Parameter estimation in the stochastic Morris-Lecar neuronal model with particle filter methods

Susanne Ditlevsen, Adeline Samson

Parameter estimation in the stochastic Morris-Lecar neuronal model with particle filter methods

SUSANNE DITLEVSEN
Department of Mathematics, University of Copenhagen
Universitetsparken 5, DK-2100 Copenhagen
susanne@math.ku.dk

ADELINE SAMSON
Université Paris Descartes, Laboratoire MAP5 UMR CNRS 8145
45 rue des St Peres, 75006 Paris
Adeline.Samson@parisdescartes.fr

June 25, 2012

Abstract

Parameter estimation in two-dimensional diffusion models with only one coordinate observed is highly relevant in many biological applications, but a statistically difficult problem. In neuroscience, the membrane potential evolution in single neurons can be measured at high frequency, but biophysical realistic models have to include the unobserved dynamics of ion channels. One such model is the stochastic Morris-Lecar model, where random fluctuations in conductance and synaptic input are specifically accounted for by the diffusion terms. It is defined through a non-linear two-dimensional stochastic differential equation with state dependent noise on the non-observed coordinate. The coordinates are coupled, i.e. the unobserved coordinate is non-autonomous, and we are therefore not in the more well-behaved situation of a hidden Markov model. In this paper, we propose a sequential Monte Carlo particle filter algorithm to impute the unobserved coordinate, and then estimate parameters maximizing a pseudo-likelihood through a stochastic version of the Expectation-Maximization algorithm. The performance is evaluated in a simulation study, and it turns out that even the rate scaling parameter governing the opening and closing of ion channels of the unobserved coordinate can be reasonably estimated. Also an experimental data set of intracellular recordings of the membrane potential of a spinal motoneuron of a red-eared turtle is analyzed.

Keywords: Sequential Monte Carlo; diffusions; pseudo likelihood; Stochastic Approximation Expectation Maximization; motoneurons; conductance-based neuron models; membrane potential

1 Introduction

In neuroscience, it is of major interest to understand the principles of information processing in the nervous system, and a basic step is to understand signal processing and transmission in single neurons.

Therefore, there is a growing demand for robust methods to estimate biophysical relevant parameters from partially observed detailed models. Statistical inference from experimental data in biophysically detailed models of single neurons is difficult. Data are typically of either of two types: extracellular recordings, where only the spike times are observed, i.e. a point process, or intracellular recordings, where the membrane potential is recorded at high frequency, e.g. observations are taken every 0.1 ms. The data are naturally described by single point models, neglecting morphological properties of the neuron, since no spatial information is available in the data, and all neuronal properties are collapsed into a single point in space. Even so, the models can be multi-dimensional and have far more parameters than the data can provide information about. Often these models are compared to experimental data by hand-tuning to reproduce the qualitative behaviors observed in experimental data, but without any formal statistical analysis.

The complexity of neuronal single cell models ranges from detailed biophysical descriptions, exemplified most prominently in the Hodgkin-Huxley model, to simplified one-dimensional diffusion models. The Hodgkin-Huxley model (Hodgkin and Huxley (1952)) is a conductance-based model, which consists of four coupled differential equations, one for the membrane voltage, and three equations describing the gating variables that model the voltage-dependent sodium and potassium channels. Conductance-based models provide a minimal biophysical interpretation for an excitable cell, including the voltage dependence or time-dependent nature of the conductance due to movement of ions across ion channels in the membrane. Several simplifications of the Hodgkin-Huxley model has been proposed, most commonly in order to reduce the four-dimensional system to a two-dimensional system, using the time scale separations by treating the fast variables as instantaneous variables by a quasi steady-state approximation, and by collapsing variables with nearly identical dynamics.

The Morris-Lecar model (Morris and Lecar (1981)) has often been used as a good, qualitatively quite accurate, two-dimensional model of neuronal spiking. It is a conductance-based model like the Hodgkin-Huxley model, introduced to explain the dynamics of the barnacle muscle fiber. It can be considered a prototype for a wide variety of neurons. The model is given by two coupled first order differential equations, the first modeling the membrane potential evolution, and the second equation modeling the activation of potassium current. If both current and conductance noise should be taken into account, the stochastic Morris-Lecar model arises, where diffusion terms have been added on both coordinates, governed by Wiener processes. Typically, the membrane potential will be measured discretely at high frequency, whereas the second variable cannot be observed. Our goal is to estimate parameters of the model from discrete observations of the first coordinate. This includes estimation of a central rate parameter characterizing the channel kinetics of the unobserved component, and the two diffusion parameters representing the amplitude of the noise.

In Huys *et al.* (2006) up to 10^4 parameters are estimated in a detailed multi-compartmental single neuron model. Only parameters entering linearly in the loss function are considered, and channel kinetics are assumed known. It is a quadratic optimization problem solved by least squares, and shown to work well for low noise and high frequency sampling. When either the discretization step or the noise increase, a bias is introduced. In Huys and Paninski (2009) they extend the estimation to allow for measurement noise, first smoothing the data by a particle filter, and then maximizing the likelihood through a Monte Carlo EM-algorithm.

In this paper, we propose to approximate the non-linear two-dimensional model through an Euler-Maruyama scheme to obtain a tractable pseudo-likelihood. Then we consider the statistical model as

an incomplete data model, where the unobserved part is imputed by a Sequential Monte Carlo (SMC) algorithm. This is not straightforward for the type of conductance based model, we are studying. It is a multi-dimensional coupled stochastic differential equation (SDE), where only one coordinate is observed. This means that the unobserved coordinate is non-autonomous, and the model does not fit into the hidden Markov model framework. Furthermore, the diffusion is not time reversible. Thus, we cannot directly apply the SMC algorithm proposed in Del Moral *et al.* (2001), which assumes the non-observed data autonomous, nor the algorithms proposed in Fearnhead *et al.* (2008), which assumes the drift to be of gradient type. In the ergodic case this corresponds to a time reversible diffusion. In the particle filter proposed in our paper, the observed coordinate is not re-simulated. Finally, we maximize the pseudo-likelihood, based on an Euler-Maruyama approximation of the SDE defining the model, through a Stochastic Approximation Expectation-Maximization (SAEM) algorithm, where the unobserved data are imputed at each iteration of the algorithm. We prove that the estimator obtained from this combined SAEM-SMC algorithm converges with probability one to a local maximum of the pseudo-likelihood. We also prove that the pseudo-likelihood converges to the true likelihood as the time step between observations go to zero.

The paper is organized as follows: In Section 2 the model is presented, the special noise structure, we are considering, is motivated, and the pseudo likelihood arising from the Euler-Maruyama approximation is found. In Section 3 we present the proposed estimation procedure and the assumptions needed for the convergence results to hold. In Section 4 we conduct a simulation study to document the performance of the method, and in Section 5 we apply the method on an experimental data set of intracellular recordings of the membrane potential of a motoneuron of an adult turtle. Proofs and technical results can be found in the Appendix.

2 Stochastic Morris-Lecar model

2.1 Exact diffusion model

We consider a stochastic Morris-Lecar model including both current and channel noise, defined as

$$\left\{ \begin{array}{rcl} dV_t &=& f(V_t,U_t)dt + \gamma d\tilde{B}_t, \\ dU_t &=& b(V_t,U_t)dt + \sigma(V_t,U_t)dB_t, \end{array} \right. \tag{1}$$

where

$$\begin{aligned} f(V_t,U_t) &= \frac{1}{C}\left(-g_{Ca}m_\infty(V_t)(V_t-V_{Ca}) - g_K U_t(V_t-V_K) - g_L(V_t-V_L) + I\right), \\ b(V_t,U_t) &= \left(\alpha(V_t)(1-U_t) - \beta(V_t)U_t\right), \\ m_\infty(v) &= \frac{1}{2}\left(1+\tanh\left(\frac{v-V_1}{V_2}\right)\right), \\ \alpha(v) &= \frac{1}{2}\phi\cosh\left(\frac{v-V_3}{2V_4}\right)\left(1+\tanh\left(\frac{v-V_3}{V_4}\right)\right), \\ \beta(v) &= \frac{1}{2}\phi\cosh\left(\frac{v-V_3}{2V_4}\right)\left(1-\tanh\left(\frac{v-V_3}{V_4}\right)\right), \end{aligned}$$

and the initial condition (V_0,U_0) is random with density $p(V_0,U_0)$. Processes $(\tilde{B}_t)_{t\geq t_0}$ and $(B_t)_{t\geq t_0}$ are independent Brownian motions. The variable V_t represents the membrane potential of the neuron at

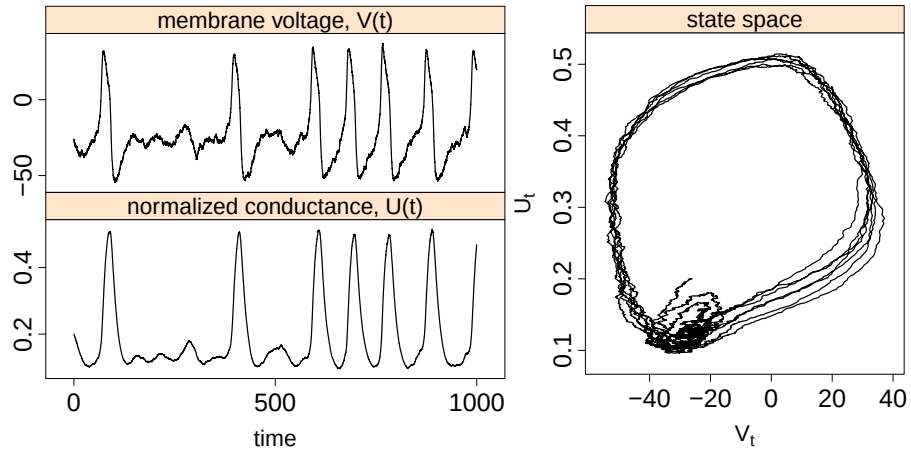

Figure 1: Simulated trajectory of the stochastic Morris-Lecar model: (V_t) as a function of time (left, top), (U_t) as a function of time (left, bottom), and (U_t) against (V_t) (right). Parameters are given in Section 4. Time is measured in ms, voltage in mV, the conductance is normalized between 0 and 1.

time t, and U_t represents the normalized conductance of the K^+ current. This is a variable between 0 and 1, and could be interpreted as the probability that a K^+ ion channel is open at time t. The equation for $f(\cdot)$ describing the dynamics of V_t contains four terms, corresponding to Ca^{2+} current, K^+ current, a general leak current, and the input current I. The functions $\alpha(\cdot)$ and $\beta(\cdot)$ model the rates of opening and closing, respectively, of the K^+ ion channels. The function $m_\infty(\cdot)$ represents the equilibrium value of the normalized Ca^{2+} conductance for a given value of the membrane potential. The parameters V_1, V_2, V_3 and V_4 are scaling parameters; g_{Ca}, g_K and g_L are conductances associated with Ca^{2+}, K^+ and leak currents, respectively; V_{Ca}, V_K and V_L are reversal potentials for Ca^{2+}, K^+ and leak currents, respectively; C is the membrane capacitance; ϕ is a rate scaling parameter for the opening and closing of the K^+ ion channels; and I is the input current.

Parameter γ scales the current noise. Function $\sigma(V_t, U_t)$ models the channel or conductance noise. We consider the following function that ensures that U_t stays bounded in the unit interval if $\sigma \leq 1$ (Ditlevsen and Greenwood, 2012)

$$\sigma(V_t, U_t) = \sigma\sqrt{2\frac{\alpha(V_t)\beta(V_t)}{\alpha(V_t)+\beta(V_t)}U_t(1-U_t)}.$$

A trajectory of the model is simulated in Fig. 1. The peaks of (V_t) corresponds to spikes of the neuron.

Data consist in discrete measurements of (V_t) while (U_t) is not measured. We denote $t_0 \leq t_1 \leq \ldots \leq t_n$ the discrete observation times. We denote $V_i = V_{t_i}$ the observation at time t_i and $V_{0:n} = (V_{t_0}, \ldots, V_{t_n})$ the vector of all the observed data. We focus on the estimation of parameters from observations $V_{0:n}$. Let $\theta \in \Theta \subseteq \mathbb{R}^p$ be the vector of parameters to be estimated. We will consider two cases: Estimation of all identifiable parameters of the observed coordinate, where all parameters of the unobserved channel dynamics are assumed known except for the volatility parameter, $\theta = (g_{Ca}, g_K, g_L, V_{Ca}, V_K, I, \gamma, \sigma)$; and estimation of $\theta = (g_{Ca}, g_K, g_L, V_{Ca}, V_K, I, \gamma, \sigma, \phi)$, where also the rate parameter of the unobserved coordinate is estimated. Note that C is a proportionality factor of the conductance parameters and thus unidentifiable, as well as the constant level in $f(\cdot)$ is given by $g_L V_L + I$, and thus V_L (or I) is unidentifiable. An ideal estimator for θ is the maximum likelihood estimator, obtained by maximizing

the likelihood of $V_{0:n}$. However, this likelihood is intractable, as explained below.

The likelihood function $p(V_{0:n};\theta)$ of model (1) has a complex form for several reasons. Let us first write the likelihood function of the ideal case where the second coordinate (U_t) is also discretely observed. Let $U_{0:n}$ denote a realization of (U_t) at observation times $t_0,\ldots,t_n$. Since the vector (V_i,U_i) is Markovian, the likelihood $p(V_{0:n},U_{0:n};\theta)$ can be written as a product of conditional densities

$$p(V_{0:n},U_{0:n};\theta)=p(V_0,U_0;\theta)\prod_{i=1}^{n}p(V_i,U_i|V_{i-1},U_{i-1};\theta),$$

where $p(V_i,U_i|V_{i-1},U_{i-1};\theta)$ is the transition density of (V_i,U_i) given (V_{i-1},U_{i-1}). Unfortunately, the density $p(V_i,U_i|V_{i-1},U_{i-1};\theta)$ has no explicit form because the diffusion is highly non-linear. Therefore, even when (U_t) is discretely observed at the same time points as (V_t), the likelihood is not explicit and the maximum likelihood estimator is not available. In this ideal case, minimum contrast estimators based on the Euler-Maruyama approximation of the diffusion have been proposed by Kessler (1997).

When the second coordinate (U_t) is not observed, the estimation is much more difficult. Indeed, although (V_t,U_t) is Markovian, the single coordinate (V_t) is not Markovian. The process (U_t) is a latent or hidden variable which has to be integrated out to compute the likelihood

$$p(V_{0:n};\theta)=\int\ldots\int p(V_0,U_0;\theta)\prod_{i=1}^{n}p(V_i,U_i|V_{i-1},U_{i-1};\theta)dU_0\ldots dU_n. \tag{2}$$

Again, the transition density $p(V_i,U_i|V_{i-1},U_{i-1};\theta)$ is generally not available and has to be approximated. We therefore introduce an approximate diffusion based on the Euler-Maruyama scheme.

2.2 Approximate diffusion

Let Δ denote the step size between two observation times. Here we assume that Δ does not depend on i. The extension to unequally spaced observation times is straightforward. The Euler-Maruyama approximation of model (1) leads to a discretized model defined as follows

$$\begin{aligned} V_{i+1} &= V_i+\Delta f(V_i,U_i)+\sqrt{\Delta}\,\gamma\,\tilde{\eta}_i, \\ U_{i+1} &= U_i+\Delta b(V_i,U_i)+\sqrt{\Delta}\sigma(V_i,U_i)\eta_i, \end{aligned} \tag{3}$$

where $(\tilde{\eta}_i)$ and (η_i) are independent centered Gaussian variables. To ease readability the same notation (V_i,U_i) is used for the original and the approximated processes. This should not lead to confusion, as long as the transition densities are distinguished, as done below. The likelihood of the approximate model is equal to

$$p_\Delta(V_{0:n};\theta)=\int\ldots\int p(V_0,U_0;\theta)\prod_{i=1}^{n}p_\Delta(V_i,U_i|V_{i-1},U_{i-1};\theta)dU_0\ldots dU_n, \tag{4}$$

where $p_\Delta(V_i,U_i|V_{i-1},U_{i-1};\theta)$ is the transition density of model (3):

$$\begin{aligned} p_\Delta(V_i,U_i|V_{i-1},U_{i-1};\theta) &= \frac{1}{\sqrt{2\pi\Delta}\gamma}\exp\left(-\frac{(V_i-V_{i-1}-\Delta f(V_{i-1},U_{i-1}))^2}{2\Delta\gamma^2}\right) \\ &\times \frac{1}{\sqrt{2\pi\Delta}\sigma(V_{i-1},U_{i-1})}\exp\left(-\frac{(U_i-U_{i-1}-\Delta b(V_{i-1},U_{i-1}))^2}{2\Delta\sigma^2(V_{i-1},U_{i-1})}\right). \end{aligned}$$

We aim at estimating the maximum of the likelihood (4) of the approximate model, which corresponds to a pseudo-likelihood for the exact diffusion.

3 Estimation method

The multiple integrals of equation (4) are difficult to handle. Thus, it is not possible to maximize the likelihood directly.

A solution is to consider the statistical model as an incomplete data model. The observable vector $V_{0:n}$ is then part of a so-called complete vector $(V_{0:n}, U_{0:n})$, where $U_{0:n}$ has to be imputed. Simulation under the smoothing distribution $p_\Delta(U_{0:n}\,|V_{0:n};\theta)dU_{0:n}$ is likely to be a complex problem, and direct simulation of the non-observed data $(U_{0:n})$ is not possible. A Sequential Monte Carlo (SMC) algorithm, also known as Particle Filtering, provides a way to approximate this distribution (Doucet *et al.*, 2001). We have adapted this algorithm to handle a coupled two-dimensional SDE, i.e. the unobserved coordinate is non-autonomous. Thus, we are not in the more well-behaved situation of a hidden Markov model.

To maximize the likelihood of the complete data vector $(V_{0:n}, U_{0:n})$, we propose to use a stochastic version of the Expectation-Maximization (EM) algorithm, namely the SAEM algorithm (Delyon *et al.*, 1999). Thus, we combine the SAEM algorithm with the SMC algorithm, where the unobserved data are filtered at each iteration step, to estimate the parameters of model (3). Details on the SMC are given in Section 3.1, and the SAEM algorithm is presented in Section 3.2. Convergence of this new SAEM-SMC algorithm is stated in Section 3.3.

3.1 SMC, a particle filter algorithm

In this section, the aim is to approximate the distribution $p_\Delta(U_{0:n}|V_{0:n};\theta)dU_{0:n}$, for a fixed value of θ. When included in the stochastic EM algorithm, this value will be the current value $\widehat{\theta}_m$ of the parameter at the given iteration. The SMC algorithm provides a set of K particles $(U_{0:n}^{(k)})_{k=1\ldots K}$ and weights $(W_{0:n}^{(k)})_{k=1\ldots K}$ approximating the conditional distribution $p_\Delta(U_{0:n}|V_{0:n};\theta)dU_{0:n}$ (see Doucet *et al.*, 2001, for a complete review). The SMC method relies on proposal distributions $q(U_i|V_i, V_{i-1}, U_{i-1};\theta)dU_i$ to sample what we call particles from these distributions. We write $V_{0:i} = (V_0, \ldots, V_i)$ and likewise for $U_{0:i}$.

***Algorithm* 1 (SMC algorithm)**

- At time $i = 0$: $\forall\, k = 1, \ldots, K$
  1. *sample* $U_0^{(k)}$ *from* $p(U_0|V_0;\theta)$
  2. *compute and normalize the weights:*

$$w_0\left(U_0^{(k)}\right) = p\left(V_0, U_0^{(k)};\theta\right), \quad W_0\left(U_0^{(k)}\right) = \frac{w_0\left(U_0^{(k)}\right)}{\sum_{k=1}^K w_0\left(U_0^{(k)}\right)}$$

- At time $i = 1, \ldots, n$: $\forall\, k = 1, \ldots, K$

  1. *resample the particles, i.e. sample the indices* $A_{i-1}^{(k)}$ *from a multinomial distribution such that*

  $$P(A_{i-1}^{(k)} = l) = W_i(U_{0:i-1}^{(l)}), \quad \forall\, l = 1, \ldots, K$$

  and put $U_{0:i-1}^{'(k)} = U_{0:i-1}^{(A_{i-1}^{(k)})}$.

  2. *sample* $U_i^{(k)} \sim q\left(\cdot|V_i, V_{i-1}, U_{i-1}^{'(k)}; \theta\right)$ *and set* $U_{0:i}^{(k)} = (U_{0:i-1}^{'(k)}, U_i^{(k)})$

  3. *compute and normalize the weights:*

$$\begin{aligned} w_i\left(U_{0:i}^{(k)}\right) &= \frac{p_\Delta\left(V_{0:i}, U_{0:i}^{(k)}; \theta\right)}{p_\Delta\left(V_{0:i-1}, U_{0:i-1}^{'(k)}; \theta\right) q\left(U_i^{(k)}|V_i, V_{i-1}, U_{0:i-1}^{'(k)}; \theta\right)} \\ W_i(U_{0:i}^{(k)}) &= \frac{w_i\left(U_{0:i}^{(k)}\right)}{\sum_{k=1}^K w_i\left(U_{0:i}^{(k)}\right)} \end{aligned}$$

Finally, the SMC algorithm provides an empirical measure

$$\Psi_{n,\theta}^K = \sum_{k=1}^K W_n(U_{0:n}^{(k)}) \mathbf{1}_{U_{0:n}^{(k)}}$$

which is an approximation to the smoothing distribution $p_\Delta(U_{0:n}|V_{0:n}; \theta) dU_{0:n}$. Here, $\mathbf{1}_x$ is the Dirac measure in x. A draw from this distribution can be obtained by sampling an index k from a multinomial distribution with probabilities W_n and setting the draw $U_{0:n}$ equal to $U_{0:n} = U_{0:n}^{(k)}$.

As in any importance sampling method, the choice of the proposal distribution q is crucial to ensure good convergence properties. The first classical choice of q is $q(U_i|V_i, V_{i-1}, U_{i-1}; \theta) = p_\Delta(U_i|V_{i-1}, U_{i-1}; \theta)$, i.e. the transition density. In this case, the weight reduces to $w_i\left(U_{0:i}^{(k)}\right) = p_\Delta(V_i|U_i^{(k)}, V_{i-1}, U_{0:i-1}^{(k)}; \theta)$. A second choice for the proposal is $q(U_i|V_i, V_{i-1}, U_{i-1}; \theta) = p_\Delta(U_i|V_i, V_{i-1}, U_{i-1}; \theta)$, i.e. the conditional distribution. In this case, the weight reduces to $w_i\left(U_{0:i}^{(k)}\right) = p_\Delta(V_i|V_{i-1}, U_{0:i-1}^{(k)}; \theta)$. Transition densities and conditional distributions are detailed in Appendix A for the approximate model (3). When the two Brownian motions are independent, as we assume, the two choices are equivalent.

We can compare this particle filter with previous filters proposed in the literature. Del Moral *et al.* (2001) propose a particle filter for a two-dimensional SDE, where the second equation is autonomous. Although the first coordinate is observed at discrete times, Del Moral *et al.* (2001) propose to simulate the $V_{0:n}$ at each iteration of the filter with the proposal $p_\Delta(V_i, U_i|V_{i-1}, U_{i-1})$. The weights are computed with a bounded function of the difference between the observed value V_i and the simulated particles $V_i^{(k)}$. Fearnhead *et al.* (2008) generalise this particle filter with any proposal distribution to simulate the V_i at each iteration. The weights then reduce to a Dirac mass of the difference between the observed value V_i and the simulated particles $V_i^{(k)}$, which is likely to be almost surely equal to zero. To avoid this problem of zero weights, the SMC algorithm proposed here is slightly different as $V_{0:n}$ is not re-simulated.

The convergence of Algorithm 1 is studied in Appendix D.

3.2 SAEM algorithm

The EM algorithm (Dempster *et al.*, 1977) is useful in situations where the direct maximization of the marginal likelihood $\theta \to p_\Delta(V_{0:n}\ ;\theta)$ is more difficult than the maximization of the conditional expectation of the complete likelihood

$$Q(\theta|\theta') = \mathbb{E}_\Delta\left[\log p_\Delta(V_{0:n}, U_{0:n};\theta)|V_{0:n};\theta'\right],$$

where $p_\Delta(V_{0:n}, U_{0:n};\theta)$ is the likelihood of the complete data $(V_{0:n}, U_{0:n})$ of the approximate model (3) and the expectation is under the conditional distribution of $U_{0:n}$ given $V_{0:n}$ with density $p_\Delta(U_{0:n}|V_{0:n};\theta')$. The EM algorithm is an iterative procedure: at the mth iteration, given the current value $\widehat{\theta}_{m-1}$ of the parameter, the E-step is the evaluation of $Q_m(\theta) = Q(\theta\,|\,\widehat{\theta}_{m-1})$, while the M-step updates $\widehat{\theta}_{m-1}$ by maximizing $Q_m(\theta)$. To fulfill convergence conditions of the algorithm, we consider the particular case of a distribution from an exponential family. More precisely, we assume:

(M1) The parameter space Θ is an open subset of $\mathbb{R}^p$. The complete likelihood $p_\Delta(V_{0:n}, U_{0:n};\theta)$ belongs to a curved exponential family, i.e.

$$\log p_\Delta(V_{0:n}, U_{0:n};\theta) = -\psi(\theta) + \langle S(V_{0:n}, U_{0:n}), \nu(\theta)\rangle\,,$$

where ψ and ν are two functions of θ, $S(V_{0:n}, U_{0:n})$ is known as the minimal sufficient statistic of the complete model, taking its value in a subset $\mathcal{S}$ of $\mathbb{R}^d$, and $\langle\cdot,\cdot\rangle$ is the scalar product on $\mathbb{R}^d$.

The approximate Morris-Lecar model (3) satisfies this assumption when we assume the scaling parameters V_1, V_2, V_3, V_4 known. Details of the sufficient statistic S are given in Appendix B.

Under assumption (M1), the E-step reduces to the computation of $\mathbb{E}_\Delta\left[S(V_{0:n}, U_{0:n})|V_{0:n};\widehat{\theta}_{m-1}\right]$. When this expectation has no closed form, Delyon *et al.* (1999) propose the Stochastic Approximation EM algorithm (SAEM) replacing the E-step by a stochastic approximation of $Q_m(\theta)$. The E-step is then divided into a simulation step (S-step) of the non-observed data $(U_{0:n}^{(m)})$ with the conditional density $p_\Delta(U_{0:n}\ |V_{0:n};\widehat{\theta}_{m-1})$ and a stochastic approximation step (SA-step) of $\mathbb{E}_\Delta\left[S(V_{0:n}, U_{0:n})|V_{0:n};\widehat{\theta}_{m-1}\right]$. We write s_m for the approximation of this expectation. Iterations of the SAEM algorithm are written as follows:

***Algorithm* 2** (SAEM algorithm)

- Iteration 0: *initialization of* $\widehat{\theta}_0$ *and set* $s_0 = 0$.
- Iteration $m \geq 1$:

S-Step: *simulation of the non-observed data* $(U_{0:n}^{(m)})$ *conditionally on the observed data from the distribution* $p_\Delta(U_{0:n}|V_{0:n};\widehat{\theta}_{m-1})dU_{0:n}$.

SA-Step: *update* s_{m-1} *using the stochastic approximation scheme:*

$$s_m = s_{m-1} + a_{m-1}\left[S(V_{0:n}, U_{0:n}^{(m)}) - s_{m-1}\right] \tag{5}$$

where $(a_m)_{m\in\mathbb{N}}$ *is a sequence of positive numbers decreasing to zero.*

M-Step: *update of* $\widehat{\theta}_m$ *by:*

$$\widehat{\theta}_m = \arg\max_{\theta\in\Theta} \left(-\psi(\theta) + \langle s_m, \nu(\theta)\rangle\right).$$

At the S-step, the simulation under the smoothing distribution $p_\Delta(U_{0:n}\ |V_{0:n};\widehat{\theta}_{m-1})dU_{0:n}$ is done by SMC, as explained in Section 3.1. We call this algorithm the SAEM-SMC algorithm. The standard errors of the parameter estimators can be evaluated through the Fisher information matrix. Details are given in Appendix C.

An advantage of the SAEM algorithm is the low-level dependence on the initialization $\widehat{\theta}_0$, due to the stochastic approximation of the E-step. The other advantage of the SAEM algorithm over a Monte-Carlo EM algorithm is the computational time. Indeed, only one simulation of the hidden variables $U_{0:n}$ is needed in the simulation step while an increasing number of simulated hidden variables is required in a Monte-Carlo EM algorithm.

3.3 Convergence of the SAEM-SMC algorithm

The SAEM algorithm we propose in this paper is based on an approximate simulation step performed with an SMC algorithm. We prove that even if this simulation is not exact, the SAEM algorithm still converges towards the maximum of the likelihood of the approximated diffusion (3). This is true because the SMC algorithm has good convergence properties.

Let us be more precise. We introduce a set of convergence assumptions which are the classic ones for the SAEM algorithm (Delyon *et al.*, 1999).

(M2) The functions $\psi(\theta)$ and $\nu(\theta)$ are twice continuously differentiable on Θ.

(M3) The function $\bar{s} : \Theta \longrightarrow \mathcal{S}$ defined by $\bar{s}(\theta) = \int S(v,u)p_\Delta(u|v;\theta)dv\,du$ is continuously differentiable on Θ.

(M4) The function $\ell_\Delta(\theta) = \log p_\Delta(v,u,\theta)$ is continuously differentiable on Θ and

$$\partial_\theta \int p_\Delta(v,u;\theta)dv\,du = \int \partial_\theta p_\Delta(v,u;\theta)dv\,du.$$

(M5) Define $L : \mathcal{S}\times\Theta \longrightarrow \mathbb{R}$ by $L(s,\theta) = -\psi(\theta) + \langle s, \nu(\theta)\rangle$. There exists a function $\hat{\theta} : \mathcal{S} \longrightarrow \Theta$ such that

$$\forall\theta\in\Theta, \quad \forall s\in\mathcal{S}, \quad L(s,\hat{\theta}(s)) \geq L(s,\theta).$$

(SAEM1) The positive decreasing sequence of the stochastic approximation $(a_m)_{m\geq 1}$ is such that $\sum_m a_m = \infty$ and $\sum_m a_m^2 < \infty$.

(SAEM2) $\ell_\Delta : \Theta \to \mathbb{R}$ and $\hat{\theta} : \mathcal{S} \to \Theta$ are d times differentiable, where d is the dimension of $S(v,u)$.

(SAEM3) For all $\theta \in \Theta$, $\int ||S(v,u)||^2 p_\Delta(u|v;\theta)du < \infty$ and the function $\Gamma(\theta) = Cov_\theta(S(\cdot,U_{0:n}))$ is continuous, where the covariance is under the conditional distribution $p_\Delta(U_{0:n}|V_{0:n};\theta)$.

(SAEM4) For any positive Borel function f

$$\mathbb{E}_\Delta(f(U_{0:n}^{(m+1)})|\mathcal{F}_m) = \int f(u)p_\Delta(u|v,\widehat{\theta}_m)du,$$

where $\{\mathcal{F}_m\}$ is the increasing family of σ-algebras generated by the random variables $s_0, U_{0:n}^{(1)}$, $U_{0:n}^{(2)}, \ldots, U_{0:n}^{(m)}$.

Assumptions (M1)-(M5) ensure the convergence of the EM algorithm when the E-step is exact (Delyon *et al.*, 1999). Assumptions (M1)-(M5) and (SAEM1)-(SAEM4) together with the additional assumption that $(s_m)_{m\geq 0}$ takes its values in a compact subset of $\mathcal{S}$ ensure the convergence of the SAEM estimates to a stationary point of the observed likelihood $p_\Delta(V_{0:n};\theta)$ when the simulation step is exact (Delyon *et al.*, 1999).

Here the simulation step is not exact and we have three additional assumptions on the SMC algorithm to bound the error induced by this algorithm and prove the convergence of the SAEM-SMC algorithm.

(SMC1) The number of particles K used at each iteration of the SAEM algorithm varies along the iteration: there exists a function $g(m) \to \infty$ when $m \to \infty$ such that $K(m) \geq g(m)\log(m)$.

(SMC2) The function S is bounded uniformly in u.

(SMC3) The functions $p_\Delta(V_i|U_i, V_{i-1}, U_{i-1};\theta)$ are bounded uniformly in θ.

Theorem 1. *Assume that (M1)-(M5), (SAEM1)-(SAEM3), and (SMC1)-(SMC3) hold. Then, with probability 1,* $\lim_{m\to\infty} d(\widehat{\theta}_m, \mathcal{L}) = 0$ *where* $\mathcal{L} = \{\theta \in \Theta, \partial_\theta \ell_\Delta(\theta) = 0\}$ *is the set of stationary points of the log-likelihood* $\ell_\Delta(\theta) = \log p_\Delta(V_{0:n};\theta)$.

Theorem 1 is proved in Appendix D. Note that assumption (SAEM4) is not needed thanks to the conditional independence of the particles generated by the SMC algorithm, as detailed in the proof. Similarly, the additional assumption that $(s_m)_{m\geq 0}$ takes its values in a compact subset of $\mathcal{S}$ is not needed, as it is directly satisfied under assumption (SMC2).

We deduce that the SAEM algorithm converges to a (local) maximum of the likelihood under standard additional assumptions (LOC1)-(LOC3) proposed by Delyon *et al.* (1999) on the regularity of the log-likelihood $\ell_\Delta(V_{0:n};\theta)$ that we do not recall here.

Corollary 1. *Under the assumptions of Theorem 1 and additional assumptions (LOC1)-(LOC3), the sequence* $\widehat{\theta}_m$ *converges with probability 1 to a (local) maximum of the likelihood* $p_\Delta(V_{0:n};\theta)$.

Proof is given in Appendix D.

In practice, the SAEM algorithm is implemented with a fixed number of particles, although an increasing number is needed to prove the theoretical convergence. Assumption (SMC1) provides the order of magnitude of the number of particles needed to obtain satisfactory results for a given number of iterations of the SAEM algorithm.

Assumptions (M1)-(M5) are classically satisfied. Assumption (SAEM1) is easily satisfied by choosing properly the sequence (a_m). Assumptions (SAEM2) and (SAEM3) depend on the regularity of the model. They are satisfied for the approximate Morris-Lecar model.

Assumption (SMC2) is satisfied for the approximate Morris-Lecar model because the variables U are bounded between 0 and 1 and the variables V are fixed at their observed values. This would not have been the case with the filter of Del Moral *et al.* (2001), which resimulates the variables V at each iteration. Assumption (SMC3) is satisfied if we require that γ is strictly bounded away from zero; $\gamma \geq \epsilon > 0$.

3.4 Properties of the approximate diffusion

The SAEM-SMC algorithm provides a sequence which converges to the set of stationary points of the log-likelihood $\ell_\Delta(\theta) = \log p_\Delta(V_{0:n}; \theta)$. The following result aims at comparing this likelihood, which corresponds to the Euler approximate model (3), with the true likelihood $p(V_{0:n}; \theta)$ in (2).

The result is based on the bound of the Euler approximation which has been proved by Bally and Talay (1996a). Their result holds under the following assumption

(H1) Functions f, b, σ are 2 times differentiable with bounded derivatives with respect to u and v of all orders up to 2.

Let us assume we apply the SAEM algorithm on an approximate model obtained with an Euler scheme of step size $\delta = \Delta/L$. Then we have the following result

Theorem 2. *Under assumption (H1), there exists a constant C, independent of θ, such that for any $\theta \in \Theta$, and any vector $V_{0:n}$*

$$|p(V_{0:n}; \theta) - p_\delta(V_{0:n}; \theta)| \leq C \frac{1}{L} n\Delta$$

Proof is given in Appendix D.

Assumption (H1) does not hold for the Morris-Lecar model, but if the state variables were bounded, assumption (H1) would hold. This is the case for U_t, whereas V_t is not. Nevertheless, the tails of V_t behave as an Ornstein-Uhlenbeck process, and could be truncated at an arbitrary large value, and the behavior would be approximately the same.

4 Simulation study

Parameter values of the Morris-Lecar model used in the simulations are the same as those of Tateno and Pakdaman (2004) for a class II membrane, except that we set the membrane capacitance constant to $C = 1\,\mu\text{F/cm}^2$, which is the standard value reported in the literature. Conductances and input current were correspondingly changed, and thus, the two models are the same. The values are: $V_K = -84$ mV, $V_L = -60$ mV, $V_{Ca} = 120$ mV, $C = 1\,\mu\text{F/cm}^2$, $g_L = 0.1\,\mu\text{S/cm}^2$, $g_{Ca} = 0.22\,\mu\text{S/cm}^2$, $g_K = 0.4\,\mu\text{S/cm}^2$, $V_1 = -1.2$ mV, $V_2 = 18$ mV, $V_3 = 2$ mV, $V_4 = 30$ mV, $\phi = 0.04$ ms^{-1}. Input is chosen to be $I = 4.5\,\mu\text{A/cm}^2$. Initial conditions of the Morris-Lecar model are $V_{t_0} = -26$ mV, $U_{t_0} = 0.2$. The volatility parameters are $\gamma = 1$ mV ms$^{-1/2}$, $\sigma = 0.03$ ms$^{-1/2}$. Trajectories are simulated with time step $\Delta = 0.1$ ms and $n = 2000$ points, and either $\theta = (g_{Ca}, g_K, g_L, V_{Ca}, V_K, I, \gamma, \sigma)$ or $\theta = (g_{Ca}, g_K, g_L, V_{Ca}, V_K, I, \gamma, \sigma, \phi)$ are estimated on each simulated trajectory. A hundred repetitions are

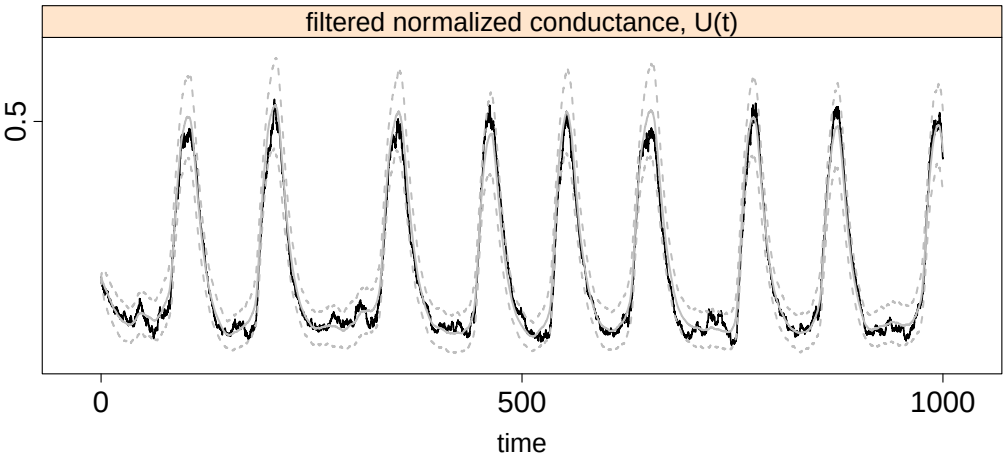

Figure 2: Filtering of (U_t) with the particle filter algorithm (200 particles): hidden simulated trajectory of the Morris-Lecar model (U_t) (black), mean filtered signal (grey full drawn line), 95% confidence interval of filtered signal (grey dashed lines).

used to evaluate the performance of the estimators. An example of a simulated trajectory (for $n = 10000$) is given in Figure 1.

4.1 Filtering results

The Particle filter aims at filtering the hidden process (U_t) from the observed process (V_t). We illustrate its performance on a simulated trajectory, with parameter θ fixed at its true values. The SMC Particle filter algorithm is implemented with $K = 100$ particles and the transition density as proposal. Figure 2 presents the result of the Particle filter procedure. The true hidden process, the mean filtered signal and its 95% confidence interval are plotted. The filtered process appears satisfactory. The confidence interval includes the true hidden process (U_t).

4.2 Estimation results

The performance of the SAEM-SMC algorithm is illustrated on 100 simulated trajectories. The SAEM algorithm is implemented with $m = 100$ iterations and a sequence (a_m) equal to 1 during the 30 first iterations and equal to $a_m = 1/(m - 30)^{0.8}$ for $m > 30$. The SMC algorithm is implemented with $K = 50$ particles at each iteration of the SAEM algorithm. Two types of initialization of the SAEM algorithm are used. The first is a random initialization of $\widehat{\theta}_0$ centered around the true value: $\widehat{\theta}_0 = \theta_{true} + \theta_{true}/3\mathcal{N}(0, 1)$. The second type is a random initialization of $\widehat{\theta}_0$ not centered around the true value: $\widehat{\theta}_0 = \theta_{true} + 0.1 + \theta_{true}/3\mathcal{N}(0, 1)$.

An example of the convergence of the SAEM algorithm for one of the iterations is presented in Fig. 3. It is seen that the algorithm converges for most of the parameters in very few iterations to a neighborhood

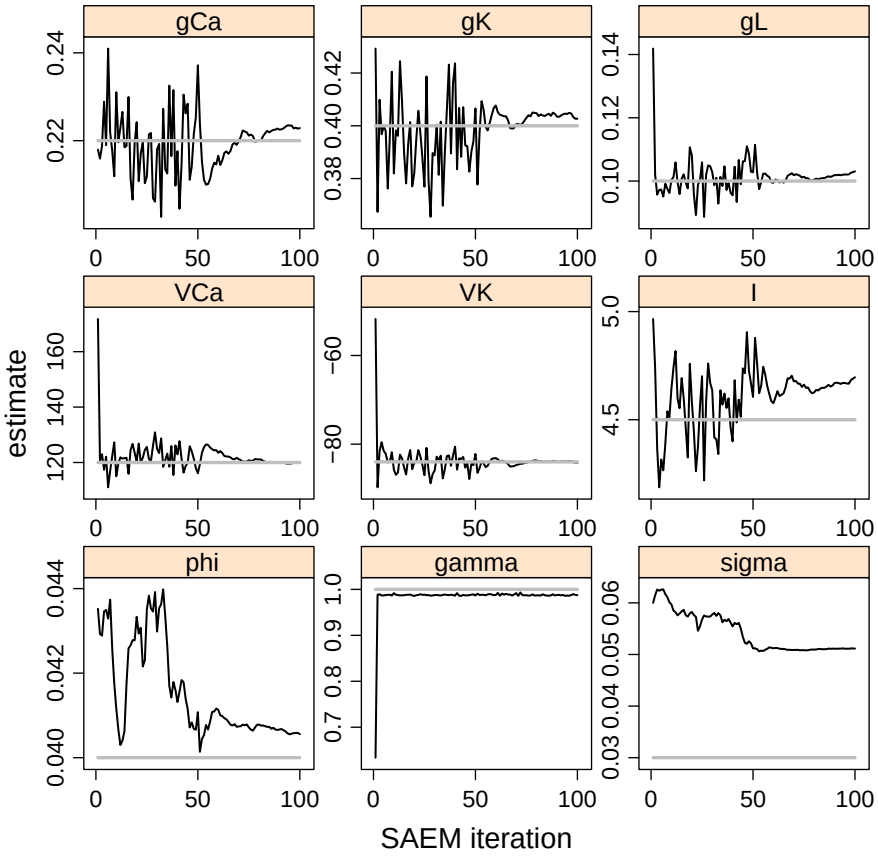

Figure 3: Convergence of the SAEM algorithm for the 9 estimated parameters on a simulated data set. True values used in the simulation are given by the gray lines.

of the true value, even if the initial values are far from the true ones. Only for ϕ and σ more iterations are needed, which is expected since these two parameters appear in the second, non-observed coordinate.

The SAEM algorithm is used to estimate either $\theta = (g_{Ca}, g_K, g_L, V_{Ca}, V_K, I, \gamma, \sigma)$ (ϕ fixed at the true value) or $\theta = (g_{Ca}, g_K, g_L, V_{Ca}, V_K, I, \gamma, \sigma, \phi)$. The SAEM estimators are compared with the pseudo maximum likelihood estimator obtained if both V_t and U_t were observed. Results are given in Table 1. The parameters are well estimated in this ideal case. When only V_t is observed, the initialization of the SAEM algorithm has low influence on the results. The parameters are well estimated, except the parameter σ which is biased. The estimation of ϕ, which is the only parameter in the drift of the hidden coordinate U_t, is good and does not deteriorate the estimation of the other parameters. In Fig. 4 we show boxplots of the estimates of the nine parameters for the three estimation settings; both coordinates observed, or only one observed with either ϕ fixed at the true value, or also estimated. All parameters appear well estimated, except for σ, which is only well estimated when both coordinates are observed. As expected, the variances of the estimators of ϕ and σ hugely increase when only one coordinate is observed, but interestingly, the variance of the parameters of the observed coordinate do not seem much affected by this loss of information.

	Parameters								
Estimator	g_L	g_{Ca}	g_K	σ	γ	V_K	ϕ	V_{Ca}	I
true values	0.100	0.220	0.400	0.030	1.00	-84.00	0.040	120.00	4.400
With both V_t and U_t observed (pseudo maximum likelihood estimator)									
mean	0.101	0.219	0.411	0.030	0.996	-83.20	0.040	121.97	4.539
RMSE	0.017	0.019	0.041	0.001	0.019	7.61	0.001	8.50	0.560
With only V_t observed (SAEM estimator)									
ϕ fixed at the true value, $\widehat{\theta}_0$ centered at the true value									
mean	0.093	0.226	0.440	0.060	1.004	-77.01	–	119.61	3.884
RMSE	0.016	0.022	0.067	0.031	0.017	9.80	–	10.00	0.836
ϕ estimated, $\widehat{\theta}_0$ centered around the true value									
mean	0.092	0.213	0.414	0.058	1.000	-85.72	0.046	123.90	4.550
RMSE	0.021	0.022	0.124	0.029	0.019	10.19	0.018	10.90	1.714
ϕ estimated, $\widehat{\theta}_0$ not centered around the true value									
mean	0.090	0.225	0.464	0.059	1.003	-78.622	0.041	119.677	4.060
RMSE	0.021	0.024	0.144	0.029	0.017	9.459	0.013	10.218	1.028
estimated									
SE	0.016	0.019	0.042	0.001	0.016	4.96	0.001	7.31	0.561

Table 1: Simulation results obtained from 100 simulated Morris-Lecar trajectories ($n = 2000$, $\Delta = 0.1$ ms). Four estimators are compared: 1/ The pseudo maximum likelihood estimator in the ideal case where both V_t and U_t are observed; 2/ SAEM estimator when only V_t is observed with the SAEM initialization at a random value centered around the true value θ, ϕ fixed at the true value; 3/ SAEM estimator when only V_t is observed with the SAEM initialization at a random value centered around the true value θ, ϕ estimated; 4/ SAEM estimator when only V_t is observed with the SAEM initialization at a random value not centered around the true value θ. An example of standard errors (SE) estimated with the SAEM-SMC algorithm on one single simulated dataset is also given.

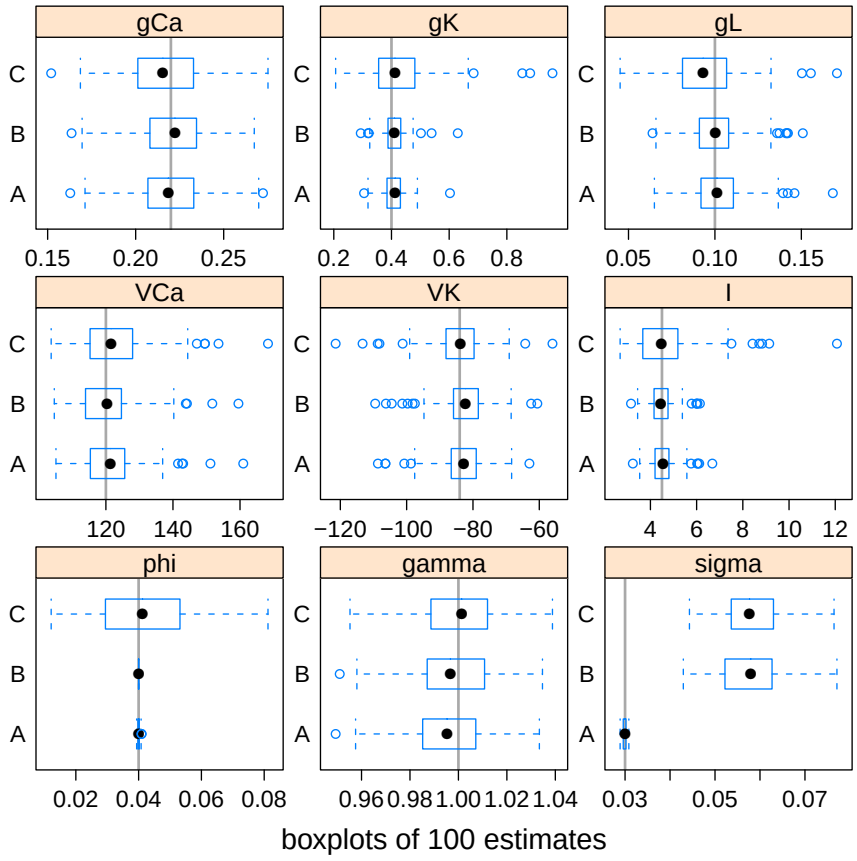

Figure 4: Boxplots of 100 estimates from simulated data sets for the 9 parameters. True values used in the simulations are given by the gray lines. A: Both V_t and U_t are observed. B: Only V_t is observed, ϕ is fixed at the true value. C: Only V_t is observed, ϕ is also estimated.

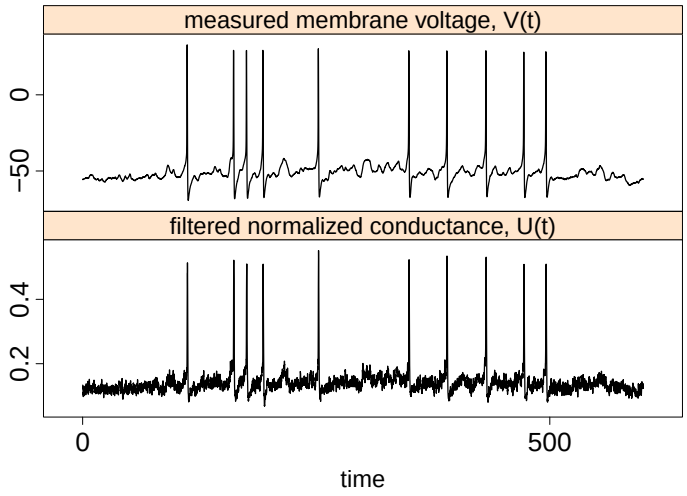

Figure 5: Observations of the membrane potential in a spinal motoneuron of an adult red-eared turtle during 600 ms (upper panel), and the filtered hidden process of the normalized conductance associated with K^+ current (lower panel) for the estimated parameters with the scaling parameters fixed at $V_1 = -2.4$ mV, $V_2 = 36$ mV, $V_3 = 4$ mV and $V_4 = 60$ mV.

The SAEM-SMC algorithm provides estimates of the standard errors (SE) of the estimators (see Appendix C). These should be close to the RMSE obtained from the 100 simulated datasets. As an example, the SE for one dataset estimated by SAEM are reported in the last line of Table 1. The estimated SE are satisfactory for most of the parameters, but tends to underestimate. The worst SE estimate is the one corresponding to σ. This might be explained by the fact that this parameter is estimated with bias.

5 Intracellular recordings from a turtle motoneuron

The membrane potential from a spinal motoneuron in segment D10 of an adult red-eared turtle (*Trachemys scripta elegans*) was recorded while a periodic mechanical stimulus was applied to selected regions of the carapace with a sampling step of 0.1 ms (for details see Berg *et al.* (2007, 2008)). The turtle responds to the stimulus with a reflex movement of a limb known as the *scratch reflex*, causing an intense synaptic input to the recorded neuron. Due to the time varying stimulus, a model for the complete data set needs to incorporate the time-inhomogeneity, as done in Jahn *et al.* (2011). The data can only be assumed stationary during short time windows, which is required for the Morris-Lecar model with constant parameters. Therefore, we only analyze a short trace where the input is approximately constant during an on-cycle (following Jahn *et al.* (2011)). The analyzed data are plotted in Fig. 5, together with a filtered trace of the unobserved coordinate.

First the model was fitted with the values of the scaling parameters V_1–V_4 as in Section 4. Most of the estimates are reasonable and in agreement with the expected order of magnitudes for the parameter values, except for the V_{Ca} reversal potential, which in the literature is reported to be around 100–150

Parameter	g_L	g_{Ca}	g_K	σ	γ	V_K	ϕ	V_{Ca}	I
With $V_1 = -1.2$ mV, $V_2 = 18$ mV, $V_3 = 2$ mV, $V_4 = 30$ mV									
Estimate	-0.962	9.555	7.280	0.091	3.003	-89.862	2.916	44.705	-2.511
SE	0.001	0.049	0.051	0.000	0.001	4.684	0.002	0.597	0.089
With $V_1 = -2.4$ mV, $V_2 = 36$ mV, $V_3 = 4$ mV, $V_4 = 60$ mV									
Estimate	1.292	11.564	18.631	0.095	2.694	-65.582	2.683	106.368	-65.162
SE	–	0.028	0.081	–	0.025	0.434	0.120	0.440	0.506

Table 2: Parameter estimates obtained from observations of the membrane potential of a spinal motoneuron of an adult red-eared turtle during 600 ms for two different sets of scaling parameters.

mV (estimated to 44.7 mV), and the leak conductance, which is estimated to be negative. Conductances are always non-negative. This is probably due to wrong choices of the scaling constants V_1–V_4. For the parameters of the model given in Section 4, the average of the membrane potential V_t between spikes is around -26 mV, whereas the average of the experimental trace between spikes is around -56 mV, a factor two larger. We therefore rerun the estimation procedure fixing V_1–V_4 to twice the value from before, which provides approximately the same equilibrium values of the normalized Ca^{2+} conductance, $m_\infty(\cdot)$, and the rates of opening and closing of K^+ ion channels, $\alpha(\cdot)$ and $\beta(\cdot)$, as in the theoretical model. Results are presented in Table 2. In this case all parameters are reasonable and in agreement with the expected order of magnitudes. In Fig. 6 the convergence of the SAEM algorithm is presented. As in the simulated data examples, it is seen that the algorithm converges for the parameters of the observed coordinate in very few iterations to a neighborhood of some value. Only for the parameters of the unobserved coordinate, ϕ and σ, more iterations are needed. For two parameters, g_L and σ, the estimated variances were negative, but very small in absolute values. This can be due to numerical instabilities, and should be interpreted as being close to zero. The SEs are probably underestimated, though, as shown in the simulation study.

6 Discussion

To the authors knowledge, this is the first time the rate parameter of the unobserved coordinate, ϕ, is estimated from experimental data. It is comforting to observe that the estimated value do not seem to be very sensitive to the choice of scaling parameters. Other parameters, like the conductances and the reversal potentials, are more sensitive to this choice, and should be interpreted with care.

The estimation procedure builds on the pseudo likelihood, which approximates the true likelihood by an Euler scheme. This approximation is only valid for small sampling step, i.e. for high frequency data, which is the case for the type of neuronal data considered here. If data were sampled less often, a possibility could be to simulate diffusion bridges between the observed points, and apply the estimation procedure to an augmented data set consisting of the observed data and the imputed values.

There are several issues that deserve further study. First, it is important to understand the influence of the scaling parameters $V_1 - V_4$, and how to estimate them for a given data set. The model is not exponential in these parameters (assumption (M1)) and new estimation procedures have to be considered. Secondly, one should be aware of the possible misspecification of the model. More detailed models incorporating further types of ion channels could be explored, but increasing the model complexity might

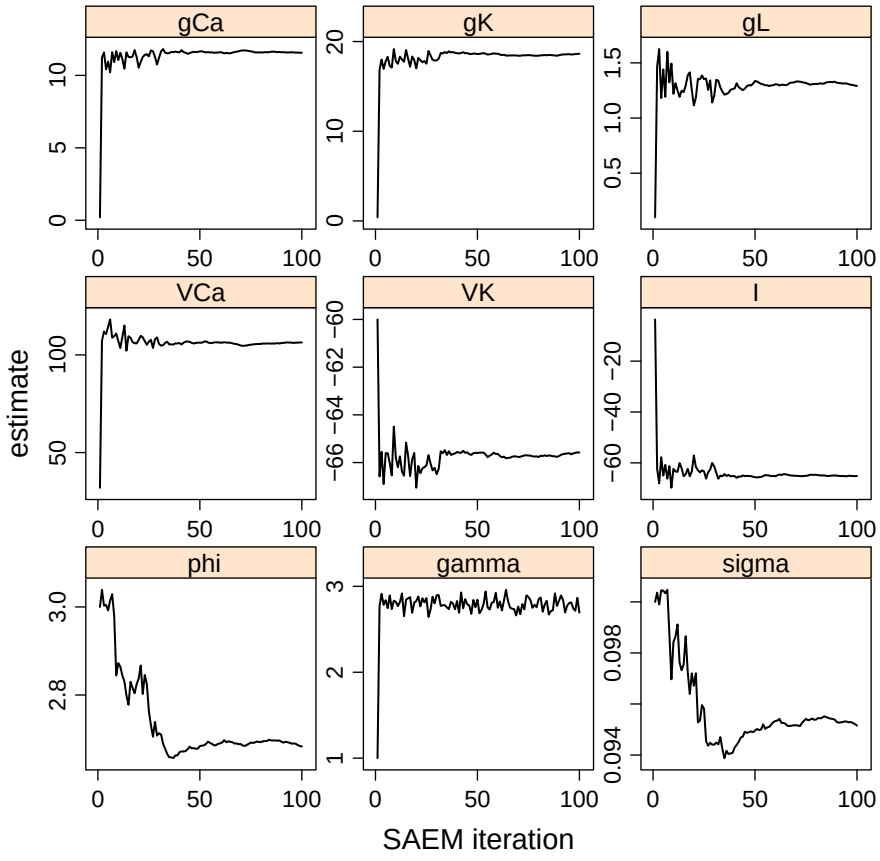

Figure 6: Convergence of the SAEM algorithm for the 9 estimated parameters on the experimental data set consisting in observations of the membrane potential of a spinal motoneuron of an adult red-eared turtle during 600 ms.

deteriorate the estimates, since the information contained in only observing the membrane potential is limited. Furthermore, the sensitivity on the choice of tuning parameters of the algorithm, like the decreasing sequence of the stochastic approximation, (a_m), number of SAEM iterations, and the number of particles in the SMC algorithm, needs further investigation. Finally, an automated procedure to find starting values for the procedure is warranted.

Acknowledgments

The authors are grateful to Rune W. Berg for making his experimental data available. S. Ditlevsen is supported by grants from the Danish Council for Independent Research | Natural Sciences. A. Samson is supported by Grants from the University Paris Descartes PCI.

Appendix

A Distributions of approximate model (3)

Consider the general approximate model

$$\begin{pmatrix} V_{i+1} \\ U_{i+1} \end{pmatrix} = \begin{pmatrix} V_i \\ U_i \end{pmatrix} + \Delta \begin{pmatrix} f(V_i, U_i) \\ b(V_i, U_i) \end{pmatrix} + \sqrt{\Delta} \begin{pmatrix} \gamma & \rho \\ \rho & \sigma(V_i, U_i) \end{pmatrix} \begin{pmatrix} \tilde{\eta}_i \\ \eta_i \end{pmatrix}$$

where ρ is the correlation coefficient between the two Brownian motions or perturbations.

The distribution of the couple (V_{i+1}, U_{i+1}) conditionally on (V_i, U_i) is

$$\begin{pmatrix} V_{i+1} \\ U_{i+1} \end{pmatrix} \Bigg| \begin{pmatrix} V_i \\ U_i \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} V_i + \Delta f(V_i, U_i) \\ U_i + \Delta b(V_i, U_i) \end{bmatrix}, \Delta \begin{bmatrix} (\gamma^2 + \rho^2) & \rho(\gamma + \sigma(V_i, U_i)) \\ \rho(\gamma + \sigma(V_i, U_i)) & (\sigma^2(V_i, U_i) + \rho^2) \end{bmatrix} \right)$$

The marginal distributions of V_{i+1} conditionally on (V_i, U_i) and U_{i+1} conditionally on (V_i, U_i) are

$$\begin{aligned} V_{i+1}|V_i, U_i &\sim \mathcal{N}\left(V_i + \Delta f(V_i, U_i), \Delta(\gamma^2 + \rho^2)\right) \\ U_{i+1}|V_i, U_i &\sim \mathcal{N}\left(U_i + \Delta b(V_i, U_i), \Delta(\sigma^2(V_i, U_i) + \rho^2)\right) \end{aligned} \tag{6}$$

The conditional distributions of V_{i+1} conditionally on (U_{i+1}, V_i, U_i) and U_{i+1} conditionally on (V_{i+1}, V_i, U_i) are

$$\begin{aligned} V_{i+1}|U_{i+1}, V_i, U_i &\sim \mathcal{N}\left(m_V, Var_V\right) \\ U_{i+1}|V_{i+1}, V_i, U_i &\sim \mathcal{N}\left(m_U, Var_U\right) \end{aligned} \tag{7}$$

where

$$\begin{aligned} m_V &= V_i + \Delta f(V_i, U_i) + \frac{\rho(\gamma + \sigma(V_i, U_i))}{\sigma^2(V_i, U_i) + \rho^2}(U_{i+1} - U_i - \Delta b(V_i, U_i)) \\ Var_V &= \Delta(\gamma^2 + \rho^2) - \frac{\Delta\rho^2(\gamma + \sigma(V_i, U_i))^2}{\sigma^2(V_i, U_i) + \rho^2} \\ m_U &= U_i + \Delta b(V_i, U_i) + \frac{\rho(\gamma + \sigma(V_i, U_i))}{\gamma^2 + \rho^2}(V_{i+1} - V_i - \Delta f(V_i, U_i)) \\ Var_U &= \Delta(\sigma^2(V_i, U_i) + \rho^2) - \frac{\Delta\rho^2(\gamma + \sigma(V_i, U_i))^2}{\gamma^2 + \rho^2} \end{aligned}$$

The distributions in (6) and (7) are equal when the Brownian motions are independent, i.e. when $\rho = 0$.

B Sufficient statistics of the approximate model (3)

We detail some sufficient statistics functions. Consider the $n \times 6$-matrix

$$X = \left(-V_{0:(n-1)}, -m_\infty(V_{0:(n-1)})V_{0:(n-1)}, -U_{0:(n-1)}V_{0:(n-1)}, U_{0:(n-1)}, \mathbf{1}, m_\infty(V_{0:(n-1)})\right)$$

where $\mathbf{1}$ is the vector of 1's of size n. Then the vector

$$S_1(V_{0:(n-1)}, U_{0:(n-1)}) = (X'X)^{-1}\, X'\, (V_{1:n} - V_{0:(n-1)})$$

is the sufficient statistic vector corresponding to the parameters $\nu_1(\theta) = (g_L, g_{Ca}, g_K, g_K V_K, g_L V_L + I, g_{Ca}V_{Ca})$, where $'$ denotes transposition.

The sufficient statistics corresponding to $\nu_2(\theta) = 1/\gamma^2$ are

$$\sum_{i=1}^{n} (V_i - V_{i-1})\, U_{i-1}, \quad \sum_{i=1}^{n} U_{i-1}^2, \quad \sum_{i=1}^{n} (V_i - V_{i-1})\, V_{i-1} m_\infty(V_{i-1}),$$

$$\sum_{i=1}^{n} (V_i - V_{i-1})\, U_{i-1}V_{i-1}, \quad \sum_{i=1}^{n} U_{i-1}^2 V_{i-1}^2.$$

The sufficient statistics corresponding to $\nu_3(\theta) = 1/\sigma^2$ are

$$\sum_{i=1}^{n} (U_i - U_{i-1})^2, \quad \sum_{i=1}^{n} (U_i - U_{i-1})\left(\alpha(V_{i-1})(1 - U_{i-1})/\phi - \beta(V_{i-1})U_{i-1}/\phi\right),$$

$$\sum_{i=1}^{n} \left(\alpha(V_{i-1})(1 - U_{i-1})/\phi - \beta(V_{i-1})U_{i-1}/\phi\right)^2.$$

The sufficient statistics corresponding to ϕ is also explicit but more complex and not detailed here.

C Fisher information matrix

The standard errors (SE) of the parameter estimators can be evaluated as the diagonal elements of the inverse of the Fisher information matrix estimate. Its evaluation is difficult because it has no analytic form. We adapt the estimation of the Fisher information matrix, proposed by Delyon *et al.* (1999) and based on the Louis missing information principle.

The Hessian of the log-likelihood $\ell_\Delta(\theta)$ can be expressed as:

$$\begin{aligned}\partial^2_\theta \ell_\Delta(\theta) &= \mathbb{E}\left[\partial^2_\theta L(S(V_{0:n},U_{0:n}),\theta)|V_{0:n},\theta\right] \\ &+ \mathbb{E}\left[\partial_\theta L(S(V_{0:n},U_{0:n}),\theta)\ (\partial_\theta L(S(V_{0:n},U_{0:n}),\theta))'|V_{0:n},\theta\right] \\ &- \mathbb{E}\left[\partial_\theta L(S(V_{0:n},U_{0:n}),\theta)|V_{0:n},\theta\right]\ \mathbb{E}\left[\partial_\theta L(S(V_{0:n},U_{0:n}),\theta)|V_{0:n},\theta\right]'.\end{aligned}$$

The derivatives $\partial_\theta L(S(V_{0:n},U_{0:n}),\theta)$ and $\partial^2_\theta L(S(V_{0:n},U_{0:n}),\theta)$ are explicit for the Euler approximation of the Morris-Lecar model. Therefore we implement their estimation using the stochastic approximation procedure of the SAEM algorithm. At the mth iteration of the algorithm, we evaluate the three following quantities:

$$\begin{aligned}G_{m+1} &= G_m + a_m\left[\partial_\theta L(S(V_{0:n},U^{(m)}_{0:n}),\theta) - G_m\right] \\ H_{m+1} &= H_m + a_m\left[\partial^2_\theta L(S(V_{0:n},U^{(m)}_{0:n}),\theta) + \partial_\theta L(S(V_{0:n},U^{(m)}_{0:n}),\theta)\ (\partial_\theta L(S(V_{0:n},U^{(m)}_{0:n}),\theta))' - H_m\right] \\ F_{m+1} &= H_{m+1} - G_{m+1}\ (G_{m+1})'.\end{aligned}$$

As the sequence $(\hat\theta_m)_m$ converges to the maximum of the likelihood, the sequence $(F_m)_m$ converges to the Fisher information matrix.

D Proof of the convergence results

We start by a Lemma which generalizes the result of Del Moral *et al.* (2001) to the particle filter we propose. Then we prove Theorem 1, Corollary 1 and Theorem 2.

D.1 Convergence results of Algorithm 1

Let us introduce some notation. For any bounded Borel function $f : \mathbb{R} \mapsto \mathbb{R}$, we denote $\pi_{n,\theta} f = \mathbb{E}_\Delta\left(f(U_n)|V_{0:n};\theta\right)$, the conditional expectation under the exact smoothing distribution $p_\Delta(U_{0:n}|V_{0:n};\theta)$ of the approximate model, and $\Psi^K_{n,\theta} f = \sum_{k=1}^K f(U_n^{(k)}) W_{n,\theta}(U^{(k)}_{0:n})$, the conditional expectation of f under the empirical measure $\Psi^K_{n,\theta}$ obtained by the SMC algorithm for a given value of θ.

The following lemma is an extension of the result of Del Moral *et al.* (2001) to a particle filter adapted to a non-autonomous equation for the second coordinate of the system and in which $V_{0:n}$ is not resimulated.

Lemma 1. *Under assumption (SMC3), for any $\varepsilon > 0$, and for any bounded Borel function f on $\mathbb{R}$, there exist constants C_1 and C_2, independent of θ, such that*

$$\mathbb{P}\left(\left|\Psi^K_{n,\theta} f - \pi_{n,\theta} f\right| \geq \varepsilon\right) \leq C_1 \exp\left(-K\frac{\varepsilon^2}{C_2\|f\|^2}\right) \tag{8}$$

where $\|f\|$ is the sup-norm of f.

Proof. We omit θ in the proof for clarity. The conditional expectation $\pi_n f$ can be written

$$\pi_n f \quad = \quad \frac{\int \mu(U_0) \prod_{i=1}^n p_\Delta(V_i, U_i | V_{i-1}, U_{i-1}) f(U_n) dU_0 \dots dU_n}{\int \mu(U_0) \prod_{i=1}^n p_\Delta(V_i, U_i | V_{i-1}, U_{i-1}) dU_0 \dots dU_n} \tag{9}$$

where μ is the distribution of U_0. We have

$$\pi_0 f = \int f(u) \mu(du).$$

Consider for $i = 1, \dots, n$, the kernels H_i from $\mathbb{R}$ into itself by

$$H_i f(u) = \int p_\Delta(V_i, u' | V_{i-1}, u) f(u') du'. \tag{10}$$

Then π_n can be expressed recursively by

$$\pi_n f = \frac{\pi_{n-1} H_n f}{\pi_{n-1} H_n 1}.$$

These kernels are extensions of the kernels considered by Del Moral *et al.* (2001). Note that the denominator of (9) is $\mu H_1 \cdots H_n 1 = p_\Delta(V_{0:n})$, which is different from 0 since it is normal, and bounded following from assumption (SMC3). We write $\nu_n = \mu H_1 \cdots H_n 1$ for this constant conditioned on the observed values $V_{0:n}$. Also (10) is bounded, i.e. $H_i 1(u) \leq C$ for all $u \in \mathbb{R}$ and $i = 1, \dots, n$, for some constant C. It directly follows that $\mu H_1 \cdots H_{i-1} 1 \leq C^{i-1}$. Furthermore, we obtain the bound

$$\mu H_1 \cdots H_i 1 \geq \frac{\mu H_1 \cdots H_{i+1} 1}{C} \geq \cdots \geq \frac{\nu_n}{C^{n-i}}.$$

Finally, using the above bounds and that π_{i-1} is a transition measure, we obtain

$$\frac{\nu_n}{C^{n-1}} \quad \leq \quad \pi_{i-1} H_i 1 \leq C. \tag{11}$$

Consider the SMC sampled particles in algorithm 1. Define the two empirical measures obtained at time i: $\Psi_i'^K = \frac{1}{K} \sum_{k=1}^K \mathbf{1}_{U_{0:i}^{'(k)}}$ and $\Psi_i^K = \sum_{k=1}^K W_i(U_{0:i}^{(k)}) \mathbf{1}_{U_{0:i}^{(k)}}$. We also decompose the weights and write $\Upsilon_i^K f = \frac{1}{K} \sum_{k=1}^K f(U_i^{(k)}) w_i(U_{0:i}^{(k)})$. Then $W_i(U_{0:i}^{(k)}) = w_i(U_{0:i}^{(k)}) / (K \Upsilon_i^K 1)$ and $\Psi_i^K f = \Upsilon_i^K f / \Upsilon_i^K 1$.

Recall the following general result (Del Moral *et al.*, 2001) for $\xi_1, \dots, \xi_K$ random variables, which conditioned on a σ-field $\mathcal{G}$ are independent, centered and bounded $|\xi_k| \leq a$. Then for any $\varepsilon > 0$ we have

$$\mathbb{P}\left(\left| \frac{1}{K} \sum_{k=1}^K \xi_k \right| \geq \varepsilon \right) \quad \leq \quad 2 \exp\left(-K \frac{\varepsilon^2}{2a^2} \right). \tag{12}$$

Let f be a bounded function on $\mathbb{R}$. Then under assumption (SMC3)

$$\Psi_i'^K f - \Psi_i^K f = \frac{1}{K} \sum_{k=1}^K \left(f(U_i^{'(k)}) - \Psi_i^K f \right) = \frac{1}{K} \sum_{k=1}^K \xi_k$$

fulfills the conditions for (12) to hold with $a = 2\|f\|$, since $\mathbb{E}(f(U_i^{'(k)})|\mathcal{G}) = \Psi_i^K f$, where $\mathcal{G}$ is the σ-algebra generated by $U_{0:i}^{(k)}$. Thus, for any $\varepsilon > 0$ we obtain

$$\mathbb{P}\left(\left|\Psi_i^{'K} f - \Psi_i^K f\right| \geq \varepsilon\right) \quad \leq \quad 2\exp\left(-K\frac{\varepsilon^2}{8\|f\|^2}\right). \tag{13}$$

Likewise, as $H_i f(u) = Q_i(fw_i)(u)$ with $Q_i(f)(u) = \int q(u'|V_i, V_{i-1}, u) f(u')du'$, we have

$$\Upsilon_i^K f - \Psi_{i-1}^{'K} H_i f = \frac{1}{K}\sum_{k=1}^{K}\left(f(U_i^{(k)})w_i(U_{0:i}^{(k)}) - Q_i(fw_i)(U_{i-1}^{'(k)})\right) = \frac{1}{K}\sum_{k=1}^{K}\xi_k$$

which fulfills the conditions for (12) to hold, now with $a = 2C\|f\|$ and $\mathcal{G}$ is the σ-algebra generated by $U_{0:i-1}^{'(k)}$. Hence, for any $\varepsilon > 0$ we obtain

$$\mathbb{P}\left(\left|\Upsilon_i^K f - \Psi_{i-1}^{'K} H_i f\right| \geq \varepsilon\right) \quad \leq \quad 2\exp\left(-K\frac{\varepsilon^2}{8C^2\|f\|^2}\right). \tag{14}$$

We want to show the following two bounds

$$\mathbb{P}\left(\left|\Psi_i^K f - \pi_i f\right| \geq \varepsilon\right) \quad \leq \quad 2I_i\exp\left(-K\frac{\varepsilon^2}{8J_i\|f\|^2}\right), \quad i = 1, \ldots, n \tag{15}$$

$$\mathbb{P}\left(\left|\Psi_i^{'K} f - \pi_i f\right| \geq \varepsilon\right) \quad \leq \quad 2I_i^{'}\exp\left(-K\frac{\varepsilon^2}{8J_i^{'}\|f\|^2}\right), \quad i = 0, 1, \ldots, n \tag{16}$$

by induction on i, for some constants $I_i, I_i^{'}, J_i, J_i^{'}$ increasing with i to be computed later. Note first that since $\pi_0 = \mu$ and $U_0^{'(k)}$ are i.i.d. with law μ, then (12) with $\xi_k = f(U_0^{'(k)}) - \mu(f)$ yields (16) for $i = 0$ with $I_i^{'} = J_i^{'} = 1$. Let $i \geq 1$ and assume (16) holds for $i - 1$. We can write

$$\Psi_i^K f - \pi_i f = \frac{1}{\pi_{i-1}H_i 1}\left(\frac{\Upsilon_i^K f}{\Upsilon_i^K 1}(\pi_{i-1}H_1 1 - \Upsilon_i^K 1) + (\Upsilon_i^K f - \pi_{i-1}H_i f)\right).$$

Note that $\Upsilon_i^K 1 > 0$ because the weights w_i are strictly positive. Define $L_i f = \Upsilon_i^K f - \pi_{i-1}H_i f$ and use that $|\Upsilon_i^K f| \leq \|f\|\Upsilon_i^K 1$ (because f is bounded) and (11) to see that

$$|\Psi_i^K f - \pi_i f| \quad \leq \quad \frac{C^{n-1}}{\nu_n}\left(\|f\||L_i 1| + |L_i f|\right) \tag{17}$$

and

$$|L_i f| \quad \leq \quad |\Upsilon_i^K f - \Psi_{i-1}^{'K} H_i f| + |\Psi_{i-1}^{'K} H_i f - \pi_{i-1}H_i f|. \tag{18}$$

Assuming that (16) holds for $i - 1$ and using (14) and that $\|H_i f\| \leq C\|f\|$ yield

$$\mathbb{P}\left(|L_i f| \geq \varepsilon\right) \quad \leq \quad 2\exp\left(-K\frac{\varepsilon^2}{32C^2\|f\|^2}\right) + 2I_{i-1}^{'}\exp\left(-K\frac{\varepsilon^2}{32J_{i-1}^{'}C^2\|f\|^2}\right).$$

We obtain

$$\begin{aligned}\mathbb{P}\left(\left|\Psi_i^K f - \pi_i f\right| \geq \varepsilon\right) \quad &\leq \quad \mathbb{P}\left(|L_i 1| \geq \frac{\varepsilon\nu_n}{2C^{n-1}\|f\|}\right) + \mathbb{P}\left(|L_i f| \geq \frac{\varepsilon\nu_n}{2C^{n-1}}\right) \\ &\leq \quad 4\exp\left(-K\frac{\varepsilon^2\nu_n^2}{128C^{2n}\|f\|^2}\right) + 4I_{i-1}^{'}\exp\left(-K\frac{\varepsilon^2\nu_n^2}{128J_{i-1}^{'}C^{2n}\|f\|^2}\right)\end{aligned}$$

Hence, (15) holds with $I_i \geq 2(1 + I'_{i-1})$ and $J_i \geq 16C^{2n}J'_{i-1}/\nu_n^2 \geq 16J'_{i-1}$ since $\nu_n \leq C^n$. By (13) and (15) we then conclude that (16) also holds for i if $I'_i = 1 + I_i$ and $J'_i = 4J_i$. These conditions are fulfilled by choosing $I_i = 3^{i+1} - 3$ and $J_i = 16^i$.

Thus, (8) holds with $C_1 = 6(3^n - 1)$ and $C_2 = 8 \cdot 16^n$. This concludes the proof. □

D.2 Proof of Theorem 1

Proof. To prove the convergence of the SAEM-SMC algorithm, we study the stochastic approximation scheme used during the SA step. The stochastic approximation (5) can be decomposed into:

$$s_{m+1} = s_m + a_m h(s_m) + a_m e_m + a_m r_m$$

with

$$\begin{aligned} h(s_m) &= \pi_{n,\hat{\theta}(s_m)} S - s_m \\ e_m &= S(V_{0:n}, U_{0:n}^{(m)}) - \Psi^{K(m)}_{n,\hat{\theta}(s_m)} S \\ r_m &= \Psi^{K(m)}_{n,\hat{\theta}(s_m)} S - \pi_{n,\hat{\theta}(s_m)} S \end{aligned}$$

where we denote by $\pi_{n,\hat{\theta}(s_m)} S = \mathbb{E}_\Delta(S(V_{0:n}, U_{0:n})|V_{0:n}; \theta)$ the expectation of the sufficient statistic S under the exact distribution $p_\Delta(U_{0:n}|V_{0:n}; \theta)$, and by $\Psi^{K(m)}_{n,\hat{\theta}(s_m)} S$ the expectation of the sufficient statistic S under the empirical measure obtained with the SMC algorithm with $K(m)$ particles and current value of parameters $\hat{\theta}(s_m)$ at iteration m of the SAEM-SMC algorithm.

Following Theorem 2 of Delyon *et al.* (1999) on the convergence of the Robbins-Monro scheme, the convergence of the SAEM-SMC algorithm is ensured if we prove the following assertions:

1. The sequence $(s_m)_{m\geq 0}$ takes its values in a compact set.
2. The function $V(s) = -\ell_\Delta(\hat{\theta}(s))$ is such that for all $s \in \mathcal{S}$, $F(s) = \langle \partial_s V(s), h(s) \rangle \leq 0$ and such that the set $V(\{s, F(s) = 0\})$ is of zero measure.
3. $\lim_{m\to\infty} \sum_{\ell=1}^m a_\ell e_\ell$ exists and is finite with probability 1.
4. $\lim_{m\to\infty} r_m = 0$ with probability 1.

Assertion 1 follows from assumption (SMC2) and by construction of s_m in formula (5).

Assertion 2 is proved by Lemma 2 of Delyon *et al.* (1999) under assumptions (M1)-(M5) and (SAEM2).

Assertion 3 is proved similarly as Theorem 5 of Delyon *et al.* (1999). By construction of the SMC algorithm, the equivalent of assumption (SAEM3) is checked for the expectation taken under the approximate empirical measure $\Psi^{K(m)}_{n;\widehat{\theta}_m}$. Indeed, the assumption of independence of the non-observed variables $U_{0:n}^{(1)}, \ldots, U_{0:n}^{(m)}$ given $\widehat{\theta}_0, \ldots, \widehat{\theta}_m$ is verified. As a consequence, for any positive Borel function f, $\mathbb{E}_\Delta^{K(m)}(f(U_{0:n}^{(m+1)})|\mathcal{F}_m) = \Psi^{K(m)}_{n;\widehat{\theta}_m} f$. Then $\sum_{\ell=1}^m a_\ell e_\ell$ is a martingale, bounded in L_2 under assumptions (M5) and (SAEM1)-(SAEM2).

To verify assertion 4, we use Lemma 1. Under assumptions (SMC2)-(SMC3) and assertion 1, Lemma 1 yields that for any $\varepsilon > 0$, there exist two constants C_1, C_2, independent of θ, such that

$$\begin{aligned}\sum_{m=1}^{M} \mathbb{P}\left(|r_m| > \varepsilon\right) &= \sum_{m=1}^{M} \mathbb{P}\left(\left|\Psi^{K(m)}_{n,\hat{\theta}(s_m)} S - \pi_{n,\hat{\theta}(s_m)} S\right| \geq \varepsilon\right) \\ &\leq C_1 \sum_{m=1}^{M} \exp\left(-K(m)\frac{\varepsilon^2}{C_2\|S\|^2}\right).\end{aligned}$$

Finally, assumptions (SMC1)-(SMC2) imply that there exists a constant C_3, independent of θ, such that

$$\sum_{m=1}^{M} \mathbb{P}\left(|r_m| > \varepsilon\right) \leq C_1 \sum_{m=1}^{M} \frac{1}{m^{C_3 g(m)\varepsilon^2}}$$

which is finite when M goes to ∞. This proves the a.s. convergence of r_m to 0.

□

D.3 Proof of Corollary 1

Proof. Theorem 6 of Delyon *et al.* (1999) can be extended without difficulty to our algorithm. It proves that under assumptions of Theorem 1 and (LOC1), the sequence $\widehat{\theta}_m$ converges to a fixed point of the EM-mapping $T(\widehat{\theta}_m) = \hat{\theta}(s(\widehat{\theta}_m))$. Assumptions (LOC2)-(LOC3), Lemma 3 of Delyon *et al.* (1999) and application of Brandière and Duflo (1996) imply that the sequence $\widehat{\theta}_m$ converges with probability 1 to a proper maximum of the likelihood. □

D.4 Proof of Theorem 2

Proof. The Markov property yields

$$\begin{aligned}|p(V_{0:n};\theta) - p_\delta(V_{0:n};\theta)| &\leq \int |p(V_{0:n},U_{0:n};\theta) - p_\delta(V_{0:n},U_{0:n};\theta)|\, dU_{0:n} \\ &\leq \int \left|\prod_{i=1}^{n} p\left(V_i,U_i|V_{i-1},U_{i-1};\theta\right) - \prod_{i=1}^{n} p_\delta\left(V_i,U_i|V_{i-1},U_{i-1};\theta\right)\right| dU_{0:n} \\ &\leq \int \sum_{i=1}^{n} \left|p\left(V_i,U_i|V_{i-1},U_{i-1};\theta\right) - p_\delta\left(V_i,U_i|V_{i-1},U_{i-1};\theta\right)\right| \\ &\quad \prod_{j=1}^{i-1} p\left(V_j,U_j|V_{j-1},U_{j-1};\theta\right) \prod_{j=i+1}^{n} p_\delta\left(V_j,U_j|V_{j-1},U_{j-1};\theta\right) dU_{0:n}\end{aligned}$$

Bally and Talay (1996a,b) provide that under assumption (H1), there exist constants $C_1 > 0$, $C_2 > 0$, $C_3 > 0$, $C_4 > 0$ independent of θ such that

$$\begin{aligned}|p_\delta(V_i,U_i|V_{i-1},U_{i-1};\theta) + p(V_i,U_i|V_{i-1},U_{i-1};\theta)| &\leq C_1 e^{-C_2\|(V_i,U_i)-(V_{i-1},U_{i-1})\|^2} \\ |p_\delta(V_i,U_i|V_{i-1},U_{i-1};\theta) - p(V_i,U_i|V_{i-1},U_{i-1};\theta)| &\leq \delta\, C_3 e^{-C_4\|(V_i,U_i)-(V_{i-1},U_{i-1})\|^2}\end{aligned}$$

We deduce that for all $i = 1, \ldots, n$, there exists a constant $C > 0$ independent of θ such that

$$\int |p\left(V_i, U_i | V_{i-1}, U_{i-1}; \theta\right) - p_\delta\left(V_i, U_i | V_{i-1}, U_{i-1}; \theta\right)| \prod_{j=1}^{i-1} p\left(V_j, U_j | V_{j-1}, U_{j-1}; \theta\right)$$

$$\times \prod_{j=i+1}^{n} p_\delta\left(V_j, U_j | V_{j-1}, U_{j-1}; \theta\right) dU_{0:n} \leq C\delta$$

Finally, we get $|p(V_{0:n}; \theta) - p_\delta(V_{0:n}; \theta)| \leq Cn\delta = C\frac{1}{L}n\Delta$. □

References

Bally, V. and Talay, D. (1996a). The law of the Euler scheme for stochastic differential equations (I): convergence rate of the distribution function. *Probability Theory and Related Fields*, **104**(1), 43–60.

Bally, V. and Talay, D. (1996b). The law of the Euler scheme for stochastic differential equations (II): convergence rate of the density. *Monte Carlo Methods Appl.*, **2**, 93–128.

Berg, R. W., Alaburda, A., and Hounsgaard, J. (2007). Balanced inhibition and excitation drive spike activity in spinal halfcenters. *Science*, **315**, 390–393.

Berg, R. W., Ditlevsen, S., and Hounsgaard, J. (2008). Intense synaptic activity enhances temporal resolution in spinal motoneurons. *PLoS ONE*, **3**, e3218.

Brandière, O. and Duflo, M. (1996). Les algorithmes stochastiques contournent-ils les pièges? *Ann. Inst. H. Poincaré Probab. Statist.*, **32**(3), 395–427.

Del Moral, P., Jacod, J., and Protter, P. (2001). The Monte-Carlo method for filtering with discrete-time observations. *Probab. Theory Related Fields*, **120**(3), 346–368.

Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, **27**, 94–128.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Jr. R. Stat. Soc. B*, **39**, 1–38.

Ditlevsen, S. and Greenwood, P. (2012). The Morris-Lecar neuron model embeds a leaky integrate-and-fire model. *To appear in J. Math. Biol.*, pages 1–21. DOI: 10.1007/s00285-012-0552-7.

Doucet, A., de Freitas, N., and Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, Stat. Eng. Inf. Sci., pages 3–14. Springer, New York.

Fearnhead, P., Papaspiliopoulos, O., and Roberts, G. (2008). Particle filters for partially observed diffusions. *J. R. Statist. Soc. B*, **70**(4), 755–777.

Hodgkin, A. and Huxley, A. (1952). A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes. *J. Physiol.*, **117**, 500–544.

Huys, Q. J. M. and Paninski, L. (2009). Smoothing of, and Parameter Estimation from, Noisy Biophysical Recordings. *PLOS Computational Biology*, **5**(5).

Huys, Q. J. M., Ahrens, M., and Paninski, L. (2006). Efficient estimation of detailed single-neuron models. *J Neurophysiol*, **96**(2), 872–890.

Jahn, P., Berg, R. W., Hounsgaard, J., and Ditlevsen, S. (2011). Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *Journal of Computational Neuroscience*, **31**, 563–579.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, **24**(2), 211–229.

Morris, C. and Lecar, H. (1981). Voltage oscillations in the barnacle giant muscle fiber. *Biophys. J.*, **35**, 193–213.

Tateno, T. and Pakdaman, K. (2004). Random dynamics of the morris-lecar neural model. *Chaos*, **14**(3), 511–530.