HAL
open science

Testing biological hypotheses in site occupancy models: a Bayesian approach

Jérôme Dupuis, Jean Joachim

To cite this version:

Jérôme Dupuis, Jean Joachim. Testing biological hypotheses in site occupancy models: a Bayesian approach. 2015. hal-01241022

HAL Id: hal-01241022

https://hal.science/hal-01241022

Preprint submitted on 9 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Testing biological hypotheses in site occupancy models: a Bayesian approach

Jérôme Dupuis*

ESP/IMT Université Paul Sabatier, 118 Route de Narbonne, Toulouse, France

and

Jean Joachim

INRA-CEFS, BP 27, 31326, Castanet-Tolosan, Toulouse, France

Summary. The occupancy rate of a target species in a region divided in quadrats (or sites) is defined as the proportion of quadrats occupied by this species. This is a key quantity in site occupancy models which typically remains unknown after the data are collected, because the probability of detecting a target species in a given quadrat is < 1. Implementing tests on occupancy rates leads to a quite unusual situation, because an occupancy rate is not a statistical parameter, but a function of a discrete process partially observed. To deal with that difficulty, we adopt a Bayesian view within which the treatment of such tests turns out to be natural. We develop our approach for discrete-time site occupancy data, and we illustrate it by testing if the occupancy rate of a bird species increases over time (colonization test). A Bayesian model averaging is implemented to deal with the fact that several plausible models are viewed for the data at hand. We state a closed-form expression for the posterior probability of each model. The posterior probability of the null hypothesis (under a given model) is obtained by implementing a data augmentation algorithm. Finally, from a variety of examples, we show that the Bayesian methodology allows us to address a wide range of questions about occupancy.

* *email:* dupuis@math.univ-toulouse.fr

Keywords: Bayesian test; Colonization test; Data aumentation algorithm; Missing data model; Model averaging; Occupancy rates; Stability test.

1. Introduction

Occupancy models are now widely used in Ecology: they are a part of the *Applied Statistics* landscape, just like the well-known capture-recapture models. Since the seminal paper of MacKenzie *et al.* (2002) statisticians have developed more and more complex models. The book of MacKenzie *et al.* (2006) - which remains the book of reference when modeling occupancy - provides a detailed description of the most important occupancy models. More recent developments can be found in Royle and Kéry (2007), MacKenzie *et al.* (2009), Dupuis *et al.* (2010), Dorazio and Rodriguez (2012). For an updated review see Bailey *et al.* (2013).

A key quantity in site occupancy models is the occupancy rate (MacKenzie *et al.* 2002). It is defined as follows. Assume that a region of interest (called R) is composed of J sites (or quadrats), and let z_{jt} denote the indicator of presence of a species of interest (called target species); thus, $z_{jt} = 1$ if the target species is present in quadrat j at time t, and 0 otherwise. The occupancy rate of the target species at time t in region R is defined as the proportion of quadrats occupied by this species at this time; it is denoted by ω_t: thus one has $\omega_t = \frac{1}{J}\sum_{j=1}^{J} z_{jt}$.

The probability of detecting the presence of a species in a quadrat is typically < 1. Most often, the occupancy rate remains unknown once the data have been collected, considering that some z_{jt}'s may not have been observed. Consequently, numerous biological questions involving occupancy rates have to be formulated through statistical tests. Now, to our knowledge, no statistical procedure has been developed to deal with such tests. The objective of this paper is thus to address this issue while working within existing site occupancy models.

From a biological point of view, the statistical methodology developed in this paper allows to investigate a wide range of questions related to occupancy. This important point is illustrated from a variety of examples in Section 6. At this stage of the paper, we give only one example. When we are interested in the way occupancy rates may change over time, we may want to test whether the target species tends to colonize the region of interest, which corresponds to an increase in its occupancy rate over time. The null hypothesis below

$$\mathrm{H}_0 : \quad \omega_1 < \omega_2 < \ldots < \omega_{T-1} < \omega_T \tag{1.1}$$

is associated with the hypothesis of a colonization over the period from $t = 1$ to $t = T$. The test (1.1) will be called the *colonization test*. Of course, testing a decrease in occupancy rates could also be of interest (see Section 5). For analyzing discrete-time site occupancy data, MacKenzie *et al.* (2006), have developed a general model which assumes that the time occupancy process is directed by a non homogeneous Markov chain; see also Royle and Kéry (2007). We implement our methodology within this framework which is the one corresponding to the data sets analyzed in this paper.

Testing the hypothesis (1.1) by using a frequentist approach is clearly inappropriate. Indeed, though occupancy models are parametric models, the theory of parametric tests (eg Lehmann and Romano, 2008) does not apply here, simply because ω_t is not a parameter of the model but a function of the z_{jt}'s. This clearly constitutes a statistical difficulty which we address by adopting a Bayesian view and by seeing occupancy models as a missing data model for the computational aspects.

As already mentioned above, the target species may not have been detected in quadrat j at time t, though it is present in it at this time. In such a situation, z_{jt} remains unknown, and it is said to be missing. This is why occupancy models enter the class of missing data models which includes, for example, capture-recapture models, mixture distributions mod-

els, hidden Markov chains (eg Little and Rubin, 2002). In missing data models, inference concerns most often the parameters, since they are usually the quantity of interest. Obtaining Bayesian estimations of the parameters is usually based on a MCMC algorithm (called data augmentation algorithm) which takes advantage of the missing data structure (eg Robert and Casella, 2005). Estimating a function of the process subject to missingness (the z_{jt}'s in our case) is rarely the objective: estimating occupancy rates constituting, in a way, a notable exception. Interestingly, the data augmentation algorithm can also be used for estimating any function of the z_{jt}'s. Dupuis, Bled and Joachim (2011) have taken advantage of this observation for obtaining the Bayesian estimation of the occupancy rate in the simple case where the model involved only one occupancy rate (by species). The key idea developed in this paper is that it is also possible to use this same observation to deal with tests on occupancy rates. Indeed, in Bayesian Statistics, the answer simply relies on the posterior probability of H_0 (or of H_1 since they sum to 1). Now, calculating $\mathrm{P}(\mathrm{H}_i \,|\, \text{data})$, where $i \in \{0, 1\}$, comes down to estimating a particular function of the z_{jt}'s (see Section 4), which means that the data augmentation algorithm can again be used for obtaining the posterior probability of H_i.

For the data set we analyze in Section 5, different plausible models of biological interest can be considered. In this context, the answer to the test is tackled via a Bayesian model averaging. Such a procedure requires the computation of the posterior probability of each model. In missing data models, computing such probabilities is often intractable, and, typically, one has to use reversible jump Markov chain Monte Carlo (RJMCMC) methods for implementing a Bayesian model averaging. Here, we show how to compute exactly these posterior probabilities, so that, interestingly, the use of RJMCMC methods is not neccessary for the data at hand.

The statistical methodology we develop in this paper is motivated by biological ques-

tions involving occupancy rates; now, handling such quantities requires that J is a part of the model. When J is very large, the site occupancy models typically assume that J is theoretically infinite; as a result, in such models, the very notion of occupancy rate is no longer applicable (see the Sections 4.1 and 4.5 of the book of MacKenzie *et al.* 2006). However, our methodology can also be useful in the asymptotic framework to deal with tests involving the occupancy parameters: this important point is developed in Section 6.

2. Data description and missing data structure

2.1 *Data description*

We assume that the study region is composed of J spatial units. In the literature, such units are called sites or quadrats; we will use this latter term. In this paper we assume that quadrats are of equal size. The data collection protocol is the following. A given year (typically in spring), an experimenter visits each qudrat $K \geq 2$ times each quadrat and records the number of visit(s) during which the target species has been detected. Then this sequence is repeated in the following years. We denote by $y_{jt} \in \{0, 1, \ldots, K\}$ the number of visits during which the target species has been detected in quadrat j, at year t. This period will afterwards be referred to as time t, where $t \in \{1, \ldots, T\}$. The whole data set is denoted by $\mathbf{y}$, so that one has $\mathbf{y} = \{\mathbf{y}_t; t = 1, \ldots, T\}$ where $\mathbf{y}_t = \{y_{jt}; j = 1, \ldots, J\}$ represents the data collected at time t. For simplicity, we assume that all the quadrats are explored; however, we will indicate how to modify the data augmentation algorithm, when inference is based on $n < J$ quadrats. As far as birds species are concerned, detections are typically based on oral recognitions (see Section 6 for details).

2.2 *The missing data structure*

Recall that, for $j = 1, \ldots, J$, we denote by z_{jt} the indicator of presence of the target species in quadrat j at time t; thus, $z_{jt} = 1$ if it is present in quadrat j at this time, and 0 otherwise. It is useful to clarify the links between z_{jt} and y_{jt}. Considering that the

probability of detecting the target species in a given quadrat is < 1, the event $(y_{jt} = 0)$ covers in fact two exclusive situations: either the target species is present in quadrat j at time t but has not been detected, or it is not present (and cannot have been detected). Consequently, z_{jt} is unknown when $y_{jt} = 0$; in such a situation, z_{jt} is said to be missing. When $y_{jt} \geq 1$, the target species has been detected in quadrat j at time t during at least one visit, and one has $z_{jt} = 1$. This way of presenting the missing data structure is essentially the one adopted in Dupuis and Goulard (2011). Compared to the presentation made in Dupuis, Bled and Joachim (2010), notation is simplified.

Once the data $\mathbf{y}_t$ are available, it is important to realize that ω_t remains unknown in all cases, except one. Indeed, if the target species has been detected in the J quadrats then $\omega_t = 1$; in all other cases, ω_t remains unknown once $\mathbf{y}_t$ is available. In fact, one has $\omega_t \geq \frac{D_t}{J}$, where D_t denotes the number of quadrats in which the presence of the target species has been detected at time t. In other respects, it is important to realize that $D_t > D_{t+1}$ does not imply that $\omega_t > \omega_{t+1}$, since it is quite possible that $\omega_t < \omega_{t+1}$. Similarly, $D_t = D_{t+1}$ does not imply that $\omega_t = \omega_{t+1}$ (except if $D_t = D_{t+1} = J$).

Separating the occupancy process $\mathbf{z} = (z_{jt}; j = 1, \ldots, J; t = 1, \ldots, T)$ from the occupancy data $\mathbf{y}$ has multiple benefits. First, this allows us to clarify the missing data structure of site occupancy data. Second, it allows us to clearly separate the assumptions relating to the way the target species occupies the region R over time, from those related to detections (see the next Section). This view refers to the state-space formulation proposed by Royle and Kéry (2007). Third, the occupancy process $\mathbf{z}$ is a key ingredient of the data augmentation algorithm. An analoguous modelling has been proposed by Dupuis (1995) in multi-state capture-recapture models.

3. Modelling, models and priors

3.1 *Modelling occupancy and detectability*

• For fixed j, the directed graph $\mathcal{G}$ below provides the conditionnal independence assumptions made between the $\{z_{jt}; t = 1, \ldots, T\} \bigcup \{y_{jt}; t = 1, \ldots, T\}$.

$$\mathcal{G} \qquad \begin{array}{ccccccc} \cdots & y_{j,t-1} & & y_{jt} & & y_{j,t+1} & \cdots \\ & \uparrow & & \uparrow & & \uparrow & \\ \cdots & z_{j,t-1} & \longrightarrow & z_{jt} & \longrightarrow & z_{j,t+1} & \cdots \end{array}$$

This graph constitutes an attractive tool for providing all the conditional independence assumptions between the different random variables present in the model. For the probabilistic interpretation of a directed graph, see eg Whittaker (1990). Moreover, the graph turns out be an essential tool for calculating in a rigorous way the conditionnal distributions appearing in the missing data simulation phase of the data augmentation algorithm (see Appendix). The above graph $\mathcal{G}$ is similar to the one considered by Dupuis (1995) for analyzing migration capture-recapture data: the migration process of a given marked animal corresponding to the time occupancy process of the target species in a given quadrat. We refer to MacKenzie *et al.* (2006) and to MacKenzie *et al.* (2009) for a description of analogies and differences existing between discrete-time site occupancy models and multi-state capture-recapture models.

The following assumptions are thus made.

Assumption A1. We assume that $z_{j1}, \ldots, z_{jt}, \ldots, z_{jT}$ constitute a first order Markov chain (j being fixed).

Assumption A2. We assume that $y_{j1}, \ldots, y_{jt}, \ldots, y_{jT}$ are independent, conditionally on the z_{jt}'s (j being fixed).

A2 implies that the probabability of detecting the target species at time t in quadrat j is not directly impacted (absence of arrow between y_{jt} and $y_{j,t-1}$) by the fact that it may

have been detected in it at time $t-1$;

• Time t and quadrat j being fixed, we denote by y_{jtk} the indicator which takes the value 1 if the target species has been detected at time t in quadrat j, during the visit k. The following assumption is made.

Assumption A3. We assume that the y_{jtk}'s are independent.

• The occupancy process has two components: a spatial component and a time component. The above graph $\mathcal{G}$ provides assumptions concerning the time occupancy process $\mathbf{z}_j = (z_{jt}; j = 1, \ldots, T)$. We need now to clarify which assumptions are made between the $\mathbf{z}_j$'s.

Assumption A4. We assume that $\mathbf{z}_1, \ldots, \mathbf{z}_j, \ldots, \mathbf{z}_J$ are independent.

A4 implies that the z_{jt}'s are independent (time t being fixed). In other terms, time t being fixed, we assume that the probability that the target species occupies a given quadrat j is not impacted by what may occur in the other quadrats (presence or absence).

• Let $\mathbf{y}_j$ be denote the vector $(y_{jt}; j = 1, \ldots, T)$. The following standard assumption is made.

Assumption A5. We assume that $\mathbf{y}_1, \ldots, \mathbf{y}_j, \ldots, \mathbf{y}_J$ are independent conditionally on $\mathbf{z}$.

A5 implies that the probability of detecting the target species in quadrat j does not depend on its possible detections in the other quadrats (time t being fixed).

Assumptions A1, A2, A3, A4 and A5 are present in MacKenzie *et al.* (2006) though put in a different way.

3.2 *Models and parameters*

Once assumptions A1, A2, A3, A4 and A5 have been retained, different models result depending on whether parameters are time dependent or not. In the most general model, the parameters are time-dependent: it is parametrized as follows. We denote by q_t the probability of detecting the target species, during one visit, in quadrat j, at time t (condi-

tionally on its presence). Assumption A3 implies that $y_{jt}|z_{jt} = 1 \sim \text{Binomial}(K, q_t)$. Note that $\Pr(y_{jtk} = 1|z_{jt} = 0) = 0$. We denote by $\psi_t(r, s)$ the probability that a quadrat is in state $r \in \{0, 1\}$ at time $t + 1$ given that it is in state $r \in \{0, 1\}$ at time t. For example, $\psi_t(0, 1)$ represents the probability that the target species colonizes the quadrat j between time t and time $t + 1$. Formally, one has:

$$q_t = \Pr(y_{jtk} = 1|z_{jt} = 1) \quad \text{and} \quad \psi_t(r, s) = \Pr(z_{j,t+1} = s|z_{jt} = r)\,.$$

Finally, we denote by μ the probability that the target species is present in quadrat j at time $t = 1$. This general model is denoted symbolically $[q_t, \psi_t(r, s)]$, omitting μ for convenience. Three sub-models of biological interest for the data analyzed in Section 6 can be derived from this general model; namely: $[q, \psi(r, s)]$, $[q_t, \psi(r, s)]$, $[q, \psi_t(r, s)]$, with obvious notation.

3.3 *Prior distributions*

For simplicity, a uniform prior distribution is placed on each parameter. More generally, we can adopt beta distributions since all parameters lies in $[0, 1]$ (whatever the model); the modifications to be made to the data augmentation algorithm (described in Section 4.2) are straightforward. Recall that the beta distribution allows us to take into account prior information consisting of a prior mean and 95% credible interval (eg Dupuis, Bled and Joachim, 2011).

4. The statistical procedure and computational issues

Throughout this Section, $p(.)$ denotes a probability mass function, $\mathbb{1}_{(C)}$ represents an indicator function that takes the value 1 when the condition C is true and zero otherwise. Moreover $\pi(.)$ will represent the prior density of any parameter, and $\pi(.|\mathbf{y})$ its posterior density.

4.1 *The testing procedure*

Let us first recall how one proceeds to implement a parametric statistical test in Bayesian statistics. Let $\theta \in \Theta$ denote the global parameter of a statistical model. Imagine that we aim to test

$$\mathrm{H}_0 : \ \theta \in \Theta_0 \quad \text{versus} \quad \mathrm{H}_1 : \ \theta \in \Theta_1 \tag{4.1}$$

where $\{\Theta_0, \Theta_1\}$ is a partition of Θ such that $\Pr(\mathrm{H}_0)$ and $\Pr(\mathrm{H}_1)$ are > 0. In Bayesian statistics the answer to the test (4.1) is based on $\Pr(\mathrm{H}_0 \,|\, \text{data})$ which is interpreted as the degree of evidence in favor of H_0. This practice is justified in decision-theoretic terms. Indeed, in this framework, implementing a test consists in estimating $\mathbb{1}_{(\theta \in \Theta_0)}$ under a given loss function; now, under the quadratic loss (which is the standard loss), the estimate of $\mathbb{1}_{(\theta \in \Theta_0)}$ is $\mathrm{E}(\mathbb{1}_{(\theta \in \Theta_0)} \,|\, \text{data})$ which coincides with $\Pr(\mathrm{H}_0 \,|\, \text{data})$; see eg the Section 5.4 of the book of Robert (2007). When analyzing the data, we will adopt this decision-theoretic framework for which the Bayesian answer to a test thus simply consists in reporting the posterior probability of H_0; of course, one may choose to report $\Pr(\mathrm{H}_1 \,|\, \text{data})$ as well, since $\Pr(\mathrm{H}_0 \,|\, \text{data}) + \Pr(\mathrm{H}_1 \,|\, \text{data}) = 1$. An alternative is to use the Bayes factor (usually denoted by B_{01}) which is defined as the ratio of the posterior odds $\Pr(\mathrm{H}_0 \,|\, \text{data})/\Pr(\mathrm{H}_1 \,|\, \text{data})$ to the prior odds $\Pr(\mathrm{H}_0)/\Pr(\mathrm{H}_1)$. But this quantity (which lies in $]0, \infty[$) is far more difficult to interpret than a probability, even if some useful guidelines exist (eg Kass and Raftey, 1995).

The test (1.1) is on a function of the occupancy process $\mathbf{z}$ (since ω_t is a function of the z_{jt}'s), not on the parameter θ. It is a quite atypical situation which requires a specific statistical treatment. We denote by $\boldsymbol{\omega}$ the occupancy rates vector $(\omega_t; t = 1, \ldots, T)$ and by Ω the space in which $\boldsymbol{\omega}$ takes its values. For testing the null hypothesis

$$\mathrm{H}_0 : \ \boldsymbol{\omega} \in \Omega_0 \quad \text{versus} \quad \mathrm{H}_1 : \ \boldsymbol{\omega} \in \Omega_1 \,, \tag{4.2}$$

where Ω_0 and Ω_1 consitute a partition of Ω, we advocate a Bayesian view, arguing that it leads to a natural statistical answer (see further). Our statistical approach relies on the following observation: before collecting the data, the process $\mathbf{z}$ is unknown, exactly as θ. Hence, the idea of applying to the test (4.2) the statistical procedure adopted for the test (4.1). Taking this view leads thus to basing the answer to the test (4.2) on $\Pr(\boldsymbol{\omega} \in \Omega_0 \,|\, \text{data})$.

The calculation of $\Pr(\boldsymbol{\omega} \in \Omega_0 \,|\, \text{data})$ is based on a set of remarks. Let Z denote the space in which $\mathbf{z}$ takes its values (note that Z is a finite set with 2^{JT} elements). The partition of Ω in Ω_0 and Ω_1 results in a partition of Z in Z_0 and Z_1 such that $\Phi(Z_0) = \Omega_0$ and $\Phi(Z_1) = \Omega_1$, where the function Φ is defined by $\boldsymbol{\omega} = \Phi(\mathbf{z})$. Consequently, computing $\Pr(\boldsymbol{\omega} \in \Omega_0 \,|\, \text{data})$ comes down to estimating a simple function of $\mathbf{z}$, namely the indicator $\mathbb{1}_{(\mathbf{z} \in Z_0)}$, since $\Pr(\boldsymbol{\omega} \in \Omega_0 \,|\, \text{data})$=$\Pr(\mathbf{z} \in Z_0 \,|\, \text{data})$= $\mathrm{E}(\mathbb{1}_{(\mathbf{z} \in Z_0)} |\, \text{data})$. This last remark will be used in the following Section.

4.2 *The data augmentation algorithm*

The data augmentation algorithm is an MCMC algorithm which was initially devised by Tanner and Wong (1987); it has been widely used in the Bayesian analysis of missing data models (eg Robert and Casella, 2004). The data augmentation algorithm is characterized by the fact that each step of the algorithm includes two phases: a parameter simulation phase and a missing data simulation phase.

As far as the second phase is concerned and in our context, $\mathbf{z}$ is partitioned in $\mathbf{z}_{mis} = \{z_{jt} | y_{jt} = 0\}$ and $\mathbf{z}_{obs} = \{z_{jt} | y_{jt} \geq 1\}$. Simulation occurs only when $z_{jt} \in \mathbf{z}_{mis}$. The data augmentation algorithm proceeds as follows: starting with an initial value $\mathbf{z}^{(0)}$ (consistent with $\mathbf{y}$), for $l \geq 1$ we iterate steps of the form:

$$\theta^{(l)} \sim \pi(\theta | \mathbf{z}_{mis}^{(l-1)}, \mathbf{y}) \qquad \mathbf{z}_{mis}^{(l)} \sim p(\mathbf{z}_{mis} | \mathbf{z}_{mis}^{(l-1)}, \theta^{(l)}, \mathbf{y}). \tag{4.3}$$

Dupuis (1995) has precisely used this scheme for conducting the Bayesian analysis of

multiple-state capture-recapture data. This scheme is in fact associated with a component-by-component simulation of the missing z_{jt}'s. An alternative is to simulate the missing data by blocks: see for example Dupuis and Schwarz (2006) in a capture-recapture set-up. When this option is retained, the above scheme simplifies as follows:

$$\theta^{(l)} \sim \pi(\theta|\mathbf{z}_{mis}^{(l-1)}, \mathbf{y}) \qquad \mathbf{z}_{mis}^{(l)} \sim p(\mathbf{z}_{mis}|\theta^{(l)}, \mathbf{y}). \tag{4.4}$$

Afterwards, we will focus on the scheme (4.3) which is the most used frequently in practice.

For the models considered in Section 3.2, it is possible to use the popular software WinBUGS for implementing the scheme (4.3). An alternative is to write one's own program: it is the choice made in this paper. This requires us to compute all the necessary conditional distributions in the MCMC algorithm. Royle and Kéry (2007) also made this choice to deal with the general model $[q_t, \psi_t(r, s)]$, but in a context different from ours. Indeed, Royle and Kéry (2007) worked within an asymptotic framework where J is theoretically infinite. In this framework, the very notion of occupancy rate is obviously not applicable. As a result, the algorithm of Royle and Kéry (2007) needs to be adapted to deal with issues of occupancy rates. All that concerns the way we implement the data augmentation algorithm is in the Appendix. In particular, the situation in which inference is based on $n < J$ sampled quadrats is considered in this Appendix.

The posterior probability of the null hypothesis (under a given model m) is obtained by applying the ergodic theorem. For large enough L, one has:

$$\Pr(\mathrm{H}_0\,|m\,, \mathbf{y}) \approx \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}_{(\boldsymbol{\omega}^{(l)} \in \Omega_0)}$$

where $\boldsymbol{\omega}^{(l)} = (\omega_t^{(l)}; t = 1, \ldots, T)$ and where

$$\omega_t^{(l)} = \frac{1}{J}\left[D_t + \sum_{j \in \mathcal{J}_t} \mathbb{1}_{(z_{jt}^{(l)}=1)} \right];$$

in the above expression, D_t denotes the number of quadrats in which the presence of the target species has been detected at time t, and $\mathcal{J}_t$ denotes the set of indices j for which z_{jt} is missing (t being fixed).

4.3 *The Bayesian model averaging procedure*

Assume that M plausible models have been considered as being liable to have generated the data at hand $\mathbf{y}$. We denote by $p(m)$ the prior probability of model m, by θ_m the global parameter of model m, and by $\pi(\theta_m)$ the prior density of θ_m. In this context, we propose to implement a Bayesian model averaging procedure to deal with the test (4.2). Such a procedure is standard; it involves a quantity of interest which is typically a parameter defined across all the M models (see eg Kass and Raftery, 1995). In our case, the quantity of interest is $\Pr(\mathrm{H}_0)$. The model averaging is based on the following equality:

$$\Pr(\mathrm{H}_0|\mathbf{y}) = \sum_{m=1}^{M} p(m|\mathbf{y}) \Pr(H_0|m, \mathbf{y})$$

where $p(m|\mathbf{y})$ represents the posterior probability of model m, and $\Pr(\mathrm{H}_0|\mathbf{y}, m)$ the posterior probability of H_0 under model m. In the previous section, we have explained how to compute $\Pr(H_0|m, \mathbf{y})$ via a data augmentation algorithm. What remains is to compute $p(m|\mathbf{y})$. In missing data models, this task is often untractable, which is why a model averaging procedure is typically undertaken via RJMCMC methods (eg Robert and Casella, 2004). Here we provide a closed-form expression for each $p(\mathbf{y}|m)$ from which the $p(m|\mathbf{y})$'s are immediately derived.

By applying the Bayes formula, one has:

$$p(m|\mathbf{y}) = \frac{p(\mathbf{y}|m)p(m)}{\sum_{m=1}^{M} p(\mathbf{y}|m)p(m)}$$

where

$$p(\mathbf{y}|m) = \int_{\Theta_m} L(\theta_m; \mathbf{y})\pi(\theta_m) \, \mathrm{d}\,(\theta_m)$$

and where $L(\theta_m; \mathbf{y})$ denotes the likelihood of data $\mathbf{y}$ under model m. Considering that

$$L(\theta_m; \mathbf{y}) = p(\mathbf{y}|\theta_m) = \sum_{\mathbf{z}_{mis}} p(\mathbf{y}, \mathbf{z}_{mis}|\theta_m) = \sum_{\mathbf{z}_{mis}} L(\theta_m; \mathbf{y}, \mathbf{z}_{mis})$$

where $L(\theta_m; \mathbf{y}, \mathbf{z}_{mis})$ denotes the likelihood of the completed data $(\mathbf{y}, \mathbf{z}_{mis})$ under model m, and that

$$\int_{\Theta_m} \sum_{\mathbf{z}_{mis}} L(\theta_m; \mathbf{y}, \mathbf{z}_{mis}) \pi(\theta_m) \, \mathrm{d}\,(\theta_m) = \sum_{\mathbf{z}_{mis}} \int_{\Theta_m} L(\theta_m; \mathbf{y}, \mathbf{z}_{mis}) \pi(\theta_m) \, \mathrm{d}\,(\theta_m)\,,$$

one observes that computing exactly $p(\mathbf{y}|m)$ requires in fact exact computation of the integral $\int_{\Theta_m} L(\theta_m; \mathbf{y}, \mathbf{z}_{mis}) \pi(\theta_m) \, \mathrm{d}\,(\theta_m)$. Data $\mathbf{y}$ being fixed, this integral depends only on model m and on $\mathbf{z}_{mis}$; it is denoted by $I_m(\mathbf{z}_{mis})$.

The above computational developments are quite general and apply to any missing data model. In the framework of the four models considered in Section 3, $I_m(\mathbf{z}_{mis})$ can be written down in a closed form, whatever m and $\mathbf{z}_{mis}$. We develop below the calculations of $I_m(\mathbf{z}_{mis})$ for the general model $[q_t, \psi_t(r,s)]$. First, we need to calculate the likelihood of the completed data. Considering assumptions A1, A2, A3, A4 and A5, it is easy to check that

$$L(\theta; \mathbf{y}, \mathbf{z}_{mis}) = \mu^{V_1}(1-\mu)^{J-V_1} \left[\prod_{t=1}^{T} q_t^{U_t} (1-q_t)^{KV_t - U_t} \right] \left[\prod_{t=1}^{T-1} \prod_{r,s \in \{0,1\}} \psi_t^{W_t(r,s)}(r,s) \right]$$

where

$$U_t = \sum_{j=1}^{J} y_{jt}, \quad V_t = \sum_{j=1}^{J} z_{jt} \quad \text{and} \quad W_t(r,s) = \sum_{j=1}^{J} \mathbb{1}_{(z_{jt}=r, z_{j,t+1}=s)} \qquad (4.5).$$

Now, it is straightforward to see that $I_m(\mathbf{z}_{mis}) = P_\mu P_q P_\psi$ with

$$P_\mu = B(1+V_1, 1+J-V_1) \qquad P_q = \prod_{t=1}^{T} B(1+U_t, 1+KV_t - U_t)$$

and

$$P_\psi = \prod_{t=1}^{T-1} B\left(1+W_t(0,1), 1+W_t(0,0)\right) \; B((1+W_t(1,0), 1+W_t(1,1))$$

where $B(.,.)$ denotes the coefficient of the Beta distribution; in these expressions, the count V_t and $W_t(r,s)$ are derived from the completed data $(\mathbf{z}_{mis}, \mathbf{y})$, and the count U_t from $\mathbf{y}$. Consequently, computing exactly $I_m(\mathbf{z}_{mis})$ does not raise any difficulty, since the value of $B(.,.)$ is provided by most software.

This way of proceeding yields the exact value of $p(m|\mathbf{y})$. It avoids the use of RJMCMC methods, but it may be time consuming when the number of missing values is large. For the data set analyzed in this paper, it is not the case. If it is, we advocate the use of RJMCMC methods. Indeed, if one takes advantage of the parallel existing between time site occupancy models and multi-state capture-recapure models mentioned in Section 3.1, $\Pr(H_0|\mathbf{y})$ should be easily obtained using the RJMCMC algorithm developed by King and Brooks (2000) in a multi-state capture-recapure set-up (provided that some modifications are made in it). For large enough J, an alternative to RJMCMC methods is to use approximations of $p(m|\mathbf{y})$; see eg Kass and Raftery (1995). We do not go further in these directions.

5. An illustration

5.1 *Description of the Nesquive wood data set and the biological hypotheses*

The data considered in this paper are a part of a large data set collected from 2003 to 2015 for studying the bird species population present in the wood of Nesquive (located near Toulouse in France). Only the period from 2011 to 2015 here is concerned. The protocol is similar to the one described in detail in Dupuis, Bled and Joachim (2010) and in Dupuis and Goulard (2011); therefore only the main points are given here. It has been divided into 10 quadrats of equal size (250 m x 250 m); all the quadrats have been visited. We consider that, for a given year $a \in \{2011, \ldots, 2015\}$, a given species occupies quadrat j if at least one individual belonging to this species has nested in quadrat j, during year a. Information about the presence of nesting species is provided by acoustic recognition

of singing males according to the following procedure. The researcher spent a prescribed time (twenty minutes in our study) at each station (in the center of quadrat), listening to birds. More precisely, data have been collected according to the following point count protocol: each 20-minute session has been sliced into four subsessions of 5 minutes each, during which the experimenter records whether the target species has been detected, or not. Each slice is the equivalent of a visit, therefore $K = 4$; as in Dupuis, Bled and Joachim (2010) and in Dupuis and Goulard (2011).

In this paper, we focus on two bird species: the chiffchaff and the Bonelli warbler. For each of these two target species, Table 1 provides the number D_a of quadrats where they have been detected in year $a \in \{2011, \ldots, 2015\}$.

[Table 1 about here.]

As far as the chiffchaff is concerned, the increase in the D_a's over the period 2012 - 2014 suggests a colonization over this period, and leads us naturally to test the following null hypothesis

$$\mathrm{H}_0 : \quad \omega_{2012} < \omega_{2013} < \omega_{2014} \,. \tag{5.1}$$

Of course, the year 2011 could be included in H_0, since it is quite possible that $\omega_{2011} < \omega_{2012}$ (even if we do not have $D_{2011} < D_{2012}$). In fact, the posterior probability of the corresponding H_0 is very small, and it is the above option which is relevant to test. Concerning the period 2011 - 2012 during which the number of quadrats in which the chiffchaff has been detected remained stable ($D_{2011} = D_{2012} = 7$), this is the following test:

$$\mathrm{H}_0 : \quad \omega_{2011} = \omega_{2012} \quad \text{versus} \quad \mathrm{H}_1 : \quad \omega_{2011} \neq \omega_{2012} \,. \tag{5.2}$$

which is relevant. Such a test will be called a *stability test*. Concerning the period 2014 - 2015, we observe that $D_{2014} = D_{2015} = 10$. As already mentioned in Section 2, the fact

that $D_{2014} = D_{2015} = J$ implies that the corresponding occupancy rates are known; one has $\omega_{2014} = \omega_{2015} = 1$ (this is why ω_{2015} has been excluded from the test (5.1)). In other terms, we can state that, with probability 1, the occupancy rate remained stable over the period 2014 - 2015.

As far as the Bonelli warbler is concerned, the decrease in the D_a's suggests to test a decrease in occupancy rates over the period 2011 - 2015, and thus to consider the following null hypothesis $\mathrm{H}_0 : \quad \omega_{2011} > \omega_{2012} > \omega_{2013} > \omega_{2014} > \omega_{2015}$. In fact, it is much more relevant to split the period 2011 - 2015 into two parts by testing separately the hyptohesis

$$\mathrm{H}_0 : \quad \omega_{2011} > \omega_{2012} > \omega_{2013} > \omega_{2014}, \tag{5.3}$$

and the hypothesis:

$$\mathrm{H}_0 : \omega_{2014} > \omega_{2015}. \tag{5.4}$$

This way of proceeding will allow to highlight that there is a strong evidence in favor of the null hypothesis (5.3). If we had worked with the null hypothesis $\mathrm{H}_0 : \quad \omega_{2011} > \omega_{2012} > \omega_{2013} > \omega_{2014} > \omega_{2015}$ we would have totally missed this interesting result concerning the period 2011 - 2014; the year 2015 (during which the Bonelli warbler has not been detected) impeding the understanding of this period (see the Section 5.2.2 for details).

5.2 *Results.*

For both data sets (one by target species), we have implemented the Bayesian averaging model procedure described in Section 4.3; the four models involved in this procedure are $[q, \psi(r,s)]$,$[q_t, \psi(r,s)]$, $[q, \psi_t(r,s)]$ and $[q_t, \psi_t(r,s)]$. We have taken $p(m) = 1/4$, putting thus an equal prior weight on each model. When inference takes place within a model selection procedure, Burnham and Anderson (1998) recommend that at least one model fit the data adequately. We believe that this common-sense recommandation also applies to an averaging model procedure, whether it is classical or Bayesian. The fit has been assessed by

computing a Bayesian p-value (eg King *et al.*, 2010). Recall that, a model being considered, values of the Bayesian p-value close to 0 or 1 leads to suspect the goodness-of-fit of the model. For both data sets, the fit of the model $[q_t, \psi(r,s)]$ is acceptable: the Bayesian p-values obtained are 0.19 for the chichaff, and 0.21 for the Bonelli warbler (calculations having been performed by using the deviance as discrepancy function). The calculation of $p(\mathbf{y}|m)$ has been performed on a PC, by using the software C; the computational time is very short (it does not exceed 15 seconds). The Bayesian answers to the tests (5.1) and (5.2) have been based on the data set involving the five years, in the aim of not lose information; recall indeed that, for fixed j, the y_{jt}'s are not independent (due to Assumption A1). This remark also applies for the tests (5.3) and (5.4).

5.2.1 The chiffchaff Table 2 below provides, for each model m, the value of the integrated likelihood (that is $p(\mathbf{y}|m)$), its posterior probability (that is $p(m|\mathbf{y})$), and the posterior probabilities of the null hypotheses (5.1) and (5.2) respectively denoted by $\Pr(H_0^{[1]}|m,\mathbf{y})$ and $\Pr(H_0^{[2]}|m,\mathbf{y})$ where $\mathbf{y}$ represents the chiffchaff data set.

[Table 2 about here.]

We first observe that the posterior probabilities of the models $[q, \psi_t(r,s)]$ and $[q, \psi(r,s)]$ are very small, and that the model $[q_t, \psi(r,s)]$ clearly dominates. From Table 2, we deduce that the averaged posterior probability of the stability test is 0.87, and that the one of the colonization test is 0.89.

5.2.2 The Bonelli warbler Table 3 below provides, for each model m, the value of the integrated likelihood, its posterior probability and the posterior probabilities of the null hypotheses (5.3) and (5.4) respectively denoted by $\Pr(H_0^{[3]}|m,\mathbf{y})$ and $\Pr(H_0^{[4]}|m,\mathbf{y})$ where $\mathbf{y}$ represents the Bonelli warbler data set.

[Table 3 about here.]

We first observe that the models where the detection probabilities depend on time (namely the models $[q_t, \psi_t(r,s)]$ and $[q_t, \psi(r,s)]$) clearly dominate: the cumulative sum of their posterior model probabilities reaches 0.95. From Table 3, we deduce that the averaged posterior probability of the hypothesis (5.3) is 0.98 and that the averaged posterior probability of the hypothesis (5.4) is 0.60. This second result may seem somewhat surprising, insofar as $D_{2014} = 3$ (and therefore $\omega_{2014} \geq 0.3$) and $D_{2015} = 0$. In fact, it explains mainly by the fact that the Bonelli warbler has not been detected in the Nesquive wood in 2015, which generates some substantial uncertainty about ω_{2015}. The small size of J contributes also to this result, but in a much less extent. Similar observations has been made in Dupuis *et al.* (2010) while estimating the occupancy rates of not detected species. However that may be, in 2015, all the z_{jt}'s are missing and information contained in the data about ω_{2015} is thus poor. For assessing the impact of such a situation, we provide the posterior distribution of ω_{2015} and we compare it to the one of ω_{2014} (this is performed under the dominating model $[q_t, \psi(r,s)]$): see Figures 1 and 2 below.

[Figure 1 about here.]

We observe that the essential of the posterior distribution of ω_{2014} is concentrated at $\omega_{2014} = 0.3$ since $\Pr(\omega_{2014} = 0.3|\mathbf{y}) \approx 0.80$ (more precisely, 0.77). Note that this peak coincides with D_{2014}/J.

[Figure 2 about here.]

Conversely, we observe that the posterior distribution of ω_{2015} is much more widespread. Unlike the year 2014, the peak of the posterior distribution of ω_{2015} does not coincide with D_{2015}/J; we have in fact two peaks (of comparable magnitude) which occurs for $\omega_{2015} = 0.2$

and $\omega_{2015} = 0.3$. Note that the posterior probability that $\omega_{2015} \geq 0.3$ is relatively high (0.49 precisely). All these elements contribute to the observed result for $\Pr(H_0^{[4]}|m, \mathbf{y})$ when $m = [q_t, \psi(r, s)]$. The examination of the posterior distributions of ω_{2014} and of ω_{2015} under the model $[q_t, \psi_t(r, s)]$ leads to very similar observations (details are omitted).

For information, the averaged posterior probability of the null hypothesis $\mathrm{H}_0 : \quad \omega_{2011} > \omega_{2012} > \omega_{2013} > \omega_{2014} > \omega_{2015}$ is 0.56; this result clearly conceals the fact that the hypothesis of a decrease of occupancy rates during the period 2011 - 2014 has a strong support (namely 0.98).

6. Conclusion

The tests developed in this paper allowed us to investigate important biological questions concerning the evolution of occupancy rates over time; each test being associated with a specific behavior of the target species: either a decrease, or an increase or stability in its occupancy rates. We believe that it would not have been possible outside the Bayesian paradigm, simply because the test is on the occupancy rates, not on the model parameter. In this last Section, we highlight some of the potential of our statistical approach (which consists in basing inference on the posterior probability of the null hypothesis averaged over a set of plausible models). We develop this point along two lines. In Section 6.1, we examine to what extent our approach could be useful for the asymptotic framework (already mentioned in the introduction and in Section 4.2). In Section 6.2, J is finite and we explore some of the numerous biological questions related to occupancy which can be addressed by our approach (with straightforward adaptations). In some situations, implementing our approach, will require either substantial modifications (from a computational point of view) or will consitute real computational challenges. This particular point is addressed in Section 6.3.

6.1 *The asymptotic framework*

In some surveys, the size of R is very large compared to the size of the site, and, as a result J is very large: for comments on the size of J see MacKenzie *et al.*, (2006), MacKenzie *et al.* (2009), as well as Dupuis and Goulard (2011). In this asymptotic framework, J is no more a part of the model and the very notion of occurancy rate is not applicable; however, the question arises as if the colonization test (1.1) could be reformulated in this framework. Considering the limit of ω_t when $J \longrightarrow \infty$ provides an answer (this limit is denoted by ψ_t in MacKenzie *et al.*, (2006) and by λ_t in this paper). Indeed, if one assumes that the z_{jt}'s are independent and identically distributed, ω_t converges in probability to $\lambda_t = E[z_{jt}]$ which thus represents the probability that the target species is present in a quadrat, at time t. Therefore, in the asymptotic framework, the test (1.1) may be reformulated as follows:

$$\mathrm{H}_0 : \quad \lambda_1 < \lambda_2 < \ldots < \lambda_{T-1} < \lambda_T \,. \tag{6.1}$$

The test now involves the occupancy parameters λ_t instead of the occupancy rates ω_t. The quantity λ_t is not a parameter of models but it is clearly a transversal quantity. It can be indeed expressed as a function of the transition probabilities by using recursive formulae. Indeed, for all $1 \leq t \leq T-1$ one has: $\lambda_{t+1} = \psi_t(1,1)\lambda_t + \psi_t(0,1)(1-\lambda_t)$ in models $[q, \psi_t(r,s)]$ and $[q_t, \psi_t(r,s)]$, and $\lambda_{t+1} = \psi(1,1)\lambda_t + \psi(0,1)(1-\lambda_t)$ in models $[q, \psi(r,s)]$ and $[q_t, \psi(r,s)]$; moreover, one has $\lambda_1 = \mu$. The data augmentation algorithm (in its asymptotic version; see Appendix) allows us to obtain $\Pr(\mathrm{H}_0|m, \mathbf{y})$, while the model averaging procedure remains unchanged. It is of interest to point out that implementing the parametric test (6.1) by adopting a Bayesian view is thus straightforward, whereas its classical treatment does not seem trivial at all.

In the asymptotic framework, the stability test considered in Section 5 becomes:

$$\mathrm{H}_0 : \quad \lambda_1 = \lambda_2 = \ldots = \lambda_{T-1} = \lambda_T. \tag{6.2}$$

if stability is tested for the period from $t = 1$ to $t = T$. The Bayesian treatment of such a parametric test requires us to modify the prior considered in Section 3.3 to ensure that $\Pr(\mathrm{H}_0) > 0$ (hitherto, this condition was always satisfied). Even if the theory to deal with such a point null hypothesis is well documented (eg Robert, 2007), its effective implementation involves specific developments which are beyond the scope of this paper. However, some authors have argued that a point null hypothesis of type $\theta = \theta_0$ (where θ denotes a real parameter) does not necessarily well represent the question of interest, whereas a null hypothesis of type $|\theta - \theta_0| \leq \epsilon$ (where ϵ is small) is often more realistic and more sensitive: see eg Berger and Delampady (1987), or Dupuis (1997), for motivations. If this point of view is proving to be relevant for the practitioner, the question of interest associated with the null hypothesis (6.2) may be reformulated via the following null hypothesis:

$$\mathrm{H}_0 : \quad |\lambda_{t_2} - \lambda_{t_1}| \leq \epsilon \quad \text{for all } t_1 \text{ and } t_2 \text{ in } \{1,\ldots,\mathrm{T}\} \tag{6.3}$$

where $\epsilon \in]0, 1[$ represents typically a small value fixed by the practitioner (for example 0.1 or 0.05 to fix ideas). The null hypothesis (6.3) means that the distance between any two occupancy parameters does not exceed ϵ. Contrary to (6.2) and interestingly, the statistical methodology developped in this paper allows us to deal with the quasi-stability test (6.3), as for the test (6.1). This is essentially because the prior appearing in Section 3.3 does not need to be modified since $\Pr(\mathrm{H}_0)$ is now > 0.

6.2 *Some of the biological questions addressed by our methodology*

Example 1. Assume that one considers that the target species is endangered if, at a given time t, its occupancy rate is smaller than a threehold ω^*. The statistical test below is of interest:

$$\mathrm{H}_0 : \quad \omega_t < \omega^* \quad v.s. \quad \mathrm{H}_1 : \quad \omega_t \geq \omega^* . \tag{6.4}$$

Example 2. In multi-state occupancy models, the status of the quadrat does not only

concern the presence of the target species. Let us illustrate this point from a situation considered by Mackenzie *et al* (2009). These authors defined three states for z_{jt}, as follows: $z_{jt} = 1$ if the species is present and breeds successfully in quadrat j at time t; $z_{jt} = 2$ if it is present but fails to breed; the meaning of $z_{jt} = 0$ is unchanged. In such a framework, a simple issue is to test the following null hyptohesis:

$$\mathrm{H}_0 : \omega_{1,t} > \omega_{2,t} \quad \text{for all} \quad t \in \{1, \ldots, T\} \tag{6.5}$$

where $\omega_{r,t}$ denote the proportion of quadrats having the breeding status r, where $r \in \{1, 2\}$.

Example 3. Two species called s_1 and s_2 are now involved, and we would like to test whether, at a given time t, the occupancy rate of species s_1 is greater than the one of species s_2. The associated test is:

$$\mathrm{H}_0 : \quad \omega_{s_1,t} < \omega_{s_2,t} \quad v.s. \quad \mathrm{H}_1 : \quad \omega_{s_1,t} \geq \omega_{s_2,t} \tag{6.6}$$

with obvious notation.

In the two first examples the data augmentation sheme (4.3), as well as our model averaging procedure, apply. In the third example, this remark also holds, on the condition that species s_1 and s_2 occupy the quadrats independently.

6.3 *Two computational challenging issues*

We assume in this paper that detectability and occupancy parameters do not depend on quadrats. However, sometimes, time specific covariates are available at the quadrat level. When they are discrete, implementing the data augmentation algorithm and the model averaging procedure only requires simple adaptations. When the available covariates are continuous, a logit formulation is typically adopted. In a such situation, the data augmentation algorihm (4.3) still applies, but Hastings-Metropolis steps have to be included in it for updating the logit parameters (Dorazio and Rodriguez, 2012). As far as the model averaging procedure is concerned, there is no longer a closed-form expression

for the posterior probability of each model and, typically, a RJMCMC algorithm has to be implemented. In example 3 of Section 6.2, we have considered a test which involves two species s_1 and s_2, and we have assumed that they occupy the quadrats independently. If this assumption is removed, implementing the data augmentation algorithm as well as a model averaging procedure, becomes clearly much more difficult and constitutes a true computational challenge. In this paper, a simple assumption is adopted for the spatial occupancy process. Recall that, for any fixed t, we assume that $z_{1t}, \ldots, z_{jt}, \ldots, z_{Jt}$ are independent (see Assumption A4). Introducing some spatial correlations between the z_{jt}'s will typically be done via an autologistic formulation, and inference will take place within hidden Markov field models (eg Cressie and Wikle, 2011). Implementing a model averaging procedure within such a framework is clearly another computational challenge.

7. References

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses *Statistical Science.* **2**, 317-352.

Burnham, K. P. and Anderson D., R. (1998). *Model Selection and Inference. A Practical Information-Theoretic Approach.* Springer. New York.

Cressie, N., and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data.* Wiley.

Dorazio, R. M. and Rodriguez, C. (2012). A Gibbs sampler for a Bayesian analysis of site-occupancy data *Methods In Ecology and Evolution* **3**, 1093-1098

Dupuis, J. A. (1995). Bayesian estimation of movement and survival probabilities from capture-recapture data. *Biometrika* **82**, 761-772.

Dupuis, J. A. (1997). Bayesian test of homogeneity for Markov chains. *Statistics and Probability Letters* **31**, 333-338.

Dupuis, J.A and Schwarz, J.C. (2006). The stratified Jolly-Seber capture-recapture model: a Bayesian approach. *Biometrics* **67**, 290-298

Dupuis, J.A., Bled, F., and Joachim, J. (2010). Estimating occupancy rate of spatially rare or hard to detect species: a conditional approach. *Biometrics* **67**, 290-298

Dupuis, J.A. and Goulard, M. (2011). Estimating species richness from quadrat sampling data: a general approach. *Biometrics* **67**, 1489-1497

Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C. (1995). A Bayesian model averaging: a tutorial. *Statistical Science* **14**, 382-417.

Kass, R.E. and Raftery, A. (1995). Bayes factor. *Journal of the American Statiscal Association.* **90**, 773-795.

Lehmann, E.L. and Romano, J.P. (2008). *Testing Statistical Hypothesesis.* Third Edition. Springer (New York).

Little, J. and Rubin, D. (2002). *Statistical Analysis with Missing data.* 2nd Edition. Wiley.

King, R. and Brooks S.P. (2002). Bayesian model discrimination for multiple strata capture-recapture data. *Biometrika* **89**, 785-806.

King, R., Morgan, B.J.T., Gimenez, O. and Brooks, S.P. (2010). *Bayesian Analysis for Population Ecology.* Chapman Hall.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle A.J. and Langtimm, A.C. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**, 2248-2255.

MacKenzie, D.I., Nichols, J.D., Seamans, M. E. and Guiterrez, R. J. (2009). Modeling species occurence dynamics with multiple states and imperfect detection . *Ecology*, **90**, 823835.

MacKenzie, D.I. and Bailey, L,L. (2004). Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics* **9**, 300-318

MacKenzie, D.I., Nichols, J.D., Royle J.A., Pollock, K.H., Bailey, L.L., Hines, J.E. (2006) *Occupancy Estimation and Modeling.* Elsevier, Amsterdam.

Robert, C.P. and Casella, G. (2004). *Monte-Carlo Statistical Methods*. Edition 2nd. Springer-Verlag.

Robert, C.P. (2007). *The Bayesian Choice: From Decision-Theoretic Fundations to Computational Implementation*. Springer-Verlag.

Royle, J. A. and Kéry, M. (2007). A Bayesian state-space formulation of dynamic occupancy models. *Ecology*, **88**, 18131823.

Tanner, M. and Wong, M. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statiscal Association.* **85**, 699-704.

Whittaker, J. (1990). *Graphical models in Applied Mathematical Multivariate Statistics*. Wiley.

Appendix

First, we assume that J is finite, and we show how to implement the data augmention algorithm corresponding to the schema (4.3). It is done by distinguish two cases, depending on whether $n < J$ quadrats have been visited, or whether all the quadrats have been visited ($n = J$). We show in particular how to use the graph of the model for calculating the distribution according to a missing z_{jt} has to be simulated. To compute this distribution we will use the separation theorem and the local Markov property. We refer to the book of Whittaker (1990) for details: the separation theorem being stated in Section 3.3, and the local Markov property in Section 3.4.

Second, we consider the asymptotic framework.

We will limit ourselves to the general model $[q_t, \psi_t(r,s)]$ because the developments concerning the three sub-models $[q, \psi(r,s)]$, $[q_t, \psi(r,s)]$, $[q, \psi_t(r,s)]$ are very similar.

1. *The case J is finite.*

1.1. *The case $n = J$.*

The data augmentation algorithm is a Gibbs sampling implemented on the vector $(\theta, \mathbf{z}_{mis})$ where the components of $\mathbf{z}_{mis}$ are ordered according to the increasing j's, and for fixed j, according to the increasing t's.

• The missing data simulation phase.

A missing z_{jt} is simulated according to:

$$p(z_{jt} \,|\mathbf{z}_{mis} \setminus \{z_{jt}\}, \mathbf{y}, \theta). \tag{1}$$

Due to A4 and A5, one has: $p(z_{jt} \,|\mathbf{z}_{mis} \setminus \{z_{jt}\}, \mathbf{y}, \theta) = p(z_{jt} \,|\mathbf{z}_j^{mis} \setminus \{z_{jt}\}, \mathbf{y}_j, \theta)$ where $\mathbf{z}_j^{mis}$ represents the components of $\mathbf{z}_j$ which are missing. Consequently, simulating a missing z_{jt} requires us to compute the conditional distribution:

$$p(z_{jt} \,|G_j \setminus \{z_{jt}\}, \theta) \tag{2}$$

where $G_j = \{z_{jt}; t = 1, \ldots, T\} \bigcup \{y_{jt}; t = 1, \ldots, T\}$. A repeated use of the Bayes formula, of the separation theorem, and of the local Markov property will allow us to significantly simplify (2). The separation theorem and the local Markov property are usually stated in undirected graphs, but these properties also apply to the directed graph $\mathcal{G}$ (which appears in Section 3.1) because it meets the *Wermuth condition*: see Section 3.5 of the book of Whittaker (1990) for this last point.

For convenience, θ is afterwards omitted in the conditionings. Considering the graph $\mathcal{G}$ and by using the local Markov property, one has:

$$p(z_{jt} \,|\, G_j \setminus \{z_{jt}\}) = p(z_{jt}|z_{j,t+1}, y_{jt}, z_{j,t-1}) \quad \text{where } 2 \leq t \leq T-1. \tag{3}$$

Using the Bayes formula, we deduce that:

$$p(z_{jt}|z_{j,t+1}, y_{jt}, z_{j,t-1}) \propto p(y_{jt}|z_{j,t+1}, z_{jt}, z_{j,t-1}) p(z_{jt}|z_{j,t+1}, z_{j,t-1}). \tag{4}$$

Now one has $p(y_{jt}|z_{j,t+1}, z_{jt}, z_{j,t-1}) = p(y_{jt}|z_{jt})$ since z_{jt} separates y_{jt} and $\{z_{j,t+1}, z_{j,t-1}\}$. We thus deduce that:

$$p(z_{jt}|z_{j,t+1}, y_{jt}, z_{j,t-1}) \propto p(y_{jt}|z_{jt}) p(z_{jt}|z_{j,t+1}, z_{j,t-1}). \tag{5}$$

Using again the Bayes formula, then the fact that z_{jt} separates $z_{j,t+1}$ and $z_{j,t-1}$, we find that:

$$p(z_{jt}|z_{j,t+1}, z_{j,t-1}) \propto p(z_{jt}|z_{j,t-1}) p(z_{j,t+1}|z_{jt}), \tag{6}$$

so that, finally, one has:

$$p(z_{jt}|z_{j,t+1}, y_{jt}, z_{j,t-1}) \propto p(z_{jt}|z_{j,t-1}) p(y_{jt}|z_{jt}) p(z_{j,t+1}|z_{jt}), \tag{7}$$

from which we deduce a unique and general expression for the conditional distribution of a missing z_{jt} (see the equation (8) below). It is easy to modify (7) to deal with the

two particular cases $t = 1$ and $t = T$. Note that, interestingly, (7) applies without any modification to the multi-state occupancy models. Interestingly, this way of conducting the calculations easily applies to more complex graphs than $\mathcal{G}$. For example, if a second order Markov chain is assumed for the occupancy process $(z_{jt}; t = 1, \ldots, T)$, it is easy to check that $p(z_{jt} \,|\, G \setminus \{z_{jt}\})$ is proportional to:

$$p(z_{jt}|z_{j,t-2}, z_{j,t-1})p(y_{jt}|z_{jt})p(z_{j,t+1}|z_{jt}, z_{j,t-1})p(z_{j,t+2}|z_{jt}, z_{j,t+1}).$$

Royle and Kery (2007) used (3) for computing the conditional distribution of z_{jt}: that requires us to compute the probability that $z_{jt} = 1|y_{jt} = 0$ for each neighborhood $\{z_{j,t+1}, z_{j,t-1}\}$ of z_{jt}. The benefit of simplifications (4), (5), (6) and (7) is proving to be quite interesting in multi-state occupancy issues. Indeed, if $a \geq 3$ denotes the number of states, a^2 probabilities will have to be calculated for obtaining the conditionnal distribution of z_{jt} from (3); if a second order Markov chain is assumed for the occupancy process, a^4 will have to be calculated.

Returning to our model, we provide below the distribution of a missing z_{jt}. Three cases have now to be considered.

- Let $2 \leq t \leq T - 1$. If $z_{jt} \in \mathbf{z}_{mis}$ it has to be simulated according to the distribution $z_{jt}|z_{j,t-1}, y_{jt}, z_{j,t+1}, \theta$ where $y_{jt} = 0$. From (7) we deduce that:

$$\Pr(z_{jt} = 1|z_{j,t-1} = a, y_{sj} = 0, z_{j,t+1} = b, \theta) = \frac{\psi_{t-1}(a,1)(1-q_t)^K \psi_t(1,b)}{\psi_{t-1}(a,1)(1-q_t)^K \psi_t(1,b) + \psi_{t-1}(a,0)\psi_t(0,b)} \tag{8}$$

- If $z_{jT} \in \mathbf{z}_{mis}$, it is simulated according to the distribution of $z_{jT}|z_{j,T}, y_{jT}, \theta$ where $y_{jT} = 0$. It is easy to check that:

$$\Pr(z_{jT} = 1|z_{j,T-1} = a, y_{jT} = 0, \theta) = \frac{\psi_{T-1}(a,1)(1-q_T)^K}{\psi_{T-1}(a,1)(1-q_T)^K + \psi_{T-1}(a,0)} \tag{9}$$

- If the initial state $z_{j1} \in \mathbf{z}_{mis}$, it has to be simulated according to the distribution

$z_{j1}|z_{j,2}, y_{j1}, \theta$ where $y_{j1} = 0$. It is easy to check that:

$$\Pr(z_{j1} = 1|z_{j,2} = b, y_{j1} = 0, \theta) = \frac{\mu(1-q_1)^K \psi_1(1,b)}{\mu(1-q_1)^K \psi_1(1,b) + (1-\mu)\psi_1(0,b)}. \tag{10}$$

We need now to clarify the conditionings. At step (l) of the algorithm, the conditioning which appears for example in (8) is as follows: $z_{j,t-1} = 1$ if $y_{j,t-1} \geq 1$ and $z_{j,t-1} = z_{j,t-1}^{(l)}$ if $y_{j,t-1} = 0$; $z_{j,t+1} = 1$ if $y_{j,t+1} \geq 1$ and $z_{j,t+1} = z_{j,t-1}^{(l-1)}$ if $y_{j,t+1} = 0$; $\theta = \theta^{(l)}$. This a consequence of the (natural) order adopted for ranking the missing z_{jt}'s (see above). A similar principle is applied to the two other cases: $t = 1$ and $t = T$.

• The parameter simulation phase.

Considering the expression of the likelihood of the completed data $L(\theta; \mathbf{y}, \mathbf{z}_{mis})$ appearing in Section (4.3) it is easy to check that, at step (l) of the algorithm, simulation of θ proceeds as follows:

$$q_t^{(l)}|\mathbf{y}, \mathbf{z}_{mis}^{(l-1)} \sim \mathcal{B}e\left(1 + U_t, 1 + KV_t^{(l-1)} - U_t\right),$$

$$\mu^{(l)}|\mathbf{y}, \mathbf{z}_{mis}^{(l-1)} \sim \mathcal{B}e\left(1 + V_1^{(l-1)}, 1 + J - V_1^{(l-1)}\right)$$

and

$$\psi_t^{(l)}(1,0)|\mathbf{y}, \mathbf{z}_{mis}^{(l-1)} \sim \mathcal{B}e\left(1 + W_t^{(l-1)}(1,0), 1 + W_t^{(l-1)}(1,1)\right)$$

$$\psi_t^{(l)}(0,1)|\mathbf{y}, \mathbf{z}_{mis}^{(l-1)} \sim \mathcal{B}e\left(1 + W_t^{(l-1)}(0,1), 1 + W_t^{(l-1)}(0,0)\right)$$

The counts U_t, V_t, $W_t(.,.)$ have been defined in Section 4.3. Note that the counts V_t and $W_t(.,.)$ involve as well observed z_{jt}'s and missing z_{jt}'s, contrary to the count U_t which is deduced from the data $\mathbf{y}$.

1. 2. *The case $n < J$.*

In some surveys, inference is based on $n < J$ sampled quadrats, and the above data augmentation algorithm has to be modified. First, note that z_{jt} is necessarily missing

when the quadrat j is not a part of the sampled quadrats; consequently, such a z_{jt} has to be simulated (recall that ω_t involves as well sampled and unsampled quadrats). It is straightforward to check that the formulae (8), (9) and (10) apply without any modification to sampled and unsampled quadrats, providing that the convention $q_t = 0$ is adopted for the unsampled quadrats. As far as the parameter simulation phase is concerned, only the definitions of the counts U_t and V_t which appear in the simulation of q_t have to be modified, as follows: the sums are now over the n sampled quadrats (instead of J).

2. The asymptotic framework ($J = \infty$).

Inference on the parameters of interest is also based on n sampled quadrats. In the asymptotic framework, $\mathbf{z}_{mis}$ is defined as the missing z_{jt}'s where j represents a sampled quadrat, while in the previous situation (J is finite and $n < J$) the vector $\mathbf{z}_{mis}$ involved the sampled quadrats for which $y_{jt} = 0$ as well as the unsampled quadrats. In fact, in the asymptotic framework, inference is only based on the sampled quadrats (the unsampled quadrats playing no part in the inference). That constitutes a key difference with the finite case $n < J$ and both situations have not to be confused. To implement the data augmentation algorithm in the asymptotic framework, there is no modification to the algorithm as it described in 1.1, except that J has to be replaced by n in the definitions of the counts U_t, V_t and $W_t(r, s)$ appearing in (4.5).

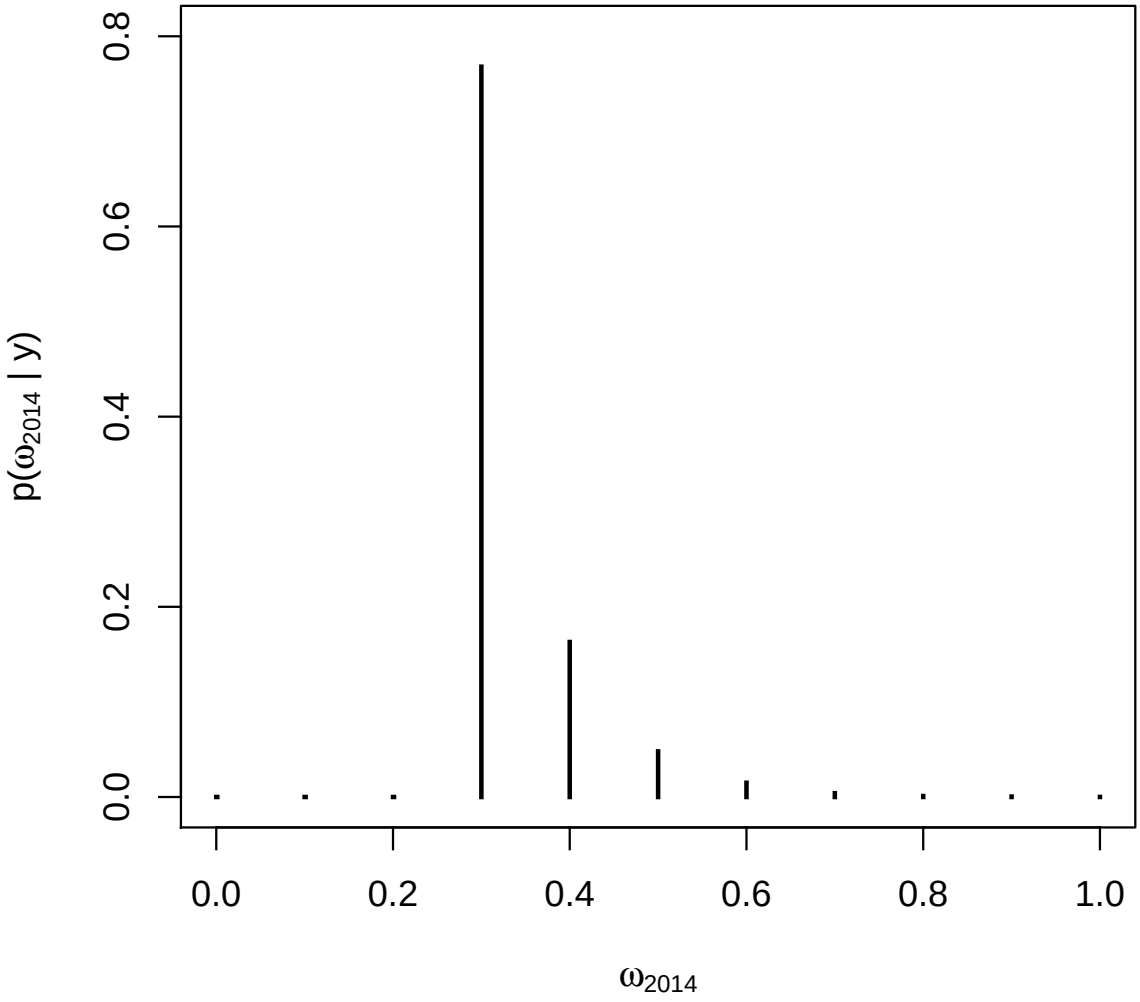

Figure 1. Posterior distribution of ω_{2014} for the Bonelli warbler under the dominating model $[q_t, \psi(r, s)]$.

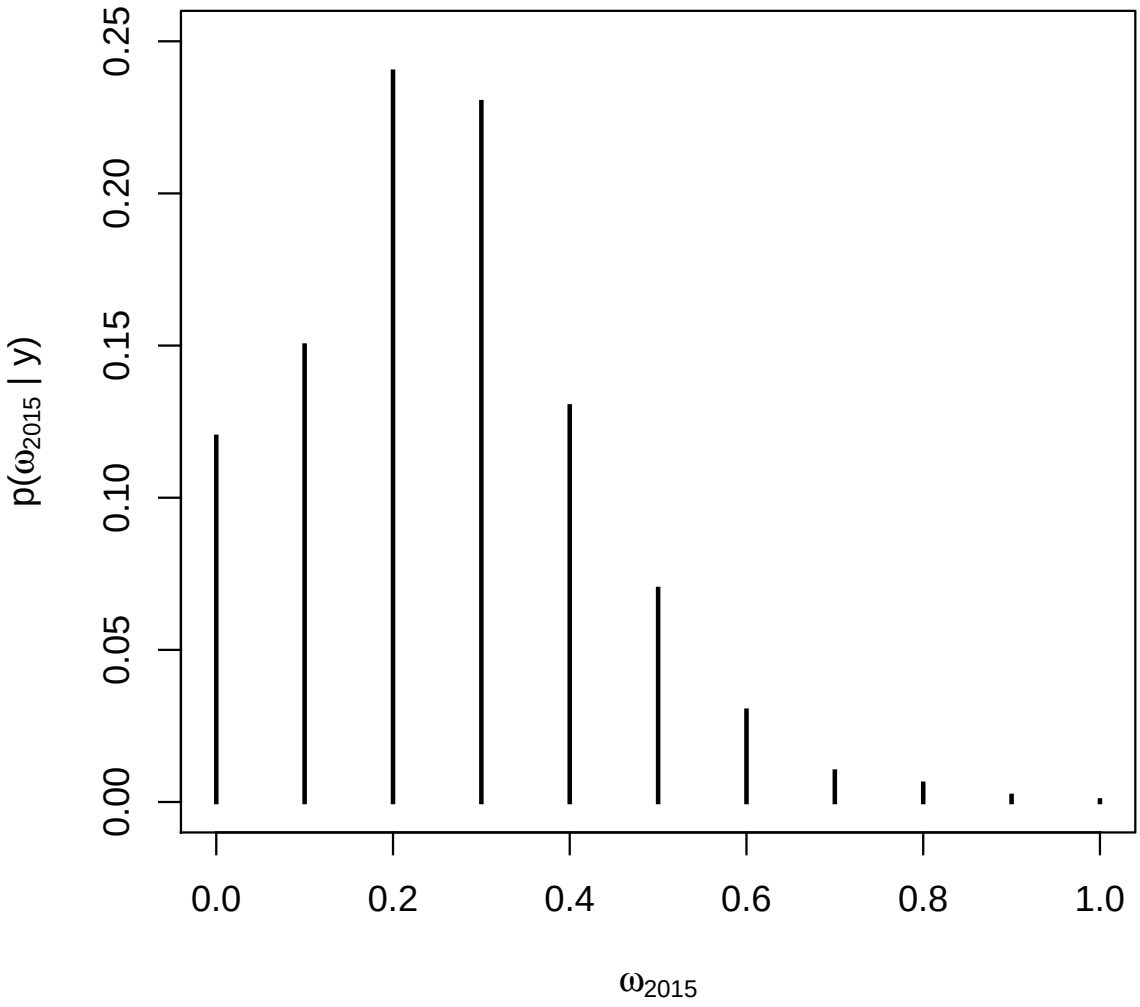

Figure 2. Posterior distribution of ω_{2015} for the Bonelli warbler under the dominating model $[q_t, \psi(r, s)]$.

Table 1

Number of quadrats in which the chiffchaff (respectively the Bonelli warbler) has been detected from 2011 *to* 2015

	2011	2012	2013	2014	2015
chiffchaff	7	7	9	10	10
Bonelli warbler	10	9	6	3	0

Table 2

Values of $p(\mathbf{y}|m)$, $p(m|\mathbf{y})$, $Pr(H_0^{[1]}|m,\mathbf{y})$, $Pr(H_0^{[2]}|m,\mathbf{y})$ *for the chiffchaff data set* $\mathbf{y}$

m	$p(\mathbf{y}\|m)$	$p(m\|\mathbf{y})$	$\Pr(H_0^{[1]}\|m,\mathbf{y})$	$\Pr(H_0^{[2]}\|m,\mathbf{y})$
$[q,\psi(r,s)]$	$5.053887e-44$	0.0008	0.99	0.98
$[q_t,\psi(r,s)]$	$5.228145e-41$	0.8269	0.91	0.86
$[q,\psi_t(r,s)]$	$1.010226e-43$	0.0016	0.99	0.99
$[q_t,\psi_t(r,s)]$	$1.079162e-41$	0.1707	0.85	0.89

Table 3

Values of $p(\mathbf{y}|m)$, $p(m|\mathbf{y})$, $Pr(H_0^{[3]}|m,\mathbf{y})$, $Pr(H_0^{[4]}|m,\mathbf{y})$ for the Bonelli warbler data set $\mathbf{y}$

m	$p(\mathbf{y}\|m)$	$p(m\|\mathbf{y})$	$\Pr(H_0^{[3]}\|m,\mathbf{y})$	$\Pr(H_0^{[4]}\|m,\mathbf{y})$
$[q,\psi(r,s)]$	$5.093365e-44$	0.02	0.99	0.99
$[q_t,\psi(r,s)]$	$1.382167e-42$	0.52	0.98	0.58
$[q,\psi_t(r,s)]$	$6.989075e-44$	0.03	0.99	0.58
$[q_t,\psi_t(r,s)]$	$1.134595e-42$	0.43	0.97	0.60