Biometrics 64, 1170–1177
December 2008

DOI: 10.1111/j.1541-0420.2008.00987.x

Bayesian Analysis of Wildlife Age-at-Harvest Data

Paul B. Conn,[1,2,*] Duane R. Diefenbach,[3] Jeffrey L. Laake,[4] Mark A. Ternent,[5] and Gary C. White[6]

[1]Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology,
Colorado State University, Fort Collins, Colorado 80523, U.S.A.
[2]Current address: National Marine Fisheries Service, NOAA Southeast Fisheries Science Center, 101 Pivers Island
Road, Beaufort, North Carolina 28516, U.S.A.
[3]U.S. Geological Survey, Pennsylvania Cooperative Fish and Wildlife Research Unit, Pennsylvania State
University, 113 Merkle Lab, University Park, Pennsylvania 16802, U.S.A.
[4]National Marine Mammal Laboratory, Alaska Fisheries Science Center, National Marine Fisheries Service, 7600
Sand Point Way, NE, Seattle, Washington 98115-0070, U.S.A.
[5]Pennsylvania Game Commission, 2001 Elmerton Ave., Harrisburg, Pennsylvania 17110, U.S.A.
[6]Department of Fish, Wildlife, and Conservation Biology, Colorado State University,
Fort Collins, Colorado 80523, U.S.A.
[*]*email:* Paul.Conn@noaa.gov

Summary. State and federal natural resource management agencies often collect age-structured harvest data. These data represent finite realizations of stochastic demographic and sampling processes and have long been used by biologists to infer population trends. However, different sources of data have been combined in ad hoc ways and these methods usually failed to incorporate sampling error. In this article, we propose a "hidden process" (or state-space) model for estimating abundance, survival, recovery rate, and recruitment from age-at-harvest data that incorporate both demographic and sampling stochasticity. To this end, a likelihood for age-at-harvest data is developed by embedding a population dynamics model within a model for the sampling process. Under this framework, the identification of abundance parameters can be achieved by conducting a joint analysis with an auxiliary data set. We illustrate this approach by conducting a Bayesian analysis of age-at-harvest and mark-recovery data from black bears (*Ursus americanus*) in Pennsylvania. Using a set of reasonable prior distributions, we demonstrate a substantial increase in precision when posterior summaries of abundance are compared to a bias-corrected Lincoln–Petersen estimator. Because demographic processes link consecutive abundance estimates, we also obtain a more realistic biological picture of annual changes in abundance. Because age-at-harvest data are often readily obtained, we argue that this type of analysis provides a valuable strategy for wildlife population monitoring.

Key words: Abundance; Age-at-harvest; Black bear; Cohort model; Mark-recovery model; Recruitment; State-space model; Survival.

1. Introduction

The age, class, and sex of harvested animals are often collected for monitoring populations as part of wildlife management programs. Such data are readily collected and used to inform management decisions. However, stochasticity associated with sampling and demographic processes is often ignored, and trends in harvest data are often used to infer population trends. Unfortunately, this approach may lead to flawed inferences about population status if trends in harvest data are related to trends in harvest or reporting rates rather than abundance trends. For instance, even with a standardized reporting system, hunter reporting rates have changed over time for the white-tailed deer harvest (*Odocoileus virginianus*) reporting system in Pennsylvania (Rosenberry, Diefenbach, and Wallingford, 2004).

Robust procedures for analyzing age-at-harvest data typically require auxiliary data to help model the harvest process (Deriso, Quinn, and Neal, 1985; Gove et al., 2002). In fisheries, statistical catch-at-age models use data from research vessel surveys or catch-effort surveys to help estimate model parameters (Fournier and Archibald, 1982; Dupont, 1983; Deriso et al., 1985). Alternatively, data from marked animals may be employed more directly to estimate the parameters (such as survival and sampling probability) needed to explain changes in age-at-harvest data over time or space (e.g., Maunder, 2001; Gove et al., 2002). The latter approach may be more useful in terrestrial wildlife applications because abundance is typically lower and the probability of encountering marked animals is substantially higher.

We describe a Bayesian approach to modeling age-at-harvest data for wildlife populations when auxiliary data from a marking study are available to help model the harvest process. We are motivated by a sampling problem involving black bear (*Ursus americanus*) in Pennsylvania, where age-structured tagging and harvest records were available. One way of incorporating such information is through the use of a state-space or "hidden process" model (Newman et al., 2006). This type of approach is increasingly common in fisheries applications (e.g., McAllister and Ianelli, 1997; Meyer and Millar, 1999; Lewy and Nielson, 2003), but has seen less use in wildlife (but see Besbeas et al., 2002; Buckland et al., 2004; Newman et al., 2006). Previous approaches to analyzing age-at-harvest data for terrestrial species have typically relied on multinomial models for cohort abundance through time, with initial population size as an index and success probabilities determined by functions of survival and harvest parameters (Laake, 1992; Gove et al., 2002). Although a reasonable and useful starting place, it is difficult to incorporate fecundity/recruitment, a "+" age class, or random effects based on frequentist analysis with this type of model structure.

In the following development, we start by describing a general model for the analysis of age-at-harvest data that remove these restrictions. Conceptually, it is composed of two parts: an observation model and a population dynamics model. After constructing an appropriate likelihood and specifying a set of prior distributions, we carry out a Bayesian analysis using Markov chain Monte Carlo (MCMC). We illustrate the application of our model with a demographic analysis of female black bears in Pennsylvania from 1986 to 1999; data from a mark-recovery study during the same time period were used to inform the estimation of survival and harvest parameters.

2. Model Development

2.1 *Data Requirements*

The fundamental data needed for age-at-harvest analysis are an age-at-harvest matrix, **C**, which summarizes annual harvests by sex and age class. We assume that there is no error associated with aging techniques up to some threshold age, A, after which individuals are grouped into a "+" category. For the purposes of this article, we further assume that **C** only includes data from the female portion of the population, although extensions to males are relatively straightforward. We also assume that the investigator has additional data from marked individuals to help model the processes of survival and harvest.

2.2 *Models for the Sampling Process*

We assume that age-at-harvest counts are realizations of a stochastic process, so that we may write

$$[\mathbf{C} \mid \mathbf{N}, \mathbf{h}] = \mathrm{I}(\mathbf{C}) \prod_{t=1}^{T} \prod_{a=1}^{A-1} [C_{ta} \mid N_{ta}, h_{ta}] \times \left[C_{1A}^{+} \,\middle|\, N_{1A}^{+}, h_{1A}\right] \prod_{t=2}^{T} \left[C_{tA}^{+} \,\middle|\, N_{tA} + N_{t,A+1}^{+}, h_{tA}\right], \tag{1}$$

where $N_{ta}^{+} = \sum_{a}^{\infty} N_{ta}, C_{ta}^{+} = \sum_{a}^{\infty} C_{ta}$, $[G \mid H]$ gives the conditional distribution of G given H, **N** denotes the vector of all N_{ta} variables for which $t > 1$ ($t = 2, 3, \ldots, T$; $a = 1, 2, \ldots, A + 1$), **h** denotes the vector of all h_{ta} parameters ($t = 1, 2, \ldots, T$; $a = 1, 2, \ldots, A$), and $\mathrm{I}(\mathbf{C})$ is an indicator function used to disallow cohort abundances that are less than harvest counts:

$$\mathrm{I}(\mathbf{C}) = \prod_{t=1}^{T-1} \prod_{a=1}^{A-1} \mathrm{I}(C_{ta} \leq N_{ta} - N_{t+1,a+1}) \times \prod_{t=1}^{T-1} \mathrm{I}\big(C_{tA}^{+} \leq N_{tA} + N_{t,A+1}^{+} - N_{t+1,A+1}^{+}\big).$$

Remaining notation is defined in Table 1. Because all survivors from N_{1A}^{+} will be part of $N_{2,A+1}^{+}$ the following year, there is no need for a $N_{1,A+1}^{+}$ parameter. The $N_{t,A+1}^{+}$ latent variables are needed in later years because of the nature of the survival process and the need for the final two abundance variables to share the same success probability. The need for this construct should be more apparent when examining the joint probability mass function (PMF) for abundance in the following section.

Table 1

Definitions of parameters, latent variables, and statistics used in the joint age-at-harvest, mark-recovery likelihood

Parameters and variables	
S_{ta}	Probability that an age a individual survives to time t+1 given it was alive at time t
h_{ta}	Probability that an age a individual is harvested and reported to wildlife personnel in [t,t+1], given that it was alive at time t
f_{ta}	Per breeder recruitment rate over [t, t+1], with reference to the number of age a breeders in the population at time t and the number of new recruits at time t+1
N_{ta}	Number of age a individuals in the population in year t immediately prior to harvest. The N_{1a} are parameters while the remaining $N_{ta}(t > 1)$ are treated as latent variables
N_{ta}^{+}	$\sum_{a}^{\infty} N_{ta}$
Statistics	
C_{ta}	Number of age a individuals that are harvested and reported to wildlife personnel in year t
C_{tA}^{+}	$\sum_{a=A}^{\infty} C_{ta}$
M	Total number of individuals marked and released over the course of the experiment
H_k	Encounter history for individual k
t_{k1}	Year in which animal k is first captured, marked, and released
t_{k2}	Year in which animal k is harvested and reported, if encountered again
I_k	Indicator variable equal to 1 if animal k is harvested and reported at some time, 0 otherwise
a_{kt}	Age of animal k at time t
A	Age at which an individual's age cannot be reliably distinguished from older age classes
T	Duration of the study (e.g., years)

We intentionally formulated (1) to be compatible with the mark-recovery model proposed by Brownie et al. (1985). This formulation assumes that the PMF for $[C_{ta}\,|\,N_{ta},\ h_{ta}]$ is given by

$$C_{ta} \sim \text{Binomial}(N_{ta}, h_{ta}).$$

Other formulations are certainly possible. For instance, Conn (2007) considered the case where harvest counts were conditional on the number of individuals that die in an interval, which is consistent with the mark-recovery model given by Seber (1982). One reviewer outlined an alternative observation model consistent with the work of Gove et al. (2002) which took into account both harvest and natural mortality. Both of these approaches impose natural constraints on harvest and natural mortality parameters and remove the need to incorporate indicator functions into the observation model. However, the Brownie parameterization is flexible with regard to the timing of harvest and natural mortality, and leads to better MCMC mixing than with the Seber parameterization (Conn, 2007). The potential for $\hat{h}_{ta} > (1 - \hat{S}_{ta})$ does not appear to be a factor that influences estimator performance, at least over a reasonable range of simulation inputs (Conn, 2007).

2.3 *Population Process Models*

The observation model (1) conditions on a number of unobserved latent abundance terms. Here, we characterize a general class of wildlife population dynamics models that provide further structure to these variables. First, we condition on $\mathbf{N_1}$, the vector of age-specific population sizes in year one immediately prior to harvest (where $\mathbf{N_t} = \{N_{t1},\ N_{t2}, \ldots, N_{t,A-1},\ N^{+}_{tA}\}$). We then write the joint PMF of abundance in year two as

$$\begin{aligned}[\mathbf{N_2}\,|\,\mathbf{N_1}, \mathbf{f_1}, \mathbf{S_1}] &= [N_{21}\,|\,\mathbf{N_1}, \mathbf{f_1}] \\ &\quad \times [N_{22}\,|\,N_{11}, S_{11}] \cdots [N_{2A}\,|\,N_{1,A-1}, S_{1,A-1}] \\ &\quad \times \left[N^{+}_{2,A+1}\,\middle|\,N^{+}_{1A}, S_{1A}\right].\end{aligned}$$

Here, $\mathbf{S_t} = \{S_{t1},\ S_{t2}, \ldots, S_{tA}\}$, the vector of survival probabilities in year t, and $\mathbf{f_t} = \{f_{t1},\ f_{t2}, \ldots, f_{tA}\}$, the vector of recruitment process intensities in year t. Remaining notation is defined in Table 1.

Joint PMFs for subsequent years are similar, but an additional allowance is made for $N^{+}_{t,A+1}$ when $t > 1$:

$$\begin{aligned}[\mathbf{N_{t+1}}\,|\,\mathbf{N_t}, \mathbf{f_t}, \mathbf{S_t}] &= [N_{t+1,1}\,|\,\mathbf{N_t}, \mathbf{f_t}] \\ &\quad \times [N_{t+1,2}\,|\,N_{t1}, S_{t1}] \cdots [N_{t+1,A}\,|\,N_{t,A-1}, S_{t,A-1}] \\ &\quad \times \left[N^{+}_{t+1,A+1}\,\middle|\,N_{tA} + N^{+}_{t,A+1}, S_{tA}\right]. \qquad (2)\end{aligned}$$

Note that this formulation implies that there is no immigration to or emigration from the harvestable population. As such, age-specific population structure is modeled as a first-order Markov process. Conditional on the vector of abundances in the first year and parameters for survival and recruitment processes, we may thus write the probability of all future age- and time-specific abundances as

$$[\mathbf{N}\,|\,\mathbf{N_1}, \mathbf{S}, \mathbf{f}] = \prod_{t=1}^{T-1} [\mathbf{N_{t+1,}}\,|\,\mathbf{N_t}, \mathbf{f_t}, \mathbf{S_t}],$$

where $\mathbf{f}$ and $\mathbf{S}$ denote the vector of all f_{ta} and S_{ta} parameters $(t = 1, 2, \ldots, T-1,\ a = 1, 2, \ldots, A)$, respectively.

In the preceding formulation, population size in year $t + 1$ consists of individuals who have survived from year t as well as new recruits to the population. As written, the number of new recruits depends on abundance in the previous year; however, in some cases, individuals may not enter the harvestable population for several years after they are born. If this is the case, we may simply condition on the augmented vector $[\mathbf{N_1}, N_{21}, \ldots, N_{A_r+1,1}]$, replacing $[N_{t+1,1}\,|\,\mathbf{N_t},\ \mathbf{f_t}]$ with $[N_{t+A_r+1,1}\,|\,\mathbf{N_t}, \mathbf{f_t}]$. Here, A_r gives the age at which animals are recruited to the population at risk of harvest. Choices of PMFs for survival and recruitment will depend on the population in question, but we suspect that binomial and Poisson models will commonly be appropriate for each process, with possible overdispersion incorporated via random effects.

2.4 *Likelihood*

We suggest that inference be based on the likelihood

$$L_1 = [\mathbf{C}, \mathbf{N}\,|\,\mathbf{N_1}, \mathbf{S}, \mathbf{h}, \mathbf{f}] = [\mathbf{C}\,|\,\mathbf{N}, \mathbf{h}][\mathbf{N}\,|\,\mathbf{N_1}, \mathbf{S}, \mathbf{f}]. \qquad (3)$$

We retain the latent variables $\mathbf{N}$ in the likelihood because of the computational difficulty in integrating them out, and because they may be of considerable interest to biologists (Link, Royle, and Hatfield, 2003). Indeed, predictions of total female population size in year t may be made with the quantity $\hat{N}_t = \hat{N}^{+}_{tA} + \sum_{a=1}^{A-1} \hat{N}_{ta}$.

2.5 *Auxiliary Data*

Even with highly constrained models, (3) is overparameterized (for an exploration of parameter identification in similarly structured models, see Gove et al., 2002). To generate sensible estimates of model parameters, additional information is needed. Gove et al. (2002) suggested basing inference on a joint likelihood similar to

$$L = L_1 L_2, \qquad (4)$$

where L_2 gives the likelihood for an auxiliary data set pertaining to survival and harvest parameters, such as from a radio-telemetry or mark-recovery study. A similar approach was used in the joint analysis of census and mark-recovery data (Besbeas et al., 2002). No matter what data set is modeled, it is imperative that the likelihood for auxiliary data be structured in such a manner so that survival and harvest parameters can be shared between L_1 and L_2. Even so, data limitations will often require the use of reduced parameter models. While parameter identification is not explored in great depth here, we hope that the following example will provide some intuition about what types of reduced parameter models will be useful in practice.

3. Example

We collected statewide marking and harvest records for black bears in Pennsylvania for the period 1986–1999 from the Pennsylvania Game Commission (PGC; Web Table 1). During this time period, PGC personnel captured bears throughout their range, determined their sex and age, and released them with individually identifiable metal ear tags. Marking occurred between March and November each year. Following a three-day hunting season in November, hunters were required to present all harvested bears to PGC check stations, where age and sex information was collected and the identity of marked bears was recorded. For further details on sampling protocols, see Diefenbach, Laake, and Alt (2004).

Several unique features of the bear data motivated us to elaborate on the general model for age-at-harvest data presented in Section 2. First, we set recruitment parameters for cubs and yearlings to 0, as Pennsylvania black bear typically do not become pregnant before age two (Kordek and Lindzey, 1980; Alt, 1989). Second, we replaced the recruitment terms $[N_{t+1,1} \mid \mathbf{N_t}, \mathbf{f_t}]$ in (2) and (3) with $[N_{t+1,1} \mid \mathbf{N_t} - \mathbf{C_t}, \mathbf{f_t}]$. In the case of black bears in Pennsylvania, the probability that a hunter reports a harvested bear to wildlife officials is thought to be near 1.0. As such, subtracting the observed harvest from population size, the previous year represents the population size immediately after harvest. Conditioning on this quantity should remove one potentially time-varying component from the recruitment process; formulating prior distributions for $\mathbf{f}$ should also be simplified in this case, although this consideration is beyond the scope of the article.

We assumed binomial models for survival and harvest processes, and a Poisson model for the recruitment process. As such, we may write the age-at-harvest likelihood as

$$\begin{aligned} L_1 &= L(\mathbf{C}, \mathbf{N} \mid \mathbf{S}, \mathbf{h}, \mathbf{f}, \mathbf{N_1}) \\ &= \prod_{t=1}^{T-1} \frac{\exp(-\lambda_t)\lambda_t^{N_{t+1,1}}}{N_{t+1,1}!} \\ &\quad \times \prod_{t=1}^{T-1}\prod_{a=1}^{A-1} \binom{N_{ta}}{N_{t+1,a+1}} S_{ta}^{N_{t+1,a+1}}(1-S_{ta})^{D_{ta}} \\ &\quad \times \binom{N_{1A}^{+}}{N_{2,A+1}^{+}} S_{1A}^{N_{2,A+1}^{+}}(1-S_{1A})^{D_{1A}} \\ &\quad \times \prod_{t=2}^{T-1} \binom{N_{tA}+N_{t,A+1}^{+}}{N_{t+1,A+1}^{+}} S_{tA}^{N_{t+1,A+1}^{+}}(1-S_{tA})^{D_{tA}} \\ &\quad \times \prod_{t=1}^{T}\prod_{a=1}^{A-1} \binom{N_{ta}}{C_{ta}} h_{ta}^{C_{ta}}(1-h_{ta})^{N_{ta}-C_{ta}} \\ &\quad \times \binom{N_{1A}^{+}}{C_{1A}^{+}} h_{1A}^{C_{1A}^{+}}(1-h_{1A})^{X_{1A}} \\ &\quad \times \prod_{t=2}^{T} \binom{N_{tA}+N_{t,A+1}^{+}}{C_{tA}^{+}} h_{tA}^{C_{tA}^{+}}(1-h_{tA})^{X_{tA}}. \end{aligned}$$

Here,

$$D_{ta} = \begin{cases} N_{ta} - N_{t+1,a+1}, & a < A \text{ or } i = 1, \\ N_{tA} + N_{t,A+1}^{+} - N_{t+1,A+1}^{+}, & \text{otherwise}, \end{cases}$$

$$X_{ta} = \begin{cases} N_{ta} - C_{ta}, & a < A \text{ or } t = 1, \\ N_{tA} + N_{t,A+1}^{+} - C_{tA}^{+}, & \text{otherwise} \end{cases}$$

and

$$\lambda_t = \sum_{a=3}^{A+} f_{ta} X_{ta}.$$

For this study, $A = 6$ and $T = 14$ (individuals that were cubs immediately prior to harvest were given an age index of 1).

In addition to age-at-harvest data, we also compiled mark-recovery histories for all females marked over the course of the study. However, we directly modeled encounter histories instead of using minimum sufficient statistics (e.g., Brownie et al., 1985). We write this likelihood as

$$L_2 \propto \prod_{k=1}^{M} \Pr(H_k),$$

where

$$\Pr(H_k) = \begin{cases} h_{t_{k2},a_{k},t_{k2}} \displaystyle\prod_{t=t_{k1}}^{t_{k2}-1} S_{t,a_{k},t}, & I_k = 1, \\ 1 - h_{t_{k1},a_{k},t_{k1}} - \displaystyle\sum_{i=t_{k1}}^{T-1} \left(\prod_{j=t_{k1}}^{i} S_{j,a_{kj}} \right) h_{i+1,a_{k},i+1}, & I_k = 0, \end{cases}$$

and definitions of statistics and parameters are given in Table 1.

We made all the common assumptions typical for mark-recovery studies (Williams, Nichols, and Conroy, 2002): animals behave independently, marks are not lost or overlooked, and there is no individual heterogeneity (at least that cannot be explained by age class). Further, we assumed that there is no mortality between the time a bear is marked and the harvest season. Several of these assumptions may be violated to some degree; for instance, sampling effort to mark black bears was distributed between March and November, although mortality was thought to be low during this period for individuals greater than one year of age. Nevertheless, the recovery rates of cubs were likely underestimated, ostensibly causing a positive bias in the number of new recruits each year. Fates of members of family groups (females with cubs) were not independent, which would cause negative bias in variance estimates. Tag loss occurred, but was of small magnitude for females (Diefenbach and Alt, 1998). As suggested in Section 2, we based inference on (4), which assumes that mark-recovery and age-at-harvest data are independent. Nevertheless, we included data from marked individuals in the age-at-harvest matrix so that total population abundance could be estimated. In some settings, this assumption may also result in measures of uncertainty that are too precise (see Conn, 2007).

In total, we fit four models to the data, which varied by the number of fixed and random effects on the logits of survival and recovery rate, and on the log of recruitment rate (Table 2). Models for S_{ta}, h_{ta}, and f_{ta} were linear on a transformed scale, consistent with current practice in the capture–recapture literature (cf. Lebreton et al., 1992). In each case, all parameters were identifiable (e.g., Conn, 2007); in our experience, identifiability of survival and recovery parameters from mark-recovery data alone served as a reasonable indicator for whether abundance and recruitment rates in the joint model could be identified. This is important because parameter identification for product multinomial models (including mark-recovery models) is a topic that has received

Table 2

Models fit to age-at-harvest and mark-recovery data sorted by estimated DIC, where ΔDIC gives the difference in DIC from the highest ranked model. When time is included in the model structure, demographic parameters are modeled as random effects, while age is always modeled as a fixed effect. Also presented are Bayesian p-values (p_{B}) for each model.

Model name	Logit(S_{ta})	Logit(h_{ta})	Log(f_{ta})	ΔDIC	p_{B}
$S(a+t)h(a+t)f(t)$	$\gamma_a + \alpha_t$	$\phi_a + \upsilon_t + \epsilon_{ta}$	$\beta + \kappa_t$	0.0	0.67
$S(a+t)h(a+t)f(\cdot)$	$\gamma_a + \alpha_t$	$\phi_a + \upsilon_t + \epsilon_{ta}$	β	1.8	0.74
$S(a)h(a)f(t)$	γ_a	$\phi_a + \epsilon_{ta}$	$\beta + \kappa_t$	15.4	0.56
$S(a)h(a)f(\cdot)$	γ_a	$\phi_a + \epsilon_{ta}$	β	20.4	0.48

considerable attention in the literature (e.g., Catchpole and Morgan, 1997; Giminez, Choquet, and Lebreton, 2004).

Preliminary analysis indicated substantial differences in observed and expected harvest data when a model such as

$$\mathrm{Logit}(h_{ta}) = \upsilon_t + \phi_a$$

was fit to the data. These differences likely resulted from annual changes in the temporal distribution of denning dates as well as the percentage of a given cohort that was pregnant. Pregnant females typically den prior to the rest of the population, and thus are more likely to be unavailable for harvest during the hunting season. It was difficult to model these biological processes directly, so we followed the approach of Barry et al. (2003) and included overdispersion terms, ϵ_{ta}, in the formulation for harvest rates. These were modeled as random effects on the logit scale and were assumed to have a normal distribution with mean zero and precision parameter $\tau_{\boldsymbol{\epsilon}}$. Similarly, when included in the model structure, year effects (i.e., α_t, υ_t, and κ_t) were normally distributed random effects on the logit scale with mean zero and precision parameters $\tau_{\boldsymbol{\alpha}}$, $\tau_{\boldsymbol{\upsilon}}$, and $\tau_{\boldsymbol{\kappa}}$, respectively. Remaining parameters (Table 2) were all modeled as fixed effects.

Conducting a Bayesian analysis required that we specify prior distributions for model parameters. In particular, we chose a diffuse, improper prior for abundance (Link et al., 2003), priors for fixed effects that were approximately uniform when transformed to (0,1) space, a prior for recruitment process intensity that was relatively flat over all biologically plausible values, and a prior for the precision of random effects that precluded clumping at 0 and 1 on transformed parameter spaces. Priors were given by

$$\begin{aligned}
&[N_{1j}] \propto c,\\
&[\gamma_j],\ [\phi_j] \sim \mathrm{Normal}(0,3),\\
&[\beta] \sim \mathrm{Normal}(0.25,1),\ \text{ and}\\
&[\tau_\alpha],\ [\tau_\upsilon],\ [\tau_\kappa],\ [\tau_\epsilon] \sim \mathrm{Gamma}(0.1,0.1).
\end{aligned}$$

We used Gibbs sampling to sequentially update each parameter and latent variable. The full conditionals for precision parameters were available in closed form, and were simulated directly as

$$[\tau_\alpha \mid \boldsymbol{\alpha}] \sim \mathrm{Gamma}\left(\frac{T-1}{2} + 0.1, \frac{\sum \alpha_t^2}{2} + 0.1\right),$$

$$[\tau_\upsilon \mid \boldsymbol{\upsilon}] \sim \mathrm{Gamma}\left(\frac{T}{2} + 0.1, \frac{\sum \upsilon_t^2}{2} + 0.1\right),$$

$$[\tau_\kappa \mid \boldsymbol{\kappa}] \sim \mathrm{Gamma}\left(\frac{T-1}{2} + 0.1, \frac{\sum \kappa_t^2}{2} + 0.1\right), \quad \text{and}$$

$$[\tau_\epsilon \mid \boldsymbol{\epsilon}] \sim \mathrm{Gamma}\left(\frac{YA}{2} + 0.1, \frac{\sum \epsilon_{ta}^2}{2} + 0.1\right).$$

All remaining parameters and variables were updated with Metropolis–Hastings steps. Proposals in a given iteration were normally distributed with a mean at the previous iteration's parameter value, and a standard deviation was chosen to achieve a 35–40% acceptance rate. Proposals outside the support of abundance classes were automatically rejected. This solution worked reasonably well for these data; for details on alternatives for highly constrained problems see Conn (2007).

For each model, we ran two independent Markov chains of length one million with overdispersed starting values. In all cases, Gelman–Rubin statistics (Gelman et al., 2004) indicated convergence after 100,000–200,000 iterations. Nevertheless, we treated the first 500,000 iterations as a burn-in and combined the second halves of each chain to generate a sample of one million from the posterior distribution. To save disk space, this sample was thinned to 200,000 by recording every fifth iteration. Marginal posterior distributions were then summarized by calculating moments and 90% HPD Bayesian credible intervals. Deviance information criterion (DIC; Spiegelhalter et al., 2002) was also calculated for purposes of model selection.

We implemented a goodness-of-fit test based on a Bayesian p-value (Gelman et al., 2004) for the age-at-harvest portion of the likelihood. For a given sample i from the posterior distribution, we simulated harvest data, $\mathbf{C}_t^{\mathrm{rep}}$, given $\boldsymbol{\theta}_t$, the set of parameter values at iteration t. Next, deviance for the age-at-harvest portion of the likelihood was calculated for observed, $\mathrm{D}_t(\mathbf{C}, \boldsymbol{\theta}_t)$, and for simulated, $\mathrm{D}_t(\mathbf{C}_t^{\mathrm{rep}}, \boldsymbol{\theta}_t)$, data. The Bayesian p-value was then obtained as

$$p_{\mathrm{B}} = \frac{1}{K}\sum_{t=1}^{K} I_{[0,\,\infty)}\left[\mathrm{D}_t\left(\mathbf{C}_t^{\mathrm{rep}}, \boldsymbol{\theta}_t\right) - \mathrm{D}_t(\mathbf{C}, \boldsymbol{\theta}_t)\right],$$

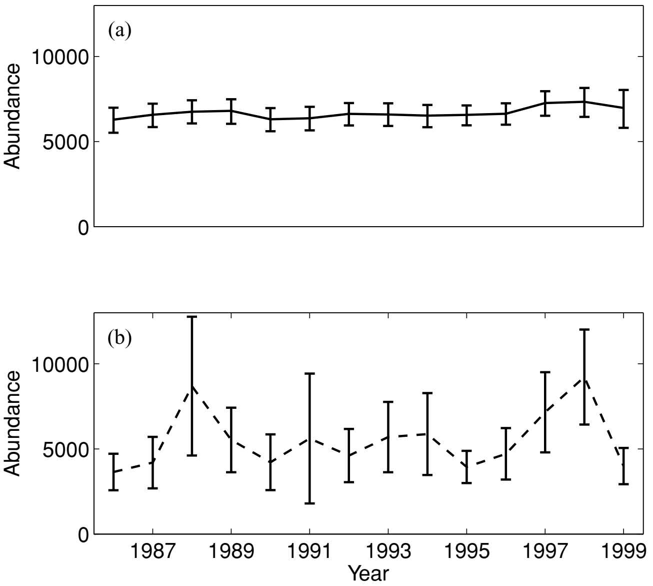

Figure 1. Abundance of female black bears in Pennsylvania from 1986 to 1998 as (a) predicted from the joint age-at-harvest and mark-recovery analysis, and (b) estimated with the bias-corrected LP estimator. Posterior means and 90% Bayesian credible intervals are presented from the highest ranked DIC model, while point estimates and 90% asymptotic confidence intervals are presented for the LP estimator.

where K denotes the total number of samples in which age-at-harvest data is simulated, and $I_{\Omega}(x)$ denotes an indicator function for the set Ω. In our case, we let $K = 200$, spacing samples evenly across Markov chain iterations.

When fit to the data, models including time varying survival and recovery rates were strongly favored by DIC, with some support for temporal effects on recruitment (Table 2). Posterior summaries from the highest ranked model indicated female black bear abundance increased in Pennsylvania from 1986 to 1999 (Figure 1). To contrast our estimator with another commonly used abundance estimator, we computed year-specific bias-corrected Lincoln–Petersen (LP) estimates of abundance, together with accompanying variances (Seber, 1982). Point estimates from the LP estimator had greater standard errors (Figure 1). Further, consecutive point estimates using the LP approach were often biologically implausible (e.g., due to constraints on reproductive capacity; see Alt, 1989). Diefenbach et al. (2004) noted this tendency, suggesting that annual changes in the availability of pregnant females for harvest could lead to a high degree of variability in single season estimators of abundance. In addition to abundance, we also summarized posterior distributions for survival probability (Figure 2), recovery probability (Figure 3), and recruitment rate (Figure 4), three quantities of fundamental interest to population biologists and wildlife managers. These estimates reconfirm that the black bear population in Pennsylvania is one of the most productive in the United States (Alt, 1989).

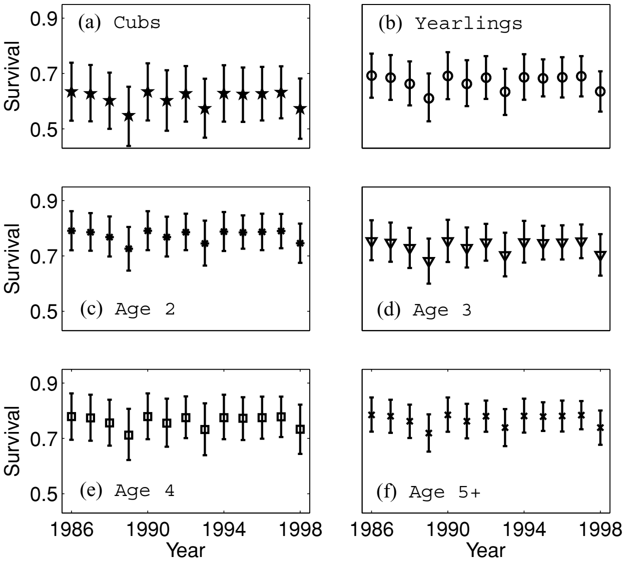

Figure 2. Female black bear survival ($\hat{S}_{ta}$) in Pennsylvania from 1986 to 1998 as estimated from age-at-harvest and mark-recovery data. Posterior means and 90% Bayesian credible intervals are presented from the highest ranked DIC model.

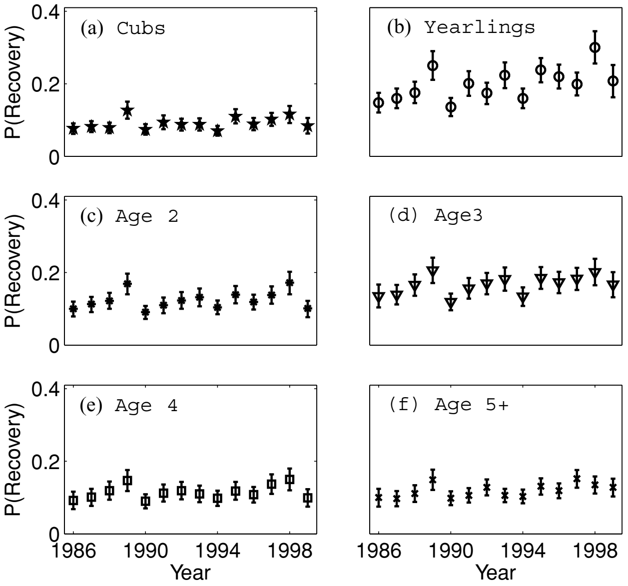

Figure 3. Recovery probabilities ($\hat{h}_{ta}$) for female black bears in Pennsylvania from 1986 to 1998 as estimated from a joint analysis of age-at-harvest and mark-recovery data. Posterior means and 90% Bayesian credible intervals are presented from the highest ranked DIC model.

4. Discussion

In this article, we were able to estimate a number of demographic parameters by jointly modeling the age structure of harvests and mark-recovery data. By adopting a state-space formulation, we were able to conduct the analysis using a likelihood structured in a hierarchical fashion. This allowed us to

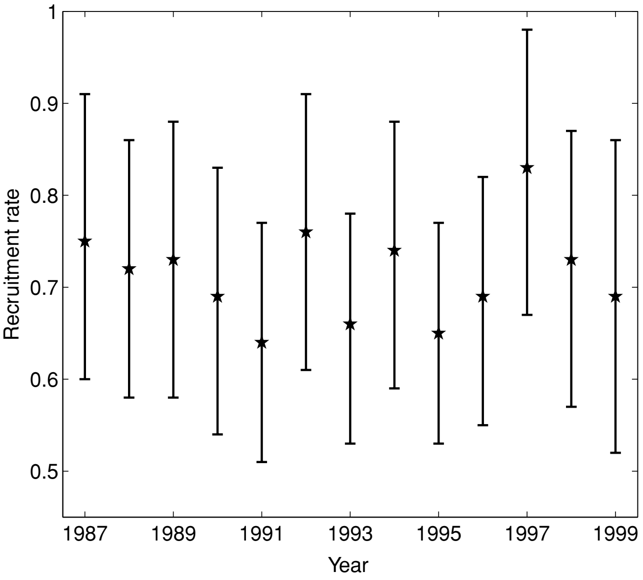

Figure 4. Estimated recruitment rate ($\hat{f}_{ta}$) for Pennsylvania black bear from 1987 to 1999. Presented are posterior means and 90% Bayesian credible intervals from the highest ranked DIC model.

improve upon previously proposed models for analyzing age-at-harvest data for wildlife. In particular, we specified population models with a self-loop for older age classes as well as an explicit recruitment process. Further, the Bayesian approach to the problem allows one to properly model stochasticity in abundance classes and to specify prior distributions that constrain parameters to biologically realistic values. While we did not do so here, density-dependent processes also could be contemplated, either by specifying process covariances or by the introduction of functional relationships between parameters.

Previous analyses of these data using a Horvitz–Thompson type estimator (Diefenbach et al., 2004) sometimes yielded conflicting point estimates and could not incorporate pertinent information about black bear biology. For instance, annual changes in point estimates of abundance often were dramatically different from what biologists would expect based on the knowledge of survival, fecundity, and movement of the species. Also, age-specific abundances were often biologically unreasonable for a system relatively closed to emigration and immigration; for instance, the abundance estimate of yearlings in year two could be greater than that for the number of cubs in year one. In contrast, our approach led to parameter estimates that were internally consistent because biological processes were embedded into the estimation process. This feature, along with the use of extra data from marked animals, led to greater precision and less temporal variation in point estimates than the LP estimator.

With regard to Pennsylvania black bear, future work should explore the tenability of several key model assumptions, such as the assumption of no preharvest mortality following marking. This assumption will be violated to some degree, particularly for cubs that are marked in dens during March, and may induce positive bias in abundance estimates. Tag loss could also induce bias when multiple years are analyzed within the mark-recovery framework and the lack of independence of fates of family groups (females with cubs) may introduce unmodeled overdispersion.

Finally, we note that the modeling framework we developed can be extended to incorporate additional data sources. For instance, if recaptures are available in addition to recoveries, one may simply replace L_2 with another likelihood (e.g., Burnham, 1993). Likewise, information from radio-telemetry studies could be incorporated (e.g., Gove et al., 2002). Extensions to handle aging misclassification are also readily available (Conn and Diefenbach, 2007). We believe this flexibility is essential for biologists, who are often confronted with a diverse array of data sources and sampling challenges.

5. Supplementary Materials

Web Table 1 referenced in Section 3 is available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

ACKNOWLEDGEMENTS

We thank the PGC and the numerous personnel who gathered data on black bears. Funding for the first author was provided by the Colorado Division of Wildlife and by NSF-IGERT grant DGE-0221595003. Comments and suggestions from K. P. Burnham, P. F. Doherty Jr., and two anonymous reviewers greatly enhanced the content of this article.

REFERENCES

Alt, G. L. (1989). Reproductive biology of female black bears and early growth and development of cubs in northeastern Pennsylvania. Ph.D. Thesis, West Virginia University, Morgantown, West Virginia.

Barry, S. C., Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2003). The analysis of ring-recovery data using random effects. *Biometrics* **59,** 54–65.

Besbeas, P., Freeman, S. N., Morgan, B. J. T., and Catchpole, E. A. (2002). Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* **58,** 540–547.

Brownie, C., Anderson, D. R., Burnham, K. P., and Robson, D. S. (1985). Statistical inference from band recovery data: A handbook. U.S. Fish and Wildlife Resource Publication (156), Washington, DC.

Buckland, S. T., Newman, K. B., Thomas, L., and Koester, N. (2004). State-space models for dynamics of wild animal populations. *Ecological Modelling* **171,** 157–175.

Burnham, K. P. (1993). A theory for combined analysis of ring recovery and recapture data. In *Marked Individuals in the Study of Bird Populations*, J. D. Lebreton and P. M. North (eds), 199–213. Basel, Switzerland: Birkhauser.

Catchpole, E. A. and Morgan, B. J. T. (1997). Detecting parameter redundancy. *Biometrics* **84,** 187–196.

Conn, P. B. (2007). Bayesian analysis of age-at-harvest data, with focus on wildlife monitoring programs. Ph.D. Thesis, Colorado State University, Fort Collins, Colorado.

Conn, P. B. and Diefenbach, D. R. (2007). Adjusting age and stage distributions for misclassification errors. *Ecology* **88,** 1977–1983.

Deriso, R. B., Quinn, T. J. I., and Neal, P. R. (1985). Catch-age analysis with auxiliary information. *Canadian Journal of Fisheries and Aquatic Sciences* **42,** 815–824.

Diefenbach, D. R. and Alt, G. L. (1998). Modeling and evaluation of ear tag loss in black bears. *Journal of Wildlife Management* **62,** 1292–1300.

Diefenbach, D. R., Laake, J. L., and Alt, G. L. (2004). Spatio-temporal and demographic variation in the harvest of black bears: Implications for population estimation. *Journal of Wildlife Management* **68,** 947–959.

Dupont, W. D. (1983). A stochastic catch-effort method for estimating animal abundance. *Biometrics* **39,** 1021–1033.

Fournier, D. and Archibald, C. P. (1982). A general theory for analyzing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences* **39,** 1195–1207.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd edition. Boca Raton, Florida: Chapman and Hall.

Giminez, O., Choquet, R., and Lebreton, J. D. (2004). Parameter redundancy in multistate capture-recapture models. *Biometric Journal* **45,** 704–722.

Gove, N. E., Skalski, J. R., Zager, P., and Townsend, R. L. (2002). Statistical models for population reconstruction using age-at-harvest data. *Journal of Wildlife Management* **66,** 310–320.

Kordek, W. S. and Lindzey, J. S. (1980). Preliminary analysis of female reproductive tracts from Pennsylvania black bears. *International Conference on Bear Research and Management* **4,** 159–161.

Laake, J. L. (1992). Catch-effort models and their application to elk in Colorado. Ph.D. Thesis, Colorado State University, Fort Collins, Colorado.

Lebreton, J. D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs* **62,** 67–118.

Lewy, P. and Nielson, A. (2003). Modelling stochastic fish stock dynamics using Markov chain Monte Carlo. *ICES Journal of Marine Science* **60,** 743–752.

Link, W. A., Royle, J. A., and Hatfield, J. S. (2003). Demographic analysis from summaries of an age-structured population. *Biometrics* **59,** 778–785.

Maunder, M. N. (2001). Integrated tagging and catch-at-age analysis (ITCAAN). In *Spatial Processes and the Management of Fish Populations*, G. H. Kruse, N. Bez, A. Booth, M. W. Dorn, S. Hills, R. N. Lipcius, D. Pelletier, C. Roy, S. J. Smith, and D. Witherell (eds). Alaska Sea Grant College Program Report No. AK-SG-01-02, Fairbanks: University of Alaska.

McAllister, M. K. and Ianelli, J. N. (1997). Bayesian stock assessment using catch-age data and the sampling-resampling importance algorithm. *Canadian Journal of Fisheries and Aquatic Sciences* **54,** 284–300.

Meyer, R. and Millar, R. B. (1999). Bayesian stock assessment using a state-space implementation of the delay difference model. *Canadian Journal of Fisheries and Aquatic Sciences* **56,** 37–52.

Newman, K. B., Buckland, S. T., Lindley, S. T., Thomas, L., and Fernàndez, C. (2006). Hidden process models for animal population dynamics. *Ecological Applications* **16,** 74–86.

Rosenberry, C. S., Diefenbach, D. R., and Wallingford, B. D. (2004). Reporting-rate variability and precision of white-tailed deer harvest estimates in Pennsylvania. *Journal of Wildlife Management* **68,** 860–869.

Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters*, 2nd edition. Caldwell, NJ: Blackburn Press.

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

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and Management of Animal Populations*. San Diego, California: Academic Press.

Received March 2007. *Revised November* 2007.
Accepted November 2007.