Bernoulli **9**(3), 2003, 373–393

Model selection for hidden Markov models of ion channel data by reversible jump Markov chain Monte Carlo

MATHISCA C.M. DE GUNST[1] and BARRY SCHOUTEN[2]

[1]*Department of Mathematics, Faculty of Exact Sciences, Free University, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands. E-mail: degunst@cs.vu.nl*
[2]*Statistics Netherlands, Prinses Beatrixlaan 428, 2273 XZ Voorburg, The Netherlands. E-mail: bstn@cbs.nl*

Ion channels are proteins that are located in the membranes of cells and are capable of conducting ions through the membrane. An ion channel is not always 'open' for transport. The ion channel molecule may reside in several configurations, some of which correspond to an open channel and others to a closed channel. The transitions of the channel between the different configurational states have a random nature. Markov processes are often used to describe this randomness. In practice, there often exist a number of candidate Markov models. The objective of this paper is the selection of a Markov model from a finite collection of such models. We propose a Bayesian setting in which the model indicator itself is viewed as a random variable, and we develop a reversible jump Markov chain Monte Carlo (MCMC) algorithm in order to generate a sample from the posterior distribution of the model indicator given the data of a single-channel recording. A hidden Markov model is used to incorporate the correlated noise in recordings and the effects of filters that are present in the experimental set-up. The reversible jump MCMC sampler is applied to both simulated and recorded data sets.

Keywords: Markov chain Monte Carlo; maximum a posteriori estimator; model identification; posterior distribution; single-channel recordings

1. Introduction

Ion channels are proteins that are located in the membranes of cells and are capable of conducting ions through the membrane. An ion channel is not always 'open' for transport. The ion channel molecule may reside in several configurations, some of which correspond to an open channel and others to a closed channel. The transitions of the channel between the different configurational states are of a random nature. Markov processes are often used to describe this randomness. The passage of ions through the membrane implies a small electric current that can be measured. Noise and filtering, however, usually have a serious impact on the resulting recordings. Therefore, hidden Markov models are used to incorporate these effects.

De Gunst *et al.* (2001) dealt with Bayesian inference from ion channel data using hidden Markov models. Here we are concerned with the problem of model identification in that

1350–7265

setting, and the objective is to select a Markov model for the gating mechanism of the ion channel from a set of candidate Markov models. For hidden Markov models the number of states of the underlying Markov chain is one of the model components that need to be identified. When all states play the same role, this number determines the model, otherwise the structure of the chain, i.e. the possible transitions between states, needs to be identified too. For ion channel models we are in the latter situation, unless there are only two states.

In a non-Bayesian context model identification is usually based upon criteria that make some correction of the log-likelihood in order to prevent larger models from being chosen. See, for instance, Horn (1987) and Ball and Sansom (1989) in the area of ion channel modelling. Rydén (1995) uses maximum split data estimates for the estimation of the number of states in a hidden Markov model.

The Bayesian setting allows for model selection by considering the model itself as a random variable. The ratio of the posterior densities times the inverse ratio of the prior densities of two different models, the so-called Bayes factor, is often used as a measure to quantify how much one model is to be favoured over another model; see, for instance, Kass and Raftery (1995). In our case, the Bayes factor of the two models cannot be derived analytically. Numerical methods should therefore be used to approximate the Bayes factors.

We will apply Markov chain Monte Carlo (MCMC) techniques to sample from the posterior density of a model. This will enable us to estimate Bayes factors for different sets of models. As different models may have different numbers of parameters and different state spaces for the ion channel states, the usual MCMC methods cannot be used to update the model. Instead, reversible jump MCMC techniques introduced by Green (1995) can be used to match dimensions. Hodgson and Green (1999) have developed a reversible jump sampler to investigate the discrimination of several simple Markov models for ion channel kinetics. Their approach only allows for Markov models with sojourn lengths that are independent, and cannot easily be extended to the non-independent case. Ball *et al.* (1999) used a model that is more general, and used a reversible jump sampler to update the number of sojourns, but did not include model updating in their algorithm. The model used by Hodgson and Green (1999) includes state-independent, correlated noise with a fixed variance. The hidden Markov model that we will use includes correlated noise, state-dependent excess noise and filtering, and allows for more general state spaces.

In Section 2 we will pose the model selection problem, and in Section 3 the reversible jump sampler is constructed and an example of a model selection problem is given. Section 4 motivates our choice of model jump proposal probabilities. In the last two sections some results of the application of the reversible jump sampler to simulated and real data sets are presented and discussed.

2. The model selection problem

We use the hidden Markov model as defined in de Gunst *et al.* (2001). Let $\{X(t)\}_{t\geqslant 0}$ be a continuous-time, irreducible, stationary Markov chain representing the state of the ion channel. Its state space S is finite and is the union of the set of closed states S_c and the set of open states S_o. The sampling interval is Δ, and we let $X_i = X(i\Delta)$ for $0 \leqslant i \leqslant T$.

Furthermore, let $\{Y_t\}_{0\leqslant t\leqslant T}$ denote the observed electrical current and $\{C_t\}_{0\leqslant t\leqslant T}$ be an AR(p) process describing the correlated, state-independent noise. For $(r \vee p) \leqslant t \leqslant T - r$, the model has the form

$$Y_t = \sum_{k=-r}^{r} \gamma_k \mu(X_{t-k}) + C_t + \sigma(X_t)\delta_t, \tag{1}$$

with $C_t = \sum_{i=1}^{p} \phi_i C_{t-i} + \epsilon_t$ and

$$\mu(s) = \begin{cases} \mu_c & \text{if } s \in S_c, \\ \mu_o & \text{if } s \in S_o, \end{cases} \qquad \sigma(s) = \begin{cases} \sigma_c & \text{if } s \in S_c, \\ \sigma_o & \text{if } s \in S_o. \end{cases}$$

The innovations ϵ_t are assumed to be independent and normally distributed with mean zero and variance σ_ϵ^2; $\{\delta_t\}_{0\leqslant t\leqslant T}$ is white standard Gaussian noise. The processes $\{X_t\}_{0\leqslant t\leqslant T}$, $\{\delta_t\}_{0\leqslant t\leqslant T}$ and $\{\epsilon_t\}_{0\leqslant t\leqslant T}$ are assumed independent. The innovations variance of the AR(p) process is σ_ϵ^2.

In the following the probability density of a variable Z in z will simply be denoted by $p(z)$, unless this is not clear from the context. Furthermore, for any sequence $a_1, a_2, \ldots, a_T$, we let a_l^{l+k} denote the vector $(a_l, a_{l+1}, \ldots, a_{l+k})$, $1 \leqslant l \leqslant l+k \leqslant T$. For notational convenience we will use y, x and c instead of $y = y_{(r\vee p)}^{T-r}$, $x = x_{\rho-r}^{T}$ and $c = c_{\rho-p}^{T-r}$, with $\rho = r \vee p$. The vector $q = (q_1, \ldots, q_l)'$ contains the non-zero transition rates, and

$$\xi = (\mu_c, \mu_o, \sigma_c^2, \sigma_o^2, \phi_1, \ldots, \phi_p, \sigma_\epsilon^2)',$$

$$\theta = (\xi, q)'.$$

Besides θ, the state space and the structure of the underlying Markov process are also unknown. The weights $\gamma_{-r}, \ldots, \gamma_r$ are assumed to be known and fixed. Our goal is to select the most likely model from a collection of biologically plausible models and to estimate its unknown parameter values based on observations y. Suppose that model M is to be selected from a collection of $\mathcal{M}$ models, so $M \in \{1, 2, \ldots, \mathcal{M}\}$. The differences between the models are the underlying Markov models that describe the gating mechanism of the ion channel. The number of states and the permitted transitions between states may vary over the different Markov models. For all models the noise and filtering are modelled identically. Also current levels and noise variances are assumed to be the same. For illustration purposes we will discuss a simple example of a model selection problem in Section 3.6.

We choose a non-informative prior distribution for the model indicator M, $p(M = m) = 1/\mathcal{M}$, for $1 \leqslant m \leqslant \mathcal{M}$. The parameter priors $p(\theta|M = m)$ are chosen such that, given model M, all parameters are independent. The transition rates q have gamma priors, the current levels μ and the autoregression coefficients ϕ have normal priors, and the variances σ^2 and σ_ϵ^2 have inverse gamma distributions as priors.

Model m will be chosen so as to maximize the posterior density $p(m|y)$, the so-called maximum a posteriori (MAP) estimate, and Bayes factors are used to summarize the evidence in favour of model m against other models. The Bayes factor in favour of model m_1 against model m_2 is defined by

$$B(m_1, m_2) = \frac{p(m_1|y)}{p(m_2|y)} \frac{p(m_2)}{p(m_1)} = \frac{p(y|m_1)}{p(y|m_2)}. \tag{2}$$

The prior ratio in (2) cancels, since the prior is non-informative, and the Bayes factors can be computed via either likelihoods or posterior distributions. For the hidden Markov model (1) the likelihoods $p(y|m)$ and the posteriors $p(m|y)$ have complicated forms and should be approximated numerically. An alternative is the use of Monte Carlo methods. De Gunst *et al.* (2001) have constructed a Gibbs sampler that generates samples from $p(c, x, \theta|y, m)$ for a particular model m. The samples of two Gibbs sampler runs for two different models may be used to approximate (2) with the weighted likelihood bootstrap method of Newton and Raftery (1994). However, since a Gibbs sampler run is then needed for every model in the set of candidate models, this is not a computationally tractable method of estimating the posterior densities in our case.

In order to be able to compare several models using only a single run of an MCMC sampling method, we will combine the approach described by Green (1995) and the Gibbs sampler to construct a reversible jump sampler that generates a Markov chain $(c, (M, q, x), \xi|y)^i$ with invariant distribution $p(c, (M, q, x), \xi|y)$. The model M, the vector of transition rates q and the ion channel states x are denoted as a triplet because transition rates and channel states only have a meaning in connection with the model. The state space of the Markov process of model m is denoted by $S_m = S_{m,c} \cup S_{m,o}$.

We remark that Kehagias (1996) has developed an algorithm to compute the MAP estimate if the observations Y come from a finite set, and that Frühwirth-Schnatter (1995) has also used Monte Carlo methods to do Bayesian model selection for state space models that, except for the presence of the Markov process $\{X_t\}_{t\in\mathbb{N}_0}$, are generalizations of (1).

3. A reversible jump sampler

3.1. The algorithm

We propose a sampler with blocks c, (m, q, x), x, q and ξ, which thus consists of five updating steps in each iteration. Our reversible jump sampler algorithm has the following form:

1. Construct starting values ξ^0 and (m^0, q^0, x^0) for ξ and (m, q, x). Let $i = 0$.
2. Sample c^{i+1} from $p(c|(m^i, q^i, x^i), \xi^i, y) = p(c|(m^i, x^i), \xi^i, y)$.
3. Sample $(m^{i+1}, \tilde{q}, \tilde{x})$ from $p((m, q, x)|c^{i+1}, \xi^i, y)$.
4. Sample x^{i+1} from $p(x|c^{i+1}, (m^{i+1}, \tilde{q}), \xi^i, y) = p(x|c^{i+1}, (m^{i+1}, \tilde{q}), (\mu, \sigma^2)^i, y)$.
5. Sample q^{i+1} from $p(q|c^{i+1}, (m^{i+1}, x^{i+1}), \xi^i, y) = p(q|(m^{i+1}, x^{i+1}))$.
6. Sample ξ^{i+1} from $p(\xi|c^{i+1}, (m^{i+1}, q^{i+1}, x^{i+1}), y) = p(\xi|c^{i+1}, (m^{i+1}, x^{i+1}), y)$. Return to step 2 with $i := i + 1$.

Steps 1, 2, 4, 5 and 6 are described in detail in de Gunst *et al.* (2001). Steps 4, 5 and 6 may be interchanged. Any model can be used as a starting model m^0. Since the number of models $\mathcal{M}$ is usually not large, convergence to the desired invariant distribution will not be affected too much by this choice.

The only new step is step 3. Since in this step a jump may be made between models having

different dimensions, in the sense that the number of free transition rates and the sizes of the state spaces of the two models may differ, step 3 makes the sampler a jump sampler. We would like to design the sampler so that it satisfies detailed balance and we have convergence to the desired invariant distribution and can apply ergodic theorems (see Gilks *et al.* (1996)). However, due to the possible difference in dimensionality this is not straightforward. As mentioned above, we will follow the technique proposed in Green (1995). First of all, we have to define how to jump between the models, by means of defining proposal probabilities for all model changes. For each model there are in principle $\mathcal{M}$ so-called move types, one for each model transition. The proposal probabilities of the move types are allowed to depend on the present state of the reversible jump sampler Markov chain.

For dimension matching for each pair of models a number of additional random variables are sampled. The number of additional variables equals the difference in the number of free transition rates of the two models. Then two mutually inverse bijections are defined between the sets of transition rates – one in each direction – where the smallest set is augmented with the additional variables. The transition rates $\tilde{q}$ of the model that is proposed as the next model then follow from the bijection from the transition rates of the current to the transition rates of the proposed model.

Next, the ion channel states need to be adapted to the proposed model. Since the number and meaning of states may change with the model, care should be taken in the way these states are adapted.

Finally, as the proposed (m, q, x) are sampled from a more or less arbitrary, but hopefully cleverly chosen, proposal distribution, an acceptance–rejection step is performed in order to obtain the correct invariant distribution. A transition to a new model m and the adapted q and x are accepted in such a way that it is a sample from $p((m, q, x)|c^{i+1}, \xi^i, y)$.

In summary, the model updating step 3 in the reversible jump sampler consists of the following three steps:

3(i) Sample a model m from some model jump proposal distribution.

3(ii) Sample, if necessary for dimension matching, additional variables and adapt the transition rates q and the ion channel states x to the new model m to obtain $\tilde{q}$ and $\tilde{x}$.

3(iii) Calculate the acceptance probability of the transition to the model m and accept the move to m with this probability.

In the following subsections we will explain the updating of the model in more detail. It can be proved (see Schouten 2000) that the transition probability kernel of the generated Markov process converges in total variation norm to the desired invariant distribution and that the Markov process is geometrically ergodic. Examples of reversible jump MCMC in different contexts can be found in Richardson and Green (1997) and Robert *et al.* (2000).

3.2. The model jump proposal probabilities

For each model there are $\mathcal{M}$ move types, the one belonging to the model itself included, which change the model into one of the $\mathcal{M}$ models. We require that $(m, \tilde{q}, \tilde{x})$ be sampled

from $p((m, q, x)|c^{i+1}, \xi^i, y)$. To speed up convergence we let the model jump proposal probabilities depend on (m^i, q^i, x^i), and not on c^{i+1} and ξ^i. The acceptance–rejection step described below will correct for this. The proposal probabilities for the different jumps are denoted by $r(m|(m^i, q^i, x^i))$ for candidate m, and sum to one. If the model m that is proposed equals the present model m^i, then q and x will not be altered and we will see later that it will then be accepted with probability one. Otherwise, the transition rates and the ion channel states need to be adjusted to be consistent with the proposed model. In general, jumps from a model to any other model are allowed, although it sometimes may be convenient to limit the number of possible transitions. For example, one may only allow transitions to models where either a state is added or one of the states is deleted.

We choose the dependence of the model jump proposal probabilities on (m^i, q^i, x^i) in such a way that jumps to another model are more likely when x^i does not fit (m^i, q^i) very well. For this we use the information in $z^i = 1_{S_{m^i,o}}(x^i)$, the current open–closed process, and base the proposal probability for a jump on m^i, q^i and z^i.

The open–closed process $\{Z_t\}_{t\in\mathbb{N}_0}$ is an aggregated Markov model. The process takes the value 0 if the Markov model resides in a closed state and 1 if the Markov model resides in an open state. From the Shannon–McMillan–Breiman theorem (see, for instance, Durrett 1996, Section 6.5) it follows that, whenever this process is stationary and ergodic, its scaled log-likelihood converges almost surely to a constant, the negative entropy $H(m, q)$:

$$-\frac{1}{T}\log p(Z_1^T|(m, q)) \to H(m, q) \qquad \text{for } T \to \infty \text{ a.s.} \tag{3}$$

In Section 4 we will prove that (3) holds and show that, when the underlying Markov chain describing the gating mechanism contains only one open state, the entropy can be computed. For this reason we let the proposal probabilities depend on the relative deviation of the negative scaled log-likelihood of the present model m^i to its entropy $H(m^i, q^i)$. Define

$$A^i = A((m^i, q^i), z^i) = \left(\frac{H(m^i, q^i)}{H(m^i, q^i) - (1/T)\log p(z|(m^i, q^i))}\right)^2. \tag{4}$$

It follows from (3) that the better x^i fits (m^i, q^i), the larger is A^i. We define the proposal probabilities of the model jumps in the following way

$$r(m|(m^i, q^i), z^i) = \begin{cases} \dfrac{1}{A^i + b_1}\left(A^i + \dfrac{B_1 B_2}{\mathcal{M} - 1 + B_2}\right) & \text{if } m = m^i, \\ \dfrac{1}{A^i + b_1}\left(\dfrac{B_1}{\mathcal{M} - 1 + B_2}\right) & \text{if } m \neq m^i, \end{cases} \tag{5}$$

where the constant B_1 determines the influence of 'how well z^i fits (m^i, q^i)' and should in some sense depend on the length T relative to the mixing conditions of the underlying Markov chain $\{X_t\}_{t\in\mathbb{N}_0}$, and B_2 determines 'how strong is our preference for sticking to our current model'. B_1 and B_2 may be chosen different for every model m, but are fixed in advance. In Section 4 we will discuss the choice of the model jump proposal probabilities in more detail.

If the state space contains more than one open state the proposal probabilities need to be

defined in a different way. We will only consider cases, though, where the number of open states is one. We remark that in these cases the use of uniform proposal distributions results in lower acceptance rates than when (5) is used.

3.3. The adaptation of the transition rates to the new model

As mentioned earlier, a transition to a new model is based on a bijection between the current model and the proposed model, and for every pair of models m_1 and m_2 two bijections need to be defined, one for the move from m_1 to m_2 and one for the move from m_2 to m_1. The two bijections are each other's inverse, so that for each pair of models m_1 and m_2 we only need to construct one bijection f_{m_1,m_2}. We assume for convenience that the number of free transition rates of m_1 is larger than or equal to that of m_2. Let $d = d(m_1, m_2)$ denote the difference in number. Then we introduce d independent Unif(0, 1) distributed random variables $u_1, \ldots, u_d$, and

$$q_2 = f_{m_1,m_2}(u, q_1),$$

where $u = (u_1, \ldots, u_d)'$. Other probability distributions may be chosen for the u_i as well, but this complicates the procedures. The function f_{m_1,m_2} is chosen such that the means of closed and open sojourns under models m_1 and m_2 are equal. A similar concept is used in Hodgson (1999). If d is larger than 2, then higher moments or other statistics, such as the probability of a transition to a particular state, should also be used to construct a bijection. However, these may have complex forms, which complicates the construction of a bijection. We found that, even for simple models, the construction of a bijection between the sets of transition rates of those models can be rather difficult.

Explicit formulae are, of course, different for each particular pair of models. We will illustrate the above by working through a simple example in Section 3.6. When there is a larger collection of possibly more complex candidate models the bijections are constructed similarly.

3.4. The adaptation of the ion channel states to the new model

If we let x^i denote the current ion channel states in model m^i, then we propose to sample the new ion channels states x in the proposed model (m, q) from $p(x|(m, q), z^i)$. This means that the open–closed process $Z_{\rho-r}^T$ does not change when proposing a move to a different model. Surprisingly, as we will see below, if the new ion channel states x are sampled in this way, all terms in the acceptance probability that involve x cancel. Hence, in the proposed set-up we do not need to adapt the ion channel states to the new model, as we throw them away immediately in step 4 of the reversible jump sampler. Another advantage is that this way we only need starting values for $Z_{\rho-r}^T$ and not for the exact states $X_{\rho-r}^T$ of the ion channel. However, if steps 4 and 5 are interchanged or if in step 4 another sampling algorithm is used, the ion channel states do need to be adapted in order to compute the acceptance probability of the proposed jump, and starting values for the ion channel states need to be given. The latter can be done by generating a sample from $p(x|(m^0, q^0), z^0)$,

having chosen starting values m^0, q^0 and z^0. A sample from $p(x|(m, q), z)$ can be generated analogously to step 4.

3.5. The acceptance probabilities

In updating step 3(iii) of the reversible jump sampler let us again assume for convenience that the proposed model m is 'bigger' than the current model m^i, i.e. the number of transition rates of m is larger than that of m^i. As usual, the proposed state (m, q, x) is accepted with probability

$$\alpha((m, q, x)^i, (m, q, x)) = 1 \wedge \frac{p((m, q, x)|c^{i+1}, \xi^i, y)\tilde{p}((m^i, q^i, x^i)|(m, q, x))}{p((m^i, q^i, x^i)|c^{i+1}, \xi^i, y)\tilde{p}((m, q, x)|(m^i, q^i, x^i))}$$

$$\equiv 1 \wedge R, \tag{6}$$

where $\tilde{p}$ denotes the proposal density. If the proposal is rejected then we set $(m^{i+1}, \tilde{q}, \tilde{x}) = (m^i, q^i, x^i)$, otherwise we set $(m^{i+1}, \tilde{q}, \tilde{x}) = (m, q, x)$. It is immediately clear that if $m = m^i$ then the proposal will always be accepted. The posterior $p((m, q, x)|c, \xi, y)$ is proportional to the product of the likelihood $p(c, \xi, y|(m, q, x))$ and the prior $p(m, q, x)$. The likelihood ratio is 1, since $p(c, \xi, y|(m, q, x)) = p(c, \xi, y|(m, q), z)$ and the open–closed process is the same for the current and the proposed model. The prior $p(m, q, x)$ factorizes as $p(x|(m, q))p(q|m)p(m)$. The terms $p(m)$ cancel in the prior ratio, because all models have equal prior probability. The proposal ratio is given by

$$\frac{\tilde{p}((m^i, q^i, x^i)|(m, q, x))}{\tilde{p}((m, q, x)|(m^i, q^i, x^i))} = \frac{r(m^i|(m, q, x))p(x^i|(m^i, q^i), z^i)}{r(m|(m^i, q^i, x^i))p(x|(m, q), z^i)h(u)}|J_{m^i,m}(u, q^i)|.$$

Here, $J_{m^i,m}(u. q^i)$ is the Jacobian of $f_{m^i,m}$ in $(u. q^i)$, which results from the adaptation of the transition rates, r is the proposal distribution of the model jumps defined in (5), and h is the probability density of the additional variables u. Since the additional variables are sampled independently from the Unif(0, 1) distribution, h cancels out. If x is allowed under z, meaning that the open–closed process corresponding to x equals z, then

$$p(x|(m, q), z) = \frac{p(x|(m, q))}{p(z|(m, q)}.$$

Since x is constructed in such a way that the open–closed process does not change, this is true. Furthermore, the model jump proposal probabilities r depend only on the open–closed process and not on the exact states. Therefore, all terms concerning x and x^i cancel and, for $m \neq m^i$, (6) reduces to

$$\alpha((m^i, q^i, x^i), (m, q, x)) = 1 \wedge \frac{p(z|(m, q))p(q|m)r(m^i|(m, q, x))}{p(z|(m^i, q^i))p(q^i|m^i)r(m|(m^i, q^i, x^i))}|J_{m^i,m}(u, q^i)|.$$

The reversed jump from m to m^i has acceptance probability $\min(1, R^{-1})$.

3.6. Example

Suppose that a model needs to be selected from the following two models based on the recording y: the one-gate model 1 written as

$$\mathcal{C} \underset{q_{1,2}}{\overset{q_{1,1}}{\rightleftharpoons}} \mathcal{O}$$

and the two-gate (fast–slow) model 2 written as

$$\mathcal{C}_1 \underset{q_{2,2}}{\overset{q_{2,1}}{\rightleftharpoons}} \mathcal{O} \underset{q_{2,4}}{\overset{q_{2,3}}{\rightleftharpoons}} \mathcal{C}_2$$

In this example we need $d = 2$ additional variables u_1 and u_2 in order to construct a bijection for all possible transition rates q_1 in model 1 to all possible transition rates q_2 in model 2, and $q_2 = f_{1,2}(u_1, u_2, q_1)$ and $(u_1, u_2, q_1) = f_{2,1}(q_2)$. Let G_1 (O_1) and G_2 (O_2) denote the length of a closed (open) sojourn for models 1 and 2, respectively. The distributions of these variables are mixtures of exponential distributions. We have

$$\mathbb{E}G_1 = \frac{1}{q_{1,1}}, \qquad \mathbb{E}G_2 = \frac{q_{2,1}}{q_{2,2}+q_{2,3}}\frac{1}{q_{2,1}} + \frac{q_{2,3}}{q_{2,2}+q_{2,3}}\frac{1}{q_{2,4}},$$

$$\mathbb{E}O_1 = \frac{1}{q_{1,2}}, \qquad \mathbb{E}O_2 = \frac{1}{q_{2,2}+q_{2,3}}.$$

The bijections $f_{1,2}$ and $f_{2,1}$ are constructed in such a way that $\mathbb{E}G_1 = \mathbb{E}G_2$ and $\mathbb{E}O_1 = \mathbb{E}O_2$. We use the bijections

$$f_{1,2}(u_1, u_2, q_{1,1}, q_{1,2}) = \left(q_{1,1}\frac{u_1}{u_2},\ u_1 q_{1,2},\ (1-u_1)q_{1,2},\ q_{1,1}\frac{1-u_1}{1-u_2}\right), \tag{7}$$

and

$$f_{2,1}(q_{2,1}, q_{2,2}, q_{2,3}, q_{2,4}) = \left(\frac{q_{2,2}}{q_{2,2}+q_{2,3}},\ \frac{q_{2,2}q_{2,4}}{q_{2,2}q_{2,4}+q_{2,1}q_{2,3}},\ \frac{q_{2,1}q_{2,4}(q_{2,2}+q_{2,3})}{q_{2,1}q_{2,3}+q_{2,2}q_{2,4}},\ q_{2,2}+q_{2,3}\right). \tag{8}$$

The role of u_1 and u_2 in $f_{1,2}$ is to a certain extent arbitrary; only two variables are determined when it is imposed that the expectations of the closed and open sojourns are equal under both models. However, when we constructed the bijections, it turned out that (7) is a quite natural choice. It can easily be seen that both (7) and (8) are bijections. After some calculations it follows that

$$|J_{1,2}(u, q)| = \frac{q_{1,1}q_{1,2}u_1(1-u_1)}{u_2^2(1-u_2)^2} \quad \text{and} \quad |J_{2,1}(q)| = \frac{q_{2,1}q_{2,2}q_{2,3}q_{2,4}}{(q_{2,1}q_{2,3}+q_{2,2}q_{2,4})^3}.$$

Clearly, $p(x|(1, q), z)$ is either 1 or 0, and for a jump from $(1, q_1, x_1)$ for model 1 to $(2, q_2, x_2)$ for model 2 we obtain

$$R = \frac{p(x_2|(2, q_2))p(q_2|2)r(1|(2, q_2, x_2))}{p(x_1|(1, q_1))p(q_1|1)r(2|(1, q_1, x_1))p(x_2|(2, q_2), z)} \frac{q_{1,1}q_{1,2}u_1(1-u_1)}{u_2^2(1-u_2)^2}.$$

4. The choice of the model jump proposal probabilities

First, we will compute the entropy for the case when there is only one open state. Then we will empirically analyse the rate of convergence of the scaled log-likelihood to the entropy and discuss how to incorporate the relative deviation into the model jump proposal probabilities.

4.1. The entropy of aggregated Markov models with a single open state

When there is a single open state $\mathcal{O}$, we know the exact state of the Markov process X_t when $Z_t = 1$. Since we assume that the Markov process $\{X_t\}_{t\in\mathbb{N}_0}$ is irreducible, aperiodic and has a finite state space, it is positive recurrent and ergodic.

Let the probabilities $p_{101}(k)$ and $p_{100}(k)$, for $k \geqslant 0$, be defined by

$$p_{101}(k) = p(X_{k+1} = \mathcal{O}|X_0 = \mathcal{O}, X_1^k \in S_c), \tag{9}$$

$$p_{100}(k) = p(X_{k+1} \in S_c|X_0 = \mathcal{O}, X_1^k \in S_c), \tag{10}$$

whenever $p(X_0 = \mathcal{O}, X_1^k \in S_c) > 0$, and otherwise set both $p_{100}(k)$ and $p_{101}(k)$ to zero. By convention, $0 \log 0 = 0$. The following theorem identifies the entropy $H(m, q)$ defined by (3) in the case of a single open state, so that it can be computed.

Theorem 4.1. *If $\{X_t\}_{t\in\mathbb{N}_0}$ is a stationary and ergodic Markov chain with a finite state space $S = S_c \cup \{\mathcal{O}\}$ and $Z_t = 1_{\{\mathcal{O}\}}(X_t)$, then*

$$\frac{1}{T}\log p(Z_1^T|m, q) \to \sum_{k=0}^{\infty} \pi(\mathcal{O})(p_{101}(k)\log p_{101}(k) + p_{100}(k)\log p_{100}(k)) \prod_{i=0}^{k-1} p_{100}(i)$$

for $T \to \infty$ a.s.

Proof. Let the states in S be labelled $1, \ldots, K$ and let the Kth state be $\mathcal{O}$. It can be shown that, if $\{X_t\}_{t\in\mathbb{N}_0}$ is stationary and ergodic, then so is $\{Z_t\}_{t\in\mathbb{N}_0}$. The one-sided stochastic process $\{Z_t\}_{t\in\mathbb{N}_0}$ can be extended to a two-sided stationary, ergodic stochastic process $\{Z_t\}_{t\in\mathbb{Z}}$. Following Blackwell (1957), the negative entropy $H(m, q)$ is then given by

$$H(m, q) = -\int_{p\in P} \psi_0(p)\log\psi_0(p) + \psi_1(p)\log\psi_1(p)\mathrm{d}\Upsilon(p),$$

where Υ is the distribution of the conditional distribution of X_0 given $Z_{-\infty}^0$, which is defined on the set $P = \{(p_1, \ldots, p_K) : p_i \geqslant 0, \sum_{i=1}^K p_i = 1\}$, and $\psi_0(p) =$

$\sum_{i=1}^{K} p_i p(X_1 \in S_c | X_0 = i)$ and $\psi_1(p) = \sum_{i=1}^{K} p_i p(X_1 = \mathcal{O} | X_0 = i)$. Blackwell (1957) proved that

$$\{(p(X_n = 1 | Z_{-\infty}^n), \ldots, p(X_n = K - 1 | Z_{-\infty}^n), p(X_n = K | Z_{-\infty}^n))\}_{n \in \mathbb{Z}}$$

is a stationary Markov process with invariant distribution Υ. Its state space P equals

$$P = \{(0, \ldots, 0, 1), (p(X_0 = 1 | Z_0 = 0, Z_{-1} = 1), \ldots, p(X_0 = K - 1 | Z_0 = 0, Z_{-1} = 1), 0),$$

$$(p(X_0 = 1 | Z_{-1}^0 = 0, Z_{-2} = 1), \ldots, p(X_0 = K - 1 | Z_{-1}^0 = 0, Z_{-2} = 1), 0), \ldots\}$$

$$\equiv \{\chi_0, \chi_1, \ldots\},$$

and hence is countable. Therefore,

$$H(m, q) = -\sum_{k=0}^{\infty} (\psi_0(\chi_k) \log \psi_0(\chi_k) + \psi_1(\chi_k) \log \psi_1(\chi_k)) \Upsilon(\chi_k).$$

It can easily be seen that $\psi_0(\chi_k) = p_{100}(k)$, $\psi_1(\chi_k) = p_{101}(k)$, for all $k \in \mathbb{N}_0$. Furthermore,

$$\Upsilon(\chi_k) = p(Z_{-k} = 1, Z_{-k+1} = 0, \ldots, Z_0 = 0) = \pi(\mathcal{O}) \prod_{i=1}^{k-1} p_{100}(i),$$

for all $k \in \mathbb{N}$, and $\Upsilon(\chi_0) = \pi(\mathcal{O})$, which completes the proof. □

Theorem 4.1 does not give the limiting entropy in closed form, and only simplifies its computation. In the following we will always approximate the infinite sum by taking a large number of terms.

4.2. The choice of constants in the model jump proposal probabilities

In order to choose the constants B_1 and B_2 in (5) we have investigated the rate of convergence of the scaled log-likelihood to the entropy $H(m, q)$. To this end we performed 100 simulations of two different aggregated Markov models for various lengths of recording. Both models are three-state models like the one in Section 3.6, but have different transition rates. The values of the transition rates were $q = (10^{-2}, 10^{-2}, 0.1, 0.5)'$ and $q = (10^{-4}, 10^{-3}, 10^{-2}, 0.3)'$. The transition rates for the first model are chosen in such a way that open and closed holding times are relatively short compared to the sampling interval. In the second model the transition rates are chosen more realistically in the sense that long closed holding times are likely to occur. The entropy H for the two models is -9.89×10^{-2} and -3.40×10^{-3} respectively, hence there is more 'information' in recordings of the first model. Tables 1 and 2 contain the sample averages and standard deviations for the two models for different lengths of recordings. From these tables it is clear that the scaled log-likelihood converges to the theoretical entropy for both models. For the first model convergence is faster, as was to be expected.

We notice from the standard deviations in Tables 1 and 2 that, if the length of a recording is increased by a certain factor, then the sample standard deviation is decreased

Table 1. The sample means and standard deviations of the simulations of the first model ($H = -9.89 \times 10^{-2}$)

	Sample	
Length T	Mean	Standard deviation
1×10^3	-9.50×10^{-2}	2.47×10^{-2}
5×10^3	-9.91×10^{-2}	1.23×10^{-2}
2.5×10^4	-9.88×10^{-2}	4.49×10^{-3}
1.25×10^5	-9.89×10^{-2}	2.50×10^{-3}

Table 2. The sample means and standard deviations of the simulations of the second model ($H = -3.40 \times 10^{-3}$)

	Sample	
Length T	Mean	Standard deviation
2×10^4	-3.21×10^{-3}	2.73×10^{-3}
1×10^5	-3.70×10^{-3}	1.46×10^{-3}
5×10^5	-3.37×10^{-3}	6.51×10^{-4}
2.5×10^6	-3.42×10^{-3}	2.82×10^{-4}

by approximately the square root of this factor. From this observation and the fact that histograms had a Gaussian shape, it can be conjectured that

$$\sqrt{T}\left(-\frac{1}{T}\log p(Z_1^T) - H(m, q)\right) \xrightarrow{d} N(0, \tilde{\sigma}^2) \qquad \text{for } T \to \infty, \tag{11}$$

with $\tilde{\sigma}^2$ some positive constant. Indeed, (11) can be proved using Theorem 8.1 of Serfozo (1975) with τ_n the time of the nth return to the open state, $\xi(t) = \log p(Z_1^t)$, $A(t) = 0$, $L(t) = 1$, $\gamma = 1/2$ and a the expected length of a closure. Hence the inverse of the square root of $A((m, q), z)$ in (4) can be approximated by a normal distribution. We have applied a Kolmogorov–Smirnov test to test the null hypothesis of normally distributed scaled log-likelihoods, and all bootstrapped p-values exceeded 0.05 except for the simulations of the second model with $T = 2 \times 10^4$.

Based on the foregoing, we choose B_1 in the following way. If T is the length of the recording and q^0 are the starting values for the transition rates of model m^0, then we compute $H(m^0, q^0)$ and fit a normal distribution to 100 realizations of $(1/T)\log p(Z_1^T)$. If τ is the standard deviation of the fitted normal distribution, then B_1 is chosen to be 2 or 3 times

$$\left(\frac{H(m, q)}{\xi_{1-\alpha/2\tau}}\right)^2, \tag{12}$$

with ξ_α the α-quantile of the standard normal distribution. For example, let $T = 5 \times 10^5$ and $\alpha = 0.05$ then (12) is 6.82 and 2.40×10^3 for model 1 and model 2, respectively. Hence, B_1 will be chosen much larger for model 2.

When a recording is noisier, it is more difficult to estimate the transition rates of the underlying Markov chain and, therefore, also to select a model. In that case if model jumps are proposed very often, i.e. when B_2 is chosen small, then the convergence properties of the Markov process that is generated by the reversible jump sampler, and especially those of the sampled transition rates, are rather poor. We have chosen B_2 to be approximately twice the inverse of the signal-to-noise ratio of the recording.

5. Results from simulated and recorded samples

We have investigated a number of model schemes for both simulated and recorded data sets. Below we will discuss the results of two such schemes, shown in Figures 1 and 2, for three simulated data sets and one recorded data set. The recorded data set is an inside-out patch measurement of a single potassium outward rectifier in a barley leaf protoplast. The data were sampled at a fixed holding potential (−96 mV). See Vogelzang (1996) for experimental details. If moves between some of the models are not allowed, then corresponding model jump proposal probabilities are taken to be zero. Scheme 1 is particularly interesting in the context of the ion channel data that we have at our disposal and is used to choose the number of gates in the mechanism that generate long closures. Let $\mathcal{M}_L$ denote the model with L such gates, i.e. the first model in scheme 1 is $\mathcal{M}_{L_1}$. Biologists conjecture that the long closures of the potassium outward rectifier are caused by a number of independent and identical gates. The leftmost state of model $\mathcal{M}_L$, for instance, corresponds to L closed independent, identical gates. Since these gates act independently, the transition rates can simply be added. The second scheme can be used to test whether

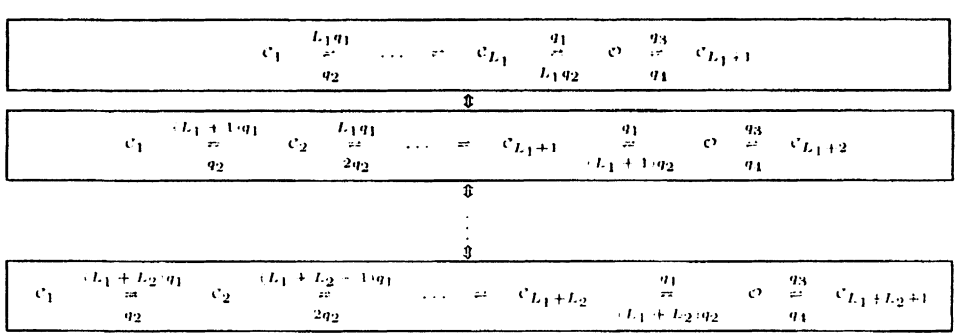

Figure 1. Model scheme 1 (simulations 1, 2 and 3, recording).

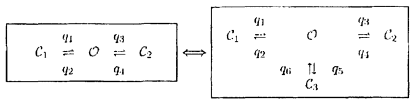

Figure 2. Model scheme 2 (simulation 1).

two or three dependent closed states are responsible for the closures. The first model in scheme 2 is $\mathcal{M}_1$.

Simulations 1, 2 and 3 are from models $\mathcal{M}_4$, $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. The transition rates in the Markov processes as well as the noise parameters are given in Table 3. The transition rates for simulation 1 were chosen such that they resemble those of the recorded data set, and simulation 1 contains both very long and very short closures. All data sets have a low noise level, except for simulation 2. We have analysed two different sections of the recording, both of length $T = 2 \times 10^5$. The simulations have length $T = 2 \times 10^5$, except for simulation 1, which has length $T = 1 \times 10^5$. Scheme 1 has been applied to simulation 1 with $L_1 = 1, L_2 = 4$, to simulation 2 with $L_1 = 1, L_2 = 3$, to simulation 3 with $L_1 = 1, L_2 = 3$, to the first section of the recording with $L_1 = 1, L_2 = 4$ and to the second section of the recording with $L_1 = 2, L_2 = 3$. Since none of the jump proposals to model $\mathcal{M}_1$ was accepted for the first section of the recording, we deleted this model from the model scheme. The number of models is $L_2 + 1$. We report the results of the application of scheme 2 to simulation 1. We found similar results for other simulations and recordings.

The bijection for a transition from model $\mathcal{M}_k$ to $\mathcal{M}_l$, with $l, k > 0$, if allowed, is

$$f_{\mathcal{M}_k,\mathcal{M}_l}(q_1, q_2, q_3, q_4) = \left(\frac{k}{l}\frac{q_2}{((q_1+q_2)/q_1)^{k/l}-1}, \frac{k}{l}q_2, q_3, q_4\right),$$

Table 3. The values of the transition rates, the model indicator, the noise parameters and the model jump proposal parameters B_1 and B_2. The current levels are given in picoamperes, and the noise variances in square picoamperes

Model	Sample				
$\mathcal{M}_i$	Simulation 1	Simulation 2	Simulation 3	Recording	Recording
i	($\mathcal{M}_4$)	($\mathcal{M}_1$)	($\mathcal{M}_2$)	I	II
q	(3.6, 2.8, 130, 7000)	(100, 100, 100, 1000)	(10, 10, 100, 2000)	–	–
M	4	1	2	–	–
μ	(0, −3)	(0, −0.85)	(0, −1.25)	–	–
σ^2	(0.001, 0.11)	(0.005, 0.05)	(0.005, 0.1)	–	–
σ_c^2	0.048	0.40	0.023	–	–
ϕ	(−0.28, −0.10, −0.05)	(−0.30, −0.10, −0.05)	(−0.23, −0.17, −0.08)	–	–
B_1	500	3000	500	500	500
B_2	3	15	6	3	3

and the bijection from model 1 to model 2 in scheme 2 is

$$f_{1,2}(u_1, u_2, q_1, q_2, q_3, q_4) = \left(q_1, q_2, u_1 q_3, \frac{u_1 q_4}{u_2}, (1-u_1)q_3, \frac{(1-u_1)q_4}{1-u_2}\right),$$

where u_1 and u_2 are two Unif(0, 1) distributed additional variables. Bijection $f_{1,2}$ can easily be inverted to obtain the reverse bijection $f_{2,1}$.

Reversible jump sampler runs have been generated of lengths $N = 5000$ for simulations 1 and 3, and the first recorded section. For the other data sets $N = 4000$. The parameter values of B_1 and B_2 are also given in Table 3.

Figure 3 shows the reversible jump sampler outcomes for the model indicator and scaled histograms. For the histograms we have used a burn-in period between 1500 and 2500, chosen by eye. Diagnostic checking of the convergence of the generated Markov chain is usually difficult for reversible jump MCMC methods (see Brooks and Giudici 1999), since there are only few transitions to the models that are less likely. One may, of course, remove

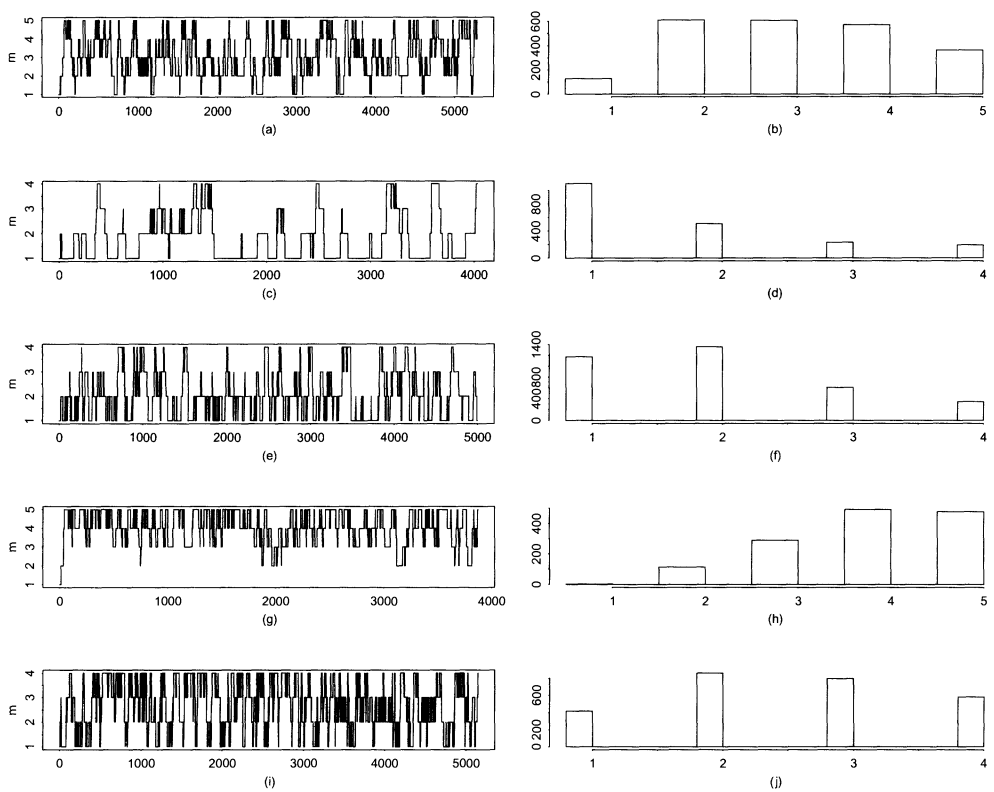

Figure 3. The reversible jump sampler runs for model M with the corresponding scaled histograms for simulation 1 ((a) and (b)), simulation 2 ((c) and (d)), simulation 3 ((e) and (f)) and two sections of the recording ((g) and (h), (i) and (j)).

these models from the set of candidate models, but that is a very opportunistic choice. Brooks and Giudici (1999) propose multiple runs of the sampler. However, this is not computationally feasible in our case. We have checked the convergence of the model indicator using the gibbsit software of Raftery and Lewis (1992) with satisfactory results. Furthermore, since we adapt the transition rates in a proposed model by fixing the mean closed and open sojourn lengths, these means can also be used to diagnose convergence. The mean closed and open sojourn lengths using the sampled transition rates for simulation 2 are shown in Figure 4. The gibbsit software applied to the closed and open means gave satisfactory results. The outcomes for c, x and the parameters θ are not shown, but similar results to those in de Gunst *et al.* (2001) were found here.

Table 4 contains the estimates for the posterior densities of the model indicator M for scheme 1, the scaled histogram counts. Also given in Table 4 are the MCMC standard errors. Except for simulation 1, the maximum posterior estimates agree with the models we have simulated from. The estimated posterior densities for simulation 1 and the recording are rather flat, which was to be expected since the underlying Markov processes mix slowly. In most cases the standard errors are rather large, so that results need to be interpreted with care.

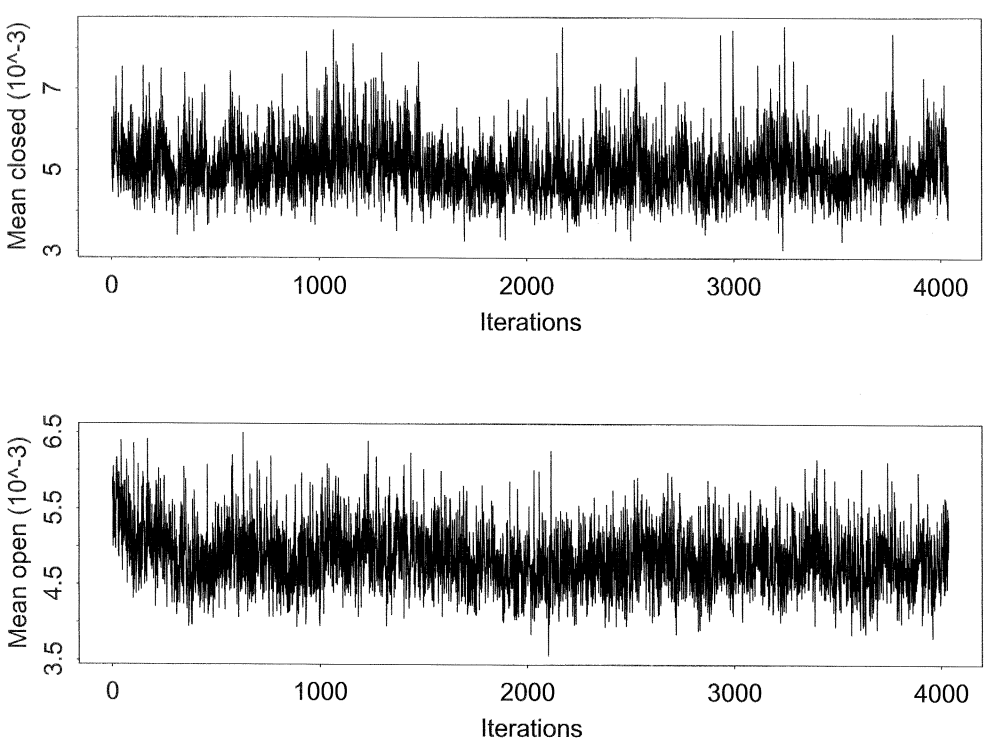

Figure 4. The mean closed and open sojourn lengths for the sampled transition rates for simulation 2.

Table 4. The estimates for the posterior densities of the model indicator M in scheme 1 and the corresponding MCMC standard errors

Model	Sample				
$\mathcal{M}_i$ i	Simulation 1 ($\mathcal{M}_4$)	Simulation 2 ($\mathcal{M}_1$)	Simulation 3 ($\mathcal{M}_2$)	Recording I	Recording II
1	0.056 (0.024)	0.541 (0.060)	0.337 (0.048)	0.000	0.157 (0.024)
2	0.268 (0.033)	0.251 (0.038)	0.389 (0.052)	0.083 (0.042)	0.325 (0.038)
3	0.267 (0.037)	0.113 (0.037)	0.175 (0.032)	0.212 (0.040)	0.299 (0.028)
4	0.251 (0.016)	0.096 (0.035)	0.099 (0.026)	0.358 (0.040)	0.219 (0.049)
5	0.159 (0.035)	–	–	0.347 (0.057)	–

The Bayes factors of the MAP estimates for the model parameter M versus the other models are given in Table 5, together with approximate standard errors that were computed by the delta method. According to Jeffreys's (1961) scale the evidence in favour of a model is substantial, strong and decisive if the Bayes factor is larger than 3.2, 10 and 100, respectively. From Table 5 we can see that, except in one instance, the first section of the recording, there is only substantial or no evidence against the other models. As could be expected from the MCMC standard errors in Table 4, the standard errors for the Bayes factors are large as well.

We have also applied a sign test to test the hypothesis that the posterior density $p(m|y)$ gives equal probability to two models, m_1 and m_2 say. If $\{m^i\}_{1 \leqslant i \leqslant N}$ are the generated

Table 5. The Bayes factors B, their approximate standard errors and the p-values of the sign test of the maximum posterior estimate versus the other models

	Sample				
	Simulation 1 ($\mathcal{M}_4$) MAP = 2	Simulation 2 ($\mathcal{M}_1$) MAP = 1	Simulation 3 ($\mathcal{M}_2$) MAP = 2	Recording I MAP = 4	Recording II MAP = 2
B	4.79 (3.00) 1.00 (0.08) 1.07 (0.23) 1.69 (0.69)	2.16 (0.63) 4.79 (0.80) 5.64 (2.88)	1.15 (0.28) 2.22 (0.62) 3.93 (1.43)	∞ 4.31 (2.49) 1.69 (0.42) 1.03 (0.21)	2.07 (0.35) 1.09 (0.17) 1.48 (0.49)
p	2×10^{-4} 0.99 0.84 0.19	3×10^{-2} 2×10^{-4} 2×10^{-4}	0.46 3×10^{-4} $1 - 10^{-6}$	0 7×10^{-1} 4×10^{-2} 0.89	6×10^{-4} 0.60 0.14

samples of the model indicator M, then we define M_i by $M_i = 1_{m_1}(m^i) - 1_{m_2}(m^i)$. It can be proved that the Markov chain generated is geometrically ergodic. Since it is also stationary, the process $M_1, M_2, \ldots, M_N$ is also stationary and geometrically ergodic. This means that central limit theorems hold. Testing whether $p(m_1|y) = p(m_2|y)$ is the same as testing whether the medians of the M_i are non-zero, for which we have used the statistic $t(M_1, \ldots, M_N) = \sum_{i=1}^{N} M_i(\sqrt{N\sigma_M^2})^{-1}$. The variance is $\sigma_M^2 = \mathrm{var}(M_i) + 2\sum_{k=1}^{\infty} \mathrm{cov}(M_1, M_{k+1})$, which needs to be estimated from the data. Under the null hypothesis $t(M_1^N)$ converges to a standard normal distribution. Table 5 also contains the resulting p-values for the maximum posterior estimates against the other models. When the Bayes factors are larger than 2 then the p-values are very small, and the null hypothesis of equal posterior probabilities of two models will be rejected.

We ran the reversible jump sampler of scheme 2 for simulation 2. Figure 5 shows the results for the model indicator M and the parameters θ. When the sampler makes an

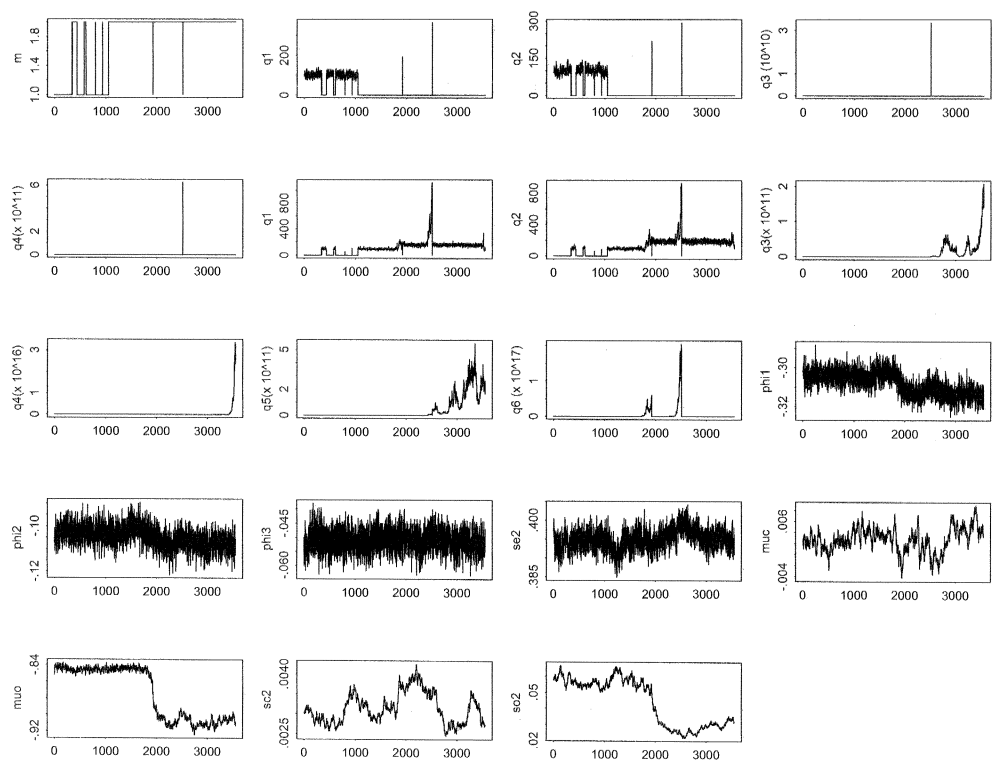

Figure 5. A reversible jump sampler run for simulation 2 in scheme 2. From top to bottom and from left to right, the samples of the model indicator M, the transition rates of both models, the AR parameters ϕ and σ_ϵ^2, the current levels μ and the noise variance σ^2 are depicted. The transition rates of the model other than the current one are set to zero. The samples of μ_c and μ_o are given in picoamperes, those of σ_ϵ^2, σ_c^2 and σ_o^2 in square picoamperes.

excursion to one model, the transition rates of the other model are set to zero. From the samples it is clear that until iteration step 1000 the sampler favours model 1. After that a transition to model 2 is proposed and accepted, and the transition rates that are sampled for model 2 suddenly increase rapidly to very high values. Proposed transitions to model 1 are then almost always rejected. Clearly, the values that are sampled, of the order of 10^{17} for q_6, are not physically realistic. The sampled variances σ_o^2 and σ_ϵ^2 and the open current level μ_o decrease somewhat at the same time. The sampling of large values for the transition rates, therefore, seems to be caused by the fact that large values of the 'open' noise are seen as very short excursions to state $\mathcal{C}_3$. This undesired effect did not occur when the reversible jump sampler was applied to simulations from the second model in scheme 2.

The computation time that is needed to make a single run of the reversible jump sampler depends on the number of models in the scheme, the relative likeliness of these models and the complexity of the 'largest' model in the scheme. Computation times are, therefore, longer than for the Gibbs sampler for a fixed model. Since we ran the Gibbs sampler for at least one of the models in every scheme, so that the starting values can be chosen close to the modes of the posterior densities, the computation times were only slightly longer than for the Gibbs sampler. On a Sun workstation (with 128 MB of RAM and an UltraSPARC 8 processor) the reversible jump sampler took between one and several days, which is rather long. However, in general the selection of a model only needs to be done a small number of times for recordings that are representative for a larger collection of recordings.

6. Discussion

Our proposed reversible jump sampler for the selection of a gating mechanism from a set of candidate gating mechanisms gave promising results for simulated data and recorded data. Its mixing properties are reasonable, and convergence seems to be rather fast. The computation times are usually not much longer than the computation times of the Gibbs sampler for a fixed model. In most cases we recovered the original model when we applied the sampler to simulated data. Clearly, the latter also depends on the amount of information that is present in a recording. The slower the mixing of the underlying Markov process, the longer the recordings needed to perform a sensible model selection. Even for recordings that have a length of $T = 2 \times 10^5$ the posterior density of the model indicator can be rather flat. However, the identifiability of models in a scheme is a problem of the statistical model and the data, and not of the method used.

The sampler proposed by Hodgson and Green (1999) cannot be generalized in a straightforward way to Markov processes with more than one open state, because their sampling algorithm is based on the independence of the lengths of consecutive closed and open sojourns, which is lost when open states are added. Although in its present form our reversible jump sampler is not applicable to Markov processes with more than one open state, additional open states can be incorporated in a natural way. The model jump proposal probabilities are presently based on the entropy of the corresponding aggregated Markov model. The entropy of aggregated Markov models with more than one open state cannot be easily identified. The sampler can be adapted, though, to general Markov processes by

choosing model jump probabilities that do not depend on the model, which is only a small change to the algorithm. We have compared the model jump proposal probabilities of (5) to uniform proposal probabilities, i.e. that do not depend on the model, and it turned out that in those cases where the probabilities of (5) were used the reversible jump sampler had better mixing properties. The construction of bijections between the transition rates of the candidate Markov models is, of course, more complicated if extra open states are added.

There are some drawbacks to the method, apart from the rather long computation times. Diagnosing convergence can be difficult, especially when models are proposed that are less likely. Also, the construction of bijections between models can be quite difficult, since the moments of closed and open sojourn lengths do not always have tractable forms. The choice of bijections is very important. The proposed transition rates need to be probable in the new model, otherwise many proposals will be rejected. Especially when models of different dimension are proposed, the choice of bijections is crucial. Furthermore, it may occur that the sampler gets stuck in high values of the transition rates whenever a model has more freedom, i.e. more Markov states, than the 'true' model.

Finally, we remark that Hodgson (1999) proposed simulated tempering to improve the mixing properties of the reversible jump sampler. We have not tried this, but it may also improve our sampler.

Acknowledgements

The authors would like to thank Michel Dekking and Ronald Meester for their helpful comments and advice. Furthermore, the authors thank the associate editor and a referee for their detailed comments. We are especially grateful to the referee for pointing us to the paper of Serfozo (1975) in connection with the proof of conjecture (11).

References

Ball, F.G. and Sansom, M.S.P. (1989) Ion-channel gating mechanisms: model identification and parameter estimation from single channel recordings. *Proc. Roy. Soc. Lond. Ser. B*, **236**, 385–416.

Ball, F.G., Cai, Y., Kadane, J.B. and O'Hagan, A. (1999) Bayesian inference for ion channel gating mechanisms directly from single channel recordings, using Markov chain Monte Carlo. *Proc. Roy. Soc. Lond. Ser. A*, **455**, 2879–2932.

Blackwell, D. (1957) The entropy of finite-state Markov chains. In *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, pp. 13–20. Prague: Publishing House of the Czechoslovak Academy of Sciences.

Brooks, S.P. and Giudici, P. (1999) Convergence assessment for reversible jump MCMC simulations. In J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds), *Bayesian Statistics 6*, pp. 733–742. Oxford: Clarendon Press.

de Gunst, M.C.M., Künsch, H.R. and Schouten, J.G. (2001) Statistical analysis of ion channel data using hidden Markov models with correlated noise and filtering. *J. Amer. Statist. Assoc.*, **96**, 805–815.

Durrett, R. (1996) *Probability: Theory and Examples*. Belmont, CA: Duxbury Press.

Frühwirth-Schnatter, S. (1995) Bayesian model discrimination and Bayes factors for linear Gaussian state space models. *J. Roy. Statist. Soc. Ser. B*, **57**, 237–246.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (eds) (1996) *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Green, P.J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **4**, 711–732.

Hodgson, M.E.A. (1999) A Bayesian restoration of an ion channel signal. *J. Roy. Statist. Soc. Ser. B*, **61**, 95–114.

Hodgson, M.E.A. and Green, P.J. (1999) Bayesian choice among Markov models of ion channels using Markov chain Monte Carlo. *Proc. Roy. Soc. Lond. Ser. A*, **455**, 3425–3448.

Horn, R. (1987) Statistical methods for model discrimination. *Biophys. J.*, **51**, 255–263.

Jeffreys, H. (1961) *Theory of Probability* (3rd edn). Oxford: Oxford University Press.

Kass, R.E. and Raftery, A.E. (1995) Bayes factors. *J. Amer. Statist. Assoc.*, **90**, 773–795.

Kehagias, A. (1996) Bayesian classification of hidden Markov models. *Mathematical and Computer Modelling*, **23**, 25–43.

Newton, M.A. and Raftery, A.E. (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Roy. Statist. Soc. Ser. B*, **56**, 3–48.

Raftery, A.E. and Lewis, S.M. (1992) How many iterations in the Gibbs sampler? In J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds), *Bayesian Statistics 4*. Oxford: Clarendon Press.

Richardson, S. and Green, P.J. (1997) On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B*, **59**, 731–792.

Robert, C.P., Rydén, T. and Titterington, D.M. (2000) Bayesian inference in hidden Markov models through reversible jump Markov chain Monte Carlo. *J. Roy. Statist. Soc. Ser. B*, **62**, 57–75.

Rydén, T. (1995) Estimating the order of hidden Markov models. *Statistics*, **26**, 345–354.

Schouten, J.G. (2000) Stochastic modeling of ion channel kinetics. Doctoral thesis, Dept of Exact Sciences, Div. of Mathematics and Computer Science, Free University of Amsterdam, Netherlands.

Serfozo, R.F. (1975) Functional limit theorems for stochastic processes based on embedded processes. *Adv. in Appl. Probab.*, **7**, 123–139.

Vogelzang, S.A. (1996) The state of plasma membrane polarization in plant cells: an interplay between ion channel and pump activity. Doctoral thesis, Dept of Plant Biology, University of Groningen, Netherlands.

Received April 2001 and revised December 2002