

Testing the Markov condition in ion channel recordings

J. Timmer* and S. Klein

Freiburger Zentrum für Datenanalyse und Modellbildung, Albertstrasse 26-28, D-79104 Freiburg, Germany

(Received 25 July 1996; revised manuscript received 2 December 1996)

A statistical test is presented to decide whether data are adequately described by probabilistic functions of finite state Markov chains ("hidden Markov models") as applied in the analysis of ion channel data. Particularly, the test can be used to decide whether a system obeys the Markov condition. Simulation studies are performed in order to investigate the sensitivity of the proposed test against violations of the model assumptions. The test can be applied analogously to Markov models. [S1063-651X(97)15003-6]

PACS number(s): 87.10.+e, 02.50.Ga

I. INTRODUCTION

Ion channels are large proteins located in the membranes of cells. They serve for signal transmission and regulate the concentration of ions in the cell. The channels open and close in a stochastic manner dependent on external conditions such as trans-membrane voltage difference, concentration of ligands, or mechanical stress. In general the channels have several states in which they are closed and open, respectively. They might even possess open states with different conductivity. The noisy current in the range of pA through single channels can be measured by the patch clamp technique [1].

Analyzing data from ion channels generally relies on the assumption of a Markovian dynamics. This holds for inferring the number of channel states and mean dwell times by fitting exponentials to dwell time histograms [2,3], for explicit modeling of low-pass filtered records by Markov models [4,5], and also for analyzing unfiltered records by hidden Markov models [6–11].

In many cases, however, it is not evident from empirical data whether the system actually obeys the Markov condition. For two reasons, this assumption has given rise to a lively discussion [12–14]. On the one hand, the information about the validity of this condition can provide valuable insight into the system under investigation [15,16]. On the other hand, conclusions drawn from a model that does not fit the process that has produced the data are very likely to lead to erroneous results. Thus, it is desirable to test whether the process is adequately described by the selected (hidden) Markov model.

We propose a test to perform this task. It is based on the asymptotic distributions of the log-likelihood that holds if the model is valid. A deviation from the expected distribution provides a test for the model. In order to evaluate the sensitivity of the proposed test against a violation of the null hypothesis four simulation studies were performed where the assumption of an underlying hidden Markov model is violated in various manners. A fifth simulation study shows that the test is also useful to estimate the minimum number of

states in a Markov model needed to be compatible with the data.

The paper is organized as follows: In the next section, we briefly review the hidden Markov model. In Sec. III the test statistic is introduced. The power of test is evaluated by simulation studies in Sec. IV. As presented, the test applies to hidden Markov models, however, it can be applied analogously to Markov models.

II. HIDDEN MARKOV MODELS

Hidden Markov models (HMM), introduced in [17] and used in diverse fields such as speech recognition [18] and ion channel analysis, are generalizations of Markov models that allow one to include observational noise. HMMs can be formulated in continuous-time and discrete-time versions. Following [7] we chose the latter. The results also hold for continuous-time models.

A stationary hidden Markov model is given by an unobservable process \mathbf{X}_t , which can take one of the s states for every point t in time. The probabilities for a change from a state i to a state j are described by a time-independent transition matrix (a_{ij}) ($i, j = 1, \dots, s$). Since each row of the matrix is normalized to unity, the $s \times s$ matrix (a_{ij}) has $s(s-1)$ free parameters. The observations Y_t are determined by the output probability densities of each of the s states. These densities are described by parameter vectors ϕ_i ($i = 1, \dots, s$). For example, the density functions $f(y, \phi_i)$ can be given by Gaussian distributions with different means and variances.

For ease of notation, the parameters of the hidden Markov model are arranged in a single parameter vector θ . Its dimension is denoted by r . For example, in the case of s states with Gaussian output probabilities the model has $r = s(s-1) + 2s = s^2 + s$ parameters.

Given an observed time series $\mathbf{Y}_{1..N} = Y_1, \dots, Y_N$ of length N , the parameter vector θ can be estimated by a maximum likelihood procedure [19–21]. For the calculation of the log-likelihood function

$$L_N(\mathbf{Y}_{1..N}, \theta) = \ln P[\mathbf{Y}_{1..N} | \theta] \quad (1)$$

the so-called forward probabilities to find the system in state i at time t given the data up to time t are defined by

*Author to whom correspondence should be addressed. FAX: +49/761/203-5967. Electronic address: jeti@fdm.uni-freiburg.de

$$\alpha_i(t, \mathbf{Y}_{1\dots t}, \boldsymbol{\theta}) = P[X_t = i, \mathbf{Y}_{1\dots t} | \boldsymbol{\theta}]. \quad (2)$$

They can be calculated using the recursion

$$\alpha_i(t, \mathbf{Y}_{1\dots t}, \boldsymbol{\theta}) = \sum_{j=1}^s \alpha_j(t-1, \mathbf{Y}_{1\dots t-1}, \boldsymbol{\theta}) a_{ij}(\boldsymbol{\theta}) f(Y_t, \phi_j) \quad (3)$$

and lead to the log-likelihood function by

$$L_N(\mathbf{Y}_{1\dots N} | \boldsymbol{\theta}) = \ln \sum_{i=1}^s \alpha_i(N, \mathbf{Y}_{1\dots N}, \boldsymbol{\theta}). \quad (4)$$

An estimate $\hat{\boldsymbol{\theta}}_N$ can be obtained by maximizing $L_N(\mathbf{Y}_{1\dots N} | \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ either by nonlinear optimization or by the expectation-maximization algorithm, i.e., the Baum-Welsh reestimation formulas [19,22]. Here, all numerical calculations have been performed by the latter method as described in [20] since it behaves numerically more stably than nonlinear optimization. For ease of notation we suppress the dependence of $L_N(\mathbf{Y}_{1\dots N} | \boldsymbol{\theta})$ on $\mathbf{Y}_{1\dots N}$ in the following.

III. THE TEST STATISTIC

In this section we introduce the statistics to test the adequacy of a given hidden Markov model to describe an observed time series.

Under mild regularity conditions, the difference between the maximum likelihood estimators $\hat{\boldsymbol{\theta}}_N$ and the true parameters $\boldsymbol{\theta}_0$ are generally believed due to central limit theorems to converge to a normal distribution

$$\sqrt{N}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_N) \sim \mathcal{N}(0, \Sigma), \quad (5)$$

with, asymptotically,

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} L_N(\hat{\boldsymbol{\theta}}_N) \rightarrow -\frac{1}{N} \Sigma_{ij}^{-1}. \quad (6)$$

This has been proven for independent random variables (see, e.g., [23]), Markov models [24], and hidden Markov models with discrete output probabilities [17]. For hidden Markov models with continuous output probabilities, up to now, the consistency of the maximum likelihood estimators [25], the local asymptotic normality in the sense of Le Cam [26,27], and the asymptotic normality of maximum split data likelihood estimators has been shown [28]. The proof of asymptotic normality of the maximum likelihood estimators in hidden Markov models is announced [29].

Given the asymptotic normality of the estimators of Eqs. (5), the distribution of the maximum log-likelihood $L_N(\hat{\boldsymbol{\theta}}_N)$, which is itself a random variable, can be derived by a Taylor expansion (see [23] for a detailed discussion):

$$\begin{aligned} L_N(\boldsymbol{\theta}_0) &= L_N(\hat{\boldsymbol{\theta}}_N) + \frac{\partial}{\partial \theta_i} L_N(\hat{\boldsymbol{\theta}}_N) (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_N) \\ &+ \frac{1}{2} (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_N) \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_N(\hat{\boldsymbol{\theta}}_N) (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_N) \\ &+ \mathcal{O}(|\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_N|^3). \end{aligned} \quad (7)$$

The second term on the right-hand side vanishes due to the estimation procedure. Neglecting higher order terms, solving for $2[L_N(\hat{\boldsymbol{\theta}}_N) - L_N(\boldsymbol{\theta}_0)]$ and using Eqs. (5) and (6) yields

$$2[L_N(\hat{\boldsymbol{\theta}}_N) - L_N(\boldsymbol{\theta}_0)] \sim \chi_r^2. \quad (8)$$

This relation holds asymptotically if the model is specified correctly. The number N of data needed to reach the asymptotic regime depends on the process. Simulation studies not presented here show that Eq. (8) holds if each transition between the states has occurred at least 10 times.

For the test we estimate $\boldsymbol{\theta}_0$ based on the whole time series of length N and denote this estimate by $\hat{\boldsymbol{\theta}}_N$. Then, the time series is divided in K parts of length $M = N/K$. For each these parts we estimate the parameters $\hat{\boldsymbol{\theta}}_M$ and evaluate the log-likelihoods $L_M(\hat{\boldsymbol{\theta}}_M)$ and $L_M(\hat{\boldsymbol{\theta}}_N)$. Asymptotically, i.e., for $N \rightarrow \infty$, $M \rightarrow \infty$, but $M/N \rightarrow 0$, the distribution of $2[L_M(\hat{\boldsymbol{\theta}}_M) - L_M(\hat{\boldsymbol{\theta}}_N)]$ is given by

$$2[L_M(\hat{\boldsymbol{\theta}}_M) - L_M(\hat{\boldsymbol{\theta}}_N)] \sim \chi_r^2. \quad (9)$$

By the proposed procedure we obtain K samples of the χ_r^2 distribution if the model is valid. In order to judge whether Eq. (9) holds, we apply the Kolmogorov-Smirnov test for the consistency of an empirical distribution with a proposed theoretical distribution [30]. The Kolmogorov-Smirnov-test statistic is denoted by \mathcal{Z} in the following.

IV. EVALUATION OF THE POWER OF THE TEST

In this section, we evaluate the power of the above proposed test; i.e., we investigate the sensitivity of the test against a violation of the null hypothesis that the data were produced by a hidden Markov model. Of course, it is not possible to consider all imaginable alternative hypotheses. One has to restrict oneself to a reasonable class of alternative hypotheses. We choose four alternative hypotheses that violate the model assumptions: (1) Nonstationary transition probabilities; (2) dwell time dependent transition probabilities; (3) a fractal model; (4) refractory time. Finally, we show that the proposed test enables one to estimate the smallest number of states of the Markov process compatible with the data.

A. Model definition

In order to evaluate the power of the test numerically we chose a hidden Markov model with three states and Gaussian output probability functions representing, e.g., one ion channel with two different conductance levels. The transition matrix A is given by

$$A = \begin{pmatrix} 0.90 & 0.05 & 0.05 \\ 0.06 & 0.92 & 0.02 \\ 0.03 & 0.02 & 0.95 \end{pmatrix}. \quad (10)$$

The means and the variances of the Gaussian output probability functions were chosen to be

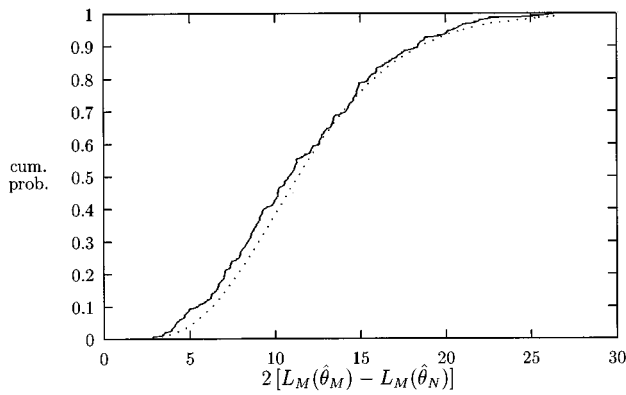


FIG. 1. The empirical cumulative distribution of $2[L_M(\hat{\theta}_M) - L_M(\hat{\theta}_N)]$ (solid line) and the expected cumulative χ^2_{12} distribution (dotted line) for the process defined by Eqs. (10) and (11).

$$\begin{aligned} \mu_1 &= 0.0, & \sigma_1^2 &= 0.1, \\ \mu_2 &= 1.0, & \sigma_2^2 &= 0.1, \\ \mu_3 &= 2.0, & \sigma_3^2 &= 0.1. \end{aligned} \quad (11)$$

The dimension r of the parameter vector θ is 12. We simulated time series of length $N=150,000$ and divided it into $K=150$ time series of length $M=1000$ to perform the test. To apply the test the resulting time series must be long enough for the asymptotic results to be valid. If the off-diagonal elements of the transition matrix are of similar magnitude, as a rule of thumb, this condition is met, if the time series have a length of at least

$$M = 10s\tau_{\max}, \quad (12)$$

with s the number of states and τ_{\max} the largest dwell time. For the chosen model, the dwell times are 10 and 12.5, respectively, 20 units of time.

Figure 1 shows the expected cumulative χ^2_{12} distribution according to Eq. (9) and the empirical cumulative distribution for the chosen process. It indicates a good qualitative agreement of the two distributions. In order to quantify this, we counted for 200 realizations of the process the number of cases where the hypothesis of consistency of the two distributions was rejected by the Kolmogorov-Smirnov test at a significance level of 5%. This results in an actual rejection rate of 4.5%, indicating that the asymptotic regime is reached for the chosen situation.

B. Power of the test

To investigate the power of a test, usually, for different degrees of violation of the null hypothesis on the order of 1000 times series are realized, the test is performed, and the fraction of rejected null hypothesis given a certain significance level α is calculated in dependence of the degree of violation. However, this procedure to evaluate the power requires an enormous computational effort to obtain a good approximation of the underlying smooth behavior since for the chosen model and number of data the maximization of the log-likelihood for a single time series requires ca. 45 min

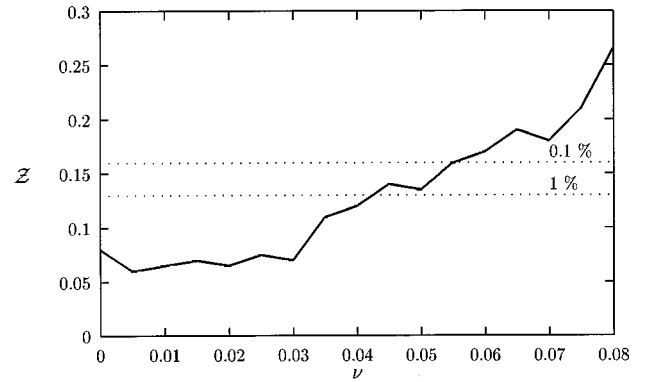


FIG. 2. The effect of drifting transition probabilities. Shown is the averaged test statistic Z of the Kolmogorov-Smirnov test for increasing drift rates ν . The 1% and the 0.1% significance levels are marked.

on an IBM 6000 RISC workstation. Therefore, we chose another way to display the power of the test. Instead of counting the simulation runs with rejected null hypothesis we average the test statistic of 10 realizations for each degree of violation of the null hypothesis to approximate the smooth curve. This procedure estimates the mean of the distribution of the test statistic for the alternative hypotheses. Simulation studies show that these distributions of the test statistic are symmetric and that their variance is rather constant. Therefore, the mean calculated here corresponds to the median of the distributions and is related monotonically to the fraction calculated usually. Thus, this procedure yields essentially the same information as the canonical method that requires only 1% of the computational effort.

We now discuss the different simulation studies to evaluate the power of the proposed test.

1. Nonstationary transition probabilities

In order to investigate the sensitivity against violations of the stationarity assumption, nonstationarity of the transition probability of the first state is introduced by

$$\tilde{a}_{11}(t) = a_{11} - \frac{(s-1)\nu t}{N}, \quad (13)$$

$$\tilde{a}_{1j}(t) = a_{1j} + \frac{\nu t}{N} \quad (j=2, \dots, s), \quad (14)$$

where s again denotes the number of states. This time dependency of the transition probabilities causes a decreasing dwell time of the first state. The drift rate ν serves as the parameter for the null hypothesis violation. As outlined above, we judge the performance of the test by averaging the test statistic of ten simulations for every degree of the null hypothesis violation. Figure 2 shows the averaged test statistic Z of the Kolmogorov-Smirnov test with increasing violation of the null hypothesis and the 1% and 0.1% levels of significance, respectively. A change of 10% over the whole observation time in the dwell probability of one of three states is detectable by the proposed test.

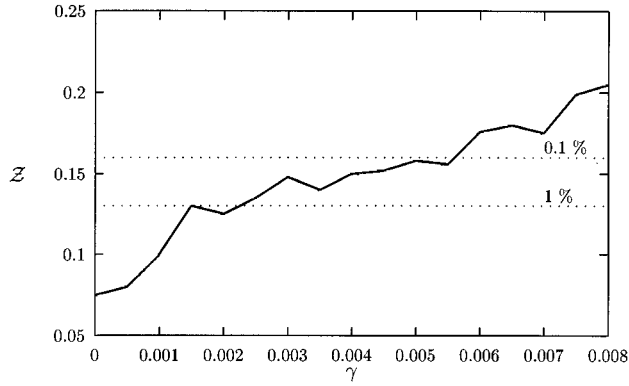


FIG. 3. The effect of dwell-time-dependent transition probabilities. Shown is the averaged test statistic \mathcal{Z} for increasing degrees γ of the null hypothesis violation. The 1% and the 0.1% significance levels are marked.

2. Dwell-time-dependent transition probabilities

The Markov condition, stating that the transition probabilities between the states do not depend on the time already spent in the states, is violated by increasing the probability to leave any state proportional to the time t_{in} already spent in the state. The proportionality constant γ parametrizes the violation of the null hypothesis.

$$\tilde{a}_{ii}(t_{in}) = a_{ii} - (s-1)\gamma t_{in}, \quad (15)$$

$$\tilde{a}_{ij}(t_{in}) = a_{ij} + \gamma t_{in} \quad (i \neq j). \quad (16)$$

Figure 3 shows the result of the simulation. A change of more than 1% per time step for the dwell probabilities leads to the rejection of the hypothesis that the time series was generated by a Markov process. During the simulation, it was controlled that the condition $0 < \tilde{a}_{ij}(t_{in}) < 1$ was not violated.

3. A fractal model

Another possibility to violate the Markov condition is given by the fractal models [15]. For these models, the dwell probability increases with the time t_{in} already spent in the state. The transition probabilities of a fractal model are given by

$$\tilde{a}_{ii}(t_{in}) = 1 - (1 - a_{ii})t_{in}^{1-D}, \quad (17)$$

$$\tilde{a}_{ij}(t_{in}) = a_{ij}t_{in}^{1-D}, \quad i \neq j, \quad (18)$$

where D is the fractal dimension that parametrizes the violation of the Markov condition. For $D=1$ the Markov model results. The result of the simulation in Fig. 4 reveals that for the given model a fractal dimension of e.g., 1.1 will lead to a rejection of a Markovian process. On the other hand, a dimension larger than 1.1 can be excluded if the test does not reject the model.

4. Refractory time

Finally, the Markov condition is violated by introducing a refractory time, i.e., a minimal time that the process has to spend in a state. To simulate such processes we used the

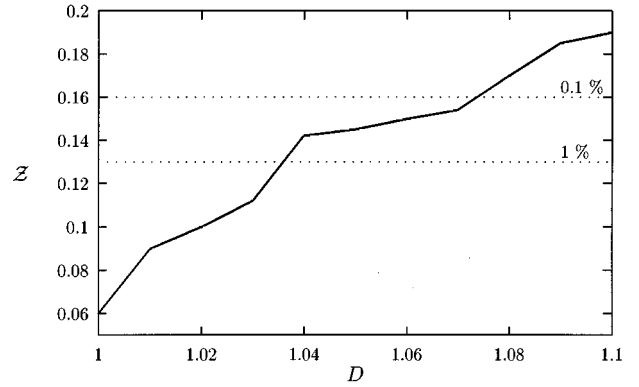


FIG. 4. Violation of the Markov condition by a fractal model. Shown is the averaged test statistic \mathcal{Z} for increasing fractal dimension D .

model according to Eq. (10) but forced the state to stay for the time τ_{ref} before the dynamics were applied. Figure 5 displays the results. Note that the dwell times of the chosen model were 10 and 12.5, respectively 20 units of time, so that the considered type of violation is only detectable if it amounts to 50% of the shortest dwell time. In summary, the test enables a detection of different types of violations of the Markov conditions.

C. Estimating the minimum number of states

So far, the number of states s of the Markov process was assumed to be known. Since the number of assumed states \hat{s} determines the degrees of freedom r of the model, the proposed test can be applied to infer the number of states of the process under investigation. This is done by comparing the left-hand side of Eq. (9) with the χ^2_r distribution with degrees of freedom \hat{r} corresponding to the assumed model, e.g., in the case of a Gaussian model: $\hat{r} = \hat{s}^2 + \hat{s}$. Figure 6 displays the results. Hidden Markov models with an increasing number of states are fitted to data from the model Eq. (10) with three states. The test enables a determination of the (correct) smallest number of states that can describe the process. Note that models with more than three states are also detected as being consistent with the data.

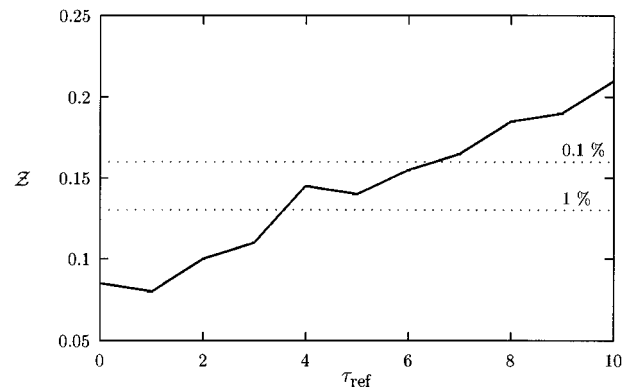


FIG. 5. The effect of refractory time. Shown is the averaged test statistic \mathcal{Z} for increasing refractory times τ_{ref} .

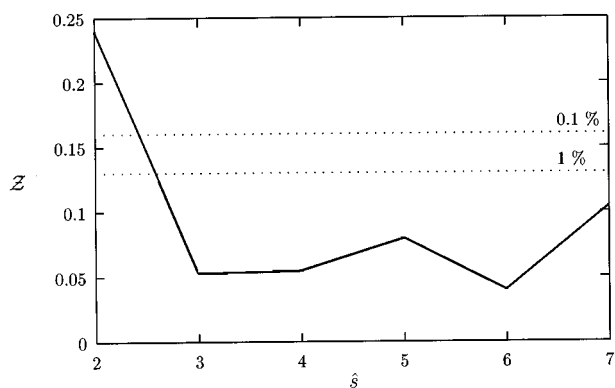


FIG. 6. Determining the number of states. Shown is averaged test statistic Z for hidden Markov models with different number of states \hat{s} applied to time series that were generated by a hidden Markov model with three states. The 1% and the 0.1% significance levels are marked.

V. DISCUSSION

Markov and hidden Markov models are increasingly used in the analysis of patch clamp ion channel data. In many cases their adequacy for a given system has been assumed, but not tested using empirical data. If a record is a realization of a hidden Markov process, the asymptotic distribution of the log-likelihood is a χ_r^2 distribution, its number of degrees of freedom r being given by the number of model parameters. Thus, a test for the consistency of the empirical distribution of a fitted model with the theoretical distribution provides a test whether the time series may be considered as a realization of a hidden Markov process.

Based on the asymptotic distribution of the log-likelihood, we have introduced such a test. The test is analogously applicable to test Markov models. In order to investigate how

sensitive the test is to detect a violation of the assumed model, we performed four simulation studies where we modified a hidden Markov process continuously in different ways to become non-Markovian. The sensitivity of the proposed test depends on how the model assumption of a stationary Markov process is violated: The test has shown to be very sensitive if the violation results from drifting transition probabilities, from dwell-time-dependent transition probabilities, or a fractal model. For example, a fractal dimension of 1.7 as reported in [15] would lead to a highly significant rejection of the Markov model used in the simulation study. The test is less sensitive to detect refractory times in the system that retard the beginning of the Markovian dynamics.

Furthermore, the proposed test can be used to estimate the minimum number of states in the Markov process necessary to describe the data.

In applications, performing simulation studies as presented will reveal which degree of violation of the model assumptions is consistent with the fitted model and which degrees of violation can be excluded if the model cannot be rejected.

The test is suited for analyzing data recorded under steady state conditions as in the case of ligand-dependent ion channels. For voltage-dependent channels where numerous trials for a certain pulse protocol are recorded these single trials determine the length M in the protested test. Further, it can help to decide whether observed changes in inactivation dynamics [31] are consistent with statistical fluctuations in a fitted model or have to be treated explicitly as modal gating between two different dynamics.

ACKNOWLEDGMENTS

We would like to thank A. Wilts and U.-P. Hansen for valuable comments on an earlier version of this manuscript.

-
- [1] E. Neher and B. Sakmann, *Nature* **260**, 799 (1976).
 [2] S. Kijima and H. Kijima, *J. Theor. Biol.* **128**, 423 (1987).
 [3] D. Colquhoun and F. J. Sigworth, in *Single Channel Recordings*, edited by B. Sakmann and E. Neher (Plenum Press, New York, 1995), p. 483.
 [4] R. Horn and K. Lange, *Biophys. J.* **43**, 207 (1983).
 [5] R. Horn and C. A. Vandenberg, *J. Gen. Physiol.* **84**, 505 (1984).
 [6] S. H. Chung, J. B. Moore, L. Xia, and L. S. Premkumar, and P. W. Gage, *Philos. Trans. R. Soc. London B* **329**, 265 (1990).
 [7] S. H. Chung, V. Krishnamurthy, and J. B. Moore, *Philos. Trans. R. Soc. London B* **334**, 357 (1991).
 [8] D. R. Fredkin and J. A. Rice, *Proc. R. Soc. London B* **249**, 125 (1992).
 [9] D. R. Fredkin and J. A. Rice, *Biometrics* **48**, 427 (1992).
 [10] J. D. Becker, J. Honerkamp, J. Hirsch, E. Schlatter, and R. Greger, *Pflug. Arc.* **224**, 328 (1994).
 [11] A. Albertsen and U. P. Hansen, *Biophys. J.* **67**, 1393 (1994).
 [12] S. J. Korn and R. Horn, *Biophys. J.* **54**, 871 (1988).
 [13] O. B. McManus, C. E. Spivak, A. L. Blatz, D. S. Weiss, and K. L. Magleby, *Biophys. J.* **55**, 383 (1989).
 [14] L. S. Liebovitch, *Biophys. J.* **55**, 373 (1989).
 [15] L. S. Liebovitch, J. Fischbarg, and J. P. Koniarek, *Math. Biosci.* **84**, 37 (1987).
 [16] P. Sansom, F. G. Ball, C. J. Kerry, R. McGee, R. L. Ransey, and P. N. P. Usherwood, *Biophys. J.* **56**, 1229 (1989).
 [17] L. E. Baum and T. Petrie, *Ann. Math. Stat.* **40**, 1554 (1966).
 [18] L. R. Rabiner, *IEEE Trans. Inf. Theory* **37**, 257 (1991).
 [19] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, *Ann. Math. Stat.* **41**, 164 (1970).
 [20] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, *Bell Syst. Tech. J.* **62**, 1035 (1983).
 [21] W. Qian and D. M. Titterton, *Philos. Trans. R. Soc. London A* **377**, 407 (1991).
 [22] A. P. Dempster, N. M. Laird, and D. R. Rubin, *J. Roy. Stat. Soc. B* **39**, 1 (1977).
 [23] D. R. Cox and D. V. Hinkley, *Theoretical Statistics* (Chapman and Hall, New York, 1974).
 [24] P. Billingsley, *Statistical Inference for Markov Processes* (University of Chicago Press, Chicago, 1961).

- [25] B. G. Leroux, *Stoch. Process. Applic.* **40**, 127 (1992).
[26] L. LeCam and G. Yang, *Asymptotics in Statistics. Some Basic Concepts* (Springer, New York, 1990).
[27] P. J. Bickel and Y. Ritov (unpublished).
[28] T. Ryden, *Ann. Stat.* **22**, 1884 (1994).
[29] T. Ryden (personal communication).
[30] L. Sachs, *Applied Statistics* (Springer, Berlin, 1984).
[31] C. C. Cannon and S. M. Strittmatter, *Neuron* **10**, 317 (1993).