

Bayesian Estimation of Abundance for Open
Populations with Covariate Dependent Capture and
Survival Probabilities

DRAFT

Simon J Bonner and Carl J Schwarz

October 25, 2009

1 Introduction

Covariates that vary both between individuals and over time present a challenge in the analysis of data from capture-recapture studies. Such quantities can only be measured for the subset of individuals captured on each sampling occasion and so a large proportion of the values appearing in the likelihood function will be unknown. Moreover, these values cannot be considered missing at random and simply ignored because of the covariate's potential effect on the probability that an individual is observed. This is of particular concern if the covariate is continuous because the likelihood will depend on high-dimensional integrals which make it impossible to evaluate analytically and very slow to evaluate numerically. Bonner and Schwarz (2006) present a Bayesian solution using the complete-data likelihood of the conditional Cormack-Jolly-Seber model (CJS) to estimate the effects of such covariates on the survival probabilities of the marked individuals; however, the method does not allow estimation of the population's size. The current work extends the Bayesian approach to the unconditional Jolly-Seber model (JS) in order to estimate the size of a population from capture-recapture data when the probabilities of capture or survival are functions of continuous, time-dependent, individual covariates.

The JS and CJS provide the basis for most models of open-population capture-recapture data (Jolly, 1965; Seber, 1965). As the names suggest, the two are closely related. Both model the recapture of marked individuals in terms of two sets of parameters: the capture probabilities (the probability that an individual alive on one sampling occasion is captured) and survival probabilities (the probability that an individual survives from one occasion to the next). The difference between the models is that the CJS conditions on the first release of each marked individual. This avoids any assumptions regarding the behaviour of the marked individuals, but the model provides no information about the size of the population. The JS also models the first capture of the marked individuals which allows estimation of the population

1 size on each occasion, but imposes assumptions on the behaviour of the unmarked
2 individuals.

3 The critical assumption of the CJS is that the probabilities of capture and survival
4 are the same for all marked individuals. The JS further assumes that all individuals,
5 marked and unmarked, share the same probability of capture. While these assump-
6 tions may be reasonable for small or isolated populations, many factors may affect an
7 individual's probability of capture or survival and failure to account for such differ-
8 ences has long been known to bias estimates of abundance (Gilbert, 1973; Carothers,
9 1973). In general, there are two strategies to account for these differences. The first
10 is to model the variations between individuals as the result of unexplained, random
11 heterogeneity. Specific methods include the large sample bias correction of Hwang
12 and Chao (1995), the simulation approach of Pledger and Efford (1998), and the
13 explicit random effects models of Otis, Burnham, White, and Anderson (1978) and
14 Pledger, Pollock, and Norris (2003). However, these methods have seen limited prac-
15 tical application. The models of Hwang and Chao (1995) and Pledger and Efford
16 (1998) impose restrictive assumptions and Link and Barker (2005) demonstrate that
17 the distribution of the random effect is not identifiable so that different models may
18 fit the observed data equally well but produce disparate estimates of abundance.

19 The more widely applied approach is to model the differences in the capture and
20 survival probabilities as functions of measurable covariates. Lebreton, Burnham,
21 Clobert, and Anderson (1992) present a unified generalized linear modelling (GLM)
22 framework to incorporate the effects of completely observable covariates in conditional
23 open-population models, including the CJS. McDonald and Amstrup (2001) provide
24 an *ad hoc* extension of this method to the JS to allow estimation of population size.
25 As in the original JS, the estimate of the population size on any occasion is obtained
26 in two stages by first estimating the unique capture probability for each individual
27 captured and then computing a Horvitz-Thompson type estimate of abundance.

1 Covariates that vary both between individuals and over time will not be com-
2 pletely observed for the marked individuals and so cannot be included in either of
3 these methods. The effects of discrete covariates can be assessed with the multi-state
4 CJS, originally formulated for three capture occasions by Arnason (1973) and later
5 generalized and adapted to the JS by Schwarz, Schweigert, and Arnason (1993) and
6 Dupuis and Schwarz (2007). In essence, the multi-state model assumes that the the
7 values of the covariates (states) for each individual obey the Markov property, and
8 likelihood contributions are formed by summing over all possible combinations of the
9 missing covariate values weighted according to their probability of occurrence. The
10 model of Bonner and Schwarz (2006) applies the same concept to incorporate contin-
11 uous covariates into the CJS, but the resulting likelihood requires high-dimensional
12 integrals in place of sums. This makes the likelihood difficult to evaluate, and so
13 Bayesian inference was applied via the complete data likelihood and Markov chain
14 Monte Carlo sampling (MCMC). A similar approach was used in King, Brooks, and
15 Coulson (2008).

16 The objective of the current work is to extend the method of Bonner and Schwarz
17 (2006) to the JS and other unconditional open-population models. We show that
18 careful selection of the prior distribution leads to an approximate Bayesian procedure
19 that divides the problem into two simple steps, first drawing information about the
20 capture probabilities from only the marked individuals and then applying these results
21 to make inference about abundance. In section 2, we develop our approach for a
22 general open-population model involving discrete covariates, extend this to models
23 involving continuous covariates, and end with specific details for the JS type model.
24 Sections 3 and 4 present an application of the JS type model to the study of Soay
25 sheep (*Ovis aries*) on the Isle of Hirta and a simulation study based on the results.
26 We conclude with a discussion of the method and its extension to more complicated
27 data.

2 Methods

2.1 Notation

The prior density for parameter θ and its posterior density given the data X are denoted as $\pi_\theta(\cdot)$ and $\pi_\theta(\cdot|X)$. Random variables whose densities depend on the limiting argument, δ , are identified by subscripts, e.g. θ_δ . A conditional random variable whose distribution depends on δ is placed inside parentheses, e.g. $(\theta|X)_\delta$. Densities and expectations that depend on the value of δ are identified by subscripts, e.g. $\pi_\theta^\delta(\cdot)$ and $E_\theta^\delta(\theta|X)$ are the prior density and posterior mean of θ_δ .

2.2 Development for Discrete Covariates

We begin by deriving our method for a capture-recapture study comprising $T > 2$ capture occasions in which the probability of capture (and possibly other parameters) depends on a discrete covariate that varies between individuals and over time. This is equivalent to assuming that on each capture occasion the population can be divided into K groups of individuals who share the same probability of capture. Let U_{kt} represent the total number of unmarked individuals in group k on occasion t , p_{kt} the capture probability for the individuals in this group, and u_{kt} the number of these individuals that are captured. The $K \times t$ matrices containing the entries U_{kt} , u_{kt} , and p_{kt} are represented as \mathbf{U} , \mathbf{u} , and \mathbf{p} . Data for the i^{th} captured individual is summarized in two vectors of length T : the capture history, $\boldsymbol{\omega}_i$, defined so that $\omega_{it} = 1$ if individual i was captured on occasion t and 0 otherwise, and the vector of group memberships, \mathbf{x}_i , which is only partially observed so that $x_{it} = k$ if the individual was captured in group k on occasion t and is missing otherwise. Combined data from all n individuals captured at least once during the study are represented in the pair of $n \times t$ matrices $\boldsymbol{\Omega}$ and \mathbf{X} . For notational convenience, we imagine that the information in $\boldsymbol{\Omega}$ and \mathbf{X} can be divided into two parts, the data from the first capture of each individual,

1 which is summarized in \mathbf{u} , and the information conditional on the first release of each
 2 marked individual, denoted by $\mathbf{\Omega}^*$ and \mathbf{X}^* .

3 Following the development of Seber (1965) the likelihood function for an open-
 4 population experiment can be constructed as the product of two components, the
 5 first modelling the initial capture of each individual and the second modelling the
 6 subsequent events. Conditioning on the number of unmarked individuals alive in
 7 each group on each occasion and assuming independence between individuals, the
 8 probability of capturing u_{kt} individuals from group k on occasion t follows a binomial
 9 distribution. The corresponding likelihood contribution is:

$$L_{1kt}(U_{kt}, p_{kt}|u_{kt}) = \binom{U_{kt}}{u_{kt}} p_{kt}^{u_{kt}} (1 - p_{kt})^{U_{kt} - u_{kt}}$$

10 and the first likelihood component is simply the product:

$$L_{1kt}(\mathbf{U}, \mathbf{p}|\mathbf{u}) = \prod_{t=1}^T \prod_{k=1}^K L_{1kt}(U_{kt}, p_{kt}|u_{kt}).$$

11 The exact form of the second component will depend on the assumed dynamics of the
 12 population but is not needed in the theoretical development and is denoted simply by
 13 $L_2(\mathbf{p}, \boldsymbol{\theta}|\mathbf{\Omega}^*, \mathbf{X}^*)$. This component must be a function of \mathbf{p} , and may also depend on
 14 an further set of parameters not appearing in L_1 , denoted by $\boldsymbol{\theta}$, but may only depend
 15 on the data conditional on first release of each marked individual ($\mathbf{\Omega}^*$ and \mathbf{X}^*). For a
 16 study in which group membership is fixed and individuals can leave between any pair
 17 of occasions, L_2 would be the product of the K CJS likelihood contributions for each
 18 group and $\boldsymbol{\theta}$ the $K \times T - 1$ matrix of survival probabilities. If individuals could move
 19 between groups following a Markov process then L_2 would represent the likelihood of
 20 a multi-state CJS model and $\boldsymbol{\theta}$ would contain both both the survival and transition
 21 probabilities (Williams, Nichols, and Conroy, 2002, ch. 17 for an overview of these
 22 models).

1 We define the prior distribution by assuming independence between the three sets
 2 of parameters, \mathbf{U} , \mathbf{p} and $\boldsymbol{\theta}$. We also assume that the elements of \mathbf{U} are independent
 3 and identically distributed *a priori* such that $\pi_{\mathbf{U}}(\mathbf{U}) = \prod_{k,t} \pi_U(U_{kt})$. Equating U_{kt}
 4 with the number of trials in a binomial experiment, the standard non-informative prior
 5 for this parameter is the improper Jeffrey’s prior with density $\pi_U(U) \propto 1/U$. This
 6 prior has been used in previous Bayesian analysis of capture-recapture and population
 7 census data (Dupuis and Schwarz, 2007; King and Brooks, 2001), but presents a
 8 problem in that it does not exist if $U = 0$ and the resulting posterior is undefined if
 9 $u_{kt} = 0$ for for any k and t . To avoid this, we approximate Jeffrey’s prior by the (still
 10 improper) distribution with density:

$$\pi_U^\delta(U) \propto \begin{cases} 1/\delta & U = 0 \\ 1/U & U > 0 \end{cases} \quad (1)$$

11 for some $\delta \ll 1$.

12 Bounding the prior density at $U_{kt} = 0$ in this way does generate a well-defined
 13 posterior for all possible \mathbf{u} , but introduces a further problem in that $E(U_{kt}|\boldsymbol{\Omega}, \mathbf{X})$
 14 may be infinite if $u_{kt} = 0$, even for seemingly reasonable priors on p_{kt} (Appendix A
 15 shows that this may occur if p_{kt} is assigned a uniform prior over all of $(0, 1)$). To avoid
 16 this, we further bound the prior density of p_{kt} away from 0 for all k and t . Specifically,
 17 given a target prior distribution $\pi_{\mathbf{p}}(\mathbf{p})$ defined over all of $(0, 1)^{KT}$ we define:

$$\pi_{\mathbf{p}}^\delta(\mathbf{p}) \propto \pi_{\mathbf{p}}(\mathbf{p}) \cdot \prod_{t=1}^T \prod_{k=1}^K \mathbf{1}(p_{kt} > \delta).$$

18 The prior distribution for the remaining parameters in L_2 need not depend on δ
 19 and is denoted by $\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. The full prior density for a given value of δ is defined by

1 $\pi^\delta(\mathbf{U}, \mathbf{p}, \boldsymbol{\theta}) \propto \prod_{k,t} (\pi_p^\delta(p_{kt})\pi_U^\delta(U_{kt})) \cdot \pi_\theta(\boldsymbol{\theta})$ and the posterior density by:

$$\pi^\delta(\mathbf{U}, \mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{X}) \propto L_1(\mathbf{U}, \mathbf{p}|\mathbf{u}) \cdot L_2(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) \cdot \pi^\delta(\mathbf{U}, \mathbf{p}, \boldsymbol{\theta}).$$

2 Our method derives from studying the distribution of $(\mathbf{U}, \mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{X})_\delta$ as $\delta \rightarrow 0$ and

3 $\pi_U(\cdot)$ approaches Jeffrey's prior.

4 Our first claim is that the contribution of L_1 to the marginal posterior distribution
5 of \mathbf{p} and $\boldsymbol{\theta}$ can be ignored provided δ is small. Specifically, we prove:

6 **Theorem 1.1**

7 As $\delta \rightarrow 0$:

$$(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{X})_\delta \xrightarrow{D} (\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*)_\delta$$

8 where the random variable $(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*)_\delta$ has density:

$$\pi_{\mathbf{p}, \boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) \propto L_2(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) \cdot \pi_p^\delta(\mathbf{p})\pi_\theta(\boldsymbol{\theta}).$$

9

10 Application of Scheffé's Lemma then shows that for small δ this distribution can be
11 approximated by replacing $\pi_p^\delta(\mathbf{p})$ by the target prior $\pi_p(\mathbf{p})$. The exact result is:

12 **Theorem 1.2**

13 As $\delta \rightarrow 0$

$$(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*)_\delta \xrightarrow{D} \mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*$$

14 where the density of $\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*$ is proportional to:

$$L_2(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) \cdot \pi_p(\mathbf{p})\pi_\theta(\boldsymbol{\theta}). \tag{2}$$

15

16 Combining Theorems 1.1 and 1.2 via the triangle inequality we obtain:

1 **Theorem 1**

2 As $\delta \rightarrow 0$

$$(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\Omega}, \mathbf{X})_\delta \xrightarrow{D} \mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\Omega}^*, \mathbf{X}^*.$$

3

4 The implication of this result is that the marginal posterior distribution of $(\mathbf{p}, \boldsymbol{\theta})_\delta$
 5 can be approximated by ignoring both the first component of the likelihood and the
 6 lower bound on prior distribution of \mathbf{p} , provide that δ is small enough. Full proofs of
 7 these results are provided in Appendix B.

8 Our second claim concerns the full conditional distribution of the number of un-
 9 marked individuals alive occasion t (i.e., the distribution of $U_t = \sum_{k=1}^K U_{kt}$ given the
 10 data, \mathbf{p} , and $\boldsymbol{\theta}$). Because \mathbf{U} is *a priori* independent of \mathbf{p} and $\boldsymbol{\theta}$ and appears only in the
 11 first component of the likelihood, it is immediate that the full conditional distribution
 12 of \mathbf{U} is proportional to:

$$L_1(\mathbf{U}, \mathbf{p} | \mathbf{u}) \pi_{\mathbf{U}}^\delta(\mathbf{U}) = \prod_{k,t} L_{1kt}(U_{kt}, p_{kt} | u_{kt}) \pi_U^\delta(U_{kt}).$$

13 and hence that the individual U_{kt} are independent *a posteriori*. We then find that:

- 14 1. $(U_{kt} | u_{kt}, p_{kt})_\delta \sim \text{Neg. Bin.}(u_{kt}, p_{kt})$ for k and t such that $u_{kt} > 0$, completely
 15 independent of δ , and
 16 2. $(U_{kt} | u_{kt}, p_{kt})_\delta \xrightarrow{P} 0$ as $\delta \rightarrow 0$ for those k and t such that $u_{kt} = 0$.

17 Application of Slutsky's Theorem then yields:

18 **Theorem 2**

19 As $\delta \rightarrow 0$:

$$(U_t | \mathbf{u}, \mathbf{p})_\delta \xrightarrow{D} \sum_{\{k,t:u_{kt}>0\}} U_{kt} | u_{kt}, p_{kt}$$

20 where $U_{kt} | u_{kt}, p_{kt}$ are independent random variables with distribution *Neg. Bin.* (u_{kt}, p_{kt}) .

21

1 This theorem implies that, for δ small enough, the groups for which $u_{kt} = 0$ can be
 2 ignored in making inference about U_t , and, moreover the posterior distribution of the
 3 remaining U_{kt} has a very simple form given the capture probabilities. Proofs of these
 4 results are provided in Appendix C.

5 A further important result concerns the convergence of moments for the posterior
 6 distribution. Given that $(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\Omega}, \mathbf{X})_\delta \xrightarrow{D} \mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\Omega}^*, \mathbf{X}^*$, the Portmanteau Theorem
 7 immediately implies that for any m $E^\delta(p_{kt}^m | \boldsymbol{\Omega}, \mathbf{X}) \rightarrow E(p_{kt}^m | \boldsymbol{\Omega}^*, \mathbf{X}^*)$ as $\delta \rightarrow 0$ because
 8 p_{kt}^m is bounded in $(0,1)$. This result combined with the conditional independence of
 9 the U_{kt} and the fact that $U_{kt} | u_{kt}, p_{kt} \sim \text{Neg. Bin.}(u_{kt}, p_{kt})$ then yields:

10 1. $E^\delta(U_t | \boldsymbol{\Omega}, \mathbf{X}) = \sum_{k,t | u_{kt} > 0} E^\delta(U_{kt} | \boldsymbol{\Omega}, \mathbf{X})$ converges to:

$$\sum_{\{k,t | u_{kt} > 0\}} E_{p_{kt}}[E_{U_{kt}}(U_{kt} | u_{kt}, p_{kt}) | \boldsymbol{\Omega}^*, \mathbf{X}^*] = \sum_{\{k,t | u_{kt} > 0\}} u_{kt} E_{p_{kt}}(p_{kt}^{-1} | \boldsymbol{\Omega}^*, \mathbf{X}^*) \quad (3)$$

11 and

12 2. $\text{Var}^\delta(U_t | \boldsymbol{\Omega}, \mathbf{X}) = \sum_{k,t | u_{kt} > 0} \text{Var}^\delta(U_{kt} | \boldsymbol{\Omega}, \mathbf{X})$ converges to:

$$\begin{aligned} & \sum_{\{k,t | u_{kt} > 0\}} \text{Var}_{p_{kt}}[E_{U_{kt}}(U_{kt} | u_{kt}, p_{kt}) | \boldsymbol{\Omega}^*, \mathbf{X}^*] + E_{p_{kt}}[\text{Var}_{U_{kt}}(U_{kt} | u_{kt}, p_{kt}) | \boldsymbol{\Omega}^*, \mathbf{X}^*] \\ &= \sum_{\{k,t | u_{kt} > 0\}} u_{kt} \text{Var}_{p_{kt}}(p_{kt}^{-1} | \boldsymbol{\Omega}^*, \mathbf{X}^*) + u_{kt} E_{p_{kt}}((1 - p_{kt})p_{kt}^{-2} | \boldsymbol{\Omega}^*, \mathbf{X}^*) \end{aligned}$$

13 where expected values of functions of p_{kt} are computed with respect to the posterior
 14 density of \mathbf{p} defined by equation (2).

15 Our conclusion is that approximate Bayesian inference for $(\mathbf{U}, \mathbf{p}, \boldsymbol{\theta})_\delta$ given $\boldsymbol{\Omega}$ and
 16 \mathbf{X} can be obtained in two stages, each providing information about a subset of the
 17 parameters and depending on a subset of the data. For example, the posterior mean of
 18 U_t in equation (3) is approximated by computing the posterior means of $p_{1t}^{-1}, \dots, p_{Kt}^{-1}$
 19 conditional on the first release of each marked individual and then summing the
 20 Horwitz-Thompson like contributions $u_{kt} E_{p_{kt}}(p_{kt}^{-1} | \boldsymbol{\Omega}^*, \mathbf{X}^*)$. In practice, quantities

1 like $E(p_{kt}^{-1}|\boldsymbol{\Omega}^*, \mathbf{X}^*)$ will usually be intractable and instead can be estimated by sam-
 2 pling from the marginal posterior distribution via MCMC. Our results then imply
 3 that a sample of size D from the full posterior can be constructed in two separate
 4 steps:

- 5 1. for $d = 1, \dots, D$ generate $\tilde{\mathbf{p}}_i^d, \tilde{\boldsymbol{\theta}}_i^d$ from the distribution whose density is propor-
 6 tional to (2), and
- 7 2. for each d generate $\tilde{U}_{kt}^d \sim \text{Neg. Bin.}(u_{kt}, \tilde{p}_{kt}^d)$ for each k and t such that $u_{kt} = 0$.

8 The values $\tilde{U}_t^d = \sum_{\{k,t:u_{kt}>0\}} \tilde{U}_{kt}^d$, $d = 1, \dots, D$, then comprise a sample whose quan-
 9 tiles, mean, and variance approximate the quantiles, mean, and variance of $(U_t|\boldsymbol{\Omega}, \mathbf{X})_\delta$
 10 for δ very close to 0.

11 Although it might suffice to estimate U_t in some studies, estimating the total
 12 population size on occasion t requires further information about the number of marked
 13 individuals alive, M_t . If samples from $(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*)$ are generated by MCMC then
 14 a sample of values from the posterior of M_t can be obtained by defining a set of
 15 indicator variables, s_{it} $i = 1, \dots, n$, $t = 1, \dots, T$, that track when each marked
 16 individual is alive (as in the method of Royle, Dorazio, and Link, 2007). A complete
 17 set of indicators, \tilde{s}_{it}^d , will be generated on each MCMC iteration and the values
 18 $\tilde{M}_{t,d} = \sum_{i=1}^n \tilde{s}_{it,d}$ and $\tilde{N}_{t,d} = \tilde{U}_{t,d} + \tilde{M}_{t,d}$, $d = 1, \dots, D$ provide samples from the
 19 posterior distributions of M_t and N_t respectively.

20 **2.3 Extension to Continuous Covariates**

21 We extend our method to studies involving a continuous covariate, $x \in (a, b)$, by
 22 imagining that the range of the covariate can be partitioned into a large number of
 23 short intervals each containing only one value of the covariate. While this is impossible
 24 for a variable that is truly continuous, in practice, any variable can only be measured
 25 on a discrete scale and so it is conceptually possible to define a partition of intervals

1 which contain at most one observable value. We then apply the method for the
 2 discrete case above: first obtaining inference for the capture probabilities from the
 3 data conditional on the first release of each marked individual and then conditionally
 4 estimating the population size on each occasion.

5 To make this concrete, suppose the probability that individual i is captured on
 6 occasion t , given that it is alive, is $p(x_{it}|\beta_t)$ where $x_{it} \in (a, b)$ is the value of the
 7 covariate and β_t a set of (possibly time-dependent) parameters. Let $\pi_\beta(\beta)$ denote
 8 the prior density of $\beta = (\beta_1, \dots, \beta_T)$ and $L_2(\beta, \theta|\Omega^*, \mathbf{X}^*)$ the component of the
 9 likelihood modelling the data conditional on the first release of each marked indi-
 10 vidual. Following the procedure above, we simulate $\tilde{\beta}^d$ and $\tilde{\theta}^d$ from the posterior
 11 density $\pi_{(\beta, \theta)}(\beta, \theta|\Omega^*, \mathbf{X}^*) \propto L_2(\beta, \theta|\Omega^*, \mathbf{X}^*)\pi_\beta(\beta)\pi_\theta(\theta)$ via MCMC, incorporating
 12 the indicator variables s_{it}^d to compute \tilde{M}_t^d . Letting \mathcal{U}_t denote the set of individuals
 13 first captured on occasion t , we then generate $\tilde{U}_{it}^d \sim \text{Neg. Bin.}(1, p_t(x_{it}|\tilde{\beta}_t^d))$ for each
 14 $i \in \mathcal{U}_t$, set $\tilde{U}_t^d = \sum_{i \in \mathcal{U}_t} \tilde{U}_{it}^d$, and treat $\tilde{N}_t^d = \tilde{M}_t^d + \tilde{U}_t^d$, $d = 1, \dots, D$ as a sample from
 15 the posterior distribution of N_t .

16 2.4 Alternative Priors

17 It is possible in theory to apply the marginalization procedure for any specification
 18 of $\pi_U(\cdot)$ and $\pi_{\mathbf{p}}(\cdot)$. Summing the posterior density over U_{kt} for all k and t will
 19 always produce the marginal density of \mathbf{p} and θ , though the result will rarely be
 20 tractable or independent of \mathbf{u} . Moreover, two restrictions on the prior are needed
 21 to produce sensible results in applications with continuous covariates. First, it is
 22 necessary that $\pi_U(0) = 1/\delta$ with δ very small (conceptually $\delta \rightarrow 0$). Otherwise,
 23 $E(U_{kt}|u_{kt}, p_{kt})$ will be non-ignorably greater than 0 for those k and t such that $u_{kt} = 0$
 24 and $E(U_t|\mathbf{u}, \mathbf{p}) = \sum_{k=1}^K E(U_{kt}|u_{kt}, p_{kt})$ will be unreasonably large (tending to ∞ as

1 K increases). Second, our argument requires that:

$$\delta \sum_{U=1}^{\infty} (1 - \delta)^U \pi_U(U) \rightarrow 0$$

2 as $\delta \rightarrow 0$ so that $(U_{kt}|u_{kt}, p_{kt})_{\delta} \xrightarrow{P} 0$ if $u_{kt} = 0$ and values of the covariate which
 3 are not observed on occasion t can be ignored when making inference about U_t (see
 4 Appendix C for details).

5 One family of improper distributions that satisfies these requirements is given by:

$$\pi_U^{\delta}(U) \propto \begin{cases} 1/\delta & U = 0 \\ (1 - \alpha)^U / U & U > 0 \end{cases} \quad (4)$$

6 for $\alpha \in (0, 1)$. Similar calculations to those in the appendices show that as $\delta \rightarrow$
 7 0, $(\mathbf{p}, \boldsymbol{\theta} | \mathbf{u}, \boldsymbol{\Omega}, \mathbf{x})_{\delta}$ converges in distribution to a random variable whose density is
 8 proportional to:

$$L_2(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\Omega}^*, \mathbf{X}^*) \pi_{\mathbf{p}}(\mathbf{p}) \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \prod_{\{k,t:u_{kt}>0\}} \left(1 + \frac{\alpha}{p(1 - \alpha)} \right)^{-u_{kt}}$$

9 and that $U_{kt}|u_{kt}, p_{kt} \sim \text{Neg. Bin.}(u_{kt}, \alpha + p_{kt}(1 - \alpha))$ for those k and t such that
 10 $u_{kt} > 0$. In comparison to the posterior distribution derived in section 2.2, the new
 11 distribution assigns its mass to higher values of p_{kt} and smaller values of U_{kt} , not
 12 surprising as $(1 - \alpha)^U / U < 1/U$ for all $\alpha < 1$.

13 2.5 Application to the Jolly-Seber Model

14 To implement our method for any specific open-population model, one need only
 15 define the second component of the likelihood and the corresponding vector of pa-
 16 rameters, $\boldsymbol{\theta}$. Here, we provide details for the JS type model which is applied in the
 17 analysis of the Soay sheep data in section 3. Following Bonner and Schwarz (2006)

1 we define $L_2(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\Omega}^*, \mathbf{X}^*)$ as the product of complete data likelihood contributions
 2 for each marked individual. Let $a_i = \min\{t : \omega_{it} = 1\}$ denote the first occasion
 3 that individual i is captured and $s_{it} = 1$ if individual i is alive on occasion t and 0
 4 otherwise. The complete data likelihood contribution for individual i is:

$$L_{2i}(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\omega}_i, \mathbf{x}_i) = \prod_{t=a_i}^T \left[\begin{aligned} & \left(p(x_{it})^{\omega_{it}} (1 - p(x_{it}))^{1-\omega_{it}} \right)^{s_{it}} \\ & \cdot \left(\phi(x_{it})^{s_{it}} (1 - \phi(x_{it}))^{1-s_{it}} \right)^{s_{i,t-1}} \\ & \cdot f(x_{i,a_i+1}, \dots, x_{iT} | x_{i,a_i}) \end{aligned} \right]$$

5 where $f(x_{a_i+1}, \dots, x_{iT} | x_{a_i})$ denotes the density of $x_{i,a_i+1}, \dots, x_{iT}$ given x_{i,a_i} , $p_t(x_{it})$
 6 the probability that individual i is captured on occasion t given that it is alive and
 7 $\phi_t(x_{it})$ the probability that the individual survives to occasion $t+1$. The full likelihood
 8 contribution is then $L_2(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\Omega}^*, \mathbf{X}^*) = \prod_{i=1}^n L_{2i}(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\omega}_i, \mathbf{x}_i)$ and on each iteration of
 9 the MCMC algorithm a complete data set is generated by updating the unknown
 10 values of x_{it} and s_{it} conditional on all other random variables.

11 In section 3, we assume $p(\cdot)$ and $\phi(\cdot)$ to be linear functions of the covariate on the
 12 logistic scale with occasion specific intercepts but constant slopes for each age class
 13 (defined in the following section). That is:

$$\begin{aligned} \text{logit}(p(t, a, x)) &= \beta_{0ta} + \beta_{1a}x \\ \text{logit}(\phi(t, a, x)) &= \gamma_{0ta} + \gamma_{1a}x. \end{aligned} \tag{5}$$

14 for occasion t and age class a . Bonner and Schwarz (2006) model the density of the
 15 covariate by assuming that the distribution of $x_{i,t+1} - x_{i,t}$ is the same for all individuals
 16 alive on occasion t . This is reasonable in their analysis of the effect of body mass on
 17 the meadow vole (*Microtus pennsylvanicus*) because the data was limited to adults.
 18 The analysis of the Soay sheep presented in Section 3 includes individuals of all ages,

1 and so it is necessary to incorporate a more complicated model of body mass which
2 accounts for the growth of young sheep. The specific model we have selected is
3 based on the Ludwig von Bertalanffy (LVB) growth model with normally distributed,
4 additive errors described by Quinn and Deriso (1999, pg. 156). Given x_{it} we model
5 $x_{i,t+1}$ by:

$$x_{i,t+1}|x_{it} \sim N(x_{\infty}(1 - e^{-r}) + x_{it}e^{-r}, \sigma^2)$$

6 where x_{∞} is the asymptotic asymptotic body mass, $r > 0$ the rate parameter, and σ^2
7 the error variance. Given these definitions, the vector of parameters occurring only in
8 the second component of the likelihood is $\theta = (\gamma_{01}, \dots, \gamma_{0,T-1}, \gamma_1, x_{\infty}, r, \sigma^2)$.

9 **3 Application**

10 In this section, we apply our method to study the dynamics of a population of Soay
11 sheep living on the Scottish Island of Hirta. Hirta is a small island in the St. Kilda
12 Archipelago 200 km off the west coast of mainland Scotland. The sheep were originally
13 introduced to an even smaller island, Soay, though exactly when and by whom is not
14 clear. By 1930 the population of the islands was reduced to the point that the
15 remaining inhabitants were evacuated. At this time, all of the sheep were removed
16 from the islands except for a small number that were transferred from Soay to Hirta
17 where they established a feral population (Clutton-Brock and Pemberton, 2004).

18 Since 1986, the sub-population of sheep living near the old town site, Village Bay,
19 have been the subject of an intense study including annual capture and marking ses-
20 sions, visual censuses, and searches for dead animals. This study has provided much
21 information about the population's dynamics. There are no predatory or competing
22 species on the island, and so changes in the population depend mostly on competition
23 between the sheep. In years when food is plentiful and the population size is low,
24 the sheep are able to find food easily and the population grows rapidly. In years

1 when food is scarce or the population is high, there is strong competition amongst
2 the sheep and many die. Smaller, younger sheep have higher mortality rates because
3 they are more susceptible to the food shortages and males have lower survival rates
4 than females because of the extra expenditures associated with the fall rut (see the
5 papers in Clutton-Brock and Pemberton, 2004, for further details).

6 As an example of our method, we applied our model to the data collected from the
7 captures of the female sheep every August from 1986 to 2000. The data set comprised
8 capture histories for 833 sheep of which 488 (59%) were captured and weighed on one
9 occasion only (primarily because of the high mortality among lambs), 150 (17%)
10 on two occasions, and 205 (25%) on three or more. An average of 56 individuals
11 were marked in each of the 15 years, with counts ranging from 21 in 1986 to 96 in
12 1987. Following Catchpole, Morgan, and Tavecchia (2008) we have categorised the
13 individuals into 4 age classes: lambs (year of birth), yearlings (second year of life),
14 adults (ages 3-7), and seniors (8 years or older).

15 Our analysis compared the fit of four different models combining the possible
16 effects of age and body mass on the capture and survival probabilities. The most
17 general model, Model 1, allowed for the effects of both age and body mass on the
18 capture and survival probabilities by including age and occasion specific intercepts
19 and age dependent slopes for both $p(\cdot)$ and $\phi(\cdot)$, as defined in equation (5). Prior
20 distributions were defined separately for each age class such that:

$$\beta_{0ta} \sim N(\mu_{\beta,a}, \tau_{\beta,a}^2) \beta_{1a} \sim N(0, 10000)$$

21 and similar for γ_{0ta} and γ_{1a} . Note that the capture probability for the lambs is
22 not estimable, because lambs cannot be previously marked, and so we imposed the
23 restrictions $\beta_{0t1} = \beta_{0t2}$ and $\beta_{11} = \beta_{12}$. The remaining three models were obtained by
24 restricting the parameters of this model. Model 2 assumed no effect of age on either

1 the probability of capture or survival ($\beta_{0t1} = \dots = \beta_{0t4}$, $\beta_{11} = \dots = \beta_{14}$ and the
2 same for γ_{0ta} and γ_{1a}), Model 3 no effect of body mass ($\beta_{1a} = \gamma_{1a} = 0$ for all a),
3 and Model 4 no effect of age or body mass ($\beta_{0t1} = \dots = \beta_{0t4}$, $\gamma_{0t1} = \dots = \gamma_{0t4}$, and
4 $\beta_{1a} = \gamma_{1a} = 0$ for all a).

5 Samples from the posterior distributions for all four models were generated through
6 MCMC sampling implemented in OpenBUGS. Markov chains were run for 500,000
7 iterations total, realizations from the first 100,000 iterations were discarded as burn-
8 in, and the remaining realizations were thinned every 40 iterations to yield a sample
9 of size 10,000 from the posterior distribution. BUGS code for Model 1 is included in
10 the supplementary materials.

11 Table 1 presents the values of the deviance information criteria (DIC) used to
12 compare the four models (Spiegelhalter, Best, Carlin, and van der Linde, 2002).
13 The DIC increased monotonically as the complexity of the models decreased and the
14 smallest DIC value was produced the model including the effects of both age and
15 body mass (Model 1). DIC values for Models 2 and 3 including only one of age and
16 body mass were greater by 132 and 85 respectively providing very strong evidence
17 that both are important predictors of capture and/or survival.

18 The estimated survival and capture probabilities from the best fitting model are
19 plotted as functions of time for fixed values of body mass in Figure 1 and as functions
20 of the covariate for fixed time in Figure 2. Posterior summaries for the model pa-
21 rameters are provided in Table 2. Capture probabilities generally decreased with the
22 age of the sheep but were fairly consistent over time. There was some indication that
23 the capture probability of the lambs and yearlings decreased with increasing body
24 mass, but the 95% CIs of β_{1ta} were relatively wide and covered 0 for all age classes.
25 Survival probabilities for the seniors and adults were also close to constant over time,
26 varied considerably between years for the lambs and yearlings. Higher body mass
27 had a positive effect on the survival probabilities of all ages of sheep. This effect was

1 strongest for the lambs and yearlings, and weakest for the adults and seniors.

2 Posterior summary statistics for the number of female sheep alive and the number
3 recruited in each year from 1987–2000 are plotted in Figure 3. Reliable estimates
4 for 1986 could not be obtained because there was no direct information concerning
5 the capture-probabilities in this year. The population of female sheep seems to have
6 increased over the 15 year period with sharp declines followed by rebounds in 1989,
7 1992 and perhaps in 1999, though data from more recent years would be needed to
8 confirm a rebound. The number of female lambs born each year was fairly steady
9 from 1988 onward with the exception of decreases in 1989, 1992 and possibly 2000.
10 The posterior mean for the total number of sheep alive during at least one sampling
11 occasion was 954 with 95% CI (876, 1105).

12 Previous estimates of the survival probabilities by Catchpole, Morgan, Coulson,
13 Freeman, and Albon (2000) show very similar patterns of survival by age and time.
14 Estimated survival probabilities were highest for the adults and yearlings and lowest
15 for lambs. Estimates for all ages were lower than usual in the periods 1988-1989,
16 1991-1992 and 1994-1995, with the lambs being affected most and the adults least.
17 Estimates of the size of the female population also showed similar trends with low
18 numbers and few births in 1986, 1989 and 1992.

19 Fit of the LVB growth model was assessed by computing a posterior predictive
20 p-value (also called a Bayesian p-value) (Meng, 1994; Gelman, Carlin, Stern, and
21 Rubin, 2003, pg. 157-177). The discrepancy measure we selected compared the body
22 masses observed on each recapture with their expected values under the LVB model:

$$\frac{1}{n^{recap}} \sum_{i=1}^n \sum_{\{t:t>a_i, \omega_{it}=1\}} \left(\frac{z_{it} - E(z_{it}|\boldsymbol{\theta}, z_{it_i^-})}{\sigma \sqrt{t - t_i^-}} \right)^2$$

23 where $n^{recap} = \sum_{i=1}^n \sum_{t=a_i+1}^T \omega_{it}$ denotes the total number of recaptures over all in-
24 dividuals and $t_i^- = \max\{s = 1, \dots, t : \omega_{is} = 1\}$ the last occasion prior to t that

1 individual i was captured. The resulting p-value, .75, provided no evidence against
2 the fit of the LVB growth model.

3 [Table 1 about here.]

4 [Table 2 about here.]

5 [Figure 1 about here.]

6 [Figure 2 about here.]

7 [Figure 3 about here.]

8 **4 Simulation**

9 The performance of the new method was assessed through a simulation study based
10 on the results obtained in section 3. To simplify the simulation, data was generated
11 from a single age model with parameters based on the estimates for the yearling sheep,
12 which indicated a strong effect of the covariate on both the capture and survival
13 probabilities. We also present results from a second simulation scenario based on the
14 same parameter estimates but with capture probabilities reduced by approximately
15 one half. For comparison, all simulated data sets were analysed with both the new
16 model and a Bayesian implementation of the standard JS model ignoring the effect
17 of the covariate.

18 In each of the two scenarios, we generated 100 data sets comprising capture his-
19 tories for a population of 1000 individuals over 15 capture occasions. Data for each
20 individual were generated by first simulating the time of birth, then the values of the
21 covariate, and finally the survival and capture events. Time of birth was randomly
22 selected from a period of twenty time units, starting 5 units before the first capture
23 occasion to establish an adult population. The value of the covariate at birth was

1 simulated from a normal distribution with mean 11.6 and standard deviation 2.5, the
2 observed mean and standard deviation of the body mass of the Soay sheep lambs.
3 Values of the remaining parameters for the two scenarios are provided in Tables 3.
4 All models were fit via MCMC simulation in OpenBUGS. Markov chains were run for
5 50,000 iterations in total with the first 10,000 discarded as burn-in, and the remaining
6 40,000 thinned every 8 iterations.

7 Table 3 summarizes the performance of the new method in estimating the param-
8 eters of the capture probability, survival probability, and LVB growth model. The
9 posterior means for all parameters were negligibly biased in both scenarios, except
10 for the parameters τ^p and τ^ϕ whose posterior distributions were highly skewed. In
11 the first scenario, the 95% credible intervals covered their respective parameters for
12 at least 95/100 of the simulated data sets. Coverage of the 95% credible intervals was
13 slightly lower than nominal for some parameters in the second simulation scenario,
14 including both β_1 and γ_1 .

15 Figure 4 compares the error in the posterior mean population size on each occasion
16 for the new model and the Bayesian implementation of the standard JS model. Pos-
17 terior means from the model accounting for the effect of the covariate had negligible
18 bias (less than 1% on average in the first scenario an 3% in the second, including the
19 estimates from the first capture occasion which are affected by confounding between
20 the initial capture and survival probabilities). The 95% credible intervals covered the
21 true population size for more than 95% of the data sets in the first scenario and 87%
22 in the second. In comparison, posterior means from model ignoring the covariate
23 were positively biased by 6.6% in the first scenario and 23% in the second and the
24 95% credible intervals covered the true values for only 40% and 58% of the data sets
25 in the first and second scenarios respectively (even when results for the first occasion
26 was excluded).

27 [Table 3 about here.]

[Figure 4 about here.]

5 Discussion

The objective of this work was to develop a method to estimate the size of an open-population from mark-recapture data while properly accounting for the effects of individual, time-dependent covariates on the probability of capture and other demographic parameters. The simulation results in Section 4 clearly demonstrate the importance of including the effects of such quantities. In particular, the second scenario shows that estimates of the population size may be biased by more than 20% if the covariate is ignored. By conditioning the likelihood on the set of unmarked individuals alive on each occasion and selecting a specific prior for these parameters, our method yields a procedure that separates inference into two, intuitive steps. The posterior distribution of the capture probabilities and other demographic parameters is approximated conditional on the first release of each marked individual in the first step and then used to draw inference about the population size in the second. Section 2.5 details the method for the JS model, but the technique applies to any model which conditions on the set of unmarked individuals including recapture-resighting-recovery and more complex structures like the robust design of Pollock (1982).

One surprising result is that our method does not require a model of the distribution of the covariate for the animals that are never captured. This is important because, as Link and Barker (2005) demonstrated for completely unobserved random effects, different distributional assumptions may fit the data equally well but produce very different inferences about the population size. Of course, there is a trade-off between the bias and precision of the estimates of the populations size. Models which specify the complete distribution of the covariate will produce more precise estimates, but will be biased if the modelling assumptions are incorrect. Our method does not

1 suffer from this possible source of bias, but estimates will be less precise, in effect
2 because they allow for complete uncertainty about the unconditional distribution of
3 the covariate.

4 One drawback of our method is that the parameter N_t does not appear in the
5 likelihood and so it is not possible to incorporate prior knowledge concerning the
6 population size directly. A way to overcome this may be to employ a variation of
7 the Bayesian melding algorithm of (Poole and Raftery, 2000). In essence, Bayesian
8 melding provides a method of reweighting samples from the posterior distribution of
9 a complex model to account for prior beliefs about derived parameters which do not
10 appear in the likelihood. We believe that a similar approach could be applied for our
11 model and are currently exploring this possibility.

12 Another approach that has been applied to a wide range of capture-recapture
13 studies involving covariates (though not individual, time-dependent, and continuous)
14 is the data augmentation (DA) strategy of Royle et al. (2007). In essence, DA provides
15 a technique for sampling from the posterior distribution for a model involving an
16 unknown number of multinomial trials (e.g., the number of individuals ever available
17 for capture). If the number of trials is considered random then the set of unknowns,
18 including the covariate values for the unobserved individuals, will not be fixed. A
19 sample from the posterior distribution of such a model could be obtained with the
20 reversible jump MCMC algorithm of Green (1995), but this can be very complicated
21 to implement. Royle et al. (2007) avoid this by augmenting the data with a large,
22 fixed number of unobserved individuals. A complete data set is then generated on
23 each iteration of their MCMC algorithm by simulating the data for each of these
24 individuals including the values of the covariate and whether or not it belongs to the
25 true population. Realizations of the population size are generated simply by summing
26 the number of individuals in the true population on each iteration.

27 The same strategy could be applied to capture-recapture analysis involving con-

1 tinuous, time-dependent, individual covariates, but we perceive two disadvantages.
2 First, simulating complete data for the individuals in the augmented data set will
3 require modelling the covariate’s unconditional distribution. Second, the computa-
4 tional complexity of the DA method depends on the size of the augmented data set
5 and the method will be inefficient if the population size is very large (populations of
6 Pacific salmon which migrate along the Fraser and Columbia Rivers each year number
7 in the millions). In contrast, our solution only requires a model of the covariate’s dis-
8 tribution conditional on the value observed when an individual is first captured, and
9 its computational complexity depends only on the number of individuals captured.

10 We conclude with one important caveat to the application in section 3. Al-
11 though the trends in the estimated population size match well with previous results,
12 the absolute estimates are in fact smaller in every year. This is caused by forcing
13 $\beta_{0t1} = \beta_{0t2}$ and $\beta_{11} = \beta_{12}$. The capture probabilities of the yearlings decline with
14 body mass ($\beta_{12} < 0$) and extrapolating to the lower body masses observed for the
15 lambs results in significantly higher capture probabilities (approximately .80 for the
16 lambs versus .65 for the yearlings). We know from the complete data set includ-
17 ing recaptures and resightings that this overestimates the capture probabilities of the
18 lambs which leads to underestimation of the number of births and the population size.
19 Unfortunately, the capture probabilities for the lambs cannot be estimated without
20 some assumption relating the lambs to the other age classes. Alternatives would be
21 to completely exclude the lambs from the estimate of population size or to model the
22 intercept of the capture probability as a linear function of age.

23 **Acknowledgements**

24 This work was initiated while Simon Bonner was a PhD student at Simon Fraser
25 University and completed after he began a post-doctoral position at the University

1 of British Columbia. Both authors are funded by grants from the National Science and
2 Engineering Research Council of Canada.

3 We thank all of the past and current members of the Soay sheep project for
4 allowing the use of the data collected from the sheep of St. Kilda, the National Trust
5 for Scotland and the Scottish Natural Heritage who have supported this project, and
6 the army units stationed on St. Kilda and Benbecula that provided assistance to
7 the project. We are particularly grateful to Tim Coulson who answered our many
8 questions about the Soay sheep, Ted Catchpole who provided the formatted data,
9 and Byron Morgan who originally introduced Simon to the Soay sheep project.

10 **References**

11 A. N. Arnason. The estimation of population size, migration rates, and survival in a
12 stratified population. *Research in Population Ecology*, 15:1–8, 1973.

13 S. J. Bonner and C. J. Schwarz. An extension of the Cormack-Jolly-Seber model for
14 continuous covariates with application to *Microtus pennsylvanicus*. *Biometrics*, 62
15 (1):142–149, March 2006.

16 A. D. Carothers. The effects of unequal catchability on Jolly-Seber estimates. *Bio-*
17 *metrics*, 29(1):79–100, 1973.

18 E. A. Catchpole, B. J. T. Morgan, T. N. Coulson, S. N. Freeman, and S. D. Albon.
19 Factors influencing soay sheep survival. *Applied Statistics*, 49(4):453–472, 2000.

20 E. A. Catchpole, B. J. T. Morgan, and G. Tavecchia. A new method for analysing dis-
21 crete life-history data with missing values. *Journal of the Royal Statistical Society*
22 – *Series B*, 70:445–460, 2008.

23 T. H. Clutton-Brock and J. M. Pemberton, editors. *Soay Sheep: Dynamics and*
24 *Selection in an Island Population*. Cambridge University Press, Cambridge, 2004.

- 1 J. A. Dupuis and C. J. Schwarz. A Bayesian approach to the multistate Jolly-Seber
2 capture-recapture model. *Biometrics*, 63(4):1015–1022, 2007.
- 3 A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*.
4 Chapman and Hall, 2nd edition, 2003.
- 5 R. O. Gilbert. Approximations of the bias in the Jolly-Seber capture-recapture model.
6 *Biometrics*, 29(3):501–526, 1973.
- 7 P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian
8 model determination. *Biometrika*, 82(4):711–7332, 1995.
- 9 W.-D. Hwang and A. Chao. Quantifying the effects of unequal catchabilities on
10 Jolly-Seber estimators via sample coverage. *Biometrics*, 51(1):128–141, 1995.
- 11 G. M. Jolly. Explicit estimates from capture-recapture data with both death and
12 immigration-stochastic model. *Biometrika*, 52:225–247, 1965.
- 13 R. King and S. P. Brooks. On the bayesian analysis of population size. *Biometrika*,
14 88(2):317–336, 2001.
- 15 R. King, S. P. Brooks, and T. Coulson. Analyzing complex capture-recapture data
16 in the presence of individual and temporal covariates and model uncertainty. *Bio-*
17 *metrics*, 64:1187–1195, 2008.
- 18 J.-D. Lebreton, K. P. Burnham, J. Clobert, and D. R. Anderson. Modelling survival
19 and testing biological hypotheses using marked animals: A unified approach with
20 case studies. *Ecological Monographs*, 62(1):67–118, March 1992.
- 21 W. A. Link and R. J. Barker. Modeling association among demographic parameters
22 in analysis of open population capture-recapture data. *Biometrics*, 61:46–54, 2005.

- 1 T. L. McDonald and S. C. Amstrup. Estimation of population size using open capture-
2 recapture models. *Journal of Agricultural, Biological, and Environmental Statistics*,
3 6(2):206–220, 2001.
- 4 X.-L. Meng. Posterior predictive p-values. *The Annals of Statistics*, 22(3):1142–1160,
5 1994.
- 6 D. L. Otis, K. P. Burnham, G. C. White, and D. R. Anderson. Statistical inference
7 from capture data on closed animal populations. *Wildlife Monographs*, 62:3–135,
8 1978.
- 9 S. Pledger and M. Efford. Correction of bias due to heterogeneous capture probability
10 in capture- recapture studies of open populations. *Biometrics*, 54(3):888–898, 1998.
- 11 S. Pledger, K. H. Pollock, and J. L. Norris. Open capture-recapture models with
12 heterogeneity: I. Cormack-Jolly-Seber model. *Biometrics*, 59:786–794, 2003.
- 13 K. H. Pollock. A capture-recapture design robust to unequal probability of capture.
14 *Journal of Wildlife Management*, 46(3):752–757, 1982.
- 15 D. Poole and A. E. Raftery. Inference for deterministic simulation models: The
16 Bayesian melding approach. *Journal of the American Statistical Association*, 95
17 (452):1244–1255, 2000.
- 18 T. J. Quinn, II and R. B. Deriso. *Quantitative Fish Dynamics*. Oxford University
19 Press, Inc., New York, 1999.
- 20 J. A. Royle, R. M. Dorazio, and W. A. Link. Analysis of multinomial models with
21 unknown index using data augmentation. *Journal of Computational and Graphical*
22 *Statistics*, 16(1):67–85, 2007.
- 23 C. J. Schwarz, J. F. Schweigert, and A. N. Arnason. Estimating migration rates using
24 tag-recovery data. *Biometrics*, 49(1):177–193, Mar. 1993.

1 G. A. F. Seber. A note on the multiple recapture census. *Biometrika*, 52:249–259,
2 1965.

3 D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures
4 of model complexity and fit (with discussion). *Journal of the Royal Statistical*
5 *Society B*, 64:583–640, 2002.

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

8 Appendices

9 **A Expectation of U_{kt} when $\pi_{p_{kt}}(p) = 1(0 < p < 1)$**

10 To demonstrate what may happen if $\pi_{\mathbf{p}}(\cdot)$ is positive over all of $(0, 1)^{KT}$ when
11 $\pi_{U_{kt}}(0) = 1/\delta$, let $\pi_{\mathbf{p}}(\mathbf{p}) = \prod_{k=1}^K \prod_{t=1}^T 1(0 < p_{kt} < 1)$. If no individuals are cap-
12 tured in group k on occasion t then p_{kt} will not appear in the second component of
13 the likelihood, and the posterior density of U_{kt} and p_{kt} will be defined by:

$$\pi_{U_{kt}, p_{kt}}^{\delta}(U, p | \boldsymbol{\Omega}, \mathbf{X}) \propto \begin{cases} \frac{1}{\delta} & U = 0 \\ \frac{(1-p)^U}{U} & U > 0 \end{cases}.$$

14 Integrating with respect to p and computing the normalizing constant produces the
15 marginal posterior density for U_{kt} :

$$\pi_{U_{kt}}^{\delta}(U_{kt} | \boldsymbol{\Omega}, \mathbf{X}) = \begin{cases} \frac{1}{1+\delta} & U = 0 \\ \frac{\delta}{U(U+1)(1+\delta)} & U > 0 \end{cases}$$

1 and hence:

$$\begin{aligned} E^\delta(U_{kt}|\mathbf{\Omega}, \mathbf{X}) &= \frac{\delta}{1+\delta} \sum_{U=1}^{\infty} \frac{1}{(U+1)} \\ &= \infty \end{aligned}$$

2 for all $\delta > 0$. However, if $\pi_{\mathbf{p}}(\mathbf{p}) \propto \prod_{k=1}^K \prod_{t=1}^T 1(\delta < p < 1)$ then:

$$E^\delta(U_{kt}|\mathbf{\Omega}, \mathbf{X}) = \frac{\delta \sum_{U=1}^{\infty} \frac{(1-\delta)^{U+1}}{U+1}}{1 + \delta \sum_{U=1}^{\infty} \frac{(1-\delta)^{U+1}}{U(U+1)}}$$

3 which is finite for all $\delta > 0$ and tends to 0 as $\delta \rightarrow 0$.

4 **B Proof of Theorem 1**

5 **Proof of Theorem 1.1**

6 **Lemma 1.1.a** *The sequence:*

$$\sum_{U=1}^{\infty} \frac{(1-p)^U}{U}$$

7 *converges for all $p \in (0, 1)$ and is a decreasing function of p .*

8 *Proof.* For fixed p , the ratio of consecutive terms is:

$$(1-p) \frac{U}{U+1}$$

9 which tends to $(1-p) < 1$ as $U \rightarrow \infty$. The ratio test then implies that the series is

10 absolutely convergent for all $p \in (0, 1)$.

11 Differentiating the series with respect to p yields:

$$-\sum_{U=1}^{\infty} (1-p)^{U-1} = -\frac{1-p}{p}$$

1 which is negative for all $p \in (0, 1)$. Hence, the sum is decreasing as a function of
 2 p . □

3 **Lemma 1.1.b** *The function:*

$$g(\delta) = \delta \sum_{U=1}^{\infty} \frac{(1-\delta)^U}{U}$$

4 converges to 0 as $\delta \rightarrow 0$ from above.

5 *Proof.* By application of l'Hopital's rule:

$$\begin{aligned} \lim_{\delta \rightarrow 0^+} g(\delta) &= \lim_{\delta \rightarrow 0^+} \delta^2 \sum_{U=1}^{\infty} (1-\delta)^{U-1} \\ &= \lim_{\delta \rightarrow 0^+} \delta \\ &= 0 \end{aligned}$$

6 □

7 Let $\epsilon > 0$. For fixed δ , the posterior marginal density of \mathbf{p} and $\boldsymbol{\theta}$ is given by:

$$\begin{aligned} \pi_{\mathbf{p}, \boldsymbol{\theta}}^{\delta}(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\Omega}, \mathbf{X}) &\propto \sum_{U_{11}=u_{11}}^{\infty} \cdots \sum_{U_{KT}=u_{KT}}^{\infty} L_1(\mathbf{U}, \mathbf{p} | \mathbf{u}) \cdot L_2(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\Omega}^*, \mathbf{X}^*) \cdot \pi^{\delta}(\mathbf{U}, \mathbf{p}, \boldsymbol{\theta}) \\ &= \prod_{k,t} \left(\sum_{U_{kt}=u_{kt}}^{\infty} L_{1kt}(U_{kt}, p_{kt} | u_{kt}) \pi_U^{\delta}(U_{kt}) \right) \cdot L_2(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\Omega}^*, \mathbf{X}^*) \pi_{\mathbf{p}}^{\delta}(\mathbf{p}) \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}). \end{aligned}$$

8 For those k and t such that $u_{kt} > 0$:

$$\begin{aligned} \sum_{U_{kt}=u_{kt}}^{\infty} L_{1kt}(U_{kt}, p_{kt} | u_{kt}) \pi_U^{\delta}(U_{kt}) &= \sum_{U_{kt}=u_{kt}}^{\infty} \binom{U_{kt}}{u_{kt}} p_{kt}^{u_{kt}} (1-p_{kt})^{U_{kt}-u_{kt}} \cdot \frac{1}{U_{kt}} \\ &= \frac{1}{u_{kt}} \sum_{U_{kt}=u_{kt}}^{\infty} \binom{U_{kt}-1}{u_{kt}-1} p_{kt}^{u_{kt}} (1-p_{kt})^{U_{kt}-u_{kt}} \\ &= \frac{1}{u_{kt}} \end{aligned}$$

- 1 noting that the final sum is the probability that a negative binomial random variable
 2 with parameters u_{kt} and p_{kt} is greater than or equal to u_{kt} , which is 1. Since u_{kt} is part
 3 of the observed data, this contribution to the posterior can be ignored completely.
 4 For those k and t such that $u_{kt} = 0$:

$$\begin{aligned} \sum_{U_{kt}=u_{kt}}^{\infty} L_{1kt}(U_{kt}, p_{kt}|u_{kt})\pi_U^\delta(U_{kt}) &= \frac{1}{\delta} + \sum_{U_{kt}=u_{kt}}^{\infty} \frac{(1-p_{kt})^{U_{kt}}}{U_{kt}} \\ &\in \left(\frac{1}{\delta}, \frac{1+g(\delta)}{\delta} \right) \end{aligned}$$

- 5 for all $p \in (\delta, 1)$ by Lemma 1.1.a. Hence:

$$\begin{aligned} \pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{X}) &< \frac{\prod_{\{k,t:u_{kt}=0\}} \left(\frac{1}{\delta} + g(\delta)\right) L_2(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) \pi_{\mathbf{p}}^\delta(\mathbf{p}) \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\int \int \prod_{\{k,t:u_{kt}=0\}} \left(\frac{1}{\delta}\right) L_2(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) \pi_{\mathbf{p}}^\delta(\mathbf{p}) d\mathbf{p} d\boldsymbol{\theta}} \\ &< (1+g(\delta))^{KT} \frac{L_2(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) \pi_{\mathbf{p}}^\delta(\mathbf{p}) \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\int \int L_2(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) \pi_{\mathbf{p}}^\delta(\mathbf{p}) d\mathbf{p} d\boldsymbol{\theta}} \\ &= (1+g(\delta))^{KT} \pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) \end{aligned}$$

- 6 By Lemma 1.1.b $\exists \delta_1 > 0$ such that $\delta < \delta_1$ implies:

$$g(\delta) < (1+\epsilon)^{1/KT} - 1$$

- 7 which implies that:

$$\pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{X}) < (1+\epsilon) \pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*).$$

- 8 Integrating over the rectangle $(\delta, p_{11}) \times \dots \times (\delta, p_{KT})$ then implies:

$$\begin{aligned} \Pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{X}) &< (1+\epsilon) \Pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) \\ &< \Pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) + \epsilon. \end{aligned}$$

1 Similarly:

$$\pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{X}) > (1 + g(\delta))^{-KT} \pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*)$$

2 and by Lemma 1.1.b $\exists \delta_2 > 0$ such that $\delta < \delta_2$ implies:

$$g(\delta) < (1 + \epsilon)^{1/KT} - 1$$

3 which in turn implies that:

$$\Pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{X}) > \Pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) + \epsilon.$$

4 Thus, $\forall \delta < \min(\delta_1, \delta_2)$:

$$|\Pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{X}) - \Pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*)| < \epsilon$$

5 which is the definition of convergence in distribution.

6 **Proof of Theorem 1.2**

7 Note that $\pi_{\mathbf{p},\boldsymbol{\theta}}^\delta(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) \rightarrow \pi_{\mathbf{p},\boldsymbol{\theta}}(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*)$ for all \mathbf{p} and $\boldsymbol{\theta}$ such that $\pi_{\mathbf{p},\boldsymbol{\theta}}(\mathbf{p}, \boldsymbol{\theta}|\boldsymbol{\Omega}^*, \mathbf{X}^*) >$

8 0. Theorem 1.2 is then an immediate consequence of Scheffe's Lemma (see e.g. ?,

9 pg. ?).

1 C Proof of Theorem 2

2 *Proof.* For any k and t such that $u_{kt} > 0$ the full conditional distribution of U_{kt} has
 3 density:

$$\begin{aligned}\pi_{U_{kt}}^\delta(U_{kt}|u_{kt}, p_{kt}) &\propto \binom{U_{kt}}{u_{kt}} p^{u_{kt}} (1-p)^{U_{kt}-u_{kt}} \cdot \frac{1}{U_{kt}} \\ &\propto \binom{U_{kt}-1}{u_{kt}-1} p^{u_{kt}} (1-p)^{U_{kt}-u_{kt}}.\end{aligned}$$

4 which is the density of a negative binomial random variable with parameters u_{kt} and
 5 p_{kt} , independent of δ .

6 For any k and t such that $u_{kt} > 0$ the full conditional probability that $U_{kt} = 0$ is:

$$\frac{\binom{1}{\frac{1}{\delta}}}{\left(\frac{1}{\delta} + \sum_{U=1}^{\infty} \frac{(1-p_{kt})^U}{U}\right)} = \left(1 + \delta \sum_{U=1}^{\infty} \frac{(1-p_{kt})^U}{U}\right)^{-1} \rightarrow 1$$

7 as $\delta \rightarrow 0$ for all $p_{kt} \in (0, 1)$, which implies that $U_{kt} \xrightarrow{P} 0$.

8 Slutsky's theorem then implies that as $\delta \rightarrow 0$:

$$(U_t|\mathbf{u}, \mathbf{p})_\delta = \sum_{k,t} (U_{kt}|u_{kt}, p_{kt})_\delta \xrightarrow{D} \sum_{\{k,t:u_{kt}>0\}} U_{kt}|u_{kt}, p_{kt}$$

9 where $U_{kt}|u_{kt}, p_{kt} \sim \text{Neg. Bin.}(u_{kt}, p_{kt})$. □

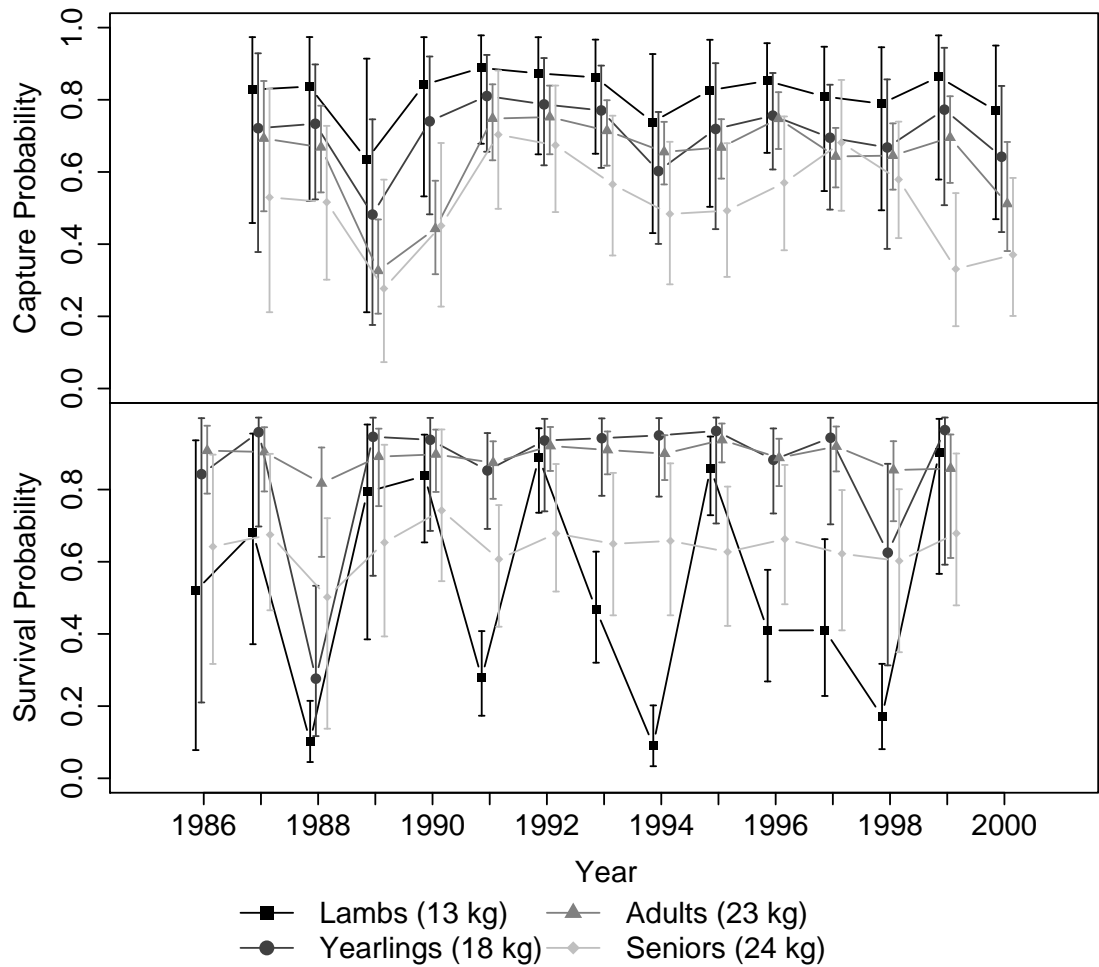


Figure 1: Effect of time on the capture and survival probabilities of the female sheep. The plotted values summarize the posterior distribution of the capture probability (top) and survival probability (bottom) of sheep with the average observed body mass for each age class, as indicated in the legend. For each age \times year the point represents the posterior mean and the error bars indicate the extents of the 95% CIs.

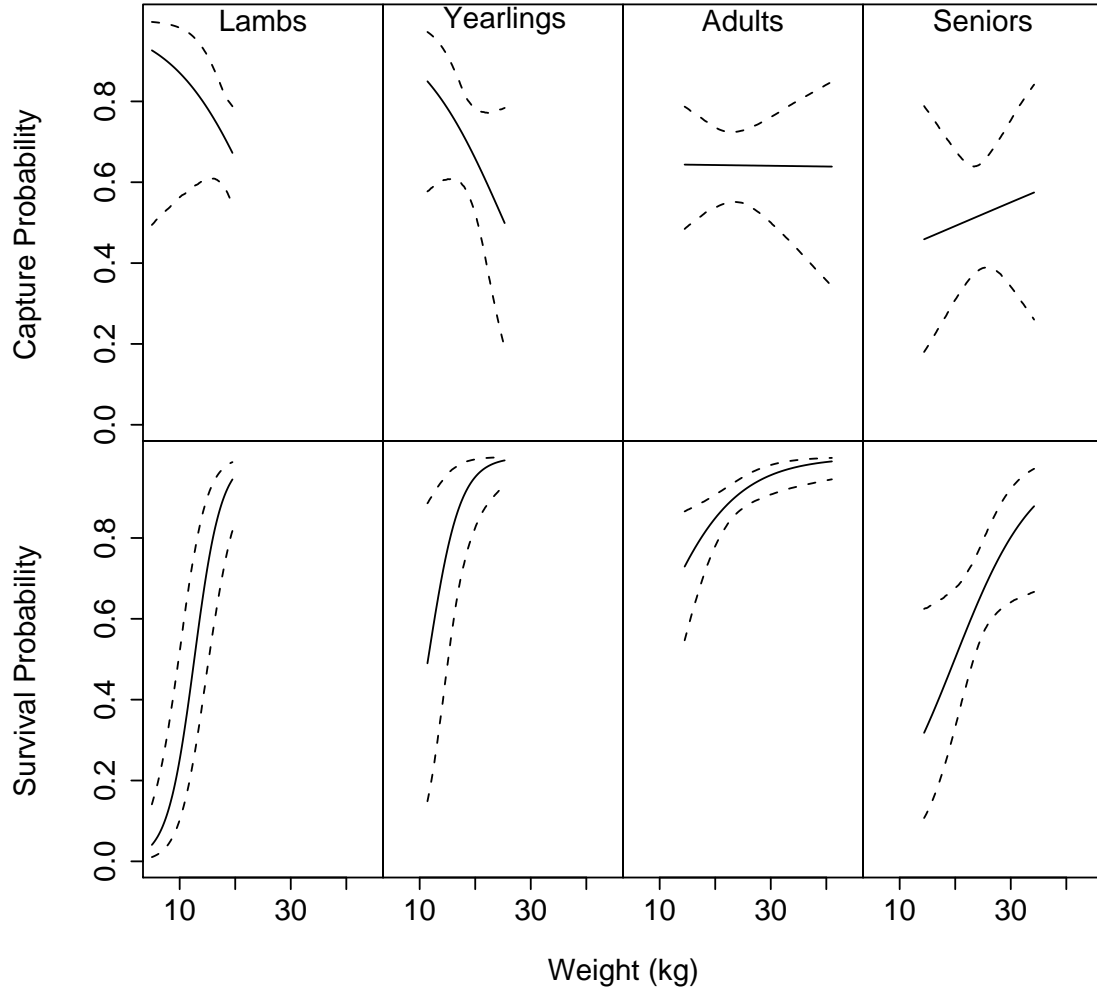


Figure 2: Effect of body mass on the capture and survival probabilities of the female sheep. The solid lines indicate the posterior mean capture probability (top) and survival probability (bottom) as a function of body mass, with the intercept parameters equal to the hierarchical mean for each age group. The dotted lines indicate the extents of the pointwise 95% CIs. For each age class, the posterior distributions are only summarized over the range of observed covariate values.

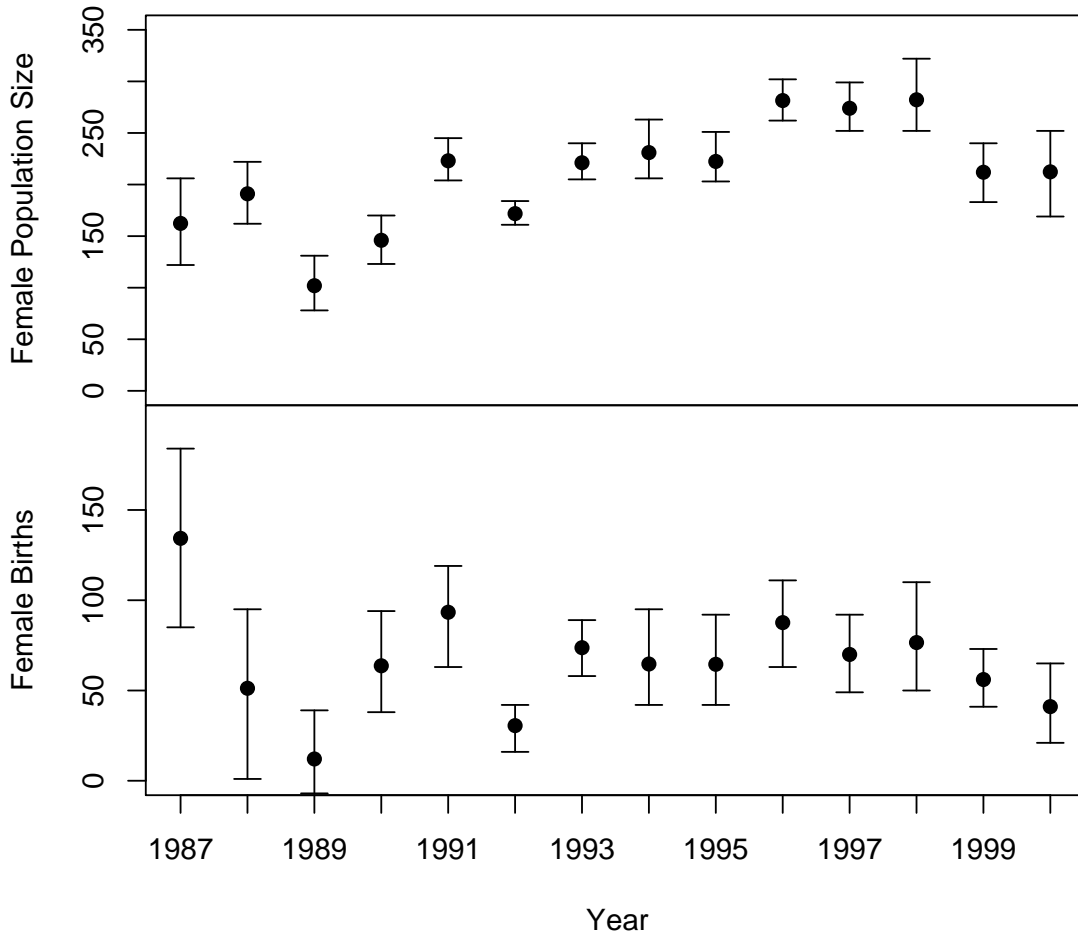


Figure 3: Estimated female population size and number of female births by year. Points in the top plot indicate the posterior mean population size for each year with 95% CIs represented by the error bars. Points in the bottom plot indicate the posterior mean number of births per year with 95% CIs represented by the error bars.

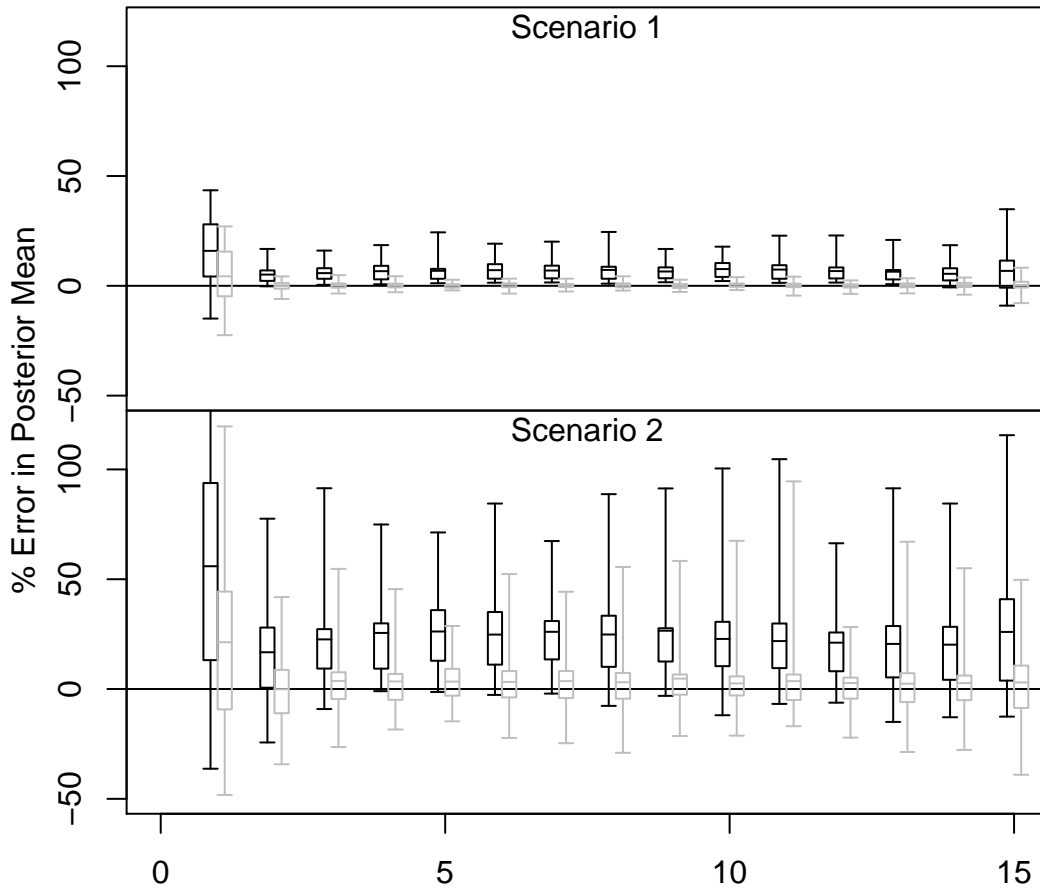


Figure 4: Error in the posterior mean estimate of population size for the two simulation scenarios. For each of the scenarios, the black symbols represent the distribution of errors for the standard JS model and the grey symbols the distribution of errors for the modelling including the covariate. The boxes extend between the first and third quantiles of the error, and the whiskers between the 2.5 and 97.5 percentiles. The line through each box represents the mean error (bias).

Table 1: DIC selection criterion for the four models fit to the female Soay sheep data.

Model	DIC	p_D
1) $p(t \times w \times a), \phi(t \times w \times a)$	3612.1	66.1
2) $p(t \times a), \phi(t \times a)$	3697.0	59.7
3) $p(t \times w), \phi(t \times w)$	3744.2	29.3
4) $p(t), \phi(t)$	3937.9	21.4

Table 2: Posterior summary statistics for parameters of Model 4 including the effects of both age and body mass on the capture and survival probabilities. Statistics provided are the posterior mean and 95% CI in parentheses. Parameters of the covariate distribution are common to all sheep. The remaining parameters vary by age class.

Par.	Lambs	Yearlings	Adults	Seniors
μ_β		3.16(-0.44, 7.14)	0.60(-0.97,2.08)	-0.50(-3.72,2.81)
τ_β		0.64(0.08, 1.21)	0.63(0.34,0.98)	0.73(0.14,1.30)
β_1		-0.12(-0.33, 0.07)	0.00(-0.06,0.06)	0.02(-0.12,0.14)
μ_γ	-5.23(-7.10,-3.48)	-4.02(-8.43,-0.21)	-0.95(-2.98,1.08)	-2.76(-5.88,0.22)
τ_γ	1.79(0.96, 2.76)	1.74(0.60, 3.29)	0.55(0.06,1.11)	0.49(0.04,1.33)
γ_1	0.41(0.29, 0.53)	0.35(0.14, 0.61)	0.13(0.05,0.22)	0.14(0.01,0.27)

Table 3: Results of the simulations based on the model fit to the Soay sheep. Performance of the model in the two scenarios was assessed by the percent bias of the posterior mean, and the percent width and coverage of the 95% credible interval of the key model parameters.

Par.	Scenario 1				Scenario 2			
	Truth	Bias	Width	Cover.	Truth	Bias	Width	Cover.
μ^p	3.5	-1.1	54.8	96.0	0.6	0.6	395.1	91.0
τ^p	0.6	4.0	95.0	97.0	0.6	0.6	98.8	98.0
β_1	-0.1	1.4	74.0	95.0	0.1	-0.1	144.2	89.0
μ^ϕ	-3.5	2.7	69.7	95.0	-3.5	-3.4	83.7	96.0
τ^ϕ	1.7	11.6	72.0	95.0	1.7	1.9	68.9	99.0
γ_1	0.3	0.9	19.1	98.0	0.3	0.3	34.6	92.0
z_∞	24.5	0.0	1.9	98.0	24.5	0.0	2.7	94.0
r_z	0.5	0.0	11.9	97.0	0.5	0.0	19.3	94.0
σ_z	2.0	-0.2	5.6	99.0	2.0	0.0	10.0	90.0