

Estimation of Ion Channel Kinetics from Fluctuations of Macroscopic Currents

Luciano Moffatt

Instituto de Química Física de los Materiales, Medio Ambiente y Energía, Consejo Nacional de Investigaciones Científicas y Técnicas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

ABSTRACT For single channel recordings, the maximum likelihood estimation (MLE) of kinetic rates and conductance is well established. A direct extrapolation of this method to macroscopic currents is computationally prohibitive: it scales as a power of the number of channels. An approximated MLE that ignored the local time correlation of the data has been shown to provide estimates of the kinetic parameters. In this article, an improved approximated MLE that takes into account the local time correlation is proposed. This method estimates the channel kinetics using both the time course and the random fluctuations of the macroscopic current generated by a homogeneous population of ion channels under white noise. It allows arbitrary kinetic models and stimulation protocols. The application of the proposed algorithm to simulated data from a simple three-state model on nonstationary conditions showed reliable estimates of all the kinetic constants, the conductance and the number of channels, and reliable values for the standard error of those estimates. Compared to the previous approximated MLE, it reduces by a factor of 10 the amount of data needed to secure a given accuracy and it can even determine the kinetic rates in macroscopic stationary conditions.

INTRODUCTION

Ion channels are multimeric transmembrane proteins that regulate the passage of ions through the cellular membrane (1). They are able to couple their kinetics to an external factor (e.g., the concentration of an agonist, the membrane potential, tension, etc.). Since the changes they generate in the membrane potential propagate much faster than the diffusion of molecules (2), ion channels have a central role in the fast processing of information and in the fast coordination of the body. Thanks to the patch-clamp technique (3), it is possible to follow in real time the ion current that passes through a single channel. In that way it was observed that ion channels have at least two conformational states—one that allows the passage of ions (open) and another that does not (closed). Their change from one state to the other is faster than standard electrophysiological recording methods (3). After the continuous recording of several thousand openings and closings, it was found that the distribution of the durations of the openings and the closings was multimodal, suggesting that the open and closed states were aggregates of conformational states with different stability (4). Those facts can be interpreted quantitatively in terms of a Markov process, stochastic and memoryless (5,6). In a Markov process, the system is defined only by its present state and not by its past. At each interval of time, there is a probability of switching to a different state; this probability is dependent on the current state of the system and usually some external variable (like concentration of an agonist, membrane potential, etc.).

The statistical properties of single channel recordings have been studied with great detail (5–15). To find the kinetic

rates that best describe a set of experiments, the method of choice (11,13,15,16) is the maximum likelihood estimation (MLE). The likelihood function is defined as the probability of obtaining the data, conditional to a given value of the model parameters; and as such, it makes use of the average time course and the random fluctuations of the data. MLE uses standard multivariate optimization methods (17) to find the parameters for which likelihood is maximal.

For single channel recordings, the likelihood function is determined by the forward recursive algorithm of hidden Markov modeling (18), which is described in Theory and Algorithm. For the special case where the instrumental noise is much greater than the fluctuations in the number of channels, the logarithm of the likelihood function of the macroscopic currents generated by a big (i.e., >1000) number of channels is numerically homologous to the negative of the squared residuals sum; this approach makes use only of the time course of the data.

The construction of the likelihood function of macroscopic currents in the general case is done by extrapolating the single channels MLE. This extrapolation involves building a meta-channel in which each state of the meta-channel represents a particular distribution of states of the population of channels (11,15). The statistics of this meta-system is also Markovian and the likelihood function can be determined in the same way as in the single channel case. I label this approach microscopic-recursive (Micro R) to emphasize the fact that it distinguishes the discrete entity of each channel, and the fact that it uses a recursive algorithm to determine the likelihood. This approach provides exact values for the likelihood, but at the price of a computation cost that increases as the $2 \times (n_{\text{states}} - 1)$ power of the number of channels, which is computationally prohibitive even for a small number of channels. Therefore, the realistic alternative

Submitted November 15, 2006, and accepted for publication February 1, 2007.

Address reprint requests to Luciano Moffatt, Tel.: 54-11-4576-3378, ext. 130; E-mail: lmoffatt@qi.fcen.uba.ar.

Editor: Richard W. Aldrich.

© 2007 by the Biophysical Society

0006-3495/07/07/74/18 \$2.00

doi: 10.1529/biophysj.106.101212

is to find algorithms that provide approximate values of the likelihood.

For instance, a simplified likelihood function can be obtained by first approximating the distribution of states by a continuous distribution and then neglecting the statistical dependence of successive measurements, considering each sample independent of the others (19). We call this approach macroscopic-nonrecursive (Macro NR) to emphasize the fact that the discrete entity of the channels is blurred and the fact that the underlying algorithm is not recursive, i.e., the calculation of each predictive value does not depend on the previous ones. This approximation makes use of the variance of the random fluctuations of the data, but it does not make use of its local time correlation. It will be studied in detail in this article.

Another approximation is the covariance fitting (20), which fits both the magnitude of the recorded current and the strength of the correlation between different time points. It maximizes a general multivariate likelihood function (20). However, this method scales as the square of the number of samples and therefore is limited in the number of points it can use: it cannot analyze an entire data set, only a subset of it (20). In this article, I propose a new algorithm, macroscopic-recursive (Macro R), that both approximates the likelihood function of the whole data set and takes into account the local time correlation of the data. This algorithm provides better estimates of the kinetic parameters than the Macro NR. This method makes use of the recursive nature of the correlation between measurements and therefore it scales linearly with the number of samples.

This article is structured as follows. In Theory and Algorithm, I describe the three algorithms in the context of Markov kinetics, hidden Markov modeling, and Bayesian statistics. In Results, I provide numerical simulations to show the relative performance of Macro R and Macro NR methods in the tasks of 1), approximating the exact likelihood value provided by the Microscopic algorithm; 2), estimating the kinetic parameters, number of channels and conductance from simulated data; and 3), providing the covariance of the estimates. In Discussion, I will examine the scope of applicability of the new algorithm and future avenues of research.

THEORY AND ALGORITHM

In this section, I will first introduce the Markov approach as a simplified model of ion channel kinetics. Next, I will bring in a second element necessary to build a description of a multichannel system: the observer, with a brief comment on Bayesian and frequentist approaches. Although channels are independent from each other, statistical correlations between the measurements performed on them arise as a consequence of the observer being incapable of distinguishing the contributions of each individual channel (7). This fact makes the description of the Markov multichannel system more

complicated than the simple enumeration of the state of each channel (11,15). Then, I will attack the central problem of this article: determining the kinetic parameters from a set of electrophysiological recordings by using hidden Markov modeling (HMM) combined with MLE. After a brief introduction to HMM, I will describe the three algorithms studied in the present article for determining the likelihood function: Micro R (13,15), Macro NR (19), and Macro R (this study). Then the application of MLE for the kinetic parameters is straightforward by using standard quasi Newton maximizing algorithms (17). The Fisher information matrix provides reliable approximations of the covariance of the estimates.

Markov kinetic modeling of the behavior of a single channel

In the Markov approach to channel kinetics (5), the continuum of structural conformations of ion channels is partitioned on a finite number k of conformational states, $s(t) = s_i$, $1 \geq i \geq k$. Each conformational state may have different physicochemical properties; the easiest to measure is the conductance through the pore. The $k \times 1$ column vector γ with elements γ_i represents the amount of current that crosses through a channel that is currently in the state s_i . In the simplest kinetic scheme, there are two states: one that conducts ions (open) and another that does not (closed). More sophisticated schemes consider multiple closed states and multiple open states of the same or different conductance.

Channels are dynamic and change their state at random. Those conformational changes are usually modeled by the simplest stochastic mechanism: a Markov process. In a Markov process, at each time interval there is a constant probability that the channel will change from one state to another. This probability is dependent only on the current state of the channel, no matter how long the channel has been in the present state. It is the perfect memoryless system; the system instantaneously forgets all its past.

The probability of each one of the possible transitions between the different states that the channel might undergo is described by a different rate constant,

$$q_{ij} = \lim_{\Delta t \rightarrow 0} \frac{Pr(s(t + \Delta t) = s_j | s(t) = s_i)}{\Delta t}, \quad (1)$$

where q_{ij} is the probability rate that the channel will change from the state i to the state j .

Since Markov processes are stochastic, we need to introduce the use of probabilities for ion channel kinetics. We define then the $(1 \times k)$ state probability vector $\mathbf{p}(t) = (p_1(t), \dots, p_k(t))$ with state probabilities $p_i(t) = Pr(s(t) = s_i)$, which denotes at a particular time the probability of the channel being in each one of the k conformational states.

Here, I make a digression. There are two main approaches to probabilities. In the frequentist approach, we think about a population of channels evolving stochastically from a given

starting point; the state probability vector indicates the probability that a randomly chosen channel from an infinite population of channels is in a given state. This approach is useful for describing a system that is evolving by itself without any perturbations from the observer. In the Bayesian approach, the key concept is that the state probability vector is not a property of the channel itself (the channel is at a single conformational state at each time); it is a measure of the uncertainty that the observer has about the channel state. Therefore, we do not have to refer to a hypothetical population of channels, but to the channel we are studying and the information we have as observers of the channel. The advantage of doing that (see below) is that we can use the concept of the state probability vector in the deduction of the recursive likelihood.

The evolution of the state probability vector \mathbf{p} in the absence of further observations is thus determined by the Kolmogorov differential equation

$$\frac{\partial \mathbf{p}(t)}{\partial t} = \mathbf{p}(t) \cdot \mathbf{Q}, \quad (2)$$

where \mathbf{Q} is an $n \times n$ rate matrix, with off-diagonal elements q_{ij} being the rate constants of the transition between states $i \rightarrow j$. Each diagonal element q_{ii} is equal to $q_{ii} = -\sum_{j \neq i} q_{ij}$, the negative sum of the off-diagonal elements of the i^{th} row. In this way the sum of each row is zero, and therefore the sum of probabilities is kept constant at one. When there is no direct transition between states $i \rightarrow j$, $q_{ij} = 0$, this matrix \mathbf{Q} summarizes the kinetic properties of the channel itself.

For $t > 0$, the solution of Eq. 2 is given by the Chapman-Kolmogorov solution:

$$\mathbf{p}(t) = \mathbf{p}(0) \cdot \exp(\mathbf{Q}t). \quad (3)$$

This equation allows us to estimate the probability state vector as a function of time, given the fact that no more observations are made on the channel state. There are two ways to define the exponential of a matrix: in terms of a Taylor series,

$$\exp(\mathbf{Q}t) = \mathbf{I} + \mathbf{Q}t + \frac{\mathbf{Q}^2}{2!}t^2 + \frac{\mathbf{Q}^3}{3!}t^3 + \dots, \quad (4)$$

or by decomposing \mathbf{Q} into its vector of eigenvalues $\boldsymbol{\lambda}$ and matrix of eigenvectors \mathbf{V} ,

$$\exp(\mathbf{Q}t) = \exp(\mathbf{V} \cdot \boldsymbol{\lambda}^D t \cdot \mathbf{V}^{-1}) = \mathbf{V} \cdot \exp(\boldsymbol{\lambda}^D t) \cdot \mathbf{V}^{-1}, \quad (5)$$

where the notation $\boldsymbol{\lambda}^D$ indicates the diagonal operator, which builds a diagonal matrix with the vector $\boldsymbol{\lambda}$. The exponential of a diagonal matrix is equal to the diagonal of the exponential of each element.

Markov modeling of an ensemble of channels

In this section, I will define, for an ensemble of channels, the same elements of the Markov approach that I have defined for a single channel: the definition of the state of the system,

the codification of the information the observer has about the system, and the equations that describe the evolution of the system.

Approaches to model an ensemble of channels

When there are multiple channels in a patch under voltage-clamp conditions, the behavior of each channel is independent from the others. Therefore, the average behavior of a population of multiple channels is well described by Eq. 3, addressing single channels modeling. However, when repeated measurements are performed on a population of channels that are indistinguishable, a local time correlation appears and a more sophisticated modeling is needed to deal with it. There are two approaches to model a population of channels. The microscopic approach is the exact discrete description of the system. The macroscopic is a continuous approximation of the system.

Microscopic approach

In the microscopic description we assume, beforehand, the number of channels N . The state of the system is described by the microscopic state vector \mathbf{n} , a $1 \times k$ vector of positive integers listing the number of channels that are in each one of the k conformational states (15). The conductance of each ensemble state is $\mathbf{n} \cdot \boldsymbol{\gamma}$, the sum of the number of channels at each state weighted by its conductance.

To describe the probability vector, we first build the occupancy index matrix \mathbf{N} , an $M \times k$ matrix, whose rows $\{\mathbf{n}_i\}$ are each one of the

$$M = \binom{N+k-1}{N}$$

possible values that \mathbf{n} is allowed to assume. Associated with this matrix is the microscopic probability vector, \mathbf{p}^m , a $1 \times M$ vector that lists the probabilities for each of the possible ensemble states. The matrix product $\mathbf{p}^m \cdot \mathbf{N}$ gives the mean number of channels that are in each state.

The evolution of the vector m is described by an homology of Eq. 2 as

$$\frac{\partial \mathbf{p}^m(t)}{\partial t} = \mathbf{p}^m(t) \cdot \mathbf{Q}^m, \quad (6)$$

where the elements of \mathbf{Q}^m are the ensemble rate constants,

$$q_{vw}^m = \lim_{\Delta t \rightarrow 0} \frac{Pr(\mathbf{n}(t + \Delta t) = \mathbf{n}_w | \mathbf{n}(t) = \mathbf{n}_v)}{\Delta t}, \quad (7)$$

which are the rate constants for the transitions between the state \mathbf{n}_v to the state \mathbf{n}_w . Since, in the limit, the probability of more than one channel changing its state is zero, only a small number of ensemble transitions are different from zero—namely, the ones that involve the transition of only one channel:

$$q_{vw}^m = \begin{cases} n_{vi}q_{ij} & n_{wi} = n_{vi} - 1; \quad n_{wj} = n_{vj} + 1; \quad \sum_s^k \delta(n_{vs}, n_{ws}) = k - 2 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Here, n_{vi} indicates the number of channels in s_i when the $\mathbf{n} = \mathbf{n}_v$; $\delta(n, m) = 1$ if $n = m$, otherwise $\delta(n, m) = 0$. In this equation, \mathbf{n}_v is the starting ensemble state and \mathbf{n}_w is the ending ensemble state. If the only difference between \mathbf{n}_w and \mathbf{n}_v is that there is one more channel in s_j and one less channel in s_i , the ensemble rate constant is equal to the rate constant of the only transition that occurred (q_{ij}) multiplied by the original number of channels that were in state s_i (n_{vi}). Otherwise the ensemble rate constant is zero. Diagonal values of \mathbf{Q}^m are calculated as the negative sum of the off-diagonal elements of the i^{th} row. Equation 6 is solved using the Chapman-Kolmogorov solution (Eq. 3).

Macroscopic approach

In the continuous macroscopic approximation, we describe the state of the system by the macroscopic state vector, $\mathbf{r} = \mathbf{n}/N$, a $1 \times k$ vector of real numbers listing the ratio of occupancies of each one of the states (19). Note that we do not refer to a number of channels but to a proportion of channels in each state, the proportion can be any real number between 0 and 1. To calculate the conductance of the ensemble of channels, we use the expression $N\mathbf{r} \cdot \boldsymbol{\gamma}$ —the average conductance multiplied by the number of channels.

The distribution of \mathbf{r} can be approximated by a multivariate Normal,

$$Pr(\mathbf{r}_t) = N\left(\mathbf{r} - \boldsymbol{\mu}_t, \frac{1}{N}\boldsymbol{\Sigma}_t\right) = N^{k/2}(2\pi)^{-k/2}|\boldsymbol{\Sigma}_t|^{-1/2} \exp\left(-\frac{N}{2}(\mathbf{r} - \boldsymbol{\mu}_t) \cdot \boldsymbol{\Sigma}_t^{-1} \cdot (\mathbf{r} - \boldsymbol{\mu}_t)^T\right), \quad (9)$$

where T is the transpose operator and we denote the mean vector and the dispersion matrix by $E(\mathbf{r}_t) = \boldsymbol{\mu}_t$ and $\text{var}(\mathbf{r}_t) = E((\mathbf{r}_t - \boldsymbol{\mu}_t)^T \cdot (\mathbf{r}_t - \boldsymbol{\mu}_t)) = \boldsymbol{\Sigma}_t/N$. In this way, the number of parameters that describes the distribution of r is only $(k^2 - k)/2$.

Classical hidden Markov modeling

The hidden Markov modeling (HMM) approach deals with the problem of using the evolution of an observable signal to guess the state of an unobservable Markov system (18). Originally developed in 1960s (21,22), it has been successfully used for single channel kinetics analyses in the 1990s (14,23), where the observable signal is the current that passes through the channel.

HMM distinguishes (18) four elements: 1), a Markov model that describes the evolution rules of the hidden system; 2), an observational model that describes the relationship between the hidden state and the observable signal; 3), a se-

quence of hidden states \mathbf{s}_t ; and 4), a sequence of observations \mathbf{y}_t .

A given Markov model is defined by the transition probabilities between the hypothetical states. The key property that defines a process as a Markov model is its absence of memory, i.e., the fact that its future evolution only depends on its present state:

$$Pr(s_{t+1} = s_j | s_t = s_i) = \lambda(s_i, s_j, \boldsymbol{\theta}) = a_{ij}. \quad (10)$$

It is important to differentiate between what constitutes the architecture of the Markov model, λ , and the numerical parameters of the model, $\boldsymbol{\theta}$. The architecture of the model refers to the way it is built, i.e., the list of states, the allowed transitions between them, and the equations that map the model parameters onto the transition probabilities between states. Usually the architecture of the model is assumed, whereas the parameters are extracted from the data. However, it is possible to compare different architectures in their ability to predict the data.

In a similar way, the architecture of the observational model, κ , refers to the equations that relate the observational parameters, $\boldsymbol{\varepsilon}$, with the probability of observing a given value of a signal given a specific state of the system:

$$Pr(y_t = y | s_t = s_i) = \kappa(y, s_i, \boldsymbol{\varepsilon}). \quad (11)$$

Two estimates of usual interest in HMM are

1. The likelihood function, the probability that the model Λ (short for λ and κ) with parameters $\boldsymbol{\Theta}$ ($\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$) might produce the signal y :

$$\ell(\boldsymbol{\Theta}) = P(y_1, \dots, y_t | \boldsymbol{\Theta}, \Lambda). \quad (12)$$

Note that this probability is independent of the actual evolution of the system. The probabilities are summed up over all the possible evolutions.

2. The maximum likelihood estimates of the parameters $\boldsymbol{\Theta}$, given the model Λ :

$$\boldsymbol{\Theta}^{\text{MLE}} = \arg \max_{\boldsymbol{\Theta}} P(y_1, \dots, y_t | \boldsymbol{\Theta}, \Lambda). \quad (13)$$

This is the value of the parameters that produces the maximum likelihood.

A third estimate, offered below, is beyond the usual scope of HMM, and will not be treated in this article.

3. The most likely model Λ , independent of the actual parameters:

$$\Lambda^{\text{MLE}} = \arg \max_{\Lambda} Pr(y_1, \dots, y_t | \Lambda). \quad (14)$$

Likelihood function and forward algorithm

To obtain the likelihood of observing the signal y given a model Λ , we use the observational model κ to determine the relationship between the observed signal and the state of the system and the Markov model λ to determine the probability of the states' sequences. A conceptually simple way to determine the likelihood of the Markov process is to sum over all the possible sequences of states (18), the product of the probability of obtaining the signal given a sequence of states by the probability of obtaining that particular sequence of states:

$$Pr(y_1, \dots, y_t | \Theta) = \sum_s Pr(y_1, \dots, y_t | s_1, \dots, s_t, \epsilon) Pr(s_1, \dots, s_t | \Theta). \quad (15)$$

This procedure is not practical, given the exponentially growing number of the possible sequences of states. A much better approach takes advantage of the fact that Markov processes are memoryless by using a recursive algorithm (18). The initialization of the algorithm is provided by an a priori estimate of the state probability vector based on equilibrium considerations:

$$\alpha_0(s_i) = Pr(s_0 = s_i, \Theta). \quad (16)$$

The i^{th} value is defined as

$$\alpha_t(s_i) = Pr(y_1 \dots y_t, s_t = s_i | \Theta), \quad (17)$$

the induction is made by

$$\alpha_{t+1}(s_j) = Pr(y_{t+1} | s_{t+1} = s_j, \epsilon^2) \times \sum_i Pr(s_{t+1} = s_j | s_t = s_i, \Theta) \alpha_t(s_i), \quad (18)$$

and the termination is the sum

$$\ell(\Theta, y_1, \dots, y_T) = Pr(y_1, \dots, y_T | \Theta) = \sum_i \alpha_T(s_i). \quad (19)$$

Using Eqs. 10 and 11 and assuming Gaussian noise, the induction step is

$$\alpha_{t+1}(s_j) = N(y_{t+1} - \gamma_j, \epsilon^2) \sum_i a_{ij} \alpha_t(s_i), \quad (20)$$

where N indicates the Normal probability density function and a_{ij} are the elements of $\mathbf{A} = \exp(Q\Delta t)$.

Bayesian formulation of the forward algorithm

It is insightful to describe the forward algorithm in Bayesian terms. The formulation is also extended for a continuous state set. The likelihood of the joint measurement $y_1 \dots y_T$ can be decomposed as the product of the incremental likelihoods due to each measurement y_i :

$$\ell_T = \ell(\Theta | y_1 \dots y_T) = \prod_{t=1}^T \ell'_t, \quad (21)$$

$$\ell'_t = Pr(y_t | y_1 \dots y_{t-1}, \Theta). \quad (22)$$

We define the prior probability vector as the information the observer has about the system before each measurement by the equation

$$p_t^*(s_i) = Pr(s_t = s_i | y_1 \dots y_{t-1}, \Theta). \quad (23)$$

If we have a continuous space, the prior probability function is defined as

$$p_t^*(\mathbf{r}) = Pr(s_t = \mathbf{r} | y_1 \dots y_{t-1}, \Theta), \quad (24)$$

where \mathbf{r} is a macroscopic state vector. After we perform the measurement, we can calculate the likelihood of this particular measurement by multiplying the conditional probability of the measurement by the prior probability vector

$$\ell'_t = \sum_i Pr(y_t | s_t = s_i, \epsilon^2) \cdot p_t^*(s_i). \quad (25)$$

If we use the continuous approximation, we multiply by the prior probability function and we integrate over the entire state probability space:

$$\ell'_t = \int_s Pr(y_t | s_t = \mathbf{r}, \epsilon^2) \cdot p_t^*(\mathbf{r}) \cdot d\mathbf{r}. \quad (26)$$

The posterior probability vector is the information the observer has about the state of the system updated after the measurement y_t is made:

$$p_t(s_i) = Pr(s_t = s_i | y_1 \dots y_t, \Theta) = \frac{Pr(y_t | s_t = s_i, \epsilon^2)}{\ell'_t} \cdot p_t^*(s_i). \quad (27)$$

The posterior probability function is

$$p_t(\mathbf{r}) = Pr(s_t = \mathbf{r} | y_1 \dots y_t, \Theta) = \frac{Pr(y_t | s_t = \mathbf{r}, \epsilon^2)}{\ell'_t} p_t^*(\mathbf{r}). \quad (28)$$

Now we use the Markov transition to find the state of the system at time $t + \Delta t$. For the discrete case,

$$p_{t+\Delta t}^*(s_j) = \sum_i Pr(s_{t+\Delta t} = s_j | s_t = s_i, \Theta) p_t(s_i). \quad (29)$$

If we use the continuous approximation, we integrate over the entire state probability space:

$$p_{t+\Delta t}^*(\mathbf{u}) = \int_s Pr(s_{t+\Delta t} = \mathbf{u} | s_t = \mathbf{r}, \Theta) p_t(\mathbf{r}) d\mathbf{r}. \quad (30)$$

In this way, the Bayesian formulation of the forward algorithm provides us with three meaningful variables: the partial likelihood, the prior probability, and the posterior probability vector or function. The partial likelihood indicates the probability of each particular measurement conditional to the previous ones. The prior and posterior probabilities tell us what we know about the state probability function just before and after each measurement. Partial likelihoods are statistically independent by construction, a very desirable statistical property that can be used for multiple purposes. One of them is to find an estimate of the Fisher information matrix.

Likelihood function of multichannel preparations

Nonrecursive macroscopic algorithm

This algorithm estimates the likelihood of the data neglecting the local time correlation:

$$\ell(\Theta, y_1, \dots, y_T) \approx \prod_{t=1}^T \text{Pr}(y_t | \Theta). \quad (31)$$

The likelihood of the joint measurements y_1, \dots, y_T is approximated by the product of the likelihood of each measurement. Although this approximation can substantially underestimate the actual likelihood of the data if there is a strong local time correlation, it is also computationally inexpensive and it has been proved to provide reliable estimates of the kinetic rates, number of channels and, with a substantial amount of data, of the conductance (19). The likelihood of each measurement is calculated by a direct comparison between the actual data y_t and the expected mean evolution of the data, y_t^{fit}

$$\begin{aligned} \ell(\Theta, y_t) &= N(y_t - y_t^{\text{fit}}, \sigma_t^2) \\ &= \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(y_t - y_t^{\text{fit}})^2}{2\sigma_t^2}\right), \end{aligned} \quad (32)$$

$$y_t^{\text{fit}} = N\mu_t \cdot \gamma, \quad (33)$$

$$\sigma_t^2 = N\gamma^T \cdot (\mu_t^D - \mu_t^T \cdot \mu_t) \cdot \gamma, \quad (34)$$

$$\mu_t = \mu_0 \cdot \exp(\mathbf{Q}t), \quad (35)$$

$$\mu_0 = \mathbf{u}_{1 \times k} \cdot (\mathbf{Q} + \mathbf{U}_{k \times k})^{-1}, \quad (36)$$

where μ_t represents the expectation of \mathbf{r} ; the macroscopic state vector at time t ; and $\mathbf{u}_{1 \times k}$ and $\mathbf{U}_{k \times k}$ are a unit vector and a unit matrix.

Microscopic recursive algorithm

In the microscopic recursive algorithm, the forward algorithm is applied straightforward.

The initial prior state probability vector \mathbf{p}^{m*} can be found by the formula,

$$\mathbf{p}_0^{m*} = \mathbf{u}_{1 \times M} \cdot (\mathbf{Q}^m + \mathbf{U}_{M \times M})^{-1}, \quad (37)$$

where \mathbf{Q}^m is the transition-rate probability matrix and $\mathbf{u}_{1 \times M}$ is a $1 \times M$ vector of ones and $\mathbf{U}_{M \times M}$ is a $M \times M$ matrix of ones.

The incremental likelihood is found summing over all the elements of the prior probability vector

$$\ell_t' = \sum_j^M N(y_t - \mathbf{n}_j \cdot \gamma, \varepsilon^2) p_{j*}^m, \quad (38)$$

where \mathbf{n}_j is the j^{th} row of \mathbf{N} , the occupancy index matrix.

Each element of the posterior probability vector is found by the equation

$$p_{jt}^m = \frac{N(y_t - \mathbf{n}_j \cdot \gamma, \varepsilon^2)}{\ell_t'} p_{j*}^m. \quad (39)$$

The prior probability at time $t + \Delta t$ is obtained by the vector product

$$\mathbf{p}_{t+\Delta t}^{m*} = \mathbf{p}_t^m \cdot \exp(\mathbf{Q}^m \Delta t). \quad (40)$$

Macroscopic recursive algorithm

The initial distribution of states is obtained from equilibrium conditions, i.e., when

$$\text{Pr}(\mathbf{r}, 0) = N\left(\mathbf{r} - \mu_0^*, \frac{1}{N} \Sigma_0^*\right), \quad (41)$$

$$\mu_0^* = \mathbf{u}_{1 \times k} \cdot (\mathbf{Q} + \mathbf{U}_{k \times k})^{-1}, \quad (42)$$

$$\Sigma_0^* = \mu_0^{*D} - \mu_0^{*T} \cdot \mu_0^*. \quad (43)$$

The incremental likelihood is obtained by integrating the product of the prior and the measurement probability (Eq. 26)

$$\ell_t' = \int_s N(y_t - N\mathbf{r}_t \cdot \gamma, \varepsilon^2) N\left(\mathbf{r}_t - \mu_t^*, \frac{1}{N} \Sigma_t^*\right) d\mathbf{r}_t, \quad (44)$$

where s indicates that the integral is made over all the possible states of \mathbf{r} . The result of this integral is

$$\ell_t' = N(y_t - y_t^*, \sigma_t^2), \quad (45)$$

where

$$y_t^* = N\mu_t^* \cdot \gamma, \quad (46)$$

$$\sigma_t^2 = \varepsilon^2 + N\gamma^T \cdot \Sigma_t^* \cdot \gamma. \quad (47)$$

Each posterior is obtained by normalizing the product of the prior and the measurement probability (Eq. 28).

$$p_t(\mathbf{r}) = \frac{N(y_t - N\mathbf{r}_t \cdot \gamma, \varepsilon^2)}{\ell_t'} N\left(\mathbf{r}_t - \mu_t^*, \frac{1}{N} \Sigma_t^*\right). \quad (48)$$

The result of this product is

$$p_t(\mathbf{r}) = N\left(\mathbf{r}_t - \mu_t, \frac{1}{N} \Sigma_t\right), \quad (49)$$

where

$$\mu_t = \mu_t^* + \frac{(y_t - y_t^*)}{\sigma_t^2} \cdot \gamma^T \cdot \Sigma_t^*, \quad (50)$$

$$\Sigma_t = \Sigma_t^* - \frac{N}{\sigma_t^2} \Sigma_t^* \cdot \gamma \cdot \gamma^T \cdot \Sigma_t^*. \quad (51)$$

Let $\mathbf{A} = \exp(\mathbf{Q} \Delta t)$, from the Markovian model, we know that

$$E(\mathbf{r}_{t+\Delta t} | y_1, \dots, y_t) = \mu_t \cdot \mathbf{A} \quad (52)$$

and from the multinomial distribution,

$$\text{var}(\mathbf{r}_{t+\Delta t}|y_1 \dots y_t, \mathbf{r}_t) = ((\mathbf{r}_t \cdot \mathbf{A})^D - \mathbf{A}^T \cdot \mathbf{r}_t^D \cdot \mathbf{A})/N. \quad (53)$$

Then, from well-known results on conditional expectation, we can calculate the prior probability at time $t + \Delta t$,

$$\begin{aligned} \boldsymbol{\mu}_{t+\Delta t}^* &= E(\mathbf{r}_{t+\Delta t}|y_1 \dots y_t) = E\{E(\mathbf{r}_{t+\Delta t}|y_1 \dots y_t, \mathbf{r}_t)\} \\ &= E(\mathbf{r}_t \cdot \mathbf{A}), \end{aligned}$$

i.e.,

$$\boldsymbol{\mu}_{t+\Delta t}^* = \boldsymbol{\mu}_t \cdot \mathbf{A}, \quad (54)$$

and

$$\begin{aligned} \text{var}(\mathbf{r}_{t+\Delta t}|y_1 \dots y_t) &= \Sigma_{t+\Delta t}^*/N = E\{\text{var}(\mathbf{r}_{t+\Delta t}|y_1 \dots y_t, \mathbf{r}_t)\} + \text{var}\{E(\mathbf{r}_{t+\Delta t}|y_1 \dots y_t, \mathbf{r}_t)\} \\ &= E\{((\mathbf{r}_t \cdot \mathbf{A})^D - \mathbf{A}^T \cdot \mathbf{r}_t^D \cdot \mathbf{A})/N|y_1 \dots y_t\} + \text{var}\{\mathbf{r}_t \cdot \mathbf{A}|y_1 \dots y_t\} \\ &= ((\boldsymbol{\mu}_t \cdot \mathbf{A})^D - \mathbf{A}^T \cdot \boldsymbol{\mu}_t^D \cdot \mathbf{A})/N + \mathbf{A}^T \cdot (\Sigma_t/N) \cdot \mathbf{A}, \end{aligned}$$

i.e.,

$$\Sigma_{t+\Delta t}^* = (\boldsymbol{\mu}_t \cdot \mathbf{A})^D - \mathbf{A}^T \cdot \boldsymbol{\mu}_t^D \cdot \mathbf{A} + \mathbf{A}^T \cdot \Sigma_t \cdot \mathbf{A}. \quad (55)$$

Maximum likelihood estimation of the kinetic parameters

The maximum likelihood estimation is a method for parameter fitting that can be used if we have a model that provides the likelihood of a given set of kinetic parameters and observations. It finds the set of parameters that maximizes the probability of obtaining the results. However, instead of maximizing likelihood, what is usually done is to minimize the negative of the logarithm of the likelihood (LL):

$$\boldsymbol{\theta}^{\text{MLE}} = \arg \min_{\boldsymbol{\theta}} (-LL) = \arg \min_{\boldsymbol{\theta}} (-\log(\ell(\Lambda, y_1 \dots y_T))). \quad (56)$$

To find the point where the negative of LL is minimal, we use a quasi-Newton algorithm. The idea of those methods is that the log-likelihood function is well approximated near its maximum by a second-order Taylor series,

$$\begin{aligned} -LL(\boldsymbol{\theta}) &\approx -LL(\boldsymbol{\theta}_0) + \mathbf{G}(\boldsymbol{\theta}_0)^T \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ &\quad + \frac{1}{2} \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \cdot \mathbf{H}(\boldsymbol{\theta}_0) \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \end{aligned} \quad (57)$$

where \mathbf{G} and \mathbf{H} are the Gradient vector of first derivatives and Hessian matrix of second derivatives of the negative of the log-likelihood respect to the parameters vector. The difference with the optimal point and the optimal log-likelihood would be approximated by

$$\boldsymbol{\theta}^* - \boldsymbol{\theta}_0 \approx -\mathbf{H}(\boldsymbol{\theta}_0)^{-1} \cdot \mathbf{G}(\boldsymbol{\theta}_0), \quad (58)$$

$$LL^* - LL_0 \approx \frac{1}{2} \mathbf{G}(\boldsymbol{\theta}_0)^T \cdot \mathbf{H}(\boldsymbol{\theta}_0)^{-1} \cdot \mathbf{G}(\boldsymbol{\theta}_0). \quad (59)$$

As calculating \mathbf{H} is computationally expensive, we approximate it by using the observed behavior of $LL(\boldsymbol{\theta})$ to build up curvature information to make an approximation to \mathbf{H} using an appropriate updating technique. Here we used the formula of Broyden, Fletcher, Goldfarb, and Shanno (BFGS) (17),

$$\mathbf{H}_{i+1} = \mathbf{H}_i + \frac{\Delta \mathbf{G} \cdot \Delta \mathbf{G}^T}{\Delta \mathbf{G}^T \cdot \Delta \boldsymbol{\theta}} - \frac{\mathbf{H}_i^T \cdot \Delta \boldsymbol{\theta}^T \cdot \Delta \boldsymbol{\theta} \cdot \mathbf{H}}{\Delta \boldsymbol{\theta}^T \cdot \mathbf{H} \cdot \Delta \boldsymbol{\theta}}, \quad (60)$$

where $\Delta \boldsymbol{\theta} = \boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}_i$.

If the quadratic form is a good approximation, the likelihood tends to a multivariate Gaussian

$$\ell(\boldsymbol{\theta}) \approx C \cdot \exp\left(\frac{1}{2}((\boldsymbol{\theta} - \boldsymbol{\theta}^*)^T \cdot \mathbf{H} \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}^*))\right), \quad (61)$$

where C is a constant. Therefore the Hessian matrix estimates the covariance matrix of the fitted parameters, $\text{cov}(\boldsymbol{\theta}) = -\mathbf{H}^{-1}$. Another way to estimate the covariance matrix is to use the Fisher information matrix (**FIM**), the variance of the Gradient of the log-likelihood function. Under certain conditions, **FIM** estimates the mean of the Hessian. **FIM** can be obtained from the gradient of the logarithm of the partial likelihood of each sample (LL'_i):

$$\mathbf{FIM} = E(\nabla LL^T \cdot \nabla LL) = \sum_i^{\text{nsamples}} (\nabla LL'_i{}^T \cdot \nabla LL'_i). \quad (62)$$

Simulations

Stochastic simulations were generated using the ligand-gated kinetic model of two closed states joined by an open state (Fig. 1) after a previous study (19) to allow an easy comparison. The instrumental noise was set to 1 pA at 50 KHz as well as the unitary conductance. No open channel noise was included in the model. The parameter values used for the simulation were called the true values as opposed to estimated values. Nonstationary and stationary conditions were simulated. In stationary conditions ligand concentration was held constant at 1 M and the initial states of the channels were randomly chosen from the equilibrium conditions. In nonstationary conditions, the concentration was also constant at 1 M, but at the beginning of the simulation all the channels were unliganded and closed (C_1). Two set of simulations were run. In the first, I compared the values of the likelihood function provided by the Macro NR, Macro R, and Micro R algorithms for 20 channels. In the second, I compared only Macro NR and Macro R algorithms in their MLE for 100 channels.

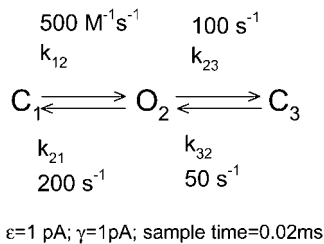


FIGURE 1 Kinetic model used in this article. It has a single binding step (k_{12}) and a “desensitized” state (C_3). It is taken from a previous work (19) to make their comparison easier.

RESULTS

Recursive and nonrecursive algorithm performances

The main point of this article is to present a new algorithm that approximates the likelihood of macroscopic currents. This new algorithm, the macroscopic recursive (Macro R), was compared against two other algorithms: one that provides the exact value of the likelihood, the microscopic recursive (Micro R) (15); and one that approximates it, the macroscopic nonrecursive (Macro NR) (19).

In the Macro NR approach, the log-likelihood is calculated by summing up the log-likelihood of each measurement, as if each sample point were an independent event. Both Micro R and Macro R, on the other hand, calculate the likelihood by summing up the incremental log-likelihood of each sample point. The Macro R method approximates LL , assuming that the state probability vector follows a multinormal distribution and the likelihood of each sample is obtained from the difference between the measured current and a predicted current. In the Microscopic method, all the possible combination of channels states are considered independently and the direct likelihoods of each one are summed up weighted by its prior probability; in this way the true value of the likelihood is obtained (true under the assumptions of the observational and Markov model). The microscopic method is computationally prohibitive for maximum likelihood estimation; it is included here as a “gold standard”. There is a fourth algorithm, covariance fitting (20), which calculates the likelihood of a subsample of the trace. This method analyzes the information available on the subsample by calculating the likelihood against a multivariate normal distribution. This method was not analyzed here.

In Fig. 2 the two approximated macroscopic algorithms are compared with the exact microscopic algorithm in their ability to provide the likelihood function of the 40-ms simulated trace generated by 20 channels at a sampling rate of 50 KHz. The Macro R algorithm needed ~ 16 ms to compute on a Pentium V the approximated value of the likelihood function of the simulated trace while the Macro NR needed only 5.6 ms. To determine an exact value, the computing time used by the Micro R algorithm is two orders-of-magnitude higher: 2730

ms. Software routines for the three algorithms are available on request from the author. Based only on the provided parameters, the Macro NR algorithm makes a prediction about the entire temporal course of the measured signal (Fig. 2 A). Both recursive algorithms, on the other hand, adjust their prediction of each measurement, taking into account the previous one (Fig. 2, B and C). As a consequence of that, recursive algorithms follow much more closely the simulated measurements (Fig. 2, B and C, compare *solid* and *shaded* lines) whereas the nonrecursive algorithm only follows the general trend (Fig. 2 A). Therefore, whenever there are more channels open than the expected, they will continue to be for some time, since the channels do not equilibrate instantaneously. As a consequence of that, a local time correlation will be present on the residual variation remaining in the data after the response expected by the nonrecursive algorithm has been subtracted (Fig. 2 D). This is something that can be seen as a clear low frequency component in the fast Fourier transform of the residuals (Fig. 2 M). Recursive algorithms completely abolish that time correlation (Fig. 2, E, F, N, and O). Another important improvement is the fact that the expected standard deviation of the measurements is much lower in the recursive algorithms (Fig. 2, H and I versus G) and, therefore, the likelihood is greater (Fig. 2, K and L versus J). This is because nonrecursive algorithms do not consider distribution of states previous to each measurement. In this way, whenever there is a random departure from the expected number of channels, this departure is considered many times for the likelihood calculation until it eventually relaxes. In the case of recursive algorithms, each random departure is considered only once.

If we look then into the deduced evolution of the prior probability ratios (Fig. 3, μ_1, μ_2, μ_3), we see that Macro NR follows the general trend, whereas the Macro R is almost undistinguishable from the Micro R. When we compare the evolution of the elements of the covariance matrix (Fig. 3, $\Sigma_{11}-\Sigma_{33}$), we see that the Macro NR differs greatly and that the Macro R follows the general trend of the Micro R, but there is a high frequency component of the Micro R that is not followed by the Macro R approximation. This reminds us of the fact that the value of each measurement is present in the actualization of the mean (Eq. 50), but not in the actualization of the covariance matrix (Eq. 51).

Log-likelihood distribution

The three algorithms provide a value for the likelihood. However, if we run several stochastic simulations of the data with the same parameters, we will find different values for the likelihood. In this way we can construct the distribution of the likelihood. What is the expected distribution of the log-likelihood? By the central limit theorem, the distribution of the log-likelihood should be approximately normal and the variance should be approximately the sum of the expected variance of the partial log-likelihoods. As the variance of the log-likelihood of a normal distribution is just

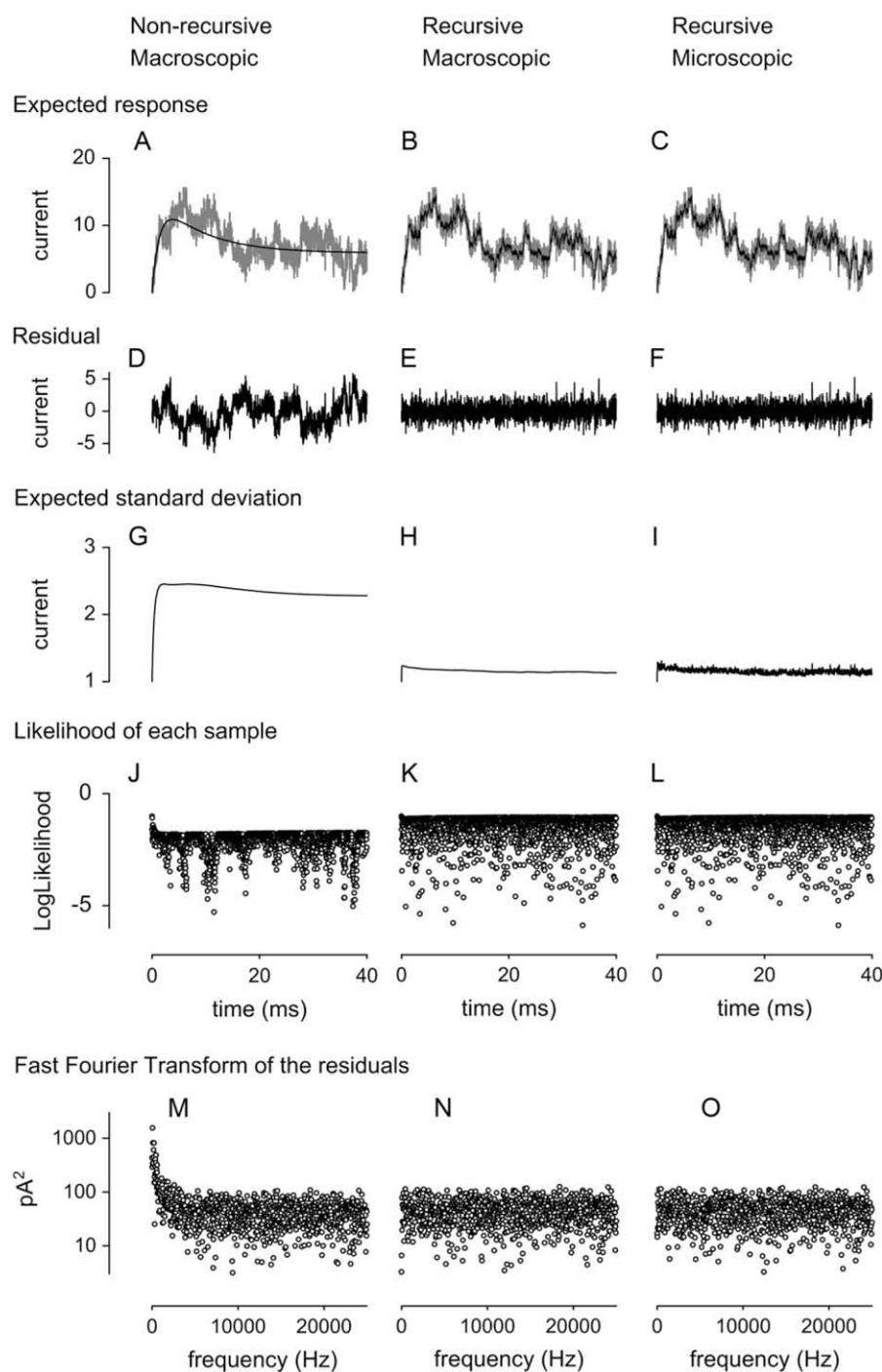


FIGURE 2 Comparison of macroscopic-nonrecursive, macroscopic-recursive, and microscopic-recursive algorithms in their ability to predict the response of the 20 simulated channels during 40 ms at 50 KHz (2000 samples). (A–C) Simulated measured response trace (*shaded*) and response expected by each algorithm (*solid representation*). Nonrecursive algorithm makes its prediction based only on the parameters value; it ignores the data. Recursive algorithms adjust the prediction of each data point based on the previously measured one. (D–F) Residual variation resulting from the deduction of the expected response. There is a time correlation in nonrecursive that is absent in recursive algorithms. (G,H) Each algorithm predicts also the standard deviation of the residuals (i.e., it provides an estimate of the square of the residuals). Note that macroscopic nonrecursive predicts a greater standard deviation and that microscopic recursive predicted a noisier trace. (J–L) Each algorithm also calculates an incremental likelihood at each sample point. Note that the lower values in the macroscopic nonrecursive algorithm. (M–O). The Fourier transform of the residuals shows a clear low frequency component in nonrecursive.

one-half, then the expected variance of the log-likelihood is just half the number of samples. I found that the cumulative distribution of the log-likelihood was closely matched by a normal distribution with a variance of $n_{\text{samples}}/2$ in the case of the recursive algorithms. In the case of the Macro NR algorithm the variance of the log-likelihood was $\sim 73\times$ higher than $n_{\text{samples}}/2$, suggesting that the “effective” sample size might be $73\times$ smaller.

Macroscopic recursive converges to microscopic recursive as the number of channels increases

Although small, there was a significant difference in the log-likelihood calculated by the microscopic and macroscopic recursive algorithms. To find out how this difference scales with the number of channels, we numerically measured the difference in the log-likelihood as the number of channels increases. The difference between microscopic and macroscopic

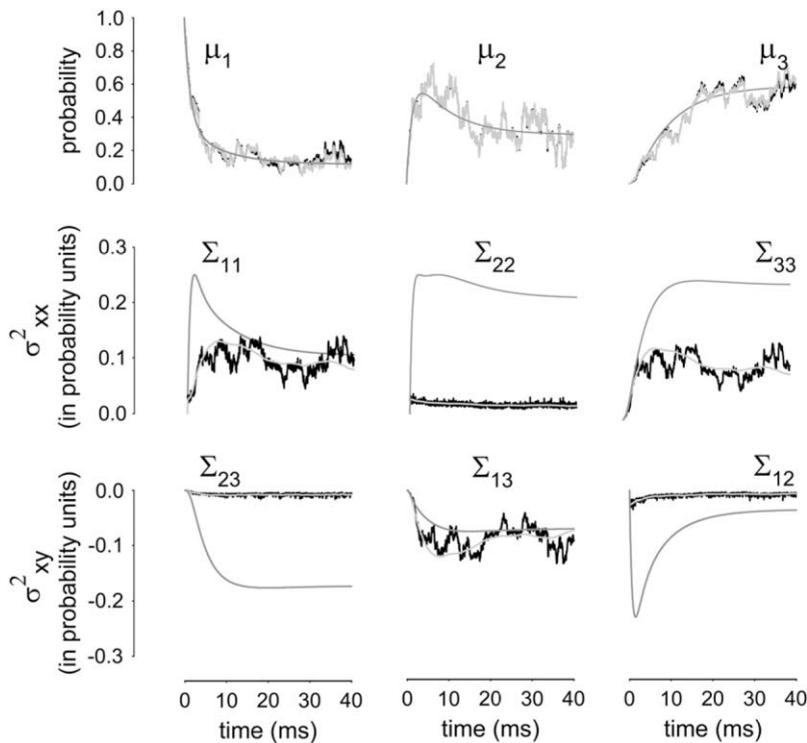


FIGURE 3 Comparison of the macroscopic nonrecursive (dark shaded), macroscopic recursive (shaded), and microscopic recursive (solid) algorithms in the evolution of the components of the mean and covariance of the state probability function. Data from the simulations analyzed on Fig. 2. The macroscopic nonrecursive differed strongly from the recursive algorithms. Both recursive algorithms were quite similar in the expected mean and they differed in the expected variance, the microscopic recursive showing much more variation than the macroscopic recursive.

decreases as the number of channels increases (Fig. 4) in an inverse relationship: it is reduced by half after each doubling in the number of channels. For 100 channels the difference is expected to be approximately two likelihood units per trace.

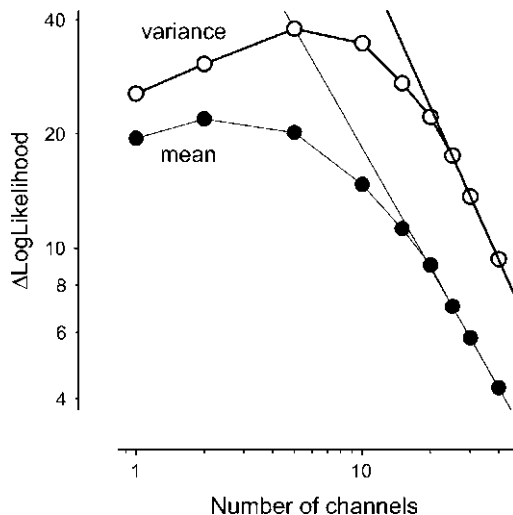


FIGURE 4 Macroscopic likelihood underestimation decreases with the number of channels. The log-log plot represents the mean and variance of the difference in the likelihood between the approximated macroscopic-recursive method and the exact microscopic-recursive method. The likelihood was calculated for single 40-ms, 50-KHz traces as a function of the number of channels. The lines indicate the linear regression of the last three points. The slope values were -1.05 for the mean and -1.33 for the variance. The variance was $2.4\times$ greater than the mean for 40 channels.

That means that, for 1000 traces, the difference would be ~ 2000 likelihood units.

Application of an algorithm for maximum likelihood estimation

The whole purpose of finding the log-likelihood function is to use it to feed a nonlinear maximization algorithm to find the optimal kinetic parameters that define the ion channel behavior. Since we want to analyze only the conditions where the approximations made by the macroscopic algorithms are expected to be valid, we increase the number of channels to 100 and we use the likelihood function of sets of 1000 independent traces. The algorithm was fed with a random perturbation of the true parameter values. The values of the following parameters were optimized: N , the number of channels, γ , the unitary conductance, and the kinetic rates k_{12} , k_{21} , k_{23} , and k_{32} . The value of ε , the instrumental error, was not optimized; its true value was used for the likelihood calculations. In real experiments, the instrumental error can be obtained from traces where there are no openings.

In Fig. 6 we show an example of the evolution of the Quasi Newton minimization algorithm under those conditions using the Macro R estimation of the likelihood function on non-stationary conditions. At the beginning of each step, the log-likelihood is calculated (Fig. 5 A, black line) for a given set of values of the parameters being optimized (Fig. 5 C; each parameter is color-coded). Then the gradient is calculated by finite differentiating (Fig. 5 B, color codes are the same) and the Hessian is updated by the BFGS formula (Eq. 60). Then a new value of the parameter estimates is calculated by

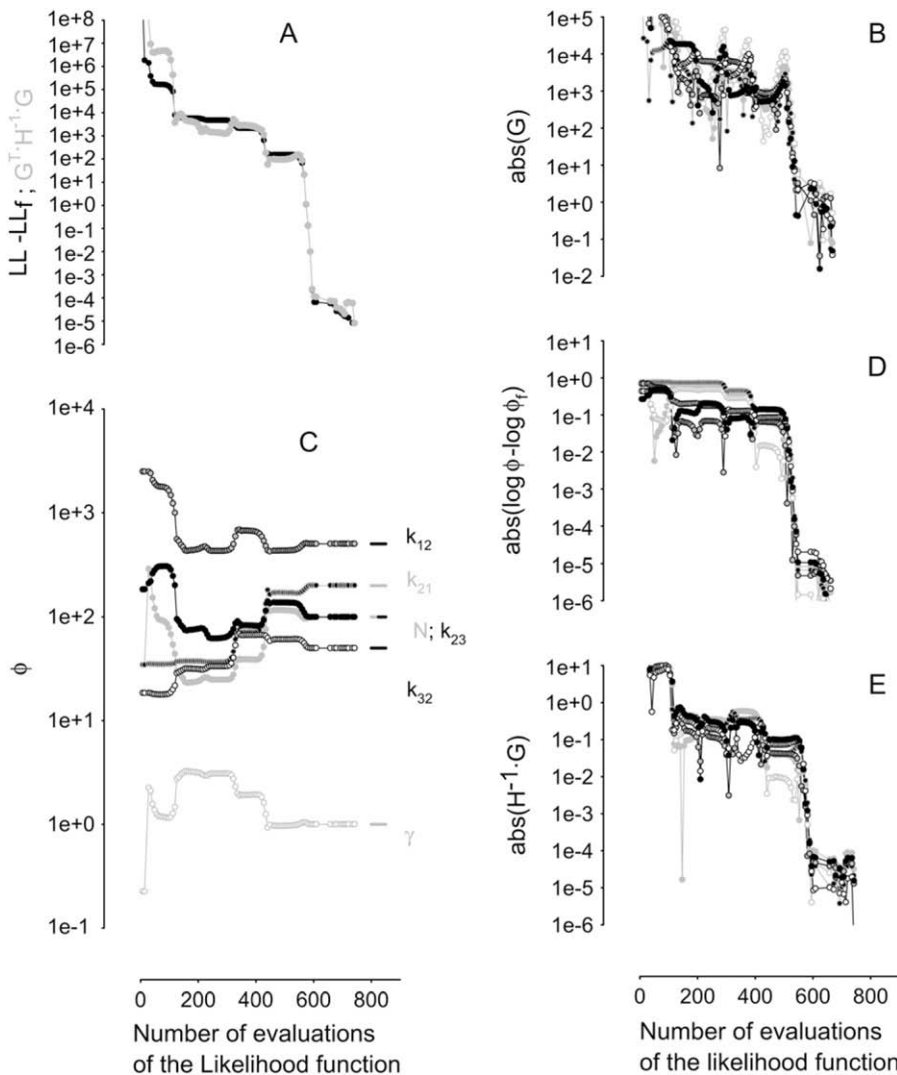


FIGURE 5 Convergence of the macroscopic-recursive MLE of 1000 simulated traces of 100 channels in nonstationary conditions. Initial parameters were chosen at random and after 1000 likelihood function evaluations, the algorithm converged to values close to the true parameters. (A) Evolution of the excess in log-likelihood as compared to the final value. In shading is the estimation of the log-likelihood excess obtained by Eq. 59, $\Delta LL = G^T H^{-1} G$, where G and H are the gradient (obtained by finite differentiation) and the Hessian (obtained by the BFGS approximation) of the log-likelihood. (B) Evolution of the gradient of the log-likelihood with respect to the parameters. (C) Evolution of the parameter estimates, short lines indicate the true values. (D) Evolution of the absolute difference between the current parameter value and the parameter value to which the algorithm will converge. (E) Evolution of the absolute value of the application of Eq. 58 $\Delta \theta = H^{-1} G$, which estimates the values shown in panel D. Symbols coding for different parameters were the same in panels B–E: N (edge, shaded; fill, shaded), γ (edge, shaded; fill, open), k_{21} (edge, shaded; fill, solid), k_{12} (edge, solid; fill, shaded), k_{32} (edge, solid; fill, open), and k_{23} (edge, solid; fill, solid).

performing a line search on the direction given by Eq. 58. The absolute values of the components of this direction are given in Fig. 5 D. Successive steps of the algorithm provide better and better estimates, as judged by the decrease in the negative of the log-likelihood (Fig. 1 A). An important issue is when to stop the algorithm. A first possibility is to wait until the absolute value of the gradient (Fig. 6 B) is smaller than a given amount. However, the value of the gradient increases with the increase in the sample size, so a criterion independent from the sample size would be more useful. The vector-valued gradient can be used to estimate the log-likelihood excess (i.e., current log-likelihood minus the log-likelihood of the optimal parameters) by applying Eq. 59 and taking the BFGS approximated value of the Hessian. This value is shown as the shaded line in Fig. 5 A. After running the minimization algorithm for 800 function evaluations, no more changes in the LL were found. In this way, a final value was obtained for the log-likelihood and the estimates of the parameters. We use then this value to estimate the difference

between the log-likelihood and the optimal log-likelihood and see if the estimation made by Eq. 57 is good. Fig. 6 A shows that the estimation is fairly good for values of log-likelihood under 10–100. In the same way, we can compare the absolute value of the difference between the running value of parameters (Fig. 5 E) and the final estimation of them with the value calculated by Eq. 58 (Fig. 5 D). The estimation is good for values in the 0.01–0.0001 range. The discrepancy at lower values is probably due to the rounding error at calculating the gradient by finite differentiating.

Convergence of recursive and nonrecursive algorithms in stationary and nonstationary conditions

To determine its robustness, recursive MLE was performed on 25 more sets of data generated on nonstationary conditions. For each set of simulated data, the starting guess of the estimated parameters were randomly chosen from values between $10\times$ lower or $10\times$ greater than the true values. Fig. 6 A

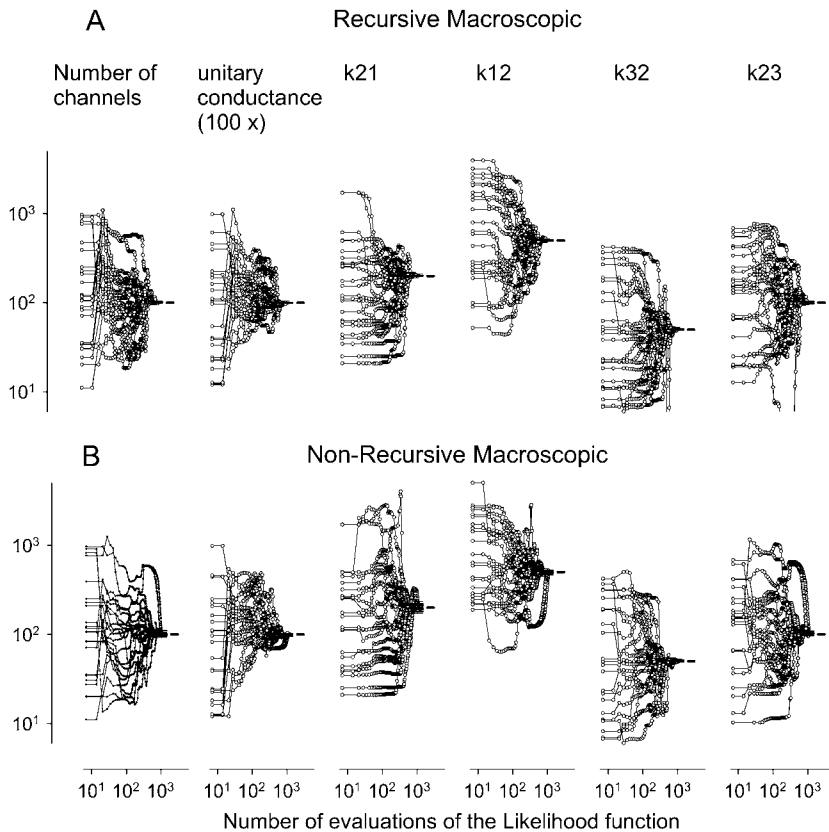


FIGURE 6 Convergence of the MLE on nonstationary conditions. (A,B) Both macroscopic-recursive and macroscopic-nonrecursive algorithms converge. Each trace represents the evolution of the nonlinear maximization of the likelihood of a different set of 1000 simulated traces of 100 channels. Each point represents an iteration of the maximization algorithm. Values estimated by the macroscopic-recursive algorithm are closer to ones used for the simulation of the data (depicted as short lines).

represents the evolution of the quasi-Newton algorithm estimation of the number of channels, conductance, k_{12} , k_{21} , k_{23} , and k_{32} for each one of the data sets. After 2000 evaluations of the likelihood function, the algorithm found, in all the cases, a value very close to the true (Fig. 6 A, small lines). The nonrecursive MLE was also applied to the same data sets: it was also efficient in finding the parameters but with a higher error rate (Fig. 7 B).

The same experimental simulation was performed on stationary conditions. In stationary conditions, the tested model is symmetrical to exchanging C_1 by C_3 . The only difference between C_1 and C_3 is that k_{12} is $10 \times k_{32}$ and k_{21} is $2 \times k_{23}$. As we randomly choose the starting value of the rates, it is sometimes possible that the starting value of k_{32} is greater than the starting value of k_{12} ; in those cases, the algorithm prefers to increase its value rather than to decrease it, therefore swapping C_1 by C_3 . That happened in 3 out of 25 simulations (see shaded lines in Fig. 7 A) where the estimated value of k_{21} converged to the true value of k_{23} , the estimated value of k_{12} converged to the true value of k_{32} , and vice versa. In the other 22 sets, all the kinetic values converged to themselves. To find out the factor determining that the algorithm swaps C_1 by C_3 , the trajectories of the parameters estimates were further analyzed by plotting k_{12} against k_{32} (Fig. 8). Trajectories conformed to two clear groups: those that swapped C_1 by C_3 (shaded line), and those that did not (solid line). The dashed line represents $y = x$; points above it indicate that k_{32} is $> k_{12}$, which was the case of all the

simulations that swapped states. There were two starting points (open triangles) where, even though k_{32} was slightly greater than k_{21} , the algorithm did not swap. In those cases, it was found that k_{12} was $> k_{32}$ (as is the case in the true parameters).

Therefore, after correcting the variation due to swapping states, the recursive algorithm is remarkably precise in returning the true values of the parameters. It is completely different in the case of the nonrecursive algorithm. In stationary conditions, there are no changes in the expected mean current, so the nonrecursive algorithm has no means to estimate the kinetic rates. It is not a surprise, then, to find that the nonrecursive algorithm failed completely in estimating the kinetic parameters (Fig. 7 B). This failure can be seen in the fact that the initial variability that existed in the fed values of the parameters remained after the competition of the minimization algorithm. Interestingly, some convergence was found in the value of the unitary conductance (Fig. 7 B).

Covariance of the adjusted parameters: theoretical and numerical results

One of the advantages of a good approximation to the likelihood function is that the inverse of the Hessian matrix provides an estimate of the covariance matrix of the maximized parameters (see Eq. 61). However, to provide meaningful information, the Hessian has to be positive-definite (i.e., all the eigenvalues have to be real and positive), something that

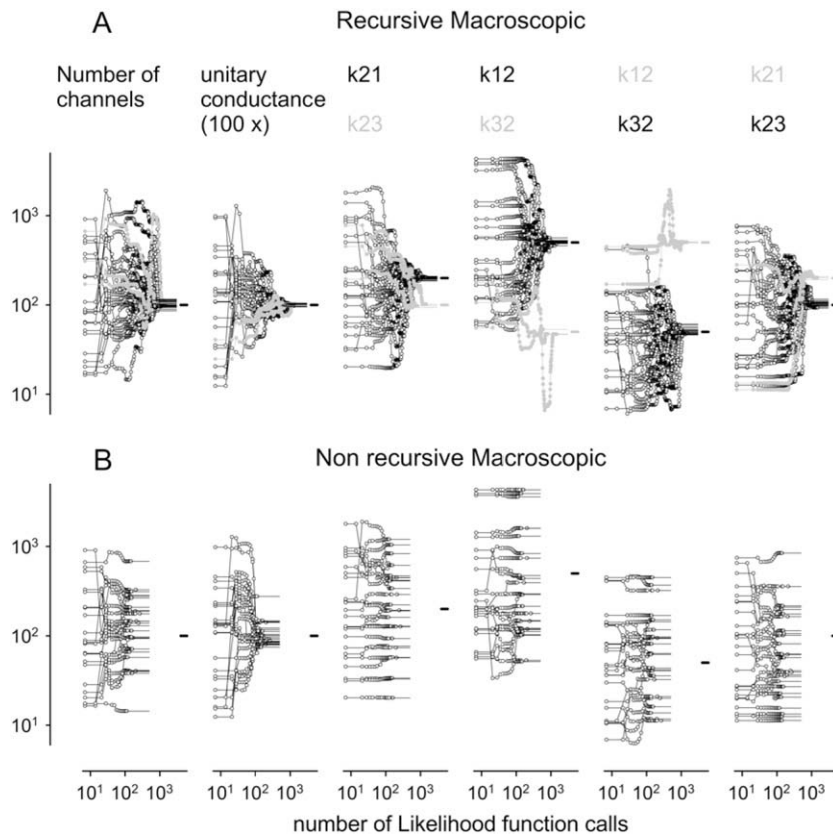


FIGURE 7 Convergence of the MLE on stationary conditions. (A) The macroscopic-recursive algorithm converges. As the model was symmetrical to swapping C_1 by C_3 , in three out of 25 simulations (*shaded*), the values of the kinetic parameters exchanged values: k_{12} by k_{32} and k_{21} by k_{23} and vice versa. (B) Nonrecursive algorithm failed to converge.

was not found always to be the case. Fortunately there is another estimate of the Covariance, the Fisher Information matrix, which is positive-definite by construction and, also important, computationally cheaper (it needs only the numer-

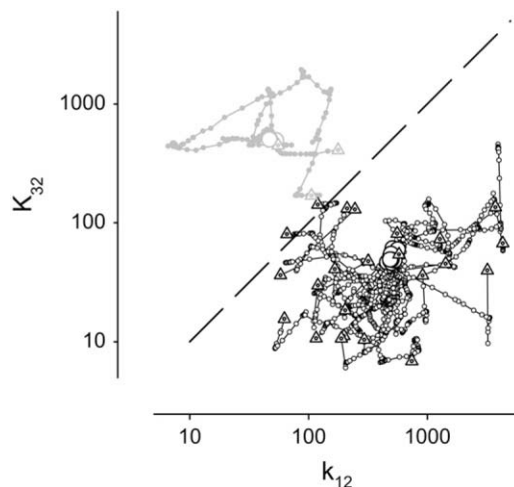


FIGURE 8 Convergence of recursive MLE on stationary conditions. Trajectory of the pair of kinetic parameters that differed the most, k_{12} and k_{32} . Each point indicates an iteration of the algorithm and a circle indicates the final parameter estimation. In most cases (*solid*) the true parameter pair was found, but in three cases (*shaded*) the algorithm chose to swap the values. In all those cases, the initial pair of values was above the line $y = x$, meaning that was bigger than the initial value of k_{12} .

ical partial derivatives of first instead of second order). The covariance estimated in this way provides an accurate description of the correlation structure of the estimated parameters both in nonstationary (Fig. 9) and stationary (Fig. 10) conditions. The information available to the algorithm in stationary and nonstationary conditions is different in nature: in stationary conditions, the algorithm uses only the information present in the random fluctuations, whereas in nonstationary conditions, the algorithm also uses the time course of the response after the stimulation. The macroscopic recursive algorithm performed equally well in describing the covariance of the estimates in both situations.

Accuracy and sensitivity of recursive and nonrecursive methods

The accuracy and sensitivity of the recursive and nonrecursive methods were analyzed in relation with an increased sample size. As it can be seen in the left panels of Fig. 11, the accuracy increased with the number of traces included in the analysis, for the three tested situations: recursive-nonstationary; recursive-stationary; and nonrecursive-nonstationary. Furthermore, the sensitivity also increased with the sample size (Fig. 11, *right panel*), doubling the sensitivity each time the sample size is quadrupled. If we compare the recursive and nonrecursive algorithms at nonstationary conditions, we see a greater accuracy and higher sensitivity in the first case, with the only exception of the slowest rate, k_{32} , where the

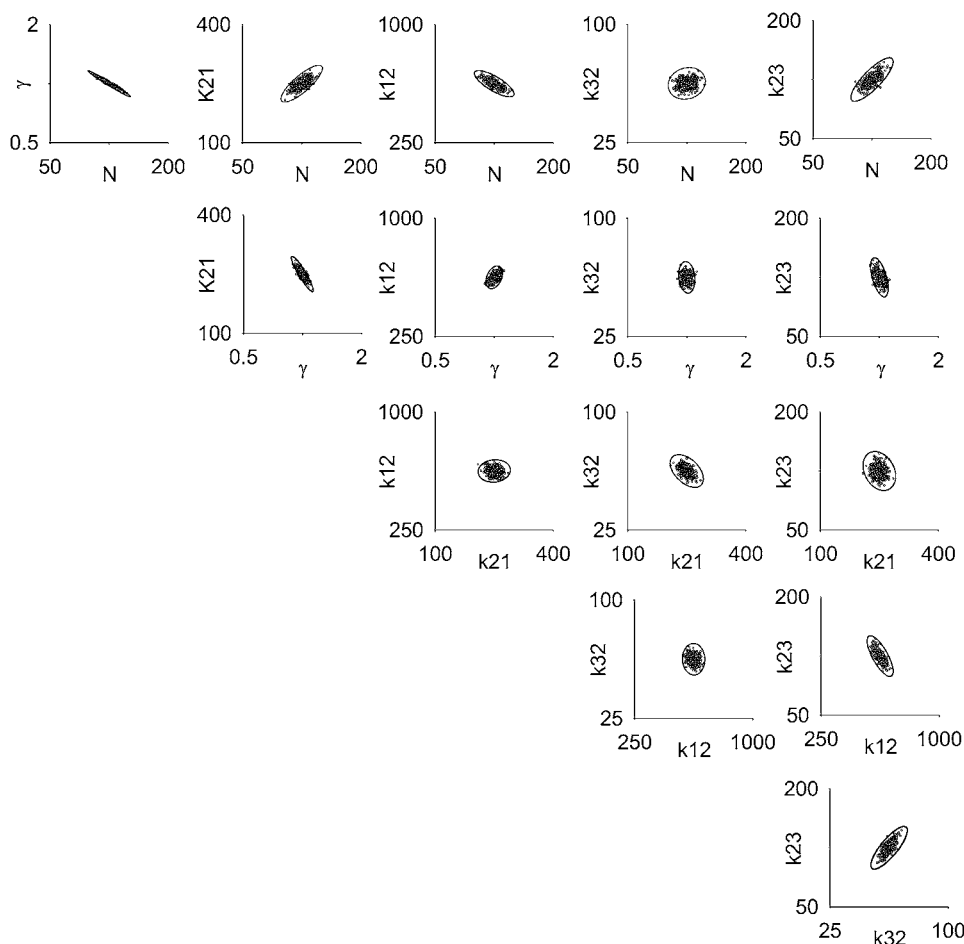


FIGURE 9 Scatter plots of the parameters estimated by the macroscopic-recursive algorithm in nonstationary conditions (100 traces). In nonstationary conditions, the algorithm uses information coming both from the random fluctuations in the number of open channels as well as from the time course of the mean number of open channels after the perturbation. Ellipses represent five standard deviations from the mean and were obtained from the inverse of the Fisher information matrix (the variance of the gradient vector, calculated from the variance of the derivative of the incremental likelihood) which estimates the covariance of the estimates.

sensitivity was the same. To guarantee the same minimum sensitivity, we need $\sim 10\times$ more traces using the nonrecursive algorithm. If we compare nonstationary and stationary conditions using the recursive algorithm, we find, for all the estimates, more sensitivity at nonstationary conditions.

The solid lines represent the expected standard deviation estimated from the Fisher Information Matrix in the recursive algorithm and from the FIM divided by 73 (the expansion in the variance of the likelihood) in the nonrecursive algorithm, and the points represent the standard deviation of the actual parameter estimates. As can be seen in the figure, the estimation of the standard deviation is remarkably accurate, with the notable exception of the slowest rate k_{32} in the nonrecursive method, where the standard deviation is approximately double that expected.

DISCUSSION

The kinetic analysis of electrophysiological recordings is easier when the number of channels is either one or very high. In those cases there are powerful and well-understood methods, especially for single channels (11–15,24). With the continuously increasing computing power available, interest in

statistical methods that can deal with a regular number of channels is increasing (25). In that sense, finding an MLE method for small populations of channels is highly desirable. The main problem to do that is to formulate the likelihood of a macroscopic trace. The exact solution (given by the recursive microscopic algorithm) has been known since the adaptation of MLE for single channels (15), but it is computationally prohibitive, since it scales as the $2(k-1)$ power of the number of channels. Therefore, efficient approximations are needed. A first step in that direction was made by proposing a non-recursive algorithm that approximates the likelihood function of macroscopic currents (19). This method was successful in retrieving all kinetic rates, conductance, and number of channels of a simple kinetic model, after providing an enormous number of traces. However, their method ignores all the information present in the local time correlation of the data (19), and therefore it does not provide reliable error rates of the estimates as present results show. Another approach directly attacked that problem by making a direct fit of the covariance matrix of the measured data (20). In this way, a much better approximation to the likelihood function and therefore reliable estimates of the error rates can be obtained. However, since this approach involves calculating the covariance matrix of the

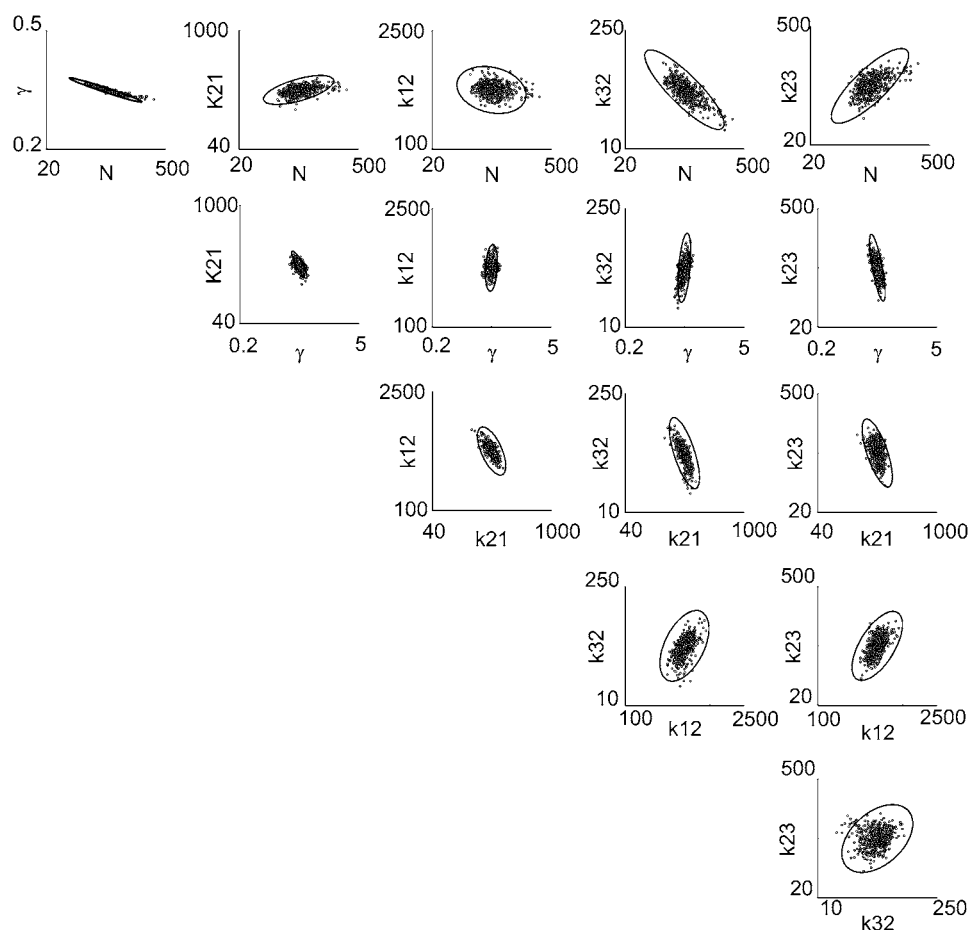


FIGURE 10 Scatter plots of the parameters estimated by the recursive macroscopic algorithm in stationary conditions (100 traces), where the algorithm has to rely only on the random fluctuations of the response. Same conventions as in Fig. 9.

full set of measurements, it can only be applied to a subset of the measurements.

Present work integrates both approaches by postulating a new algorithm, macroscopic and recursive, which approximates the likelihood of the macroscopic currents including the information present in the local time correlation of the data. This new algorithm is computationally affordable and does not scale with the number of channels, although it is more expensive computationally than the nonrecursive one. The simulations presented here have shown that the proposed macroscopic recursive algorithm provides likelihood values close to the exact microscopic recursive algorithm and that both values converge with increasing number of channels. Consequently, MLE provides values that converge with increasing number of traces to the true parameters values used originally to generate the data. As has been previously shown (19), the nonrecursive macroscopic method also converge, although it needed $\sim 10\times$ more traces to assure the same maximal error in the estimates. The proposed algorithm provides reliable values for the covariance matrix of the estimates, something that the nonrecursive method is not able to do. Interestingly, approximate values for the error rates in the estimates were found by taking the values provided by the

nonrecursive algorithm and multiplying them by the expansion factor found in the variance of the log-likelihood. However, the error rates thus found were not right for all the estimates, showing that this is not a reliable way to obtain the error rates in a general way.

It was surprising to find that the proposed method allows the recovery of six parameters (four kinetic rates + number of channels and conductance) in stationary conditions (Fig. 8 A). Where does the algorithm get the information? Equilibrium noise analysis (7,26) could provide four independent parameters (two time constants with their coefficients) and the mean current could provide a fifth. Where the sixth come from? Although the channels are in equilibrium, they spend some time away from it. So, it is likely that the algorithm makes use of that variation to get an estimate of the well-known relationship between variance and mean current that is used in nonstationary analysis of noise (27).

In fact, the same problem is present in the ability of the nonrecursive method to get the same six parameters in nonstationary conditions (Fig. 7 B and (19)). In this case, there are again four parameters provided by the temporal course of the relationship (two time constants with their coefficients) and one by the variance. The sixth is probably provided by

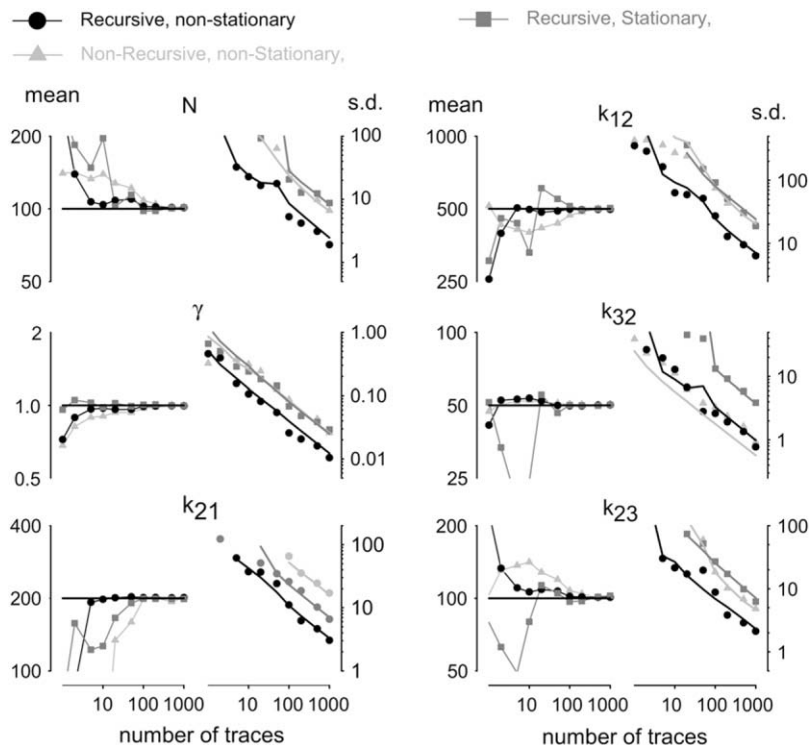


FIGURE 11 Comparison of the accuracy and sensitivity of macroscopic recursive in nonstationary and stationary conditions and of nonrecursive in nonstationary conditions. The thick lines show the theoretical sensitivity predicted by the Fisher Information Matrix alone (for the recursive algorithm) or corrected (for the nonrecursive algorithm) by the increase in the log-likelihood variance (~ 73). To get errors of $\leq 10\%$ for all the parameters, the macroscopic-recursive method needed 100 traces in nonstationary conditions and 1000 traces in stationary conditions. The macroscopic-nonrecursive method needed 1000 traces in nonstationary conditions.

the relationship between the variance and mean current, which is subject to variations induced by the changes in the agonist concentration.

Present work validates the macroscopic recursive algorithm with a simple kinetic scheme of three states. More complex schemes would involve the task of recovering a greater number of parameters. For each extra state, there would be an extra time constant with its coefficient. Therefore, if the number of added parameters in the complex scheme is $< 2 \times$, the number of extra states, the Macro R algorithm should be able to recover all the parameters even in stationary conditions, as long as a sufficient number of traces were provided.

Bayesian approach to HMM

In this article, the classical forward algorithm that calculates the likelihood function in hidden Markov modeling is presented in a Bayesian way. That allows us to replace the “alphas” of the old terminology by three meaningful variables: the prior and posterior state probability functions that indicate the knowledge the observer has about the distribution of states before and after each measurement and the incremental likelihood of each measurement, which indicates how likely is each measurement, after the deduction of previous ones. The partial likelihoods are statistically independent by construction, a very nice statistical property that allows estimating the Fisher Information Matrix, which under regularity conditions is an estimator of the Hessian and, therefore, of the inverse of the covariance of the parameters estimates.

Conditions where the algorithms can be applied and how to interpret them

There are several aspects of the real world electrophysiology that are not taken into account in either of the algorithms presented here. First, the noise of electrophysiological traces is usually not Gaussian white but colored, and there is always the effect of the low-pass filter used to clean the signal. There are published methods (10,14) that deal with colored noise and filtering that might be extrapolated to multiple channels. Second, in the modeling presented here, the measurements are considered to be instantaneous, where electrophysiological measurements always result from some time-averaging. At a single channel this problem leads to a failure in the detection of brief openings and it has been heavily studied (11,28–32). In macroscopic analysis, this problem leads to the underestimation of the fluctuations at high frequency. Building observational models that take into account this fact is necessary to increase the time resolution of the kinetic modeling. This problem is especially significant for multiple channels recording, since the fluctuation kinetics increases linearly with the number of channels. A formulation of the algorithms that can deal with time-averaged signals is therefore fundamental to the widespread use of the algorithms presented here. Both problems are connected and they might be solved together.

Macroscopic-recursive MLE makes two more assumptions:

The first assumption is that the state probability function is well approximated by a multivariate normal distribution. The error made by this approximation is higher when the

number of channels is small. This can be seen in the increased difference between microscopic and recursive algorithms (Fig. 4) at lower number of channels. For this reason, the method was tested with 100 channels with a mean of ~ 30 open channels. The method might be biased at a smaller number of channels. More sophisticated approximations to the state probability function are necessary to overcome this limitation. Alternatively, it is possible that, by integrating on longer times, the approximation to normality holds.

The second assumption is that the distribution of the partial likelihood, with respect to the measurement, is normal. This assumption starts to be violated if the instrumental noise is less than the unitary conductance.

The nonrecursive method makes one more assumption: that there is no local time correlation in the residuals. There are two ways to fulfill this assumption. One way is to measure at intervals greater than the expected correlation time. However, that implies measuring at times greater than the kinetic constants, and therefore completely missing the kinetics. The other way is to have more instrumental noise than kinetic noise. However, the information obtained from the variance is also lost; the method in this case is equal to a least-squares sum, where no information about the number of channels and conductance can be obtained. So, there is no useful way for this method to fulfill its assumptions. However, the method seems to be able to recover the parameters in the tested situation. The only effect of the local time correlation seems to be in the estimation of the error rates of the parameter estimates (and therefore probably in the ability to distinguish between alternative models). The fact that a simple correction, multiplying by the variance inflation factor determined in Fig. 4, was enough to correct most (but not all) of the error estimates (Fig. 11) suggests that it might be possible to estimate the error rates in a relatively simple way, making the nonrecursive algorithm an attractive alternative to the recursive algorithm, when computing power is a problem.

Author thanks Shlomo Dellal and Rachel Tittle for improving the English of the manuscript and an anonymous referee for a very thorough review with many valuable suggestions (including a didactic deduction of Eqs. 54 and 55).

This work was supported by CONICET and ANPCyT (grant No. PICT 06-12345).

REFERENCES

- Hille, B. 2001. *Ion Channels of Excitable Membranes*. Sinauer, Sunderland, MA.
- Hodgkin, A. L., and A. F. Huxley. 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117:500–544.
- Hamill, O. P., A. Marty, E. Neher, B. Sakmann, and F. J. Sigworth. 1981. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.* 391:85–100.
- Colquhoun, D., and B. Sakmann. 1981. Fluctuations in the microsecond time range of the current through single acetylcholine receptor ion channels. *Nature*. 294:464–466.
- Colquhoun, D., and A. G. Hawkes. 1981. On the stochastic properties of single ion channels. *Proc. R. Soc. Lond. B Biol. Sci.* 211:205–235.
- Colquhoun, D., and A. G. Hawkes. 1982. On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 300:1–59.
- Colquhoun, D., and A. G. Hawkes. 1977. Relaxation and fluctuations of membrane currents that flow through drug-operated channels. *Proc. R. Soc. Lond. B Biol. Sci.* 199:231–262.
- Sigworth, F. J., and S. M. Sine. 1987. Data transformations for improved display and fitting of single-channel dwell time histograms. *Biophys. J.* 52:1047–1054.
- Milne, R. K., G. F. Yeo, B. W. Madsen, and R. O. Edeson. 1989. Estimation of single channel kinetic parameters from data subject to limited time resolution. *Biophys. J.* 55:673–676.
- Venkataramanan, L., and F. J. Sigworth. 2002. Applying hidden Markov models to the analysis of single ion channel activity. *Biophys. J.* 82:1930–1942.
- Qin, F., A. Auerbach, and F. Sachs. 1996. Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* 70:264–280.
- Qin, F., A. Auerbach, and F. Sachs. 1997. Maximum likelihood estimation of aggregated Markov processes. *Proc. Biol. Sci.* 264:375–383.
- Qin, F., A. Auerbach, and F. Sachs. 2000. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophys. J.* 79:1915–1927.
- Qin, F., A. Auerbach, and F. Sachs. 2000. Hidden Markov modeling for single channel kinetics with filtering and correlated noise. *Biophys. J.* 79:1928–1944.
- Horn, R., and K. Lange. 1983. Estimating kinetic constants from single channel data. *Biophys. J.* 43:207–223.
- Fredkin, D. R., and J. A. Rice. 1992. Maximum likelihood estimation and identification directly from single-channel recordings. *Proc. Biol. Sci.* 249:125–132.
- Shanno, D. F. 1970. Conditioning of quasi-Newton methods for function minimization. *Math. Comput.* 24:647–656.
- Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*. 77:257–286.
- Milescu, L. S., G. Akk, and F. Sachs. 2005. Maximum likelihood estimation of ion channel kinetics from macroscopic currents. *Biophys. J.* 88:2494–2515.
- Celentano, J. J., and A. G. Hawkes. 2004. Use of the covariance matrix in directly fitting kinetic parameters: application to GABAA receptors. *Biophys. J.* 87:276–294.
- Baum, L. E., and T. Petrie. 1966. Statistical inference for probabilistic functions of finite state Markov Chains. *Ann. Math. Stat.* 37:1554–1563.
- Baum, L. E., T. Petrie, G. Soules, and N. Weiss. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.* 41:164–171.
- Becker, J. D., J. Honerkamp, J. Hirsch, U. Frobe, E. Schlatter, and R. Greger. 1994. Analyzing ion channels with hidden Markov models. *Pflügers Arch.* 426:328–332.
- Colquhoun, D., C. J. Hatton, and A. G. Hawkes. 2003. The quality of maximum likelihood estimates of ion channel rate constants. *J. Physiol.* 547:699–728.
- Cannon, R. C., and G. D'Alessandro. 2006. The ion channel inverse problem: neuroinformatics meets biophysics. *PLoS Comput. Biol.* 2:0862–0869.
- Neher, E., and C. F. Stevens. 1977. Conductance fluctuations and ionic pores in membranes. *Annu. Rev. Biophys. Bioeng.* 6:345–381.
- Sigworth, F. J. 1981. Interpreting power spectra from nonstationary membrane current fluctuations. *Biophys. J.* 35:289–300.

28. Ball, F. G., S. S. Davies, and M. S. Sansom. 1990. Single-channel data and missed events: analysis of a two-state Markov model. *Proc. Biol. Sci.* 242:61–67.
29. Crouzy, S. C., and F. J. Sigworth. 1990. Yet another approach to the dwell-time omission problem of single-channel analysis. *Biophys. J.* 58:731–743.
30. Magleby, K. L., and D. S. Weiss. 1990. Estimating kinetic parameters for single channels with simulation. A general method that resolves the missed event problem and accounts for noise. *Biophys. J.* 58:1411–1426.
31. Blatz, A. L., and K. L. Magleby. 1986. Correcting single channel data for missed events. *Biophys. J.* 49:967–980.
32. Hawkes, A. G., A. Jalali, and D. Colquhoun. 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 Biol. Sci.* 337:383–404.