Biom. J. 37 (1995) 1, 39 – 63 Akademie Verlag

Evaluation and Estimation of Various Markov Models with Applications to Membrane Channel Kinetics

G. W. PULFORD, J. C. GALLANT, R. A. KENNEDY*, S. H. CHUNG

Department of Chemistry and *Research School of Information Science & Engineering
Australian National University

Summary

Hidden Markov modelling is a powerful and efficient digital signal processing strategy for extracting the maximum likelihood model from a finite length sample of noisy data. Assuming the number of states in the model is known, then the state levels, transition probabilities, initial state distribution and the noise variance can be estimated. We investigate the applicability of this technique in membrane channel kinetics not only as a parameter estimator, but also as an aid to discriminating between various model types according to their statistical likelihood. We survey three representative classes of channel dynamics, namely: aggregated Markov models, semi-Markov models (with asymptotically convergent transition probabilities), and coupled Markov models; reformulating each within a discrete-time hidden Markov model framework. We then provide numerical evidence of the effectiveness of the procedure using simulated channel data and hence show that the correct model, as well as the model parameters, can be discerned. We also demonstrate that the model likelihood can be used to indicate the approximate number of states in the model.

Key words: Channel dynamics; Model evaluation; Ion channels; Hidden Markov models; Maximum likelihood estimation.

1. Introduction

It is generally accepted that the current fluctuations in biological membrane channels can be modelled adequately as a function of a finite-state Markov process. Opinions vary as to exactly what kind of model is "best suited" to describing real patch-clamp data. Among the models which have been proposed are aggregated Markov models (COLQUHOUN and HAWKES, 1977), fractal models (LIEBOVITCH et al., 1986; KORN and HORN, 1988; LIEBOVITCH and SULLIVAN, 1987; LIEBOVITCH et al., 1987), semi-Markov models which include fractal models (BALL et al., 1991 and 1993), hidden Markov models (FREDKIN and RICE, 1992a and 1992b) and coupled Markov models (KENNEDY and CHUNG, 1992). Other examples may be found in SANSOM et al., (1989). With so many competing models it is expedient to decide upon a means of ranking them, thereby eliminating implausible models.

Our work is motivated by two practical considerations. Firstly, having postulated a particular model, we need to estimate (in an optimal sense) the model parameters based on the observation sequence. Secondly, given a set of alternative models for the channel kinetics, we seek an analytical scheme able to discriminate between them by assessing their plausibility. Here we demonstrate that the hidden Markov model (HMM) technique fulfills both of these requirements. We will restrict ourselves to models which can be classed as or adequately represented by discrete-time Markov chains with a possible aggregation of states. Lest the reader be led to believe that this is too restrictive an assumption, we point out that the previously cited models can be treated within our framework. For models which are continuous in time, for example those of Colquhoun and Hawkes (COLQUHOUN and HAWKES, 1977; 1981; 1982), we sample at a sufficiently high rate to recover a discrete-time model that is an accurate approximation. Hidden Markov models have already been applied successfully in other areas of signal processing (RABINER, 1989), and details concerning their implementation may be found in CHUNG et al. (1990, 1991).

Traditionally, identification of membrane channel models has focussed on techniques involving the numerical fitting of the interval histograms or the power spectrum of the noisy channel data. These approaches are often not robust when the data are noisy and are limited in the amount of information they can yield about the process parameters. For instance, fitting a sum of exponential distributions to an interval histogram will only reveal the mean durations (or, equivalently, the diagonal transition probabilities) and relative component amplitudes of the process. Fitting Lorentzian functions to the power spectrum will again only partially identify the Markov chain. Furthermore, such approaches only work when the relevant time constants, i.e., mean durations differ sufficiently. Some other approaches which have been tried include maximum likelihood estimation (BALL and SANSOM, 1989), which suffers an exponential complexity problem, and correlation functions (LIEBOVITCH et al., 1987), which have only been applied to identify binary Markov chains.

On the other hand, hidden Markov modelling is a computationally efficient maximum likelihood estimator which is robust in the presence of white noise and applicable to processes with many states. Moreover, HMM estimation yields a complete description of the process in that it can identify the initial distribution, state levels and transition probabilities of the underlying Markov chain; and it evaluates the model likelihood, thus providing a sensitive measure of the goodness of fit. A disadvantage of the technique–a common drawback for many identification schemes–is that the number of states in the underlying Markov process must be known a priori. We will see that this problem can be overcome, in an empirical sense, by running several estimators with different numbers of model states and then locating the "knee" in the log-likelihood curve from which the most appropriate number of model states may be estimated.

After a brief discussion of some broad classes of Markov models, we outline in Section 3 three representative examples of channel dynamics, and reformulate each model within the HMM framework. We will present the discrete-time formulation of these models as it is our intention to apply digital signal processing techniques from hidden Markov model identification. In section 6 we give numerical examples of these three HMM types and show that the HMM algorithm can be used to discriminate between them. We will see that, with appropriate interpretation, the HMM technique is a powerful tool not only in its ability to estimate the model parameters, but also in identifying which type of Markov model is more plausible in a statistical likelihood sense.

2. Representative Markov Models

2.1 *The Three Models For Comparison*

The three representative types of HMM that we have chosen for this exposition can be classed under the broad headings of Colquhoun and Hawkes models (COLQUHOUN and HAWKES, 1977; 1981; 1982), coupled Markov models (KENNEDY and CHUNG, 1992), and semi-Markov models. The unobservable or hidden process is assumed to be a discrete-time Markov chain with a finite state space. For detailed exposition of the theory of such processes, see BHARUCHA-REID (1960), GANTMACHER (1974), HOWARD (1960), and ISAACSON and MADSEN (1985). The standard HMM framework does not generally allow for chains with time-varying transition probabilities. However, in the particular case of a semi-Markov model (FELLER, 1966; COX, 1962; SMITH, 1955), where the transition probabilities depend on the time since the last transition, we can approximate the process by one in which significant time variation is limited to a maximum number of lags, and then concatenate the (discrete) time and state to form a homogeneous vector Markov chain. Thus a semi-Markov model can be approximated by a vector Markov model (i.e., a Markov chain with vector state) whenever the transition probabilities can be considered constant for large time lags.

A special case of discrete-time semi-Markov models, which will interest us here, is the so-called "fractal model" (LIEBOVITCH et al., 1986; LIEBOVITCH, 1989; KRISHNAMURTHY et al., 1991) whose kinetic rates have an exponential form. We will use this model to illustrate how HMM estimation techniques can be applied to identify the parameters of a non-stationary process. Fractal models, unfortunately, have the disadvantage that their transition probabilities cannot usually be expressed in terms of elementary functions. We will return to a discussion of semi-Markov models in Section 3.3.

The preceding discussion should give the reader some idea of the range of possible underlying models encompassed by the Markov model framework. We now proceed with a description of some mechanisms by which these signal processes can be "hidden" or observed indirectly. Firstly, the physical process can only be measured in the presence of corrupting noise which we model as white Gaussian noise with zero mean. Secondly, in keeping with the application of HMM techniques to membrane channel models, in which the stochastic process represents the channel current, it is traditional to assume that the observable process has only 2 states–corresponding to an open or closed channel. This suggests an aggregation or many-to-one mapping of the underlying model states. An exception to this binary aggregation is the case of a coupled Markov model where a single channel is modelled as comprising a number of binary pores, each of which may be open or closed. This type of model has generally more than 2 states whose physical manifestation is the presence of conductance sublevels in the membrane current. Of course, if the individual pores are not independent but function in sympathy, i.e., they are tightly coupled, then, especially in the presence of noise, these multiple levels need not all be visible in the data record. This property of coupled Markov models will be apparent from the simulated channel data in Section 6. We can now state formally the three categories of HMM which will concern us in the remainder of the article. These are:

(CMM) Coupled Markov models which are representative of first order, n-state ($n \geqq 2$), homogeneous Markov chains.

(CHM) Discrete-time Colquhoun and Hawkes models, representative of first order, multi-state, homogeneous HMMs with aggregation to two observable states.

(SMM) Binary semi-Markov models with asymptotically convergent transition probabilities which can be approximated by homogeneous vector Markov models.

In all cases, we assume that only noisy measurements of the underlying stochastic process are available. We will consider the three representative classes of hidden Markov models in more detail in the following sections.

2.2 *Notation*

We define in Table 1 some notation and abbreviations that will appear in the rest of the paper. We use bold face to denote vectors and matrices.

Table 1
Notation and abbreviations

Symbol	Description
x_k	scalar Markov chain
y_k	scalar Markov chain with additive noise
$\mathbf{z}_k$	vector Markov chain for semi-Markov model
$\mathbf{X}_k$	vector Markov chain for coupled Markov model
s_i	state level i, $i=1, \ldots, N$
$\mathbf{S}_i$	$N-1$-vector of state levels
$\mathbf{P}$	transition probability matrix
$\mathbf{Q}$	intensity matrix
N	number of states
L	maximum time lag since last transition
K	number of data points
t	continuous time
k	discrete time index, $k=0, 1, 2, \ldots$
τ	time since last transition
T	sampling time
HMM	Hidden Markov Model
CHM	Colquhoun and Hawkes Model
CMM	Coupled Markov Model
SMM	Semi-Markov Model

3. Model Descriptions

3.1 *The Coupled Markov Model*

The channel is viewed as consisting of a number $N-1$ of identical pores which may be either open or closed. The pores may function independently of one another, in the decoupled case; or as a single unit, in the fully coupled case; or be interdependent to some extent, in the partially coupled case. If $x_k^{(l)}$ is the state of pore l ($l=1, \ldots, N-1$) at time k, taking values s_1 when the channel is closed or s_2 when the channel is open, then the total current is proportional to

$$x_k = x_k^{(1)} + \cdots + x_k^{(N-1)} = \underbrace{[1, \ldots, 1]}_{N-1} \mathbf{X}_k, \tag{1}$$

where the $N-1$-vector $\mathbf{X}_k$ is defined as

$$\mathbf{X}_k = [x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(N-1)}]', \tag{2}$$

denoting the transpose of $\mathbf{x}$ by $\mathbf{x}'$. This has 2^{N-1} possible values $\{\mathbf{S}_i\}_{i=1}^{2^{N-1}}$,

$$\mathbf{S}_1 = [s_1, s_1, \ldots, s_1]'; \quad \mathbf{S}_2 = [s_2, s_1, \ldots, s_1]', \ldots, \quad \text{etc.}$$

Thus x_k has N possible values since the component currents are binary. In the decoupled case, each $x_k^{(l)}$ is a binary Markov chain with transition matrix

$$\mathbf{P} = (p_{ij}) = \begin{bmatrix} \zeta & 1-\zeta \\ 1-\varrho & \varrho \end{bmatrix}, \tag{3}$$

where the transition probabilities are defined independently of l by

$$p_{ij} = \Pr(x_{k+1}^{(l)} = s_j \mid x_k^{(l)} = s_i). \tag{4}$$

The $2^{N-1} \times 2^{N-1}$ transition matrix for the vector process $\mathbf{X}_k$ is given by the Kronecker product of the individual transition probability matrices:

$$\mathbf{P}_I = \underbrace{\mathbf{P} \otimes \mathbf{P} \otimes \cdots \otimes \mathbf{P}}_{N-1 \text{ terms}}. \tag{5}$$

In the fully coupled case, the state of each pore is allowed to be different initially, but all pores synchronise, i.e., all open or close together at the next sampling instant. In this case, the vector process $\mathbf{X}_k$ has $2^{N-1} \times 2^{N-1}$ transition matrix

$$\mathbf{P}_F = \begin{bmatrix} \zeta & 0 & \cdots & 0 & 1-\zeta \\ \delta & 0 & \cdots & 0 & 1-\delta \\ \vdots & \vdots & & \vdots & \vdots \\ \delta & 0 & \cdots & 0 & 1-\delta \\ 1-\varrho & 0 & \cdots & 0 & \varrho \end{bmatrix} \tag{6}$$

where δ can be chosen according to

$$\delta = \frac{1-\zeta}{2-\varrho-\zeta} \tag{7}$$

to reduce the number of free parameters in the model. Even with this restriction, a wide variety of signal behaviours can be generated by the CMM model.

The partially-coupled system is described by the convex combination of the decoupled and fully-coupled systems, so that its transition matrix is given by

$$\mathbf{P}_C = (1-\kappa)\,\mathbf{P}_I + \kappa \mathbf{P}_F, \quad \kappa \in [0, 1]. \tag{8}$$

The parameter κ governs the amount of coupling between individual pore chains. When $\kappa = 0$ the chains are decoupled (independent), whereas for $\kappa = 1$ they are fully-coupled. The channel current x_k resulting from this partially-coupled model can be shown to be Markov and is obtained from the vector process X_k by aggregating states with equal numbers of open pores. Thus the transition matrix of x_k is related to (8) by

$$\mathbf{P}_{CMM} = \mathbf{L}_0 \mathbf{P}_C \mathbf{R}_0, \tag{9}$$

where $\mathbf{L}_0$ and $\mathbf{R}_0$ are respectively $N \times 2^{N-1}$ and $2^{N-1} \times N$ constant matrices determined by the number of pores L. For a given number of pores, there are N channel states in CMM and the free parameters are $\{\zeta, \varrho, \kappa\}$, with δ determined by (7). In addition, given a matrix $\mathbf{P}_{CMM}$, it is possible to estimate the model parameters using a recursive least squares technique whose details need not concern us here.

3.2 *The Discrete-Time Colquhoun and Hawkes Model*

The traditional ion channel based on the work of Colquhoun and Hawkes represents the channel current as an N-state, continuous-time Markov chain aggregated to a binary process $x(t)$ in which t denotes the time. The N states, numbered from 1 to N correspond to conformations of the channel macromolecule which cannot be observed directly, but are responsible for the channel openings. Not all the conformational states communicate, so that there are disallowed transitions. Because the N underlying states have been aggregated into just open and closed channel states, the model is non-Markovian. The continuous-time formulation requires the specification of a constant $N \times N$ intensity matrix $\mathbf{Q}$ which determines the matrix of transition probabilities $\mathbf{P}(t) = (p_{ij}(t))$ with

$$p_{ij}(t) = \Pr(x(t) = j \mid x(0) = i), \tag{10}$$

via the equation $\mathbf{P}(t) = \exp(\mathbf{Q}t)$, with initial condition $\mathbf{P}(0) = \mathbf{I}$.

We can view a discrete-time Markov chain as the process whose transition matrix is obtained from that of the continuous-time process by sampling via

$$\mathbf{P} = \exp(\mathbf{Q}T) \tag{11}$$

with state $x_k = x(kT)$ and transition probabilities $p_{ij} = \Pr(x_{k+1} = j \mid x_k = i)$. The sampling period T is chosen to give a reasonable approximation to the continuous-time system. If T is sufficiently small, we may take the discrete-time transition probabilities corresponding to forbidden transitions as zero.

Suppose the states are ordered so that states $\{1, \ldots, n\}$ are closed and states $\{n+1, \ldots, N\}$ are open. In the absence of noise, the aggregated process is given by

$$y_k = \begin{cases} s_1 & \text{if} \quad x_k \in \{1, \ldots, n\} \\ s_2 & \text{if} \quad x_k \in \{n+1, \ldots, N\}, \end{cases} \tag{12}$$

where s_i are the currents due to closed ($i=1$) and open ($i=2$) channels. The continuous-time Colquhoun and Hawkes model can be slightly more general than the discrete-time version presented here.

For clarity, we take the simple 3-state example from COLQUHOUN and HAWKES (1977) in which state 1 may only transit to state 3 by passing through the intermediate state 2. The relevant transition matrix is of the form

$$\mathbf{P}_{CH} = \begin{bmatrix} p_{11} & 1-p_{11} & 0 \\ p_{21} & 1-p_{21}-p_{23} & p_{23} \\ 0 & 1-p_{33} & p_{33} \end{bmatrix} \tag{13}$$

which, as indicated, has 4 independent parameters. States 1 and 2 are aggregated to the closed channel state and state 3 is the open channel state. We will return to this model in the section on numerical simulation.

3.3 *The Semi-Markov Model*

Semi-Markov models are well known in the area of renewal theory (COX, 1962) and are characterised by transition probabilities which depend on the time since the last transition which we denote by τ. Thus τ is the local time which is reset at each transition. A semi-Markov model can be characterised by a time-varying intensity matrix $\mathbf{Q}(\tau)$ (under certain weak conditions on the semi-Markov kernel). The transition matrix is the solution of the Chapman-Kolmogorov differential equations:

$$\frac{d\mathbf{P}(\tau)}{dt} = \mathbf{Q}(\tau)\,\mathbf{P}(\tau) = \mathbf{P}(\tau)\,\mathbf{Q}(\tau), \quad \mathbf{P}(0) = \mathbf{I}. \tag{14}$$

By analogy with the theory of linear systems (CALLIER and DESOER, 1991; KAILATH, 1980), the state transition matrix $\boldsymbol{\Phi}(\tau, \tau_0)$ for the continuous-time system (14), defined by the equations

$$\begin{aligned} \boldsymbol{\Phi}(\tau, \tau_0) &= \mathbf{Q}(\tau)\,\boldsymbol{\Phi}(\tau, \tau_0) \\ \boldsymbol{\Phi}(\tau_0, \tau_0) &= \mathbf{I}, \end{aligned} \tag{15}$$

yields the discrete-time system matrix

$$\mathbf{P}_k = \boldsymbol{\Phi}((k+1)\,T, k\,T) \tag{16}$$

when sampled with period T.

We can approximate the semi-Markov model, making it amenable to HMM estimation, by assuming that after some time LT, where L is called the lag, the transition probabilities can be regarded as constant. In other words, for all i, j,

$$p_{ij}(\tau_k) = \begin{cases} p_{ij}(\tau_k) & \text{if} \quad \tau_k \leqq LT \\ p_{ij}(LT) & \text{if} \quad \tau_k > LT. \end{cases}$$

We now specialise to the two-state case as it is more relevant to channel modelling. By defining a vector state

$$\mathbf{z}_k = [\tau_k, x_k], \tag{17}$$

whose values we number from 1 to $2L$ according to the rule

$$\langle \mathbf{z}_k \rangle = i + (j-1)N \quad \text{if} \quad \tau_k = 1, \quad x_k = s_j, \tag{18}$$

for $i = 1, \ldots, L$ and $j = 1, 2$, we arrive at a homogeneous Markov chain with $2L$ states. The transition matrix $\mathbf{P}_{SMM}$ of this chain, whose (m, n) element is defined by

$$\Pr(\langle \mathbf{z}_{k+1} \rangle = n | \langle \mathbf{z}_k \rangle = m); \quad m, n = 1, \ldots, 2L,$$

has the structure shown below:

$$\left[\begin{array}{cccc|cccc}
0 & p_{11}[1] & \ldots & 0 & p_{12}[1] & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & p_{12}[2] & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\
0 & 0 & \ldots & p_{11}[L-1] & p_{12}[L-1] & 0 & \ldots & 0 \\
0 & 0 & \ldots & p_{11}[L] & p_{12}[L] & 0 & \ldots & 0 \\
\hline
p_{21}[1] & 0 & \ldots & 0 & 0 & p_{22}[1] & \ldots & 0 \\
p_{21}[2] & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\
p_{21}[L-1] & 0 & \ldots & 0 & 0 & 0 & \ldots & p_{22}[L-1] \\
p_{21}[L] & 0 & \ldots & 0 & 0 & 0 & \ldots & p_{22}[L]
\end{array}\right] \tag{19}$$

where the transition probabilities of the discrete-time semi-Markov model are defined as[1)]

$$p_{ij}[m] \triangleq p_{ij}(mT) = \Pr(x_{k+1} = s_j | x_k = s_i); \quad m = 1, \ldots, L, \; j = 1, 2, \tag{20}$$

with $\tau_m = mT$ being the time spent in the current state s_i which is updated as

$$\tau_{k+1} = \begin{cases} \tau_k + 1 & \text{if} \quad x_{k+1} = x_k \quad \text{and} \quad \tau_k < L \\ \tau_k & \text{if} \quad x_{k+1} = x_k \quad \text{and} \quad \tau_k = L \\ 1 & \text{otherwise.} \end{cases} \tag{21}$$

[1)] We use the symbol $\triangleq$ to signify defined equality.

In practice, the choice of lag L will trade off the quality of the approximation and the computational burden associated with estimating the $2L$ parameters in $\mathbf{P}_{SMM}$, not to mention the length of the data record required for estimation when L is large. Note that we do not constrain the individual transition probabilities $p_{ij}(\tau_k)$ to follow the trajectories predicted by the model (16). We merely constrain the structure of the transition matrix during the estimation. Effectively, this means that the estimation procedure is blind to the actual form of time dependence of the underlying process in the sense that arbitrary time variation in $p_{ij}(\tau_k)$ is permitted. We will treat these parameter estimation questions more fully in the following sections. As an example of a semi-Markov model, we now describe the fractal model. We mention that other such models are possible, depending on the behaviour of the transition probabilities, for instance, these may exhibit a damped oscillation before settling down to a steady value.

In continuous time, a fractal model is a finite-state semi-Markov process whose intensity matrix is given by

$$\mathbf{Q}(\tau)=\begin{bmatrix} -k_1\tau^{1-D_1} & k_1\tau^{1-D_1} \\ k_2\tau^{1-D_2} & -k_2\tau^{1-D_2} \end{bmatrix} \tag{22}$$

for some constants $k_i>0$, the kinetic rate constants; and D_i with $1\leqq D_i<2$, called the fractal dimensions. Thus the intensities are supposed to vary inversely with the time since the last transition between open and closed states. It can be shown that the transition probabilities of a fractal model become constant for large lags so we may approximate the process by a homogeneous Markov chain.

We follow the prescription described before to obtain a discrete-time approximation of the continuous-time system. For simplicity, we assume that $D_1=D_2\triangleq D$ to obtain an explicit formula for the transition probabilities of the discrete-time chain:

$$\begin{aligned}\mathbf{P}_k &= \exp\left\{\int_{kT}^{(k+1)T}\mathbf{Q}(\sigma)\,d\sigma\right\} \\ &= \frac{1}{k_1+k_2}\begin{bmatrix} k_2+k_1c(kT) & k_1(1-c(kT)) \\ k_2(1-c(kT)) & k_1+k_2c(kT) \end{bmatrix}\end{aligned} \tag{23}$$

where

$$c(\tau_k)=\exp\left(-\frac{k_1+k_2}{2-D}(\tau_k^{2-D}-\tau_{k-1}^{2-D})\right).$$

The general behaviour of the transition matrix for a fractal model is illustrated in this special case–namely that the probability of remaining in a given state increases monotonically as a function of the time spent in that state, becoming constant for large time lags. We will use the fractal model for data generation in Section 6.

4. HMM Processing Framework

4.1 *Model Parameters*

The quantities which characterise a discrete-time, n-state, hidden Markov model with state variable x_k and observable output y_k are the following:

(1) The number of states n and the corresponding state levels $s_1, \ldots, s_n$.

(2) A vector of initial state probabilities π_0 with components

$$\pi_0(i) = \Pr(x_0 = s_i), \quad i = 1, \ldots, n.$$

(3) A matrix of constant transition probabilities $\mathbf{P} = (p_{ij})$ where

$$p_{ij} = \Pr(x_{k+1} = s_j | x_k = s_i); \quad i, j = 1, \ldots, n.$$

(4) A matrix of observation symbol probabilities $\mathbf{B} = (b_i(y_k))$, $i = 1, \ldots, n$, where, for computational purposes, y_k is quantized to M levels and $b_i(y_k) = \Pr(y_k | x_k = s_i)$.

The observation process y_k is given by

$$y_k = x_k + w_k, \tag{24}$$

where x_k is the n-state, homogeneous, Markov chain characterised by (1), (2) and (3) above, and w_k is a sequence of independent[2)] random variables which, as before, we assume to be taken from a white Gaussian density with zero mean and variance σ^2, thus

$$b_i(y_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_k - s_i)^2}{2\sigma^2}\right\}. \tag{25}$$

We denote a HMM by the triple $\lambda = (\mathbf{P}, \mathbf{B}, \pi_0)$ on the understanding that the number of states n has been chosen prior to estimation of the model. More general formulations of hidden Markov models are possible, but the present framework is adequate for our demonstration.

4.2 *HMM Parameter Estimation*

The usual HMM framework allows us to obtain maximum likelihood estimates of the model λ, i.e., estimates of the initial state probabilities, the transition probabilities, the state levels, the noise variance and the state sequence. For our

[2)] The standard HMM framework does not allow for filtered observations or coloured noise.

purposes of comparing the various models, we will be content with estimating the transition probabilities and evaluating the likelihood function

$$\Pr(Y_K|\lambda), \tag{26}$$

in which Y_K denotes the sequence of observed data $\{y_1, \ldots, y_K\}$. The maximum likelihood model $\hat{\lambda}$ has the property that

$$\Pr(Y_K|\hat{\lambda}) \geqq \Pr(Y_K|\lambda), \ \forall \lambda.$$

An efficient (recursive) procedure for computing the likelihood of a given model–a calculation which would otherwise require a sum over all n^K possible state sequences, is given by the "forward-backward" procedure. This is an iterative procedure for computing the likelihood in terms of the "forward" variables $\alpha_k(i)$, defined by

$$\alpha_k(i) = \Pr(y_1, \ldots, y_k, x_k = s_i|\lambda), \quad i = 1, \ldots, n, \tag{27}$$

which yields the likelihood as

$$\Pr(Y_K|\lambda) = \sum_{i=1}^{n} \alpha_K(i).$$

We will not supply further details of these equations here, but instead refer the interested reader to RABINER (1989) (and references therein) for a fuller treatment. Similarly, the "backward" variables $\beta_k(i)$ are defined as

$$\beta_k(i) = \Pr(y_{k+1}, \ldots, y_K|x_k = s_i, \lambda). \tag{28}$$

Both the forward and backward variables are needed to obtain the model parameter estimates via the Baum-Welch re-estimation formulae (BAUM and PETRIE, 1966; BAUM et al., 1970). The HMM algorithm is a multi-pass strategy computing, at each pass i through the data, the variables (27), (28) given the model λ_i and using these to evaluate the current model likelihood as well as to re-estimate the model parameters. This procedure yields an updated model λ_{i+1} which is assumed during the next pass through the data. The crucial property of this re-estimation process is that the likelihood of the sequence of models $\{\lambda_i\}$ is a non-decreasing function of the number of passes i, with the consequence that the algorithm converges to a local maximum of the likelihood function (26). It should be noted that the complexity involved in HMM processing is of the order of $n^2 K$ operations per pass.

Given a HMM λ_i, it is possible to compute the "most probable" underlying state sequence $\{\hat{x}_k\}$, although the solution will depend on the criterion of optimality which is adopted. Two common interpretations are the (single symbol)

maximum a posteriori (MAP) criterion and the maximum likelihood sequence estimator (MLSE). We will not, however, elaborate these concepts here, as knowledge of the likelihood and associated transition probabilities is sufficient for our purposes in the model evaluation.

4.3 *General Remarks*

A few general comments seem in order before we discuss the technique for model comparison. The first of these concerns the computation of model statistics. It should be noted that all statistics of interest can be calculated once the model has been identified. In particular, the computation of channel properties such as mean durations, spectra and histograms is a straightforward matter with the model in hand. Since these quantities can be obtained in an analytical manner, they can be used to validate the model by comparison with the statistics computed directly from the noisy data.

Secondly, it is easy to incorporate models such as the discrete-time Colquhoun and Hawkes model which contain disallowed transitions into the HMM framework. All that is required is the setting of the relevant initial transition probabilities to zero. The HMM framework also allows for Markov models some of whose states have been aggregated. We need only constrain the state levels of the aggregated states to be equal during processing, although, in practice, such models tend to be difficult to identify due to the lack of information concerning transitions between states within the same aggregate.

We saw before that the Markov model framework is able to encompass a variety of different channel models, providing that these can be represented, or approximated, by homogeneous Markov chains. However, in the standard HMM framework, it is not generally a simple matter to include functional constraints on the transition probabilities. The reason is that such constraints imply a different parametrisation of the transition matrix. The conventional HMM re-estimation formulae are the end product of a constrained optimisation problem resulting from the Expectation Maximization algorithm (BAUM, et al., 1970; TITTERINGTON, et al., 1985). In the standard formulation, the only constraints are that the estimation transition probabilities must form a stochastic matrix–they are otherwise free to vary. Any additional constraint, due to other model parametrisations, e.g. CMM, or known time-variation, e.g. fractal models, would need to be included before computing the re-estimation formulae. Disallowed transition constraints may be an exception, and can be incorporated simply by setting the required elements of P to zero during estimation. In this article, we will not attempt to redesign the HMM procedure to allow for models with probability constraints more general than forbidden transitions. However, we caution that in certain instances even models with forbidden transitions may lead to problems in estimating the intensity matrix (FREDKIN and RICE, 1992b).

In the interests of completeness, we describe a technique for evaluating the approximate model likelihood when the parametrisation of the model does imply additional functional dependence between its transition probabilities, as is the case for both CMM and the fractal model. We emphasize that this technique is laborious in practice as it necessitates an extra parameter estimation step after the HMM processing. For this reason, we have not included this last step in the simulation examples. We first incorporate any structural constraints anticipated in the model, by setting the required elements of the transition matrix to zero, e.g., as displayed in the reformulation of the semi-Markov model (19). We then proceed with the HMM processing without further constraints on the p_{ij}, until convergence of the algorithm is observed and yields the model estimate. We can then "fit" the assumed model to the estimated model in a least squares sense. In the case of CMM, this means obtaining values of the parameters $\{\zeta, \varrho, \kappa\}$ that give a best sum-of-squares fit to the estimated transition matrix. This is largely an ad hoc procedure except we note that if the model accurately represents the real process the fitting of the model parameters is a sound procedure. The transition probabilities of the process can be compared with those of the model using the estimated parameters to judge (both qualitatively and quantitatively) the accuracy of this extra parameter estimation step.

Once the particular model parameters have been obtained from the estimated model, one can perform a further pass of HMM processing, using the transition matrix corresponding to the fitted model. Thus it is possible to obtain (at least approximately) a likelihood value for the specific model in question, which will be less than that of the unconstrained model. This brings us to an important question which we discuss in the next section, namely: how should we compare models of different types with differing numbers of free parameters?

5. Ethic For Model Comparison

5.1 *Overview*

It is well known that the error in fitting a model to a given set of data decreases with the number of free parameters in the model. Thus it makes sense, in selecting a model from a set of models with different numbers of parameters, to penalise models having too many parameters. This is perhaps the most subjective part of any parameter estimation scheme. To give a well known example, in the identification of ARMA systems (LJUNG, 1987) many criteria are available to decide upon the "best" model, among the more common are Akaike's information theoretic criterion (AKAIKE, 1974) and the Schwarz criterion (SANSOM et al., 1989 and SCHWARZ, 1978). In summary, the choice of model from a set of possible models invariably depends on the way we choose to discriminate between them. Often, the selection criterion is combination of a measure of how

well the model fits the data, e.g. the likelihood; and the complexity of the model, e.g. the number of parameters. The question of how to penalise a HMM for having an excessive number of free parameters is an area of current research and one proposed criterion for model order selection is the compensated likelihood approach (FINESSO, 1990). Researchers in this area often focus on asymptotic estimators of the number of states. In practice the sample size is finite, and these estimators may exhibit bias in some cases.

5.2 *Log-Likelihood Comparison Criterion*

In this paper, we will appeal to the principle of parsimony in comparing models with different numbers of free parameters but similar structures. We measure the goodness of fit by evaluating the log likelihood of the model, and weigh this against what is to be gained by increasing the number of parameters, which generally increases the likelihood. Therefore, if a plot of log likelihood versus model order shows a "knee" for a certain model order[3], we would prefer this model to one of higher order. For a CMM, the model order is equal to N, although the number of free parameters is three for a fixed model order. In the case of a fractal model, the number of free parameters is four, but the order of the model depends on the value of the lag L we choose in the implementation (19). For a semi-Markov model of this form the order is $2L$ and there are $2L$ free parameters–the $p_{ii}(k)$, $k = 1, \ldots, L$, $i = 1, 2$. In the latter case there is an additional caveat–the estimates of the time-varying transition probabilities should form a "reasonably smooth" curve, i.e., one that is consistent with the transition probabilities of the assumed semi-Markov model. Note that a model with a large number of states but whose transition matrix structure is constrained need not necessarily have a higher likelihood than a model with fewer states and an unconstrained transition matrix.

In addition, we can ascertain a limited amount of information from the structure of the estimated transition matrix. If the HMM estimates indicate that certain states have very small average residence times or near zero transition probabilities, then this implies that we should delete these states from our model. We can now give some numerical examples illustrating the effectiveness of HMM processing both as an estimator and as an aid to discriminating between various model types.

---

[3] We are using "order" to mean the dimension of the transition matrix in a particular HMM implementation. This should not be confused with the order of the Markov chain, which is unity for all the models we consider.

6. Numerical Estimation Examples

6.1 *Simulation Environment and Parameters*

The HMM algorithm as well as the data generation module were coded in C for execution on a SUN workstation. The noisy simulated data were scaled by a factor of 204.8, equal to the product of amplifier and ADC gains, and then quantized to 12-bit integer values. To avoid numerical underflow in the calculations, the scaling and initialisation procedures described in CHUNG et al. (1990) were used. In all simulation examples, the standard deviation of the additive, white Gaussian noise was set to 0.5, and the length K of the data segment was 100000 points. For each chain, the closed state level was taken to be zero ($s_1 = 0$) and the fully open state level (s_2) as -2 in dimensionless units. For brevity of simulation, the state levels and noise variance were assumed to be known. In practice it is reasonable to expect that initial estimates of these quantities are available from visual examination or histograms of the data record prior to HMM processing. The accuracy of such estimates will affect the number of passes needed for convergence of the HMM algorithm.

6.2 *Data Generation*

The following models were used to generate simulated channel data:

(1) A 3-state ($N=3$) coupled Markov model with parameters $\zeta=0.99$, $\varrho=0.97$, $\delta=0.25$, $\kappa=0.4$; and state levels 0, -1, -2, with corresponding transition matrix:

$$\mathbf{P}_{CMM} = \begin{bmatrix} 0.9841 & 0.0119 & 0.0041 \\ 0.3178 & 0.5764 & 0.1058 \\ 0.0125 & 0.0349 & 0.9525 \end{bmatrix}. \tag{29}$$

(2) A discrete-time Colquhoun and Hawkes model with transition matrix:

$$\mathbf{P}_{CH} = \begin{bmatrix} 0.999 & 0.001 & 0 \\ 0.025 & 0.95 & 0.025 \\ 0 & 0.03 & 0.97 \end{bmatrix} \tag{30}$$

and state levels 0, 0, -2.

(3) A fractal semi-Markov model with fractal dimensions $D_1 = D_2 = 1.95$ and rate coefficients $k_1 = 0.5$, $k_2 = 1{,}0$, with state levels 0, -2 and lag of 10 samples.

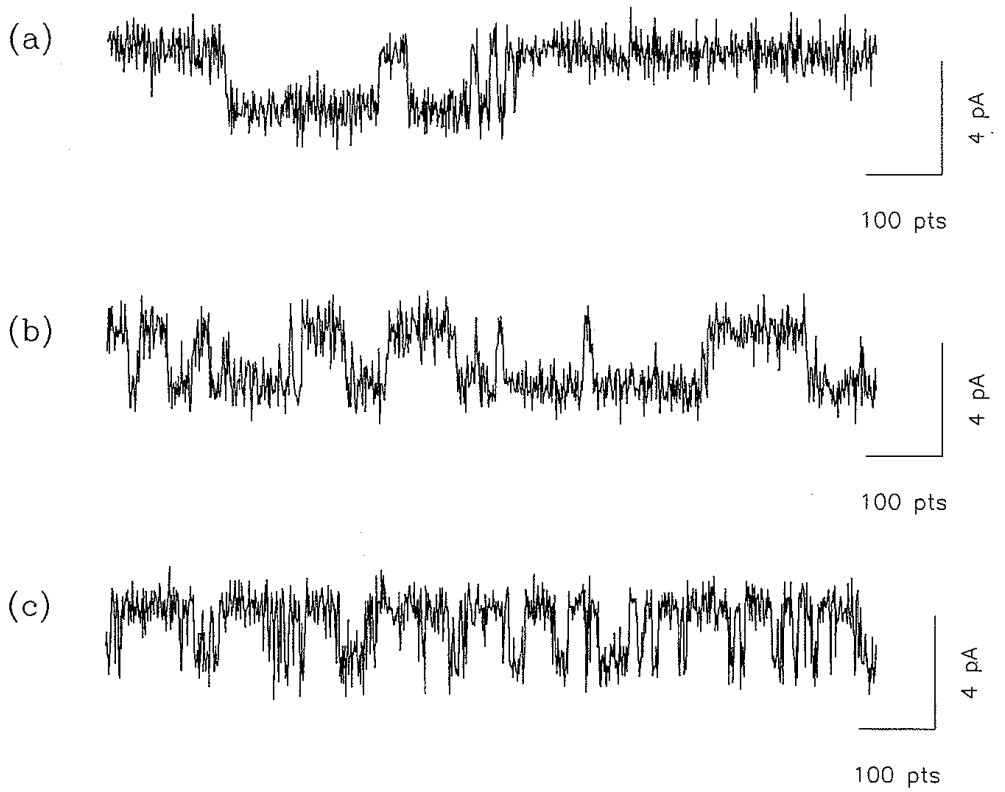

Fig. 1. Short segments of data generated according to three different signal models. Signal sequences were generated using (a) the 3-state coupled Markov model; (b) the 3-state Colquhoun-Hawkes model (b); and (c) a fractal semi-Markov model with a lag of 10 points. To make the data appear realistic, zero-mean Gaussian noise was added to all three signal sequences.

Figures 1a–1c show the noisy data generated by the models listed above. Although it may be argued that the trained eye could determine which signal corresponds to a particular model class, this distinction becomes increasingly difficult at low signal-to-noise ratios. Moreover, histogram-based techniques are quite sensitive to the level of noise. The results presented in the next part should convince the reader of the power of HMM-based techniques in the identification of model type and model parameters.

6.3 *Simulation Results*

For each of the three data sets in turn, and for a variety of model types, HMM processing was carried out until convergence of the log likelihood was observed. The assumed model types were 2 and 3 state CMMs (which have no disallowed transitions); a 3-state CHM; and SMMs with lag between 4 and 20 samples.

For computational simplicity, we illustrate the method using CMMs and CHMs having up to 3 states, and use a relatively high signal-to-noise ratio. The

log likelihoods obtained for models in the same class always display a "knee" as the number of states increases, and this property is adequately illustrated by the SMM examples. We emphasize that the technique is applicable to chains with more than 3 states and to data with much lower signal-to-noise ratios. Of course, such cases require considerably more computation. For this reason also, the SMM simulations, which involve chains with up to 32 states, were only carried out over 20 passes, which is adequate for convergence of the log likelihood to 3 decimal places.

The bar graphs of Fig. 2 and tables 2–4 show the log likelihoods for the various model types obtained by HMM processing of the three data sets. The correct model is indicated by a dagger (†). Concentrating on the first of these, Table 2 and Fig. 2a, which show the log likelihoods for the CMM data, it is clear that the 3-state CMM (with no disallowed transitions) gives the best fit of all the models tested. Recall that the fit should improve monotonically as the number of states and hence free parameters in the model increases. The other models provide a fit to the data which is significantly worse–a difference of about 10^{60} in likelihood. The estimated transition matrix for the 2- and 3-state CMMs were

$$\hat{\mathbf{P}}_{2CMM}=\begin{bmatrix}0.9820 & 0.0180\\ 0.0282 & 0.9718\end{bmatrix};\quad \hat{\mathbf{P}}_{3CMM}=\begin{bmatrix}0.9824 & 0.0125 & 0.0050\\ 0.0967 & 0.5305 & 0.3728\\ 0.0176 & 0.0335 & 0.9488\end{bmatrix}.$$

These may be compared with the generating transition matrix (29). The differences in the 3×3 case are attributable to the finite length of the record–10^5 points–and the relatively low signal-to-noise ratio. Fitting the estimated transition matrix to give a true coupled Markov model resulted in only a small decrease in the log likelihood, although this is not always the case.

Table 2

Log likelihood values and number of passes for CMM data. The number of passes shown is the number required for convergence to 7 decimal places of the log likelihood, except where indicated by a star (*) in which case the convergence is to 3 decimal places only.

Assumed Model Type	Log Likelihood	No. of Passes
2-state CMM	−268732.49	12
† 3-state CMM	−268018.40	35
3-state CHM	−268643.17	700
SMM lag 4	−268637.78	20*
SMM lag 6	−268637.10	20*
SMM lag 8	−268636.74	20*
SMM lag 10	−268635.73	20*
SMM lag 12	−268633.35	20*
SMM lag 14	−268632.51	20*
SMM lag 16	−268630.33	20*

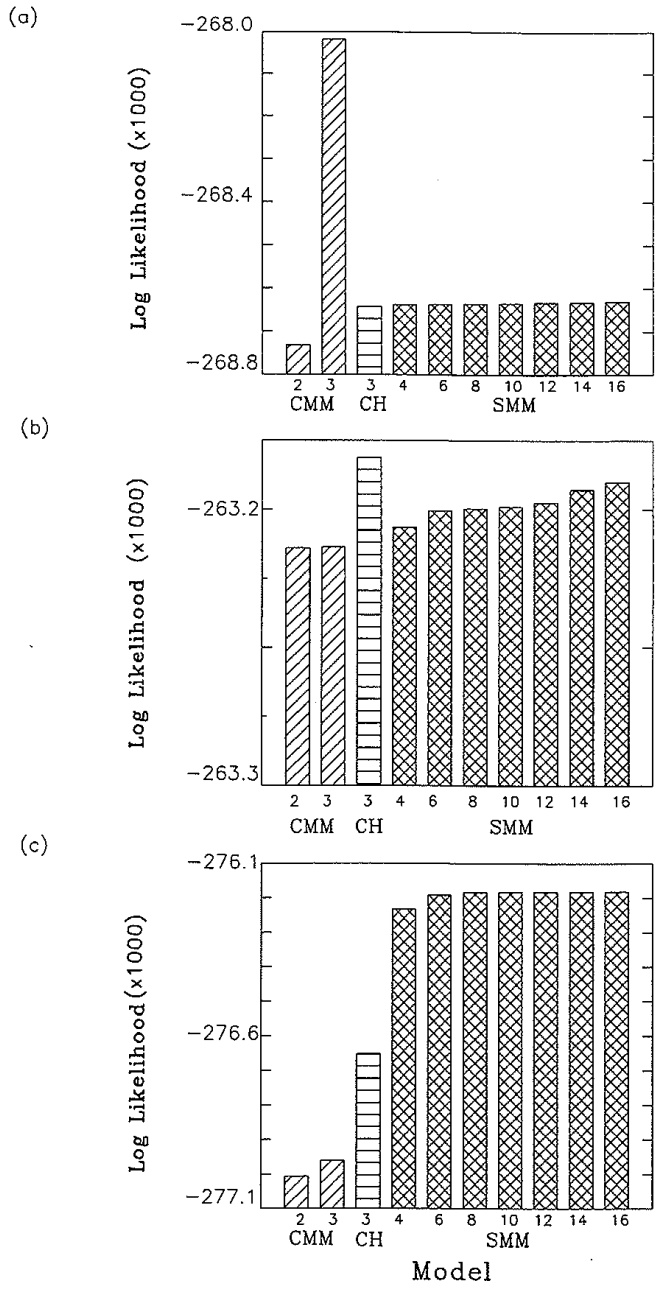

Fig. 2. Log likelihoods for models based on the CMM, CHM or SMM data. The observation sequence generated by the coupled Markov model was analyzed using a HMM tailored to the three different signal models.
(a) The likelihood that the data were generated by a coupled Markov model (cross hatched) is the largest, as indicated on the bar graph. The true model was a 3-state coupled Markov model. (b) The highest likelihood is obtained when the observation sequence is assumed to be generated by a Colquhoun-Hawkes model (horizontal shading), as was the case in this simulation. (c) Using a HMM based on the assumption that the underlying generating mechanism is fractal in nature gives the highest likelihood values (cross hatches). As the lag of the SMM is increased, the model likelihood increases, but only marginally once the true model order (lag 10) is reached.

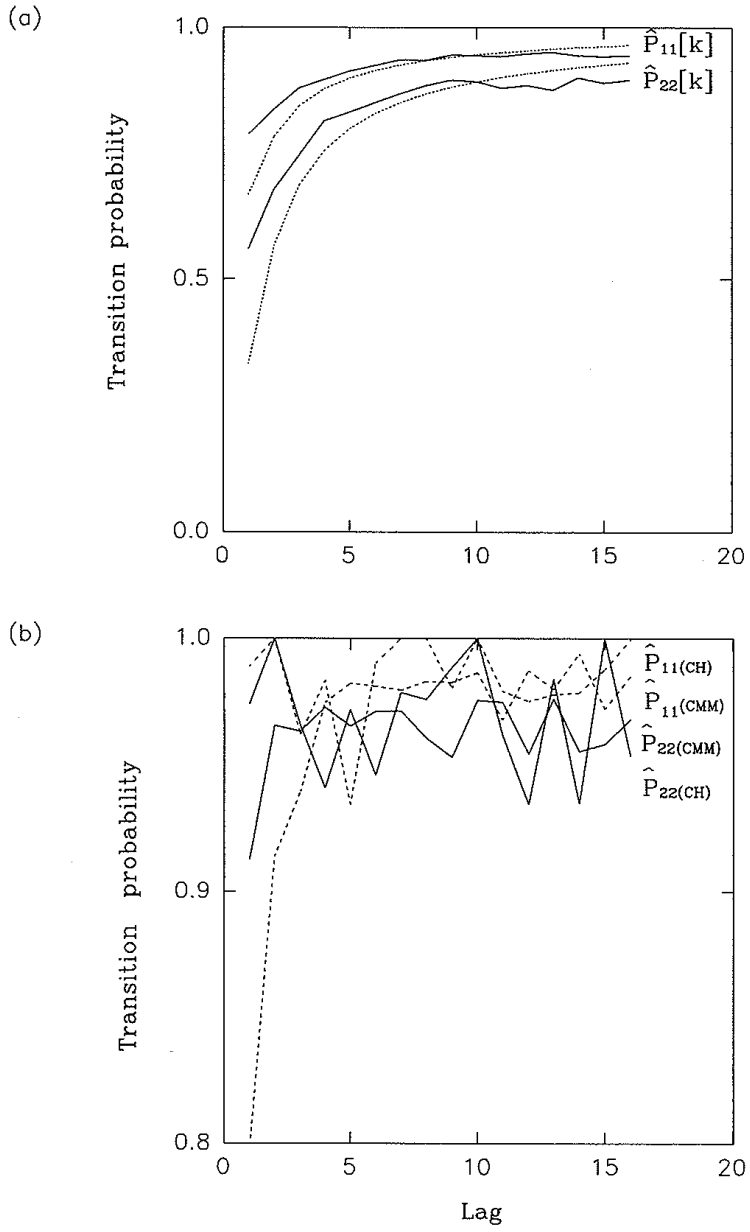

Fig. 3. Semi-Markov model estimates for the data generated by the semi-Markov model, Colquhoun-Hawkes model and coupled Markov model.
(a) Estimated transition probabilities for the fractal semi-Markov model are plotted against lag. Thus, by employing a HMM based on the assumption of time-varying transition probabilities, it is possible to estimate the form of the time variation. The closeness of the estimated curves (solid lines) to their true values (dotted lines) indicates the accuracy of the technique.
(b) Semi-Markov model estimates for the Colquhoun and Hawkes and CMM transition probabilities are plotted against lags. Although the estimates vary wildly, their variation is bounded in the range 0.9 to 1.0. One has the general impression that a model with constant transition probabilities would have been more appropriate, as is the case for both model types.

Turning to the CHM results in Table 3 and Fig. 2b, we see that the 3-state CHM gives the highest log likelihood of those tested, although the differences are less marked than in the previous example. The estimated transition matrix for the 3-state CHM was

$$\hat{\mathbf{P}}_{CH} = \begin{bmatrix} 0.9987 & 0.0013 & 0 \\ 0.03448 & 0.9400 & 0.0252 \\ 0 & 0.0357 & 0.9643 \end{bmatrix},$$

which is close to the true transition matrix (30), indicating that the HMM algorithm was able to estimate transition probabilities between states in the same aggregated class. The estimated transition matrix for the 3-state CMM was

$$\hat{\mathbf{P}}_{3CMM} = \begin{bmatrix} 0.9984 & 0.0007 & 0.0009 \\ 1 & 0 & 0 \\ 0.0353 & 0 & 0.9647 \end{bmatrix},$$

implying that the state 2 at -1 was in fact not statistically supported by the data. The estimated transition probabilities for the lag 16 SMM based on the CHM data are shown in Fig. 3b. The estimates are unsteady but point to the fact that the true model did in fact have constant transition probabilities.

Table 3
Log likelihood values and number of passes for CHM data.

Assumed Model Type	Log Likelihood	No. of Passes
2-state CMM	−263213.88	10
3-state CMM	−263213.56	600
† 3-state CHM	−263181.23	661
SMM lag 4	−263206.41	20
SMM lag 6	−263200.42	20*
SMM lag 8	−263199.63	20*
SMM lag 10	−263198.87	20*
SMM lag 12	−263197.28	20*
SMM lag 14	−263192.72	20*
SMM lag 16	−263190.06	20*

The model likelihoods for the SMM data (generated by a fractal model with lag 10) are displayed in Table 4 and Fig. 2c. Clearly the semi-Markov models have the upper hand in this case. The estimated time-varying transition probabilities are shown in Fig. 3a and these clearly demonstrate a smooth, exponential increase, as expected for a fractal model. The roughness of the estimates for large

Table 4
Log likelihood values and number of passes for SMM data.

Assumed Model Type	Log Likelihood	No. of Passes
2-state CMM	−277007.17	11
3-state CMM	−276961.06	700
3-state CHM	−276652.27	450
SMM lag 4	−276232.65	12
SMM lag 6	−276191.14	12
SMM lag 8	−276183.83	12
† SMM lag 10	−276183.09	14
SMM lag 12	−276182.90	13
SMM lag 14	−276182.10	15
SMM lag 16	−276182.07	15

values of lag is a result of the shortness of the data record. The estimated transition matrices for the CMM and CHM,

$$\hat{\mathbf{P}}_{2CMM}=\begin{bmatrix}0.9199 & 0.0801\\ 0.2318 & 0.7682\end{bmatrix};\quad \hat{\mathbf{P}}_{CHM}=\begin{bmatrix}0.8853 & 0.1147 & 0\\ 0.1810 & 0.6083 & 0.2103\\ 0 & 0.2373 & 0.7627\end{bmatrix}.$$

on the other hand would lead us to suspect that a homogeneous Markov model was perhaps inappropriate for this data set. This example serves to demonstrate the "knee phenomenon" alluded to earlier. The log likelihood for the SMM models increases monotonically: rapidly at first (for lags 4–8), and only incrementally for lags above 10. The actual lag was 10, so that the likelihood curve is an indicator of the order of the model. Similar arguments can be brought to bear on the number of states in homogeneous Markov models.

7. Discussion

We have presented a method based on hidden Markov models that is applicable not only in the estimation of model parameters for a given model structure, but also in comparing models from different classes. In illustrating this concept, we chose three types of models that are relevant in the field of biological channel modelling, namely, coupled Markov models, discrete-time Colquhoun and Hawkes models, and semi-Markov models. The CHM class is representative of aggregated, homogeneous Markov models with disallowed transitions, while SMMs can deal with time-varying transition probabilities. The simulation results support the empirical observation that the log likelihood is a satisfactory measure of proximity of the model to the data, and additionally that the knee in the likelihood curve gives a fair indication of the most suitable number of states or order of the model. We do not claim to have demonstrated the latter point in

an analytical sense, rather it is an observation that appears to be valid at least for the limited class of models we consider.

We repeat that, during estimation, we can constrain the structure of the transition matrix by setting some elements to zero, but the transition probabilities are otherwise freely varying. A further fitting step would be required to determine the parameters of a CMM or fractal model which are close in some sense to the estimated transition matrix. An example of this model fitting may be found in KENNEDY and CHUNG (1992). It should be clear, however, that the optimal model structure can be estimated by our method. Strictly speaking, the incorporation of models which are parametrised by fewer quantities than the set of their transition probabilities involves a redesign at the EM algorithm level, and this is a subject for future research.

A further model, not explicitly treated here, is the higher order scalar Markov chain, in which the transition probabilities depend on more than one past state. A higher order Markov chain can always be recast as a first-order vector Markov chain using the simple observation that for any stochastic process $\{x_k\}$

$$\begin{aligned}\Pr(x_{k+1} | x_k, \ldots, x_{k-N}) &= \Pr(x_{k+1}, x_k, \ldots, x_{k-N+1} | x_k, \ldots, x_{k-N}) \\ &= \Pr(\mathbf{X}_{k+1} | \mathbf{X}_k),\end{aligned} \tag{31}$$

where $\mathbf{X}_k = (x_k, \ldots, x_{k-N})$ (see, e.g., BILLINGSLEY (1961)). In other words, by relabelling the states we can obtain a first-order Markov chain. Owing to the replication of some elements in $\mathbf{X}_{k+1}$ and $\mathbf{X}_k$ some transition probabilities in the resulting first-order process for the vector z_k will be zero. A similar remark concerning the structure of the transition matrix holds in the case of the fractal model. This has important computational consequences in that only the non-zero transition probabilities need be updated in the parameter estimation scheme, despite the usually large number of states. Our treatment of semi-Markov models parallels the latter in its reformulation of the problem as a vector Markov chain with a particular structure.

The implication of our results for real channel currents is that it should now be possible to ascertain what type of signal model is most appropriate and consistent with the data. Theoretically, the best model in the HMM framework is the one which maximises the likelihood function of the data. It is hoped that this mathematical tool, with appropriate application, may add to our understanding of the physical mechanisms underlying ion transport in cell membranes.

Acknowledgement

We would like to thank Dr. Vikram Krishnamurthy for his helpful consultation on some technical aspects of hidden Markov modelling. This work was supported in part by grants from the Ramaciotti Foundation and NH&MRC of Australia.

References

AKAIKE, H., 1974: A new look at the statistical model identification. IEEE Trans. Auto. Control **AC-19,** 716–723.

BALL, F., and 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., MILNE, R. K., and YEO, G. F., 1991: Aggregated semi-Markov processes incorporating time interval omission. Adv. Appl. Prob. **23,** 727–797.

BALL, F., YEO, G. F., MILNE, R. K., EDESON, R. O., MADSEN, B. W., and SANSOM, M. S. P., 1993: Single ion channel models incorporating aggregation and time interval omission. Biophys. J. **64,** 357–374.

BAUM, L. E., and PETRIE, T., 1966: Statistical inference for probabilistic functions of finite state Markov chains. Ann. Math. Stat. **37,** 1554–1563.

BAUM, L. E., PETRIE, T., SOULES, G., and WEISS, N., 1970: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann. Math. Stat. **41,** 164–171.

BHARUCHA-REID, A. T., 1960: *Elements of the Theory of Markov Processes and Their Applications.* New York: McGraw-Hill.

BILLINGSLEY, P., 1961: *Statistical Inference for Markov Processes.* Chicago: University of Chicago Press.

CALLIER, F. M., and DESOER, C. A., 1991: *Linear System Theory.* New York: Springer-Verlag.

CHUNG, S. H., MOORE, J. B., XIA, L., PREMKUMAR, L. S., and GAGE, P. W., 1990: Characterization of single channel currents using digital signal processing techniques based on hidden Markov models. Phil. Trans. R. Soc. Lond. **B 329,** 265–285.

CHUNG, S. H., KRISHNAMURTHY, V., and MOORE, J. B., 1991: Adaptive processing techniques based on hidden Markov model for characterizing very small channel currents buried in noise and deterministic interferences. Phil. Trans. R. Soc. Lond. **B 334,** 357–384.

COLQUHOUN, D., and HAWKES, A. G., 1977: Relaxation of membrane currents that flow through drug-operated channels. Proc. R. Soc. Lond. **B 199,** 231–262.

COLQUHOUN, D., and HAWKES, A. G., 1981: On the stochastic properties of single ion channels. Proc. R. Soc. Lond. **B 211,** 205–235.

COLQUHOUN, D., and 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.

COX, D. R., 1962: *Renewal Theory.* London: Methuen.

FELLER, W., 1966: *An Introduction to Probability Theory and its Applications.* Volume 2. New York: John Wiley.

FINESSO, L., 1990: *Consistent Estimation of the Order for Markov and Hidden Markov Chains.* Ph. D. Dissertation, University of Maryland, U.S.A.

FREDKIN, D. R., and RICE, J. A., 1992a: Bayesian restoration of single channel patch clamp recordings. Biometrics **48,** 427–448.

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

GANTMACHER, F. R., 1974: *The Theory of Matrices,* Vol. 2. New York: Chelsea Publishing Company.

HOWARD, R. A., 1960: *Dynamic Programming and Markov Processes.* New York: John Wiley.

ISAACSON, L., and MADSEN, R. W., 1985: *Markov Chains, Theory and Applications.* New York: John Wiley.

KAILATH, T., 1980: *Linear Systems.* Englewood Cliffs, N.J.: Prentice-Hall.

KENNEDY, R. A., and CHUNG, S. H., 1992: Identification of coupled Markov chain models with application. Proc. 31st Conf. Decision & Control **4,** 3529–3534.

KORN, S. J., and HORN, R., 1988: Statistical discrimination of fractal and Markov models of single-channel gating. Biophys. J. **54,** 871–877.

KRISHNAMURTHY, V., MOORE, J. B., and CHUNG, S. H., 1991: On hidden fractal signal processing. Signal Processing **24,** 177–192.

LIEBOVITCH, L. S., FISCHBARG, J., and KONIAREK, J. P., 1986: Optical correlation functions applied to the random telegraph signal: how to analyze patch clamp data without measuring the open and closed times. Math. Biosciences **78,** 203–215.

LIEBOVITCH, L. S., FISCHBARG, J., KONIAREK, J. P., TODOROVA, I., and WANG, M., 1987: Fractal model of ion-channel kinetics. Biochim. et Biophys. Acta **896,** 173–180.

LIEBOVITCH, L. S., 1989: Testing fractal and Markov models of ion channel kinetics. Biophys. J. **55,** 373–377.

LIEBOVITCH, L. S., and SULLIVAN, J. M., 1987: Fractal analysis of a voltage-dependent potassium channel from cultured mouse hippocampal neurons. Biophys. J. **52,** 979–988.

LJUNG, L., 1987: *System Identification: Theory for the User*. New Jersey: Prentice Hall.

RABINER, L. R., 1989: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77** (no. 2). 257–285.

SANSOM, M. S. P., BALL, F. G., KERRY, C. J., MCGEE, R., RAMSEY, R. L., and USHERWOOD, N. R., 1989: Markov, fractal, diffusion, and related models of ion channel gating – a comparison with experimental data from two ion channels. Biophys. J. **56,** 1229–1243.

SCHWARZ, G., 1978: Estimating the dimension of a model. Ann. Statist. **6,** 461–464.

SMITH, W. L., 1955: Regenerative stochastic processes. Proc. Roy. Soc. Lond. **A 232,** 6–31.

TITTERINGTON, D. M., SMITH, A. F. M., and MAKOV, U. E., 1985: *Statistical Analysis of Finite Mixture Distributions*. New York: John Wiley.

Received, Nov., 1993
Revised, Febr., 1994
Accepted, March, 1994

Dr. SHIN-HO CHUNG
The Australian National University
Protein Dynamics Unit
Dept. of Chemistry
Canberra ACT 0200
Australia