J. theor. Biol. (2001) **208,** 000–000
doi:10.1006/jtbi.2000.2230, available online at http://www.idealibrary.com on IDEAL®

Model Selection in Non-nested Hidden Markov Models for Ion Channel Gating

MIRKO WAGNER AND JENS TIMMER

Freiburger Zentrum für Datenanalyse und Modellbildung, Eckerstr. 1, *D*-79104 *Freiburg, Germany*

(*Received on* 9 *June* 2000, *Accepted in revised form on* 2 *November* 2000)

An important task in the application of Markov models to the analysis of ion channel data is the determination of the correct gating scheme of the ion channel under investigation. Some prior knowledge from other experiments often allows to reduce the number of possible models significantly. If these models are nested, standard statistical procedures, like likelihood ratio testing, provide reliable selection methods. In the case of non-nested models information criteria like AIC, BIC, etc., are used. However, it is not known if any of these criteria provide a reliable selection method and which is the best one in the context of ion channel gating. We provide an alternative approach to model selection in the case of non-nested models with an equal number of open and closed states. The models to choose from are embedded in a properly defined general model. Therefore, we circumvent the problems of model selection in the non-nested case and can apply model selection procedures for nested models.

1. Introduction

The physiological research on ion channels focuses on uncovering the correlation between the structure of a channel protein and its physiological function. Biochemical studies and cloning experiments provide detailed static information about the structure of the channel protein and its sub-units (Hille, 1992; Aidley & Stanfield, 1996). The dynamics of an ion channel are determined by conformational changes of the sub-units of the channel protein or binding of ligand molecules. Therefore, the identification of the sub-units allows assumptions to be made about possible mechanisms which govern the dynamics of the opening and closing of an ion channel and which are compatible with the identified sub-units. Gating schemes summarize these assumptions about the dynamical behavior by specifying the number of open and closed states and the allowed transitions between the states.

Besides the electrophysiological characterization, patch clamp recordings provide the additional complementary information about dynamical features of ion channels needed to discriminate between different assumptions about the true gating scheme of an ion channel (Sakmann & Neher, 1995). Thus, the analysis of recorded patch clamp data requires statistical procedures to infer the gating scheme from the data.

Markov chains in continuous time have proven to be a suitable model class to describe the transitions among the unobserved states in an ion channel (Colquhoun & Hawkes, 1977, 1982). As the ion channel current is observed and not the states themselves, only an aggregated image of the underlying process is available and so the measured currents are modelled either by an aggregated Markov chain (Ball & Rice, 1992) or by its generalization, a hidden Markov model (Chung *et al.*, 1990; Fredkin & Rice, 1992; Chung

0022–5193/01/000000 + 12 $35.00/0

& Kennedy, 1996; Michalek *et al*., 1999, 2000) depending on the signal-to-noise ratio. In the following, we do not distinguish between aggregated Markov chains and hidden Markov models and generally use the term hidden Markov model for both cases.

For nested gating schemes, i.e. one gating scheme is a sub-model of the other gating scheme, possibly after applying a similarity transformation (Kienker, 1989), likelihood ratio testing provides an appropriate method for model selection. The selection criterion of likelihood ratio testing is a principle of parsimony because likelihood ratio testing favors the simpler gating scheme due to the null hypothesis that the gating scheme with the smaller number of parameters is the true one. In addition, likelihood ratio testing has favorable asymptotic properties under the null hypothesis. Its test statistic asymptotically follows a χ^2 distribution with a number of degrees of freedom given by the difference of the number of parameters in both models, and, in particular, the distribution is independent of the unknown true parameters. The main assumptions for this result are: the gating schemes are not misspecified, all parameters are identifiable both under the null hypothesis and under the alternative and the true parameters are in the interior of the parameter space (Cox & Hinkley, 1974). These assumptions, however, may sometimes not be fullfilled and the problem of likelihood ratio tests under non-standard conditions may arise, for example, if some transition rates are on the boundary of the parameter space or not identifiable under the null hypothesis (Self & Liang, 1987; Wagner *et al*., 1999).

Likelihood ratio testing is not directly applicable for the model selection of non-nested gating schemes. In order to use this method from the nested case, the non-nested gating schemes have to be embedded in a general model. Arbitrary complex gating schemes, however, cannot serve this purpose because of the following difficulties: firstly, the number of identifiable parameters in hidden Markov models is limited to 2 times the number of open states times the number of closed states (Fredkin *et al*., 1983; Fredkin & Rice, 1986); secondly, gating schemes are typically embedded in other gating schemes by constraining certain transition rates to zero, so that these transition rates are part of the boundary of the parameter space. For selecting between different gating schemes it is not necessary that a general model be interpreted as a gating scheme, it is sufficient that this model provides a parameterization of the likelihood functions of all proposed gating schemes and that these gating schemes are not on the boundary of the parameter space.

In the following section, we develop a partial solution by deriving a parameterization of the likelihood functions for hidden Markov models with the only restriction that these models must have the same number of open and closed states. Section 5 presents some simulation studies where we apply the results of Section 2. In Section 6, we compare the proposed selection method to common procedures based on information criteria.

2. Notation and Model Equivalence

In this section, we develop the notation used throughout the subsequent sections and state a result on the equivalence of two hidden Markov models needed in Section 4 for the inference on hidden Markov models.

We assume, that we observe a continuous process at discretely sampled time points with sampling time Δt. We denote with $y_1, \ldots, y_T$ the observation sequence of length T, the unobserved sequence of the background states with $x_1, \ldots, x_T$. The state space of the unobserved Markov chain has dimension n, and is partitioned into n_o open states and n_c closed states with $n = n_o + n_c$. For convenience, the states are ordered: the states $1, \ldots, n_o$ are the open states, followed by the closed states: $n_o + 1, \ldots, n$. $\mathcal{O}$ denotes the set of indices of the open states $\{1, \ldots, n_o\}$, $\mathscr{C}$ denotes the set of indices of the closed states $\{n_o + 1, \ldots, n\}$. The generator matrix Q determines the dynamics of the Markov chain. It is partitioned according to the open and closed states:

$$Q = \left(\begin{array}{c|c} Q_{oo} & Q_{oc} \\ \hline Q_{co} & Q_{cc} \end{array} \right). \tag{1}$$

The transition probability matrix A for the discretely sampled process is obtained by the

matrix exponential of $Q\Delta t$:

$$A = \exp(Q\Delta t). \tag{2}$$

The observation sequence $y_1, \ldots, y_T$ is conditionally independent given the unobserved sequence of background states $x_1, \ldots, x_T$:

$$\begin{aligned} &p(y_1, \ldots, y_T | x_1, \ldots, x_T) \\ &\quad = p(y_1|x_1) \ldots p(y_T|x_T). \end{aligned} \tag{3}$$

Furthermore, the conditional density of observing y given background state x simplifies in the case of ion channel gating to

$$p(y|x) = g(y|x) = \begin{cases} g_o(y) & \text{if } x \in \mathcal{O} \\ g_c(y) & \text{if } x \in \mathcal{C}. \end{cases} \tag{4}$$

For aggregated Markov chains, $g(y|x)$ is the indicator function of the open states, for hidden Markov models $g_o(y)$ and $g_c(y)$ may be Gaussian densities taking into account a different noise in the open states and the closed states, respectively, besides the different conductance levels.

The probability density for an observation sequence $y_1, \ldots, y_T$ follows from the Markov property and the conditional independence of the observation sequence (Baum & Petrie, 1966; Baum *et al.*, 1970):

$$\begin{aligned} p(y_1, \ldots, y_T) = &\sum_{x_1, \ldots, x_T = 1}^{n} \pi_{x_1} g(y_1|x_1) A_{x_1, x_2} \\ &\times g(y_2|x_2) \ldots A_{x_{T-1}, x_T} g(y_T|x_T), \end{aligned} \tag{5}$$

where π denotes the initial distribution of the Markov chain.

We now transform eqn (5) taking the partitioning into open and closed states into account. Each of the index variables $a_i \in \{\mathcal{O}, \mathcal{C}\}$, $i = 1, \ldots, T$, denotes either the whole index set of the open or the closed states. Then, A_{a_i, a_j} is the submatrix of A formed from the index sets a_i and a_j and $A_{a_i, a_l} A_{a_l, a_j}$ denotes the matrix multiplication between the sub-matrices A_{a_i, a_l} and A_{a_l, a_j}. The probability density for an observation sequence $y_1, \ldots, y_T$ is now given by

$$\begin{aligned} p(y_1, \ldots, y_T) = &\sum_{a_1, \ldots, a_T \in \{\mathcal{O}, \mathcal{C}\}} (\pi_{a_1} A_{a_1, a_2} \ldots A_{a_{T-1}, a_T} 1_{a_T}) \\ &\times (g_{a_1}(y_1) \ldots g_{a_T}(y_T)), \end{aligned} \tag{6}$$

where π_{a_1} is a row vector with the entries of π from the index set a_1, 1_{a_T} is a column vector of ones with the dimension of the index set a_T. Note that in the case of aggregated Markov chains, there is only one non-zero summand in eqn (6). A generalization of eqn (6) for an arbitrary number of aggregation classes is the starting point for a general investigation of the identifiability problem in hidden Markov models (Ito *et al.*, 1992).

In Kienker (1989) it is shown that the outcomes of two aggregated Markov models are statistically not distinguishable if their generator matrices are related by a similarity transformation with a transformation matrix of the following form:

$$S = \left(\begin{array}{c|c} S_{oo} & 0 \\ \hline 0 & S_{cc} \end{array} \right), \tag{7}$$

where S is invertible and the rows S are normalized to one. This result may also be derived from eqn (6) and thereby extended to the case of hidden Markov models. The probability density is invariant under transformations of the transition probability matrix A and the initial distribution π which leave the coefficients $\pi_{a_1} A_{a_1, a_2} \ldots A_{a_{T-1}, a_T} 1_{a_T}$ in eqn (6) numerically invariant. Assuming that two hidden Markov models with an equal output conditional density $g(y|x)$ but different generator matrices Q, Q' and initial distributions π, π' are related by a transformation matrix S: $Q' = S^{-1} Q S$ and $\pi' = \pi S$, then the probability densities of both models are the same because the transition probability matrices are related by the same similarity transformation $A' = S^{-1} A S$ due to eqn (2) and the coefficients in eqn (6) agree:

$$\begin{aligned} &\pi'_{a_1} A'_{a_1, a_2} \ldots A'_{a_{T-1}, a_T} 1_{a_T} \\ &\quad = (\pi_{a_1} S_{a_1, a_1}) \underbrace{(S^{-1}_{a_1, a_1} A_{a_1, a_2} S_{a_2, a_2})}_{=1} \\ &\quad \ldots (S^{-1}_{a_{T-1}, a_{T-1}} A_{a_{T-1}, a_T} \underbrace{S_{a_T, a_T}}_{=1}) 1_{a_T} \\ &\quad = \pi_{a_1} A_{a_1, a_2} \ldots A_{a_{T-1}, a_T} 1_{a_T}. \end{aligned} \tag{8}$$

3. Likelihood Parameterizations

Any model-selection procedure requires the estimation of the model parameters as a preliminary step. A gating scheme determines a model in the case of ion channel gating and so its transition rates have to be estimated from the measured data. This can be achieved by the maximum likelihood method (Horn & Lange, 1983; Fredkin & Rice, 1992; Albertsen & Hansen, 1994; Qin *et al*., 1997; Michalek & Timmer, 1999). In addition, the estimation of the transition rates is often burdened with difficulties like time-interval omission (Roux & Sauve, 1985; Ball *et al*., 1993; Qin *et al*., 1996; Colquhoun *et al*., 1996) or colored noise on the data (Venkataramanan *et al*., 1998a,b; Michalek *et al*., 2000). In the case of aggregated Markov models, time-interval omission requires methods for missed event correction, which has been solved exactly by Hawkes *et al*. (1990, 1992).

In the following, we emphasize the fact that the likelihood function L depends only through the transitions probability matrix A on the transition rates by writing: $L(A(Q)) = p(y_1, \ldots, y_T)$. Under mild regularity conditions the maximum likelihood estimator for hidden Markov models is asymptotically normally distributed (Bickel *et al*., 1998). 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.

In the following paragraphs, we derive a parameterization of the likelihood functions of all hidden Markov models which have the same number of open and closed states. The derivation is divided into two steps. Firstly, we investigate the special case of models which follow the law of detailed balance, and then we drop this restriction and generalize the parameterization to arbitrary hidden Markov models.

3.1. MODELS WHICH OBEY DETAILED BALANCE

The gating of an ion channel is subject to the principle of detailed balance in the absence of an external energy source (Song & Magleby, 1994). Under these conditions the sub-matrices Q_{oo} and Q_{cc} of the generator matrix can be diagonalized with real eigenvalues (Fredkin *et al*., 1983; Kijima & Kijima, 1987). The same line of reasoning can be applied to the sub-matrices A_{oo} and A_{cc} of the transition probability matrix $A = \exp(Q\Delta t)$. Therefore, we can choose the sub-matrices S_{oo} and S_{cc} of a similarity transformation S to be the transformation matrices diagonalizing A_{oo} and A_{cc}. By applying this transformation to the transition probability matrix A, we define the following matrix:

$$A^{(S)}(Q) = \left(\begin{array}{ccc|ccc} a_1^{(o)} & \ldots & 0 & & & \\ \vdots & \ddots & \vdots & & S_{oo}^{-1} A_{oc} S_{cc} & \\ 0 & \ldots & a_{n_o}^{(o)} & & & \\ \hline & & & a_1^{(c)} & \ldots & 0 \\ S_{cc}^{-1} A_{co} S_{oo} & & & \vdots & \ddots & \vdots \\ & & & 0 & \ldots & a_{n_c}^{(c)} \end{array}\right), \quad (9)$$

the scalars $a_1^{(o)}, \ldots, a_{n_o}^{(o)}$ and $a_1^{(c)}, \ldots, a_{n_c}^{(c)}$ are the eigenvalues of the sub-matrices A_{oo} and A_{cc}, respectively. In general, the matrix $A^{(S)}(Q)$ is not a transition probability matrix because its entries may be negative, but the rows are still normalized to one. With regard to model selection, this matrix possesses the following important property because of eqns (8) and (6):

$$L(A^{(S)}(Q)) = L(A(Q)). \quad (10)$$

Due to the row normalization $A^{(S)}$ has $2n_o n_c$ parameters, namely the entries in $A_{oc}^{(S)}$ and $A_{co}^{(S)}$. The likelihood function of an arbitrary hidden Markov model with n_o open states and n_c closed states following the principle of detailed balance, depends on the transition rates only through these $2n_o n_c$ independent parameters of $A^{(S)}$. Therefore, all likelihood functions of hidden Markov models satisfying the law of detailed balance are embedded in the function space which is parameterized by $A^{(S)}$. We denote this space by $\mathscr{F}$. In addition, we again find the result that the maximum number of identifiable parameters in hidden Markov models is bounded by $2n_o n_c$ (Fredkin *et al*., 1983; Kienker, 1989). Since the entries in the sub-matrices $A_{oc}^{(S)}$ and $A_{co}^{(S)}$ are not required to be positive, a gating scheme cannot be part of the boundary of this parameter space and so the problem of tests under non-standard conditions can be avoided (Self & Liang, 1987; Wagner *et al*., 1999).

3.2. THE GENERAL CASE

If the gating of an ion channel does not follow the principle of detailed balance, the generator matrix of the underlying Markov process need not have real eigenvalues. Then the matrix $A^{(S)}$ does not provide a parameterization of all possible likelihood functions of hidden Markov models with n_o open states and n_c closed states including the models which may violate the law of detailed balance. In the general case, there is no single parameterization of all possible likelihood functions as in the case of detailed balance.

In the following, we will show that a small number of non-overlapping parameterizations is sufficient to cover the whole functions space of likelihood functions. This result is based on a lemma of unitary spectral theory: any $n \times n$-matrix M with real-valued components which is diagonalizable with some complex eigenvalues is equivalent to a diagonal matrix of the following form (Lang, 1987):

$$S^{-1}MS = \begin{pmatrix} \lambda & & & & & \\ & \ddots & & & 0 & \\ & & \lambda_r & & & \\ & & & Z_{r+1} & & \\ & 0 & & & \ddots & \\ & & & & & Z_{r+k} \end{pmatrix}, \quad (11)$$

$\lambda_j, 1 \leqslant j \leqslant r$, are the real eigenvalues of A, Z_j, $r+1 \leqslant j \leqslant r+k$, are 2×2 matrices which contain the real and imaginary parts of the complex eigenvalues: $\lambda_j = \alpha_j + \mathrm{i}\beta_j$ of A: $Z_j = \begin{pmatrix} \alpha_j & \beta_j \\ -\beta & \alpha_j \end{pmatrix}$. As the complex conjugate of a non-real eigenvalue is also an eigenvalue, k is one-half of the total number of non-real eigenvalues. S is the real-valued transformation matrix for this operation.

If the gating of an ion channel is out of thermodynamic equilibrium and the sub-matrices A_{oo} and A_{cc} of the transition probability matrix A have some complex eigenvalues, we can choose S_{oo} and S_{cc}, respectively, in a similarity transformation S to be the transformation matrices which transform A_{oo} and A_{cc} to the form of eqn (11):

$$A^{(S)}_{r_o,r_c} = \left(\begin{array}{c|c} A^{(S)}_{r_o,r_c})_{oo} & S_{oo}^{-1} A_{oc} S_{cc} \\ \hline S_{cc}^{-1} A_{co} S_{oo} & (A^{(S)}_{r_o,r_c})_{cc} \end{array} \right) . \quad (12)$$

The subscripts r_o and r_c denote the number of real eigenvalues of A_{oo} and A_{cc} and we identify $A^{(S)}_{n_o,n_c}$ with $A^{(S)}$ from eqn (9). Analogous to the case of detailed balance, the likelihood function can be expressed in terms of $A^{(S)}_{r_o,r_c}$:

$$L(A^{(S)}_{r_o,r_c}(Q)) = L(A(Q)) . \quad (13)$$

$A^{(S)}_{r_o,r_c}$ has $2n_on_c$ independent parameters, namely the sub-matrices $(A^{(S)}_{r_o,r_c})_{oc}$ and $(A^{(S)}_{r_o,r_c})_{co}$, because the rows of $A^{(S)}_{r_o,r_c}$ are standardized to one and the 2×2 Z matrices are antisymmetric.

For a given number r_o of real eigenvalues of A_{oo} and a given number r_c of real eigenvalues of A_{cc}, $A^{(S)}_{r_o,r_c}$ only parameterizes a sub-set of the possible likelihood functions of arbitrary hidden Markov models with n_o open and n_c closed states. So, we enlarge the function space $\mathscr{F}$ introduced in the previous section by the union of the non-overlapping parameterizations $A^{(S)}_{r_o,r_c}$. For example, the set of parameterizations: $(A^{(S)}, A^{(S)}_{3,1}, \quad A^{(S)}_{1,3}, A^{(S)}_{1,1})$ covers the function space $\mathscr{F}$ for $n_o = n_c = 3$. In particular, note that only for more than two open states or more than two closed states complex eigenvalues can occur.

4. Model Selection

In this section, we use the parameterizations of the likelihood function derived in Section 3 to develop a model selection procedure based on the likelihood ratio statistic LR. We assume that a small number r of non-nested hidden Markov models $\mathrm{HMM}_1, \ldots, \mathrm{HMM}_r$ with the same number of open states n_o and closed states n_c, respectively, are promising candidates for being the true model for an observation sequence $y_1, \ldots, y_T$. Each hidden Markov model HMM_j represents a set of likelihood functions and these functions are all contained in the function space $\mathscr{F}$. Furthermore, we assume that the observation sequence $y_1, \ldots, y_T$ is generated by an element of $\mathscr{F}$ denoted by $A^{(\mathrm{true})}$: $A^{(\mathrm{true})} \in \mathscr{F}$.

The selection is divided into two steps: first, we test the null hypothesis that the data are consistent with one of the models $\mathrm{HMM}_1, \ldots, \mathrm{HMM}_r$ against the alternative that another model not contained in the given selection with the n_o open states and n_c closed states, respectively, is the true

model:

$$\mathrm{H}_0: \quad A^{(\mathrm{true})} \in \bigcup_{i=1}^{r} \mathrm{HMM}_i,$$

$$\mathrm{H}_1: \quad A^{(\mathrm{true})} \in \mathscr{F} \Big\backslash \bigcup_{i=1}^{r} \mathrm{HMM}_i. \qquad (14)$$

If we cannot reject the null hypothesis, we proceed in the second step with the selection among the models under consideration using the likelihood ratio statistic LR.

The transition rates of each model to choose from are parameterized by a parameter vector θ_i of dimension k_i for $1 \leqslant i \leqslant r$. In the simplest case, the elements of each parameter vector θ_i are the transition rates themselves. The number of parameters must be less than or equal to $2n_on_c$. Each generator matrix Q_i is a function of the parameter vector θ_i: $Q_i = Q_i(\theta_i)$.

As a prerequisite for the selection procedure, the following result on the asymptotic distribution of the likelihood ratio under the null hypothesis is needed: we assume that the observation sequence $y_1, \ldots, y_T$ is generated by one of the given hidden Markov models, e.g. HMM_j. The parameter vector θ_j as well as the independent entries in the matrix $A^{(S)}$ can be estimated by the maximum likelihood method. $\hat{\theta}_j$ denotes the parameter vector maximizing $\ln L(A(Q_j(\hat{\theta}_j)))$ and $\hat{A}^{(S)}$ denotes the matrix maximizing $\ln L(A^{(S)})$. In the general case, $\hat{A}^{(S)}$ has to be searched in all possible parameterizations $A^{(S)}_{r_o,r_c}$. Then the two-fold log likelihood ratio LR is asymptotically χ^2-distributed with a number of degrees of freedom df_j given by the difference of $2n_on_c$ and the number of parameters k_j [Cox & Hinkley, 1974):

$$\mathrm{LR}_j = 2[\ln L(\hat{A}^{(S)}) - \ln L(A(Q_j(\hat{\theta}_j)))] \overset{T\to\infty}{\sim} \chi^2_{\mathrm{df}_j},$$

$$\mathrm{df}_j = 2n_on_c - k_j. \qquad (15)$$

If the hidden Markov model HMM_j does not contain the true model for the data, the distribution of likelihood ratio LR_j moves towards infinity for the number of data points T going to infinity and is asymptotically normally distributed (Vuong, 1989).

4.1. STEP ONE: TEST FOR CONSISTENCY

Under the null hypothesis, exactly one of the LR_j follows the $\chi^2_{\mathrm{df}_j}$-distribution whereas the remaining likelihood ratios are large compared to the likelihood ratio of the true model for a sufficiently large number T of data points. Thus, we use the vector $(\mathrm{LR}_1, \ldots, \mathrm{LR}_r)$ as test statistic and choose for each component a $(1-\alpha)$-quantile $q_{1-\alpha,j}$ according to the corresponding $\chi^2_{\mathrm{df}_j}$-distribution. We reject the null hypothesis if all likelihood ratios LR_j are greater than the chosen quantile $q_{1-\alpha,j}$.

For a test of at most size β, in principle, the α for the quantiles $q_{1-\alpha,j}, 1 \leqslant j \leqslant r$, has to be adjusted according to the following equation for the rejection probability under the null hypothesis:

$$P_{H_0}(\mathrm{LR}_1 > q_{1-\alpha,1}, \ldots, \mathrm{LR}_r > q_{1-\alpha,r}) \leqslant \beta. \qquad (16)$$

Although eqn (16) cannot be solved for α, α can be chosen to be equal to β without changing the actual size of test too dramatically for a sufficiently large number T of data points. This is due to the following inequality assuming that, for ease of notation, the model HMM_1 contains the true model:

$$P_{H_0}(B_1, \ldots, B_r) \leqslant \beta,$$

$$\alpha P_{H_0}(B_2, \ldots, B_r | B_1) \leqslant \beta,$$

$$\alpha \leqslant \underbrace{\frac{\beta}{P_{H_0}(B_2, \ldots, B_r | B_1)}}_{\nearrow 1 \text{ for } T\to\infty} \qquad (17)$$

with $B_j = \{\mathrm{LR}_j > q_{1-\alpha,j}\}$. Since the complements of $B_2, \ldots, B_r$ converge to the empty set under the above-stated assumption, $P_{H_0}(B_2, \ldots, B_r | B_1)$ converges to one for an increasing number of data points T.

4.2. STEP TWO: SELECTION

A successful "step one" gives confidence that one of the models $\mathrm{HMM}_1, \ldots, \mathrm{HMM}_r$ is the true one. The likelihood ratios are now used to select a model: if exactly one likelihood ratio LR_{j_0} is smaller than the corresponding quantile

$q_{1-\alpha,j_0}$ and all other likelihood ratios exceed their quantiles, we decide for the model HMM_{j_0}. We interpret the event that more than one likelihood ratio is smaller than the corresponding quantile as an indication that the number of data points is not sufficient to distinguish reliably between the models to choose from.

In the next section, we demonstrate the feasibility of the proposed selection procedure in a simulation study.

5. Simulation Study

We exemplify the model selection procedure proposed in Section 4 by the following simulation study. We suppose that the ion channel under investigation has two open states and two closed states justified by hypothetical prior knowledge, e.g. some previous experiments. Furthermore, we make the assumption that we can summarize this prior knowledge into the following hypotheses about the gating scheme:

$$\text{Gating scheme 1: } O_1 \rightleftharpoons O_2 \rightleftharpoons C_1 \rightleftharpoons C_2, \quad (18)$$

$$\text{Gating scheme 2: } O_1 \rightleftharpoons C_1 \rightleftharpoons C_2 \rightleftharpoons O_2. \quad (19)$$

Gating scheme 1 has one gateway state; successive open and closed dwell times, therefore, are independent (Fredkin *et al.*, 1983). In gating scheme 2, the two-dimensional distribution of successive open and closed dwell times does not factor into the product of the one-dimensional dwell time distributions and for this reason the two models are not nested. As both gating schemes do not contain any loops, both models always satisfy the principle of detailed balance and the matrix $A^{(S)}$ [eqn (9)] provides a sufficient parameterization of all possible likelihood functions of both models. Gating schemes 1 and 2 have six transition rates and matrix $A^{(S)}$ has eight independent entries, namely the four entries in $A_{oc}^{(S)}$ and in $A_{co}^{(S)}$. Thus, the difference in the number of degrees of freedom between the general model and the two specific models is two. In summary, our model selection procedure requires the following steps for this simulation study.

Preliminary step: estimate the eight parameters $A_{oc}^{(S)}$ and $A_{co}^{(S)}$ of the general model, the transition rates of gating schemes 1 and 2 and calculate for each of gating schemes 1 and 2 the two-fold log likelihood ratios LR_1 and LR_2 [eqn (15)] between the general model and the specific gating scheme.

Step 1: *test for consistency*: for a test of approximate size α, choose a $1-\alpha$ quantile $q_{1-\alpha}$ of the χ^2-distribution with two degrees of freedom. Reject the null hypothesis that either gating scheme 1 or gating scheme 2 is the true model if both two-fold log likelihood ratios, LR_1 and LR_2, are above the chosen quantile $q_{1-\alpha}$. If the null hypothesis is not rejected, proceed with "step 2", otherwise neither gating scheme 1 nor gating scheme 2 is an acceptable model for the given data.

Step 2: *selection*: select gating scheme 1 if the two-fold log likelihood ratio LR_1 for gating scheme 1 is smaller than or equal to the quantile $q_{1-\alpha}$ as well as the two-fold log likelihood ratio LR_2 for gating scheme 2 is greater than the quantile $q_{1-\alpha}$. Select gating scheme 2 accordingly. If both two-fold log likelihood ratios LR_1 and LR_2 fall below the quantile $q_{1-\alpha}$, we take this as an indication that the number of data points is not sufficient to distinguish reliably between gating schemes 1 and 2.

In particular, we have four possible outcomes of our model-selection procedure: none of the gating schemes is acceptable, either gating scheme 1 or gating scheme 2 is selected or the number of data points does not allow a reliable distinction between the two models.

We investigate the model selection procedure for an increasing measurement time: 66, 131, 262, and 524 s with a sampling rate of 1 kHz. Accordingly, the number of data points varies from: $T = 2^{16}, 2^{17}, 2^{18}$ to 2^{19}. Gating scheme 1 should be the true model with the following generator matrix:

$$\left(\begin{array}{c|c} Q_{oo} & Q_{oc} \\ \hline Q_{co} & Q_{cc} \end{array}\right) = \left(\begin{array}{cc|cc} -75 & 75 & 0 & 0 \\ 150 & -280 & 130 & 0 \\ \hline 0 & 100 & -250 & 150 \\ 0 & 0 & 70 & -70 \end{array}\right) \text{ (in Hz).} \quad (20)$$

For our model selection procedure, the entries in the sub-matrices $A_{oc}^{(S)}$ and $A_{co}^{(S)}$ have to be estimated from the data. As we know the true generator matrix given by eqn (20) in our simulation study, we calculate the true matrix $A^{(S)}$ corresponding to the true generator matrix in the following in order to exemplify the calculations involved in our model selection procedure.

First of all, the transition probability matrix A is given by eqn (2):

$$A = \exp\left(Q\frac{1}{1000}s\right)\left(\begin{array}{c|c} A_{oo} & A_{oc} \\ \hline A_{co} & A_{cc} \end{array}\right)$$

$$\approx \left(\begin{array}{cc|cc} 0.9326 & 0.0632 & 0.0040 & 0.0002 \\ 0.1263 & 0.7653 & 0.1003 & 0.0080 \\ \hline 0.0062 & 0.0772 & 0.7882 & 0.1285 \\ 0.0001 & 0.0029 & 0.0600 & 0.9370 \end{array}\right). \quad (21)$$

The matrix $A^{(S)}$ and the transition probability matrix A are related by the similarity transformation S that diagonalizes the sub-matrices A_{oo} and A_{cc}:

$$S = \left(\begin{array}{cc|cc} 1.0998 & -0.0998 & 0 & 0 \\ 0.6743 & 0.3257 & 0 & 0 \\ \hline 0 & 0 & 0.7349 & 0.2651 \\ 0 & 0 & 1.0839 & -0.0839 \end{array}\right). \quad (22)$$

Thus, the matrix $A^{(S)}$ corresponding to the true generator matrix is given by

$$A^{(S)} = S^{-1}AS$$

$$\approx \left(\begin{array}{cc|cc} 0.9714 & 0 & 0.0218 & 0.0069 \\ 0 & 0.7266 & 0.2080 & 0.0654 \\ \hline 0.0157 & 0.0066 & 0.9777 & 0 \\ 0.1783 & 0.0742 & 0 & 0.7475 \end{array}\right). \quad (23)$$

For each number of data points, we simulate 1000 noisy recordings. We approximate the observational noise actually found in experiments by a white Gaussian process. The ratio between the standard deviation of the noise and the difference between the conductance levels of the open and closed states is set to one. For the sake of simplicity, we assume that the conductance levels and the standard deviations of the noise and the initial distribution of both gating schemes are known in advance. The initial distribution of the general model is determined from the fixed initial distribution of gating scheme 1 and the similarity transformation given by eqn (22). The transition rates of both gating schemes as well as the entries in $A_{oc}^{(S)}$ and $A_{co}^{(S)}$ are estimated directly from the simulated noisy data set by the maximum likelihood method. The maximization of the likelihood function is performed numerically by the EM algorithm (Dempster *et al.*, 1977; Meng & van Dyk, 1997; Michalek & Timmer, 1999) and a nonlinear maximization routine based on a quasi-Newton method (the subroutine E04UCF from NAG, 1997). For the calculation of the first derivatives of the likelihood function, we use Fisher's identity (Fisher, 1925; Jamshidian & Jennrich, 1997) and the "sinch"-algorithm described by Najfeld & Havel (1995) to evaluate the derivatives of the matrix exponential for estimating the transition rates.

Figure 1 shows the distribution function of the likelihood ratio statistic of gating scheme 1, LR_1, and of gating scheme 2, LR_2, for $T = 2^{16}$ and for $T = 2^{19}$. LR_1 approximately follows a χ^2-distribution with two degrees of freedom, which corresponds to eqn (15). We expect that the separation between the distributions of the true model, LR_1, and of the wrong model, LR_2 increases with the number of data points due to the dichotomy of Kakutani (Shiryaev, 1995). For the long measurement time of $T = 2^{19}$ the distributions of LR_1 and of LR_2 separate fairly well; for 2^{16} data points, there is a considerable overlap of the two distributions.

In Fig. 2, we investigate the scaling behavior of the likelihood ratio statistic for the wrong model, LR_2, with the number of data points T. In the case of independent random variables the twofold log likelihood ratio of the wrong model is asymptotically normally distributed and its mean and variance are proportional to the number of data points T (Vuong, 1989). Figure 2 illustrates that the mean and the variance of LR_2 have the same scaling behavior with the number of data points. For each T, the simulated sample of

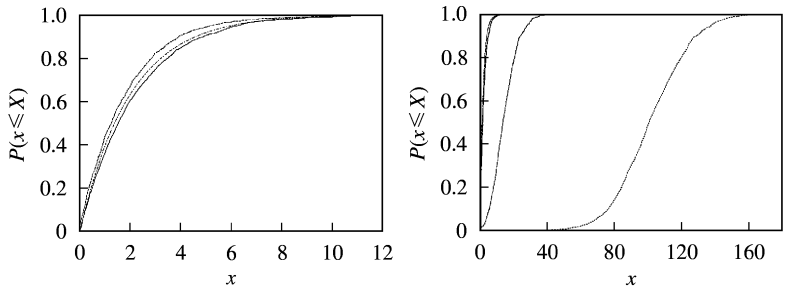

FIG. 1. Simulation study: the distribution functions of the likelihood ratio statistic for gating schemes 1 and 2, LR_1 and LR_2 for $T = 2^{16}$ and 2^{19}: x indicates either LR_1 or LR_2. The likelihood ratio statistic of the true gating scheme 1 follows asymptotically a χ^2-distribution with two degrees of freedom. The plot on the left-hand side compares the empirical distribution of LR_1 for a different number of data points with a χ^2_2-distribution: (·····) χ^2_2; (——) LR_1: 2^{19}; (- - - - -) LR_1: 2^{16}; (·····) LR_2: 2^{19}; (——) LR_2: 2^{16}.

The separation of the distributions of the true and the wrong model increases with the number of data points. The plot on the right-hand side shows the separation of the distributions of the true and the wrong model depending on the number of data points.

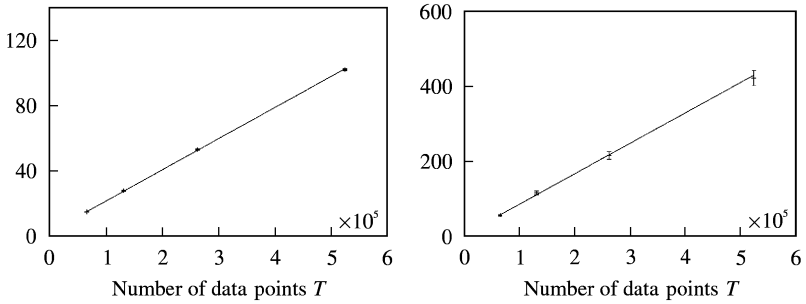

FIG. 2. Scaling behavior of the distribution of LR_2 with the number of data points T. The plot on the left-hand side shows the dependency of the mean of LR_2 on the number of data points T, the plot on the right-hand side the dependency of the variance of LR_2 on the number of data points T. The error bars indicate the standard deviation of the estimated mean and variance: (⟼) mean of LR_2; (·····) fitted line; (⟼) variance of LR_2.

LR_2-values is standardized and the empirical distribution functions of these standardized samples are compared with the standard normal distribution function in Fig. 3. Hence, the distribution of the likelihood ratio statistic of the wrong model has the same asymptotic properties as the distribution function for independent random variables. Especially, the investigation of the scaling behavior of the likelihood ratio statistic for the wrong model allows to estimate the required number of data points for which the overlap between the distributions of the true model and those of the wrong model becomes negligible.

Table 1 summarizes the results of the simulation study; the percentage for all possible events of the proposed selection procedure is shown. The parameter α which fixes the rejection probability in "step 1" of the selection procedure [see eqn (16)] is set to 5%. As indicated by eqn (17), the actual size of the test in "step 1", the first row in Table 1, does not differ significantly from the parameter α. For $T = 2^{16}$, in 11% of the cases, we cannot reject either of the two models. This is due to the significant overlap of the distributions of LR_1 and LR_2 shown in Fig. 1, so a reliable distinction between both models cannot be made under these conditions. Table 2 summarizes the results of the simulation study for "step 2" of the selection procedure; the percentage for all possible events conditioned on the acceptance of the null hypothesis is shown. For $T = 2^{18}$ and 2^{19} the overlap between the distributions of LR_1 and LR_2 is negligible and we always select the correct model. In particular, this suggests that

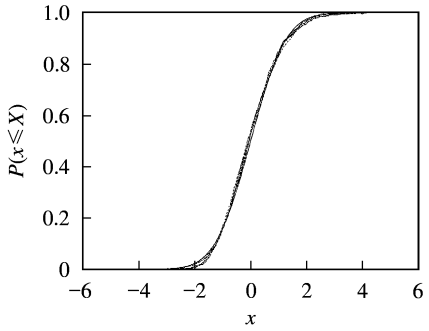

FIG. 3. Asymptotic normality of the distribution of LR_2. For each T, the simulated LR_2 values are standardized and the empirical distribution functions of these standardized samples are plotted together with the standard normal distribution function. (- - - - - - -), 2^{16}; (· · · · · · ·), 2^{17}; (· · · · · ·), 2^{18}; (- · - · - · -), 2^{19}; (——), $N(0, 1)$.

TABLE 1

Summary of results of the simulation study: the leftmost column contains the possible event of the proposed model selection procedure. Each row contains the percentage of the event mentioned in the leftmost column

T	2^{16}	2^{17}	2^{18}	2^{19}
"Accept H_0"	0.970	0.96	0.94	0.94
"Reject H_0"	0.030	0.04	0.06	0.06
"Select model 1"	0.852	0.95	0.94	0.94
"Select model 2"	0.003	0.0	0.0	0.0
"Not enough data"	0.114	0.01	0.0	0.0

TABLE 2

Summary of results of the simulation study conditioned on the acceptance of H_0. Each row contains the percentage of the event mentioned in the leftmost column

T	2^{16}	2^{17}	2^{18}	2^{19}
"Select model 1"	0.880	0.994	1	1
"Select model 2"	0.003	0.0	0.0	0.0
"Not enough data"	0.117	0.006	0.0	0.0

the rejection probability of "step 1" should be adjusted to the number of data points: the larger the number of data points, the smaller the parameter α and the smaller the actual size of the test by eqn (17) following the classical suggestion (Neyman & Pearson, 1933; Bauer *et al.*, 1988).

6. Discussion

Model selection methods are based either on some criteria which compare some kind of "goodness of fit" between several models or on hypothesis tests which test for simplifications of a general model to be still compatible with the data.

The most prominent examples for the first case are the information criteria (Akaike, 1993; Schwarz, 1978; Burnham & Anderson, 1998). The AIC and the BIC were first used by Ball & Sansom, (1989) in the context of ion channel gating. These criteria are estimates of the Kullback–Leibler distance between the true probability distribution of the data and the probability distribution of the model (Konishi & Kitagawa, 1996). A model selection procedure based on one of these criteria will choose the model with the smallest estimated Kullback–Leibler distance. Konishi & Kitagawa (1996) show, however, that these criteria do not always reliably estimate the true Kullback–Leibler distance, especially, if the models are misspecified. Moreover, these criteria do not provide a scale to assess the significance of differences between values of these criteria for different models.

The main advantage of our proposed model-selection procedure over classical information criteria is that the embedding of the gating schemes in a general model gives us the needed scale on which we can reliably compare likelihood values of different models. Moreover, this allows to decide if the proposed models are at all compatible with the measured data and if they are, to decide if the number of data points is sufficient to distinguish reliably between the models. In the context of classical information criteria, none of these decisions is possible and it is common practice to choose the model with the smallest estimated Kullback–Leibler distance.

As our proposed selection procedure is currently limited to models with the same number of open and closed states, we will investigate generalizations to drop this restriction in future work.

REFERENCES

AIDLEY, D. J. & STANFIELD, P. R. (1996). *Ion Channels: Molecules in Action*. Cambridge: Cambridge University Press.

AKAIKE, H. (1993). Information theory and an extension of the maximum likelihood principle. In: *Breakthroughs in Statistics* (Kotz, S. & Johnson, N. L., eds), Springer Series in Statistics, Vol. 1, pp. 599–624. Berlin: Springer.

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. & RICE, J. A. (1992). Stochastic models for ion channels: introduction and bibliography. *Math. Biosci.* **112,** 189–206.

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

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

BAUER, P., PÖTSCHER, B. B. & HACKL, P. (1988). Model selection by multiple test procedures. *Statistics* **1,** 39–44.

BAUM, L. E. & PETRIE, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.* **40,** 164–171.

BAUM, L. E., PETRIE, T., SOULES, G. & WEISS, N. (1970). A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.* **41,** 164–171.

BICKEL, P., RITOV, Y. & RYDÉN, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist.* **26,** 1614–1635.

BURNHAM, K. P. & ANDERSON, D. R. (1998). *Modelselection and Inference—A Practical Information–Theoretic Approach.* Berlin: Springer.

CHUNG, S.-H. & KENNEDY, R. A. (1996). Coupled Markov chain model: characterization of membrane channel currents with multiple conductance sublevels as partially coupled elementary pores. *Math. Biosci.* **133,** 111–137.

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. *Philos. Trans. R. Soc. Lond. B* **329,** 265–285.

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. (1982). On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Philos. Trans. Roy. Soc. Lond. B* **300,** 1–59.

COLQUHOUN, D., HAWKES, A. G. & SRODZINSKI, K. (1996). Joint distributions of apparent open times and shut times of single ion channels and the maximum likelihood fitting of mechanisms. *Philos. Trans. R. Soc. Lond. A* **354,** 2555–2590.

COX, D. & HINKLEY, C. (1974). *Theoretical Statistics.* London: Chapman & Hall.

DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. R. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* **39,** 1–38.

FISHER, R. (1925). Theory of statistical estimation. *Proc. Camb. Philos. Soc.* **22,** 700–725.

FREDKIN, D. R., MONTAL, M. & RICE, J. A. (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* (LeCam, L. M. & Olshen, R. A., eds), Vol. I, pp. 269–289. Belmont: Wadsworth Press.

FREDKIN, D. R. & RICE, J. A. (1986). On aggregated Markov processes. *J. Appl. Probab.* **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.

HAWKES, A. G., JALALI, A. & COLQHOUN, D. (1990). The distributions of the apparent open times and shut times in a single channel record when brief events cannot be detected. *Philos. Trans. R. Soc. Lond. A* **332,** 511–538.

HAWKES, A. G., JALALI, A. & COLQHOUN, D. (1992). Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Philos. Trans. R. Soc. Lond. B* **337,** 383–404.

HILLE, B. (1992). *Ionic Channels of Excitable Membranes*, 2nd Edn, Sinauer.

HORN, R. & LANGE, K. (1983). Estimating kinetic constants from single channel data. *Biophys. J.* **43,** 207–223.

ITO, H., AMARI, S.-I. & KOBAYASHI, K. (1992). Identifibility of hidden Markov information sources and their minimum degrees of freedom. *IEEE Trans. Inf. Theory* **38,** 324–333.

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.

KIJIMA, S. & KIJIMA, H. (1987). Statistical analysis of channel current form a membrane patch I. Some stochastic properties of ion channel or molecular systems in equilibrium. *J. theor. Biol.* **128,** 423–434.

KONISHI, S. & KITAGAWA, G. (1996). Generalised information criteria in model selection. *Biometrika* **4,** 875–890.

LANG, S. (1987). *Linear Algebra*, Undergraduate Texts in Mathematics, 3rd Edn. Berlin: Springer.

MENG, X. & VAN DYK, D. (1997). The EM algorithm—an old folk-song sung to a fast tune. *J. R. Statist. Soc. B* **59,** 511–567.

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.* **28,** 605–609.

MICHALEK, S. & TIMMER, J. (1999). Estimating rate constants in hidden Markov models by the EM algorithm. *IEEE Trans. Signal Proc.* **47,** 226–228.

MICHALEK, S., WAGNER, M. & TIMMER, J. (2000). A new approximate likelihood estimator for ARMA-filtered hidden Markov models. *IEEE Trans. Signal Proc.* **48,** 1537–1547.

NAG (1997). *Fortran Library Mark 18.* Oxford, U.K.: 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.

NEYMAN, L. & PEARSON, E. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. Roy. Soc. A* **231,** 289–337.

QIN, F., AUERBACH, A. & SACHS, F. (1996). Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* **70,** 264–280.

QIN, F., AUERBACH, A. & SACHS, F. (1997). Maximum likelihood estimation of aggregated Markov processes. *Proc. R. Soc. Lond. B* **264,** 375–383.

ROUX, B. & SAUVE, R. (1985). A general solultion to the time interval omission problem applied to single channel analysis. *Biophys. J.* **48,** 149–158.

SAKMANN, B. & NEHER, E., eds (1995). *Single-Channel Recording*. 2nd Edn. New York: Plenum Press.

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

SELF, S. G. & LIANG, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Statist. Assoc.* **82,** 605–610.

SHIRYAEV, A. (1995). *Probability*, 2nd Edn. Berlin: Springer.

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

VENKATARAMANAN, L., KUC, R. & SIGWORTH, F. (1998a). Identification of hidden Markov models for ion channel currents. Part II: state-dependent excess noise. *IEEE Trans. Signal. Proc.* **46,** 1916–1929.

VENKATARAMANAN, L., WALSH, J. L., KUC, R. & SIGWORTH, F. (1998b). Identification of hidden Markov models for ion channel currents. Part I: colored, background noise. *IEEE Trans. Signal. Proc.* **46,** 1901–1915.

VUONG, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57,** 307–333.

WAGNER, M., MICHALEK, S. & TIMMER, J. (1999). Testing for the number of states in hidden Markov models with application to ion channel data. In: *Studies in Classification, Data Analysis, and Knowledge Organization, Classification in the Information Age* (Gaul, W. & Locarek-Junge, H., eds), pp. 260–267. Berlin: Springer.

AUTHOR QUERY FORM

HARCOURT PUBLISHERS

JOURNAL TITLE: JTBI DATE:07/12/2000
ARTICLE NO. : 20002230

Queries and / or remarks

Manuscript Page/line	Details required	Author's response
1	Please indicate the corresponding author.	
26	Hille (1992) → Place of publication?	