THE ROYAL SOCIETY

Estimating transition rates in aggregated Markov models of ion channel gating with loops and with nearly equal dwell times

Mirko Wagner*, **Steffen Michalek and Jens Timmer**

Freiburger Zentrum für Datenanalyse und Modellbildung, Eckerstraße 1, D–79104 Freiburg, Germany

A typical task in the application of aggregated Markov models to ion channel data is the estimation of the transition rates between the states. Realistic models for ion channel data often have one or more loops. We show that the transition rates of a model with loops are not identifiable if the model has either equal open or closed dwell times. This non-identifiability of the transition rates also has an effect on the estimation of the transition rates for models which are not subject to the constraint of either equal open or closed dwell times. If a model with loops has nearly equal dwell times, the Hessian matrix of its likelihood function will be ill-conditioned and the standard deviations of the estimated transition rates become extraordinarily large for a number of data points which are typically recorded in experiments.

Keywords: non-identifiability; aggregated Markov models; parameter estimation; ion channel gating; maximum likelihood

1. INTRODUCTION

The so-called patch-clamp technique allows the recording of single ion channel currents (Neher & Sakmann 1976; Hamill *et al*. 1981). The observed current switches rapidly for most ion channels between two conductance levels because an ion channel performs transitions among a number of unobserved open and closed states. The transitions among these states are usually described by a Markov chain in continuous time due to the underlying physics of the transitions (Colquhoun & Hawkes 1977). For most ion channels, there are more than one open and closed state, respectively, necessary to model adequately the gating of the ion channel. Because it is only possible to observe if the channel is either in an open or a closed state, but not in which one, the observed ion current is an aggregated image of the underlying process, which is modelled by an aggregated Markov process (Colquhoun & Hawkes 1981; Colquhoun & Sigworth 1995; Fredkin *et al*. 1983). The transitions between the unobserved states correspond to physiological processes such as changes in the geometrical conformation of the channel protein or the binding of ligand molecules to receptor sites on the channel protein (Hille 1992). Therefore, not every transition between the states is possible, but a characteristic gating scheme determines the dynamics of the ion channel data. The gating scheme together with the transition rates between the allowed transitions parameterize an aggregated Markov model.

For a given gating scheme, the transition rates have to be estimated from the data. This can be achieved by the maximum-likelihood method (Fredkin & Rice 1992; Albertsen & Hansen 1994; Machalek & Timmer 1999). It was shown recently that under mild regularity conditions the maximum-likelihood estimator is asymptotically normally distributed with a covariance matrix given by the inverse of the limiting covariance matrix of the score function (Bickel *et al*. 1998). In particular, the covariance matrix of the maximum-likelihood estimator can be estimated by the inverse of the Hessian matrix of the likelihood function at the maximum-likelihood point.

Because of the aggregation of the states, the maximum number of parameters which can be estimated from the data is limited to twice the number of open states multiplied by the number of closed states (Fredkin *et al*. 1983; Fredkin & Rice 1986). For instance, a three-state model which is composed of one closed and two open states and where all states are interconnected, has six parameters. It always has the same observable outcome as a suitable chosen three-state model with only four parameters, where the states are ordered in a linear chain (Kienker 1989). Consequently, the transition rates in a three-state gating scheme with a loop are not identifiable.

However, gating schemes with one or more loops are often used to model realistic ion channel data (Horn & Lange 1983; Ball & Sansom 1989; Bates *et al*. 1990; Vandenberg & Bezanilla 1991). At least two open and two closed states are necessary to obtain a gating scheme with one loop and with all transition rates identifiable. The simplest gating scheme fulfilling these requirements is shown in figure 1. Any other gating scheme with two open and two closed states is either equivalent to the gating scheme shown in figure 1 or its transition rates are not identifiable (see figure 2). The problem of non-identifiable transition rates is aggravated if the analysis of ion channel data is only based on the marginal distributions of the open times and the closed times, respectively (Edeson *et al*. 1994).

The concept of aggregated Markov models can be generalized to hidden Markov models which also incorporate the

*Author for correspondence (mirko@fdm.uni-freiburg.de).

Received 10 May 1999 *Accepted* 22 June 1999

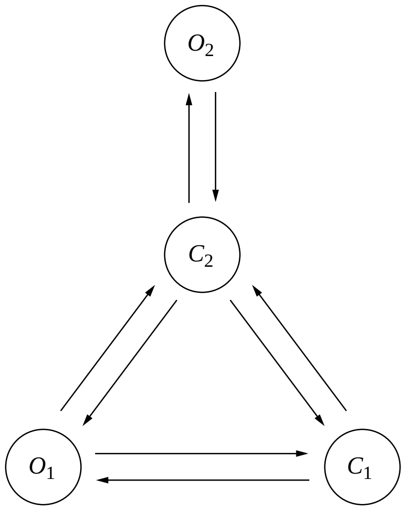

Figure 1. Loop-gating scheme: gating scheme with one loop and four states. We denote an open and a closed state by 'O' and 'C', respectively.

noise on the data (Chung *et al.* 1990, 1991; Fredkin & Rice 1997). Hidden Markov models can be further extended by taking into account the perturbation of the measured ion channel data caused by filtering (Venkataramanan *et al.* 1998; Michalek *et al.* 1998, 1999).

This paper is organized as follows. In § 2 we show for a simple four-state loop model that the transition rates are not identifiable, if the model has either equal open or equal closed times. In § 3 we examine the consequence of the non-identifiability for the estimation of the transition rates in the case that the true model has only nearly equal dwell times.

2. EQUAL DWELL TIMES

We examine a four-state aggregated Markov model with equal open times and with a gating scheme given by figure 1, called the 'loop-gating scheme' in the following. The case of equal closed times is analogous. The results derived for this simple model are also valid for every more complicated aggregated Markov model with loops which contains the loop-gating scheme as a sub-model. Moreover, we expect that for every other aggregated Markov model with loops the results are also true but harder to prove.

In the following, we assume that the ion channel current is observed at discretely sampled time points resulting in a time-series of data points Y_i, $i = 1, \ldots, N$ where every Y_i can only take two values for the two possible outcomes 'open' and 'closed'. The parameter vector of the aggregated Markov model is denoted by θ. For our analysis, we will use the likelihood function of the observed time-series $Y_1, \ldots, Y_N$:

$$L_{Y_1, \ldots, Y_N}(\theta) = P_\theta(Y_1, \ldots, Y_N). \quad (1)$$

$P_\theta(Y_1, \ldots, Y_N)$ is the joint probability distribution function of the data given the parameter vector θ. The analysis of aggregated Markov models is often not based on the likelihood function given by equation (1) but based on the likelihood function of the observed series of channel dwell times (Horn & Lange 1983; Fredkin *et al.* 1983; Kienker 1989; Ball & Sansom 1989; Colquhoun & Sigworth 1995). For the case of aggregated Markov models, we can equally use both functions. However, arguments which are based on the likelihood function of the observed data points and not on the likelihood function of the observed series of dwell times are more easily extended to the case of hidden Markov models. For practical purposes, artefacts caused by filtering have to be considered. In the case of aggregated Markov models, this can be achieved by taking into account a possible time interval omission (Ball *et al.* 1993; Colquhoun *et al.* 1996); in the case of hidden Markov models the filter artefacts can be incorporated in the signal model (Michalek *et al.* 1998, 1999).

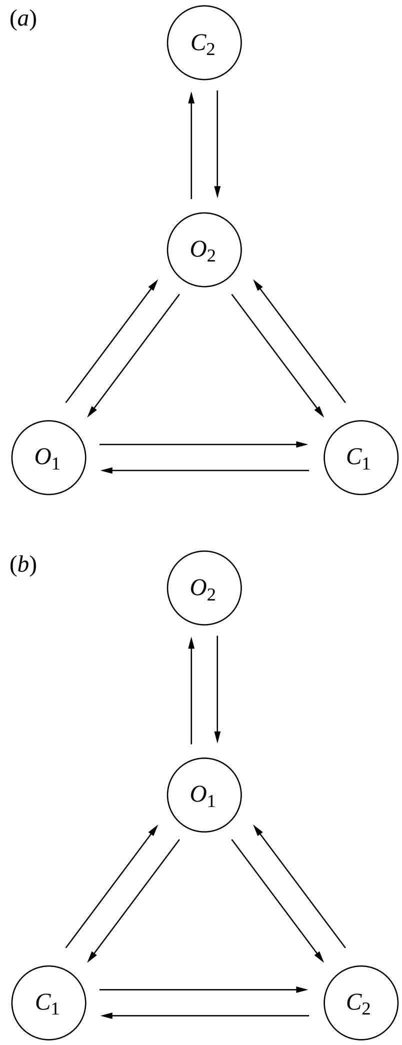

Figure 2. Possible alternatives to the loop-gating scheme: Gating scheme (a) is equivalent to the loop-gating scheme. The transition rates of gating scheme (b) are not identifiable because this gating scheme has exactly one gateway state (O_1) (Fredkin *et al.* 1983).

In the following, we assume that an aggregated Markov model with a generator matrix Q which is compatible with the loop-gating scheme, is parameterized in a way so as to impose no other constraint than ensuring equal open times. The parameters of such an aggregated Markov model are not identifiable in the following sense: it is possible to perform a slight variation of the parameters leading to a new generator matrix which is also compatible with the loop-gating scheme and whose corresponding aggregated Markov model has the same joint probability distribution as the original model. Indistinguishable aggregated Markov models are characterized by the following lemma (Kienker 1989).

Lemma 1 (Kienker 1989). *Let Q' and Q be the generator matrices describing two regular Markov models. If the two generator matrices Q and Q' are related by a similarity transformation: $Q' = S^{-1}QS$ where S is the form:*

$$\left(\begin{array}{c|c} S_{oo} & 0 \\ \hline 0 & S_{cc} \end{array}\right), \tag{2}$$

and each row of S sums to unity, then the joint probability distributions are the same for both models.

We split the proof of the non-identifiability of the transition rates into two parts. In the first part, we show that the Hessian matrix of the likelihood function is singular at the maximum-likelihood point. This result will give us an idea, which we will use in the second part, on how to obtain a family of transformation matrices $S(\epsilon)$ which depends on a continuous parameter ϵ and includes the identity transformation for $\epsilon = 0$ with the following properties: the transformed generator matrices $Q'(\epsilon) = S(\epsilon)^{-1}QS(\epsilon)$ are compatible with the loop-gating scheme, and the corresponding aggregated Markov models have the same joint probability distribution.

(a) *A sufficient condition for the singularity of the Hessian matrix*

We formulate a sufficient condition for the Hessian matrix of the likelihood function to be singular at the maximum-likelihood point. In general, the likelihood function L is a function of an arbitrary generator matrix Q: $L = L(Q)$—the obvious dependency on the data is omitted. The generator matrices of aggregated Markov models which are compatible with a certain gating scheme can be parameterized in an arbitrary manner for lemma 2 by a parameter vector θ: $Q = Q(\theta)$. Most commonly, the vector θ contains a subset of the transition rates of the aggregated Markov model.

Lemma 2. *Let $\hat{\theta}$ be a maximum-likelihood point for a given aggregated Markov model and given data. Now, consider the likelihood function as a function of all entries of the generator matrix Q: $L = L(Q)$ instead of just the parameters θ. Assume the existence of a twice continuously differentiable matrix-valued function $g(\epsilon)$ depending on a real parameter ϵ with the following properties:*

(i) $g(\epsilon = 0) = Q(\hat{\theta})$,
(ii) $g(\epsilon) = Q(\hat{\theta}) + \epsilon C_1 + \epsilon^2 C_2 + O(\epsilon^3)$, *where C_1 and C_2 denote some matrices with entries independent of ϵ,*
(iii) $L(g(\epsilon)) = L(Q(\hat{\theta}))$,
(iv) *There exists an interval $[-\epsilon_0, \epsilon_0]$ around zero so that for every ϵ in this interval the second-order approximation of $g(\epsilon)$: $Q(\hat{\theta}) + \epsilon C_1 + \epsilon^2 C_2$ is a generator matrix which is compatible with the gating scheme of the given aggregated Markov model.*

Then the Hessian matrix of the likelihood function $L(\theta)$ is singular at $\hat{\theta}$.

Lemma 2 is proven by substituting the Taylor expansion of g into (iii) and comparing coefficients corresponding to the same power of ϵ. Note that $g(\epsilon)$ for $\epsilon \neq 0$ does not need to satisfy the constraints imposed by the given parameterization; in fact, it does not even need to be a generator matrix at all.

(b) *The loop model with equal open times*

In this section we will prove that the Hessian matrix of a loop model with equal open times is singular at the maximum-likelihood points by using lemma 2. It is proven in lemma 1 that $L(S^{-1}Q\,S) = L(Q)$ for every matrix S of the same form as in equation (2). We can therefore fulfil assumption (iii) of lemma 2 by choosing the matrix-valued function $g(\epsilon)$ as

$$g(\epsilon) = S^{-1}(\epsilon)Q\,S(\epsilon), \tag{3}$$

where $S(\epsilon)$ is a transformation matrix for every ϵ in the sense of lemma 1.

By equation (3), the task to find a suitable function $g(\epsilon)$ is transfered to the problem to determine the transformation matrices $S(\epsilon)$. The further premises of lemma 2 require us to consider the Taylor expansion of the function $g(\epsilon)$. Therefore, we will expand equation (3) step by step to higher orders in the following. Because every order will pose restrictions on $S(\epsilon)$, we will start with the most general form of the transformation matrix $S(\epsilon)$ and take into account the restrictions posed on $S(\epsilon)$ for every order in the Taylor series.

To keep the calculations simple, it is useful to parameterize generator matrices compatible with the loop-gating scheme and with equal open times in the following way:

$$Q = \left(\begin{array}{c|c} Q_{oo} & Q_{oc} \\ \hline Q_{co} & Q_{cc} \end{array}\right) = \left(\begin{array}{cc|cc} -q_{13}b & 0 & q_{13} & q_{13}(b-1) \\ 0 & -q_{13}b & 0 & q_{13}b \\ \hline q_{31} & 0 & -q_{31}-q_{34} & q_{34} \\ q_{41} & aq_{31} & q_{43} & -q_{41}-aq_{31}-q_{43} \end{array}\right). \tag{4}$$

The q_{ij} terms and a are constrained to be positive; b has to be greater than unity.

Because of condition (i) of lemma 2, the transformation matrix $S(\epsilon = 0)$ must be the identity transformation. Expanding S to first order in ϵ, we can approximate S as follows: $S(\epsilon) \approx I + \mathcal{J} + \ldots$, where I denotes the identity transformation and $\mathcal{J}$ the first-order term in the Taylor expansion. Accordingly, we can rewrite equation (3) for small values of ϵ,

$$g(\epsilon) = Q' \approx Q + \underbrace{Q\mathcal{J} - \mathcal{J}Q}_{Q_1'} + \ldots. \tag{5}$$

It is a necessary condition to fulfil assumption (iv) in lemma 2 that the first-order approximation $Q + Q_1'$ must be compatible with the loop-gating scheme. This is ensured by choosing $\mathcal{J}$ as

$$\mathcal{J}(\epsilon_1, \epsilon_2) = \left(\begin{array}{c|c} \mathcal{J}_o & 0 \\ \hline 0 & \mathcal{J}_c \end{array}\right) = \left(\begin{array}{cc|cc} -a\epsilon_1 & a\epsilon_1 & 0 & 0 \\ b\epsilon_2 & -b\epsilon_2 & 0 & 0 \\ \hline 0 & 0 & -\epsilon_1 & \epsilon_1 \\ 0 & 0 & \epsilon_2 & -\epsilon_2 \end{array}\right), \tag{6}$$

depending on two real parameters ϵ_1, ϵ_2.

From lemma 1, $\mathcal{J}$ has four free parameters in general. However, the requirement to be compatible with the loop-gating scheme imposes four constraints, namely that the entries q'_{12}, q'_{21}, q'_{23} and q'_{32} of the transformed generator matrix Q' have to vanish. Therefore, one would expect that $\mathcal{J}$ is fully determined or that even a matrix $\mathcal{J}$ which meets the constraints does not exist. However, in the case of equal dwell times, the restrictions of vanishing entries (1,2) and (2,1), respectively, are always fulfilled because the sub-matrix Q_{oo} is proportional to the 2×2 identity matrix and therefore it commutes with $\mathcal{J}_{\text{o}}$. Thus, the loop-gating scheme constraints reduce the number of parameters in $\mathcal{J}$ only by two. The restrictions given by the higher-order terms in the Taylor series equation (3) will reduce the number of parameters in $\mathcal{J}$ to one, which will be shown in the following.

For the expansion of equation (3) to higher orders, we use the following parameterization of the transformation matrices S:

$$S_f(\epsilon_1,\epsilon_2) = f(\mathcal{J}(\epsilon_1,\epsilon_2)), \quad f(0)=1 \text{ and } f'(0)=1, \tag{7}$$

where f is an arbitrary analytical and invertible function to be determined in the following. This approach has the nice property that $(S_f^{-1}(\epsilon_1,\epsilon_2)QS_f(\epsilon_1,\epsilon_2))_{\text{oo}} = Q_{\text{oo}}$ for arbitrary ϵ_1,ϵ_2 because

$$\begin{aligned}(S_f^{-1}(\epsilon_1,\epsilon_2)QS_f(\epsilon_1,\epsilon_2))_{\text{oo}} &= f^{-1}(\mathcal{J}_{\text{o}})\,(-q_{13}bI)f(\mathcal{J}_{\text{o}})\\ &= f^{-1}(\mathcal{J}_{\text{o}})f(\mathcal{J}_{\text{o}})Q_{\text{oo}}\\ &= Q_{\text{oo}}.\end{aligned} \tag{8}$$

Expanding equation (3) to the second order using equation (7), we obtain

$$\begin{aligned}Q' = Q + &\underbrace{Q\mathcal{J} - \mathcal{J}Q}_{Q'_1}\\ &+\underbrace{\frac{f''}{2}Q\mathcal{J}^2 - \mathcal{J}Q\mathcal{J} + \frac{1}{2}(2-f'')\mathcal{J}^2Q}_{Q'_2} + \ldots,\end{aligned} \tag{9}$$

where f'' denotes the second derivative of f at point 0.

$Q + Q'_1 + Q'_2$ must meet the constraints by the loop-gating scheme to satisfy the assumption (iv) of lemma 2. This implies further restrictions on ϵ_1, ϵ_2 and on the second-order coefficient f'' of the Taylor expansion of f. The constraint that the sub-matrix $(Q+Q'_1+Q'_2)_{\text{oo}}$ has to be of diagonal shape, is fulfilled for arbitrary ϵ_1, ϵ_2 and f'' applying the same arguments as in equation (8). Therefore, the two remaining restrictions require the entries (2,3) and (3,2) of $Q+Q'_1+Q'_2$ to vanish. These restrictions are given by the following pair of equations:

$$\begin{aligned}\frac{1}{2}bq_{13}\epsilon_2[&(-f''+af''+2-2a)\epsilon_1\\ &+(f''+bf''+2-2b)\epsilon_2] = 0,\end{aligned} \tag{10}$$

$$\begin{aligned}\frac{1}{2}aq_{31}\epsilon_1\Bigg[&\left(-af''-2\frac{q_{41}}{q_{31}}+f''\right)\epsilon_1\\ &-(-f''+bf''+2-2b)\epsilon_2\Bigg] = 0.\end{aligned} \tag{11}$$

They can be solved by the following two independent solutions:

$$\begin{aligned}\epsilon_1 &= 0,\\ f'' &= 2;\end{aligned} \tag{12}$$

$$\begin{aligned}\epsilon_2 &= 0,\\ f'' &= 2\frac{q_{41}}{q_{31}(1-a)}.\end{aligned} \tag{13}$$

Using these solutions, we can build two independent matrix-valued functions $g_1(\epsilon)$ and $g_2(\epsilon)$:

$$\begin{aligned}g_1(\epsilon) &= f_1(\epsilon\mathcal{J}_1)^{-1}Q\,f_1(\epsilon\mathcal{J}_1)\\ &= Q + \epsilon(Q\mathcal{J}_1 - \mathcal{J}_1Q) + \epsilon^2(Q\mathcal{J}_1^2 - \mathcal{J}_1Q\mathcal{J}_1) + O(\epsilon^3),\end{aligned} \tag{14}$$

$$\begin{aligned}g_2(\epsilon) &= f_2(\epsilon\mathcal{J}_2)^{-1}\,Q\,f_2(\epsilon\mathcal{J}_2)\\ &= Q + \epsilon(Q\mathcal{J}_2 - \mathcal{J}_2Q) + \epsilon^2\Bigg(\frac{q_{41}}{q_{31}(1-a)}Q\mathcal{J}_2^2 - \mathcal{J}_2Q\mathcal{J}_2\\ &\quad + \left(1-\frac{q_{41}}{q_{31}(1-a)}\right)\mathcal{J}_2^2Q\Bigg) + O(\epsilon^3),\end{aligned} \tag{15}$$

$$\mathcal{J}_1 = \left(\begin{array}{cc|cc} 0 & 0 & 0 & 0\\ b & -b & 0 & 0\\ \hline 0 & 0 & 0 & 0\\ 0 & 0 & 1 & -1\end{array}\right), \tag{16}$$

$$\mathcal{J}_2 = \left(\begin{array}{cc|cc} -a & a & 0 & 0\\ 0 & 0 & 0 & 0\\ \hline 0 & 0 & -1 & 1\\ 0 & 0 & 0 & 0\end{array}\right), \tag{17}$$

where f_1 and f_2 denote some analytical functions, whose first-order Taylor coefficient is unity and whose second-order Taylor coefficient is given by equations (12) and (13), respectively, and is otherwise arbitrary. Both functions $g_1(\epsilon)$ and $g_2(\epsilon)$ fulfil all assumptions of lemma 2. Consequently, the Hessian matrix of the likelihood function is singular at the maximum-likelihood point.

For lemma 2 to be applicable, it is sufficient that only the terms in the Taylor series up to order 2 are compatible with the loop-gating scheme. Therefore, it is not necessary to specify any higher-order Taylor coefficients than the second-order coefficient of the functions f_1 and f_2. Moreover, the functions $g_1(\epsilon)$ or $g_2(\epsilon)$ need not be generator matrices for any finite ϵ values.

In the following section, we will show that is it possible to find special functions f_1 and f_2 so that $g_1(\epsilon)$ and $g_2(\epsilon)$ are generator matrices compatible with the loop-gating scheme for small, but finite values of the parameter ϵ.

(c) *Non-identifiability*

In §2(b) we proved the singularity of the Hessian matrix of the likelihood function at a maximum-likelihood point for the loop-gating scheme with equal open times. We constructed two different families of transformation matrices corresponding to the two solutions of equations

(10) and (11). For the proof of the singularity of the Hessian, it was sufficient that only the sum of the first three terms in the Taylor expansion of equation (3) is compatible with the loop-gating scheme, but not $Q'(\epsilon)$ itself; $Q'(\epsilon)$ is not even required to be a generator matrix at all.

The transition rates of the loop-gating scheme are not identifiable if we can find particular functions f_1 or f_2 with the following properties: the transformed matrices $Q'(\epsilon)$ are generator matrices and compatible with the loop-gating scheme. Therefore, it follows that the aggregated Markov model with generator matrix $Q'(\epsilon)$ accomplishes the assumptions of lemma 1 for every sufficiently small value of ϵ. Hence, it has the same joint probability distribution as the aggregated Markov model with generator matrix Q.

The functions f_1 and f_2 were not fully specified in § 2(b). In the following, we will exploit the freedom to choose the higher-order Taylor coefficients of the functions f_1 or f_2 to satisfy the constraints given by the loop-gating scheme. Equation (3) can be written as

$$Q'(\epsilon) = f_\alpha^{-1}(\epsilon \mathcal{J}_\alpha) Q f_\alpha(\epsilon \mathcal{J}_\alpha) = \sum_{l=0}^{\infty} \epsilon^l \sum_{\substack{i,j\\ i+j=l}} \frac{(f_\alpha^{-1})^{(i)} f_\alpha^{(j)}}{i!j!} \mathcal{J}_\alpha^i Q \mathcal{J}_\alpha{}^j, \tag{18}$$

where $\alpha = 1,2$, $f_\alpha^{(j)}$ denotes the jth Taylor coefficient and $(f_\alpha^{-1})^{(i)}$ denotes the ith Taylor coefficient of f_α^{-1}, both Taylor expansions are performed around the point zero. Note that $(f_\alpha^{-1})^{(l)}$ is fully specified by the Taylor coefficients f_α^j, $j = 0, \ldots, l$ because f_α is subject to the constraint $f_\alpha^{-1}(x) f_\alpha(x) = 1$. Because of relation (8), it need only be shown that the entries (2,3) and (3,2) in the transformed matrices $Q'(\epsilon)$ vanish. This is the case if these two entries vanish in every order $\mathcal{J}_\alpha^i Q \mathcal{J}_\alpha^j$. We will show that one of these constraints is always satisfied by the choice of the matrices $\mathcal{J}_1$ and $\mathcal{J}_2$, respectively, in equations (16) and (17); the other constraint will then lead to recursion equations for the Taylor coefficients of the functions f_1 and f_2, respectively.

Case $\alpha = 1$. We examine the solution (12). After some tedious algebra, we find that

$$(\mathcal{J}_1^i Q \mathcal{J}_1^j)_{\mathrm{co}} = (-1)^{i+j} b^j (-a\, q_{31}) \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}, \quad i,j \geqslant 1. \tag{19}$$

By equation (19), it is shown that the entry (3,2) is vanishing in every order of equation (18) independent of the function f_1.

The constraint that the entry (2,3) is zero in the lth order of equation (18) leads to the following condition for the Taylor coefficients $f_1^{(l)}$:

$$\frac{(f_1^{-1})^{(l)}}{l!} b^l + \frac{f_1^{(l)}}{l!} b + \sum_{\substack{i+j=l\\ i,j \geqslant 1}} \frac{(f_1^{-1})^{(i)} f_1^{(j)}}{i!j!} b^i = 0. \tag{20}$$

Expression (20) is a recursion equation for $f_1^{(l)}$. To solve this equation, we consider the function

$$f_1(x) = (1-x)^{-1} = \sum_{l=0}^{\infty} x^l, \quad \text{for } |x| < 1, \tag{21}$$

with Taylor coefficients $f_1^{(l)} = l!$. The Taylor coefficients of the inverse function $f_1^{-1} = 1 + x$ of order 2 and higher vanish. Therefore, it is easily shown, that $f_1(x) = (1-x)^{-1}$ solves equation (20). The corresponding transformation matrix is given by

$$S_1(\epsilon) = (1 - \epsilon \mathcal{J}_1)^{-1}. \tag{22}$$

Again after some tedious algebra, we can derive the transformed matrix $Q'(\epsilon)$:

$$Q_1'(\epsilon) = S_1^{-1}(\epsilon) Q\, S_1(\epsilon) = \begin{pmatrix} -q_{13} b & 0 & \dfrac{q_{13}(1+b\epsilon)}{1+\epsilon} & \dfrac{q_{13}(b-1)}{1+\epsilon} \\ 0 & -q_{13} b & 0 & q_{13} b \\ q_{31} & 0 & -q_{31} - \dfrac{q_{34}}{1+\epsilon} & \dfrac{q_{34}}{1+\epsilon} \\ A & \dfrac{a q_{31}(1+\epsilon)}{1+b\epsilon} & B & C \end{pmatrix} \tag{23}$$

where

$$A = \frac{1}{1+b\epsilon}(ab q_{31} \epsilon(\epsilon+1)) + \epsilon_{41} - q_{31}) + q_{41},$$

$$B = \frac{-1}{1+\epsilon}(a q_{31} \epsilon(1+\epsilon) + \epsilon(q_{41} - q_{31}) - \epsilon(q_{31} + q_{34} - q_{41} + q_{43}) - q_{43}),$$

$$C = \frac{-1}{1+\epsilon}(a q_{31}(1+\epsilon) + \epsilon(q_{34} + q_{41} + q_{43}) + q_{41} + q_{43}).$$

$Q_1'(\epsilon)$ is a generator matrix and compatible with the loop-gating scheme for all sufficiently small ϵ values and $Q_1'(\epsilon)$ is related to Q by a similarity transformation satisfying the assumptions of lemma 1. An aggregated Markov model with generator matrix $Q'_1(\epsilon)$ is therefore indistinguishable from an aggregated Markov model with generator matrix Q.

Case $\alpha = 2$. We examine the solution (13). The calculations are similar to the previous case.

$$(\mathcal{J}_2^i Q \mathcal{J}_2^j)_{\mathrm{oc}} = (-1)^{i+j-1} a^i q_{13} \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}, \quad i,j \geqslant 1. \tag{24}$$

In contrast to the previous case, the entry (2,3) is vanishing in every order of equation (18) independent of the function f_2.

We now obtain the following recursion equation for the Taylor coefficients $f_2^{(l)}$ due to the constraint that the entry (3,2) is zero in the lth order of equation (18):

$$\frac{(f_2^{-1})^l}{l!} + \frac{f_2^{(l)}}{l!} a^{l-1} - \frac{q_{41} - q_{31}}{q_{31}} \sum_{\substack{i+j=1\\ i,j \geqslant 1}} a^{j-1} \frac{(f_2^{-1})^{(i)} f_2^{(j)}}{i!j!} = 0. \tag{25}$$

We are not aware of a solution of equation (25) which can be expressed by elementary functions. The function f_2 is, however, determined by equation (25), and there exists a second family of transformation matrices $Q'_2(\epsilon) = f_2^{-1}(\epsilon \mathcal{J}_2) Q f_2(\epsilon \mathcal{J}_2)$ which is compatible with the loop-gating scheme and related to the generator matrix Q by a similarity transformation in the sense of lemma 1. Again, an aggregated Markov model with generator matrix $Q'_2(\epsilon)$ is indistinguishable from an aggregated Markov model with generator matrix Q for every sufficiently small value of ϵ.

We have shown that for every aggregated Markov model with equal open times and which is compatible to the loop-gating scheme, we can find two families of aggregated Markov models which depend continuously on a parameter ϵ and which are equivalent to the given aggregated Markov model. Therefore, the transition rates of such an aggregated Markov model can never be estimated from the measured data because there is a choice between an infinite number of models which could have produced the data equally well. The parameters of an aggregated Markov model with equal open times are only identifiable if the dimension of the parameter space is further reduced, for example some transition rates are kept fixed or some transition rates are functions of the other transition rates as in the case of imposing the principle of detailed balance on the transition rates. There are at least two constraints needed in addition to the condition of equal open times in order to obtain a parameterization whose parameters are all identifiable.

3. NEARLY EQUAL DWELL TIMES

In the following section, we consider the estimation of the transition rates in the loop model. Therefore, we drop the restriction of equal open times and we assume that all transition rates of the loop-gating scheme are parameters of an aggregated Markov model and that no further constraints are imposed on the parameters.

Given data which are generated by an aggregated Markov model with the loop-gating scheme and not with equal, but almost equal, dwell times, all transition rates are identifiable and the maximum-likelihood estimators converge to the true transition rates for the number of data points going to infinity. Furthermore, the maximum-likelihood estimators are asymptotically normally distributed. In practice, only a limited number of data are available. Therefore, we have to study the finite sample properties of the maximum-likelihood estimators. There are two typical finite sample problems. First, there are not enough data available to apply the asymptotic theory of the maximum-likelihood estimator; second, the asymptotic theory is applicable, but the estimated standard deviations of the estimated parameters are extremely large. In this section, we demonstrate by simulation studies that we encounter these problems for models with nearly equal open times because the finite sample properties of the maximum-likelihood estimators depend on the ratio of the open times as a consequence of the non-identifiability of the transition rates in the case of equal open times.

The Hessian matrix of the likelihood function at the maximum-likelihood point plays a crucial role in the asymptotic theory of the maximum-likelihood estimators, because the covariance matrix can be estimated by the inverse of the Hessian matrix (Bickel *et al.* 1998). The estimated standard deviations of the parameters will thus become large if the Hessian matrix is ill-conditioned.

The loop models with equal open times form a subspace in the parameter space of the loop models. Because we have shown in § 2(a) that the transition rates in a loop model are not identifiable if the model has equal open times, the likelihood function is constant on a set of two-dimensional (2D) manifolds in the parameter space which corresponds to loop models with equal open times. Thus, the Hessian matrix of the likelihood function is singular on this set of 2D manifolds. Moreover, the Hessian matrix is ill-conditioned in the neighbourhood of these manifolds because the condition number of a matrix depends continuously on its entries.

We therefore investigate the dependency of the standard deviation of the parameter estimator on the ratio of the open times in a simulation study. The simulated data were generated by the loop model with the following generator matrix:

$$\begin{pmatrix} -100 & 0 & 25 & 75 \\ 0 & -1/\tau_2 & 0 & 1/\tau_2 \\ 24 & 0 & -44 & 20 \\ 147.6 & 25 & 41 & -213.6 \end{pmatrix}. \tag{26}$$

All transition rates are given in Hertz. The closed dwell times are 4.5 ms and 25 ms, the open dwell times are 10 ms and τ_2, where τ_2 is varying from 80 ms to 20 ms in steps of 5 ms. For each τ_2 we simulate 32 recordings of length 420 s with a sampling rate of 5 kHz (2^{21} data points), estimate the transition rates by the maximum-likelihood method and estimate the covariance matrix by the inverse of the Hessian matrix at the maximum-likelihood point. The maximization of the likelihood function is performed numerically by the EM algorithm (Michalek & Timmer 1999) and a nonlinear maximization routine based on a quasi-Newton method (NAG 1997). For the calculation of the first derivatives of likelihood function, we use Fisher's identity (Fisher 1925; Jamshidian & Jenrich 1997) and the 'sinch'-algorithm described by Najfeld & Havel (1995) to evaluate the derivatives of the matrix exponential. The Hessian matrix is calculated numerically by using the first derivatives of the likelihood function (NAG 1997).

Figures 3 and 4 subsume the results of this simulation study. Figure 3 shows the median of the condition numbers of the Hessian matrix. The median absolute deviations of the condition number about the median condition numbers are plotted as error bars. The condition numbers are calculated in the 1-norm (Golub & VanLoan 1996, pp. 54–57). As an example, the estimated relative error of the transition rate $C_1 \rightarrow C_2$ is shown in figure 4, we denote by 'relative error' the estimated standard deviation divided by the true value of the transition rate. The error bars indicate the estimated standard deviation of the error estimate. Both figures illustrate the rise of the condition number and the estimated relative error, respectively, towards smaller values of the open time ratio as expected by the theoretical considerations. Below an open time ratio of 2, the numerical optimization of the likelihood function is not stable because of the flatness of the likelihood function near the maximum-likelihood point.

The magnitude of the estimated standard deviations depends on the length of the time-series. Thus, we examine the dependence of the estimated relative error on the length of the data set in a second simulation study.

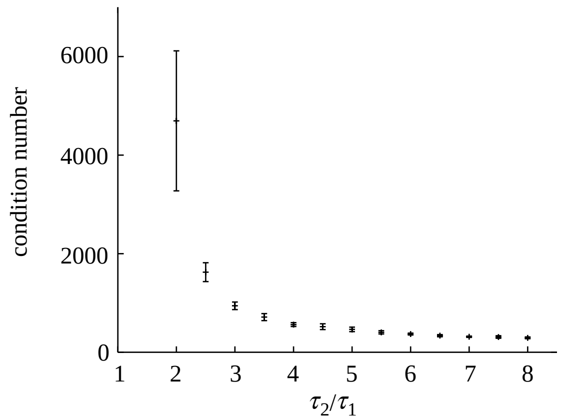

Figure 3. Condition number of the Hessian matrix. The error bars indicate the median absolute deviation of the condition number about the median condition number. The condition numbers are calculated using the 1-norm. For each τ_2 value we simulated 32 recordings of length 420 s with a sampling rate of 5 kHz (2^{21} data points).

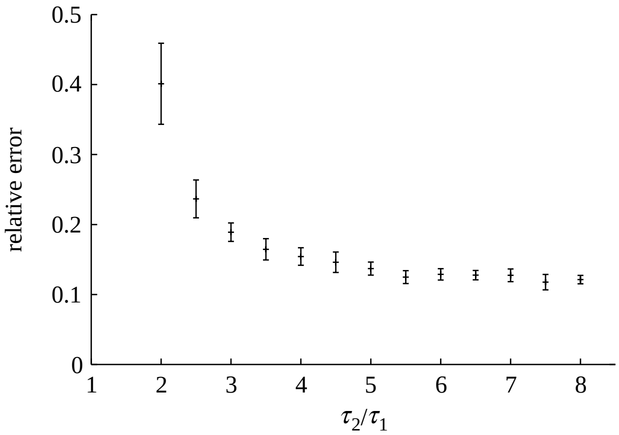

Figure 4. Estimated relative errors of the transition rate $C_1 \rightarrow C_2$. The error bars indicate the estimated standard deviation of the error estimate. For each τ_2 we simulated 32 recordings of length 420 s with a sampling rate of 5 kHz (2^{21} data points).

Using the same aggregated Markov model as in the first simulations, the length of simulated data sets is varied from 2^{16} to 2^{21}. The open time ratio is set to 3. We simulate 32 recordings for each length. Figure 5 shows the estimated relative error of the transition rate $C_1 \rightarrow C_2$ and figure 6 shows the median of the condition numbers.

For less than 2^{19} data points, corresponding to 105 s, the transition rate $C_1 \rightarrow C_2$ cannot be estimated reliably, and the asymptotic result that the maximum-likelihood estimators follow a Gaussian distribution, is not applicable.

The Hessian matrix scales asymptotically with unity divided by the number of data points (Bickel *et al.* 1998). Therefore, the condition number of the Hessian matrix is asymptotically independent of the length of the time-series. Figure 6 suggests that the Hessian matrix already obeys this scaling law, although the maximum-likelihood estimators still deviate from the Gaussian distribution if the number of data points is less than 2^{19}.

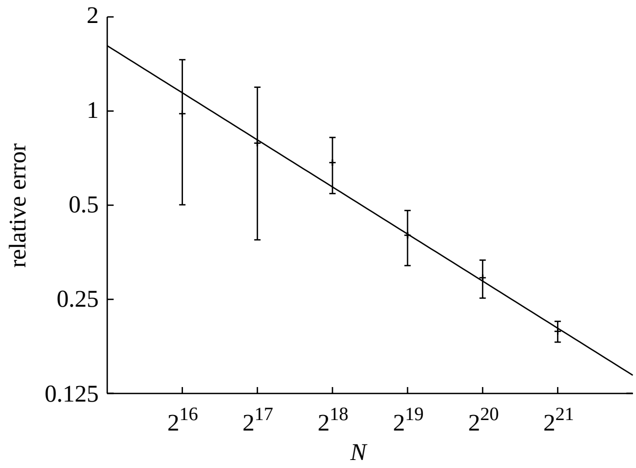

Figure 5. Estimated relative errors of the transition rate $C_1 \rightarrow C_2$ for different numbers of data points N. The error bars indicate the estimated standard deviation of the error estimate. The solid line shows the expected asymptotic scaling behaviour $1/\sqrt{N}$ of the error estimates. The ratio of the open times is $\tau_2/\tau_1 = 3$. Even though this transition rate is identifiable for $\tau_1/\tau_2 \neq 1$, the non-identifiability for $\tau_2 = \tau_1$ has an effect on estimated error for $\tau_1/\tau_2 \neq 1$.

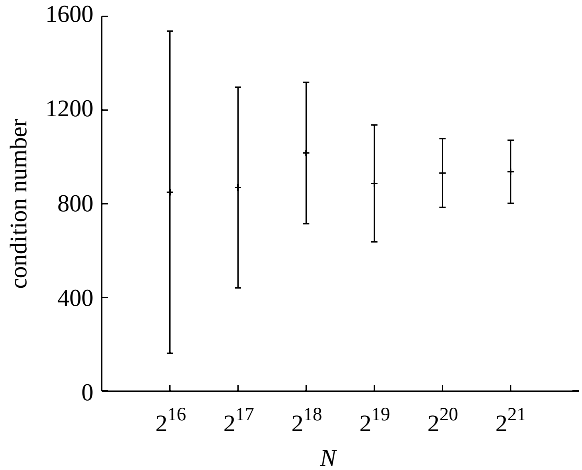

Figure 6. Condition number of the Hessian matrix for different numbers of data points N. The error bars indicate the median absolute deviation of the condition number about the median condition number. The condition numbers are calculated using the 1-norm. The ratio of the open times is $\tau_2/\tau_1 = 3$. The simulation result is consistent with the asymptotic independence of the condition number on the number of data points N.

4. DISCUSSION

We have shown the non-identifiability of the transition rates in models with the loop-gating scheme and with equal dwell times. This non-identifiability has an effect on the estimation of the transition rates in models with the loop-gating scheme if the dwell times are nearly equal. Moreover, the simulation studies indicate that an unexpectedly large number of data points are sometimes necessary to estimate the transition rates reliably.

The loop-gating scheme has served as the simplest example for a model with loops, but the results based on the loop-gating scheme can surely be carried over to more complicated aggregated Markov models with loops. Moreover, this simple example demonstrates the importance of estimating the errors of transition rates when analysing real data.

These consequences of the non-identifiability in models with equal dwell times can be avoided only by putting restrictions on the transition rates to be estimated and, so, reducing the number of parameters in the model. Because equations (10) and (11) have two solutions, at least two constraints are needed to allow the identifiability of the remaining parameters. A common constraint for models with loops is the principle of detailed balance (Song & Magleby 1994; Kienker 1989). In the case of the loop model, however, detailed balance reduces the number of parameters only by one, and a further constraint would be necessary. Therefore, even a model which is subject to the principle of detailed balance is not identifiable in the case of equal open times.

The impact of non-identifiability in models with equal dwell times on tests for detecting violations of detailed balance will be further investigated.

REFERENCES

Albertsen, A. & Hansen, U.-P. 1994 Estimation of kinetic rate constants from multi-channel recordings by a direct fit of the time series. *Biophys. J.* **67**, 1393–1403.

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

Ball, R. G., Yeo, G. F., Milne, R. K., Edeson, R. E., Madsen, B. W. & Simson, M. P. 1993 Single ion channel models incorporation aggregation and time interval omission. *Biophys. J.* **64**, 357–374.

Bates S. E., Sansom, M. S. P., Ball, F. G., Ramsey, R. L. & Usherwood, P. N. R. 1990 Glutamate receptio-channel gating—maximum likelihood analysis of gigaohm seal recordings from locust muscle. *Biophys. J.* **58**, 219–229.

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

Chung, S.-H., Moore, J., Xia, L., Premkumar, L. S. & 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. & Moore, J. B. 1991 Adaptive processing techniques based on hidden Markov models for characterizing very small channel currents buried in noise and deterministic interferences. *Phil. Trans. R. Soc. Lond.* B **334**, 357–384.

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

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

Colquhoun, D. & Sigworth, F. J. 1995 Fitting and statistical analysis of single-channel records. In *Single-channel recording*, 2nd edn (ed. B. Sakmann & E. Neher), pp. 483–587. New York: Plenum Press.

Colquhoun, D., Hawkes, A. G. & Srodzinski, K. 1996 Joint distributions of apparent open-and-shut times of single-ion channels and maximum-likelihood fitting of mechanisms. *Phil. Trans. R. Soc. Lond.* B **354**, 2555–2590.

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

Fisher, R. 1925 Theory of statistical estimation. *Proc. Camb. Phil. Soc.* **22**, 700–725.

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

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

Fredkin, D. & Rice, J. 1997 Fast evaluation of the likelihood of an HMM: ion channel currents with filtering and colored noise. Technical Report, Department of Statistics, University of California at Berkeley, No. 485, available at (http://www.stat.berkeley.edu/tech-reports/).

Fredkin, D. R., Montal, M. & Rice, J. 1983 Identification of aggregated Markovian models: application to the nicotinic acetylcholine receptor. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, vol. I (ed. L. M. LeCam & R. A. Olshen), pp. 269–289. Belmont, CA: Wadsworth Press.

Golub, G. H. & VanLoan, C. F. 1996 *Matrix computations*, 3rd edn. Johns Hopkins Studies in the mathematical sciences. Baltimore, MD: Johns Hopkins University Press.

Hamill, O. P., Marty A., Neher, E., Sakmann, B. & Sigworth, F. J. 1981 Improved patch-clamp technique for high-resolution current recoding from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–100.

Hille, B. 1993 *Ionic channels of excitable membranes*, 2nd edn. Sunderland, MA: Sinauer Associates.

Horn, R. & Lange, K. 1983 Estimating kinetic constants from single channel data. *Biophys. J.* **43**, 207–223.

Jamshidian, M. & Jennrich, R. I. 1997 Acceleration of the EM algorithm by using quasi-newton methods. *J. R. Statist. Soc.* B **59**, 569–587.

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

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

Michalek, S., Wagner, M. & Timmer, J. 1998 A new approximate likelihood estimator for ARMA-filtered hidden Markov models. *IEEE Trans. Sig. Proc.* (Submitted.)

Michalek, S., Lerche, H., Wagner, M., Mitrović, N., Schiebe, M., Lehmann-Horn, F. & Timmer, J. 1999 On identification of Na^+ channel gating schemes using moving-average filtered hidden Markov models. *Eur. Biophys. J.* (In the press.)

NAG 1997 *Fortran Library Mark 18*. Oxford, UK: The Numerical Algorithms Group Ltd.

Najfeld, I. & Havel, T. F. 1995 Derivatives of the matrix exponential and their computation. *Adv. Appl. Math.* **16**, 321–375.

Neher, E. & Sakmann, B. 1976 Single-channels currents recorded from membrane of denervated frog muscle fibres. *Nature* **260**, 799–802.

Song, L. & Magleby, K. L. 1994 Testing for microscopic reversibility in the gating of maxi K^+ using two-dimensional dwell-time distributions. *Biophys. J.* **67**, 91–104.

Vandenberg, C. A. & Bezanilla, F. 1991 A sodium channel gating model based on single channel macroscopic ionic, and gating currents in the squid giant axon. *Biophys. J.* **60**, 1511–1533.

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

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.