THE ROYAL SOCIETY

# Bayesian choice among Markov models of ion channels using Markov chain Monte Carlo

BY MATTHEW E. A. HODGSON† AND PETER J. GREEN
*Department of Mathematics, University of Bristol, Bristol BS8 1TW, UK*

*Received 7 August 1998; revised 21 December 1998; accepted 20 January 1999*

Hodgson formulates a Bayesian method of ion-channel analysis, using an alternating renewal model of channel kinetics and an autoregressive process to model patch-clamp noise. The resulting posterior is computed by a reversible jump Markov chain Monte Carlo method. Our focus here is on extending this analysis to the more physically realistic hidden Markov models of channel dynamics. Here, the states of the channel are partitioned into classes of states having the same conductance, so that patch-clamp recording is informative only about the aggregated process. This loss of information increases the challenge of making structural inference about the underlying Markov chain. The study we present, involving discrimination between a few simple models, serves to give some indication as to the feasibility of more general Markov model determination for ion channels.

**Keywords: exponential mixture distributions; hidden Markov models; ion channels; predictive distributions; reversible jump Markov chain Monte Carlo; sensitivity analysis**

## 1. Hidden Markov models for ion channels

Ion channels, large proteins spanning cell membranes, are thought to exist in a finite number of physicochemically distinct states, some of which allow the passage of current in the form of selected ions. Their kinetics are most frequently modelled by a continuous-time Markov chain in which the state space is partitioned into classes of states having the same conductance. Since it is the current through the channel which is measured (with noise), states having the same conductance are indistinguishable. A thorough exposition of the properties of hidden Markov models of single-channel kinetics is given by Colquhoun & Hawkes (1982). In theory the technique of patch-clamp recording (see, for example, Sakmann & Neher 1995) reveals which class the channel state belongs to, so that inference about the postulated 'hidden' Markov chain may be based on this so-called 'aggregated' process. Such inference is the subject of Fredkin & Rice (1986). In practice, degradation of patch-clamp records by additive noise and filtering prior to digitization further complicates the statistician's task. We restrict attention to the usual case of just two conductance classes, termed 'open' and 'closed'.

A conventional algorithm to restore the true channel current from a patch-clamp record involves thresholding followed by the deletion of apparent sojourns in either

† Present address: Biomathematics and Statistics Scotland, Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen AB15 8HQ, UK.

class of duration less than some critical value, a practice known as 'time-interval omission'. The use of the thresholded signal as data, with and without time-interval omission, in making inference about hidden Markov models has received much attention (see, for example, Ball *et al.* 1993; Edeson *et al.* 1994). An unwieldy likelihood hampers frequentist inference in the case of time-interval omission. For the simplest two-state Markov model, solutions to this difficulty are presented by Ball *et al.* (1992) and Ball & Davies (1995). For the general model, Hawkes *et al.* (1992) define 'extended' open and closed sojourn times and give asymptotic representations of their densities as exponential mixtures. From the Bayesian viewpoint, Ball *et al.* (1996) analyse discrete ion-channel data simulated from the two-state model, using Markov chain Monte Carlo (MCMC) methods to sample from the posterior distribution of the transition matrix.

Some authors base inference on raw rather than thresholded patch-clamp data (see Magleby & Weiss (1990) and Fredkin & Rice (1992*b*) for frequentist approaches). Fredkin & Rice (1992*a*) use recursive algorithms to compute marginal and global Bayes restorations of ion-channel signals based on a two-state Markov chain prior and a Gaussian 'white noise' likelihood. Hodgson (1999) treats the opening and closing of the channel as an alternating renewal process with gamma-distributed sojourn times, without incorporating a more fundamental hidden stochastic process to drive the aggregated process. He draws inference by MCMC sampling from the joint posterior distribution of the sojourn time parameters, the channel conductance levels and the binary step function representing the true channel current. The unknown number of discontinuities ('switches') of this function induces variable dimensionality in the joint posterior, the situation for which Green's (1995) reversible jump protocol was designed. The approach of Ball *et al.* (1997) is similar, except that they treat the latent process as a Markov chain and incorporate it explicitly into their model. The realized path of this Markov chain, rather than the derived aggregated process, is the variable-dimension object updated by reversible jump moves in their sampler. In addition to the two-state and three-state linear models (models *11* and *21* in the notation we introduce in the next section), they consider three-state cyclic and five-state models for the channel kinetics.

This paper extends Hodgson's methods to the more familiar and physically realistic class of hidden Markov models. Kienker (1989) studies classes of *equivalent* models; that is distinct Markov chains yielding probabilistically identical aggregated processes. A strategy for discriminating between alternative models, based on the Schwarz criterion, is presented by Ball & Sansom (1989). We shall attempt more fully Bayesian model selection from a set of four simple candidate models. The numbers of open and closed states and the transition rates of the Markov chain will all be unknown. Hence our MCMC sampler requires reversible jump moves to update the model in addition to those used by Hodgson to update the step function. Although our models are mathematically identifiable, the comprehensive inference we attempt may be unfeasible owing to the low information content of the data. Our study should reveal whether the idea of Markov model determination for ion channels has the potential to be generalized.

In § 2 we introduce the set of candidate models and state the implied sojourn time densities that are non-trivial. We also describe both the Bayesian model for ion-channel model determination and the reversible jump MCMC algorithm to sample from the posterior distribution. Section 3 illustrates the performance of our method

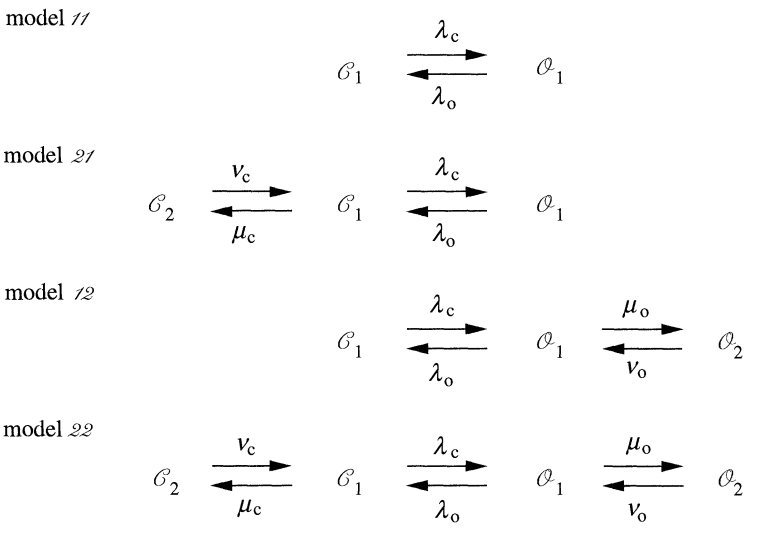


Figure 1. The candidate Markov models.

on data simulated from one of the candidate models and on the same synthetic data used by Hodgson (1999). Section 4 details an investigation into the sensitivity of results to four factors identified beforehand as being of probable importance. These factors are the true model underlying the simulated dataset, the length of the dataset, the ease of identifiability of the true model and finally the noise precision of the dataset. We conclude the paper with a discussion.

## 2. Discrimination between four simple Markov models

### (*a*) *Introduction*

The generator $Q = [q_{ij}]$ of a continuous-time Markov chain with state space $\mathcal{S}$ induces a directed graph on $\mathcal{S}$, in which there is an edge leading from $i$ to $j$ if and only if $q_{ij} > 0$. Figure 1 illustrates the four models of ion-channel kinetics considered in this paper by their associated graphs. Each edge is accompanied by the corresponding transition rate. The state space of each Markov model is partitioned into an open class $\mathcal{O}$ and a closed class $\mathcal{C}$. Classes consist of either a single state or an 'inner' state (given the subscript 1) and an 'outer' state (subscript 2). As there is only a single pathway between the classes in all models, deriving the probability density function of the sojourn time in a two-state class is a simple exercise. This enables us to integrate the latent Markov process out analytically and include only the aggregated process in our ion-channel model. We will simulate segments of raw data from the different Markov models for analysis and then sample from the joint posterior distribution of the model indicator, transition rates and aggregated process. The chief interest will be the extent to which the data 'select' the true model if we are indifferent between them *a priori*.

### (*b*) *The density of the sojourn time in a two-state class*

It has been known for many years (see, for example, Colquhoun & Hawkes 1981) that the sojourn time in a size-$n$ subset of the state space of a finite continuous-time

Markov chain has in general an exponential mixture distribution with $n$ components. Kijima & Kijima (1987) proved that the weights attached to these exponential components are always positive if the underlying Markov chain is time reversible, a condition satisfied trivially by all of our candidate models. It transpires that the probability density function $f_c(t)$ of the closed sojourn times under Markov models *21* and *22* can be written as

$$f_c(t) = (\tfrac{1}{2} - c)(a - b)\exp(-(a - b)t) + (\tfrac{1}{2} + c)(a + b)\exp(-(a + b)t), \quad t > 0, \tag{2.1}$$

where $a = \frac{1}{2}(\nu_c + \mu_c + \lambda_c)$, $b = \sqrt{a^2 - \nu_c\lambda_c}$ and $c = (\lambda_c - a)/2b$. It is easily shown that $c$ lies in the interval $(-\frac{1}{2}, \frac{1}{2})$. Closed sojourn times under models *21* and *22*, and of course open times under models *12* and *22*, therefore have the form of a mixture of two exponential distributions with positive weights.

### (c) *The Bayesian model*

Let $\mathcal{K} = \{11, 21, 12, 22\}$ be the set of candidate Markov models and $k \in \mathcal{K}$ be the model indicator. For each $k$, we build a prior model for the noiseless data. Given $k$, we denote the vector of transition rates by $\boldsymbol{\Lambda}$; hence $\boldsymbol{\Lambda}$ has dimension 2 if $k = 11$, 4 if $k = 21$ or *12* and 6 if $k = 22$. Let $\boldsymbol{x}$ be the indicator function of channel openness on the continuous interval $[0, T]$ over which the patch-clamp recording is taken. This function represents the aggregated process, since $\boldsymbol{x}(t)$ is equal to 1 if the channel state is in the open class at time $t$ and 0 otherwise. Let also $l_o$ and $l_c$ be the channel conductance levels in the open and closed classes, respectively. Then if $\boldsymbol{c}$ is the binary step function representing the true current, we have

$$\boldsymbol{c}(t) = l_c + (l_o - l_c)\boldsymbol{x}(t), \quad \forall t \in [0, T].$$

The discretized signal, namely the vector $(\boldsymbol{c}(t) : t = 0, 1, \ldots, T)$, is degraded by additive dependent noise $\boldsymbol{z}$ and then passed through a finite impulse response Gaussian filter $F$ to yield the vector $\boldsymbol{y}$ of synthetic data points:

$$\boldsymbol{y} = F(\boldsymbol{c} + \boldsymbol{z}) = F\boldsymbol{c} + F\boldsymbol{z}.$$

Here, $F$ is known and $\boldsymbol{z}$ is treated as multivariate normal with mean zero. Choosing $\boldsymbol{z}$ from the family of autoregressive processes yields a simple banded Toeplitz form for the inverse $\boldsymbol{\Sigma}^{-1}$ of the variance–covariance matrix, which is computationally convenient. Prolonged desensitized spells are evident in single-channel patch-clamp recordings, so that knowledge of the noise process masking a real signal may be considerable. In addition, the family of autoregressive processes is sufficiently large that we do not consider the specification of the form of the noise to be a major drawback of our methodology.

The prior probabilities assigned to models *11*–*22* shown in § 2 *a* are given by the four-dimensional vector $\boldsymbol{m}$. If there are two closed states ($k = 21$ or *22*), the three transition rates determining the density of the closed sojourn times are given independent gamma priors:

$$\nu_c, \mu_c, \lambda_c \sim \Gamma(\alpha, \tfrac{1}{2}\beta).$$

In the limit as $\mu_c$ tends to 0, the sojourn time distribution collapses to the $\mathrm{Exp}(\lambda_c)$ distribution. This is of course the sojourn time distribution when there is a single

closed state, in which case $\lambda_c$ is given a gamma prior distribution but with a different inverse scale parameter:

$$\lambda_c \sim \Gamma(\alpha, \beta).$$

Transition rates governing open sojourn times are assigned priors in the same manner. When there are two closed states, the expected closed sojourn time is

$$E[T] = \frac{\mu_c + \nu_c}{\nu_c \lambda_c}.$$

Therefore, halving the inverse scale parameters of the priors on the transition rates, and hence doubling their prior means, has the effect of stabilizing the location of the closed sojourn time distribution in moving from one to two closed states; this is its motivation.

Although the order constraint $l_c < l_o$ on the channel conductance levels is necessary for identifiability, it is of no practical importance in our examples. We therefore omit it, assigning independent normal priors to the levels:

$$l_c, l_o \sim N(\theta, \kappa^2).$$

The autoregressive noise process $\boldsymbol{z}$ has variance $\sigma^2$. Figure 2 illustrates the joint distribution of the Markov model indicator, transition rates, aggregated process and data by its directed acyclic graph (DAG). As is the usual convention, circles contain random quantities and squares contain fixed or observed quantities.

### (*d*) *Design of the reversible jump sampler*

We require to simulate from the joint posterior

$$p(k, \Lambda, \boldsymbol{x}, l_c, l_o \mid \boldsymbol{y}) \propto p(k)p(\Lambda \mid k)p(\boldsymbol{x} \mid \Lambda, k)p(l_c)p(l_o)p(\boldsymbol{y} \mid \boldsymbol{x}, l_c, l_o). \tag{2.2}$$

First, consider the case where the channel is closed at time $t = 0$ and there are an even number of switches between the closed and open classes in the interval of observation. Let $s = 2j$ be the number of switches and $0 < t_1 < \cdots < t_{2j} < T$ be the switch times. Then there are $j$ open sojourns on the intervals $[t_1, t_2], \ldots, [t_{2j-1}, t_{2j}]$, $(j-1)$ closed sojourns on $[t_2, t_3], \ldots, [t_{2j-2}, t_{2j-1}]$ and two 'censored' closed sojourns on $[0, t_1]$ and $[t_{2j}, T]$. The conditional distribution of $\boldsymbol{x}$ given $(\Lambda, k)$ is

$$p(\boldsymbol{x} \mid \Lambda, k) = \frac{m_c}{m_c + m_o} \frac{1 - F_c(t_1 \mid \Lambda, k)}{m_c} \prod_{i=1}^{j} f_o(t_{2i} - t_{2i-1} \mid \Lambda, k) \times \prod_{i=1}^{j-1} f_c(t_{2i+1} - t_{2i} \mid \Lambda, k) \times (1 - F_c(T - t_{2j} \mid \Lambda, k)). \tag{2.3}$$

Here, $m_c$ and $m_o$ are the mean closed and open sojourn times, so the first term on the right-hand side is the probability that the channel is closed at $t = 0$. The density and cumulative distribution functions of closed sojourn times implied by the Markov model $k$ and transition rates $\Lambda$ are denoted by $f_c(\cdot \mid \Lambda, k)$ and $F_c(\cdot \mid \Lambda, k)$, respectively; for open times we replace the 'c' subscript with 'o'. The second term on the right-hand side, the density of the 'censored' closed sojourn on $[0, t_1]$, is

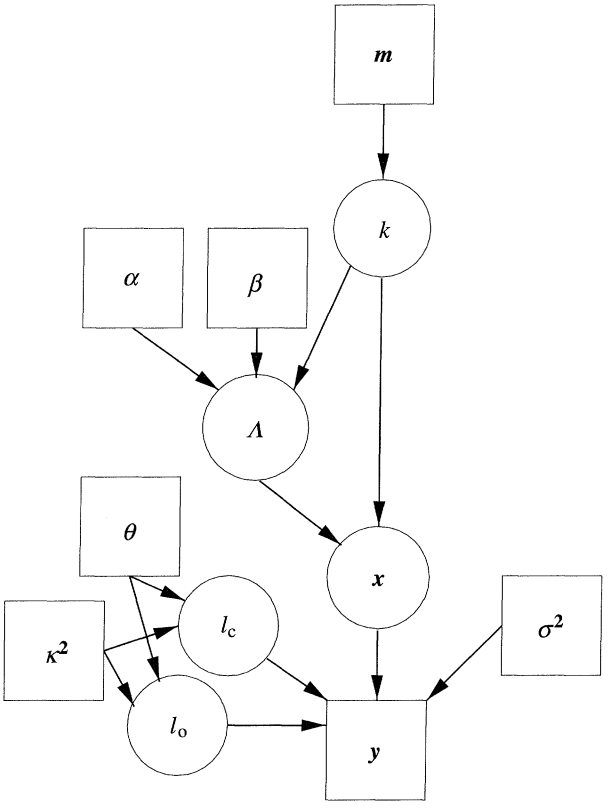


Figure 2. The DAG corresponding to the ion-channel model.

a standard result in renewal theory. The terms within the product signs are the densities of the sojourns on the intervals $[t_1, t_2]$ up to $[t_{2j-1}, t_{2j}]$, and the final term on the right-hand side is the probability that the channel remains closed from time $t_{2j}$ to time $T$. The modifications to the above expression when $\boldsymbol{x}$ has an odd number of switches, when the channel is open initially or when both occur are obvious.

The moves implemented in our sampler are

(a) updating the Markov model indicator $k$, necessarily altering $\Lambda$ simultaneously;

(b) updating the transition rates $\Lambda$ for a given $k$;

(c) updating the indicator function $\boldsymbol{x}$ for given $(k, \Lambda)$; and

(d) updating the channel conductance levels $(l_c, l_o)$.

Note that the reversible jump technique (Green 1995) is in general needed in the moves (a) and (c), as $\Lambda$ and $\boldsymbol{x}$ have variable dimensionality. We now describe move (a) in detail. Suppose the current model is the two-state model, $k = 11$. With probability 0.5, we propose to create a second closed state; otherwise we propose a second open state. In either case, the dimensionality of the corresponding transition rate vector $\Lambda$ increases from 2 to 4. Following Green's recipe, we generate a two-dimensional random vector $\boldsymbol{z} = (z_1, z_2)$ independently of the existing rates and set

$$\Lambda' = \Lambda'(\Lambda, \boldsymbol{z});$$

that is, the new transition rates $\Lambda'$ are a deterministic function of the existing rates $\Lambda$ and the random vector $\boldsymbol{z}$. Suppose that a second closed state is proposed. We shall preserve the mean sojourn times, so that the new transition rates $(\lambda_c', \mu_c', \nu_c', \lambda_o')$ are related to the existing rates $(\lambda_c, \lambda_o)$ by

$$\begin{aligned} \frac{\mu_c' + \nu_c'}{\nu_c' \lambda_c'} &= \frac{1}{\lambda_c}, \\ \lambda_o' &= \lambda_o. \end{aligned} \tag{2.4}$$

A necessary condition to satisfy the first equation is $\lambda_c' > \lambda_c$. We draw $z_1 \sim \Gamma(\rho_1, \rho_2)$ and set

$$\lambda_c' = \lambda_c \exp(z_1).$$

As both are transition rates out of the inner closed state $\mathcal{C}_1$, we relate $\lambda_c'$ and $\mu_c'$ by a multiplicative perturbation. We generate $z_2 \sim N(0, \tau^2)$ independently of $z_1$ and set

$$\mu_c' = \lambda_c \exp(z_1 + z_2).$$

Here, $\rho_1, \rho_2$ and $\tau$ are simulation parameters. The final rate $\nu_c'$ is determined by (2.4):

$$\nu_c' = \frac{\lambda_c \exp(z_1 + z_2)}{\exp(z_1) - 1}.$$

In the reverse move jumping from Markov model *21* to model *11*, which is chosen with probability 0.5 when the model is currently *21*, the inverse of the transformation specified above must be used for the new transition rates, namely

$$\begin{aligned} \lambda_c' &= \frac{\nu_c \lambda_c}{\nu_c + \mu_c}, \quad \lambda_o' = \lambda_o, \\ z_1 &= \log\left(\frac{\nu_c + \mu_c}{\nu_c}\right), \quad z_2 = \log\left(\frac{\mu_c}{\lambda_c}\right). \end{aligned}$$

From Green (1995), the appropriate acceptance probability for the move creating a second closed state is

$$\min\left\{1, \frac{p(21)p(\Lambda' \mid k = 21)p(\boldsymbol{x} \mid \Lambda', k = 21)}{p(11)p(\Lambda \mid k = 11)p(\boldsymbol{x} \mid \Lambda, k = 11)q(\boldsymbol{z})} \left|\frac{\partial \Lambda'}{\partial(\Lambda, \boldsymbol{z})}\right|\right\}, \tag{2.5}$$

where $q(\cdot)$ is the density of $\boldsymbol{z}$. The term $p(21)/p(11)$ is read directly from the vector $\boldsymbol{m}$, $p(\Lambda' \mid 21)/p(\Lambda \mid 11)$ is

$$\frac{\beta^{2\alpha}}{2^{3\alpha}(\Gamma(\alpha))^2}\left(\frac{\nu_c' \mu_c' \lambda_c'}{\lambda_c}\right)^{\alpha - 1} \exp(-\beta/2(\nu_c' + \mu_c' + \lambda_c' - 2\lambda_c)),$$

the density of $\boldsymbol{z}$ is

$$q(\boldsymbol{z}) = \frac{{\rho_2}^{\rho_1}}{\Gamma(\rho_1)\tau\sqrt{2\pi}} z_1^{\rho_1 - 1} \exp(-\rho_2 z_1 - {z_2}^2/2\tau^2),$$

and the transformation of variables from $(\lambda_c, \lambda_o, z_1, z_2)$ to $(\nu_c', \mu_c', \lambda_c', \lambda_o')$ has Jacobian

$$\left|\frac{\partial \Lambda'}{\partial(\Lambda, \boldsymbol{z})}\right| = \frac{\nu_c'^2 \lambda_c'^2}{\lambda_c^2}.$$

The term $p(\boldsymbol{x} \mid \Lambda', k = \mathit{21})/p(\boldsymbol{x} \mid \Lambda, k = \mathit{11})$ follows from (2.3); note that the open sojourn density and the mean closed sojourn time are unchanged by the proposals. The acceptance probability of the reverse move has the same form as (2.5) with the ratio terms inverted and some obvious relabelling of variables.

Proposals to create or delete the second open state $\mathcal{O}_2$ are generated in exactly the same way as for the second closed state $\mathcal{C}_2$. Given the current model $k$, it is proposed to change the number of closed states (from 1 to 2 or from 2 to 1) with probability 0.5 and change the number of open states otherwise. Hence $k$ may not change from *11* to *22* or from *21* to *12* in a single move.

Turning now to move (b), the proposed updates to the transition rates for a given model are multiplicative perturbations of their current values. If $k = \mathit{22}$ for example, we draw $z_i \sim N(0, \tau^2)$, $i = 1, \ldots, 6$ independently and set

$$\nu_{\mathrm{c}}' = \nu_{\mathrm{c}} \exp(z_1), \qquad \mu_{\mathrm{c}}' = \mu_{\mathrm{c}} \exp(z_2), \ldots, \qquad \nu_{\mathrm{o}}' = \nu_{\mathrm{o}} \exp(z_6).$$

The parameter $\tau$ takes the same value as in the model-changing move type. The acceptance probability for this Metropolis–Hastings joint update is $\min\{1, R\}$, where

$$\begin{aligned} R &= \text{full conditional ratio} \times \text{proposal ratio}, \\ &= \left( \frac{\nu_{\mathrm{c}}' \mu_{\mathrm{c}}' \lambda_{\mathrm{c}}' \lambda_{\mathrm{o}}' \mu_{\mathrm{o}}' \nu_{\mathrm{o}}'}{\nu_{\mathrm{c}} \mu_{\mathrm{c}} \lambda_{\mathrm{c}} \lambda_{\mathrm{o}} \mu_{\mathrm{o}} \nu_{\mathrm{o}}} \right)^{\alpha} \frac{p(\boldsymbol{x} \mid \Lambda', k = \mathit{22})}{p(\boldsymbol{x} \mid \Lambda, k = \mathit{22})} \\ &\quad \times \mathrm{e}^{-\beta(\nu_{\mathrm{c}}' + \mu_{\mathrm{c}}' + \lambda_{\mathrm{c}}' + \lambda_{\mathrm{o}}' + \mu_{\mathrm{o}}' + \nu_{\mathrm{o}}' - \nu_{\mathrm{c}} - \mu_{\mathrm{c}} - \lambda_{\mathrm{c}} - \lambda_{\mathrm{o}} - \mu_{\mathrm{o}} - \nu_{\mathrm{o}})/2}. \end{aligned}$$

If $k = \mathit{11}$, only two normal variates need to be drawn; if $k = \mathit{21}$ or *12*, four need to be drawn.

Move (c), updating the indicator function of channel openness $\boldsymbol{x}$, is a two-stage process. Firstly, we attempt to perturb the location of a randomly chosen discontinuity ('switch') of $\boldsymbol{x}$. Secondly, we make two independent random choices: between attempting to increase or decrease the number of switches of $\boldsymbol{x}$ and between attempting to change this number by one or two. Let the number of switches of $\boldsymbol{x}$ on the observation interval $[0, T]$ be $s$, and let their locations be $\boldsymbol{t} = (t_1, \ldots, t_s)$, where $0 < t_1 < \cdots < t_s < T$. In the first stage of updating $\boldsymbol{x}$ (the 'shift' move), we draw an existing switch location uniformly at random and propose to relocate it to a new point uniformly distributed on the interval bounded by the adjacent switches. No dimension changing is involved in the shift move. In the second stage of updating $\boldsymbol{x}$, we choose between 'type-1' and 'type-2' moves, and between 'birth' and 'death' steps within these move types. Type-1 moves attempt to create or delete a single switch at either end of $[0, T]$, while type-2 moves attempt to create or delete a single pair of adjacent switches, which corresponds to an open or closed channel sojourn. The design of the shift, type-1 and type-2 moves follows that of Hodgson (1999). We therefore omit further details, remarking only that since the distribution of the function $\boldsymbol{x}$ given the channel's kinetic parameters is now given by (2.3), the expressions for the acceptance probabilities of these moves are changed.

To update the conductance levels $(l_{\mathrm{c}}, l_{\mathrm{o}})$ in move (d) we draw independent standard normal variates $w_{\mathrm{c}}$ and $w_{\mathrm{o}}$ and propose new levels $(l_{\mathrm{c}}', l_{\mathrm{o}}')$, where $l_{\mathrm{c}}' = l_{\mathrm{c}} + \delta w_{\mathrm{c}}$, $l_{\mathrm{o}}' = l_{\mathrm{o}} + \delta w_{\mathrm{o}}$ and $\delta$ is a simulation parameter. The full conditional for the conductance levels is determined by (2.2):

$$p(l_{\mathrm{c}}, l_{\mathrm{o}} \mid \cdots) \propto p(l_{\mathrm{c}}) p(l_{\mathrm{o}}) p(\boldsymbol{y} \mid \boldsymbol{x}, l_{\mathrm{c}}, l_{\mathrm{o}}).$$

Here, '$|\cdots$' denotes conditioning on all other variables in the model. As the joint proposal distribution is symmetric, the update has the Metropolis acceptance probability $\min\{1, R\}$, where

$$\begin{aligned} R &= \text{full conditional ratio} \\ &= \text{prior ratio} \times \text{likelihood ratio}. \end{aligned}$$

The prior ratio is

$$\exp\left[-\frac{1}{2\kappa^2}\{(l'_{\mathrm{c}} - l_{\mathrm{c}})(l'_{\mathrm{c}} + l_{\mathrm{c}} - 2\theta) + (l'_{\mathrm{o}} - l_{\mathrm{o}})(l'_{\mathrm{o}} + l_{\mathrm{o}} - 2\theta)\}\right]$$

and the likelihood ratio is

$$\exp[-\tfrac{1}{2}\{Q(\boldsymbol{c}') - Q(\boldsymbol{c})\}],$$

where $\boldsymbol{c}'$ is the modified channel current implied by the proposal and

$$Q(\boldsymbol{c}) \equiv (F^{-1}\boldsymbol{y} - \boldsymbol{c})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(F^{-1}\boldsymbol{y} - \boldsymbol{c}).$$

In a single iteration of the sampler (henceforth termed a 'sweep') we implement moves (a)–(c) sequentially. The only random element of this formula is the choice of dimension-changing move with which to update $\boldsymbol{x}$. Move (d) is included in only every fiftieth cycle due to the computational expense of calculating the quadratic form $Q(\cdot)$. This subsampling, while appearing not to significantly slow convergence to the marginal posteriors for the levels, allows proportionately more time to be devoted to updating the other parameters.

## 3. Performance of the model discrimination methodology

### (*a*) *Data simulated from the maximal model*

A useful and easily implemented means of partly validating any MCMC sampler intended to simulate from a posterior distribution is to disconnect the data from the model and simulate from the prior. The output can then be compared with the chosen prior. We simulated from the prior

$$p(k, \Lambda, \boldsymbol{x}) = p(k)p(\Lambda \mid k)p(\boldsymbol{x} \mid \Lambda, k),$$

checking that the Markov model distribution $p(k)$, the priors on the transition rates $p(\Lambda \mid k)$ and the indicator function distribution $p(\boldsymbol{x} \mid \Lambda, k)$ were satisfactorily approximated by the empirical distributions. Having satisfied ourselves of the correctness of the code, we proceeded with simulating from the posterior.

We simulated a single channel dataset of $n = 2^{12} = 4096$ equispaced observation points. The 'truth' underlying these data was $k = 22$, $\nu_{\mathrm{c}} = \mu_{\mathrm{c}} = 0.05$, $\lambda_{\mathrm{c}} = 0.025$, $\lambda_{\mathrm{o}} = 0.04$ and $\mu_{\mathrm{o}} = \nu_{\mathrm{o}} = 0.02$. The choice of transition rates is arbitrary save for it yielding expected mean closed and open times (80 and 50, respectively) comparable to those in the dataset analysed in Hodgson (1999). These times (and all others in the paper) are expressed in observation index units, not physical time units. We added noise from an AR(15) process identical to that fitted to the noise in Hodgson, and then applied the same filter. Our choice of model parameters is $m = (0.25, 0.25, 0.25, 0.25)$, $\alpha = 1$, $\beta = 40$, corresponding to prior indifference

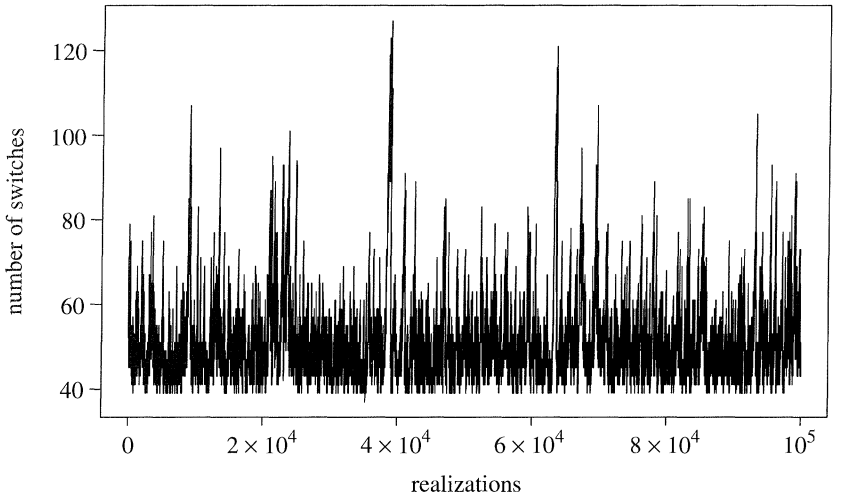


Figure 3. Trace of the number of switches.

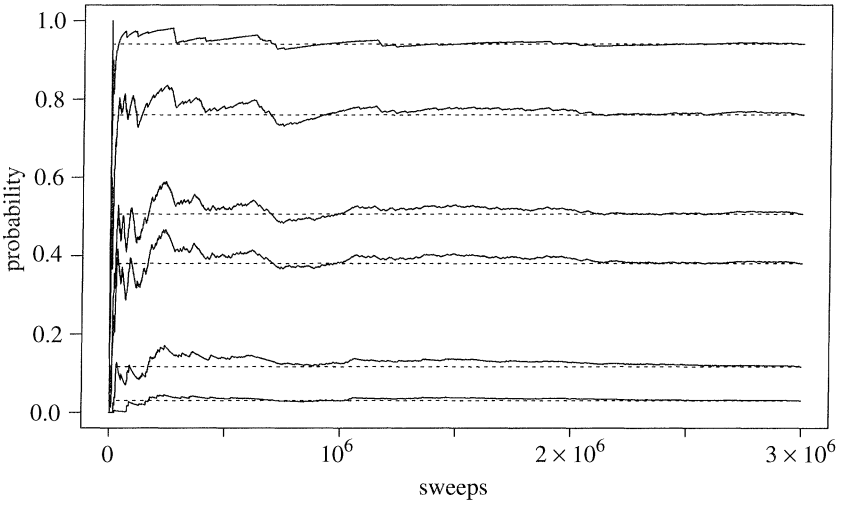


Figure 4. Quantiles diagnostic for the number of switches.

between the four Markov models and exponential priors on the transition rates. With these model parameters, we found from experimental runs that the choice $\rho_1 = \rho_2 = 2$, $\tau = 0.3$ gave a reasonable balance between 'boldness' and acceptance rate of the moves updating $k$; we do not claim that this choice is optimal. These settings of model and simulation parameters apply to all subsequent runs.

We ran the sampler for three million sweeps with no burn-in, recording realizations every 30th sweep purely due to storage constraints. This took 79 min on a 167 MHz Sun UltraSparc 1 workstation. Pilot runs had indicated that using aggressive optimization options on compiling the code cut the resulting computation time by around 60% without affecting numerical results. The trace of the number of switches of the indicator function $\boldsymbol{x}$ is shown in figure 3. The broken lines of figure 4 represent quantiles of the empirical distribution of the number of switches, while the solid lines

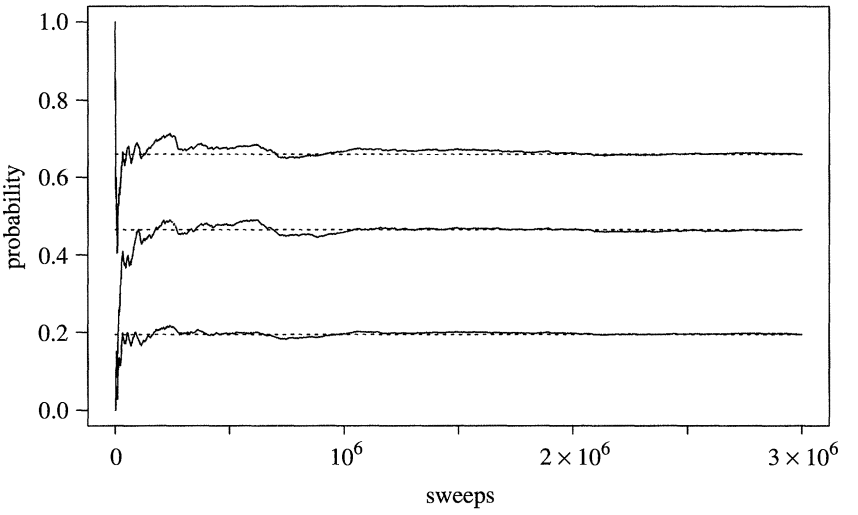


Figure 5. Quantiles diagnostic for the Markov model indicator.

trace the cumulative occupancy fractions below these quantiles. The equivalent convergence diagnostic for the Markov model indicator $k \in \{11, 21, 12, 22\}$ is shown in figure 5. Mixing over the number of switches appears to be satisfactory, and if we take the stabilization of the occupancy fractions as an indication of convergence then the quantiles plots suggest discarding as burn-in the initial 40% of the realizations, basing posterior inference on the remainder.

It would be desirable to make convincing inference about hidden Markov models for ion channels based only on weak prior information. Completely uninformative priors on the transition rates given the model, achieved by setting $\alpha = \beta = 0$ in our construction, are not possible: all the posterior probability would then be assigned to $k = 11$, the minimal model. This phenomenon is known as 'Lindley's paradox' (Lindley 1957; Bartlett 1957). Section 4 will present a sensitivity analysis with results which we believe justify our decision not to impose more diffuse priors on the transition rates (taking values of $\alpha$ less than 1 also increased time to convergence as estimated by the quantiles plots). Having chosen $\alpha$, the choice of $\beta$ expresses our prior belief about the frequency of channel openings and closings relative to the sampling frequency of the data. If all transition rates within all models took their prior mean values, the mean open and closed sojourn times would equal $\beta/\alpha$. In the light of Eisenberg and Levis's data (§ 3 *c*), we believe that the choice $\beta/\alpha = 40$ is reasonable but not critical.

Encouragingly, the strongest support in the empirical distribution of the model indicator $k$ (based on the final 60 000 recorded realizations) is for the true model, $k = 22$. The estimated posterior probabilities for models *11*, *21*, *12* and *22* are 0.193, 0.269, 0.191 and 0.347, respectively. The standard errors of these estimates were assessed by the batched means method and estimated to be 0.007, 0.009, 0.007 and 0.012, respectively. Quantiles plots for the transition rates conditional on $k$ suggested that the sampler mixes faster within than between $k$, and also faster within $k = 11$ than within more complicated models. An explanation, supported by the findings of § 3 *e*, is that when there are two open or closed states, the three-dimensional

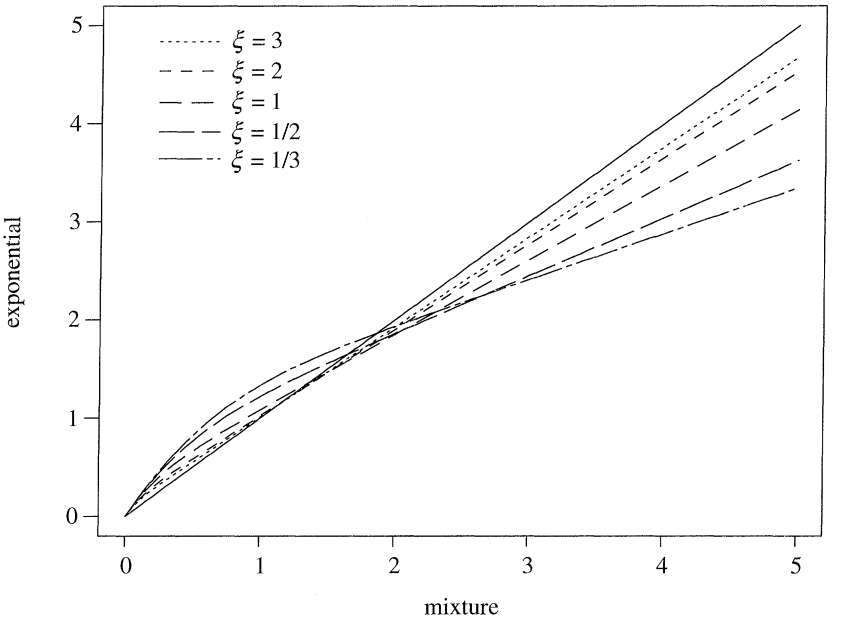


Figure 6. Departure from exponentiality for a range of values of $\xi$.

joint posterior for the relevant rates is diffuse and the dimension-preserving moves updating $\Lambda$ have a low acceptance rate.

### (*b*) *A seemingly paradoxical phenomenon*

We stated in the previous section that the choice of the six transition rates given $k = 22$ was constrained only by the loose requirement that they give realistic expected sojourn times. It is known (§ 2 *b*) that sojourn times in a two-state class have two-component exponential mixture distributions. Given that $k = 22$ is the most favoured model *a posteriori*, it is natural to ask how far these distributions 'depart' from exponentiality, and hence how remarkable this finding is.

Note that for both sojourn times, we set the true outer transition rates, that is $\nu_c$ and $\mu_c$ for closed times, to be equal. Denote by $\xi$ the ratio of these outer rates to the 'gateway' rate ($\lambda_c$ for closed times). Then $\xi = 2$ for closed times and $\xi = \frac{1}{2}$ for open times. Figure 6 illustrates the degree of departure from exponentiality for various values of $\xi$ by $Q$–$Q$ plots of the pure exponential against the two-component exponential mixture distributions. All distributions are scaled to have mean 1. It is clear that the mixture distributions assign higher probability density to high and low values and lower density to intermediate values, this trend being less marked with increasing $\xi$. Under this set-up, it is easily shown that the sojourn time distribution has coefficient of variation

$$\sqrt{1 + \frac{1}{2\xi}}.$$

Given our choice of transition rates, the theoretical open sojourn time distribution is markedly further from the exponential than that of the closed times, the respective

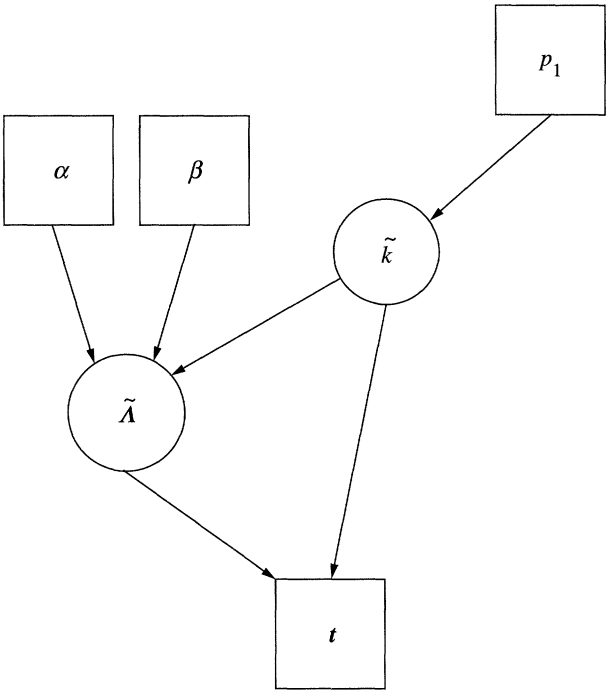


Figure 7. The DAG corresponding to the simplified model.

coefficients of variation being $\sqrt{2}$ and $\frac{1}{2}\sqrt{5}$. However, the previous section shows that $k = 21$, the model having two closed and one open state, is favoured *a posteriori* over $k = 12$, the model having a single closed and two open states. The obvious explanation is that sampling variation leads to a simulated channel signal having open times fitting a single exponential better than closed times do.

In an effort to verify or discredit this theory, we conducted off-line model discrimination experiments for the sample of completed realized closed times and then separately for the completed realized open times from the true signal underlying the dataset of the previous section. There were 27 in each sample. The associated DAG is in figure 7. In this reduced model, $\tilde{k} \in \{1, 2\}$ is the number of exponential components in the mixture distribution fitted to the sojourn times and $p_1$ is the prior probability of there being just one component. The vector $\tilde{\Lambda}$ contains the associated transition rates, and hence is one-dimensional if $\tilde{k} = 1$ and three-dimensional otherwise. The data $\boldsymbol{t}$ are the sojourn times. The priors on the rates given $\alpha$ and $\beta$, and their default values, are as before. We set $p_1 = 0.5$. It was trivial to write a program to compute the posterior under this simplified model as the necessary code is an extract from that already written. The results of these experiments were conclusive. The posterior probability $p(\tilde{k} = 2 \mid \boldsymbol{t})$ was 0.237 for the closed time sample but as high as 0.987 for the open time sample. The simple explanation of the surprising finding of the previous paragraph can be discounted.

For each $t \in \{0, 1, \ldots, T = 2^{12} - 1\}$, we may estimate the posterior probability of the channel being open at time $t$, and hence obtain a pointwise estimate of the posterior mean of $\boldsymbol{x}$, the indicator function of channel openness. Thresholding the mean function at the 50% level yields an estimate $\hat{\boldsymbol{x}}$ of the true indicator function.

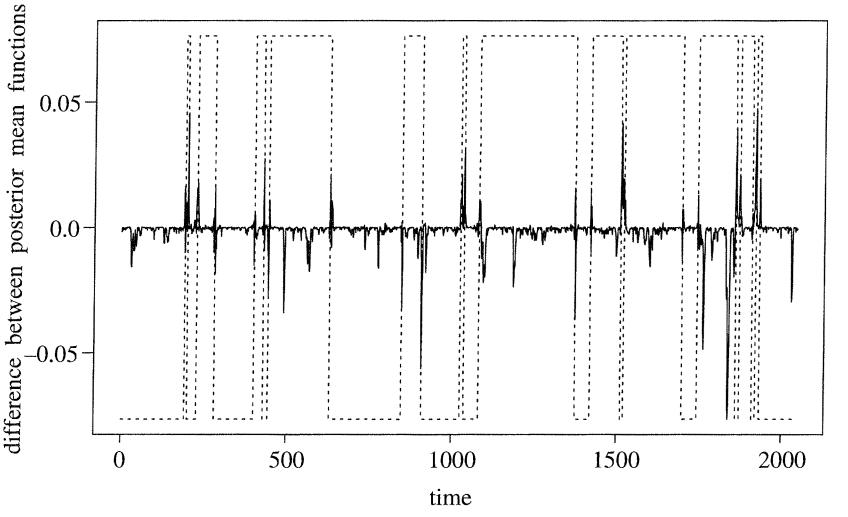


Figure 8. The difference between the posterior mean functions under models *21* and *12*.

We estimated the posterior mean functions conditional on the Markov model $k$ and thresholded them to obtain estimates of the true signal conditional on $k$. We found negligible dependence on $k$, the largest number of points of disagreement between any two estimates being 5 out of 4096. However, the threshold estimate is insensitive, suppressing much relevant information in the sample. The differences in the distributions of $\boldsymbol{x}$ under models *21* and *12* are illustrated in figure 8. For the first half of the channel record, we show as a solid curve the difference $\boldsymbol{d}$ between the posterior mean functions conditional on $k = 21$ and on $k = 12$. The dotted line represents the function $(2\hat{\boldsymbol{x}} - 1) \sup |\boldsymbol{d}_i|$. The unconditional estimate $\hat{\boldsymbol{x}}$ of the true indicator function and those conditional on $k$ are almost identical. The positive peaks of $\boldsymbol{d}$ are all in the neighbourhood of a 'switch' of $\hat{\boldsymbol{x}}$. In contrast, some negative peaks are located within lengthy closed and open sojourns. Negative peaks within closed sojourns correspond to times at which the evidence for a brief opening is stronger under model *12* than model *21*. As model *12* induces an open time distribution with coefficient of variation greater than 1, this is to be expected. Similarly, negative peaks within open sojourns correspond to times at which model *21* supports a brief closing more strongly than model *12*.

In summary, the small differences in the distributions of $\boldsymbol{x}$ under the competing models indicate some association between closed and open sojourn times. This association probably accounts for the disparity between the posterior model probabilities given the noisy data and those one would obtain given the noiseless signal.

### (*c*) *The Eisenberg and Levis data*

The data analysed by Hodgson (1999) and previously by Johnstone & Silverman (1997) are a segment of length $2^{12} = 4096$ from a time-series of $10^5$ data points simulated by the physiologists Eisenberg and Levis. The time-series represents the recorded current through a single ion channel with two conductance levels, and is intended to stimulate the development of new signal-processing techniques for this kind of data (Eisenberg 1994). Hodgson (1999) assigned an alternating renewal model

to the channel's aggregated process. It would be interesting to see which of our candidate Markov models are well supported by these data, and to compare our results with those of Hodgson.

The estimated posterior probabilities for models *11*–*22* given the Eisenberg and Levis data are 0.405, 0.119, 0.369 and 0.107, respectively. As in § 3 *a* these estimates are based on a run of three million sweeps of the reversible jump MCMC sampler, subsampling every thirtieth sweep and discarding the first 40% of the output. The standard errors of the estimates, again assessed by the batched means method, were estimated to be 0.013, 0.005, 0.012 and 0.005, respectively. It is $k = 11$, the minimal two-state model, and $k = 12$, the model with one closed and two open states, which together receive the bulk of the posterior support. It is unsurprising that there is more support for a second open state than a second closed state. Hodgson (1999) placed gamma distributions on the channel sojourn times. With independent Exp(1) priors on the closed and open time 'shape' parameters $s_0$ and $s_1$, their posterior medians were 1.88 and 0.72, respectively. Since the coefficient of variation of the gamma distribution with shape $s$ is $1/\sqrt{s}$, there is clear evidence that the open times have a higher coefficient of variation than the closed times. One would therefore anticipate the posterior distribution placing greater probability on the open times having the two-component exponential mixture distribution, which always has coefficient of variation greater than 1.

### (*d*) *Predictive distributions*

After every sampler sweep, the realized values of $(k, \Lambda)$ determine a realized closed sojourn time density $f_c(\cdot \mid k, \Lambda)$ and an open time density $f_o(\cdot \mid k, \Lambda)$. In the same manner as for other functionals of interest, averaging these densities across the MCMC output yields estimates of their posterior expectations. Averaging conditionally on a given value of $k$ gives estimates of $E\{f_c(\cdot \mid k, \Lambda) \mid k, \boldsymbol{y}\}$ and $E\{f_o(\cdot \mid k, \Lambda) \mid k, \boldsymbol{y}\}$, Bayesian predictive density estimates of the sojourn times under the given Markov model. Further averaging across $k$, weighting each model by its estimated posterior probability, gives estimates of $E\{f_c(\cdot \mid k, \Lambda) \mid \boldsymbol{y}\}$ and $E\{f_o(\cdot \mid k, \Lambda) \mid \boldsymbol{y}\}$, the unconditional predictive density estimates of the sojourn times.

We display both conditional and unconditional open time density estimates for the Eisenberg and Levis dataset in figure 9. The estimates conditional on Markov models *11* and *21*, in which there is just one open state, are very similar to each other, as are those conditional on models *12* and *22*, in which there are two open states. This suggests, contrary to the findings of § 3 *b*, that given the number of states in one class, the association between the transition rate(s) governing the distribution of sojourns in that class and the number of states in the other class is weak. The density estimates conditional on $k = 12$ and $k = 22$ have higher coefficients of variation than the estimates given $k = 11$ and $k = 21$. The overall (unconditional) predictive density estimate, given by the solid line in figure 9, is a compromise between the two classes of similar estimates. All the above remarks apply equally to the closed time density estimates (not shown), only with models *21* and *12* interchanged.

We may also compare predictive sojourn time distributions under the current set-up of a choice between hidden Markov models and under the alternating renewal model, with gamma-distributed sojourn times, of Hodgson (1999). Figure 10 shows

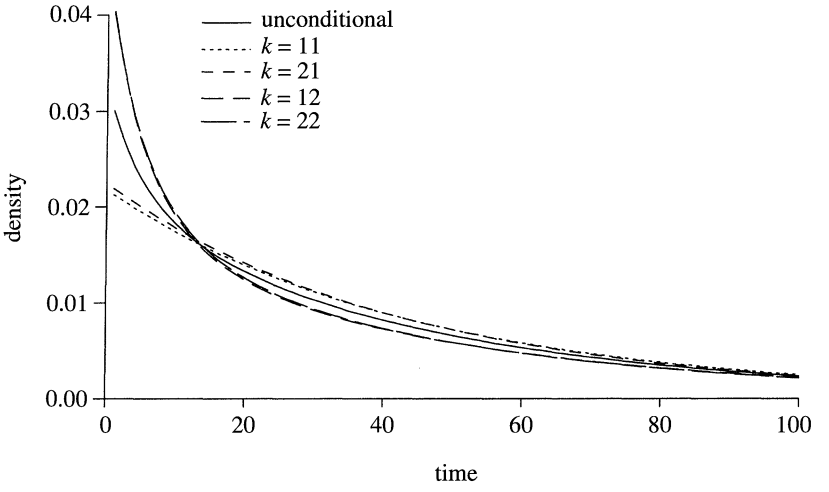


Figure 9. Bayesian predictive density estimates for open sojourn times.

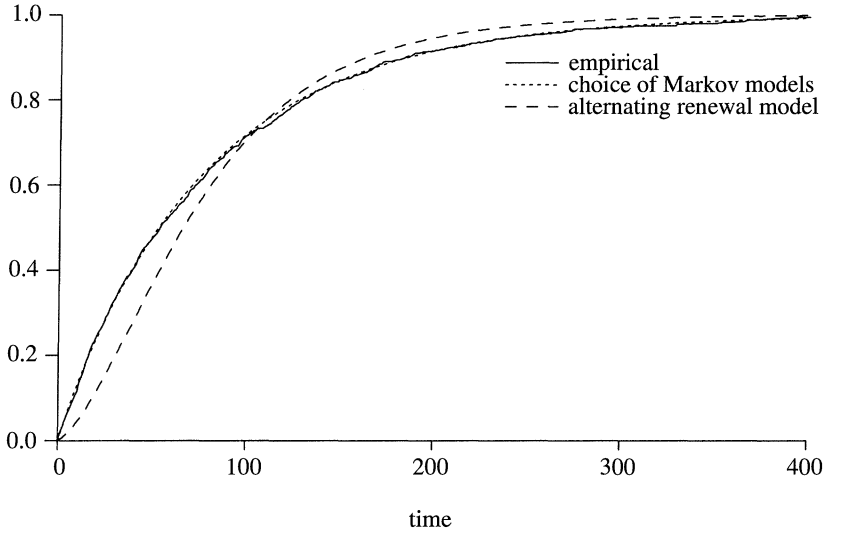


Figure 10. Predictive distribution function estimates for closed times.

the unconditional closed time distribution function estimate under the current model (dotted line) and the estimate under Hodgson's model (dashed line). Also shown, as a solid line, is the empirical distribution function of the 811 completed closed times from the entire 'true signal' record of Eisenberg and Levis. Clearly, the current model provides much the better fit. Under our model, the predictive distribution function (and its estimate) must be concave as it is a continuous mixture of exponential distribution functions. Since Hodgson places no prior constraint on the closed time shape parameter $s_0$, this need not be the case under his model. His predictive distribution has a lower coefficient of variation than ours, illustrating the fact that the bulk of the posterior support in his analysis was for values of $s_0$ greater than 1.

Table 1. *Summary of the posterior output from runs 1 and 2*

| parameter | run 1 | run 2 |
|---|---|---|
| $I\{k = 11\}$ | 0.409 (0.013) | 0.413 (0.011) |
| $I\{k = 21\}$ | 0.116 (0.005) | 0.125 (0.004) |
| $I\{k = 12\}$ | 0.373 (0.014) | 0.355 (0.009) |
| $I\{k = 22\}$ | 0.102 (0.004) | 0.107 (0.005) |
| $s$ | 68.1 (0.8) | 67.5 (0.5) |
| $\lambda_c \mid k = 11$ | 0.0125 (0.000 048) | 0.0125 (0.000 047) |
| $\lambda_o \mid k = 11$ | 0.0217 (0.000 077) | 0.0218 (0.000 087) |
| $\lambda_c \mid k = 12$ | 0.0139 (0.000 28) | 0.0136 (0.000 13) |
| $\lambda_o \mid k = 12$ | 0.051 (0.004) | 0.046 (0.002) |
| $\mu_o \mid k = 12$ | 0.055 (0.003) | 0.052 (0.002) |
| $\nu_o \mid k = 12$ | 0.068 (0.001) | 0.068 (0.001) |

### (*e*) *Assessing sampler mixing and convergence*

In § 3 *a* our decision to discard the first 40% of the MCMC output as burn-in was based on figures 3–5. None of these informal diagnostics give cause for grave concern, but it would be unwise to rely solely on them as evidence of satisfactory mixing and convergence to stationarity. There is a danger that such plots fail to spot serious difficulties, especially in the current context where there are two degrees of variable dimensionality. In this section we summarize the results of applying more formal diagnostics to our sampler output. For a fuller account of these experiments, see Hodgson (1998, § 5.4).

Table 1 compares posterior mean and standard error estimates (parenthesized) of some key parameters obtained from two independent sampler runs. In this table, $I\{E\}$ is the indicator variable associated with the event $E$ and $s$ is the number of switches of $\boldsymbol{x}$. To initialize $\boldsymbol{x}$ for the MCMC runs in §§ 3 *a* and 3 *c*, we thresholded the data at their median value and then imposed time interval omission by removing all apparent sojourns shorter than five time units. Changing the detection limit in the second stage of this algorithm changes the initial number of switches dramatically. With limits of three, five and nine time units, the estimates have 254, 44 and 16 switches, respectively; the true number is 55. In runs 1 and 2 in table 1, $\boldsymbol{x}$ was initialized by setting the detection limit to be 3 and 9, respectively. Despite their very different starting points, the results from these runs, both based on subsampling three million sweeps after every thirtieth sweep, are reassuringly similar.

We examine mixing over both the Markov model indicator $k$ and the number of switches $s$ of the indicator function, and then condition on the more likely values of $k$. This approach towards assessing the performance of our reversible jump algorithm is based on that of Richardson & Green (1997). Judging from traces of $s$ from runs 1 and 2 (not shown), the influence of the starting state does not extend beyond the first 5000 realizations (150 000 sweeps). Convergence of both the $s$ and the $k$ series is diagnosed by the test of Gelman & Rubin (1992). However, the tests of Geweke (1992) and Heidelberger & Welch (1983) are failed, except for the $k$ series from run 1, which passes the latter. We emphasize that none of these diagnostics

were designed with reversible jump output in mind, and their applicability to series of discrete variables is questionable. As for within-$k$ mixing and convergence, we restrict attention to the cases $k = 11$ and $k = 12$, attracting between them over 75% of the posterior support. Given $k = 11$, the posteriors for the transition rates $\lambda_c$ and $\lambda_o$ both have coefficient of variation less than 0.2. This indicates that the data are quite informative about the parameters of the two-state Markov model. The tests of Gelman & Rubin (1992) and Raftery & Lewis (1992) suggest rapid convergence, as does Heidelberger & Welch's (1983) test except when applied to the $\lambda_o$ series arising from run 2. Given $k = 12$, the rates governing open sojourn times ($\lambda_o, \mu_o$ and $\nu_o$) have considerably more diffuse posteriors than $\lambda_c$, the single rate governing closed times. The data carry relatively little information about these three transition rates given this model. Although both Gelman & Rubin's (1992) and Raftery & Lewis's (1992) tests diagnose convergence of the marginal distributions of all rates, both Heidelberger & Welch's (1983) and Geweke's (1992) tests are failed comprehensively by all rates for both runs. Hodgson (1998, § 5.4) provides a possible explanation for the negative results of some of the tests for within-$k$ convergence.

## 4. A sensitivity analysis

### (a) *Factors to be investigated*

In § 3 *a* we simulated a dataset based on $k = 22$. Given the settings we chose for $\boldsymbol{m}$, $\alpha$ and $\beta$, the true model receives the strongest, though by no means overwhelming, posterior support. However, this result has little significance without a qualitative understanding of some of the key factors influencing the posterior probabilities attached to the competing models. We alluded to Lindley's paradox: namely that $p(k = 11 \mid \boldsymbol{y})$ will tend to 1 as the priors on the transition rates become less and less informative. With only mildly informative priors ($\alpha = 1$, $\beta = 40$) one might anticipate that the simplest model will be 'easier' to identify than the others. The level of noise in the data and the length of the data segment are obvious factors of possible importance. Intuitively we would anticipate that lowering the noise level and lengthening the channel recording would result in a higher posterior probability on the correct value of $k$, but at the price of slower sampler mixing.

The fourth and final factor in the forthcoming study may be expressed informally as the ease with which the true model may be identified. In § 3 *b* we defined the ratio $\xi$ of the 'outer' rates to the 'gateway' rate in the case of the two-state class with equal outer rates. We saw (figure 6) that the exponential mixtures implied for the sojourn time distributions by the transition rates showed greater departure from exponentiality the lower the value of $\xi$. In this context, we would anticipate low values of $\xi$ for the open and closed times to facilitate identification of the true model $k = 22$. We cannot, however, calibrate this factor by a quantity whose definition is model dependent as the model is another factor. Calibrating this rather abstract factor is indeed problematic, and is discussed in the next section.

### (b) *Calibrating the final factor and the experimental set-up*

We shall study the effects on the posterior distribution of the model indicator $k$ of the four factors identified above. It is straightforward to define and set 'high' and 'low' levels for three of the factors. Factor A is the true model itself, $k = 11$

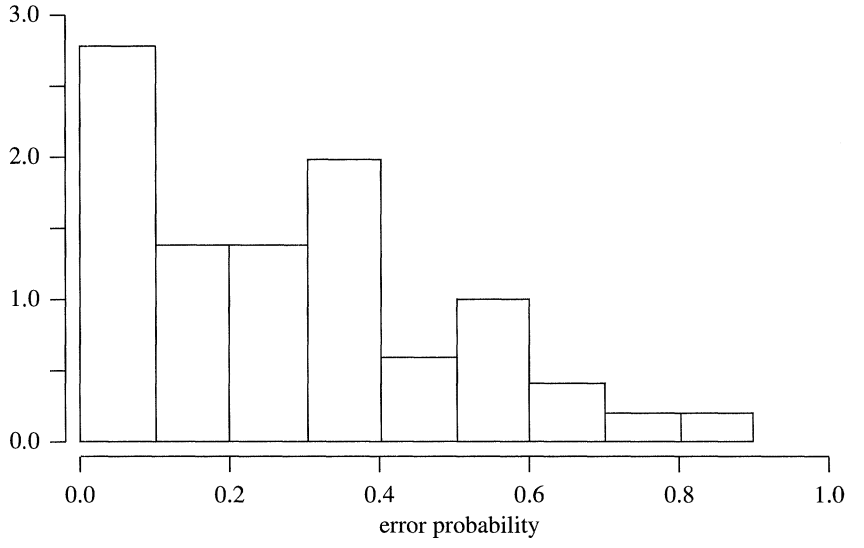


Figure 11. Histogram of error probabilities for 50 samples of 100 sojourn times.

being the high level and $k = 22$ being the low level. Factor B is the length of the dataset, with high level $n = 2^{13}$ and low level $n = 2^{12}$. Factor C is described loosely as a measure of how easily the true model is identified, with high level corresponding to cases in which the truth is easily identified. Finally factor D is the noise precision $\sigma^{-2}$, with high level equal to the precision of Eisenberg and Levis's data (Hodgson 1999) and low level equal to half the high level. By virtue of the definitions of high and low levels, each factor is expected *a priori* to have a positive main effect. Our first attempt to provide a satisfactory definition of factor C was as follows: given the true model and the transition rate(s) determining the (open or closed) sojourn time distribution, we generated a random sample of 100 such times. With sensible choices of rates, this sample size is of the same order as the number of switches in the indicator functions underlying the datasets to be analysed. We measured the infidelity of this sample to its underlying model by the posterior probability allocated to the 'incorrect' number of exponential components under the choice between pure exponential and two-component exponential mixture distributions, given prior indifference between the two (§ 3 *b* and figure 7). The posterior probability on the wrong number of components shall henceforth be called the *error probability.*

We generated 50 samples of 100 times from the closed sojourn time distribution implied by $k = 22$, $\nu_c = \mu_c = 0.0533$, $\lambda_c = 0.0267$ (so that $\xi = 2$). The resulting error probabilities are displayed in figure 11; their mean is 0.27. Clearly, the sampling variation in error probabilities for samples of this size is considerable. Similar degrees of dispersion were observed for the error probabilities derived from samples of 100 times from both the single exponential of the same mean (75) and exponential mixtures of the same mean but having different values of $\xi$. In the mixture case, mean error probabilities increase with $\xi$ as expected: the mean rises from 0.15 to 0.44 in going from $\xi = 1$ to $\xi = 3$. For a single sample of sojourn times, however, this effect is overwhelmed by the sampling variation. Hence it is

wholly unsatisfactory to calibrate factor C by mean error probabilities from samples of size 100 from the theoretical sojourn time distributions implied by the true model (level of factor A) and the transition rates. However, these findings suggest a related but superior way to assign levels to factor C. Rather than simulating samples of size 100 from the sojourn time distributions implied by the choice of Markov model and transition rates, suppose that for each combination of levels of factors A (true model) and B (record length) we simulated several noiseless channel records. We could then extract the completed sojourn times from these records and compute the error probabilities associated with these randomly sized samples. Choosing transition rates to give the same distribution for open and closed times, so that realized sojourn times need not be partitioned, simplifies matters. For each combination of A and B levels, we hand-pick two records from the set of those simulated, one yielding a 'high' error probability and the other a 'low' error probability. Having done this, we would have eight records, four yielding approximately the same 'high' error probability and four yielding approximately the same 'low' error probability. Given the high sampling variation of error probabilities associated with samples of size of order 100, it should be easy to establish two well-separated bands of error probability and simulate signals yielding error probabilities falling within one or other of these bands. The four records giving 'high' error probability will be used to generate noisy datasets corresponding to the low level of factor C (true model 'hard to identify'), and those giving 'low' error probability will generate datasets for which factor C takes its high level (true model 'easily identified'). When $k = 11$, we chose $\lambda_c = \lambda_o = 0.0133$ and when $k = 22$, we chose $\nu_c = \mu_c = \mu_o = \nu_o = 0.0533$, $\lambda_c = \lambda_o = 0.0267$. Table 2 summarizes our efforts to calibrate factor C by means of error probabilities. We follow the convention that + denotes the high level of a two-level factor and − denotes the low level. We have not endeavoured to match error probabilities corresponding to the low and high levels of factor C too closely as the forthcoming experiments are essentially qualitative.

We conduct a single replicate of the $2^4$ factorial design. To each of the eight ion-channel signals corresponding to all combinations of factors A, B and C, we added the autoregressive noise processes of variance corresponding to both levels of factor D. This yielded the 16 datasets to be analysed by our Markov model selection methodology.

### (*c*) *Results*

Summarized in table 3 are the results of the factorial experiment. As it is restricted to $[0, 1]$, we do not take the posterior probability assigned to the true model as our response variable. Instead we measure the evidence for the true model against the model corresponding to the 'incorrect' level of factor A by twice the logarithm of the Bayes factor $B_{A,-A}$. This statistic has the same scale as the classical likelihood ratio test statistic, and is given a rough calibration by Kass & Raftery (1995). Their guidelines are that values of $2 \log B_{12}$ greater than 2 provide 'positive' evidence for model 1 over model 2, and values greater than 6 provide 'strong' evidence.

The estimates of the main effects derived from table 3 are 5.05, −0.02, 1.64 and −0.09 for A, B, C and D, respectively. The interactions largest in magnitude are 5.20 for AB and −1.43 for ABC. Given that Bayes factors tend to favour parsimonious

Table 2. *Calibrating factor C by error probabilities*

| factor A | factor B | error probability | factor C |
|---|---|---|---|
| − | − | 0.375 | − |
| − | + | 0.346 | − |
| + | − | 0.278 | − |
| + | + | 0.317 | − |
| − | − | 0.143 | + |
| − | + | 0.157 | + |
| + | − | 0.141 | + |
| + | + | 0.176 | + |

Table 3. *Results of the factorial experiment*

| run | A | B | C | D | treatment combination | $2 \log B_{A,-A}$ |
|---|---|---|---|---|---|---|
| 1 | − | − | − | − | (1) | 0.67 |
| 2 | + | − | − | − | a | −1.26 |
| 3 | − | + | − | − | b | −5.05 |
| 4 | − | − | + | − | c | 0.67 |
| 5 | − | − | − | + | d | 0.14 |
| 6 | + | + | − | − | ab | 5.81 |
| 7 | + | − | + | − | ac | 2.42 |
| 8 | + | − | − | + | ad | −2.54 |
| 9 | − | + | + | − | bc | −3.50 |
| 10 | − | + | − | + | bd | −5.10 |
| 11 | − | − | + | + | cd | 1.45 |
| 12 | + | + | + | − | abc | 5.75 |
| 13 | + | + | − | + | abd | 5.93 |
| 14 | + | − | + | + | acd | 3.70 |
| 15 | − | + | + | + | bcd | −4.31 |
| 16 | + | + | + | + | abcd | 5.54 |

models when prior information on model parameters is weak, the strong positive effect of A is unsurprising. The coefficient of variation of a distribution, which for the gamma distribution with shape $\alpha$ is $1/\sqrt{\alpha}$, is a measure of its diffuseness. As $\alpha$ tends to zero, and so the priors on the transition rates of each Markov model become less informative, we would expect the positive effect of A to increase. The positive main effect of factor C is also as expected, but D has neither a significant main effect nor any significant interactions. The noise precision presumably must be altered by a factor greater than 2 before inferences about the underlying channel signals, given the low and high precision data generated from them, are significantly different. As the precision tends to zero, datasets become increasingly uninformative and so posterior beliefs increasingly reflect prior beliefs. In particular, the posterior odds in favour of one model over another will approach the prior odds, and so Bayes

factors will tend to 1. Since datasets are also less informative as they become shorter, this phenomenon of *prior shrinkage* would explain the AB interaction. When factor A is at its low level (so that the true model is $k = 22$), the Bayes factor $B_{A,-A}$ is typically less than 1, so that a shorter dataset increases $B_{A,-A}$ towards 1. Conversely, at its high level, A typically raises $B_{A,-A}$ above 1, so that a shorter dataset reduces $B_{A,-A}$ towards 1. Hence AB has a positive effect.

In summary, table 3 reveals that when B takes its high level, it is always far better to have A at its high level than at its low level. When B takes its low level, the difference in performance caused by changing the level of A is smaller, and it is better to have A at the same level as C.

## 5. Discussion and further work

This paper has described a study into the feasibility of carrying out general Markov model selection for ion-channel kinetics. It extends the work of Hodgson (1999) from an alternating renewal model to a choice of four simple Markov models. We do not, however, accomplish this by explicitly incorporating the latent Markov process into our Bayesian model (figure 2), but rather by integrating it out analytically (§§ 2 *a* and 2 *b*). This approach is convenient only for the simplest extensions of the basic two-state model, $k = 11$. Although our extensions are generally simpler than those appearing in the literature, model *21* was proposed by Castillo & Katz (1957) as the simplest mechanism for non-cooperative agonist action.

Of greatest interest in this study is the marginal posterior for $k$. General model selection for ion channels will be worth pursuing only if there is clear evidence of the ability to discriminate between these four models. In this respect, the factorial experiment of § 4 gives mixed results. When model *11* is the truth, it is favoured *a posteriori* over model *22* in six cases out of eight, yet when it is false, it is still favoured in four cases out of eight. Lindley's paradox manifests itself in the strong positive effect of factor A. Even with mildly informative exponential priors on the transition rates within each model, the bias towards the simpler model is clearly evident. Ion-channel data, at least in the quantities and subject to the noise levels of this paper, are insufficiently informative to make the ideal of Bayesian model determination with 'uninformative' priors viable. Factor C measures the practical identifiability of the true model given the noiseless channel signal. All our models are mathematically identifiable, in that distinct choices of transition rates will induce distinct distributions on the closed and open sojourn times. However, some choices of transition rates governing sojourn times in a two-state class, for example choosing $\xi$ to be large in the set-up of § 3 *b*, result in a sojourn time distribution virtually indistinguishable from a single exponential, so that the two-state class tends to degenerate into a single state. Since the parameter of a single exponential is just a scale factor, the evidence of departure from exponentiality in a sample of a given size from an exponential distribution is determined solely by sampling variation, and it is the prior on the transition rates given $k = 11$ which determines the distribution of the error probabilities in § 4 *b*. Hence factor C is rather one-sided.

Two important areas for future research are the application of the methodology to real single-channel data, and computational development. In analysing real data we will require the collaboration of physiologists to establish plausible Markov models and appropriate priors on the transition rates within each model, which may incorpo-

rate dependence between the rates. Of course, the records we simulate in this paper are orders of magnitude shorter than typical real records. Although it is not reflected in the main effect of factor B in our experiment, there is no doubt that increasing the amount of data will aid identification of the true model. However, implementation of the Bayesian method is then likely to be hampered by poorer mixing of the MCMC sampler. Perhaps by allowing different Markov models to apply to different subsections of the dataset, along the lines of the simulated tempering idea in Hodgson (1999), this difficulty could be overcome.

This research was undertaken while the first author was a PhD student at the Department of Mathematics, University of Bristol under the supervision of the second author. We thank Bob Eisenberg and Rick Levis for permission to use their data. The first author is grateful to Biomathematics and Statistics Scotland for allowing him time to revise the paper in accordance with the referees' comments.

## References

Ball, F. G. & Davies, S. S. 1995 Statistical inference for a two-state Markov model of a single-ion channel, incorporating time-interval omission. *J. R. Stat. Soc.* B **57**, 269–287.

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

Ball, F. G., Chen, A. & Sansom, M. S. P. 1992 Poisson sampling-based inference for single ion channel data with time interval omission. *Proc. R. Soc. Lond.* B **250**, 263–269.

Ball, F. G., Yeo, G. F., Milne, R. K., Edeson, R. O., Madsen, B. W. & Sansom, M. S. P. 1993 Single ion channel models incorporating aggregation and time interval omission. *Biophys. Jl* **64**, 357–374.

Ball, F. G., Cai, Y., Kadane, J. B. & O'Hagan, A. 1996 MCMC methods for discrete sojourn time ion channel data. Technical report, Statistics Group, University of Nottingham, UK.

Ball, F. G., Cai, Y., Kadane, J. B. & O'Hagan, A. 1997 Bayesian inference for ion channel gating mechanisms directly from single channel recordings, using Markov chain Monte Carlo. Technical report, Statistics Group, University of Nottingham, UK.

Bartlett, M. S. 1957 A comment on D. V. Lindley's statistical paradox. *Biometrika* **44**, 533–534.

Colquhoun, D. & Hawkes, A. G. 1981 On the stochastic properties of single ion channels. *Proc. R. Soc. Lond.* B **211**, 205–235.

Colquhoun, D. & Hawkes, A. G. 1982 On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Phil. Trans. R. Soc. Lond.* B **300**, 1–59.

del Castillo, J. & Katz, B. 1957 Interaction at end-plate receptors between different choline derivatives. *Proc. R. Soc. Lond.* B **146**, 369–381.

Edeson, R. O., Ball, F. G., Yeo, G. F., Milne, R. K. & Davies, S. S. 1994 Model properties underlying non-identifiability in single channel inference. *Proc. R. Soc. Lond.* B **255**, 21–29.

Eisenberg, R. 1994 Biological signals that need detection: currents through single membrane channels. In *Proc. 16th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Soc.* (ed. J. Norman & F. Sheppard), pp. 32a–33a.

Fredkin, D. R. & Rice, J. A. 1986 On aggregated Markov processes. *J. Appl. Prob.* **23**, 208–214.

Fredkin, D. R. & Rice, J. A. 1992*a* Bayesian restoration of single-channel patch clamp recordings. *Biometrics* **48**, 427–448.

Fredkin, D. R. & Rice, J. A. 1992*b* Maximum likelihood estimation and identification directly from single-channel recordings. *Proc. R. Soc. Lond.* B **249**, 125–132.

Gelman, A. & Rubin, D. B. 1992 Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–511 (with discussion).

Geweke, J. 1992 Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian statistics* (ed. J. M. Bernardo, A. F. M. Smith, A. P. Dawid & J. O. Berger), vol. 4, pp. 169–193. Oxford University Press.

Green, P. J. 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Hawkes, A. G., Jalali, A. & Colquhoun, D. 1992 Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Phil. Trans. R. Soc. Lond.* B **337**, 383–404.

Heidelberger, P. & Welch, P. D. 1983 Simulation run length control in the presence of an initial transient. *Operations Res.* **31**, 1109–1144.

Hodgson, M. E. A. 1998 Reversible jump Markov chain Monte Carlo and inference for ion channel data. PhD thesis, Department of Mathematics, University of Bristol, UK.

Hodgson, M. E. A. 1999 A Bayesian restoration of an ion channel signal. *J. R. Stat. Soc.* B **61**, 95–114.

Johnstone, I. M. & Silverman, B. W. 1997 Wavelet threshold estimators for data with correlated noise. *J. R. Stat. Soc.* B **59**, 319–351.

Kass, R. E. & Raftery, A. E. 1995 Bayes factors. *J. Am. Stat. Ass.* **90**, 773–795.

Kienker, P. 1989 Equivalence of aggregated Markov models of ion-channel gating. *Proc. R. Soc. Lond.* B **236**, 269–309.

Kijima, S. & Kijima, H. 1987 Statistical analysis of current from a membrane patch. I. A stochastic theory of a multi-channel system in the steady state. *J. Theor. Biol.* **128**, 423–434.

Lindley, D. V. 1957 A statistical paradox. *Biometrika* **44**, 187–192.

Magleby, K. L. & Weiss, D. S. 1990 Estimating kinetic parameters for single channels with simulation: a general method that resolves the missed events problem and accounts for noise. *Biophys. Jl* **58**, 1411–1426.

Raftery, A. E. & Lewis, S. M. 1992 How many iterations in the Gibbs sampler? In *Bayesian statistics* (ed. J. M. Bernardo, A. F. M. Smith, A. P. Dawid & J. O. Berger), vol. 4, pp. 763–773. Oxford University Press.

Richardson, S. & Green, P. J. 1997 On Bayesian analysis of mixtures with an unknown number of components. *J. R. Stat. Soc.* B **59**, 731–792 (with discussion).

Sakmann, B. & Neher, E. (eds) 1995 *Single-channel recording*, 2nd edn. New York. Plenum.