Bayesian Restoration of Ion Channel Records using Hidden Markov Models

Rafael Rosales,*† J. Alex Stark,† William J. Fitzgerald,† and Stephen B. Hladky*
*Pharmacology and †Engineering, University of Cambridge, Cambridge CB2 1QJ, United Kingdom

ABSTRACT Hidden Markov models have been used to restore recorded signals of single ion channels buried in background noise. Parameter estimation and signal restoration are usually carried out through likelihood maximization by using variants of the Baum–Welch forward–backward procedures. This paper presents an alternative approach for dealing with this inferential task. The inferences are made by using a combination of the framework provided by Bayesian statistics and numerical methods based on Markov chain Monte Carlo stochastic simulation. The reliability of this approach is tested by using synthetic signals of known characteristics. The expectations of the model parameters estimated here are close to those calculated using the Baum–Welch algorithm, but the present methods also yield estimates of their errors. Comparisons of the results of the Bayesian Markov Chain Monte Carlo approach with those obtained by filtering and thresholding demonstrate clearly the superiority of the new methods.

INTRODUCTION

The statistical analysis of single channel patch clamp records has been the subject of extensive research for over two decades. The most common scheme uses filtering and thresholding to recover the underlying process that is followed by the channel (see Colquhoun and Sigworth, 1995, for an extensive review). As a consequence, the original signal is reduced to a sequence of dwell times at a finite number of possible conductance levels separated by the thresholds. Empirical density estimates for the time spent at each level are then constructed as histograms, which are usually fitted by exponential mixtures. In the second stage of the analysis, the obtained mixtures characterized by a set of weights and rate constants are interpreted by modeling the channel dynamics as a finite state space aggregated Markov chain which is also homogeneous and has continuous (time) parameter (Colquhoun and Hawkes, 1982; Fredkin et al., 1985; Ball and Sansom, 1989). The inferences are directed toward the structure and specific value of the infinitesimal generator (the transition rate matrix) of the Markov chain. Finally, model selection between various alternatives is handled by using traditional penalized likelihood ratios. Ball and Rice (1992) provide a critical overview of the inferential difficulties of these approaches.

With only a few exceptions (Fredkin and Rice, 1992b), these methods rely heavily on the accuracy of the initial restoration step and are thus complicated when considering signal-to-noise ratios that are at the limit of filtering and thresholding. The pioneering work of Chung et al. (1990) introduced hidden Markov models with the initial aim of extracting information about current amplitudes and channel kinetics at low signal-to-noise ratios. More recently, hidden Markov models have also been considered by Venkataramanan et al. (1998) and Michalek and Timmer (1999) extending the applicability of the initial framework. Parameter estimation and signal restoration are usually carried out in an effective way by using variants of Baum's forward–backward procedures and reestimation formulas (Baum et al., 1970). In this case, estimates are represented by points in the parameter space that correspond to the coordinates of a local maximum of the likelihood.

A radically different approach for the treatment of hidden Markov models is provided by Bayesian statistics. In this paper, the initial restoration problem is addressed by modeling the observations with hidden Markov models. However, inferences are performed by using Bayesian statistics and an extension to a stochastic simulation method first proposed by Robert et al. (1993). Fredkin and Rice (1992a) also considered Bayesian restoration, but their approach is methodologically closer to Baum's maximization procedures. Bayesian inference based on stochastic search methods known as Markov Chain Monte Carlo have also been considered by Ball et al. (1999, 1997), Hodgson (1999), and Hodgson and Green (1999), however, the parameterization and their inferences are directed toward the transition rate matrix associated with a gating mechanism. The methods developed here, based on simpler procedures, are computationally efficient and supported by a solid probabilistic basis.

In this article, it is shown that the Bayesian methodology provides a useful way to deal with hidden Markov models for ion channels. Bayesian theory provides statistical grounds for the assessment of the uncertainty in all the unknowns considered in the modeling process. The next two sections present a brief overview of the problems posed by hidden Markov models and also their treatment from a Bayesian perspective. Results for various synthetic signals

Received for publication 30 March 2000 and in final form 12 December 2000.

Dr. Stark was formerly with the National Institute of Statistical Sciences, Research Triangle Park, NC 27709-4006 USA.

Address reprint requests to Rafael Rosales, IVIC, Departamento de Matematicas, Apartado 21827, Caracas 1020-A, Venezuela. Tel.: 58-2-5041412; Fax: 58-2-5041416; E-mail: rrosales@canchy.ivic.ve.

0006-3495/01/03/1088/16 $2.00

are shown in the following section, together with those obtained via standard Baum–Welch maximization. Comparisons of the Bayesian Markov Chain Monte Carlo approach against filtering and thresholding clearly demonstrate the new methods to be superior.

ION CHANNEL MODELING

Hidden Markov model

This section presents an overview of the hidden Markov model formalism that will be adopted and the notation to be used. The parameterization is consistent with the one considered by Chung et al. (1990) or Venkataramanan et al. (1998).

A hidden Markov model is defined by two interrelated stochastic processes. The first of these, representing the ion channel, is a first-order finite-state homogeneous Markov chain, $\{z^l: l \in \mathbb{R}_+\}$, on $\mathcal{C} = \{1, 2, \ldots, n\}$, with initial density λ and transition rate matrix Q. The continuous process is approximated by a discrete time version, $\{z^k: k \in T\}$, where $T = \{\delta, 2\delta, \ldots, \delta N\}$, with N as the total number of sample points and δ the sampling period of the acquisition system (hereafter we assume $\delta = 1$). In these terms $\{z^k\}$ is determined by a transition probability matrix, $A = \exp(Q\delta)$, with elements $a_{ij} = \mathbb{P}(z^k = j|z^{k-1} = i)$, $i, j \in \mathcal{C}$ (where $\mathbb{P}(X|Y)$ is the probability of X given Y). The second process is represented by the set $y = \{y^k: k \in T\}$ of independent and identically distributed random variables that constitute the observations. Each observation y^k is assumed to arise as a function of z^k, defined by a conditional density

$$\mathbb{P}(y^k|z^k = i, \theta) = d_i(y^k), \tag{1}$$

where θ is possibly a vector used to denote the parameters associated with a particular family d. Under standard normality assumptions, the function of the process is

$$d_i(y^k) \propto \exp[-(y^k - q_i)^2/2\sigma_i^2], \tag{2}$$

with q_i and σ_i^2 as the mean and the variance of the ith possible outcome associated with z^k. Assuming the number of conductance states, n, is known a priori, the model is completely specified by

$$\theta = (\lambda_i, a_{ij}, q_i, \sigma_i^2), \quad i, j \in \mathcal{C}. \tag{3}$$

This suggests an equivalence between states and conductances. However, this correspondence may be relaxed by imposing constraints on q such as $q_j = \xi$ for some value ξ, and $j \in J \subset \mathcal{C}$ if $z^k = j$, as will be shown in the section Class Dwell Times.

In this setting, each observation arises as the sum of two terms,

$$y^k = q_i + \sigma_i\eta^k, \quad \eta^k \overset{\text{iid}}{\sim} N(0, 1). \tag{4}$$

The first term at the right-hand side of Eq. 4 represents the amount of current flowing through the channel associated with the ith conductance level, whereas η^k is a normal distributed random variable with mean 0 and variance 1, representing the disturbance of the noise inherent in the recording apparatus. The symbol $\sim$ together with iid is used to denote independent and identically distributed.

The assumption of independent observations, explicitly stated as

$$\mathbb{P}(y^k|y^{k-1}, \ldots, y^1, z^k = i, \theta) = \mathbb{P}(y^k|z^k = i, \theta), \tag{5}$$

is an important simplification that leads to the analytical treatment of the model presented so far. It should be noted that the observations y^k are indirectly dependent on each other through the Markovian structure induced conditionally on $d_i(y^k)$ by the underlying process z^k.

Maximum likelihood approach

Let $Z = \{z: z^k = i; k \in T, i \in \mathcal{C}\}$ be the space of events constituted by the path realizations of the chain $\{z^k\}$, and z one of its elements, i.e., z constitutes a particular sequence of conductance states and Z the set of all possible sequences. Then, following Eq. 1 and the Markov property on z, the likelihood, which equals the probability of the observations conditional on the parameters, $L(\theta) = p(y|\theta)$, is given by

$$L(\theta) = \sum_{z\in Z} \mathbb{P}(y, z|\theta) = \sum_{z\in Z} \mathbb{P}(y|\theta, z)\mathbb{P}(z|\theta)$$

$$= \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_N=1}^{n} [\lambda_{i_1} d_{i_1}(y^1) a_{i_1 i_2} d_{i_2}(y^2) \cdots a_{i_{N-1} i_N} d_{i_N}(y^N)], \tag{6}$$

with $i_1, i_2, \ldots, i_N \in \mathcal{C}$. To simplify notation, let

$$h_i^1 = \lambda_i d_i(y^1), \quad h_{ji}^k = a_{ji} d_i(y^k), \tag{7}$$

then

$$L(\theta) = \sum_{i_1=1}^{n} \cdots \sum_{i_{N-1}=1}^{n} \sum_{i_N=1}^{n} h_{i_1}^1 \cdots h_{i_{N-1} i_N}^N. \tag{8}$$

Statistical inferences concerning θ generally take the form of point estimates, θ^*, obtained at a local maximum of $L(\theta)$,

$$\theta^* = \underset{\theta\in\Theta}{\operatorname{argmax}}\{L(\theta)\}, \tag{9}$$

with Θ usually a compact subset of some Euclidean space $\mathbb{R}^d$, more precisely $\Theta = \mathbb{R}^n \times \mathbb{R}_+^n \times [0, 1]^{n\times n+1}$. The notation "argmax$\{L(\,)\}$" refers to the point in Θ where $L(\,)$ attains a (local) maximum. In the particular case when d is normal (Eq. 2) and z is known, the required maximization is analytic. However, in patch clamp recordings, z is unknown

and inferences are directed toward both (z, θ). In this case, the maximization has to be performed numerically by considering n^N terms at Eq. 8. Evaluation of $L(\theta)$ becomes a critical issue, especially in the case of patch clamp records suitable for stationary analysis where $N > 10^6$. In the general framework, this problem was first considered by Baum et al. (1970), who derived an efficient solution reducing the evaluation to $O(n^2N)$. Baum's methods, known as the forward–backward algorithm and reestimation formulas, were first applied to patch clamp records by Chung et al. (1990).

BAYESIAN APPROACH

Bayesian analysis proceeds by considering the parameters θ as random variables on $\Theta \subset \mathbb{R}^d$. Inferences in this case require the definition of a joint density $p(y, \theta)$, which can be decomposed as $L(\theta)p(\theta)$, simply by following the definition of conditional probability. The density $p(\theta)$, known as the prior, represents our previous knowledge or belief about the value of the parameters before the observations have been examined. Once y is available, the uncertainties on θ are represented by a posterior density, $\pi(\theta|y)$, which follows from $p(y, \theta)$ and the application of Bayes Theorem as

$$\pi(\theta|y) = L(\theta)p(\theta) \Big/ \int L(\theta)p(\theta)\, d\theta. \tag{10}$$

Let $\mathbb{E}_\pi$ denote mathematical expectation with respect to the posterior and g any (square integrable) function of θ. Formal statistical inferences in this context often take the form of integrals

$$\mathbb{E}_\pi[g(\theta)] = \int g(\theta)\pi(\theta|y)\, d\theta, \tag{11}$$

and are thus based on the support of $\pi(\theta|y)$ rather than on a single point estimate. This ensures the use of probabilities for events $\theta \in C$ $(C \subseteq \Theta)$, in contrast to classical confidence procedures. An account of Bayesian theory can be found in Bernardo and Smith (1994).

Although there are clear theoretical advantages of the Bayesian approach, there are difficulties that must be overcome before they can be realized in practice. The integrals in Eqs. 10 and 11 are not analytical for the current model. This fact arises as a consequence of the structure imposed by the likelihood in Eq. 8, regardless of the specific form of $g(\theta)$ and $p(\theta)$.

Markov chain Monte Carlo for hidden Markov models

A solution to the integration task posed by Eqs. 10 and 11 is provided by Monte Carlo approximation together with Markov chain sampling. First, an ergodic Markov chain $\theta^{(0)}$, $\theta^{(1)}, \ldots$, is constructed on Θ, such that its invariant distribution corresponds to the hidden Markov model posterior $\pi(\theta|y)$. It should be made clear that this process represents a different Markov chain than that followed by the channel on Z.

More precisely, let $K_{\theta m} = \mathbb{P}(\theta^{(m)} \in B|\theta^{(m-1)} = x)$ denote the step transition probability from $\theta^{(m-1)}$ to $\theta^{(m)}$ for any $m = 1, 2, \ldots$, and $\pi^0 = \mathbb{P}(\theta^{(0)} \in A)$ an initial density for $A, B \subset \Theta$, and $x \in \Theta$. Under mild regularity conditions that ensure ergodicity

$$\lim_{m\to\infty}[K_\theta^m \pi^0] \to_d \pi(\theta|y), \tag{12}$$

for almost all initial values of $\theta^{(0)}$. Here $K_\theta^m = \mathbb{P}(\theta^{(m)} \in B|\theta^{(0)} = x)$ denotes the mth iterate of K_θ^1, that is, $K_\theta^m = K_{\theta m}K_\theta^{m-1}$, and $\to_d$ is used to denote convergence in distribution (see Billingsley, 1968). Thus, after an initial relaxation period b, successive iterations generate a sequence of samples $\theta^{(b+1)}, \ldots, \theta^{(M)}$ that are approximately distributed according to $\pi(\theta|y)$. This sequence may then be used to calculate averages that approximate the desired expectations under the posterior target density

$$\mathbb{E}_\pi[g(\theta)] \approx \frac{1}{M-b} \sum_{m=b+1}^{M} g(\theta^{(m)}), \tag{13}$$

constituting a Monte Carlo estimate of the integrals in Eq. 11.

The combination of these techniques has been used to tackle complex multidimensional problems, proving to be an effective tool where standard frequentist methods have failed. The theoretical background needed for the use of these general state chains is described in Nummelin (1984) or Meyn and Tweedie (1993); whereas their application to Bayesian statistics is presented among others by Tierney (1994), Besag et al. (1995), or Robert and Casella (1999). J. A. Stark, R. Rosales, W. J. Fitzgerald, and S. B. Hladky (manuscript in preparation) also present a different Markov chain Monte Carlo approach for the single channel restoration problem in terms of a change point model. The particular chain that will be used to generate approximate samples from the posterior in Eq. 10 is presented in detail in the next section.

Gibbs sampler

Let $p(\theta_i|\theta_{-i}, y)$ be the (full) conditional densities for each of the components θ_i, given values of the other components $\theta_{-i} = \{\theta_j : j \neq i\}$. These densities are uniquely determined by a particular posterior (Besag, 1974) and constitute the building blocks of a Markov chain Monte Carlo method known as the Gibbs sampler (Gelfand and Smith, 1990). This sampler proceeds as follows. Given an arbitrary set of

starting values $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_r^{(0)})$, draw a sample of $\theta_1^{(1)}$ from $p(\theta_1|\theta_{-1}^{(0)}, y)$, i.e.,

$$\theta_1^{(1)} \sim p(\theta_1|\theta_2^{(0)}, \ldots, \theta_r^{(0)}, y)$$

then

$$\begin{aligned} \theta_2^{(1)} &\sim p(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_r^{(0)}, y) \\ &\ldots \\ \theta_r^{(1)} &\sim p(\theta_r|\theta_1^{(1)}, \ldots, \theta_{r-1}^{(1)}, y). \end{aligned} \tag{14}$$

This completes one iteration of the sampling scheme $\theta \sim \{p(\theta_i|\theta_{-i}, y)\}$, and also a transition from $\theta^{(0)}$ to $\theta^{(1)}$. The sampling of a high dimensional vector θ has been replaced by the sampling of lower dimensional components, which is one of the key features of this algorithm.

The decomposition of $\pi(\theta|y)$ into its conditionals is rather involved because of the n^N terms in Eq. 8. Consider instead the joint $\pi(\theta, z|y)$, which leads to the following decomposition and sampling scheme:

$$(\theta, z) \sim \{p(\theta_i|\theta_{-i}, y, z), p(z|\theta, y)\}. \tag{15}$$

In this case, under the appropriate choice of priors, all the densities $p(\theta_i|\theta_{-i}, y, z)$ are analytical because they only involve one of the possible realizations in Z. This simplification was first considered by Robert et al. (1993), and by Diebolt and Robert (1994) in the general mixture distribution context. The key of this conditional decomposition resides in considering z as a random variable, much in the same way as θ, instead of integrating it out. This scheme leads to rigorous theoretical facts on the convergence of the sampler to be discussed in the section entitled Convergence.

Following Eq. 15, draws have to be produced from the full conditional for the hidden process

$$p(z|y, \theta) = h_{i_1}^1 \cdots h_{i_{N-1}i_N}^N / L(\theta) \tag{16}$$

for any $i_1, \ldots, i_{N-1}, i_N \in \mathcal{C}$. This can be achieved in different ways. A first alternative consists of the method proposed by Carter and Kohn (1994), which constitutes a stochastic version of the forward–backward algorithm. This follows by noting that $p(z|y, \theta)$ can be decomposed as

$$p(z|y, \theta) = p(z^N|y, \theta) \prod_{k=1}^{N-1} p(z^k|z^{k+1}, y_k, \theta), \tag{17}$$

with $y_k = y^1, \ldots, y^k$. Given z^{k+1}, $p(z^k|z^{k+1}, y_k, \theta)$ is a discrete distribution, which suggests the following sampling strategy. For $k = 2, \ldots, N$ and $i \in \mathcal{C}$, compute and store the optimal filter $p(z^k = i|y_k, \theta)$, then sample z^N from $p(z^N|y, \theta)$, and, for $k = N - 1, \ldots, 1$, sample z^k from $p(z^k|z^{k+1}, y_k, \theta)$, where, if $z^{k+1} = l$, for $i \in \mathcal{C}$,

$$p(z^k = i|z^{k+1} = l, y_k, \theta) = \frac{a_{il} p(z^k = i|y_k, \theta)}{\sum_{j \in \mathcal{C}} a_{jl} p(z^k = j|y_k, \theta)}. \tag{18}$$

This strategy updates (θ, z) by following the decomposition given by Eq. 15. A second option follows by considering the full conditional densities

$$p(z^k|z^{-k}, y, \theta) = p(z^k|z^{k-1}, z^{k+1}, y, \theta), \tag{19}$$

which follow from the Markov property on z. For $k = 2, \ldots, N - 1$ and for any $i, j, r \in \mathcal{C}$, these densities are

$$p(z^1 = i|y, \theta) \propto h_i^1,$$

$$p(z^k = i|z^{k-1} = j, z^{k+1} = r, y, \theta) \propto h_{ji}^k a_{ir},$$

and, for $k = N$,

$$p(z^N = i|z^{N-1} = j, \theta, y) \propto h_{ji}^N. \tag{20}$$

This sampling scheme leads to the conditional decomposition,

$$(\theta, z) \sim \{p(\theta_i|\theta_{-i}, y, z), p(z^k|y, \theta, z^{-k})\}, \tag{21}$$

where i denotes the index over the components of θ (Eq. 3). Note that the first option generates a realization of z directly from the conditional $p(z|\theta, y)$ by using a single Gibbs component. The second uses N different components and presents a more complicated correlation structure, since each z^k depends on both z^{k-1} and z^{k+1}, which leads to a slower mixing chain (see Liu et al., 1995, for the effects of correlations among sampled components). However, by adopting the latter option, it is possible to avoid any part of the forward–backward filtering procedures used in the other two cases. This fact is particularly useful in the treatment of dependent observations (R. Rosales, J. A. Stark, W. J. Fitzgerald, and S. B. Hladky, manuscript in preparation).

Priors

In this paper, we consider proper conjugate priors that attempt to be weakly informative on Θ. A prior $p(\theta)$ that belongs to a parametric family $\mathfrak{F}$ is conjugate to a given likelihood if the resulting posterior is also from this family. In this case, the required conditionals for θ also belong to $\mathfrak{F}$ (see, for example, Bernardo and Smith, 1994). This prior has been chosen for tractability. It should be noted that hidden Markov models do not allow the use of improper (noninformative) priors for λ or for each row of A (Diebolt and Robert, 1994).

Let a_i denote the ith row of A. Then, by assuming independence between priors over states and conditional independence between priors over parameters,

$$\begin{aligned} p(\theta) &= \prod_{i \in \mathcal{C}} p(\lambda, a_i, q_i, \sigma_i^2) \\ &= p(\lambda) \prod_{i \in \mathcal{C}} p(a_i) p(q_i) p(\sigma_i^2). \end{aligned} \tag{22}$$

In the case of a univariate normal likelihood for η^k and a multinomial for z, the conjugate families for θ are given by the following densities: $\lambda \sim D(b_1, \ldots, b_n)$, $a_i \sim D(e_{i1}, \ldots, e_{in})$, $q_i \sim N(m_i, s_i^2)$ and $\sigma_i^2 \sim IG(u_i, w_i)$. *IG* is used for an inverted gamma density with shape and scale u_i, w_i and hence with mean $w_i/(u_i - 1)$ whenever $u_i > 1$ and variance $w_i^2/[(u_i - 1)^2(u_i - 2)]$ if $u_i > 2$. The symbol D is used for a Dirichlet density with parameter vector x with either $x = b$ or $x = e_i$. In this case, $x_j > 0$, and the jth mean and variance for the jth component are x_j/x_0 and $x_j(x_0 - x_j)/[x_0^2(x_0 + 1)]$ where $x_0 = \Sigma_j x_j$.

The parameters m_i, s_i^2, u_i and w_i are regarded as constants that may be calculated according to the observed data range. Let the range extend from 0 to R, with R measured in physical units of current. Then, setting

$$m_i = R/2, \quad s_i^2 = R^2 \tag{23}$$

for all $i \in \mathcal{C}$, tends to produce a prior for q_i that is relatively flat over the interval specified by R. The constants u_i and w_i may be specified by setting the mean of the inverse gamma density equal to the observed variance σ_y^2 (if available), and also the variance equal to some reasonable multiple of the data range $\tilde{R}$ (i.e., $\tilde{R} = 4R$). Solving the resulting system for u_i and w_i gives

$$u_i = [(\sigma_y^2)^2 + 2\tilde{R}]/\tilde{R} \quad w_i = [(\sigma_y^2)^4 + \sigma_y^2\tilde{R}]/\tilde{R}. \tag{24}$$

Noninformative priors for the initial density and for the transition probabilities are obtained at the limit when $b_i \to 0$ and $e_{ij} \to 0$. However, care must be taken because $b_i = e_{ij} = 0$ for any $i, j \in \mathcal{C}$ results in an improper prior. Sensible choices representing weak prior information are $e_{ij} \in [0.1, 1]$ and $b_i = 10^{-6}$. In terms of the single channel record, the value for e_{ij} constitutes a statement about the frequency of the transition $i \to j$, whereas the choice for b_i represents the prior belief that the channel will start at the ith level. Due to the structure induced by the Dirichlet density for each row, informative priors for A obeying reversibility constraints imposed by Kolmogorov's criteria could also be used as those suggested for Q by Ball et al. (1999).

The full conditionals necessary to implement the Gibbs sampler follow immediately once the priors have been specified. Samples for all these densities where generated by implementing Eqs. 14, 18, and 20, and following the methods described in Fishman (1996) and Gelman et al. (1995). The explicit form for $p(\theta_i|\theta_{-i}, y, z)$ is derived in the Appendix.

Convergence

Successive iterations from the sampling scheme at Eq. 15 (or Eq. 21) produce the sequence $(\theta^{(1)}, z^{(1)}, \ldots, (\theta^{(M)}, z^{(M)})$, which is a Markov chain on $\Theta \times Z$. Let $\pi(\theta|y)$ be the marginal of $\pi(\theta, z|y)$, that is,

$$\pi(\theta|y) = \sum_{z \in Z} \pi(\theta, z|y), \tag{25}$$

and $\pi^m(\theta|y)$ the density generated at the mth Gibbs step. Let also $f(z|y)$ and $f^m(z|y)$, be the corresponding densities for z. Then,

$$\pi^m(\theta|y) = \sum_{z \in Z} \pi(\theta|z, y) f^m(z|y) \tag{26}$$

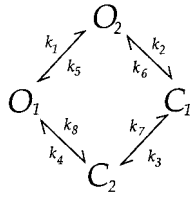
represents a relation between the marginals at each step that enables transfer of the convergence properties of $\{z^{(m)}\}$ to $\{\theta^{(m)}\}$. This fact, first elicited in Diebolt and Robert (1994) and Robert (1995), is known under the term "duality principle." Because $\{z^{(m)}\}$ is a regular discrete state space Markov chain, then it is stationary and its equilibrium distribution $f(z|y)$ is unique, that is, $\{z^{(m)}\}$ is ergodic. Although $\{\theta^{(m)}\}$ is not a Markov chain, it is straightforward to show that its stationary distribution is the marginal $\pi(\theta|y)$. In this case, due to Eq. 26, the following results hold:

a. For any starting value $\theta^{(0)}$, the process $\{\theta^{(m)}\}$ converges uniformly to $\pi(\theta|y)$ at a geometric rate.
b. For any real function g, provided $\mathbb{E}_\pi[|g(\theta)|] < \infty$, and for any stating point $\theta^{(0)}$, $\mathbb{E}_{\pi^m}[g(\theta)]$ converges uniformly at a geometric rate to $\mathbb{E}_\pi[g(\theta)]$.
c. Because $\{z^{(m)}\}$ is a stationary Markov chain, it is also φ-mixing, and hence, so is $\{\theta^{(m)}\}$. In this case a Central limit Theorem for any real function $g(\theta)$ holds, provided $\mathbb{E}_\pi[|g(\theta)|^2] < \infty$.

Proofs for a and b can be found in Theorem 1, (i)–(ii) in Robert et al. (1993). A proof for c is provided by Theorem 1 (iii) in Robert et al. (1993) and Theorem 20.1 in Billingsley (1968). An expression for the rate of convergence in a is outlined in the Appendix.

RESULTS

This section presents results obtained with the Gibbs sampler on synthetic signals of known characteristics. The following example was designed to present a challenge by providing a combination of brief events together with closely located subconductance levels. A four-state Markov chain with reversible cyclic mechanism



SCHEME 1

was considered. The symbols k_i, $i = 1, \ldots, 8$ are used to denote the transition rate constants among the states. By

setting $k_1 = 6.183$, $k_2 = 0.454$, $k_3 = 2.697$, $k_4 = 1.665$, $k_5 = 0.446$, $k_6 = 13.163$, $k_7 = 0.182$, and $k_8 = 11.812$ and the sampling period to $\delta = 0.005$, the transition probability matrix, $A = \exp(Q\delta)$, becomes

$$\begin{array}{c} \\ C_1 \\ C_2 \\ O_2 \\ O_1 \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & O_2 & O_1 \end{array} \\ \begin{bmatrix} 0.924 & 0.0129 & 0.063 & 0.000122 \\ 0.000870 & 0.991 & 0.000153 & 0.00793 \\ 0.00218 & 0.785e-4 & 0.996 & 0.00213 \\ 0.584e-4 & 0.0562 & 0.0295 & 0.914 \end{bmatrix} \end{array}$$

and the expected life times at each state are spread over two orders of magnitude, i.e., $\tau_{c_1} = 0.0551$, $\tau_{c_2} = 1.351$, $\tau_{o_1} = 0.0629$, and $\tau_{o_2} = 0.5519$. This mechanism generates long sojourns in an "open" state or in a "closed" state, which are then interrupted by brief transitions to a state of the opposite type, that is, $C_2 \to O_1 \to C_2$ or $O_2 \to C_1 \to O_2$. The conductances were set to $q_{c_1} = 0.07$, $q_{c_2} = 0.0$, $q_{o_1} = 0.14$, and $q_{o_2} = 0.21$, and white noise of mean 0.0 and standard deviation 0.1 was added to a million-point realization of z. The allowed set of transitions, together with the labels assumed for the conductances, can be seen in Fig. 1 *C*. The corresponding noisy segment is presented in Fig. 1 *A*.

These data were analyzed with the Gibbs sampler specified by Eq. 15, more precisely by using stochastic forward–backward updates (Eq. 18) with four conductance levels and 2000 iterations. The priors where specified as: $m_i = 0.36$, $s_i^2 = 0.25$, $u_i = 2$, $w_i = 1$, $e_{ij} = 0.5$, and $b_i = 1 \times 10^{-6}$, for all $i, j \in \mathcal{C}$. The sampler was started at arbitrary initial values for θ and z, namely: $q_i^{(0)} = 0.36$, $\sigma_i^{2(0)} = 0.5$, $a_{i=j}^{(0)} = 0.99$, and $a_{i\neq j}^{(0)} = 0.003$, and $\lambda_i^{(0)} = 0.25$. The initial realization, $z^{(0)}$, was obtained by sampling from the values of $A^{(0)}$ and $\lambda^{(0)}$.

General output for z and θ

The traces at Fig. 1 *B* show detail of a few of the sampled realizations for z. Long-lived sojourns with high signal-to-noise ratios are usually well located, whereas brief instances with poor resolution present some degree of uncertainty. An example of the latter is given by the short sojourns at C_1, arising from transitions from O_2 such as the one near the sample 447 in Fig. 1 *B*. Rather than showing all the samples for z, it is desirable to present a summary of their statistical properties. Figure 1 *D* presents the most frequently visited state at each time point. The actual location of the levels is assigned by taking the ergodic average for the sampled values of q, i.e., by setting $g(\theta_i) = q_i$ in Eq. 13. Different summaries for $z^{(m)}$, such as the Monte Carlo mean $\pm$ standard deviation or their Rao-Blackwellized versions (Robert and Casella, 1999) are also possible. The proportion of misclassified points computed from the most frequently visited state is 0.0315, with a large majority of these accounted for by fitted transitions that slightly precede or follow the corresponding transition in the ideal record.

The plots in Fig. 2 show samples of few components of θ against the number of iterations. It appears that the sampler reaches a stationary state within approximately 100 iterations. Further iterations (see Fig. 2 *D*, and further $>10^6$, not shown) do not seem to diverge from these values. The ergodic averages for some components of θ obtained from the last 1000 iterations are (standard deviation in parenthe-

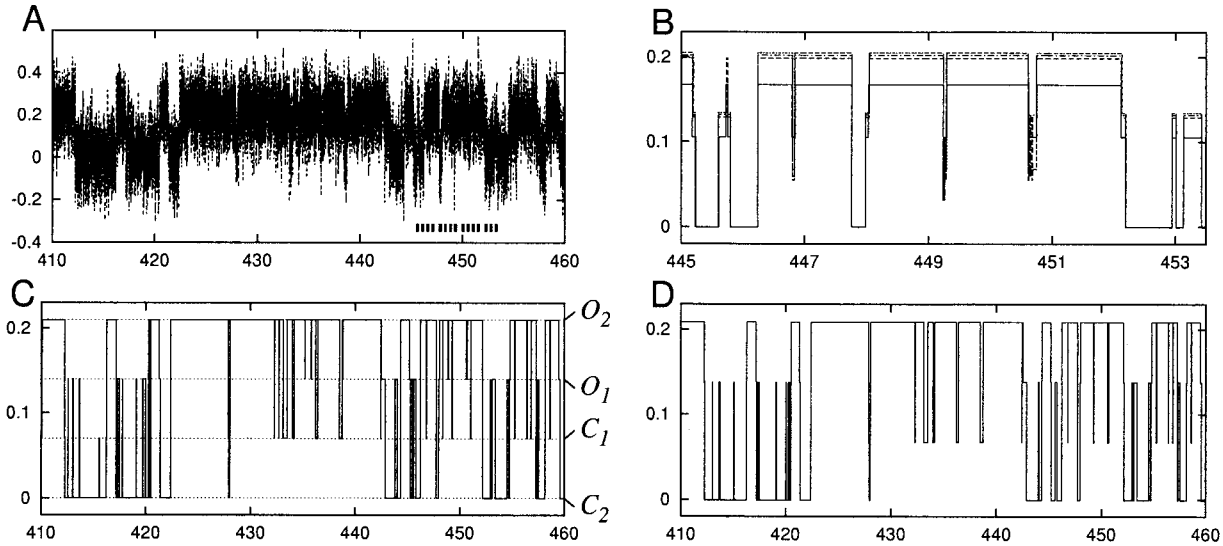


FIGURE 1 (*A*) A segment of the synthetic data generated by the mechanism in the Results section. (*B*) An ensemble of few sampled realizations, $z^{(m)}$, for m: 22, 24, 25, 47, 49 ($\times 10^2$), for the range marked in (*A*). (*C*) The ideal (noiseless, unfiltered) trace underlying the data segment shown in (*A*) indicating the labeling of the states, C_1, C_2, O_1, O_2. (*D*) The most frequently visited state during the last 1000 iterations of the Gibbs sampler.

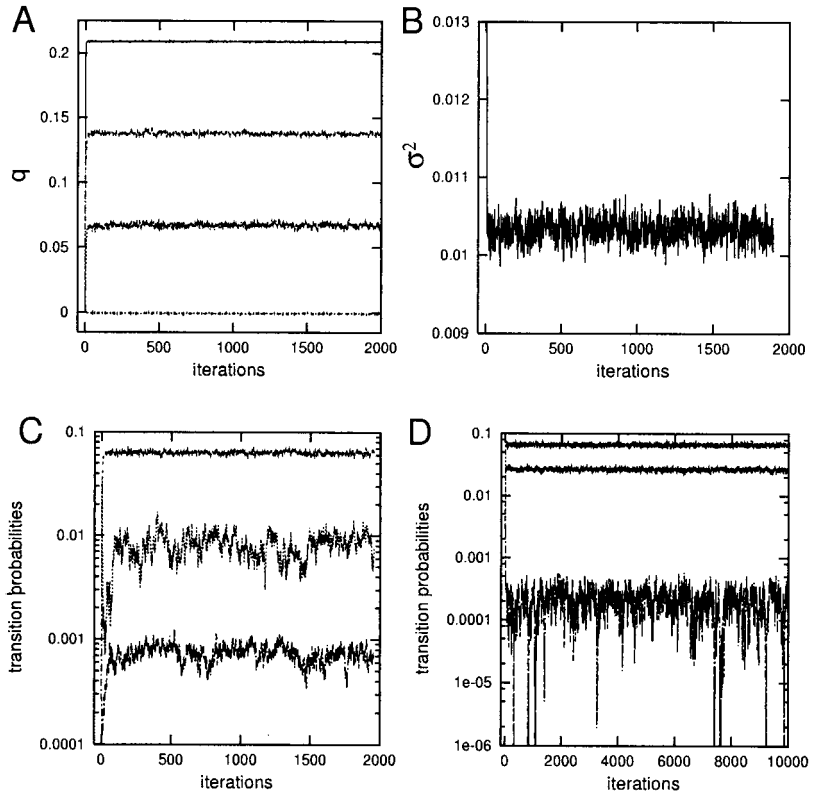


FIGURE 2 Samples for some components of θ against iteration number: (*A*) samples for the level positions q; (*B*) samples for the variance $\sigma^2_{C_1}$; (*C*) samples for $a_{C_2C_1}$ (*center*), $a_{O_1C_1}$ (*top*), $a_{C_1C_2}$ (*bottom*); (*D*) samples for $a_{O_2C_2}$ (*top*), $a_{O_1O_2}$ (*center*), $a_{C_1O_2}$ (*bottom*). The total number of iterations in (*A*), (*B*), and (*C*) were 2000, whereas (*D*) presents a longer run (10^4) that suggests stationarity for $a_{C_1O_2}$.

ses): $q_{c_1} = 0.067$ (.001), $q_{c_2} = -7.44e-4$ (1.9e − 4), $q_{o_1} = 0.138$ (8.68e − 4), and $q_{o_2} = 0.209$ (1.43e − 4), the transition probability matrix,

$$\begin{array}{c} \\ C_1 \\ C_2 \\ O_2 \\ O_1 \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & O_2 & O_1 \end{array} \\ \begin{bmatrix} 0.924 & 0.008 & 0.066 & 0.002 \\ (0.003) & (0.002) & (0.003) & (0.002) \\ 7.3e-4 & 0.991 & 2.2e-4 & 0.008 \\ (1.3e-4) & (2.3e-4) & (1.1e-4) & (2.9e-4) \\ 0.002 & 5.8e-5 & 0.996 & 0.0021 \\ (1e-4) & (5e-5) & (1.1e-4) & (1.1e-4) \\ 2.54e-4 & 0.064 & 0.027 & 0.91 \\ (3.53e-4) & (0.002) & (0.002) & (0.002) \end{bmatrix} \end{array},$$

and the variances of the noise $\sigma^2_{c_2} = 0.01$ (2.64e − 5), $\sigma^2_{c_1} = 0.01$ (1.45e − 4), and $\sigma^2_{o_1} = 0.01$ (9.64e − 5), $\sigma^2_{o_2} = 0.01$ (1.69e − 5). In general, the sampled realizations that correspond to states or transitions that are less frequent have higher variances. In this case, the uncertainties associated with the short-lived states C_1 and O_1 are higher.

For comparison, Baum's likelihood maximization was applied using the same $\theta^{(0)}$ values as those for the sampler. In this case, the estimates for the θ components reported above were: $q_{c_1} = 0.067$, $q_{c_2} = 0.003$, $q_{o_1} = 0.164$, $q_{o_2} = 0.207$, $\sigma^2_{c_2} = 0.01$, and $\sigma^2_{o_1} = 0.01$, and, for the diagonal of A: $a_{c_1c_1} = 0.964$, $a_{c_2c_2} = 0.996$, $a_{o_1o_1} = 0.968$, and $a_{o_2o_2} = 0.998$.

Kernel estimates

Some properties of the samples $\theta_i^{(b+1)}, \ldots, \theta_i^{(M)}$ can be summarized through a kernel density estimate. Denote by $E = M - b$ the effective number of iterations after burn in. This estimator is given by

$$\hat{k}(x) = \frac{1}{E} \sum_{m=b+1}^{M} \phi\{x - \theta_i^{(m)}, \omega\}, \tag{27}$$

where ϕ, known as a kernel function, is itself a density usually chosen to be unimodal and symmetric about zero. Here we consider ϕ as a normal density with mean 0.0 and standard deviation (bandwidth) $\omega > 0$. Following Eq. 27, this estimate is constructed by centering a scaled normal density at each sample of $\theta_i^{(m)}$; the actual value at any point x is the average of the $M - (b + 1)$ ordinates at that value. For the level positions, q, this estimate might be used as an alternative to the all-points conductance histogram commonly used in single channel analysis. The estimates for some components of θ including the ones for q_{c_1} and q_{c_2} are shown in Fig. 3. Just as the bin width affects the appearance of a histogram, the value of ω affects the shape of the density estimate, with larger values producing smoother estimates. For all the components of θ, this parameter is obtained as $\omega = \sigma_\theta[4/(3E)]^{1/5}$, where σ_θ denotes the sample standard deviation (see Bowman and Azzalini, 1997, p. 31).

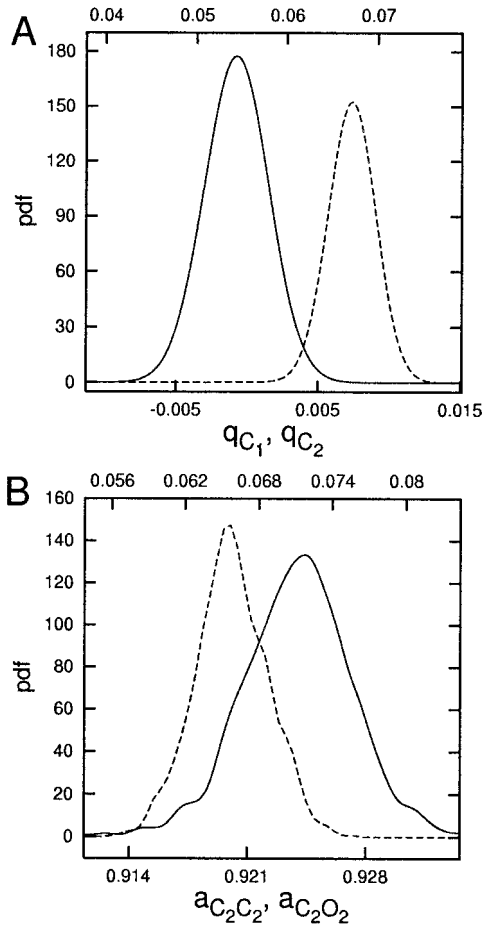


FIGURE 3 Kernel density estimates for selected posterior marginals. (*A*) Estimates for q_{c_2} (*continuous*, *bottom axis*) and q_{c_1} (*dashes*, *top axis*), both calculated from Eq. 27 with $\omega = 0.002$. (*B*) Estimates for $a_{c_2c_2}$ (*continuous*, *bottom axis*) and $a_{C_2O_2}$ (*dashes*, *top axis*); both obtained with $\omega = 0.0005$. The kernels were constructed from the last 1000 realizations.

Both figures also reflect the fact of higher uncertainty for the parameters associated to short lived sates. In any case, the posterior mode follows closely the true value.

Dwell time estimation

Single-state dwell times

The kernel density estimates for the dwell times $t_1, t_2, \ldots$ at any single state are presented in Fig. 4. These are calculated from m sampled realizations of z by following a form similar to the one in Eq. 27. Denote $t_{i,m}$ the ith sojourn of the mth realization, then the kernel is computed as

$$\hat{k}(x) = \frac{1}{E}\sum_{m=b+1}^{M}\frac{1}{H_m}\sum_{i=1}^{H_m}\phi\{x - t^*_{i,m}, \omega_i\}, \qquad (28)$$

where H_m is the number of dwell times for a particular state in $z^{(m)}$, and $t^*_{i,m} = \ln(t_{i,m})$, is the sequence of log-trans-

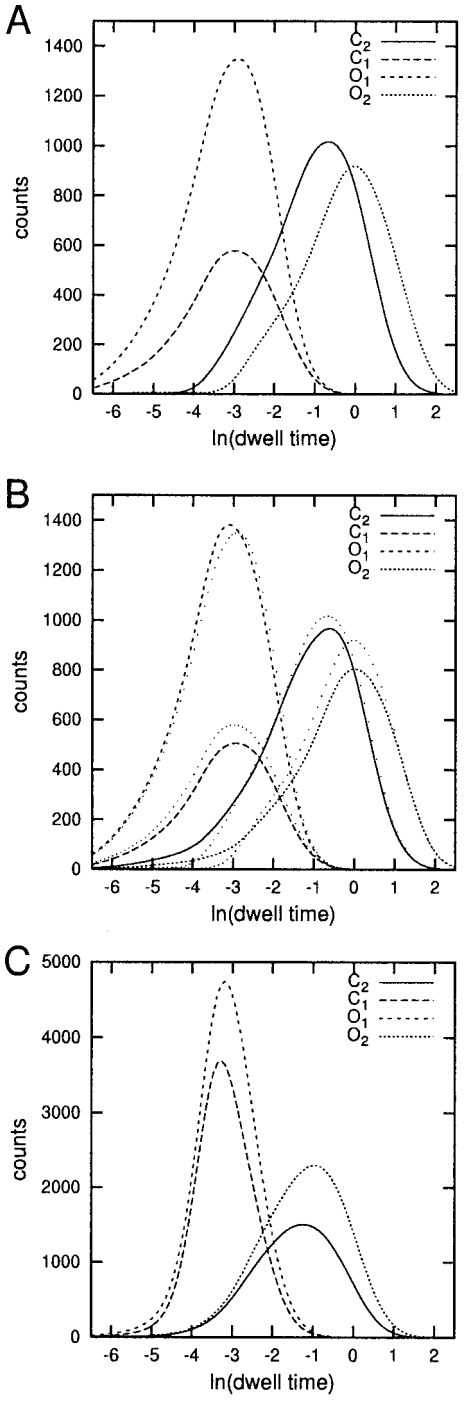


FIGURE 4 Kernels for the dwell times for each state obtained: (*A*) from the ideal trace (see Fig. 1 *C*); (*B*) from the posterior estimate produced using the Gibbs sampler; and (*C*) by low-pass filtering the trace at 5 kHz and thresholding (*C*). The curves from (*A*) are echoed in (*B*) for comparison. Note the different vertical scale in (*C*). Each trace has been scaled so that it integrates to the number of sojourns in the corresponding state. For the theoretical trace in (*A*) the counts and time constants were C_1: 1622, 0.055; C_2: 2858, 1.35; O_1: 3675, 0.063; and O_2: 2527, 0.552. The bandwidth was calculated from Eq. 29 with $g = 0.39$. For comparison with the ideal values of τ given earlier, the mean dwell times in each of the states in (*C*) are: $\tau_{c_1} = 0.050$, $\tau_{c_2} = 0.363$, $\tau_{o_1} = 0.052$, and $\tau_{o_2} = 0.438$.

formed sojourns. It is useful to use a scale-dependent bandwidth,

$$\omega_i = \max\{\ln(1 + \delta/t_{i,m}), g\}, \quad (29)$$

where g is the bandwidth or standard deviation of the normal density used for long dwell times, and $\ln[1 + \delta/t_{i,m}]$ ensures that the bandwidth for short dwell times corresponds to a time uncertainty of at least one sample interval. The choices of bandwidth and the spacing of x values for evaluation of ϕ are related. So that each dwell time will add equal weight to the density estimate, g must exceed the spacing. The densities in Fig. 4 have been scaled by the number of detected events or counts to allow comparison between the kernel estimates calculated from output of the Gibbs sampler and from the results of filtering and thresholding. For the Gibbs sampler, the density is multiplied by $H = \mathbb{E}_\pi[H_m]$. The plot in Fig. 4 *A* corresponds to the estimates for the time spent at each state, calculated from the noiseless sequence. The plot in Fig. 4 *B* shows the estimate obtained from the last 20 realizations obtained by the Gibbs sampler. These are in good agreement with the ideal results. The plot in Fig. 4 *C* shows the estimate obtained by low-pass filtering the data to 5 kHz using a digital Gaussian filter and noting the crossings of thresholds at 0.035, 0.105, and 0.175. Many of the extra counts in the large peaks for states C_1 and O_1 in Fig. 4 *C* are artifacts that represent the finite time taken for the filtered conductance to pass through the band of conductances corresponding to these states in the transitions between C_2 and O_1 and between O_2 and C_1, respectively. These artifacts render the estimates for C_1 and O_1 virtually useless. If the cut-off frequency for the low-pass filter could be increased, the artefactual counts would occur for shorter dwell times and thus might be separable from the genuine events. However, increasing the cut-off frequency increases the amplitude of the noise, which increases the frequency of another form of artefactual crossing of the thresholds. Indeed the peaks for the long-lived states C_2 and O_2 in Fig. 4 *C* contain more counts than for the ideal trace and are shifted to shorter dwell times because, even after filtering to 5 kHz, the noise produces a large number of additional crossings of the thresholds at 0.035 and 0.175.

Class dwell times

In practice, filter and threshold analysis of the data summarized in Fig. 4 would be attempted using a single threshold leading to two classes of states, $O = \{O_1, O_2\}$ above the threshold and $C = \{C_1, C_2\}$ below. To obtain the predictions of the model for these classes, the Q matrix is partitioned into four submatrices Q_{oo}, Q_{co}, Q_{oc}, and Q_{cc}; each with entries equal to the corresponding transition rates (Colquhoun and Hawkes, 1982). Figure 5, *A* and *B*, displays the theoretical densities obtained by using the spectral representation of the Q matrix (Colquhoun and Hawkes, 1982, 1995), and least square fits of two log-transformed exponentials, $\Sigma_i (W_i/\tau_i)\exp[x - \exp(x)/\tau_i]$, $i = 1, 2$; with $x = t_o$ or $x = t_c$ to the kernel estimates obtained from the Gibbs sampler for the time spent in both classes. These are calculated with Eqs. 28 and 29 after adding together the events corresponding to successive sojourns of each class. The relevant dwell times were taken from the last 20 sampled realizations. In Fig. 5 *C* the dwell times have been obtained by low-pass filtering the noisy trace with a 5-kHz digital Gaussian filter and noting a change of state whenever the signal crosses a threshold at 0.105, half way between the two middle levels. With this synthetic signal, a large majority of genuine transitions do cross the midline, which makes this an appropriate position for a single threshold. The filter frequency of 5 kHz was chosen to reduce the "false alarms" resulting from noise fluctuations to less than 5% of the total. It is possible to use relatively light filtering and tolerate a relatively high rate of noise transitions when the signal is in a state close to a threshold because these states are occupied for less than 10% of the time and genuine transitions from them are frequent. However, even at 5 kHz, more than a third of the genuine events have been missed. Further low-pass filtering leads to a substantial reduction in the number of threshold crossings observed.

The use of the Gibbs sampler to analyze data for which several states present the same conductance can be illustrated directly by considering the previous data set, but with $q_{c_1} = q_{c_2} = 0.0$ and $q_{o_1} = q_{o_2} = 0.07$. These data were analyzed with four states and the restriction $q_{c_1}^{(m)} = q_{c_2}^{(m)}$, $q_{o_1}^{(m)} = q_{o_2}^{(m)}$. In the first case, the state space is simply $\mathscr{C} = \{O, C\}$, whereas in the second, the partition $\mathscr{C} = O \cup C$ with $O = \{O_1, O_2\}$ and $C = \{C_1, C_2\}$ is implicitly assumed. The sampler was run for 2×10^4 iterations using the same priors and the same initial values for θ and z as those used previously. The results are shown in Fig. 6. Under these demanding conditions, the counts are now noticeably different from the ideal ones but the time constants are still close to their correct values. Moreover, in this case, the sampler is still able to recover the transition probability matrix,

	C_1	C_2	O_2	O_1
C_1	0.927	0.006	0.015	0.005
	(0.01)	(0.005)	(0.01)	(0.052)
C_2	$8.54e-4$	0.991	0.001	0.007
	$(7.12e-4)$	$(5.55e-4)$	(0.001)	(0.001)
O_2	$2.87e-4$	$7.9e-4$	0.996	0.003
	$(2.88e-4)$	$(6.81e-4)$	$(3.57e-4)$	$(7.8e-4)$
O_1	0.005	0.071	0.012	0.908
	(0.005)	(0.008)	(0.009)	(0.006)

The Gibbs sampler can also be used to illustrate the consequences of fitting with a wrong model by using only two states. The results are shown in Fig. 6 *D*. Comparison of

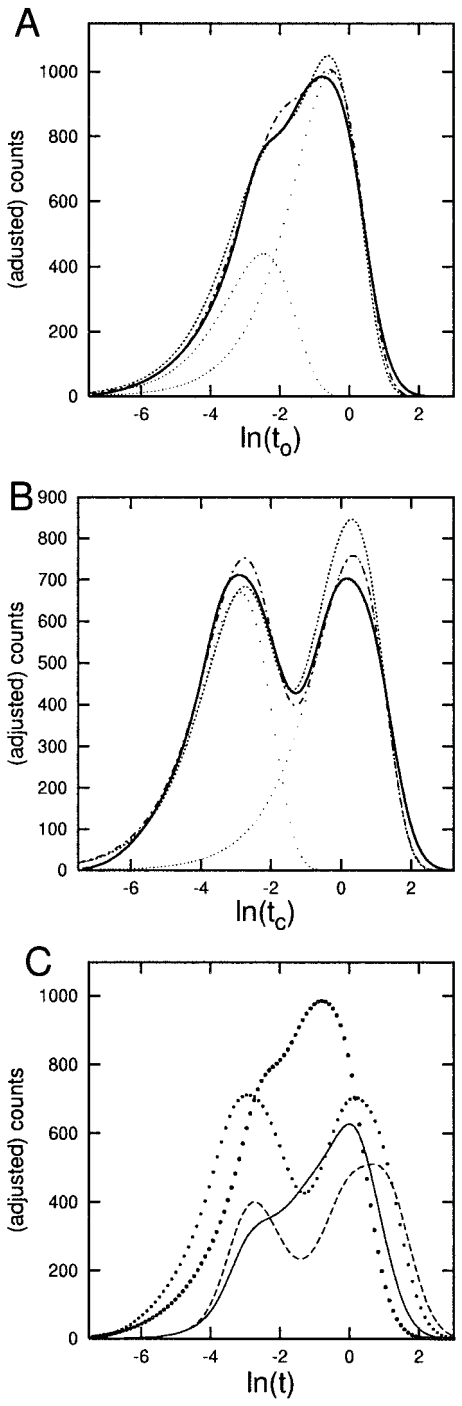


FIGURE 5 Kernels for log dwell-times (*A*) in the open class, *O*, and (*B*) in the closed class, *C*, obtained from the sojourns in the last 20 sampled realizations of $z^{(m)}$ produced by the Gibbs sampler. The bandwidths were calculated from Eq. 29 with $g = 0.39$. The kernel estimate is shown as a continuous (*bold*) line, the theoretical pdf obtained from the *Q* matrix as short dashes, and a nonlinear least squares fit of an exponential mixture (see text for details) as dot-dashes. The dots in both (*A*) and (*B*) correspond to the components of the fitted mixture. (*C*) Kernel for the dwell times in *O* (*continuous*) and *C* (*dashes*) obtained via filtering and thresholding, superimposed on the kernels shown in (*A*) and (*B*). Kernels and fits are scaled to integrate to the total number of dwell times. Theoretical pdfs in (*A*) and (*B*) are scaled to integrate to the number of dwell times present in the original noiseless data. For the kernels for the Gibbs sampler output the counts and time constants were: C_1, 1835, 0.055; C_2, 2089, 1.42; O_1, 1083, 0.068; and O_2, 2873, 0.57. A fit to the kernels presented in (*C*) (not shown) gave: $\tau_{c_1} = 0.076$, $\tau_{c_2} = 2.033$, $\tau_{o_1} = 0.097$, $\tau_{o_2} = 1.012$. Note (see also Fig. 6) that the fitted components in (*A*) and (*B*) are not estimates of the kernels for the individual states (see e.g., Colquhoun and Hawkes, 1982).

the traces for the open and closed states in Fig. 6 *D* with those for the open and closed classes in Fig. 6, *A* and *B*, shows that the predictions are biased toward those consistent with a two-state model. Many rapid transitions are missed, and the apparent time constants for the longer events are correspondingly too long ($\tau_{C_2} = 2.45$ and $\tau_{O_2} = 1.24$, as compared to 1.44 and 0.60, respectively). Attempts to use low-pass filtering and noting the crossings of a threshold at 0.035 produced much less satisfactory kernels. With filtering to 2 kHz, more than 20% of the crossings were false alarms corresponding to the noise, whereas, at 1 kHz, there were only 1158 counts in each class with almost complete failure to detect rapid events.

Joint densities

Class dwell-time histograms or kernels are useful for determining the number of components and their respective mean dwell-time durations. However, they do not display important information that is available about the way states are connected. Further insight into the relationship among states of different classes is obtained by studying the possible correlations of consecutive dwell times. For pairs of consecutive sojourns at *O* and *C*, these correlations can be displayed as a two-dimensional histogram (Magleby and Song, 1992). The surface at Fig. 7 *A* presents the theoretical joint density for the pair (x_o^*, x_c^*), which is obtained by following arguments similar to the one used to derive the unidimensional densities for each class (see Fredkin et al., 1985, Eq. 4.1). Figure 7 *B* shows the bivariate kernel estimate obtained as a multidimensional extension to Eq. 28,

$$\hat{k}(x_o^*, x_c^*) = \frac{1}{E}\sum_{m=b+1}^{M}\frac{1}{H_m}\sum_{i,j=1}^{H_m}\phi[x_o^* - u_{i,m}^*, \omega_i]\phi[x_c^* - v_{j,m}^*, \omega_j], \qquad (30)$$

with $u_{i,m}^*$ and $v_{j,m}^*$ as the log-transformed (consecutive) sojourns obtained from the last 20 iterations of the sampler for the classes *O* and *C*, respectively. The bandwidths ω_i, and ω_j are given by Eq. 29.

The highest peak in Fig. 7 *B* arises from brief sojourns in *O* followed by long ones in *C*, which, for two open and two closed states, could arise either from $O_1 \rightarrow C_2$ as in the present mechanism or from $O_1 \rightarrow C_2 \rightleftharpoons C_1$. A second peak corresponds to the occurrence of long dwell times at *O* which are adjacent to short ones in *C*, i.e., $O_2 \rightarrow C_1$ as here

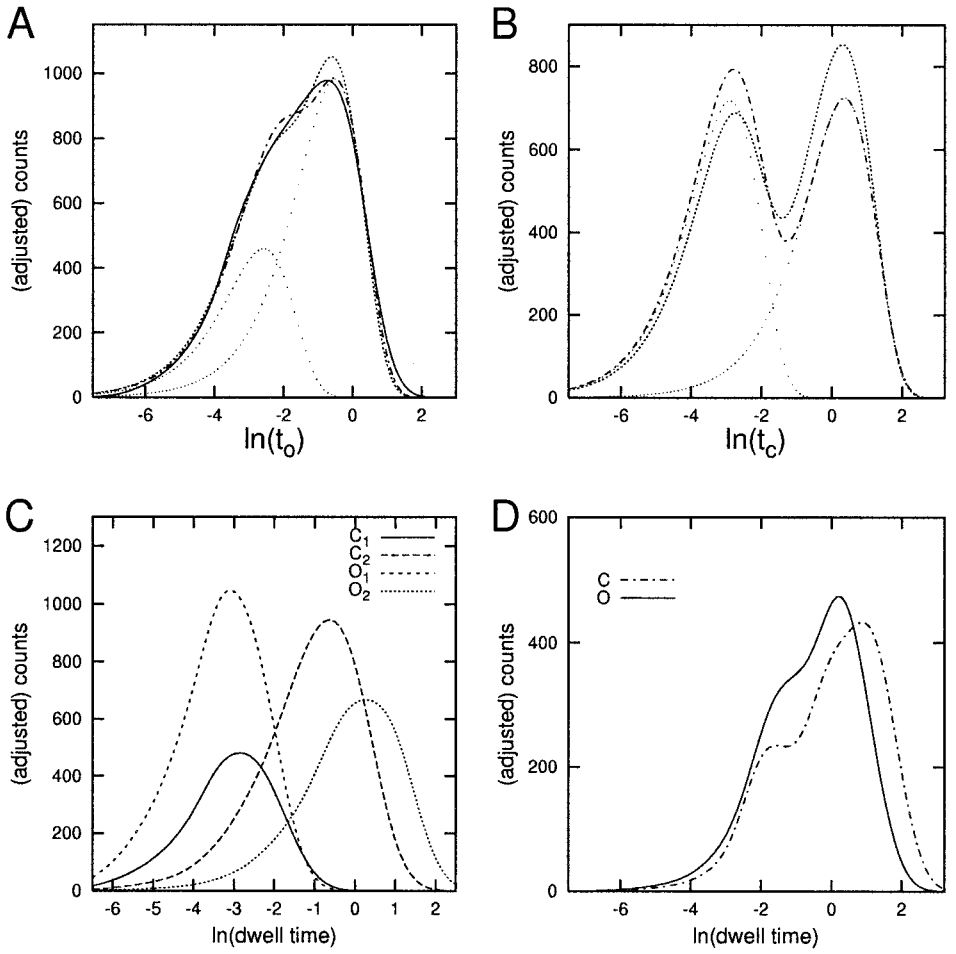


FIGURE 6 Logarithmic dwell time kernel estimates obtained for the analysis of the same mechanism as in the Results section, but with only two conductance levels, $q_{c_1} = q_{c_2} = 0.0$ and $q_{o_1} = q_{o_2} = 0.07$. Plots (*A*), (*B*), and (*C*) correspond to the estimates obtained when analyzing this data with four states with the restriction $q_{c_1}^{(m)} = q_{c_2}^{(m)}, q_{o_1}^{(m)} = q_{o_2}^{(m)}$, for $m = 1, 2, \ldots, M$. (*A*) represents the estimate for the dwell times in the open class, (*B*) for the closed class, and (*C*) the single state estimates. The plot at (*D*) presents the estimates for the class dwell times obtained with the sampler by considering only two states. For both (*A*), and (*B*) the solid (*bold*) line corresponds to the kernel estimate, short dashes correspond to the theoretical pdf obtained from Q, and dot-dashes to a nonlinear least squares fit of a two-component exponential mixture. The counts and the time constants for the fitted components gave: C_1, 1324, 0.055; C_2, 2801, 1.438; O_1, 2805, 0.077; and O_2, 2000, 0.599. In (*D*), the solid lines represents the estimate for the open state and the dashed line the one for the closed state. All the densities are rescaled to integrate respectively to the number of dwell times in each class or state. Plot (*A*) may be compared directly with Fig. 5 *A*, plot (*B*) with Fig. 5 *B* and plot (*C*) with Fig. 4 *B*.

or $O_1 \rightleftharpoons O_2 \rightarrow C_1$. A third, less prominent, maximum is constituted by long sojourns in both classes, which, in general, could correspond to direct transitions $O_2 \rightarrow C_2$ but here arise primarily from composite events of the form $O_2 \rightarrow C_1 \rightarrow C_2$, in which the middle event is so brief that it is not detected. All of these features are recovered by the Gibbs sampler as shown by the kernel estimate at Fig. 7 *B*. In general, the visible maxima correspond to the ones that present the most frequent interval combinations (including compound transitions) or those arising from components with widely separated time constants.

Magleby and Song (1992) have introduced a measure for the significance of the dependencies between successive dwell-time pairs that helps to clarify the connectivity of the states. This measure is defined as the difference between the square roots of the frequency of dwell-time pairs (x_o^*, x_c^*), and the frequency that would be expected if the successive dwell times occurred independently,

$$D(x_o^*, x_c^*) = \sqrt{p(x_o^*, x_c^*)} - \sqrt{p(x_o^*)p(x_c^*)}. \qquad (31)$$

This quantity can be obtained from the estimate in Fig. 7 *B* by marginalization. Let $ng_1 \times ng_2$ be the dimension of the grid used for the estimate at Fig. 7 *B*, then, for the (p, q)th coordinate $p_{pq}(x_o^*, x_c^*) = \hat{k}_{pq}(x_o^*, x_c^*)$, and

$$p_p(x_o^*) = \sum_{q=1}^{ng_2} \hat{k}_{pq}(x_o^*, x_c^*),$$
$$p_q(x_c^*) = \sum_{p=1}^{ng_1} \hat{k}_{pq}(x_o^*, x_c^*). \qquad (32)$$

Figure 7 *C* shows the theoretical dependence-difference, and Fig. 7 *D* the one obtained from the estimate at Fig. 7 *B*. It can be seen that significant dependencies, $D(x_o^*, x_c^*) > 0$, are related to dwell-time pairs that correspond to events $O_1 \rightarrow C_2$ and $O_2 \rightarrow C_1$. Negative dependencies exclude the events $O_1 \rightarrow C_1$ and $O_2 \rightarrow C_2$. It is therefore possible to conclude that the smallest of the three maxima in Fig. 7 *B* arises from composite transitions either $O_2 \rightarrow O_1 \rightarrow C_2$ or $O_2 \rightarrow C_1 \rightarrow C_2$.

DISCUSSION

Summary

A method for dealing with the statistical analysis of hidden Markov models for single channel records has been devel-

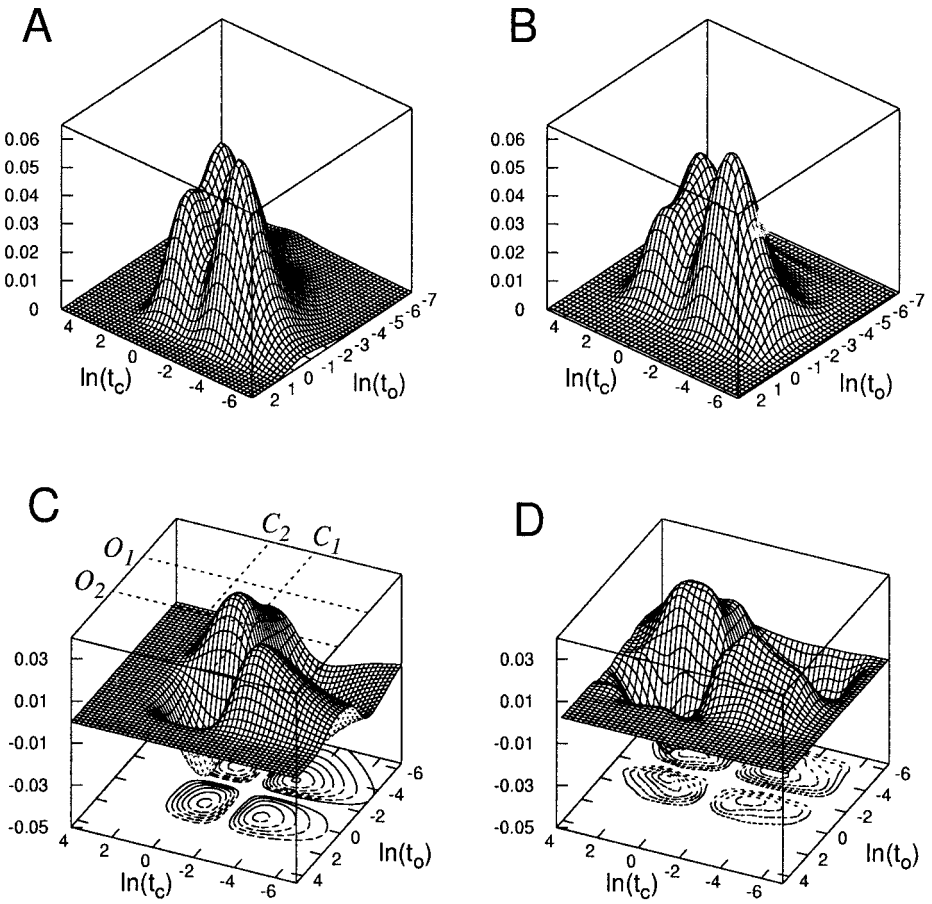


FIGURE 7 (*A*) Theoretical density for pairs of adjacent sojourns, (x_o^*, x_c^*), in *O*, *C* from the mechanism in the Results section. (*B*) Posterior kernel density estimate for (x_o^*, x_c^*), calculated from 10^5 dwell time pairs generated after iteration 1000. The estimate was obtained following Eq. 30 with both ω_o and ω_c calculated from Eq. 29 with $g = 0.39$ and a grid of 50×50. (*C*) Theoretical dependence-difference (Eq. 31). (*D*) Dependence-difference obtained from (*B*) by Eq. 32.

oped. Inferences concerning signal restoration and parameter estimation were based on the framework provided by Bayesian statistics in contrast to methods based on likelihood maximization (Baum et al., 1970; Chung et al., 1990; Venkataramanan et al., 1998; Michalek and Timmer, 1999; Fredkin and Rice, 1992a). Here, the probability distribution of all the unknowns given the observations was explicitly defined, and was the basis for all the inferences. Estimates are generated by Markov chain Monte Carlo sampling from this distribution. By analyzing signals of known characteristics, these methods have been found to produce estimates of the model parameters that are similar to those obtained using Baum's likelihood maximization.

Sampling methods provide a precise description of the modeling process. Under mild regularity conditions, the Gibbs sampler automatically provides a measure for the uncertainty of the generated estimates. This fact follows from the Central Limit Theorem, which applies as a consequence of the duality principle between $\{\theta^{(m)}\}$ and $\{z^{(m)}\}$ stated in the section, Convergence. Further results concerning the estimation of the variance are possible in this setting (Chauveau et al., 1998). Although different in nature from the Monte Carlo estimates obtained here, Bickel and Ritov (1996) and Bickel et al. (1998) have established asymptotic normality for likelihood-based estimates. However, the accuracy of the asymptotic approximation still remains unknown. Moreover, contrary to Monte Carlo estimates, the precision of the estimates in this setting depends on the length of the data sample. In the likelihood context, S. Michalek, M. Wagner, W. Vach, and J. Timmer (submitted for publication) propose $\sigma\sqrt{a_{ij}/t\lambda_i}$ as an empirical estimate for the standard deviation of a_{ij}. An alternative for the construction of confidence intervals in this setting is provided by the use of parametric bootstrap methods that require additional computations (MacDonald and Zucchini, 1997).

From a practical perspective, these methods do not impose any constraint on the position of the conductance levels, and they do not require specification of thresholds. Any information that is available about the conductance level positions is incorporated via a prior, which, as discussed in the section, Priors, may be arbitrarily vague. The prior for the levels in the example considered here was specified by setting the mean position for all the levels at the midpoint of the observed data range, and their dispersion to 3.68 times the standard deviation of the observations.

Kernel density estimates have been used to summarize the samples of the model parameters $\theta^{(m)}$ and dwell times

obtained from $z^{(m)}$. These serve much the same purpose as histograms. They have well-known advantages for display when there are relatively few data points (see i.e., Bowman and Azzalini, 1997), compared to the desired number of bins in the histogram. Kernel estimation may thus be particularly useful in the construction of complex two-dimensional empirical estimates such as the joint density of sojourn pairs. In addition, for the display of dwell times on a logarithmic axis, kernel construction is simpler than preparation of a smoothed histogram because the operations of binning, correction for the logarithmic transformation (Stark and Hladky, 2000), and smoothing are all replaced by a single step.

To illustrate the method and to demonstrate its effectiveness, a synthetic data set was constructed with four equally-spaced conductance levels and rapid transitions between nonadjacent levels. Traditional methods in which transitions are detected as threshold crossings after low-pass filtering the data are very difficult to apply to this type of data. If three thresholds are spaced between the levels, there are obvious artifacts (see Fig. 4 *C*) including false alarms when the noise reaches a nearby threshold and "transit peaks" which result from the finite time it takes the filtered trace to pass through the intermediate conductance band in a transition between nonadjacent levels. If the low-pass filtering is light (e.g., 20 kHz), then the noise frequently reaches the thresholds and the genuine transitions are swamped. If, in contrast, the data is more heavily filtered (i.e., to 5 kHz as in Fig. 4 *C*), then the time taken for a transit through the intermediate conductance band becomes comparable to the actual dwell times in some of the conductance states, and, again, there are no means for distinguishing genuine from artefactual events. In Fig. 4 *C* the filtering has eliminated approximately a third of the genuine transitions, while the "transit peaks" completely obscure the peaks for the more rapid transitions. By contrast, the Gibbs sampler recovers a faithful copy of the ideal trace from the noise, and, furthermore, the parameters of the probability matrix, A, are very close to the values used in the construction of the dataset. The mean dwell times in the states can be recovered from either the probability matrix or from fits of exponentials to the kernel density estimates for the dwell times.

In practice, analysis of these data using methods based on filtering and thresholding would use a single threshold and would thus detect dwell times in closed and open classes of states rather than in the individual states. Results based on this approach are shown in Fig. 5. More than a third of the genuine events have been missed by filter and threshold, and ~5% of the transitions that have been detected are false alarms.

Much of the difficulty inherent in fitting the data used for Figs. 4, 5, and 7 arises from transitions between nonadjacent conductances. Figure 6 shows the results obtained for the same sequence of state transitions, but with the conductances of the two closed states both zero and the conductances of the open states both 0.07. The conductance transitions were chosen to be less than the standard deviation (still 0.1) of the white noise, so that these data would present a real challenge for fitting. When the Gibbs sampler is run with four states subject to the constraints, $q_{c_1}^{(m)} = q_{c_2}^{(m)}$ and $q_{o_1}^{(m)} = q_{o_2}^{(m)}$ the shapes of the kernels for the dwell times in the conductances classes (see Fig. 6, *A* and *B*), the time constants for the single state estimates (see Fig. 6 *C*), and the transition probabilities are close to their ideal values. By contrast, when the Gibbs sampler is run with just two states, the kernel estimates for the classes are badly distorted (see Fig. 6 *D*) with poor detection of rapid events and a corresponding lengthening of the long events. Attempts to use threshold detection after sufficient low-pass filtering to reduce false alarms resulting from noise to less than 5% of the number of detected events produced no suggestion of a rapid component in the kernels.

The approach followed here is similar in spirit to those of Ball et al. (1999) and Hodgson and Green (1999). Both are concerned with restoration under a constrained hidden Markov model that follows a particular gating mechanism. The second actually takes a further step toward model selection. The principal objective here is signal restoration, and, for this, we have designed a simpler Markov Chain Monte Carlo sampler based on a discrete time approximation. As a result, the methods developed are less computationally intensive. Moreover, for the set of examples considered, the sampler seems to be less sensitive to initial conditions and does not require acceleration techniques to improve its mixing such as tempering. Possible disadvantages of our parameterization are discussed in the next section.

Our programs have been implemented in C and compiled using the GNU C compiler (gcc) on various platforms including Unix, Linux, and Windows NT without further modifications. Each iteration consisting on the update of $(\theta^{(m)}, z^{(m)})$ to $(\theta^{(m+1)}, z^{(m+1)})$ for the 10^6 samples data with four states presented in the Results section takes about 12 s on a PC with a Pentium 225-MHz processor under Linux. For the examples presented here, 1000 iterations were allowed for burn in, although for most cases a stationary point was reached at <50 iterations. The kernels shown in Figs. 4 and 6 are based on the last 20 iterations of each run.

Extensions

The results in this paper are based on a simplified version of the hidden Markov model Markov chain Monte Carlo method in which the observations are assumed to be independent and identically distributed. Real patch clamp records, however, are usually correlated due to the use of analog low-pass band filters and the presence of nonwhite background noise. It has been shown (Venkataramanan et al., 1998, among others) that standard Hidden Markov models produce biased estimates if these dependencies are not

taken into account. An extension of the Gibbs sampler to filtered and colored noise data will be presented by R. Rosales et al. (manuscript in preparation).

The underlying continuous time process followed by the channel $\{z^l\}$ is approximated by the discrete version $\{z^k\}$. In these terms, the model is parameterized through the transition probability matrix A instead of the generator Q. This produced satisfactory results for the purposes of signal restoration. However, there are several disadvantages for this choice if the main objective is to estimate the rate constants of the model, i.e., to determine Q. First, the map $Q \mapsto A = \exp(Q\delta)$, for real positive elements of Q, is not onto the space of all $n \times n$ transition probabilities. Second, if the sample interval is sufficiently long that multiple transitions can occur within one interval, the inversion can be very inaccurate or even impossible. Values that should be zero may not be, i.e., composite events via transient intermediates may be incorrectly assigned as direct transitions in the model. Third, the rate constants Q are much more convenient than the probabilities A for describing the structure of the model in displays of the possible transitions of $\{z^k\}$, such as shown by the mechanism in the Results section. Finally, the continuous time parameterization may reduce the computational effort by considering interval updates for z rather than single point ones.

Parameterization of the model using Q rather than A has been considered in the frequentist context by Michalek and Timmer (1999) and in the Bayesian framework by Ball et al. (1997). In the first approach, the transition rates are maximized numerically by considering Baum's auxiliary function for Q. The Bayesian approach follows by considering an exponential mixture likelihood for class dwell times, which, following standard theory (Colquhoun and Hawkes, 1982), is given by Q. This matrix is parameterized by k_{ij}, and, in this case, a conjugate prior for each rate is given by a Gamma density $G(\alpha_{ij}, \beta_{ij})$, with both α_{ij}, and β_{ij} assumed known. The estimation of θ and z is made by using a hybrid Markov chain Monte Carlo algorithm. The rates k_{ij} are sampled through a Metropolis–Hastings kernel, and z by one with reversible jumps similar to the one presented in J. A. Stark et al. (manuscript in preparation).

Selecting one or possibly a subset of aggregated hidden Markov models from an arbitrarily chosen family of competing candidates is perhaps still the most challenging problem in the statistical analysis of ion channel data. With few exceptions (Hodgson and Green, 1999; Ball and Sansom, 1989), this problem and the closely related question of model identifiability have not been formally addressed. In practical terms, there are instances in which physically different models produce identical statistics (Kienker, 1989; Wagner et al., 1999). In principle, Bayesian statistics provide a sound and consistent framework for model selection. The results presented here suggest that the present Bayesian Markov Chain Monte Carlo methods are suitable for signal restoration and may prove to be useful for model selection.

APPENDIX

Because the Gaussian hidden Markov model likelihood belongs to a exponential family, there is a class $\Im$ of conjugate priors for which the relevant full conditionals reduce to analytic densities (Bernardo and Smith, 1994, Prop. 5.4, p. 266). In this case, conditionals for θ under a particular realization, z, are given by

$$\begin{aligned} p(\theta_i|\theta_{-i}, y, z) &= p(\theta, y, z) \Big/ \int p(\theta, y, z)\, d\theta_i \\ &= p(y, z|\theta)p(\theta) \Big/ \int p(y, z|\theta)p(\theta)\, d\theta_i \\ &= L_z(\theta_i)p(\theta_i) \Big/ \int L_z(\theta_i)p(\theta_i)\, d\theta_i, \end{aligned} \tag{A1}$$

where $L_z(\theta_i)$ is the function formed from the likelihood under z with the terms that contain the component θ_i. The last relation holds only if θ_i is a_i or λ. When considering $\theta_i = q_i$ or $\theta_i = \sigma_i^2$, each conditional is also a function of the other component.

Conditional for σ_i^2

Following Eq. A1, explicit substitution of the likelihood and the prior $IG(u_i, w_i)$ leads to

$$p(\sigma_i^2|\theta, y, z) = \frac{\exp[-\chi(\sigma_i^2)]\sigma_i^{-2c_1}}{\int_{\mathbb{R}_+}\exp[-\chi(x)]x^{-c_1}\, dx} \tag{A2}$$

where

$$\chi(\sigma_i^2) = \left[\sum_{k=1}^{N} (y^k - q_i)^2 \mathbb{I}_i(z^k) + w_i\right] \Big/ 2\sigma_i^2,$$

$$c_1 = \sum_{k=1}^{N} \mathbb{I}_i(z^k) + u_i + 1,$$

and $\mathbb{I}_i(z^k)$ is the indicator function, i.e., $\mathbb{I}_i(z^k) = 1$ if $z^k = i$ and 0 otherwise. Setting $x = \sigma_i^2$, $c_2 = c_1 - 1$, and $c_3 = \Sigma_{k=1}^{N} (y^k - q_i)^2 \mathbb{I}_i(z^k)/2 + w_i$, the integral in Eq. A2 has the solution $c_3^{-c_2}\Gamma(c_2)$ (Box and Tiao, 1992, p. 144, Eq. A2.1.1). Substitution of this result into Eq. A2 leads to the conditional,

$$p(\sigma_i^2|\theta, y, z) = IG(\overline{u_i}, \overline{w_i}),$$

with $\overline{u_i} = c_2$, and

$$\overline{w_i} = 1/2\left[\sum_{k=1}^{N} (y^k - q_i)^2 \mathbb{I}_i(z^k) + w_i\right].$$

Conditional for q_i

Taking a normal prior with mean and variance m_i, s_i^2 leads to the following expression for the numerator of Eq. A1,

$$\exp\left[-\frac{1}{2}\frac{\sum_{k=1}^{N} (y^k - q_i)^2 \mathbb{I}_i(z^k)}{\sigma_i^2} - \frac{(q_i - m_i)^2}{s_i^2}\right]. \tag{A3}$$

The denominator follows by integrating this expression with respect to q_i on $\mathbb{R}$, which leads to

$$\frac{(2\pi\sigma_i^2)^{-n_i/2}}{s_i^2\sigma_i^2}(s_i^2 n_i + \sigma_i^2)^{1/2} e^{-(c_4+c_5)}, \tag{A4}$$

with

$$c_4 = \frac{s_i^2 \sum_{k=1}^{N} (y^k)^2 \mathbb{I}_i(z^k) + \sigma_i^2 m_i}{2\sigma_i^2 s_i^2},$$

$$c_5 = \frac{s_i^2 \sum_{k=1}^{N} y^k \mathbb{I}_i(z^k) + \sigma_i^2 m_i}{4\sigma_i^2 s_i^2 (s_i^2 m_i + \sigma_i^2)}.$$

The conditional for q_i is a normal density that results by replacing Eqs. A3 and A4 into Eq. A1,

$$p(q_i|\theta, y, z) = N(\bar{m}_i, \bar{s}_i^2),$$

with

$$\bar{m}_i = \frac{s_i^2 \sum_{k=1}^{N} y^k \mathbb{I}_i(z^k) + m_i\sigma_i^2}{s_i^2 \sum_{k=1}^{N} \mathbb{I}_i(z^k) + \sigma_i^2},$$

$$\bar{s}_i^2 = \frac{\sigma_i^2 s_i^2}{s_i^2 \sum_{k=1}^{N} \mathbb{I}_i(z^k) + \sigma_i^2}.$$

Conditionals for a_i and λ

Let $D(a_i|e_i)$ be the ith transition probability row prior, which is explicitly given by

$$D(a_i|e_i) = \frac{\Gamma(\sum_{j=1}^{n} e_{ij})}{\prod_{j=1}^{n} \Gamma(e_{ij})} a_{i1}^{e_{i1}-1} \cdots a_{in}^{e_{in}-1}.$$

Following Eq. A1, the full conditional for each row is

$$p(a_i|\theta, y, z) = \prod_{j=1}^{n} a_{ij}^{c_{ij}+e_{ij}-1} \Big/ \int_{\Lambda} \prod_{j=1}^{n} x_{ij}^{c_{ij}+e_{ij}-1}\, dx_{ij}, \tag{A5}$$

where

$$c_{ij} = \sum_{k=1}^{t-1} \mathbb{I}_i(z^k)\mathbb{I}_j(z^{k+1})$$

is the frequency of the transition $i \to j$ during z, and

$$\Lambda = \left\{ a_{ij} : a_{ij} > 0, \sum_{j=1}^{n} a_{ij} < 1 \right\}.$$

The integral in Eq. A5 has solution (Box and Tiao, 1992, p. 142, Eq. A2.1.7)

$$\Gamma\left(\sum_{j=1}^{n} c_{ij} + e_{ij}\right) \Big/ \prod_{j=1}^{n} \Gamma(c_{ij} + e_{ij}).$$

The conditional follows by substituting the expression above into Eq. A6. In this case, the conditional is

$$p(a_i|\theta, y, z) = D(c_{i1} + e_{i1}, \ldots, c_{in} + e_{in}).$$

The full conditional for λ is similarly found by considering the same prior structure as for the row a_i, but the quantities c_{ij} are replaced by $o_i = \mathbb{I}_i(z^1)$. In this case, the conditional is a Dirichlet density,

$$p(\lambda|\theta, y, z) = D(o_1 + b_1, \ldots, o_n + b_n).$$

APPENDIX B: CONVERGENCE RATE

This appendix presents an explicit form for the rate of convergence and its implications for the practical implementation of the Markov Chain Monte Carlo algorithm. Let $K_z^m = \mathbb{P}(z^{(m)} = B|z^{(0)} = A)$, the mth iterate of the transition probability for z, and $f^0(z = A|y)$ its initial density for any realizations $A, B \in Z$. Because Z is countable and $K_z^m > 0$ (by construction from Eqs. 17 or 20), then $z^{(m)}$ is ergodic, and, hence, for any $m = 1, \ldots,$

$$|f^m(z|y) - f(z|y)| \le (1 - 2K^*)^{m-1}, \tag{B1}$$

where

$$f^m(z|y) = \sum_{z \in Z} K_z^m f^0(z|y),$$

and

$$K^* = \min_{z^{(1)}, z^{(0)}} \{K_z^1\} \tag{B2}$$

(see Kemeny and Snell, 1960, Corollary 4.1.5, p. 71). Let Δ denote the L^1 norm with respect to $f(\cdot)$ or $\pi(\cdot)$, i.e.,

$$\Delta(f^m, f) = \sum_{z \in Z} |f^m(z|y) - f(z|y)|.$$

Then, following Eq. 26 and noting that Eq. B1 is pointwise satisfied for any $z^{(m)} = B$ ($B \in Z$), gives

$$\Delta(\pi^m, \pi) \le \Delta(f^m, f) \le n^N (1 - 2K^*)^{m-1}.$$

Hence, the convergence to the posterior marginal for z and θ is determined by the same rate that is governed K^*. In these terms, the number of iterations necessary to achieve an accuracy of ε for $\ln[\Delta(\pi^m, \pi)]$ is given by

$$m = [\ln(\varepsilon) - N \ln(n)]/\ln(1 - 2K^*).$$

Evaluation of the convergence rate leads, therefore, to the minimization of K_z^1 in Eq. B2. As an example, for the sampling scheme defined by Eq. 21, this is

$$K_z^1 = \prod_{k=1}^{N} p(z^k|z^{(1),l<k}, z^{(0),l>k}, \theta^{(0)}, y)$$

$$\propto h_{i_1}^1 \prod_{k=2}^{N-1} (h_{i_{k-1}i_k}^k a_{i_k i_{k+1}}) h_{i_{N-1}i_N}^N,$$

with $z^{(m),k}$ as the kth element of $z^{(m)}$. Given $\theta^{(0)}$, the sequence $i_1, \ldots, i_N$ that minimizes the above expression can be obtained recursively by a dynamic programming procedure such as the Viterbi algorithm.

This research was supported by the Maria Luisa Sanchez Scholarship awarded to R.R. by the Venezuelan Academy of Sciences, managed by Girton College in Cambridge, and Biotechnology and Biological Sciences Research Council grant 8/E03204 to S.B.H. and W.R.F. The authors also wish to thank the referees whose comments lead to an improved version of the original manuscript.

REFERENCES

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

Ball, F. G., Y. Cai, and A. O'Hagan. 1997. MCMC for hidden continuous-time Markov chains. http://www.shef.ac.uk/~st1ao/ps/hidden-m.ps.

Ball, F. G., and J. A. Rice. 1992. Stochastic models for ion channels: introduction and bibliography. *Math. Biosci.* 112:189–206.

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

Baum, L. E., T. Petrie, G. Soules, and N. Weiss. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.* 41:164–171.

Bernardo, J., and A. Smith. 1994. Bayesian Theory. Wiley, New York.

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Stat. Soc. B.* 36:192–326.

Besag, J., P. J. Green, D. Higdon, and K. L. Mergensen. 1995. Bayesian computation and stochastic systems. *Stat. Science* 10:3–66.

Bickel, P. J., and Y. Ritov. 1996. Inference in hidden Markov models I: local asymptotic normality in the stationary case. *Bernoulli.* 2:199–228.

Bickel, P. J., Y. Ritov, and T. Rydén. 1998. Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist.* 26:1614–1635.

Billingsley, P. 1968. Convergence of probability measures. Willey, New York. 174–177.

Bowman, A. W., and A. Azzalini. 1997. Applied Smoothing Techniques for Data Analysis. Chapter 2. Oxford Statistical Science Series, 18. Clarendon Press, Oxford, U.K.

Box, G. E. P., and G. C. Tiao. 1992. Bayesian Inference in Statistical Analysis. Wiley Classics Library. Wiley, New York. 142–144.

Carter, C. K., and R. Kohn. 1994. On Gibbs sampling for state space models. *Biometrika.* 81:541–553.

Chauveau, D., J. Diebolt, and C. P. Robert. 1998. Control by the central limit theorem. *In* Discretization and MCMC Convergence Assessment, No. 135 in Lecture Notes in Statistics, chapter 5, C. P. Robert, editor. Springer Verlag, New York. 99–126.

Chung, S. H., J. B. Moore, L. Xia, L. S. Premkumar, and P. W. Gage. 1990. Characterization of single channel currents using digital signal processing techniques based on hidden Markov models. *Phil. Trans. Roy. Soc. Lond. B.* 329:265–285.

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

Colquhoun, D., and A. G. Hawkes. 1995. A Q-Matrix cookbook. *In* Single-Channel Recording, chap. 20. B. Sackmann and E. Neher, editors, Plenum, New York. 589–633.

Colquhoun, D., and F. J. Sigworth. 1995. Fitting and statistical analysis of single channel records. *In* Single-Channel Recording, chap. 19, B. Sackmann and E. Neher, editors. Plenum, New York. 483–587.

Diebolt, J., and C. P. Robert. 1994. Estimation of finite mixture distributions through Bayesian sampling. *J. R. Stat. Soc. B.* 56:163–175.

Fishman, G. S. 1996. Monte Carlo, Concepts, Algorithms, and Applications. Chapters 1–6. Springer Verlag, New York.

Fredkin, D. R., M. Montal, and J. A. Rice. 1985. Identification of aggregated Markovian models: application to the nicotinic acetylcholine receptor. *In* Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, vol. 1, L. A. Le Cam and R. A. Olshen, editors. Wadsworth Press, Monterey, CA. 269–289.

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

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

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

Gelman, A., J. B. Carlin, S. S. Hal, and D. B. Rubin. 1995. Bayesian Data Analysis. Texts in Statistical Science. Chapman & Hall, London.

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

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

Kemeny, J. G., and J. L. Snell. 1960. Finite Markov Chains. Van Nostrand, Princeton, New Jersey. 70–74.

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

Liu, J. S., W. H. Wong, and A. Kong. 1995. Covariance structure and convergence rate of the Gibbs sampler with various scans. *J. Roy. Stat. Soc. B.* 57:157–169.

MacDonald, I. L., and W. Zucchini. 1997. Hidden Markov and Other Models for Discrete-valued Time Series. Monographs on Statistics and Applied Probability 70. Chapman & Hall, London.

Magleby, K. L., and L. Song. 1992. Dependency plots suggest the kinetic structure of ion channels. *Proc. R. Soc. Lond. B.* 249:133–142.

Meyn, S. P., and R. L. Tweedie. 1993. Markov Chains and Stochastic Stability. Springer Verlag, London.

Michalek, S., and J. Timmer. 1999. Estimating rate constants in hidden Markov models by the EM algorithm. *IEEE Trans. Sig. Proc.* 47:226–228.

Nummelin, E. 1984. General Irreducible Markov Chains and Non-Negative Operators. Cambridge University Press, Cambridge.

Robert, C. P. 1995. Convergence control methods for Markov chain Monte Carlo algorithms. *Statist. Sci.* 10:231–253.

Robert, C. P., and G. Casella. 1999. Monte Carlo Statistical Methods. Springer Texts in Statistics. Springer Verlag, New York.

Robert, C. P., G. Celeux, and J. Diebolt. 1993. Bayesian estimation of hidden Markov models: a stochastic implementation. *Statist. Prob. Lett.* 16:77–83.

Stark, J. A., and S. B. Hladky. 2000. Adjustments for the display of quantized ion channel dwell times in histograms with logarithmic bins. *Biophys. J.* 78:662–667.

Tierney, L. 1994. Markov chains for exploring posterior distributions. *Ann. Statist.* 22:1701–1786.

Venkataramanan, L., J. L. Walsh, R. Kuc, and F. J. Sigworth. 1998. Identification of hidden Markov models for ion channel currents. I. Colored, background noise. *IEEE Trans. Sig. Proc.* 46:1901–1915.

Wagner, M., S. Michalek, and J. Timmer. 1999. Estimating transition rates in aggregated Markov models of ion channel gating with loops and with nearly equal dwell times. *Proc. R. Soc. Lond. B.* 266:1919–1926.