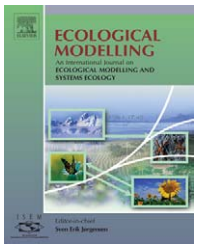
available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ecolmodel



Stochastic modelling of ecological processes using hybrid Gibbs samplers

David M. Walker[a,*], *F. Javier Pérez-Barbería*[b], *Glenn Marion*[c]

[a] *Department of Applied Mathematics, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
[b] *The Macaulay Institute, Craigiebuckler, Aberdeen AB15 8QH, United Kingdom*
[c] *Biomathematics & Statistics Scotland, James Clerk Maxwell Building, King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom*

ARTICLE INFO

Article history:
Received 15 February 2005
Received in revised form
6 April 2006
Accepted 12 April 2006
Published on line 23 June 2006

Keywords:
Stochastic modelling
Markov chain Monte Carlo
Inference
Agent-based models
Parameter estimation
Gibbs and Metropolis-Hastings sampling

ABSTRACT

Stochastic process models are useful in describing a broad range of individual-based phenomena and are increasingly being applied in ecology. However, the estimation of parameters in such models is an important issue which has typically received much less attention than the exploration of model behaviour. The difficulties of parameter estimation are compounded by the fact that in most situations the available data are incomplete in some sense. Here, we demonstrate how methods of Markov chain Monte Carlo (McMC) Gibbs sampling can be combined within reversible-jump Metropolis–Hastings McMC frameworks to produce a hybrid sampler which can be used to obtain estimates of parameters and missing data for a broad class of stochastic process rate models. We apply these methods to two stochastic models arising from the ecology of grazed ecosystems in order to display the benefits of the hybrid sampler and the usefulness of a stochastic modelling approach to experiments where limited data exist.

1. Introduction

The behaviour of many processes studied in the natural sciences can be described, or approximated, by stochastic models. For example, stochastic processes have been used in modelling biological populations (Renshaw, 1991; Bolker and Pacala, 1997; Wilson and Hassell, 1997; Matis et al., 1998; Rand, 1999; Keeling, 2000a,b), epidemics (Isham, 1991; Filipe and Gibson, 1998), chemical reactions (Van Kampen, 1992) and other applications.

Understanding the dynamics of complex natural systems and identifying behavioural states based on some set of observations is a common methodological problem in ecology. Agent-based modelling and simulation is one approach used for understanding the behaviour of complex biological and ecological systems in which agents, such as animals, interact with each other and their environment using simple local rules. A drawback of the agent-based approach is often the weak linkage between models and appropriate field data. There is a need therefore to develop methods and models which better link data with model parameters. This is especially true in ecology where field data are limited, or sparsely sampled, leading to data sets with many missing values due to, for example, bad weather conditions or mortality of the experimental subjects during the study.

A useful framework for developing agent-based and other models when limited data are available is the theory of stochastic processes (Cox and Miller, 1965). The theory pro-

* *Corresponding author.*
E-mail address: enwalker@eie.polyu.edu.hk (D.M. Walker), j.perez-barberia@macaulay.ac.uk (F.J. Pérez-Barbería),glenn@bioss.ac.uk (G. Marion).

0304-3800/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2006.04.008

vides a general and flexible framework both to describe and infer behaviour. It is a powerful tool for developing models and provides a framework to parameterize such models from data. Unfortunately, the non-linearities required to model many real world systems typically lead to stochastic processes that are intractable to analytic solution (Isham, 1991; Bolker and Pacala, 1997; Filipe and Gibson, 1998; Matis et al., 1998; Keeling, 2000a,b); however, simulation is usually straightforward (Renshaw, 1991) unless the expected number of events is extremely large. To place an agent-based model in a stochastic framework, the rules governing individual behaviour are formulated in probabilistic terms. In the special case of exponentially distributed times between events, this task simplifies to one of defining the rates at which each event occurs (note that this is an implicit assumption in many deterministic models based on ordinary differential equations). For example, a simple model for the foraging behaviour of a grazing animal would be described by the rates at which an animal: (i) grazes its current location and (ii) changes location in search of better quality or more abundant food. It would be naive to assume that the behaviour could be fully described by simple relationships of the type described here but the approach leads naturally to the formulation of models which can approximately characterize and capture certain aspects of behaviour. The stochastic approach can also point the way towards better deterministic process models by accounting for variability and spatial heterogeneity using suitable limiting processes and approximations (Whittle, 1957; Isham, 1991; Bolker and Pacala, 1997; Matis et al., 1998; Keeling, 2000a; Holmes et al., 2004; Marion et al., 2005).

In this paper, we focus on the inverse problem of statistical inference: given an observation of a process how can we obtain an indication of the range of model parameters which could plausibly explain the observations? Fitting stochastic dynamical models directly to observations allows parameter uncertainty to be treated more completely since the model itself defines the error distribution and implicitly accounts for correlations in the data. In contrast estimation based on least-squares, as often used for deterministic models, typically makes the additional assumptions that errors are uncorrelated and Gaussian.

However, a full analytic treatment of parameter estimation for dynamical stochastic systems is rarely feasible since observations of biological processes from practical experiments or field studies typically record only a subset of the information that defines the evolution of the system. In such cases, we must "integrate out" the missing information which typically leads to analytically intractable high dimensional integrals. Recent advances in computing power mean that sampling methods and, in particular, McMC (Metropolis et al., 1953; Hastings, 1970; Gelfand and Smith, 1990; Smith and Roberts, 1993; Besag and Green, 1993) are flexible enough to be used to make inferences about missing data and unknown parameters by providing robust approximations to such difficult integrals. The methods are based on Gibbs sampling, Metropolis–Hastings algorithms and the methodological advance of reversible-jump McMC which is specifically tailored to explore state spaces of varying dimension (Gelman et al., 1995; Green, 1995; Gilks et al., 1996; Gamerman, 1997). The need to sample from state spaces of varying dimension arises here because the observed data does not determine the numbers of all event types. Therefore, sampling from the range of plausible reconstructions of the missing data implies sampling over different numbers of reconstructed events. It should be noted that this approach is limited to relatively small numbers of missing events although Marion et al. (in preparation) have recently applied such methods to a case with $\sim 10^4$ missing events.

The joint estimation of parameters and missing data (also referred to as nuisance parameters) is typically conducted within the framework of Bayesian estimation (Lee, 2004) in which explicit quantification of uncertainty in model parameters (and indeed the missing data) is given by their posterior distributions with respect to the observed data and of course the model. A requirement, which should be mentioned, is the need for the selection of subjective prior distribution of parameters in the Bayesian methodology. A potential advantage of this approach is that the shape of the prior can be chosen to quantify information gained from previous studies. In many cases, however, little prior information is available and the prior distribution is often chosen to be uniform, perhaps over some range of parameters determined from the literature. In either case prior influence, lessens with more information and large observation samples mean that posterior distributions are determined largely by the data. In addition, the robustness of results to prior assumptions can be checked. The Bayesian approach coupled with McMC techniques has been applied in recent years to infer the parameters of stochastic epidemic models (Gibson, 1997; Gibson and Renshaw, 1998, 2001a,b; Renshaw and Gibson, 1998; O'Neill and Roberts, 1999).

In this paper, we will apply Bayesian methods using McMC to estimate parameters in stochastic agent-based models. In particular we describe sampling of parameters of stochastic models by both conventional methods (standard Gibbs and Metropolis–Hastings) and the "griddy-Gibbs" approximation (Tanner, 1996). In the conventional approach, we sample from the standard full conditional distributions whereas in the "griddy" method these are approximated using a parameter grid and likelihood calculations. The performance of the various approaches are compared using a simple grazing offtake model in ruminant animals, and their combined use in a hybrid sampler is presented using data from an animal movement and food selection experiment. The hybrid sampler represents a framework for parameter estimation which can be applied to a wide class of stochastic models.

The outline of the remainder of this paper is as follows: in the next section, we introduce the general stochastic modelling framework for agent-based models. We describe the methods of parameter estimation and present the reversible-jump McMC algorithm and the two methods of Gibbs sampling together with Metropolis–Hastings sampling. The combined hybrid sampler is also introduced. In Section 3, we then compare the performance of the sampling algorithms by way of application to two examples from studies of agricultural and ecological systems.

2. Markov chain Monte Carlo samplers

Here, we formulate the problem of inferring parameters from incomplete data for a time-homogeneous Markov process.

In this context, the simple rules of agent-based models can be described in terms of the set of q allowed changes or event types $\{e_i : i = 1, \ldots, q\}$, where event type e_i induces a change δs_{e_i} in the state of the system, denoted $s(t)$ at time t. In other words, if an event of type e_i occurs at time t, the state of the system immediately afterwards is $s(t) + \delta s_{e_i}$. The rate $r(e_i, s(t); \mathbf{a})$ at which event e_i occurs at time t is governed by the state of the system and the vector of model parameters $\mathbf{a}$. More generally, the parameters can depend on the time, but we shall restrict our discussion and studies to the time-homogeneous case. The total event rate at time t is $R(s(t); \mathbf{a}) = \Sigma_{i=1}^{q} r(e_i, s(t); \mathbf{a})$. These rates may be used to define a deterministic or mean-field dynamics in terms of ordinary differential equations (Renshaw, 1991). The stochastic dynamics of the corresponding time-homogeneous Markov process are defined as follows: (i) the time τ to the next event (of any type) is drawn from an exponential distribution with rate $R(s(t); \mathbf{a})$ and (ii) the event type which occurs at time $t + \tau$ is chosen to be type e_i with probability $r(e_i, s(t); \mathbf{a})/R(s(t); \mathbf{a})$ (Cox and Miller, 1965). In combination, (i) and (ii) define a stochastic updating rule such that, conditional on the state of the system being $s(t)$ at time t, the probability that an event of type e_i occurs before any other event type and does so at time $t + \tau$ is given by,

$$P(s(t + \tau) = s(t) + \delta s_{e_i} | s(t)) = r(e_i, s(t); \mathbf{a})\, e^{-\tau R(s(t); \mathbf{a})} \tag{1}$$

Suppose the timings and nature of all events which occur in the interval $[t_0, t_n]$ are observed and recorded. Then, let t_k be the time at which event k in the sequence occurs and denote its type by $E(k) \in \{e_i : i = 1, \ldots, q\}$. Suppose there are n events, then given an initial state $s(t_0)$ the finite and complete realization of the stochastic process, $S = \{s_{t_k}\}_{k=0}^{n}$ can be generated from the set of events: $\mathcal{E} = \{(E(k), t_k) : k = 1, \ldots, n\}$.

The *Likelihood* of the complete data set $\mathcal{E}, L(\mathbf{a}, \mathcal{E}) = P(\mathcal{E}|\mathbf{a}, s(t_0))$, is the probability of observing the complete sequence of events $\mathcal{E}$ given the parameters $\mathbf{a}$ and the initial configuration $s(t_0)$. It is written:[1]

$$P(\mathcal{E}|\mathbf{a}, s(t_0)) = \prod_{k=1}^{n} r(E(k), s(t_{k-1}); \mathbf{a})\, e^{-(t_k - t_{k-1}) R(s(t_{k-1}); \mathbf{a})} \tag{2}$$

and follows directly from the definition of the model via the stochastic update rule (1). Therefore, if complete data are available, Likelihood methods (Edwards, 1992) can be used to estimate model parameters. Moreover, this is also true for non-Markovian stochastic processes, although the form of the likelihood will differ from that shown in Eq. (2). In the sequel, we shall simply write the complete likelihood as $P(\mathcal{E}|\mathbf{a})$ dropping the explicit dependence on the initial condition $s(t_0)$ which may either be regarded as known and fixed or considered as an additional set of parameters to be estimated and thus incorporated into the vector $\mathbf{a}$. Note that we have already suppressed the conditional dependence of the likelihood on the model since we do not compare different models directly in this paper.

[1] In general, the final observation time T may not coincide with the occurrence of the final event at t_n. In such cases, the likelihood (2) should be multiplied by an additional term $e^{-(T - t_n) R(s(t_n); \mathbf{a})}$ describing the probability that nothing happens between t_n and T.

In the case of incomplete data, we observe a set of events $\mathcal{D}$ (the data), but there are also those hidden events $\mathcal{H}$ we do not observe. The complete realization is therefore characterized by the full set of events $\mathcal{E} = (\mathcal{D}, \mathcal{H})$. Applying Bayes' rule[2] to $P(\mathcal{E}|\mathbf{a}) = P(\mathcal{D}, \mathcal{H}|\mathbf{a})$ we obtain the joint *posterior* distribution for the parameters $\mathbf{a}$ and the unobserved events $\mathcal{H}$,

$$P(\mathbf{a}, \mathcal{H}|\mathcal{D}) = \frac{P(\mathcal{D}, \mathcal{H}|\mathbf{a}) P(\mathbf{a})}{P(\mathcal{D})} \tag{3}$$

in terms of the likelihood for complete observations (2), the *parameter prior* $P(\mathbf{a})$ and the normalisation constant $P(\mathcal{D})$. The prior distribution is typically chosen to reflect any knowledge about the parameters available before the data $\mathcal{D}$ were obtained. For example, $P(\mathbf{a})$ may be derived from previous analysis, or simply be a uniform distribution over some plausible range of parameter values as ascertained from appropriate literature. In the absence of such information, the prior is usually chosen to be some convenient form, for example, for the rate parameters considered here, an (unnormalised) flat prior on the positive real line or a gamma distribution. In addition, it is common to assume independence between the priors for each of the N components of the parameter vector, i.e., $P(\mathbf{a}) = \prod_{k=1}^{N} P(a_k)$. It is good practice to test the robustness of any analysis to prior specification.

Bayesian inference (see, e.g., Lee, 2004) is based on the posterior distribution (3) which, for a given set of data, is simply proportional to the likelihood and the prior. For example, the distribution of parameters is given by,

$$P(\mathbf{a}|\mathcal{D}) = \int_{\mathcal{H}} P(\mathbf{a}, \mathcal{H}|\mathcal{D})\, d\mathcal{H} \tag{4}$$

which is just the joint posterior (3) marginalised over the hidden events. However, this integral is typically analytically intractable and the space of possible hidden events too large to allow evaluation by quadrature. Moreover, evaluation of the normalisation constant $P(\mathcal{D})$ in (3) involves integrals of similar computational complexity. Fortunately, Markov chain Monte Carlo techniques, allow parameter samples to be drawn directly from the posterior $P(\mathbf{a}, \mathcal{H}|\mathcal{D})$ without having to calculate the normalisation constant $P(\mathcal{D})$. The Metropolis–Hastings algorithm and Gibbs sampling allow parameter samples to be drawn directly from the posterior, but since the number of unobserved events is in general unknown, in sampling over $\mathcal{H}$, the Markov chain must explore spaces of varying dimension (corresponding to the numbers of events in a given realization) requiring application of reversible jump McMC (Green, 1995). Variants on a combined algorithm allowing sampling over both parameters and missing events are described in detail below in Sections 2.1–2.5.

The samples generated from the posterior $P(\mathbf{a}, \mathcal{H}|\mathcal{D})$ using McMC allow the calculation of essentially any statistic based on the parameters, $\mathbf{a}$, and missing events, $\mathcal{H}$. For example, the marginal distribution of parameters described by Eq. (4) may be estimated by simply disregarding the sampled hidden events and forming a histogram of the sampled parameter values only. The marginal distribution of any single parame-

[2] $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.

ter (component of $\mathbf{a}$) or the joint distribution of two or more may be obtained in a similar fashion. Such estimates improve as the number of samples generated from the Markov chain increases.

2.1. *Reversible-jump Metropolis–Hastings algorithm*

To generate samples of the missing events from the posterior $P(\mathbf{a}, \mathcal{H}|\mathcal{D})$ for a given set of parameters $\mathbf{a}$, start with a set of hidden events $\mathcal{H}_0$ which are consistent with the observations $\mathcal{D}$. Let $\mathcal{H}_i$ denote the set of hidden events at the ith step and iterate the following procedure:

1. propose $\mathcal{H}_i \to \mathcal{H}'$ with probability $q(\mathcal{H}_i, \mathcal{H}')$;
2. set $\mathcal{H}_{i+1} = \mathcal{H}'$ with probability $\min\left\{1, \frac{P(\mathcal{D},\mathcal{H}'|\mathbf{a})q(\mathcal{H}',\mathcal{H}_i)}{P(\mathcal{D},\mathcal{H}_i|\mathbf{a})q(\mathcal{H}_i,\mathcal{H}')}\right\}$;
3. else $\mathcal{H}_{i+1} = \mathcal{H}_i$.

Note that since the method only makes use of relative values of the posterior $P(\mathbf{a}, \mathcal{H}|\mathcal{D})$, the acceptance probability in step 2 is straightforward to calculate as the ratio of likelihoods of complete events (see Eq. (2)) multiplied by a ratio of proposal probabilities.

The proposal probabilities allow the exploration of the space of possible hidden events. In theory $q()$ can be any distribution, e.g., uniform, however selection of the proposal distribution determines how well the chain mixes, and thus convergence time (the number of samples that must be discarded as burn-in). Therefore, in practice, it is often sensible to use q's related to the application. In our examples (see Section 3), the proposal probability ratio turns out to be equal to one but for more complicated changes to a reconstructed realization this may not be the case. For example, in Marion et al. (in preparation), changes to a realization of a herbivore grazing model includes adding or removing off-take events. Specifically, in that implementation of the Markov chain we add an off-take event to a realization of M such events, with probability $q(\mathcal{H}_i, \mathcal{H}') = 1/(t_f - t_0)$, i.e., the probability of inserting an event within a specified time interval. The reverse move is then the deletion of the event just inserted. This is from the realization now containing $M + 1$ events, and thus occurs with probability $q(\mathcal{H}', \mathcal{H}_i) = 1/(M + 1)$.

The above procedure implements a Markov chain indexed by i and it can be shown that as $i \to \infty$ the distribution of states visited by the chain is independent of i. Moreover, this distribution is in fact the desired posterior distribution (Metropolis et al., 1953; Hastings, 1970; Green, 1995). The algorithm therefore provides a way of reconstructing complete realizations of the process consistent with the data and the physical model. For this reason, the case of incomplete data sets can be catered for as plausible and consistent values of the missing data are proposed and the associated uncertainty is accounted for. Moreover, we can use a form of Gibbs sampling to obtain samples of the parameters to be estimated for each complete, or reconstructed, realization as described below.

2.2. *Gibbs samplers*

Conventional Gibbs sampling can be achieved with the help of Bayes rule. If the prior distribution $P(\mathbf{a})$ is chosen to be conjugate to the likelihood then the posterior will be a known distribution. If the parameter vector $\mathbf{a} = (a_1, a_2, \ldots, a_k) \in R^k$, then the Gibbs sampler updates $\mathbf{a}$ component by component. That is, for a given (reconstructed) complete realization of events $\mathcal{E} = (\mathcal{D}, \mathcal{H})$:

1. assign values to $\mathbf{a}^{(0)}$ (this can be drawn from the prior distribution but this is not necessary);
2. set $i = 0$;
3. repeat
   - draw $a_1^{(i+1)}$ from $P(a_1|\mathcal{E}, a_2^{(i)}, \ldots, a_k^{(i)}) \propto L(a_1, \mathcal{E}, a_2^{(i)}, \ldots, a_k^{(i)})P(a_1)$;
   - draw $a_2^{(i+1)}$ from $P(a_2|\mathcal{E}, a_1^{(i+1)}, a_3^{(i)}, \ldots, a_k^{(i)})$;
   - ...
   - draw $a_k^{(i+1)}$ from $P(a_k|\mathcal{E}, a_1^{(i+1)}, \ldots, a_{k-1}^{(i+1)})$;
   - set $i = i + 1$;
   - store every mth value of $\mathbf{a}$ after an initial burn-in period;
4. end repeat.

Note that although successive draws from the chain are correlated, it is common practice to use $m = 1$ as this avoids throwing away samples. We have presented the above update procedure element-by-element. This is not always required and a block update whereby all ($\mathbf{a}$), or a subset ($\tilde{\mathbf{a}}$), of parameters are updated simultaneously is often more efficient, in which case $\tilde{\mathbf{a}}^{(i+1)}$ is drawn from $P(\tilde{\mathbf{a}}^{(i)}|\mathcal{E})$.

This method of sampling is particularly well suited to estimating parameters which appear linearly within model process rate probabilities since selection of conjugate priors can then be straightforward. As we shall see in Section 3 for the time-homogeneous Markov process models described earlier gamma priors for linear rate parameters imply that the marginal posteriors (conditional on the other parameters and the missing events) for each parameter are also gamma distributions. If a parameter appears non-linearly then it may be non-trivial, or not possible, to select a prior resulting in a posterior which can be written in the form of a known distribution.

2.3. *Griddy-Gibbs sampler*

It is not always possible to choose a conjugate prior distribution, or the product (likelihood × prior) cannot easily be written in the form of a known distribution. If, however, the possible parameter values are constrained to take values on a finite lattice of points within the parameter space Q, then for a given realization $\mathcal{E}$, we can approximate,

$$P(\mathbf{a}|\mathcal{E}) \approx \frac{L(\mathbf{a}, \mathcal{E})}{\sum_{\mathbf{a}' \in Q} L(\mathbf{a}', \mathcal{E})}. \tag{5}$$

This approximation is coined "griddy-Gibbs" (Tanner, 1996) and the prior in this case is discrete and uniform over a number of grid points. The griddy-Gibbs sampler is similar to the conventional Gibbs sampler. It can perform block updates or be carried out element-by-element.

1. Assign values to $\mathbf{a}^{(0)}$ drawn uniformly from the grid points;
2. set $i = 0$;
3. repeat

- draw $a_1^{(i+1)}$ from
  $P(a_1|\mathcal{E}, a_2^{(i)}, \ldots, a_k^{(i)}) \approx \frac{L(a_1, \mathcal{E}, a_2^{(i)}, \ldots, a_k^{(i)})}{\sum_{a_1' \in Q} L(a_1', \mathcal{E}, a_2^{(i)}, \ldots, a_k^{(i)})}$;
- draw $a_2^{(i+1)}$ from $P(a_2|\mathcal{E}, a_1^{(i+1)}, a_3^{(i)}, \ldots, a_k^{(i)})$;
- ...
- draw $a_k^{(i+1)}$ from $P(a_k|\mathcal{E}, a_1^{(i+1)}, \ldots, a_{k-1}^{(i+1)})$;
- set $i = i + 1$;
- store every mth value of $\mathbf{a}$ after an initial burn-in period;

4. end repeat.

A nice consequence of the griddy algorithm is that the entire equilibrium distribution can be approximated by Eq. (5) and not solely estimated from the sampled parameters.

2.4. *Metropolis–Hastings parameter sampler*

A third method of parameter sampling is the Metropolis–Hastings algorithm. We introduced a more complicated reversible-jump version to explore the space of hidden events in Section 2.1 but it can also be usefully used for parameter estimation. In this case, the proposal probabilities can be chosen as (continuous) uniform distributions rather than a discrete prior in the griddy sampler. As a result, the acceptance probability is a ratio of two likelihood calculations which is a further advantage over the griddy method especially in examples where parameters, or blocks of parameters, are such that the Markov chain converges quickly to the equilibrium distribution. The Metropolis–Hastings algorithm for parameter sampling is given by,

1. assign values to $\mathbf{a}^{(0)}$;
2. set $i = 0$;
3. repeat
   - uniformly draw $\mathbf{a}'$;
   - set $\mathbf{a}^{(i+1)} = \mathbf{a}'$ with acceptance probability $\min\{1, \frac{L(\mathbf{a}', \mathcal{E})}{L(\mathbf{a}^{(i)}, \mathcal{E})}\}$;
   - else $\mathbf{a}^{(i+1)} = \mathbf{a}^{(i)}$;
   - set $i = i + 1$;
   - store every mth value of $\mathbf{a}$ after an initial burn-in period;
4. end repeat.

The above algorithm can similarly be modified to perform element-by-element updates. The uniform proposal distribution for the parameters may be inefficient and can be replaced, for example, with a sample centred on the current value. The acceptance probability in step 3 must then be modified to account for the fact the forward and reverse moves may not have equal probability. The above algorithm can also be modified to account for non-uniform priors in which case the acceptance probability becomes a ratio of likelihood $\times$ prior.

2.5. *Hybrid sampler*

The above algorithms can be combined to produce a very flexible and powerful hybrid estimation algorithm. Let $\mathbf{a}$ denote those parameters which can be estimated using the conventional Gibbs sampler, let $\mathbf{b}$ denote those parameters which can otherwise be estimated using the griddy-Gibbs sampler, and let $\mathbf{c}$ denote those parameters estimated with a Metropolis–Hastings sampler. To approximate the integral in Eq. (4), start with a realization $\mathcal{E}_0 = (\mathcal{D}, \mathcal{H}_0)$ consistent with the data, and iterate the following procedure:

1. assign values to $\mathbf{a}^{(0)}$, $\mathbf{b}^{(0)}$ and $\mathbf{c}^{(0)}$;
2. set $i = 0$;
3. repeat
   - draw $\mathbf{a}^{(i+1)}$ using the conventional Gibbs sampler with $\mathbf{b}^{(i)}$, $\mathbf{c}^{(i)}$ and $\mathcal{H}_i$;
   - draw $\mathbf{b}^{(i+1)}$ using the griddy-Gibbs sampler with $\mathbf{a}^{(i+1)}$, $\mathbf{c}^{(i)}$ and $\mathcal{H}_i$;
   - sample $\mathbf{c}^{(i+1)}$ using Metropolis–Hastings with $\mathbf{a}^{(i+1)}$, $\mathbf{b}^{(i+1)}$ and $\mathcal{H}_i$;
   - update the realization to $\mathcal{H}_{i+1}$ using the reversible-jump Metropolis–Hastings algorithm with $\mathbf{a}^{(i+1)}$, $\mathbf{b}^{(i+1)}$ and $\mathbf{c}^{(i+1)}$;
   - set $i = i + 1$;
   - store every mth value of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ after an initial burn-in period;
   - if required, store $\mathcal{H}_i$;
4. end repeat.

It is often beneficial to update the reconstructed realization many times using the reversible-jump algorithm before drawing new parameter samples. In this way, the space of hidden events can perhaps be explored more fully thus aiding convergence of the Markov chain.

The above hybrid algorithm is extremely flexible; the reversible-jump step deals with incomplete data and uncertainty in the observations; the griddy-Gibbs and Metropolis–Hastings steps allow for more complex models and parameterizations which do not lend themselves to conjugate prior selection; and the benefit of conventional Gibbs sampling is retained for those parameters which do lend themselves to conjugate prior specification. We remark that both griddy-Gibbs and Metropolis–Hastings can also be used to estimate all parameters and not just those where conjugate prior selection is difficult.

Ecological model parameters are often believed to lie in some range either determined from experiment, or by collating previous estimates from literature searches. For example, the vegetation dynamics model VegeTate (Birch et al., 2000; Birch, 2002) contains many parameters with ranges specified in this way. Despite the lengthier computer run time due to repeated likelihood calculations over a grid – the likelihood must be recalculated at every grid point for each iteration – the ecological researcher may prefer the "griddy" method as the uniform prior is perhaps a less subjective choice of prior than in the conventional method. Of course, these Bayesian methods have the advantage of allowing subjectiveness if there are strong reasons for doing so.

The hybrid procedure implements a Markov chain (indexed by i) which (asymptotically as $i \to \infty$) generates samples from the posterior distribution $P(\mathbf{a}, \mathcal{H}|\mathcal{D})$. Since we must draw the samples from the equilibrium distribution of the Markov chain, a key problem is knowing how long to let the Markov chain run before using the sample. There are a number of convergence diagnostics available (Gilks et al., 1996) but by far the most common is visual inspection of the chain output to

check for trends in the mean or variability of the samples. This is the method we rely on and we monitor time series of the parameter samples obtained by the chain after an initial burn-in.

3. Examples

Vegetation dynamics are of paramount importance in most terrestrial ecosystems, because of the close link existing between primary producers and higher trophic levels. Herbivorous mammals play a decisive role in the structure and dynamics of the vegetation, the number of animals, their habitat use, diet choice and interactions between animal species being the key factors that model vegetation (Osem et al., 2002) other than fire and human intervention. Foraging models of these systems are of special interest to ecologists because the nested levels of different scales at which they operate, from plants or parts of plants at the bite size scale to geographic scale due to, for example, migration movements of grazing ungulates (McNaughton, 1991; Farnsworth et al., 2002).

We describe the applicability of McMC using two examples of foraging behaviour at two contrasting spatial scales, the first one at the scale of bite size and the second one at the scale of an experimental area.

3.1. *Grazing simulation model*

In this example, we consider a simple non-spatial grazing system. More complex stochastic models of agricultural grazing systems have been introduced by (Marion et al., 2005) and discussed in the context of statistical inference by (Marion et al., in preparation). The model here abstractly describes an animal grazing in a paddock/cell of initial sward height $h(0)$ at rate β and is summarized as:

$r(h \to h + \delta h)$	δh	Description
βh	-1	off-take

(6)

The state space is given by the sward height at time t_n, i.e., $s(t_n) = h(t_n)$. Given information from measurements of units of the sward height we will demonstrate how the bite (off-take) rate β can be estimated using the sampling algorithms. This example is simple enough to allow detailed study of the algorithms performance.

We simulate a realization of the model using $\beta = 1.0$ and the initial sward height set to $h(0) = 20$ units. The simulation ends when the cell is grazed to zero height, i.e., 20 (bite) events have taken place. In the first instance, we shall observe *all* 20 events so that the sampling algorithms can be compared with complete data. We will then consider an incomplete data case where only the initial height and final height and times are observed. In this case, a reversible-jump step is needed to reconstruct consistent plausible realizations.

The likelihood given in Eq. (2) is represented by an exponential distribution. We know the bite rate must satisfy $\beta > 0$ so by choosing the prior distribution to be a gamma distribution $P(\beta) = Ga(a, b)$ we can show the conditional distribution $P(\beta|\mathcal{E}) = Ga(a^*, b^*)$ since the gamma distribution and exponential distributions are conjugate. Explicitly, we have,

$$L(\beta, \mathcal{E}) = \prod_{i=1}^{n} \beta h_{t_i} \, e^{-(t_i - t_{i-1})\beta h_{t_i}} = \Big[\prod_{i=1}^{n} h_{t_i}\Big] \beta^n \, e^{-\beta \sum_{i=1}^{n} \delta t_i h_{t_i}}$$

The gamma prior distribution, $Ga(a, b)$, takes the form [Lee (2004)]:

$$P(\beta) \propto \beta^{a-1} \, e^{-b\beta}.$$

The posterior distribution of β conditional on the complete history $\mathcal{E}$ is given by $P(\beta|\mathcal{E}) \propto L(\beta, \mathcal{E})P(\beta)$ and hence

$$P(\beta|\mathcal{E}) \propto \beta^{n+a-1} \, e^{-\beta(\sum_{i=1}^{n} \delta t_i h_{t_i} + b)}$$

which is *also* a gamma distribution, $Ga(a^*, b^*)$, with $a^* = n + a$ and $b^* = \sum_{i=1}^{n} \delta t_i h_{t_i} + b$. We can compare and contrast the performance of different priors (i.e., the parameters a and b) for the Gibbs sampler with different parameter grids in the griddy-Gibbs sampler.

For a complete realization, a^* and b^* are fixed once the prior is chosen. In Fig. 1, we show the effects of choosing different priors on the posterior. Similarly, the number of grid points in the griddy sampler determines the accuracy of the approximation to the posterior distribution. In Fig. 2, we show the effects of different numbers of equally spaced grid points across the range $\beta \in (10^{-4}, 2)$.

Examining Fig. 1, we see that in this example, due to the small number of events (data), the choice of gamma prior distribution parameters appears to have some influence on the posterior distribution. We notice that as the prior moves left the (estimated) posterior appears to shift to the right. This is a feature of gamma priors, most commonly discussed in introductory Bayesian texts in the context of prior specification of precision parameters. A full discussion could become technical very quickly but one can see that as the prior moves left its weight and slope (almost flat) where the posterior lies becomes less influential. We can further see this effect by examining the expressions for a^* and b^* given above. The prior moving left coincides with a and b becoming smaller meaning the data or likelihood having more influence on the posterior. This effect of prior influence would lessen as the number of data increases.

The number of grid points does have an effect on the quality of the approximated distributions in Fig. 2. Comparing Figs. 1 and 2, we see that for more refined parameter grids, for example, greater than 50 grid points, the griddy-Gibbs approximation achieves comparable results to the conventional Gibbs sampling approach. In Fig. 3, we show the results of an application of the Metropolis–Hastings algorithm using 10,000 samples. We see that the estimated distribution is similar to the conventional methods and the griddy-Gibbs sampling method with a fine grid.

The reversible-jump algorithm is designed to cope with incomplete or partially specified data information. It is highly unlikely that an experimentalist, or observer, would have the time or resources to record the time of all bite events. A more realistic situation would involve measurements of the sward

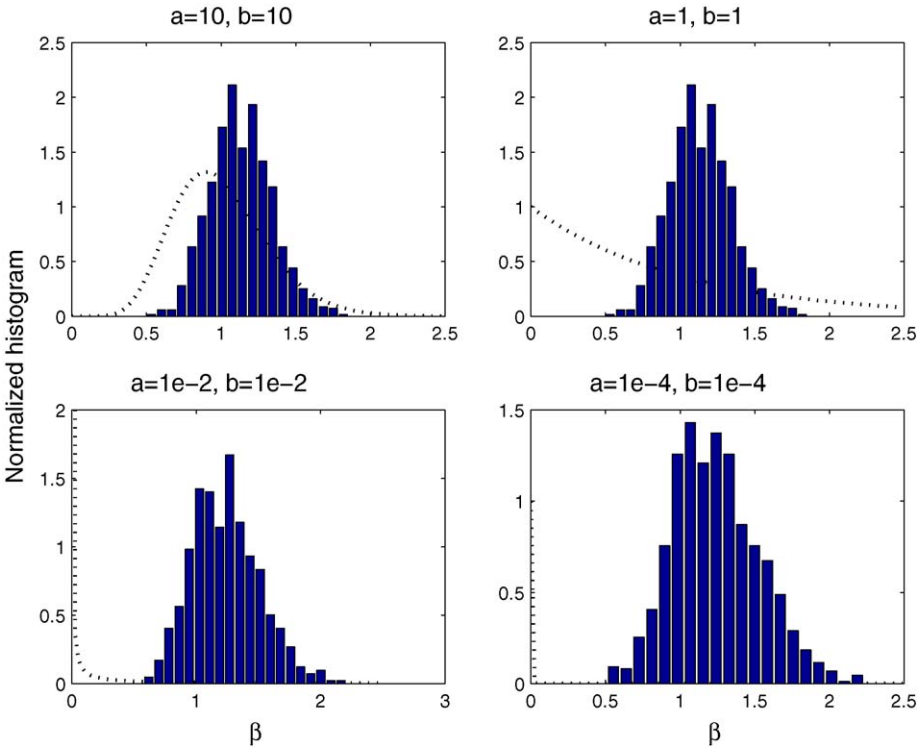


Fig. 1 – Normalized histograms from the parameter samples. The prior distributions for different gamma prior parameters and complete data are indicated by the dotted lines (see text for details).

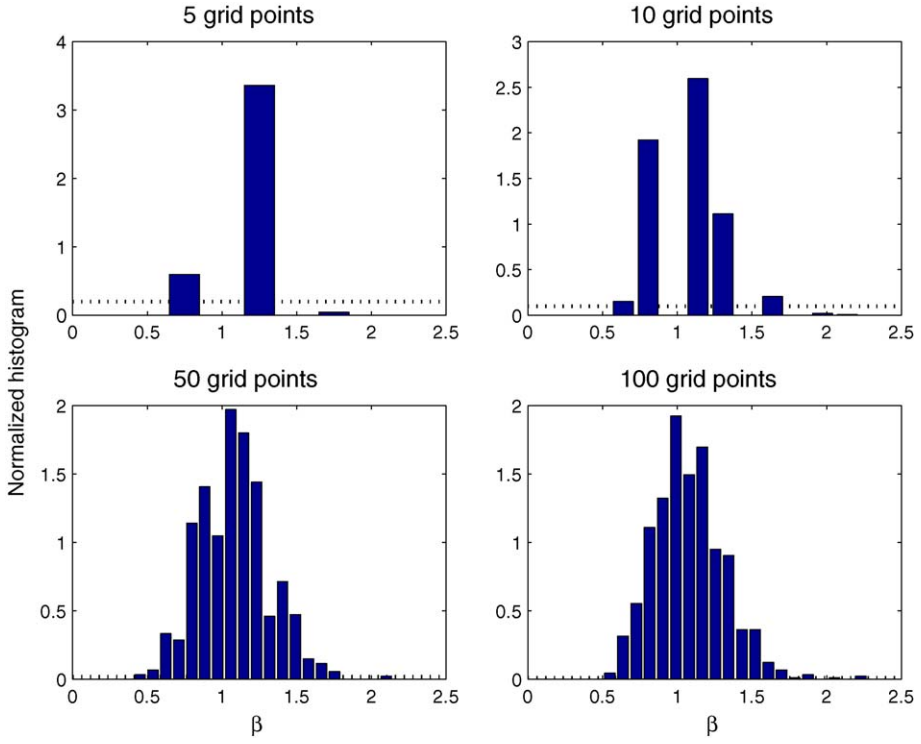


Fig. 2 – Normalized histograms to estimate the posterior distributions for different numbers of parameter grid points and complete data (see text for details). The uniform (discrete) priors are indicated by dotted lines.

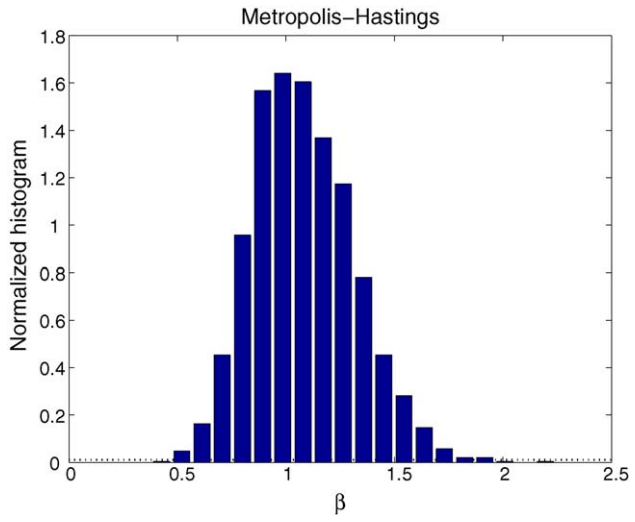


Fig. 3 – Normalized histograms from parameter samples using the Metropolis–Hastings algorithm. The prior is indicated by dotted line.

height at select times, for example, the beginning and end of each working day. A reversible-jump algorithm can be used to reconstruct realizations consistent with these observations. Three basic changes to a realization can be proposed: add a new bite event, remove an existing but unobserved event and rearrange the time of an existing event. In Marion et al. (in preparation), we develop techniques to cope with such a situation.

For the purposes of this paper, however, we will consider a simplified version of the reversible-jump realization reconstruction step. We will consider measurements of the initial sward height and the time of the final bite event, t_{20}, when the paddock is fully grazed to zero height. Since the off-take is fixed at 1 unit we know that 19 other bite events took place which we impute into an initial consistent realization at random times within $(0, t_{20})$. Therefore, the rearrange step of the reversible-jump algorithm is the only step required which has a proposal probability ratio of 1.0.

In Figs. 4 and 5, we show the results of the hybrid sampler using conventional Gibbs and griddy-Gibbs methods for parameter sampling in conjunction with the reversible-jump step. Each simulation was run for 1000 rearrange steps and sample draws. The first 600 steps were discarded as an initial burn-in, and so only the final 400 samples are used in the shown distributions. In Fig. 4, we show the estimated distribution obtained by normalizing histograms of the sampled parameters. Once again we see that the choice of gamma prior distribution parameters has some influence on the posterior distributions due to the small number of data. In Fig. 5, we show the equivalent griddy-Gibbs results. The only difference from the complete realization example above was the need to specify a wider range for the parameter grid $\beta \in (10^{-4}, 4)$ to account for the greater uncertainty of unknown event times. A 50-point parameter grid again appears to compare favourably with the conventional Gibbs sampling approach. We found similar results (not shown) for the Metropolis–Hastings sampling scheme.

3.2. *Animal movement and food selection*

We now introduce a behavioural model of animal movement based on observations of a food selection experiment involving Soay sheep (*Ovis aries*). The experiment was carried out at the Macaulay Institute's Glensaugh Research Station and was principally designed to investigate effects of food handling, sex, body and mouth size of sheep on intake and forage choices when available food is sparsely distributed.

The experimental setup consisted of 36 pegboards arranged in a 6×6 grid with 2 m between boards and walls (see Fig. 6). The pegboard is a device that simulates a complex environment in which food is sparsely distributed at the scale of the feeding station, for example, small plants on stony ground or fallen fruits and seeds. The wooden pegs surround a grass (food) pellet which can be eaten by the animal with different levels of difficulty according to the spacing of the pegs. The use of the pegboard to investigate behavioural differences in feeding Soay sheep is discussed extensively in Pérez-Barbería et al. (2004).

An animal was released alone into the pegboard arena and an observer recorded the order of pegboards visited, the transit time between pegboards, the time spent at a pegboard, and the number of pellets eaten after each pegboard visit was noted. Nine male and nine female sheep were used in the experiment.

There are two main events taking place when an animal enters the arena, these are transit events where the animal moves from pegboard to pegboard, and feeding (bite) events when the animal is stopped at a pegboard and tries to eat (independently whether the animal succeeds or not). When an animal encounters a pegboard, it will typically eat all of the pellets before moving on. If we denote the number of pellets on pegboard i by p_i, then a reasonable event rate for a feeding event in the time interval $(t, t + \delta t)$ is

$$P(\text{bite at } i) = a_i \beta (1 + p_i)\, \delta t \tag{7}$$

where a_i denotes the animals location, $a_i = 1$ if the animal is at pegboard i and $a_i = 0$ otherwise. If the pegboard has already been visited and the pellets supply was depleted, the animal appears to spend some time searching the pegboard for pellets before moving on. So, a feeding event also incorporates a visit to a pegboard if pellets are depleted. Hence, we use $(1 + p_i)$ instead of just p_i.

Animals in the arena typically move to the closest pegboard from their current location, preferring to move to adjacent boards rather than to boards along the diagonal. There are occasions, however, when the animals move to pegboards which are not near-neighbours. The animals also do not appear to make a distinction between full and empty pegboards from a distance so on occasions they tend to return to a previously eaten pegboard. A model of movement which seems appropriate is one that has been used in stochastic epidemiological models (Gibson, 1997), namely,

$$P(\text{move } i \text{ to } j, i \neq j) = a_i Z_{ij}^{-1/2\gamma} \delta t \tag{8}$$

where Z_{ij} denotes the distance from pegboard i to pegboard j. (We use the city-block or Manhatten norm in our experiments but Euclidean norms could equally well be used.)

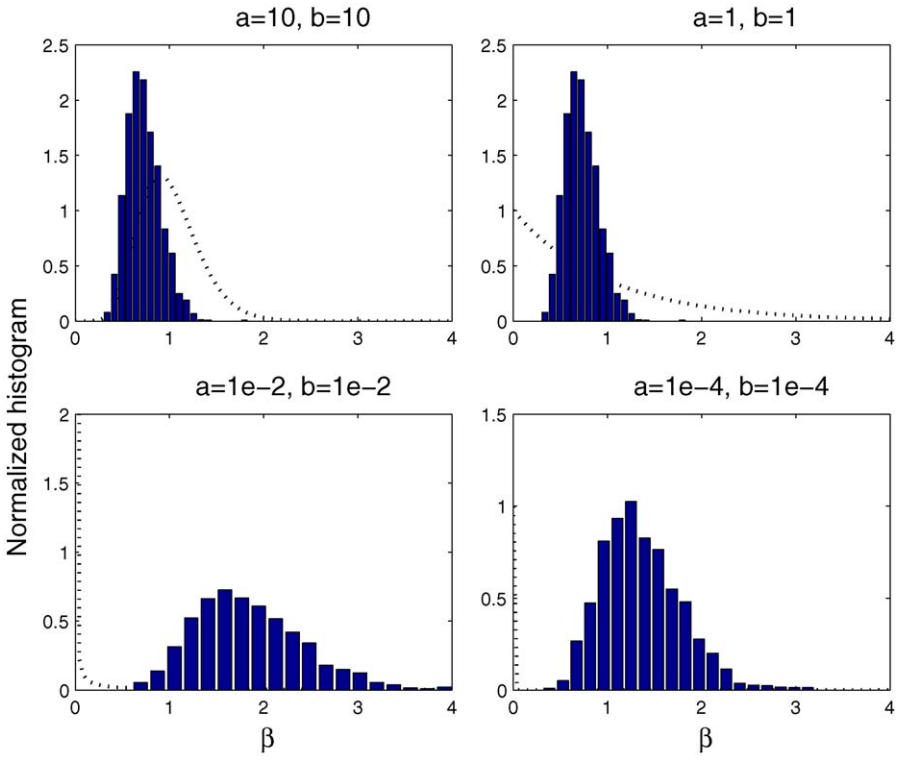


Fig. 4 – Normalized histograms using the parameter samples to estimate the posterior distributions for different gamma prior parameters in the reversible-jump algorithm. The prior is indicated by the dotted lines.

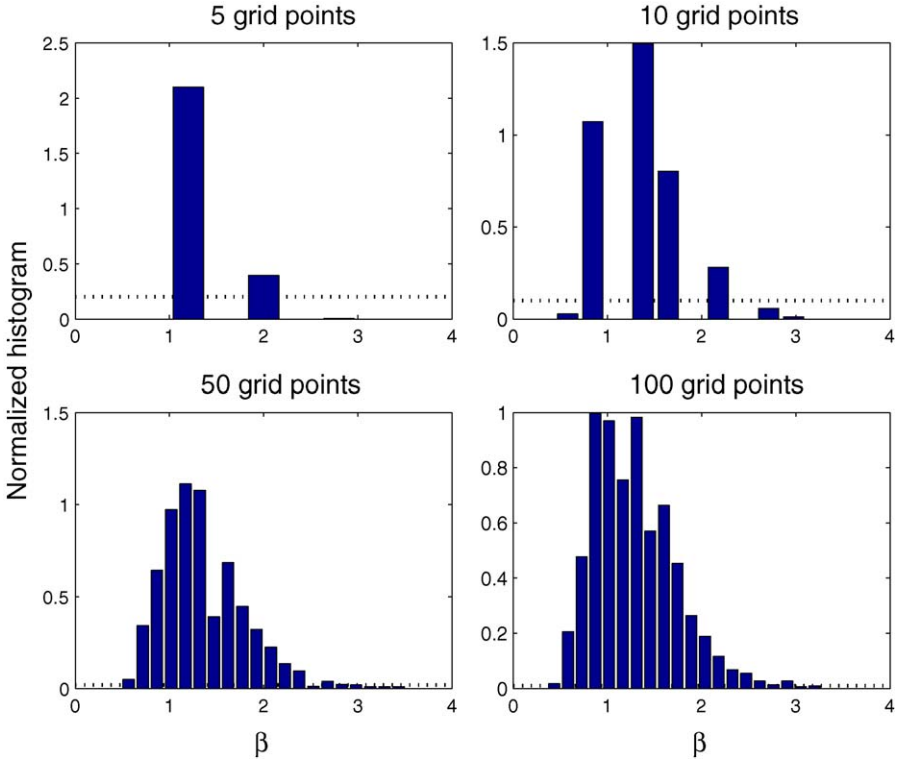


Fig. 5 – Normalized histograms estimating the posterior distributions for different numbers of parameter grid points in the reversible-jump algorithm. The (discrete) prior is indicated by the dotted line.

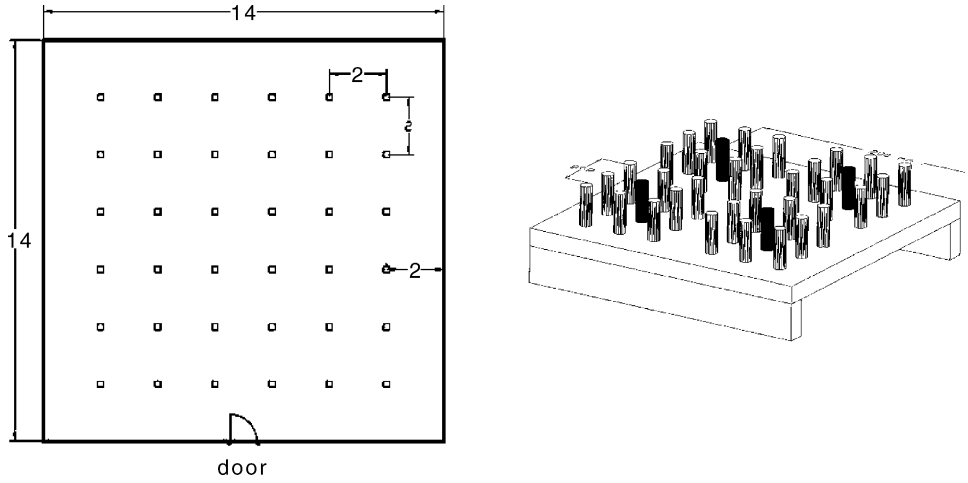


Fig. 6 – The left panel shows a schematic of the experimental arena. An animal enters the arena through the door and its path, transit time and visit time at the pegboards (squares) are recorded. The number of pellets on each pegboard at the start and end of the experiment are also recorded. The right panel is a schematic of the pegboard. The dark dowells indicate the grass pellets surrounded by wooden dowells.

The two parameters to be estimated from the measured data are β, an "off-take rate", or "pellet depletion rate" parameter, and γ, a movement rate parameter. The model can be summarized as follows:

$r(s \to s + \delta s)$	δp_i	δa_i	δa_j	Description
$a_i\beta(1+p_i)\delta t$	If $p_i > 0$, $\delta p_i = -1$	0	0	Pellet eaten at pegboard i
$a_i Z_{ij}^{-1/2\gamma}\delta t$	0	-1	$+1$	Animal moves i to j $(i \neq j)$

The above model requires a full implementation of the hybrid sampler. The off-take rate parameter can be estimated using conventional Gibbs sampling. The likelihood is of the form of Eq. (2), so using a Gamma prior $P(\beta) = Ga(a, b)$, we can show that

$$p(\beta|D) \propto L(\beta, D)p(\beta) = Ga(a^*, b^*)$$

with $a^* = N_B + a$ and $b^* = \sum_{i=1}^{N}(1+p_i)\delta t_i + b$, where N_B denotes the number of feeding events within the N event realization. The movement parameter requires an implementation of the griddy-Gibbs sampler as a conjugate prior is difficult to find or does not exist. A 200-point grid over the range $\gamma = (0.05, 0.15)$ was used. A reversible-jump step is necessary to average over the uncertainty of unknown feeding

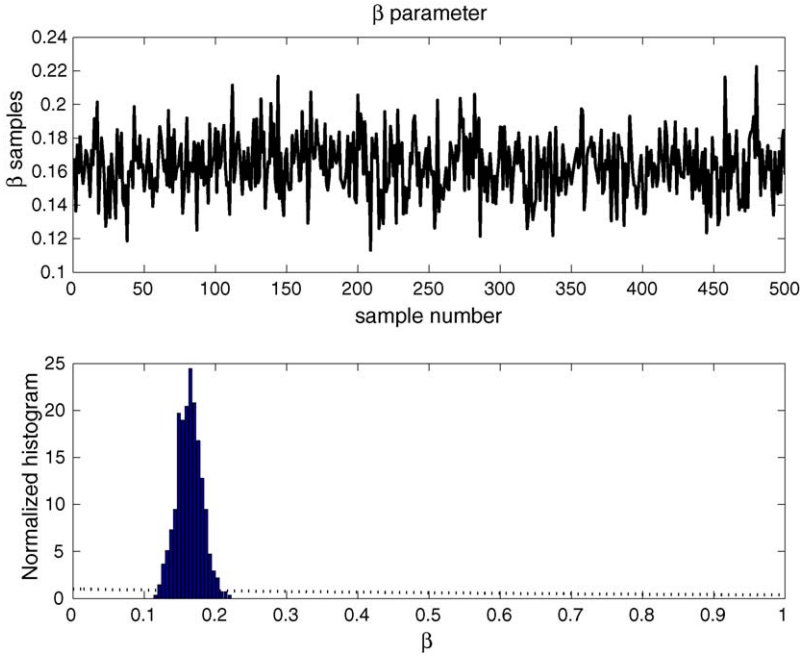


Fig. 7 – The upper plot shows the parameter samples as the algorithm progresses. The lower plot shows the normalized histogram of the samples. The prior distribution is shown as a dotted line.

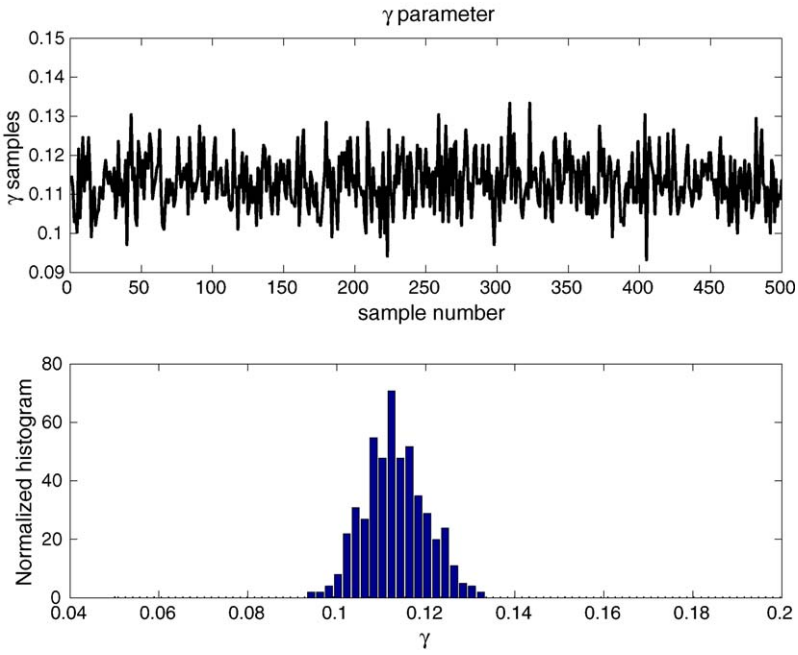


Fig. 8 – The upper plot shows the parameter samples as the algorithm progresses. The lower plot shows the normalized histogram of the parameter samples. The prior distribution is shown as a dotted line.

event times. The total number of pellets eaten is recorded for each pegboard visit together with the visitation time. A complete realization of the model requires the off-take times for each individual pellet. Since these are not recorded the unknown event times must be imputed within the relevant time intervals and then averaged over using the reversible-jump rearrange step.

For each trial, we implement the hybrid sampler and record 500 samples. The parameters for the β gamma prior were set to $a = 1$ and $b = 1$ for each trial. In Fig. 7, we show the results of analyzing the data for one trial of a female sheep marked red–black. The sample draws of β as the algorithm progresses are shown in the upper plot while the normalized histogram of the sampled values are shown alongside the prior in the lower plot. Fig. 8 shows similar plots for the movement parameter γ obtained by griddy-Gibbs sampling. The Metropolis–Hastings scheme produces similar results (not shown). All trials and animals displayed similar results. For example, Figs. 9 and 10 display the results of all five trials of the red–black female sheep. Despite there being some variability across trials the resulting (estimated) distributions exhibit consistent results.

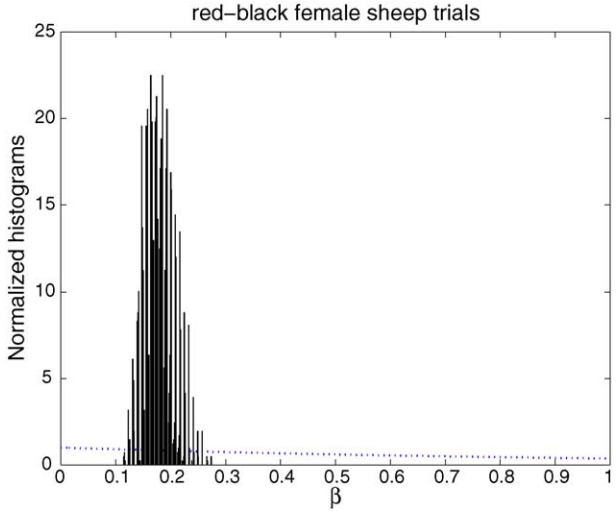


Fig. 9 – The estimated distributions of the β parameter using normalized histograms for each trial of the female sheep with red–black markings. The prior is shown by the dotted line.

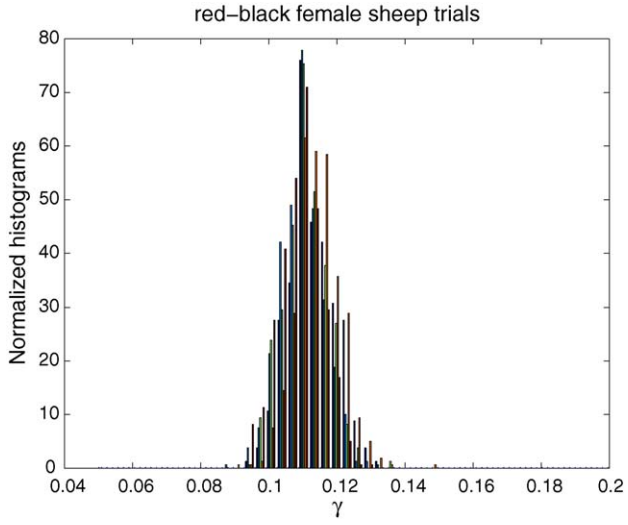


Fig. 10 – The estimated distributions of the γ parameter using normalized histograms for each trial of the female sheep with red–black markings. The prior is shown by the dotted line.

4. Summary

We have shown how different McMC sampling techniques can be combined to form a very flexible, powerful and widely applicable hybrid sampler. We compared the performance of conventional Gibbs sampling, a griddy-Gibbs approximation and a Metropolis–Hastings sampling scheme on a simple grazing system model. The griddy-Gibbs sampler was shown to be a reasonable approximation technique as long as the grid is fine enough. A very fine grid with many grid points may result in long simulation times for complex examples since the likelihood must be recalculated at every grid point for each iteration, and so we recommend using conventional Gibbs sampling where possible. The griddy-Gibbs sampling method comes into its own when it is not possible to select a conjugate prior due to the location of the parameter in the model. This is also true of the Metropolis–Hastings scheme and for examples where convergence to the desired equilibrium distribution is quick this scheme may be the preferred option.

The full power of the hybrid sampler was used with a novel application to an ecological experiment. We proposed a purely behavioural model based on the observations to demonstrate the algorithm. We made no attempt to extract or attribute physiological properties to the results. The results of this example did however highlight the success of the hybrid sampler and demonstrate the potential of using such techniques and models in ecological studies.

There are a number of stochastic process models in the literature whose transition probabilities do have biological and physical significance. For example, in a series of three papers, Berg and Shuman (1995a,b,c) develop a stochastic model of behaviour of radionuclides in forests by modelling soil and vegetation uptakes. Bearlin et al. (1999) introduced a stochastic model for seagrass, the parameters of which could be estimated using the sampling techniques outlined in this paper. A number of spatially extended and stage-structured stochastic population dynamic models have been proposed which require the estimation of important parameters (Liu et al., 2000; Yemshanov and Perera, 2002; Castañera et al., 2003; Leung and Grenfell, 2003; Fink and Kofoet, 2005; Watanabe et al., 2005a; Griebeler and Sietz, 2006). The sampling and parameter fitting technology described in this paper could be applied to these models, or sub-components of these models.

Bayesian methods of inference such as Gibbs samplers and Metropolis–Hastings sampling schemes have found a place in the estimation of parameters in ecological models. Indeed a body of work is growing which use McMC to improve and enhance ecological models (Harmon and Challenor, 1997; Dowd and Meyer, 2003; Qian et al., 2003; Rivot et al., 2004; Malve et al., 2005; Watanabe et al., 2005b).

We hope that the techniques we have outlined here can find wide application to the modelling of ecologically interesting mechanisms and to areas where incomplete data has been a hindrance to development.

Acknowledgments

We thank Iain J. Gordon and Noemie Dufourmantelle for their contributions and involvement during the design and data gathering aspect of the arena experiments. BioSS and MLURI gratefully acknowledge the financial support of the Scottish Executive Environment and Rural Affairs Department (SEERAD). The authors would also like to thank two anonymous referees whose comments helped to improve the manuscript.

REFERENCES

Bearlin, A.R., Burgman, M.A., Regan, H.M., 1999. A stochastic model for seagrass (*Zostera muelleri*) in Port Phillip Bay, Victoria, Australia. Ecol. Modell. 118, 131–148.

Berg, M.T., Shuman, L.J., 1995a. A three-dimensional stochastic model of the behavior of radionuclides in forests. I. Model structure. Ecol. Modell. 83, 359–372.

Berg, M.T., Shuman, L.J., 1995b. A three-dimensional stochastic model of the behavior of radionuclides in forests. II. Cs-137 behavior in forest soils. Ecol. Modell. 83, 373–386.

Berg, M.T., Shuman, L.J., 1995c. A three-dimensional stochastic model of the behavior of radionuclides in forests. III. Cs-137 uptake and release by vegetation. Ecol. Modell. 83, 387–404.

Besag, J., Green, P.J., 1993. Spatial statistics and Bayesian computation. J. R. Stat. Soc. B 55, 25–37.

Birch, C.P.D., 2002. The influence of position on genet growth: a simulation of a population of bracken (*Pteridium aquilinum* (L.) Kuhn) genets under grazing. Evol. Ecol. 15, 463–483.

Birch, C.P.D., Vuichard, N., Werkman, B.R., 2000. Modelling the effects of patch size on vegetation dynamics: bracken [*Pteridium aquilinum* (L.) Kuhn] under grazing. Ann. Bot. 85B, 63–76.

Bolker, B., Pacala, S.W., 1997. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. Theor. Popul. Biol. 52, 179–197.

Castañera, M.B., Aparicio, J.P., Gurtler, R.E., 2003. A stage-structured stochastic model of the population dynamics of *Triatoma infestans*, the main vector of Chagas disease. Ecol. Modell. 162, 33–53.

Cox, D.R., Miller, H.D., 1965. The Theory of Stochastic Processes. Chapman and Hall, London.

Dowd, M., Meyer, R., 2003. A Bayesian approach to the ecosystem inverse problem. Ecol. Modell. 168, 39–55.

Edwards, A.W.F., 1992. Likelihood. Johns Hopkins University Press, Baltimore.

Farnsworth, K.D., Focard, S., Beecham, J.A., 2002. Grassland–herbivore interactions: how do grazers coexist? Am. Nat. 159, 24–39.

Filipe, J.A.N., Gibson, G.J., 1998. Studying and approximating spatio-temporal models for epidemic spread and control. Phiolos. Trans. R. Soc. Lond. B 353, 2153–2162.

Fink, M., Kofoet, A., 2005. A two-dimensional stochastic model of downy mildew of radish. Ecol. Modell. 181, 139–148.

Gamerman, D., 1997. Markov Chain Monte Carlo. Chapman and Hall, London.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85, 398–409.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman and Hall, CRC.

Gibson, G.J., 1997. Markov chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology. Appl. Stat. 46, 215–233.

Gibson, G.J., Renshaw, E., 1998. Estimating parameters in stochastic compartmental models using Markov chain methods. IMA J. Math. Appl. Med. Biol. 15, 19–40.

Gibson, G.J., Renshaw, E., 2001a. Inference for immigration-death processes with single and paired immigrants. Inverse Probl. 17, 455–466.

Gibson, G.J., Renshaw, E., 2001b. Likelihood estimation for stochastic compartmental models using Markov chain methods. Stat. Comput. 11, 347–358.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall, CRC.
Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.
Griebeler, E.M., Sietz, A., 2006. The use of Markovian metapopulation models: Reducing the dimensionality of transition matrices by self-organizing Kohonen networks. Ecol. Modell. 192, 271–285.
Harmon, R., Challenor, P., 1997. A Markov chain Monte Carlo method for estimation and assimilation into models. Ecol. Modell. 101, 41–59.
Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.
Holmes, P., Brown, E., Moehlis, J., Boyacz, R., Gao, J., Hu, P., Aston-Jones, G., Clayton, E., Rajkowski, J., Cohen, J.D., 2004. Optimal decisions: from neural spikes, through stochastic differential equations, to behavior. Proceedings of the 2004 International Symposium on Nonlinear Theory and its Applications (NOLTA2004), Fukuoka, Japan, pp. 15–18.
Isham, V., 1991. Assessing the variability of stochastic epidemics. Math. Biosci. 107, 209–224.
Keeling, M.J., 2000a. Metapopulation moments: coupling, stochasticity and persistence. J. Anim. Ecol. 69, 725–736.
Keeling, M.J., 2000b. Multiplicative moments and measures of persistence in ecology. J. Theor. Biol. 205, 269–281.
Lee, P.M., 2004. Bayesian Statistics: An Introduction. Arnold, London.
Leung, B., Grenfell, B.T., 2003. A spatial stochastic model simulating a scabies epidemic and coyote population dynamics. Ecol. Modell. 166, 41–52.
Liu, Y.-C., Durrett, R., Milgroom, M.G., 2000. A spatially-structures stochastic model to simulate hetergenous transmission of viruses in fungal populations. Ecol. Modell. 127, 291–308.
Malve, O., Laine, M., Haario, H., 2005. Estimation of winter respiration rates and prediction of oxygen regime in a lake using Bayesian inference. Ecol. Modell. 182, 183–197.
Marion, G., Swain, D.L., Hutchings, M.R., 2005. Understanding foraging behaviour in spatially heterogeneous environments. J. Theor. Biol. 232, 127–142.
Marion, G., Walker, D.M., Swain, D.L., Hutchings, M.R. BioSS technical report, in preparation.
Matis, J.H., Kiffe, T.R., Parthasarathy, P.R., 1998. On the cumulants of population size for the stochastic power law logistic model. Theor. Popul. Biol. 53, 16–29.
McNaughton, S.J., 1991. Evolutionary ecology of large tropical herbivores. In: Price, P.W., Lewinsohn, T.M., Fernandes, G.W., Benson, W.W. (Eds.), Plant–animal interactions: evolutionary ecology in tropical and temperate regions. John Wiley & Sons Inc., New York.
Metropolis, M., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092.
O'Neill, P.D., Roberts, G.O., 1999. Bayesian inference for partially observed stochastic epidemics. J. R. Stat. Soc. A 162, 121–129.
Osem, Y., Perevolotsky, A., Kigel, J., 2002. Grazing effect on diversity of annual plant communities in a semi-arid rangeland: interactions with small-scale spatial and temporal variation in primary productivity. J. Ecol. 90, 936–946.
Pérez-Barbería, F.J., Walker, D.M., Gordon, I.J., 2004. Sex differences in feeding behaviour at feeding station scale in Soay sheep (*Ovis Aries*). Behaviour 141, 999–1020.
Qian, S.S., Stow, C.A., Borsuk, M.E., 2003. On Monte Carlo methods for Bayesian inference. Ecol. Modell. 159, 269–277.
Rand, D.A., 1999. Correlation equations and pair approximations for spatial ecologies. In: McGlade, J. (Ed.), Advanced Ecological Theory: Principles and Applications. Blackwell Science, Oxford.
Renshaw, E., 1991. Modelling Biological Populations in Space and Time. Cambridge University Press.
Renshaw, E., Gibson, G.J., 1998. Can Markov chain Monte Carlo be usefully applied to stochastic processes with hidden birth times? Inverse Probl. 14, 1581–1606.
Rivot, E., Prévost, E., Parent, E., Baglinière, J.L., 2004. A Bayesian state–space modelling framework for fitting a salmon stage-structured population dynamics model to multiple time series of field data. Ecol. Modell. 179, 463–485.
Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J. R. Stat. Soc. B 55, 3–23.
Tanner, M.A., 1996. Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions, third ed. Springer.
Van Kampen, N.G., 1992. Stochastic Processes in Physics and Chemistry. North Holland, Amsterdam.
Watanabe, K., Yoshimura, C., Omura, T., 2005a. Stochastic model for recovery prediction of macroinvertebrates following a pulse-disturbance in river. Ecol. Modell. 189, 396–412.
Watanabe, K.H., Lin, H.-I., Bart Jr., H.L., Martinat, P., Means, J.C., Kunas, M.L., Grimm, D.A., 2005b. Bayesian estimation of kinetic rate constants in a food-web model of polycyclic aromatic hydrocarbon bioaccumulation. Ecol. Modell. 181, 229–246.
Whittle, P., 1957. On the use of the normal approximation in the treatment of stochastic processes. J. R. Stat. Soc. B 19, 268–281.
Wilson, H.B., Hassell, M.P., 1997. Host-parasitoid spatial models: the interplay of demographic stochasticity and dynamics. Proc. R. Soc. Lond. B 264, 1189–1195.
Yemshanov, D., Perera, A.H., 2002. A spatially explicit stochastic model to simulate boreal forest cover transitions: general structure and properties. Ecol. Modell. 150, 189–209.