

Bayesian estimation of movement and survival probabilities from capture-recapture data

By JEROME A. DUPUIS

*Laboratoire de Statistique et Probabilités, Université Paul Sabatier, 118 Route de Narbonne,
31062 Toulouse, Cedex, France*

SUMMARY

The Arnason–Schwarz model is usually used for estimating survival and movement probabilities of animal populations from capture-recapture data. The missing data structure of this capture-recapture model is exhibited and summarised via a directed graph representation. Taking advantage of this structure we implement a Gibbs sampling algorithm from which Bayesian estimates and credible intervals for survival and movement probabilities are derived. Convergence of the algorithm is proved using a duality principle. We illustrate our approach through a real example.

Some key words: Bayes estimator; Capture-recapture; Directed graph; Duality principle; Gibbs sampling; Missing data; Movement probability; Survival probability.

1. INTRODUCTION

Knowledge of survival and movement probabilities is vital for analysing changes in dynamics of animal populations (Clobert & Lebreton, 1991; Hestbeck, Nichols & Malecki, 1991). Brownie et al. (1993) and Schwarz, Schweigert & Arnason (1993) propose numerical approaches for calculating the maximum likelihood estimators of these parameters when information is provided by capture-recapture data of tagged animals moving in a stratified space. It is usually assumed that capture probabilities are stratum-specific and time-specific, and that the migration process is directed by a nonhomogeneous first-order Markov chain; following Brownie et al. (1993) we refer to these assumptions as the Arnason–Schwarz model. Brownie et al. (1993) also consider a behaviour with ‘memory’ by modelling the migration process as a second-order Markov chain; we refer to this model as the Hestbeck–Brownie model. Bayesian procedures have not previously been developed in the setting of open populations, and a fortiori in open and stratified populations, although information about capture and survival parameters is often available (Pollock, 1991).

In a tag-recovery data set-up Pollock (1991) and Schwarz et al. (1993), dealing with the Arnason–Schwarz model, stressed the computational difficulties due to the missing data corresponding to the failures of capture between release and recovery. Moreover the very complicated form of the likelihood of the Hestbeck–Brownie model constrained Brownie et al. (1993) to make inference from a conditional likelihood which requires the discarding of individual capture-recapture histories including two successive capture failures. From a Bayesian point of view, the computational difficulties are even more considerable. George & Robert (1992) have suggested a missing data representation to simplify Bayesian inference through use of Monte Carlo Markov chain algorithms. More generally,

this approach has been advocated by Smith & Roberts (1993) for overcoming the prohibitive computational constraints inherent in missing data models.

The outline of the paper is as follows. The multiple capture-recapture protocol is recalled in § 2.1. In § 2.2, we show how capture-recapture data from open and moving populations belongs in the framework of missing data models. The Arnason–Schwarz model assumptions are given in § 2.3. In § 3, we describe the Gibbs sampling algorithm from which the posterior quantities of interest are derived, both for the Arnason–Schwarz model and for the Hestbeck–Brownie model. Convergence issues are examined in § 3.3. In § 4, we implement our approach with multiple capture-recapture data from a *Lacerta vivipara* population moving in a zone including three contiguous areas.

We develop our approach with the Arnason–Schwarz model when inference is based upon multiple capture-recapture data. However it also applies to the Hestbeck–Brownie model, see § 3, and with tag-recovery data.

2. THE ARNASON–SCHWARZ MODEL

2.1. The multiple capture-recapture experiment

We consider individuals from an open animal population which move in a zone, K , including k geographical strata. We denote by \dagger the location which corresponds to ‘exit’ from K : an animal in \dagger is either dead or outside K . The experimental protocol is standard; see e.g. Brownie et al. (1993). In each stratum, at time $j = 1$, captured animals receive an individual mark and are returned to the population. Then a second sample is taken: the locations of recaptured animals are recorded, the unmarked animals are marked and all are returned to the population. We repeat this procedure until the end of the experiment. We denote by m the number of capture occasions, including the first tagging period: capture-recapture data thus include $m - 1$ cohorts. We assume there is no loss on capture, no mark loss and that no observation can be collected from animals in state \dagger . Moreover, the marked sample is assumed representative; see Schwarz et al. (1993) for more details. In particular we assume that tagging does not influence the migration behaviour.

2.2. The missing data structure

We denote by y_i the capture-recapture history related to the animal i ; for instance, a possible occurrence when $k = 3$ and $m = 8$ is

$$y_i = r t . s t \quad (2.1)$$

This sequence means that the i th animal marked at time $j = 1$ in r has been recaptured at times $j = 2, 4, 5$ in areas t, s, t respectively. It has not been recaptured at times $j = 3, 6, 7, 8$, its locations at these times being therefore missing.

For (2.1), we define the corresponding process

$$x_i = 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0,$$

where $x_{(i,j)} = 1$ if the animal i is recaptured at time j and $x_{(i,j)} = 0$ otherwise. Therefore y_i can be viewed as the stacking of two processes: a process of interest, the migration process denoted by z_i , and the capture-recapture process x_i which is a missing data indicator of z_i . This missing data structure is described in § 2.3 by means of a directed graph.

We define

$$M_i = \{j > t_i | x_{(i,j)} = 0\}, \quad C_i = \{j \geq t_i | x_{(i,j)} = 1\},$$

where $1 \leq t_i \leq m-1$ is the time at which animal i has been tagged. The set of animals i such that $t_i \leq j$ is denoted by \mathcal{T}_j . We denote by $z_{(i,j)}$ the position of the animal i at time $j \geq t_i$. Let $(y_1, \dots, y_i, \dots, y_n)$ be denoted by y , and $(x_1, \dots, x_i, \dots, x_n)$ by x , where n denotes the number of tagged animals.

2.3. Assumptions and parameters

Assumption 1. The individual processes (x_i, z_i) are independently and identically distributed.

The dependence structure between x_i and z_i is summarised in graph \mathcal{G}_1 below; see e.g. Lauritzen & Spiegelhalter (1988) for details. This representation allows a clear and compact formulation of the conditional independence assumptions between these two processes:

$$\begin{array}{ccccccc} & \dots & x_{(i,j-1)} & & x_{(i,j)} & & x_{(i,j+1)} & \dots \\ \mathcal{G}_1 & & \uparrow & & \uparrow & & \uparrow & \\ & \dots & z_{(i,j-1)} & \rightarrow & z_{(i,j)} & \rightarrow & z_{(i,j+1)} & \dots \end{array}$$

Thus we assume the following.

Assumption 2. The probability of being in $r \in K_+ = K \cup \{\dagger\}$ at time j depends upon the capture and migration history until $j-1$ only through the location at time $j-1$. In particular, the migration behaviour may be modelled by a nonhomogeneous first-order Markov chain, where capture at time $j-1$ has, conditionally on $z_{(i,j-1)}$, no impact on the destination $z_{(i,j)}$.

Assumption 3. The probability of being captured at time j depends upon the capture and migration history up to j only through the location at time j . In particular, there is no trap response (Seber, 1982). This means that, given the migrations until j , the probability of this capture at time j does not depend on the captures which occurred before j .

Note that the nonresponse mechanism has to be considered non-ignorable according to the classification of Little & Rubin (1987), since the distribution of x_i , conditionally on z_i , depends on the missing locations and on the observed locations: see (3.4).

We further assume the following.

Assumption 4. Movements among areas between sampling times are unconstrained and migrations beyond the study areas are permanent.

Consistent with the previous assumptions, we introduce the following notation. For $2 \leq j \leq m$ and $r \in K_+$, denote by $p_j(r)$ the probability that a marked animal will be recaptured at time j , given that it is in r at this time. We have $p_j(\dagger) = 0$. For $1 \leq j \leq m-1$, $r \in K_+$, $s \in K_+$ denote by $q_j(r, s)$ the transition probability that an animal will be in s at time $j+1$, given that it is in r at time j . Thus, for each i the sequence $(z_{(i,j)})$ is a nonhomogeneous Markov chain with state space K_+ and known initial state $z_{(i,t_i)} = r$ for an individual tagged at time t_i in area r ; \dagger being the only absorbing state. Assumption (4) implies

$$q_j(r, s) > 0 \quad (r, s \in K), \quad q_j(\dagger, \dagger) = 1, \quad q_j(\dagger, r) = 0 \quad (r \in K).$$

In order to obtain expressions in terms of quantities of biological interest, see Brownie et al. (1993) and Hestbeck et al. (1991) for motivation, for r and s in K , we decompose the transition probability $q_j(r, s)$ as the product of a survival probability $\phi_j(r)$ and an

interstratum movement probability $\psi_j(r, s)$; namely

$$q_j(r, s) = \phi_j(r)\psi_j(r, s),$$

where

$$\phi_j(r) = \sum_{s \in K} q_j(r, s) = 1 - q_j(r, \dagger). \quad (2.2)$$

We denote by $\theta = (p, \phi, \psi)$ the parameters of the model, where

$$p = (p_j(r)), \quad \phi = (\phi_j(r)), \quad \psi = (\psi_j(r, s)).$$

3. ESTIMATION

3.1. Prior distributions

Let $\psi_j(r) = (\psi_j(r, 1), \dots, \psi_j(r, k))$. We assume that

$$p_j(r) \sim \text{Be}(a, b), \quad \phi_j(r) \sim \text{Be}(\alpha, \beta), \quad \psi_j(r) \sim \mathcal{D}_k(e_1, \dots, e_k),$$

all independently, where the hyperparameters (a, b) , (α, β) , and (e_1, \dots, e_k) are known. Thus the prior distribution $\pi(\theta)$ is proportional to

$$\prod_{\substack{2 \leq j \leq m \\ r \in K}} p_j(r)^{a-1} \{1 - p_j(r)\}^{b-1} \prod_{\substack{1 \leq j \leq m-1 \\ r \in K}} \phi_j(r)^{\alpha-1} \{1 - \phi_j(r)\}^{\beta-1} \prod_{j=1}^{m-1} \sum_{r, s \in K} \psi_j(r, s)^{e_s-1}.$$

An extension of this prior model would be to assume that these hyperparameters depend upon locations or sampling times (Castledine, 1981). This extension is considered in § 4, when dealing with the example.

3.2. The Gibbs sampling algorithm

A Gibbs sampler, using the scheme (3.1) and (3.2) below, produces two chains, $(\theta^{(l)})$, and $(z^{(l)})$, related to the simulated missing data. The chain $(\theta^{(l)})$ can be considered, for large l , as an approximate sample from the posterior distribution, namely $\pi(\theta|y)$ of θ , and can then be used to derive, via the ergodic theorem, any posterior quantity of interest: see § 3.3 for justification.

We use $f(\cdot)$ to denote a probability mass function. Starting with an initial value $z^{(0)}$, for $l \geq 1$ we iterate steps of the form as follows.

Step 1. Generate

$$\theta^{(l)} \sim \pi(\theta | z^{(l-1)}, y). \quad (3.1)$$

Step 2. Generate

$$z^{(l)} \sim f(z | z^{(l-1)}, \theta^{(l)}, y). \quad (3.2)$$

That is, for each i and for each $j \in M_i$ generate

$$z_{(i,j)}^{(l)} \sim f\{z_{(i,j)} | (z_{(i,j')}^{(l)})_{j' < j}, (z_{(i,j')}^{(l-1)})_{j' > j}, y_i, \theta^{(l)}\}. \quad (3.3)$$

If $x_{(i,j)} = 1$, no simulation is undertaken. Note that the notation $z^{(l)}$ represents only the set of simulated $z_{(i,j)}$'s at step l .

Before describing the parameter simulation step (3.1), we need the likelihood of θ for the completed data denoted by (x, z) , where $z_{(i,j)}$ represents either the observed location

if $x_{(i,j)} = 1$ or the simulated location, according to (3.3), if $x_{(i,j)} = 0$. We have

$$L(\theta|x, z) = \prod_{i=1}^n f(x_i|z_i, \theta) f(z_i|z_{(i,t_i)}, \theta).$$

It follows from Assumptions 2 and 3 that

$$\prod_{i=1}^n f(x_i|z_i, \theta) = \prod_{i=1}^n \prod_{j=t_i+1}^m f(x_{(i,j)}|z_{(i,j)}) = \prod_{\substack{2 \leq j \leq m \\ 1 \leq r \leq k}} p_j(r)^{u_j(r)} \{1 - p_j(r)\}^{v_j(r)}, \quad (3.4)$$

where

$$u_j(r) = \sum_{i \in \mathcal{J}_{j-1}} 1(x_{(i,j)} = 1, z_{(i,j)} = r), \quad v_j(r) = \sum_{i \in \mathcal{J}_{j-1}} 1(x_{(i,j)} = 0, z_{(i,j)} = r),$$

and from Assumption 2 and (2.2) that

$$\begin{aligned} \prod_{i=1}^n f(z_i|z_{(i,t_i)}, \theta) &= \prod_{i=1}^n \prod_{j=t_i+1}^{m-1} f(z_{(i,j+1)}|z_{(i,j)}) \\ &= \prod_{j=1}^{m-1} \prod_{r \in K} \left[\phi_j(r)^{w_j(r, \cdot)} \{1 - \phi_j(r)\}^{w_j(r, \dagger)} \prod_{s \in K} \psi_j(r, s)^{w_j(r, s)} \right], \end{aligned}$$

where

$$\begin{aligned} w_j(r, s) &= \sum_{i \in \mathcal{J}_j} 1(z_{(i,j)} = r, z_{(i,j+1)} = s), \\ w_j(r, \dagger) &= \sum_{i \in \mathcal{J}_j} 1(z_{(i,j)} = r, z_{(i,j+1)} = \dagger), \quad w_j(r, \cdot) = \sum_{s \in K} w_j(r, s). \end{aligned}$$

Note that the statistic $(\{u_j(r)\}, \{v_j(r)\}, \{w_j(r, s)\})$ is sufficient for $\theta = (p, \phi, \psi)$, when based upon the completed data.

For $r \in K$, simulation of θ at the step l proceeds as follows:

$$\begin{aligned} p_j^{(l)}(r)|y, z^{(l-1)} &\sim \text{Be}\{a + u_j^{(l-1)}(r), b + v_j^{(l-1)}(r)\}, \\ \phi_j^{(l)}(r)|y, z^{(l-1)} &\sim \text{Be}\{\alpha + w_j^{(l-1)}(r, \cdot), \beta + w_j^{(l-1)}(r, \dagger)\}, \\ \psi_j^{(l)}(r)|y, z^{(l-1)} &\sim \mathcal{D}_k\{e_1 + w_j^{(l-1)}(r, 1), \dots, e_k + w_j^{(l-1)}(r, k)\}. \end{aligned}$$

We do not simulate the $z_{(i,t_i)}$'s since they are known. Starting from (3.3), it follows from Assumptions 2 and 3 that $z_{(i,j)}^{(l)}$ is simulated according to

$$f(z_{(i,j)}|z_{(i,j-1)}, z_{(i,j+1)}, x_{(i,j)}, \theta^{(l)}) \quad (t_i + 1 \leq j \leq m-1), \quad (3.5)$$

$$f(z_{(i,m)}|z_{(i,m-1)}, x_{(i,m)}, \theta^{(l)}) \quad (j = m), \quad (3.6)$$

where $z_{(i,j-1)}$ represents either the observed location if $x_{(i,j-1)} = 1$ or the simulated state $z_{(i,j-1)}^{(l)}$ if $x_{(i,j-1)} = 0$, and where a similar convention is adopted for $z_{(i,j+1)}$, except that it is step $(l-1)$ which is involved.

First, if $z_{(i,j-1)}^{(l)} = \dagger$ then $z_{(i,j)}^{(l)} = \dagger$. Now consider the case $j-1 \in C_i$ or $j-1 \in M_i$ but $z_{(i,j-1)}^{(l)} \neq \dagger$. For $t_i + 1 \leq j \leq m-1$ such that $x_{(i,j)} = 0$ we have, for $s \in K_\dagger$,

$$\text{pr}(z_{(i,j)}^{(l)} = s|z_{(i,j-1)} = r, z_{(i,j+1)} = t, \theta^{(l)}) = \frac{q_{j-1}^{(l)}(r, s) \{1 - p_j^{(l)}(s)\} q_j^{(l)}(s, t)}{\sum_{u \in K_\dagger} q_{j-1}^{(l)}(r, u) \{1 - p_j^{(l)}(u)\} q_j^{(l)}(u, t)},$$

and for $j = m$ such that $x_{(i,m)} = 0$ we have

$$\text{pr}(z_{(i,m-1)}^{(l)} = s|z_{(i,m-1)} = r, \theta^{(l)}) = \frac{q_{m-1}^{(l)}(r, s) \{1 - p_m^{(l)}(s)\}}{\sum_{u \in K_\dagger} q_{m-1}^{(l)}(r, u) \{1 - p_m^{(l)}(u)\}},$$

where

$$q_j^{(l)}(r, s) = \psi_j^{(l)}(r, s) \phi_j(r)^{(l)} \quad (r, s \in K), \quad q_j^{(l)}(r, \dagger) = 1 - \phi_j^{(l)}(r) \quad (r \in K), \\ q_j^{(l)}(\dagger, r) = 0 \quad (r \in K), \quad q_j^{(l)}(\dagger, \dagger) = 1, \quad p_j^{(l)}(\dagger) = 0.$$

When dealing with the Hestbeck–Brownie model, whose assumptions are those of the Arnason–Schwarz model except that $(z_{(i,j)})$ is a second-order Markov chain, a missing location at time j would be simulated according to

$$f(z_j | z_{(j-2)}, z_{(j-1)}, x_j, z_{(j+1)}, z_{(j+2)}, \theta),$$

which, when $t_i + 2 \leq j \leq m - 2$, is proportional to

$$f(z_{(j+2)} | z_{(j+1)}, z_j) f(z_{(j+1)} | z_j, z_{(j-1)}) \{1 - p_j(z_j)\} f(z_j | z_{(j-1)}, z_{(j-2)}),$$

where the index i has been omitted for convenience.

3.3. Convergence issues

In order to establish convergence of $(\theta^{(l)})$ to $\pi(\theta | y)$ and ergodicity of $(\theta^{(l)})$ we study convergence and ergodicity of the dual Markov chain $(z^{(l)})$. Convergence, ergodicity and mixing properties of $(\theta^{(l)})$ are then deduced from those of $(z^{(l)})$ by applying a duality principle (Robert, Celeux & Diebolt, 1993; Diebolt & Robert, 1994). We thus take advantage of the fact that the dual chain $(z^{(l)})$ is simpler to study than the sequence $(\theta^{(l)})$, which is not a Markov chain.

THEOREM 3. *The sequence $(z^{(l)})$ is a first-order homogeneous ergodic Markov chain, with unique stationary distribution $f(z | y)$.*

Proof. We denote by \mathcal{H} the finite state space of $(z^{(l)})$. The sequence $(z^{(l)})$ is a first-order and homogeneous Markov chain with stationary distribution $f(z | y)$ (Robert et al., 1993), with respect to the transition kernel density

$$Q(z, z') = \text{pr}(z^{(l+1)} = z' | z^{(l)} = z, y) = \int_{\Theta} \pi(\theta | z^{(l)} = z, y) \text{pr}(z^{(l+1)} = z' | z^{(l)} = z, \theta, y) d\theta.$$

Since $(z^{(l)})$ is a finite Markov chain, we deduce ergodicity from irreducibility and aperiodicity.

Note that, since the conditional densities $\pi(\theta_s | y, z^{(l)})$ for the components (θ_s) of θ are positive on their support, we have $Q(z, z') > 0$ only if the condition

$$\text{pr}(z^{(l+1)} = z' | z^{(l)} = z, \theta, y) > 0 \tag{3.7}$$

is satisfied. Two states z and z' of \mathcal{H} do not satisfy the condition (3.7) if they include elements such as $z_{(i,j)}^{(l+1)} = \dagger$ and $z_{(i,j+1)}^{(l)} = s$, where $s \in K$. Still working with example (2.1), $Q(z', z)$ is null if, for instance,

$$z = r \, t \, \underline{r} \, s \, t \, \underline{s} \, \underline{r} \, \underline{s}, \quad z' = r \, t \, \underline{t} \, s \, t \, \underline{t} \, \underline{\dagger} \, \underline{\dagger}.$$

We underline the elements which comprise the states z and z' and, to make it clear, we recall the observed locations.

In the Appendix we show that $(z_i^{(l)})$ is irreducible and aperiodic, for a given $1 \leq i \leq n$. Irreducibility and aperiodicity of the chain $(z^{(l)})$ follow from

$$f(z^{(l+1)} | z^{(l)}, \theta, y) = \prod_{i=1}^n f(z_i^{(l+1)} | z_i^{(l)}, \theta, y_i),$$

which represents the fact that $(z^{(l)})$ is a set of n independent Markov chains $(z_i^{(l)})$. \square

Furthermore, since the chain $(z^{(l)})$ is associated with a finite state space, it is uniformly geometrically ergodic and geometrically ϕ -mixing. When the two phases of an iteration of the Gibbs sampling scheme are those given by (3.3) and (3.4), Robert et al. (1993) have established a duality principle which allows us to derive the same ergodic and mixing properties for the chain $(\theta^{(l)})$. This principle is based upon the inequality

$$\|\pi(\theta^{(l+1)}|z^{(0)}, y) - \pi(\theta|y)\|_1 \leq \|f(z^{(l)}|z^{(0)}, y) - f(z|y)\|_1,$$

where $\|\cdot\|_1$ represents the L_1 -norm associated with Lebesgue or counting measure, depending on the context.

Let h be a real function such that $\int_{\Theta} |h(\theta)| \pi(\theta|y) d\theta < \infty$. From the ergodic theorem, when L is large enough, we can approximate the posterior quantity of interest $E\{h(\theta)|y\}$ by

$$\frac{1}{L} \sum_{l=1}^L h(\theta^{(l)}). \quad (3.8)$$

In particular, $L^{-1} \sum \theta_s^{(l)}$ is a consistent estimator of $E(\theta_s|y)$, where θ_s represents $q_j(r, s)$, $\phi_j(r)$ or $p_j(r)$. Moreover, we can use the Gibbs sampler to approximate the posterior expected loss, namely the trace of the matrix $\text{var}(\theta|y)$, which represents a measure of the posterior error of estimating θ by $E(\theta|y)$ (Berger & Robert, 1990).

Also, the geometric ϕ -mixing property of $(\theta^{(l)})$ implies that a Central Limit Theorem holds (Billingsley, 1968). Let h be a real function such that $E\{h(\theta)^2|y\} < \infty$. When $(\theta^{(l)})$ is in its stationary state we have, as $L \rightarrow +\infty$,

$$L^{-\frac{1}{2}} \sum_{l=1}^L [h(\theta^{(l)}) - E\{h(\theta)|y\}] \rightarrow N(0, \sigma_h^2),$$

where

$$\sigma_h^2 = \text{var}\{h(\theta)|y\} + 2 \sum_{l=1}^{\infty} \text{cov}\{h(\theta^{(0)}), h(\theta^{(l)})|y\}.$$

Estimating the variance σ_h^2 would allow us to assess the rate of convergence in (3.8): see Geyer (1992) for a possible approach. Providing a formal stopping rule for the Gibbs sampler is a much debated issue which we do not investigate in this paper: see Geyer (1992), Gelman & Rubin (1992), Raftery & Lewis (1992) for discussion.

For the Hestbeck–Brownie model, convergence of the Gibbs sampling algorithm can be proved similarly to Theorem 3.

4. AN ILLUSTRATION

4.1. Data description

We briefly illustrate our approach on data from a study focusing on the dispersal of the common lizard, *Lacerta vivipara*. Characteristics of this species, the experimental protocol, and the field description are detailed by Massot et al. (1992); only the main points are given here. Capture-recapture data have been collected on three contiguous sites A, B, C of Mont Lozère in Cévennes, France, from June 1989 to August 1991, from two samples per year, in June and in August. The marked population includes 96 juveniles. About 50% of the locations are missing, mostly at times 4, 5 and 6.

4.2. *Modelling the prior information*

For each capture and survival parameter, the prior information is composed of the prior mean and a 95% prior credible interval. For each vector $\psi_j(r)$, prior information is composed of the prior means of the $\psi_j(r, s)$'s and a 95% prior credible interval for each $\psi_j(r, r)$. Massot et al. (1992), previous studies, and the spatial structure of the site have been taken into account in determining these quantities. We assume that the hyperparameters of the capture prior distributions do not depend on the location, but only on time j through the catch effort. For instance, the catch effort was less intensive for $j = 5$ and 6. We have assumed that the hyperparameters of the movement and survival prior distributions do not depend on the year of capture. However, we have taken into account the biological rhythms of this species, which explain the annual periodicity of these prior parameters: hibernation from September to April, search for partners and mating in early

Table 1. *Capture parameters: prior information and parameters of the prior distributions*

	2	3	Time 4	5	6
Mean	0.3	0.4	0.5	0.2	0.2
95% credible interval	[0.1, 0.5]	[0.2, 0.6]	[0.3, 0.7]	[0.05, 0.4]	[0.05, 0.4]
Beta	(6.0, 14.0)	(8.0, 12.0)	(12.0, 12.0)	(3.5, 14.0)	(3.5, 14.0)

Table 2. *Survival parameters: prior means, prior 95% credible intervals and prior distributions*

	Site A		Site B		Site C	
	$j = 1, 3, 5$	$j = 2, 4$	$j = 1, 3, 5$	$j = 2, 4$	$j = 1, 3, 5$	$j = 2, 4$
Mean	0.7	0.65	0.7	0.7	0.7	0.65
95% credible int.	[0.4, 0.95]	[0.35, 0.9]	[0.4, 0.95]	[0.4, 0.95]	[0.4, 0.95]	[0.35, 0.9]
Distribution	Be(6.0, 2.5)	Be(6.5, 3.5)	Be(6.0, 2.5)	Be(6.0, 2.5)	Be(6.0, 2.5)	Be(6.5, 3.5)

Table 3. *Movement parameters between sites A, B and C*

(a) *Prior means and prior 95% credible intervals*

Site	$j = 1, 3, 5$			Site	$j = 2, 4$		
	A	B	C		A	B	C
A	0.95 [0.8, 1.0]	0.04	0.01	A	0.9 [0.7, 1.0]	0.09	0.01
B	0.02	0.96 [0.8, 1.0]	0.02	B	0.05	0.9 [0.7, 1.0]	0.05
C	0.01	0.04	0.95 [0.8, 1.0]	C	0.01	0.09	0.9 [0.7, 1.0]

(b) *Prior distribution of $\psi_j(r)$*

Site	$j = 1, 3, 5$	$j = 2, 4$
A	$\mathcal{D}(9.5, 0.4, 0.1)$	$\mathcal{D}(9.0, 0.9, 0.1)$
B	$\mathcal{D}(0.25, 9.5, 0.25)$	$\mathcal{D}(0.5, 9.0, 0.5)$
C	$\mathcal{D}(0.1, 0.4, 9.5)$	$\mathcal{D}(0.1, 0.9, 9.0)$

Table 4. Posterior means and posterior 95% credible intervals

(a) Capture parameters					
Site	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
A	0.59 [0.45, 0.75]	0.48 [0.32, 0.67]	0.59 [0.41, 0.76]	0.19 [0.07, 0.38]	0.25 [0.10, 0.45]
B	0.60 [0.45, 0.74]	0.51 [0.33, 0.69]	0.57 [0.41, 0.74]	0.16 [0.05, 0.35]	0.21 [0.08, 0.42]
C	0.62 [0.48, 0.77]	0.49 [0.32, 0.68]	0.58 [0.41, 0.76]	0.14 [0.03, 0.32]	0.23 [0.09, 0.43]
(b) Survival parameters					
Site	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
A	0.86 [0.70, 0.98]	0.70 [0.50, 0.89]	0.84 [0.63, 0.98]	0.67 [0.40, 0.91]	0.75 [0.46, 0.97]
B	0.90 [0.77, 1.00]	0.66 [0.44, 0.88]	0.79 [0.55, 0.97]	0.64 [0.33, 0.93]	0.72 [0.42, 0.96]
C	0.89 [0.75, 0.99]	0.61 [0.41, 0.83]	0.79 [0.54, 0.97]	0.64 [0.35, 0.91]	0.73 [0.43, 0.96]
(c) Movement parameters					
Site	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
A → A	0.95 [0.86, 1.00]	0.95 [0.82, 1.00]	0.91 [0.76, 1.00]	0.92 [0.72, 1.00]	0.96 [0.80, 1.00]
A → B	0.04 [0.00, 0.15]	0.05 [0.00, 0.19]	0.09 [0.00, 0.24]	0.07 [0.00, 0.28]	0.03 [0.00, 0.18]
A → C	0.01 [0.00, 0.04]	0.00 [0.00, 0.06]	0.00 [0.00, 0.05]	0.01 [0.00, 0.09]	0.01 [0.00, 0.10]
B → A	0.07 [0.01, 0.18]	0.06 [0.00, 0.21]	0.01 [0.00, 0.10]	0.04 [0.00, 0.22]	0.02 [0.00, 0.15]
B → B	0.90 [0.78, 0.99]	0.87 [0.69, 0.99]	0.93 [0.79, 1.00]	0.91 [0.69, 1.00]	0.96 [0.80, 1.00]
B → C	0.04 [0.00, 0.13]	0.07 [0.00, 0.22]	0.06 [0.00, 0.20]	0.04 [0.00, 0.22]	0.02 [0.00, 0.15]
C → A	0.00 [0.00, 0.04]	0.01 [0.00, 0.07]	0.01 [0.00, 0.06]	0.01 [0.00, 0.09]	0.01 [0.00, 0.10]
C → B	0.10 [0.02, 0.23]	0.15 [0.03, 0.33]	0.07 [0.00, 0.23]	0.07 [0.00, 0.28]	0.03 [0.00, 0.20]
C → C	0.90 [0.78, 0.99]	0.85 [0.67, 0.98]	0.92 [0.77, 1.00]	0.92 [0.72, 1.00]	0.96 [0.79, 1.00]

May; whereas little is known about these parameters between mating and hibernation (Massot et al., 1992). We adopted beta distributions for the capture and survival parameters and Dirichlet distributions for the vector movement parameters $\psi_j(r)$. The prior information determines the hyperparameters. Tables 1, 2 and 3 list the prior information and the corresponding hyperparameters.

4.3. Results

We used a single long run of 10^6 iterations of the Gibbs sampler to derive the Bayes estimators. Table 4 lists posterior means and posterior 95% credible intervals, obtained from the 2.5% and 97.5% quantiles, for each capture, survival and movement parameter.

5. CONCLUSION

Recognising the missing data structure of the model allows us to apply Gibbs sampling and thus to propose Bayes estimators previously unavailable; thus, it is possible to incorporate prior information into the model. We can also take advantage of the missing data structure to handle inference outside the Bayesian paradigm, by implementing EM-type algorithms (Wei & Tanner, 1990).

In this paper we have assumed that the hyperparameters of the prior distributions are known, although hierarchical extensions could be considered when dealing with poor prior information, as well as to robustify the inference. In any case, the sensitivity of the estimators to the choice of the prior distribution will have to be investigated, either theoretically or through simulations.

Some extensions of the model might be considered of interest. Taking into account a trapping effect will lead to an immediate extension of the two models considered in this paper. Another extension assumes that capture at time j , conditionally on the location at time j , affects the location at $j + 1$. Moreover, it might be useful to take into account some individual components as well as characteristics of the habitat; this could be achieved through a logistic model: see e.g. Carlin & Polson (1992) for a possible Bayesian approach. Moreover, it might be necessary to relax the assumption that animals behave independently with respect to movement. Indeed, it follows from observations that, for some species, an individual migration in a given year depends on the breeding successes in the year before of the nearest neighbours; see e.g. Danchin, Cadiou & Boulinier (1993). A directed graph representation will be an appropriate tool to describe the complex probabilistic structure of these extensions and to calculate densities involved in the corresponding Gibbs sampling algorithms. Moreover, these extensions could offer a useful framework in which to investigate new missing data structures in longitudinal studies.

ACKNOWLEDGEMENT

I am grateful to the researchers of the Laboratoire d'Ecologie de l'Ecole Normale Supérieure de Paris involved with the biological aspect of this work, in particular to J. Clobert for suggesting this data set, and to M. Massot for providing prior information. I thank also C. Robert for helpful comments, S. Legendre for valuable pieces of advice in the computing part of this work, and CREST-INSEE for financial support. The suggestions of a referee and of the editor led to this much improved version of the paper.

APPENDIX

Part of proof of Theorem 3

Consider the index j' , when it exists, such that for all $j > j'$ the locations are missing. We denote by p the number of such missing locations. If j' does not exist, then $z_{(i,m)}$ has been observed and $p = 0$. If $p \leq 1$ for all z and z' in \mathcal{H} , (3.7) is satisfied.

When $p > 1$, we partition \mathcal{H} into $p + 1$ classes $\mathcal{C}_0, \dots, \mathcal{C}_a, \dots, \mathcal{C}_p$, where \mathcal{C}_0 is the state space

such that $z_{(i,j)} \in K$ for all $j \geq m - p + 1$; \mathcal{C}_p is the state space such that $z_{(i,j)} = \dagger$ for all $j \geq m - p + 1$; and \mathcal{C}_a ($1 \leq a \leq p - 1$) is the state space such that:

$$z_{(i,j)} = \dagger \quad (m - a + 1 \leq j \leq m), \quad z_{(i,j)} \in K \quad (m - p + 1 \leq j \leq m - a).$$

We assume nothing about the $z_{(i,j)}$'s when $j < j'$; however, note that such a $z_{(i,j)}$ necessarily belongs to K .

For all $0 \leq a \leq p$ and for all $z \in \mathcal{C}_a$, and $z' \in \mathcal{C}_a$, we have $Q(z, z') > 0$ since condition (3.7) is satisfied; therefore the chain $z_i^{(l)}$ is aperiodic and the (\mathcal{C}_a) are communication classes. For all $0 \leq b < a \leq p$ and for all $z \in \mathcal{C}_a$ and $z' \in \mathcal{C}_b$, condition (3.7) is also satisfied.

Finally, when $p \geq b > a \geq 0$, for all $z \in \mathcal{C}_a$ and $z' \in \mathcal{C}_b$, condition (3.7) is satisfied only if $b = a + 1$. We have, therefore,

$$\mathcal{C}_0 \rightarrow \mathcal{C}_1, \dots, \mathcal{C}_a \rightarrow \mathcal{C}_{a+1}, \dots, \mathcal{C}_{m-1} \rightarrow \mathcal{C}_m,$$

where the notation $\mathcal{C}_u \rightarrow \mathcal{C}_v$ implies that, for all $z \in \mathcal{C}_u$ and $z' \in \mathcal{C}_v$, $Q(z, z') > 0$. In the example in § 3.3, a step from z to z' could pass through the state:

$$r \ t \ t \ s \ t \ s \ t \ \dagger.$$

Therefore, given i , the chain $(z_i^{(l)})$ is aperiodic and irreducible.

REFERENCES

- BERGER, J. O. & ROBERT, C. (1990). Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface. *Ann. Statist.* **18**, 617–51.
- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. New York: John Wiley.
- BROWNIE, C., HINES, J. E., NICHOLS, J. D., POLLOCK, K. H. & HESTBECK, J. B. (1993). Capture-recapture studies for multiple strata including non-Markovian transition probabilities. *Biometrics* **49**, 1173–87.
- CARLIN, B. C. & POLSON, N. G. (1992). Monte-Carlo Bayesian methods for discrete regression models and categorical time series. In *Bayesian Statistics 4*, Ed. J. O. Berger, J. M. Bernardo, A. P. Dawid and A. F. M. Smith, pp. 577–86. Oxford: Oxford University Press.
- CASTLEDINE, B. (1981). A Bayesian analysis of multiple capture-recapture sampling for a closed population. *Biometrika* **68**, 197–210.
- CLOBERT, J. & LEBRETON, J. D. (1991). Estimation of demographic parameters. In *Bird Population Studies: their Relevance to Conservation and Management*, Ed. C. Perrins, J. D. Lebreton and G. Hiron, pp. 75–104. Oxford: Oxford University Press.
- DANCHIN, E., CADIOU, B. & BOULINIER, T. (1993). La régulation des populations de mouettes tridactyles. *La Recherche* **256**, 893–5.
- DIEBOLT, J. & ROBERT, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *J. R. Statist. Soc. B* **56**, 363–75.
- GELMAN, A. & RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7**, 457–511.
- GEORGE, E. I. & ROBERT, C. P. (1992). Capture-recapture estimation via Gibbs sampling. *Biometrika* **79**, 677–83.
- GEYER, C. J. (1992). Practical Markov chain Monte Carlo (with Discussion). *Statist. Sci.* **7**, 473–511.
- HESTBECK, J. B., NICHOLS, J. D. & MALECKI, R. A. (1991). Estimation of movement and site fidelity using mark-resight data of wintering Canada Geese. *Ecology* **72**, 523–33.
- LAURITZEN, S. L. & SPIEGELHALTER, D. J. (1988). Local computations with probabilities on graphical structures and their applications to expert systems (with Discussion). *J. R. Statist. Soc. B* **50**, 157–224.
- LITTLE, R. J. A. & RUBIN, D. B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley.
- MASSOT, M., CLOBERT, J., PILORGE, T., LECOMTE, J. & BARBAULT, R. (1992). Density dependence in the common lizard: demographic consequences of a density manipulation. *Ecology* **73**, 1742–56.
- POLLOCK, K. H. (1991). Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future. *J. Am. Statist. Assoc.* **86**, 225–38.
- RAFTERY, A. E. & LEWIS, S. (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics 4*, Ed. J. O. Berger, J. M. Bernardo, A. P. Dawid and A. F. M. Smith, pp. 763–73. Oxford: Oxford University Press.
- ROBERT, C. P., CELEUX, G. & DIEBOLT, J. (1993). Bayesian estimation of hidden Markov chains: a stochastic implementation. *Statist. Prob. Lett.* **11**, 77–88.
- SCHWARZ, C. G., SCHWEIGERT, J. F. & ARNASON, A. N. (1993). Estimating migration rates using tag-recovery data. *Biometrics* **49**, 177–193.

- SEBER, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters*, 2nd ed. New York: Macmillan.
- SMITH, A. F. M. & ROBERTS, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte-Carlo methods. *J. R. Statist. Soc. B* **55**, 3–23.
- WEI, G. C. G. & TANNER, M. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Am. Statist. Assoc.* **85**, 699–704.

[Received August 1993. Revised July 1995]