A Bayesian approach to the multi-state Jolly-Seber Capture-Recapture Model

Jerome A. Dupuis∗

Laboratoire de Statistique et Probabilités, Université Paul Sabatier,
Toulouse, France

and

Carl James Schwarz†

Department of Statistics and Actuarial Science, Simon Fraser University,
Burnaby, BC, Canada V5A 1S6

Summary. This paper considers a Bayesian approach to the multi-state extension of the Jolly-Seber model commonly used to estimate population abundance in capture-recapture studies. It extends the work of George and Robert (1992, Biometrika 79, 677-683) which dealt with the Bayesian estimation of a closed population with only a single state for all animals. A superpopulation is introduced to model new entrants in the population. Bayesian estimates of abundance are obtained by implementing a Gibbs sampling algorithm based on data augmentation of the missing data in the capture histories when the state of the animal is unknown. Moreover, a partitioning of the missing data is adopted to ensure the convergence of the Gibbs

∗email: Jerome.Dupuis@math.ups-tlse.fr
†email: cschwarz@stat.sfu.ca
sampling algorithm even in presence of impossible transitions between some states. Lastly, we apply our methodology to a population of fish to estimate abundance and movement.

KEY WORDS: abundance; bayesian estimates; capture-recapture; missing data; multi-state model; stratified Jolly-Seber model.

1. Introduction

The Jolly-Seber (Jolly, 1965; Seber, 1965) is the standard model for estimating abundance in capture-recapture experiments in open-populations. A key assumption of this model is homogeneity in capture and survival rates among animals in the population and it is well known that heterogeneity in catchability can cause bias in estimates of abundance (Pledger and Efford, 1998; Link, 2003). In some cases, this heterogeneity is associated with animals in different states such as geographic locations or behavioral modes. Schwarz and Arnason (1996) presented methodology when the states are permanent (e.g. sex), and Schwarz, Schweigert, and Arnason (1993) or Brownie et al. (1993) developed methods for non-permanent states (e.g. geographical movement) when abundance is not estimated. Arnason (1973) and Seber (1982) developed moment estimators to estimated abundance when movement among states occurred in death-only models. Dupuis (1995) and King and Brooks (2002) presented a bayesian solution for the movement case (without estimating abundance).

In this paper we develop a bayesian solution to the multi-state Jolly-Seber model. The bayesian approach avoids the serious computational difficulties of maximum likelihood estimation due to the complexity of the likelihood of
the observed data. We begin by extending the super-population approach of Schwarz and Arnason (1996) to model new entrants into the population. A missing data structure is developed that is different from that used in Dupuis (1995) and in King and Brooks (2002). Current approaches to simulate the missing data by conditioning upon the immediately preceding and following states (referred to afterwards as the component-by-component approach) can be computationally costly when the number of missing states is large, and can fail to be ergodic in presence of impossible transitions between some states. A new data augmentation scheme is proposed where missing data is simulated block-by-block rather than component-by-component that solves these problems.

2. Missing data structure, assumptions, and modeling

The experimental protocol is standard, and is not reviewed in this paper (Schwarz et al., 1993). Let $T$ represent the number of capture occasions (including the first tagging period) during the course of the study; $K = \{1, \ldots, k\}$ be the set of possible states in the study; $n$ be the total number of marked animals during the course of the study. We denote by $P$ the super-population of size $N$, i.e. all animals liable to be captured in at least one capture occasion over all time periods and states in the study. The state space $K$ is augmented with state $\dagger$ representing death, and state $*$ representing the state of an animal which has not yet joined the study. Throughout the paper, $\mathbb{I}_{(c)}$ denotes an indicator function that takes the value 1 if the condition $c$ is true and zero otherwise.
2.1 **Statistics**

While the individual capture histories are used as data input, we also make use of the following sufficient statistics:

- \( M_t(r) \) the number of animals marked at time \( t \) in state \( r \)
- \( C_{t_1}^{t_2}(r, s) \) the number of animals captured at time \( t_1 \) in state \( r \) and next captured in state \( s \) at time \( t_2 \)
- \( D_t(r) \) the number of animals captured at time \( t \) in state \( r \) and which have not been recaptured afterwards.

2.2 **The missing data structure**

Let \( z_{(i,t)} \) represent the state of animal \( i \) at time \( t \). If \( z_{(i,t)} = r \), then animal \( i \) has entered the study and is alive at time \( t \) in state \( r \in K \); if \( z_{(i,t)} = \dagger \), it is dead at time \( t \), and if \( z_{(i,t)} = * \), animal \( i \) has not yet entered the study at time \( t \). The vector \((z_{(i,t)}; t = 1, \ldots, T)\) is denoted by \( z_i \). Let \( y_i \) represent the observed capture history of animal \( i \). For example, the observed history \( y_i = (0\ 0\ 2\ 0\ 0\ 0\ 1\ 1\ 0) \) with \( T = 9 \), and \( k = 2 \), indicates that animal \( i \) was first captured, marked and released at time \( t = 3 \) in state 2, and was recaptured in state 1 at times \( t = 7, 8 \). It was not captured at times \( t = 1, 2, 4, 5, 6, \) or \( 9 \). If \( y_{(i,t)} = 0 \) then \( z_{(i,t)} \) is unknown i.e. is a missing data value. If an animal is never captured, all the \( z_{(i,t)} \)'s are missing. The key difference from previous work is the need to take into account the missing states prior to the first capture of an animal and to account for the complete set of missing states for animals never observed.

Finally, we denote by \( x_i \) the capture-process derived from \( y_i \), where \( x_{(i,t)} = 1 \) (or \( x_{(i,t)} = 0 \)) if the animal \( i \) is captured (or not captured) at time \( t \). Then \( y_{(i,t)} = r \) is equivalent to \( x_{(i,t)} = 1 \) and \( z_{(i,t)} = r \), and
\( y_{(i,t)} = 0 \) is equivalent to \( x_{(i,t)} = 0 \). From our example history above, we have \( x_i = (0 0 1 0 1 1 0 0 0) \). History \( y_i \) is thus viewed as the stacking of processes \( z_i \) and \( x_i \) (which is a missing data indicator of \( z_i \)).

Introducing the two underlying processes \( x_i \) and \( z_i \) presents some advantages. It allow us to rigorously describe the missing data structure, and facilitates the calculation of the conditional distributions involved in the Gibbs sampling algorithm by making use of some properties in directed graphs.

### 2.3 Model and Assumptions

The usual assumptions for the multi-state model are made. For simplicity, we assume that there is no loss on capture. The individual processes \((x_i, z_i)\) are independently and identically distributed; thus we assume that animals behave independently with respect to capture, survival and movement, and that movement is directed by a first-order Markov chain.

The conditional independence assumptions between the \( z_{(i,t)} \)'s and \( x_{(i,t)} \)'s are summarized in the directed graph \( G \) in Figure 1.

[Figure 1 about here.]

### 2.4 Parameters, abundances, and number of entrants

\( q_t(r, s) \) transition probability from \( r \in K \) to \( s \in K \cup \{\dagger\} \) between times \( t \) and time \( t + 1 \). \( q_t(r, s) = \Pr(z_{(i,t+1)} = s | z_{(i,t)} = r) \).

\( \phi_t(r) \), It is often convenient to decompose \( q_t(r, s) \) as the product of a survival probability \( \phi_t(r) \) and an inter-state movement probability \( \psi_t(r, s) \), namely: \( q_t(r, s) = \phi_t(r) \psi_t(r, s) \).

\( p_t(r) \) capture probability at time \( t \) in \( r \); \( p_t(r) = \Pr(x_{(i,t)} = 1 | z_{(i,t)} = r) \).
probability that an animal enters the study between \( t \) and \( t + 1 \) and is located in state \( r \) at \( t + 1 \). Thus, \( \beta_t(r) = \Pr (z_{(i,t)} = *, z_{(i,t+1)} = r) \). By convention \( \beta_0(r) = \Pr (z_{(i,1)} = r) \). \( \sum_t \sum_r \beta_t(r) = 1. \)

\[ \theta \]

the set of parameters formed by the capture, movement, survival, and entry parameters.

\[ N \]

the size of the super-population.

\[ N_t(r) \]

the number of animals in the study population located in state \( r \) at time \( t \).

\[ E_t(r) \]

the number of entrants into the study into state \( r \) between times \( t \) and \( t + 1 \). \( E_0(r) = N_1(r) \).

The relationship between the parameters for a time-dependent model is illustrated in Figure 2.

Although parameter \( N \) is not (in general) a quantity of biological interest, it is an integral part of our approach. Note that \( N_t(r) \) and \( E_t(r) \) are not parameters, but functions of \( N \) and of the \( z_{(i,t)} \)'s since \( E_t(r) = \sum_{i=1}^{N} \mathbb{I}(z_{(i,t)} = *, z_{(i,t+1)} = r) \) and \( N_t(r) = \sum_{i=1}^{N} \mathbb{I}(z_{(i,t)} = r) \).

2.5 The prior distributions placed on \( N \) and \( \theta \)

In many situations, prior information on the size of the super-population size \( N \) will be unlikely. The improper prior distribution of Jeffrey, \( \pi(N) \propto \frac{1}{N} \) is often used (King and Brooks, 2001). The main alternative to the Jeffrey prior is \( \pi(N) \propto 1 \) (e.g. Castledine, 1981). Later, we show that our results are largely unaffected by the choice between these two priors. Beta prior
distributions are placed on the $\phi_t(r)$’s and on the $p_t(r)$’s. Dirichlet prior distributions are placed on $\beta_t = (\beta_t(r); r = 1, \ldots, k; t = 0, \ldots, T - 1)$ and on $\psi_t(r, \cdot) = (\psi_t(r, s); s = 1, \ldots, k)$. All priors are assumed to be independent of each other; in particular, we assume that $\pi(\theta, N) = \pi(\theta)\pi(N)$.

3. The Gibbs sampling algorithm.

Let $\mathcal{L}(\theta, N|\mathcal{D})$ denotes the likelihood of $\theta$ and $N$ from data $\mathcal{D}$; $z_m$ denote the missing data (which is defined conditionally on $\mathcal{D}$ and $N$); and $\xi$ denotes the vector $(N, z_m)$. Starting with an initial value $\theta^{(0)}$, we iterate steps of the form $(l \geq 1)$:

$$\xi^{(l)} \sim f(\xi|\theta^{(l-1)}, \mathcal{D}) \quad \theta^{(l)} \sim \pi(\theta|\xi^{(l)}, \mathcal{D})$$

where $f(\cdot)$ denote a probability mass function. We simulate $\xi$ in two steps:

$$N \sim \pi(N|\theta, \mathcal{D}) \quad \text{and} \quad z_m \sim f(z_m|N, \theta, \mathcal{D}).$$

3.1 The missing data simulation phase (data augmentation)

The simulation of the missing state information for animal $i$ begins by partitioning the missing state information into blocks (of different types) depending upon the pattern of missing data.

The Type I block consists of sample times up to (but excluding) the time of first capture. The Type II blocks consist of the consecutive sample times where the state of the animal was not observed but is bracketed by two capture-events. The Type III block consists of the sample times after the last sample time an animal was seen to the end of the experiment. For an animal that is never captured, all the $z_{(i,t)}$’s are missing and there is a single
Type IV block consisting of all sample occasions.

For example, consider the history presented in Section 2.2. It is cut into 3 blocks of Type I, II and III respectively: $B_1(i) = \{z_{(i,1)}, z_{(i,2)}\}$, $B_2(i) = \{z_{(i,4)}, z_{(i,5)}, z_{(i,6)}\}$, $B_3(i) = \{z_{(i,9)}\}$.

Then for a given history $y_i$, let $z_m(i)$ represent the set of the missing $z_{(i,t)}$ related to animal $i$, and $B_j(i)$ is the set of blocks partitioning the missing data for animal $i$ such as $z_m(i) = \bigcup_{j=1}^{J(i)} B_j(i)$.

The distribution of the set of missing values over the entire super-population, $f(z_m|N, D, \theta)$, is found as follows. For convenience, we will drop $N$ and $\theta$ in the notation for conditioning.

First, due to the assumption of independence among animals, $f(z_m|D) = \prod_{i=1}^{N} f(z_m(i)|y_i)$.

Second, if animal $i$ has been marked, the joint distribution of the missing data blocks, $f(z_m(i)|y_i)$ is shown by Theorem 1 to be the product of the distribution of the missing values in the individual blocks. We denote by $\tilde{B}_j(i)$ the neighborhood of block $B_j(i)$ which is defined as the subset of $G$ formed by the variables directly connected to $B_j(i)$. For example, $\tilde{B}_j(i) = \{z_{(i,t_j-1)}, z_{(i,t_j+1)}\} \cup \{x_{(i,t_j)}, \ldots, x_{(i,t'_j)}\}$, if $B_j(i) = (z_{(i,t_j)}, \ldots, z_{(i,t'_j)})$ and is of type II.

**Theorem 1.** Let $i$ be designate a marked animal. Conditionally on $y_i$, the set of blocks for that animal, $B_j(i)$ are independent. Moreover:

$$f(B_j(i)|y_i) = \prod_{j=1}^{j=J(i)} f(B_j(i)|\tilde{B}_j(i))$$

**Proof.** The proof lies on local and global Markov properties in directed
graphs. It is straightforward and has been omitted.

Finally, the distribution of the missing values within each block for each animal, \( f(B_j(i)|\bar{B}_j(i)) \) differs according to the type of \( B_j(i) \) and is derived in the Web Appendix A.

Two blocks of Type I, II or III are said to be similar if they have the same type, and are considered to have the same length if they are enclosed at identical times by the same observed states. Similar blocks are simulated using a multinomial distribution in a single step. For example, the set of the missing blocks of type II, of length \( a \), and enclosed by states \( u \) and \( v \) at time \( t_1 \) and \( t_2 \), will be simulated conditionally on \( C_{t_1}^{u_2}(u,v) \), using the multinomial distribution \( M(\gamma_h; h = 1, \ldots, k^a) \) where \( \gamma_h \) represents the probability of path \( H \) in the block of missing data conditional upon the observed states \( u \) and \( v \) and is derived in the Web Appendix A. Similar derivations can be made for a block of type I, III or IV. Multinomial distributions of type I are generated conditionally on the \( M_t(r) \)'s, and those of type III conditionally on the \( D_t(r) \)'s. For the blocks of type IV (animals never seen), the (unique) multinomial is generated conditionally on \( N - n \).

The missing data \( z_m \) is simulated conditionally on \( N \) and on the sufficient statistics given earlier.

Let us return to the multinomial distribution above. For ‘large’ values of \( k \) and \( a \), the number of cells can be high, and simulation can be costly. In such a situation, we simulate using a series of \( 1 + k(a - 1) \) consecutive multinomial with \( k \) cells, rather than a unique multinomial with \( k^a \) cells; this requires storage for only \( ak \) terms instead of \( k^a \) terms. A similar strategy can be used to deal with the simulation of the other multinomials of types I, II, III, and IV.
III, and IV. Details appear in a technical report available from the authors.

3.2 The parameter simulation phase

3.2.1 Super-population $N$ To find the posterior distribution $\pi(N|\theta, D)$ we start from $\pi(N|\theta, D) \propto L(\theta, N|D)\pi(N)$. It is easy to show that:

$$
\pi(N|\theta, D) \propto \left(\frac{N}{n}\right)^{N-n} \left[1 - \sum_{t=1}^{T} \sum_{r=1}^{k} \alpha_t(r)p_t(r)\right] \pi(N) \mathbb{1}_{(N \geq n)} \tag{3.2}
$$

where $\alpha_1(r) = \beta_0(r)$ and $\alpha_{t+1}(s) = \beta_t(s) + \sum_r \alpha_t(r)(1-p_t(r))q_t(r,s)$. Note that $\lambda = \sum_t \sum_r \alpha_t(r)p_t(r)$ represents the probability for any animal of $P$ to be marked during the experiment. From (3.2) we have: $N - n \sim NB (n, \lambda)$ if $\pi(N) \propto 1/N$ and $N - n \sim NB (n+1, \lambda)$ if $\pi(N) \propto 1$. Since 1 is typically very small compared to $n$, the Bayesian estimates of any quantity of interest are practically unaffected by the choice of the non-informative prior distribution placed on $N$.

3.2.2 Other parameters The posterior distribution $\pi(\theta|N, z_m, D)$ depends upon the complete data likelihood $L(\theta, N|z_m, D)$ based on the augmented data from above. From the assumptions listed in Section 2.2, one has

$$
L(\theta, N|z_m, D) \propto L_{\phi,\psi} \times L_\beta \times L_p
$$

where

$$
L_{\phi,\psi} = \prod_{1 \leq t \leq T-1} \left\{ \phi_t(r)^{W_t(r)} \left(1 - \phi_t(r)\right)^{W_t(r,\dagger)} \right\} \prod_{s \in K} \psi_t(r,s)^{W_t(r,s)}
$$
\[
L_\beta = \prod_{0 \leq t \leq T - 1} \beta_t(r)^{E_t(r)}
\]
\[
L_p = \prod_{1 \leq r \leq k} p_t(r)^{U_t(r)} \{1 - p_t(r)\}^{V_t(r)}.
\]

where \( W_t(r, s) = \sum_{i=1}^{N} I(z_{(i,t)} = r, z_{(i,t+1)} = s); W_t(r) = \sum_{s \in K} W_t(r, s); U_t(r) = \sum_{i=1}^{N} I(x_{(i,t)} = 1, z_{(i,t)} = r); \) and \( V_t(r) = \sum_{i=1}^{N} I(x_{(i,t)} = 0, z_{(i,t)} = r). \) Since \( \pi(\theta | N, z_m, D) \) is proportional to \( \mathcal{L}(\theta, N | z_m, D) \pi(\theta), \pi(\theta) \) is conjugate for the complete data likelihood. Simulation of \( \theta \) at step \((l)\) proceeds as follows:

\[
\phi_{(l)}(r) \sim \text{Be} \left( \eta_t(r) + W_t^{(l)}(r), \mu_t(r) + W_t^{(l)}(1, \dag) \right)
\]
\[
\psi_{(l)}(r, \cdot) \sim \text{Dir} \left( e_t(r, 1) + W_t^{(l)}(r, 1), \ldots, e_t(r, k) + W_t^{(l)}(r, k) \right)
\]
\[
\beta_{(l)} \sim \text{Dir} \left( \gamma_t(r) + E_t^{(l)}(r); r = 1, \ldots, k; t = 0, \ldots, T - 1 \right)
\]

where \( \eta_t(r), \mu_t(r), e_t(r, s), \alpha_t(r), b_t(r), \) and \( \gamma_t(r) \) are known hyperparameters.

3.3 Convergence issues

**Theorem 2.** Let \( \xi = (N, z_m) \). The sequence \((\xi^{(l)})\) is a first-order Markov chain with invariant distribution the posterior distribution of \( \xi \). It is ergodic, even in presence of impossible transitions between states of \( K \).

The proof appears in the Web Appendix B. A similar result exists concerning the sequence \((\theta^{(l)})\). Any posterior quantity of interest related to \( \xi = (z_m, N) \) can thus be approximated by applying the ergodic theorem to the Markov chain \((\xi^{(l)})\).

When the component-by-component scheme is adopted (as in Dupuis, 1995 and in King and Brooks, 2002), the above result is generally not true. In some circumstances, the Markov chain generated by the Gibbs sampling can fail to be ergodic and yields wrong results. This can occur when some
transitions between states of $K$ are impossible. Lebreton and Pradel (2002) give examples of such a situation. In the Web Appendix C, we provide an example for which the Gibbs sampling based on a component-by-component missing data scheme fails to be ergodic.

4. **Constraints on parameter and identifiability issues**

Restrictions on parameters can be enforced in two ways. First, parameter identification and design matrices can be used in much the same way as in the computer program MARK (White and Burnham, 1999). Second, the prior distributions can be specified to place very strong prior belief upon the value of a parameter. Both methods can be used in the computer program developed for this project. Using strong priors to implement restrictions on parameters (e.g. that a certain component is zero) runs into difficulties when constraints are imposed on parameters which belong to a set of parameters which sum to 1 and when a Dirichlet prior distribution is placed on the vector. It is not possible, via a Dirichlet distribution, to incorporate a strong prior on some components and a weak prior on other components (Dupuis 2002). To overcome this difficulty, we reparameterize the $\beta_t(r)$. Define $\zeta_t$ as the conditional probability that an animal enters the study population between $t$ and $t + 1$ given that it enters on or after $t$, and $\beta^*_t(r)$ as the conditional probability of entering into state $r$ given that the animal entered the study population between $t$ and $t + 1$: i.e.

$$\zeta_t = \frac{\sum_{r=1}^{k} \beta_t(r) \sum_{j=t}^{T-1} \sum_{r=1}^{k} \beta_j(r)}{\sum_{j=t}^{T-1} \sum_{r=1}^{k} \beta_j(r)} \quad \text{and} \quad \beta^*_t(r) = \frac{\beta_t(r)}{\sum_{r=1}^{k} \beta_t(r)}.$$
It is easy to verify that:

$$\beta_t(r) = (1 - \zeta_0)(1 - \zeta_1)(1 - \zeta_2) \ldots (1 - \zeta_{t-1})\zeta_t \beta^*_t(r)$$

Independent Beta distributions are then put on the $\zeta_t$ and a Dirichlet distribution is put on $(\beta^*_t(1), \ldots, \beta^*_t(k))$ for each $t$.

Although the bayesian methodology does not suffer from problems caused by parameter non-identifiability it is still useful to understand which parameters are non-identifiable in the maximum likelihood framework. Gimenez, Choquet, and Lebreton (2003) present a general methodology for determining identifiability in the Arnason-Schwarz model and found a correspondence between identifiability problems in the single-state and multi-state Arnason-Schwarz models. We find a similar correspondence between confounded parameters of the Jolly-Seber model and our multi-state extension. For example, in the full model with all parameters fully time dependent, the first immigration, distribution, and capture parameters are confounded and cannot be separated without restrictions on parameters.

5. Example

Mille Lacs (Minnesota) is a large lake with a population of northern pike ($Esox lucius$). Tags were applied (approximately 2000/year) and recoveries took place on the spawning grounds between 1994 and 1998 (inclusive); only recoveries (no new tags released) occurred on the spawning grounds between 1999 and 2004 (inclusive); and tagging and recoveries resumed in 2005. A second gillnet recovery look place later in 2005. Tagging and recovery effort was not spread uniformly around the lake and preliminary plots showed some
movement between spawning sites over years. Full details are available in Bruesewitz and Reeves (2005) and Schwarz (2005) and summary statistics are presented in Web Table 1.

Tag loss was estimated through double tagging studies and is estimated to be between 2-4% per year and has been ignored.

There are two sources of heterogeneity in this study which call for a multi-state analysis. First, while most spawning grounds were targeted during tagging periods, there is variation in catchability. While effort was fixed (i.e. so many net days per spawning ground), the abundance of spawners varies considerably across spawning grounds. Second, additional recoveries, (a tribal harvest) did not occur on the same spawning grounds where tags were applied, and from 1998 to 2004 no harvest occurred on several major spawning grounds.

Consequently, the lake was divided into two states corresponding (roughly) to high and low catchability and if a tribal harvest did or did not occur. While it may be possible to create additional states, simple plots of movements showed that fish were fairly faithful to these two states with low interchange. A capture history for a fish is a 13 digit vector corresponding to year (1994-2005, with 2005 subdivided into two capture periods corresponding to the spawning time and the gillnet assessment). The complete set of capture histories is available on the Biometrics web site.

Because no new tags were applied from 1994 to 2004, models with year-specific survival rates should not be fit as year-specific survival rates are not estimable (in the likelihood sense). Consequently, only a model with constant survival rates over time were fit. Based upon a literature review, a
beta $B(8, 2)$ prior with a mean of .8 was used.

We had no strong beliefs about movement between the two states, but northern pike are known to be fairly faithful to spawning areas so Dirichlet $D(8, 2)$ (for state 1) and Dirichlet $D(2, 8)$ (for state 2) priors were imposed which give an prior belief of a .80 fidelity rate (and a .20 movement rate) that was again constant over the years.

Capture rates were believed to be modest, so Beta $B(1, 9)$ priors with a mean of .10 were used for the year-state specific capture-rates.

Recruitment to the population (small fish growing larger than the 18 inch cutoff) should occur in every year but we have no prior information on exceptionally strong or weak year classes so we modeled new entrants using a beta $B(1, 1)$ prior on initial population present ($\zeta_0$) which implies that about 50% of the super-population was initially available (in 1994) and 50% of the super-population recruited in future years. Subsequent recruitment parameters ($\zeta_1, \zeta_2, \ldots$) had $B(1, 10)$, $B(1, 9)$, etc. priors imposed representing approximate equal fractions of the remaining superpopulation entering each subsequent year. A uniform prior was used for the distribution of new recruits among the two states in each year.

Finally, loss on capture rates are variable across time, with tribal harvest and gillnet assessments having close to a 100% loss on capture. Either $B(1, 9)$ or $B(9, 1)$ were used as prior distributions as appropriate.

Twenty-thousand iterations were used for the burn-in phase; an additional 500,000 iterations were performed. Several different values for the parameters of the prior distributions and different starting points for the iterations yielded comparable results.
Fit of the model was assessed in several ways. First, we used the contingency table tests developed by Pradel, Wintrebert, and Gimenez (2003). These failed to find any evidence of lack of fit. Second, a Bayesian p-value was computed as outlined by King and Brooks (2002). Our plot (Figure ??) showed no evidence of a lack of fit with a Bayesian p-value of 0.15. Third, residual plots can be constructed comparing the observed counts of each history with its average (over the posterior distribution) expected count. Our residual plot (Figure 4) showed a good fit to the observed data except for a histories that had a single observed fish but a very small expected count. Lastly, the stratification used was very coarse, and there likely remains unaccounted heterogeneity. For example, there is some evidence that northern pike varied in catchability by length and/or sex. The ratio of the average deviances between the real and simulated datasets can be thought of as a variance-inflation factor used in maximum likelihood situations.

[Figure 3 about here.]

[Figure 4 about here.]

The estimated total (over both states) population sizes in each year is presented in Table 1 along with estimates from a standard Jolly-Seber analysis of the same data. The multi-state and standard estimates are comparable in 1994 to 1998 when a proper tagging study over all tagging areas was conducted; but appear to differ considerably in 1999 and 2000 when tribal harvest did not occur in all areas. Both methods give very imprecise estimates in 2001-2003 when the number of marks captured is very small as there are few marked fish still alive. An empirical investigation showed that minor
changes in the number of marked fish captured in these years gives rise to wide swings in the estimates of abundance for those years, but estimates in other years were unaffected. The estimates coincide again in 2005 when a large number of tags were again applied and recaptured later in that year.

[Table 1 about here.]

The remaining estimates are summarized in Table 2. Estimated yearly survival rates are about 60% and may be equal in both states. [This simpler model was not fit.] The fish have a moderate chance of mixing between the states with between 15% and 30% of each population moving to the other states between sampling occasions.

[Table 2 about here.]

6. Conclusion

This paper presented a natural extension to the Jolly-Seber model using a multi-state model. Combining the missing data and the super population approaches turns out to be a efficient strategy for obtaining the bayesian estimates.

Our paper proposes a block-by-block data augmentation scheme as opposed to a component-by-component scheme. Our method is computationally less costly and the resulting Gibbs sampling produces an ergodic Markov chain, even in presence of impossible transitions. Since we only work with complete data, our approach easily extends to more complex directed graphs than the one considered in Section 2.3. For example, it is possible to extend our block-by-block method to second-order Markov models but Theorem 1 needs to be modified.
The methods in this paper can also be used for abundance estimation in multi-state closed populations (Schwarz and Ganter, 1995) by constraining the entrance and survival parameters; and for estimation in age structured population (Pollock 1981) by constraining the transition matrix.

We did not consider model selection in this paper. This can be approached using a RJMCMC approach similar to King and Brooks (2002) or using a deviance information criterion (DIC) which is much less computationally intensive.

Computer code is available from the Biometrics website. It has been extensively exercised using simulation studies to proof the code.

7. Supplementary Material

Web Appendices and Web Tables referenced in Section 3.1 and Section 5, capture histories, and computer codes are available under the Paper Information link at the Biometrics website http://www.tibs.org/biometrics.

Acknowledgements

This work is supported by a research grant from the Canadian Natural Science and Engineering Research Council (NSERC) and contracts from the Minnesota Department of Natural Resources to CJS. JAD received support from the Fonds France Canada pour la Recherche (FFCR) sponsored by the Service scientifique de l’Ambassade de France au Canada.

8. References


Figure Legends

Figure 1:
Directed graph for the first-order Markov model.

Figure 2:
Organization of parameters in the time-dependent model. Parameters marked with an asterisk are confounded in the likelihood and require additional restrictions or strong prior information to estimate separately. The super-population size is
\[ N = \sum_{t=0}^{T-1} \sum_{r=1}^{k} E_t (r). \]

Figure 3:
Goodness of fit plot of the Tukey chi-square statistic of the real counts for histories compared to its expected count and the same statistic for simulated data. The bayesian \( p \)-value is the proportion of times the real data statistic was greater than the simulated data statistic.

Figure 4:
Standardized residual plot of the difference between the observed history counts and the average expected history counts. Many of the extreme residuals have an observed count of 1 with very small expected counts.
Figure 1. Directed graph for the first-order Markov model.
### Figure 2.

Organization of parameters in the time-dependent model. Parameters marked with an asterisk are confounded in the likelihood and require additional restrictions or strong prior information to estimate separately. The size of the super-population is

\[ N = \sum_{t=0}^{T-1} \sum_{r=1}^{k} E_t(r). \]
Figure 3. Goodness of fit plot of the Tukey chi-square statistic of the real counts for histories compared to its expected count and the same statistic for simulated data. The estimated bayesian $p$-value is the proportion of times the real data statistic was greater than the simulated data statistic computed for each iteration of the Markov Chain Monte Carlo simulation.
Figure 4. Standardized residual plot of the difference between the observed history counts and the average expected history counts. Many of the extreme residuals have an observed count of 1 with very small expected counts.
Table 1
Population estimates (thousands of fish) of northern pike from a standard Jolly-Seber and our multi-state Jolly-Seber model. The left two sets of columns are a comparison of the total population; the right two sets of columns are estimates for the individual states. *ne* indicates that parameter is not estimable in the likelihood sense.

<table>
<thead>
<tr>
<th>Year</th>
<th>Standard Jolly-Seber (likelihood)</th>
<th>Total from Multi-state Jolly-Seber (bayesian)</th>
<th>State 1 Multi-State Jolly-Seber (bayesian)</th>
<th>State 2 Multi-state Jolly-Seber (bayesian)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>SE</td>
<td>Est</td>
<td>SD</td>
</tr>
<tr>
<td>1994</td>
<td>ne</td>
<td>ne</td>
<td>ne</td>
<td>ne</td>
</tr>
<tr>
<td>1995</td>
<td>13.7</td>
<td>1.1</td>
<td>13.5</td>
<td>1.1</td>
</tr>
<tr>
<td>1996</td>
<td>20.1</td>
<td>1.3</td>
<td>20.4</td>
<td>1.3</td>
</tr>
<tr>
<td>1997</td>
<td>29.1</td>
<td>2.3</td>
<td>28.4</td>
<td>2.2</td>
</tr>
<tr>
<td>1998</td>
<td>35.7</td>
<td>3.2</td>
<td>33.7</td>
<td>3.1</td>
</tr>
<tr>
<td>1999</td>
<td>85.5</td>
<td>19.5</td>
<td>76.5</td>
<td>14.7</td>
</tr>
<tr>
<td>2000</td>
<td>81.0</td>
<td>14.9</td>
<td>76.5</td>
<td>14.7</td>
</tr>
<tr>
<td>2001</td>
<td>152.3</td>
<td>42.0</td>
<td>157.5</td>
<td>36.6</td>
</tr>
<tr>
<td>2002</td>
<td>118.2</td>
<td>37.2</td>
<td>122.7</td>
<td>25.0</td>
</tr>
<tr>
<td>2003</td>
<td>94.3</td>
<td>32.7</td>
<td>93.9</td>
<td>17.9</td>
</tr>
<tr>
<td>2004</td>
<td>75.1</td>
<td>4.8</td>
<td>67.2</td>
<td>8.9</td>
</tr>
<tr>
<td>2005</td>
<td>46.5</td>
<td>3.2</td>
<td>46.9</td>
<td>3.1</td>
</tr>
</tbody>
</table>
Table 2

*Estimates of survival and movement from the northern pike study.*

<table>
<thead>
<tr>
<th>State</th>
<th>Yearly Survival Mean (SD)</th>
<th>Estimated movement to State 1 Mean (SD)</th>
<th>Estimated movement to State 2 Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.66 (0.02)</td>
<td>0.87 (0.02)</td>
<td>0.13 (0.02)</td>
</tr>
<tr>
<td>2</td>
<td>0.58 (0.02)</td>
<td>0.29 (0.05)</td>
<td>0.71 (0.05)</td>
</tr>
</tbody>
</table>