., Alonso, R., and Crooks, K. R. (2013), "Integrated Modeling of Bilateral Photo-Identification Data in Mark-Recapture Analyses," *Ecology*. doi:[10.1890/12-1613.1](https://doi.org/10.1890/12-1613.1).
- McClintock, B. T., and Hoeting, J. A. (2010), "Bayesian Analysis of Abundance for Binomial Sighting Data with Unknown Number of Marked Individuals," *Environmental and Ecological Statistics*, 17, 317–332.
- McClintock, B. T., White, G. C., Antolin, M. F., and Tripp, D. W. (2009a), "Estimating Abundance Using Mark-Resight When Sampling Is with Replacement or the Number of Marked Individuals Is Unknown," *Biometrics*, 65, 237–246.
- McClintock, B. T., White, G. C., Burnham, K. P., and Pryde, M. A. (2009b), "A Generalized Mixed Effects Model of Abundance for Mark-Resight Data When Sampling Is Without Replacement," in *Modeling Demographic Processes in Marked Populations*, New York: Springer, pp. 271–289.
- McKelvey, K. S., and Pearson, D. E. (2001), "Population Estimation with Sparse Data: the Role of Estimators Versus Indices Revisited," *Canadian Journal of Zoology*, 79, 1754–1765.
- Plummer, M. (2003), "JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling."
- R Development Core Team (2009), *R: a Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing, URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- Samuel, M. D., Garton, E. O., Schlegel, M. W., and Carson, R. G. (1987), "Visibility Bias During Aerial Surveys of Elk in North-Central Idaho," *The Journal of Wildlife Management*, 51, 622–630.

- Schwartz, C. C., Haroldson, M. A., Cherry, S., and Keating, K. A. (2008), "Evaluation of Rules to Distinguish Unique Female Grizzly Bears with Cubs in Yellowstone," *The Journal of Wildlife Management*, 72, 543–554.
- USFWS (1993), *Grizzly bear recovery plan*, U.S. Fish and Wildlife Service, Missoula, Montana, USA.
- White, G. C. (1996), "Noremark: Population Estimation from Mark-Resighting Surveys," *Wildlife Society Bulletin*, 24, 50–52.
- (2005), "Correcting Wildlife Counts Using Detection Probabilities," *Wildlife Research*, 32, 211–216.
- White, G. C., and Burnham, K. P. (1999), "Program Mark: Survival Estimation from Populations of Marked Animals," *Bird Study*, 46, 120–139.